PREMIER REFERENCE SOURCE

Handbook of Computational Intelligence in Manufacturing and Production Management

DIPAK LAHA & PURNENDU MANDAL

Handbook of Computational Intelligence in Manufacturing and Production Management

Dipak Laha
Jadavpur University, India

Purnendu Mandal
Lamar University, USA

INFORMATION SCIENCE REFERENCE
Hershey • New York

Acquisitions Editor: Kristin Klinger
Development Editor: Kristin Roth
Senior Managing Editor: Jennifer Neidig
Managing Editor: Sara Reed
Copy Editor: Ashlee Kunkel
Typesetter: Lindsay Bergman
Cover Design: Lisa Tosheff
Printed at: Yurchak Printing Inc.

Published in the United States of America by
Information Science Reference (an imprint of IGI Global)
701 E. Chocolate Avenue, Suite 200
Hershey PA 17033
Tel: 717-533-8845
Fax: 717-533-8661
E-mail: cust@igi-global.com
Web site: http://www.igi-global.com

and in the United Kingdom by
Information Science Reference (an imprint of IGI Global)
3 Henrietta Street
Covent Garden
London WC2E 8LU
Tel: 44 20 7240 0856
Fax: 44 20 7379 0609
Web site: http://www.eurospanonline.com

Library of Congress Cataloging-in-Publication Data

Handbook of computational intelligence in manufacturing and production management / Dipak Laha & Purnendu Mandal, editors.
p. cm.
Summary: "This book focuses on new developments in computational intelligence in areas such as forecasting, scheduling, production planning, inventory control, and aggregate planning, among others. It provides cutting-edge knowledge on information technology developments for both researchers and professionals in fields such as operations and production management, Web engineering, artificial intelligence, and information resources management"--Provided by publisher.
Includes bibliographical references and index.
ISBN 978-1-59904-582-5 (hardcover) -- ISBN 978-1-59904-584-9 (ebook)
1. Production management--Technological innovations. 2. Computational intelligence. I. Laha, Dipak. II. Mandal, Purnendu.
TS155.6.H357 2007
658.5--dc22
2007024485

British Cataloguing in Publication Data
A Cataloguing in Publication record for this book is available from the British Library.

Table of Contents

Section III
Applications in Manufacturing and Production Management

Detailed Table of Contents

Section I
Computational Intelligence Methodologies

Earlier methods for solving manufacturing scheduling problems by classical optimization techniques such as linear programming and branch and bound reveal serious limitations. More advanced heuristics as well as various efficient optimization methods based on the evolutionary computing paradigm such as genetic algorithms, simulated annealing, and artificial immune system are required. This chapter briefly discusses the overview of these emerging heuristics and metaheuristics, and their applications to scheduling problems. The artificial immune system is discussed at length to add to the growing research interests.

This chapter focuses on the application of some artificial intelligence (AI) based random search algorithms, such as genetic algorithm (GA), ant colony optimization (ACO), simulated annealing (SA), artificial immune system (AIS), and tabu search (TS) in solving machine loading problem in flexible manufacturing systems. The objectives of the chapter are to make readers aware of intricate solutions that might exist in machine loading problem of FMS, and to provide examples of generic procedure for various AI based random search algorithms. The other objective is to describe the step-wise implementation of search algorithms over machine loading problem.

Production and operations management requires specific financial tools in order to carry out production planning, costing, investment appraisal, and other functions. This chapter focuses on information technology automation of financial functions adopted by production departments for forecasting production needs, production planning and control, profit volume analysis, cost analysis, and investment appraisal analysis. An attempt is made to classify various quantitative and qualitative techniques in relation to various financial aspects. Specifically, advances of neural networks, expert systems, advanced statistical analysis, operational research methods, and various hybrid techniques are presented in relation to financial considerations. A strategic alignment model is proposed for adoption of financial applications in businesses.

The decision process needs a systematic approach to structure the system requirements and highlight the management preferences while considering uncertain conditions. The analytical hierarchical process (AHP) could be employed for structuring the criteria influencing the process choice. An integrated fuzzy AHP model is proposed in this chapter and the model is analyzed within the boundary conditions of the fuzzy criteria using the Expert Choice software. The proposed model is generic in structure and is applicable to many firms.

In this chapter, an attempt has been made to determine an optimal solution of a deterministic production-inventory model that consists of single deteriorating items and a constant rate of deterioration. The model considers the lead time to be negligible and the demand rate is a ramp type function of time. Shortages are allowed and partially backlogged. During this shortage period, the backlogging rate is a variable which depends on the length of the waiting time over the replenishment period. Mathematical formulation of the problem highlighted the model as a complex nonlinear constrained optimization problem. Considering the complexities towards solution, modified real-coded genetic algorithms (elitist MRCGA) with ranking selection, whole arithmetic crossover, and nonuniform mutation on the age of the population has been developed. The proposed production-inventory model has been solved via MRCGA and simulated annealing and as well as standard optimization methods. Finally, the results are embedded with numerical example and sensitivity analysis of the optimal solution with respect to the different parameters of the system is carried out.

This chapter focuses on condition monitoring techniques in manufacturing systems. Two aspects of condition monitoring process are considered: feature extraction and condition classification. Feature extraction methods described and implemented are fractals, kurtosis and Mel-frequency cepstral coefficients. Classification methods described and implemented are support vector machines (SVM), hidden Markov models (HMM), Gaussian mixture models (GMM), and extension neural networks (ENN). The effectiveness of these features are tested using SVM, HMM, GMM, and ENN on condition monitoring of bearings and are found to give good results.

Demand forecasting of short life span products involves unique issues and challenges that cannot be fully tackled in existing software systems. SIMForecaster (a forecasting system developed at the Singapore Institute of Manufacturing Technology, Singapore) has successfully been used for many important forecasting problems in industry. This chapter identifies specific soft computing techniques, namely small world theory, memes theory, neural networks (with special structures such as binary neural networks [BNNs], bidirectional segmented memory [BSM] recurrent neural networks, and long-short-term-memory [LSTM] networks) for solving forecasting problems. It is suggested that, in addition to these neural network techniques, integrated demand forecasting systems for handling optimization problems involved in short life span products would also need some techniques in evolutionary computing as well as genetic algorithms.

This chapter provides several examples to illustrate how data mining, a key area of computational intelligence, offers a great promise to manufacturing companies. It also covers a brief overview of data warehousing as a strategic resource for quality improvement and as a major enabler for data mining applications. Although data mining has been used extensively in several industries, in manufacturing its use is more limited and new. The examples published in the literature of using data mining in manufacturing promise a bright future for a broader expansion of data mining and business intelligence in general into manufacturing.

Optimization and search methods can assist the designer at all stages of the design process. The past decade has seen a rapid growth of interest in stochastic search algorithms, particularly those inspired

This chapter examines the application of system archetypes as a systems development methodology to create simulation models. The application of system archetypes to the strategic business analysis of a healthcare system reveals that it is possible to identify the lacuna in management's strategic thinking processes. In the research study, hospital executives found that policy modification with slight variable changes helps to avoid pitfalls in systems thinking and avoid potentially cost prohibitive learning had these policies been implemented in real life.

The provision of timely, accurate, relevant, and concise information for managerial decision making has traditionally represented a challenge to information systems designers. The mass adoption of enterprise resource planning (ERP) systems has multiplied the amount of data being recorded about the movement of inventory in the supply chain. However, this online information requires much off-line manipulation in order for it to be meaningful to managers. In addition, this data are based on physical structures and business models that evolve over time, and thus inevitably a gap opens between the virtual enterprise and reality. Despite the benefits of inventory visibility and expenditure control afforded by ERP systems, managers still require data from other, nonintegrated systems. In this chapter the authors present their research on decision-making support in two manufacturing organizations, with the objective of understanding how these integrated applications support the manager in achieving goals.

This chapter focuses on hardware/software, Web-based technologies, and managerial policy options in supporting e-business. More and more, internal applications are being moved from legacy systems into a more flexible Web-based environment. The issue concerning World Wide Web technologies is important to today's businesses. Decision making in this area is complex and needs to consider carefully the characteristics and needs of the entities employing these technologies. The research presented here compares performances and costs of technologies used to serve dynamic Web content.

This chapter elaborates the basic concepts underlying the development of Web-based decision support systems (DSS). The chapter introduces a Web-based decision support system for water resources management on a basin scale and also some evolving concepts like mobile agent technology to meet the

In CAD/CAM, reverse engineering involves obtaining a CAD model from an object that already exists. An exact replica can then be produced, or modifications can be made before manufacture. Single-perspective triangulation sensors provide an inexpensive method for data acquisition. However, such sensors are subject to localized distortions caused by secondary reflections or occlusion of the returning beam, depending on the orientation of the sensor relative to the object. This chapter describes an investigation into integrating optical camera data to improve the scanning process and reduce such effects, and intelligent algorithms, based on image analysis, which identify the problem regions, so that the sensor path and orientation can be planned before the scan, thereby reducing distortions.

This chapter illustrates the use of data mining as a computational intelligence methodology for forecasting data management needs. Specifically, this chapter discusses the use of data mining with multidimensional databases for human lung cancer and forestry. The data mining is performed using four selected software of SAS® Enterprise Miner™, Megaputer PolyAnalyst® 5.0, NeuralWare Predict®, and BioDiscovery GeneSight®. The tools and techniques discussed in this chapter can be representative of those applicable in a typical manufacturing and production environment.

This chapter discusses the background of supply chain planning and execution systems, their role in an organization, and how they are aiding in collaboration. Studies show that organizations are finding creative ways to mitigate supply chain costs while maintaining operational efficiency. New approaches, technologies, and methodologies are aiding with these cost-cutting measures to drastically reduce supply chain costs and increase customer satisfaction. The chapter presents a case study on how a supply chain management system could help an organization be more effective.

This chapter focuses on the understanding of manufacturing environment and policies at a national level. In this chapter two modeling approaches are discussed for understanding the intertwined relationships among factors which influence the performance and competitiveness of manufacturing: the system dynamics approach and the quantitative survey approach. The system dynamics approach is used to

Foreword

Artificial intelligence (AI) is simply a way of providing a computer or a machine to think intelligently like human beings. Since human intelligence is a complex abstraction, scientists have only recently began to understand and make certain assumptions on how people think and to apply these assumptions in order to design AI programs. It is a vast knowledge base discipline that covers reasoning, machine learning, planning, intelligent search, and perception building.

Traditional AI had the limitations to meet the increasing demand of search, optimization, and machine learning in the areas of large, biological, and commercial database information systems and management of factory automation for different industries such as power, automobile, aerospace, and chemical plants. The drawbacks of classical AI became more pronounced due to successive failures of the decade long Japanese project on fifth generation computing machines. The limitation of traditional AI gave rise to development of new computational methods in various applications of engineering and management problems. As a result, these computational techniques emerged as a new discipline called computational intelligence (CI).

Computational intelligence terminology was originated by Professor Lotif A. Zadeh. Since its inception in early 1990s, the topic has changed to a great extent concerning its content and applications. Earlier it was concerned with the fuzzy sets, neural networks. and genetic algorithms. Now, it consists of granular computing, neural computing, and evolutionary computing along with their interactions with artificial life, chaos theory. and others. Evolutionary computational technique includes genetic algorithms, evolutionary programming, and evolutionary strategies and genetic programming. Artificial neural networks mimic the biological information system. Evolutionary computing algorithms are used for optimization problems, and fuzzy logic as a basis for representing imprecise knowledge.

Computational intelligence tools have attracted the growing interest of researchers, scientists, engineers, and managers in a number of practical applications. These applications include engineering, business, and banking. It has emerged as a relatively new field of research and has been finding more and more applications in various areas. Fuzzy set theory is more useful for reasoning with imprecise data and knowledge. Neural networks are more applicable in machine learning, whereas genetic algorithms are most suitable for the areas of search and optimization but it is not so successful in handling real time problems.

The applications of CI are diverse, including medical diagnosis, data mining, design and manufacturing, production planning and scheduling systems, robots working in hazardous environments, autonomous vehicles, image matching, and control systems, just to mention a few for the service of mankind.

There are several advantages of CI over traditional approaches. These include conceptual simplicity, broad domains of applications, better performance than classical methods on real life problems, use of knowledge management and hybridization with other methods, parallelism, and capability to solve dynamic problems.

A lot of innovation has been noticed in manufacturing and production management in recent years, becoming a very important area in business today. Production management is an interesting mixture of managing people, sophisticated technology, and the applications of computational intelligence. The handbook addresses the latest and most important issues related to production management. This handbook primarily serves as one comprehensive source of information where business managers, professors, and researchers can look for disseminate technology and ideas, and gain knowledge through a variety of research topics including theoretical, experimental, and case studies. It focuses on applications of new developments of computational intelligence tools such as artificial neural networks, genetic algorithms, and artificial immune system and swarm optimization methods to various areas of management.

The present exploration on manufacturing and production management is thoroughly edited and reviewed for which it has become a "hallmark" for the user/readers to pave the way for better managerial perspective. I am inclined to believe that the topics discussed by professors, researchers, and professional managers of international repute would be globally useful for the purpose they have been written.

Angappa "Guna" Gunasekaran, PhD
University of Massachusetts, Dartmouth

Preface

Experts now believe that world-class performances by organizations in providing high-quality cost-competitive products and services are essential for survival in today's business environment. Organizations need to attain a competitive advantage which could be achieved through effective integration of technology strategy with business strategy (Sohal, Ramsay, & Samson, 1992; Sohal, Samson, & Weill, 1991).

Information technology has significantly changed companys' business strategy (Black & Lynch, 2001, 2004**)**. During the last two decades manufacturing and information technology has forced great changes in the ways businesses manage their operations in meeting the desired cost and quality of products and services, customer demands, competition, and other challenging situations. While the 19th century gave birth to the Industrial Revolution, the 20th century saw a new kind of revolution in the Information Technology Revolution. The Information Revolution deals with the development of technologies that allow quicker and cheaper transmission of data and images, and storage and retrieval of information. Integration of resources and business units has become more effective than ever primarily due to the development of enterprise wide information systems, the Internet, and Web-based information systems. Production and operations organizations have been the forerunners in the implementation of such information systems.

There are essentially two types of technologies in manufacturing and operational organizations: core and enabling technologies. The core technology is that technology that provides leverage to the organization to fulfill its mission and grow (Laugen, Acur, Boer, & Frick, 2005). For example, Toyota's core competency is its manufacturing technology, Cannon's is its printer motor technology, British Aerospace's is wing technology, while Boeing believes its core competency to be systems integration technology. On the other hand, enabling technologies are those that facilitate or assist the core technology in doing what it does best. An example of such technologies is information technologies that run the Toyota assembly line and call center specialists who assure that problems with information technology can be mitigated. Information technologies in manufacturing companies offer both operational and strategic benefits. The strategic benefit of IT includes enhanced competitive position, improved strategic flexibility, and facilitating manufacturing globalization.

During the past decade, the role of IT in production management changed from the back-office supporting tools to a strategic role. Strategic information systems (SIS) now play a critical role in helping organizations to increase production efficiency, and to be more effective and competitive. As the business environment is changing fast, the need for newer and more effective IT/IS is arising. In fact, there has been a constant demand on IT professionals for improved methodologies, design, and applications. Accordingly, the researchers are responding to this demand through computational intelligence, particularly focusing on neural networks (Haykin, 1994; Wang & Takefuji, 1993), genetic algorithms (Davis, 1987; Deb, 2001; Goldberg, 1989), evolutionary programming (Diego & Duc Truong, 2007; Konar &

Jain, 2001), artificial immune systems (Dasgupta, 1980; De Castro & Timmis, 2002), and fuzzy systems (Zadeh, 1965; Zimmermann, 1999).

IT/IS have tremendous impact on the productivity in both manufacturing and service organizations (Roth, 1996). Companies have implemented systems such as enterprise resource planning (ERP), MRP, EDI, and so forth over time for improving their productivity. The Internet has created a brand-new outlet from which firms can market and sell their goods and services. The enormous amount of information that is now available to consumers on the Internet is mind-boggling. Improvements in the Internet and communication technologies have led to increased globalization of businesses.

Effective production management is the key to business success. Undoubtedly, newer information technologies have and will have growing influence in future of production and operations field. This handbook focuses on new developments in computational intelligence in areas such as forecasting, scheduling, production planning, inventory control, and so forth. It offers a great theoretical challenge for researchers and, from practical point of view, plays a significant role in the successful operation of different fields of production management. The application of various tools, as described in the handbook, will lead to a rapid turn-around of jobs and minimization of in-process inventory, and thereby minimizing the overall cost of production. The handbook incorporates newer efficient optimization methods that have emerged recently, based on the evolutionary computing paradigms such as genetic algorithms, neural networks, simulated annealing (Aarts & Korst, 1989; Van Laarhoven & Aarts, 1987), artificial immune systems, ant-colony algorithm (Dorigo, Caro, & Gambardella, 1997), and swarm intelligence (Kennedy & Eberhart, 2001). These tools are currently being utilized for developing efficient methodologies for different engineering and management problems.

There is yet another reason for compiling this handbook: minimizing the conceptual gap of unbalanced view of IT between IT researchers and production professionals. In spite of numerous developments in methodological areas, IT professionals are very little aware of production technologies. Following the same logic, production management professionals are not fully aware of IT related developments. This handbook primarily serves as a single source where IT researchers and production professionals can look for technologies and ideas, and knowledge through a variety of research methods including theoretical, experimental, and case studies. The handbook introduces researchers to many computing methodologies applicable in both services and manufacturing sectors. It addresses new developments in the field of production management and new information related to software, while remaining a strong focus on the fundamental concepts.

Production management and the use of information technology have both been extensively researched over recent years. There is no comprehensive study of the extent of use of information technology in production and operations management area. Most of the studies reported in the production management area have been too specific in the conventional areas such as inventory control, project management, scheduling, and so forth. New research areas have emerged due to the development of computational intelligence tools. The managerial practices have seen a direction of new development of Internet, World Wide Web, network based computing, data sharing, and data mining. In contrast to other books, this book will focus on the integration between IT and production systems, with emphasis on the applicability to real-life problems.

ORGANIZATION OF THE BOOK

The handbook is organized into three sections: Section I: Computational Intelligence Methodologies; Section II: Supply Chain and Decision Support Systems; and Section III: Applications in Manufacturing and Production Management. The book contains 23 chapters contributed by leading experts from various parts of the world.

A brief description of each of the chapters follows.

Chapter I discusses the present challenges on developing heuristics and metaheuristics for scheduling problems. Manufacturing scheduling offers a great theoretical challenge to researchers. Traditionally researchers emphasized on classical optimization methods such as linear programming and branch and bound method to solve scheduling problems. However, these methods have the limitation of tackling small-sized scheduling problems because of the consumption of high CPU time. As a result, heuristics, as well as various efficient optimization methods based on the evolutionary computing paradigms such as genetic algorithms, simulated annealing, and artificial immune systems, have been applied to scheduling problems for obtaining near optimal solutions. These computational tools are currently being utilized successfully in various engineering and management fields. The chapter briefly discusses the overview of these emerging heuristics and metaheuristics and their applications to the scheduling problems. Given the rise in attention by the researchers, more emphasis has been given to explore artificial immune systems in details.

Chapter 2 deals with the application of some artificial intelligence based random search algorithms like genetic algorithms, ant colony optimization, simulated annealing, artificial immune system, and tabu search to machine loading problems in flexible manufacturing system. Comparative performance evaluations of these techniques with the best existing heuristics based on standard benchmark dataset have been presented in this chapter.

Chapter III focuses on financial tools required in production management settings. Production and operation management requires specific financial tools in order to accomplish the functions of production planning, costing, investment appraisal, and so forth. Computational intelligence in those financial functions is needed for production forecasting, production planning and control, profit volume analysis, cost analysis, investment appraisal, and analysis. The chapter discusses advances of neural networks, expert systems, advanced statistical analysis and operational research methods, and various hybrid techniques. A strategic alignment model is derived for the adoption of financial applications in businesses.

Chapter IV investigates the decision process of manufacturing systems under uncertain conditions. The decision process needs a systematic approach to structure the system requirements and highlight the management preferences while considering vague criteria. In order to establish a suitable empirical approach for the decision process compatible with the current/future requirements, the analytical hierarchical process (AHP) is employed for structuring the criteria influencing the process choice. The application of the proposed AHP model for the selection of manufacturing process is demonstrated using numerical examples. In addition, due to dealing with vague data in the decision process, the uncertain criteria are characterized by typical fuzzy sets. The integrated fuzzy AHP is then analyzed within the boundary conditions of the fuzzy criteria using the Expert Choice software. The proposed model is intended to be generic in structure and applicable to many firms.

Within the constraints of certain shortages and backlogs in a deterministic production-inventory control model, Chapter V presents some mathematical models highlighting the complex nonlinearity constrained optimization problem with a view to achieving optimal solutions using modified real-coded genetic algorithms and simulated annealing. Some numerical examples and sensitivity analysis have been included towards achieving such optimal solutions.

Chapter VI addresses the different condition monitoring techniques using computational intelligence. The effectiveness of different aspects of condition monitoring of bearings has been tested using different techniques such as neural networks, thereby producing good results.

Chapter VII addresses the issues, challenges, and problems of demand forecasting of short lifespan products. Due to the limitation of SIMForecaster, the existing forecasting system, the authors identify some soft computing techniques for solving these problems. They also suggest the importance of evolutionary computing techniques including genetic algorithms in the context of integrated demand forecasting system.

Chapter VIII addresses the issue of data mining process and its application to manufacturing. The author suggests by illustrating some examples that data mining as a computational intelligence approach offers a great promise to manufacturing companies. He also feels that although it has been widely used in different industries, its use is limited and new to manufacturing. He also believes that data mining will occupy a mainstream application in manufacturing, thereby enhancing the capabilities in the organization.

Chapter IX presents an overview of evolutionary computing application for engineering design. An optimal design may be defined as the one that most economically meets its performance requirements. Optimization and search methods can assist the designer at all stages of the design process. The past decade has seen a rapid growth of interest in stochastic search algorithms, particularly those inspired by natural processes in physics and biology. Impressive results have been demonstrated on complex practical optimization of several schools of evolutionary computation. Evolutionary computing, unlike conventional technique, had the robustness for producing s variety of optimal solutions in a single simulation run, giving wider options for engineering design practitioners to choose from. Despite limitations, the act of finding the optimal solution for optimization problems has shown a substantial improvement in terms of reducing optimization process time and cost as well as increasing accuracy.

Chapter X presents some methodologies to capture the dynamics of supply chain, detect the changes, and thereby predict the behavior on these changes and finally define the needed modification to mitigate the unwanted behaviors and performance. The authors describe these methodologies through the integration of system dynamics, neural networks, eigen value analysis, and sensitivity analysis tools that contribute to the advancement of prediction and mitigation of undesirable supply chain behavior within short- and long-term horizons. Finally, a case study has been briefly summarized in this context.

In Chapter XI, a decision support system is proposed to analyze projects with respect to market, technicalities, and social and environmental impact in an integrated framework using analytic hierarchy process, a multiple attribute decision making technique. This not only reduces duration of project evaluation and selection, but also helps select an optimal project for the organization for sustainable development. The entire methodology has been applied to a cross-country oil pipeline project in India and its effectiveness has been demonstrated.

Chapter XII addresses the issues relating to the modeling and analysis of dynamic supply networks. The author uses the combination of network structure modeling and simulation of dynamic behavior to enhance the performance analysis of supply networks.

Chapter XIII presents a case study where system archetypes are applied to create simulation models in healthcare with a view to identifying the loop holes in management strategic thinking processes and defying these fallacies during implementation.

Chapter XIV addresses the problems and challenges of a manufacturing integrated information system for managerial decision making. The authors present their research work on decision support systems in two manufacturing organizations where ERP have been implemented successfully with a view to facilitating manager's role in bridging the gap between the ERP system in supply chain and the real-world business organization.

Chapter XV discusses the issue concerning the importance of wide Web technologies in today's business, which is playing an increasing role in the communication of people. The author compares different Web technologies to decide their best implementation with respect to performance and cost. The authors claim that a broader scope approach due to continuing developments in Web technologies is suggested for comparative analysis.

Chapter XVI elaborates the key concepts and technical issues concerning the development of Web-based decision support systems (DSS). The Web-based DSS enhances communication and decision-making capability in a distributed environment or a multiple stakeholder process. The authors present the application of Web-based DSS to water resources management on a basin scale. The authors hope that better understanding of these concepts of Web-based DSS will bring together participants like analysts, modelers, and the end users.

Chapter XVII discusses different independent component analysis (ICA) algorithms and their application to manufacturing problems. Since it was difficult to satisfy the complexity of prediction of spatial data on mineral resources and remote sensing imagery by the conventional methods, the ICA method has paved the way for futuristic research in spite of having its some limitations and disadvantages.

Chapter XVIII describes the methodology of a biologically inspired swarm intelligence technique and its application to some production management problems such as vehicle routing and motion planning of mobile robots. Computer simulation for these problems has been included.

Chapter XIX identifies the existing challenges in the application of optimization techniques for any metal cutting-based manufacturing unit. The authors review the scope and status of artificial neural networks and metaheuristic strategies in metal-cutting process. Subsequently, a solution methodology based on these tools has been proposed. Finally, the authors present a case study in a multiple response grinding process optimization problem using these tools.

Chapter XX describes an investigation of an integrated approach combining optical camera data and intelligent algorithms to overcome the limitation of single-perspective triangular sensors for laser scanning of 3D surfaces.

Chapter XXI describes the uses of data mining for forecasting data management needs for the selected biotechnology data of forest cover data and human lung cancer data set. Four data mining software have been used to obtain enhanced intelligent capabilities for biotechnology research. The proposed tools and techniques can be utilized in a typical manufacturing and production environment.

Chapter XXII addresses the importance of a networked supply chain model, which is the combination of Web and supply chain management technology. As a result, supply chain costs will be reduced along with the increase in customer satisfaction. Finally, the authors present a case study on supply chain management enhancing the effectiveness of the organization.

Chapter XXIII discusses system a dynamics modeling approach and a quantitative survey approach to model interactions in manufacturing systems. Modeling is a great tool to analyze long-term consequences of policy options in manufacturing. Models could be used for understanding the intertwined relationships among factors which influence the performance and competitiveness of manufacturing. The system dynamics approach is used to develop a conceptual model of the strategic issues that influence the performance and competitiveness of manufacturing, and the results of a quantitative survey are used to understand the actual extent of the influences of various factors in the current situation.

REFERENCES

Aarts, E.H.L., & Korst, J.H.M. (1989). *Simulated annealing and boltzman machines*. Chichester: Wiley.

Black, S.E., & Lynch, L.M. (2001). How to compete: The impact of workplace practices and information technology on productivity. *The Review of Economics and Statistics*, *83*(1), 1-28.

Black, S.E., & Lynch, L.M. (2004). What's driving the new economy? The benefits of workplace innovation. *The Economic Journal*, *114*(493).

Dasgupta, D. (1980). *Artificial immune system and their applications.* Springer Publication.

Davis, L. (1987). *Genetic algorithm and simulated annealing.* London: Pitman.

De Castro, L.N., & Timmis, J. (2002). *Artificial immune systems: A new computational intelligence approach*. London: Springer Publication.

Deb, K. (2001). *Multi-objective optimization using evolutionary algorithms.* John Wiley & Sons Publication.

Diego, A., & Duc. Truong, P. (2007). *Computational intelligence for engineering and manufacturing.* Springer Publication.

Dorigo, M., Caro, G.D., & Gambardella, L.M. (1997). Ant colony system: A cooperative learning approach to the travelling salesman problem. *IEEE Transactions on Evolutionary Computation*, *1*(1), 53-66.

Goldberg, D.E. (1989). *Genetic algorithms in search optimization and machine learning.* Reading, MA: Addison-Wesley.

Haykin, S. (1994). *Neural networks: A comprehensive foundation.* New York: Macmillan Publishing.

Konar, A., & Jain, L.C. (2001). *An introduction to computational intelligence paradigms.* Kluwer Academic Publishers.

Kennedy, J., & Eberhart, R.C. (2001). *Swarm intelligence.* Morgan Koufman Academic Press.

Laugen, B.T., Acur, N., Boer, H., & Frick, J. (2005). Best manufacturing practices: What do the best-performing companies do? *International Journal of Operations and Production Management*, *25*(2), 131-150.

Roth, A.V. (1996). Neo-operations strategy. In G. H. Gaynor (Ed.), *Handbook of technology management*. New York: McGraw-Hill.

Sohal, A.S., Ramsay, L., & Samson, D. (1992). Quality management practices in Australian industry. *Total Quality Management, 3*(3), 283-299.

Sohal, A.S., Samson, D., & Weill, P. (1991). Manufacturing and technology strategy: A survey of planning for AMT. *Computer-Integrated Manufacturing Systems*, *4*(2), 71-79.

Van Laarhoven, P.J.M., & Aarts, E.H.L. (1987). *Simulated annealing: Theory and applications*. Dordrecht, The Netherlands: Reidel.

Wang, J., & Takefuji, Y. (1993). *Neural networks in design and manufacturing.* World Scientific.

Zadeh, L.A. (1965). Fuzzy sets. *Information and Control, 8*(3), 338-353.

Zimmermann, H. J. (1999). *Practical application of fuzzy technologies.* Berlin: Springer Publication.

Acknowledgment

The editors would like to acknowledge the help of all involved in the collation and review process of the handbook, without whose support the project could not have been satisfactorily completed.

Most of the authors of chapters included in this handbook also served as referees for chapters written by other authors. Thanks go to all those who provided constructive and comprehensive reviews. Support of the Information Systems and Analysis Department at Lamar University and Mechanical Engineering Department at Jadavpur University is acknowledged for archival server space in the completely virtual online review process.

Special thanks also go to the publishing team at IGI Global, whose contributions throughout the whole process from inception of the initial idea to final publication have been invaluable. In particular to Kristin Roth, who continuously prodded via e-mail for keeping the project on schedule.

In closing, we wish to thank all of the authors for their insights and excellent contributions to this handbook.

Dipak Laha, PhD
Kolkata, India

Purnendu Mandal, PhD
Beaumont, Texas

July 2007

List of Reviewers

Simul Banerjee, Jadavpur University, India

Pankaj Barua, Assam Agricultural University, India

Ardhendu Bhattacharyya, Jadavpur University, India

Fergal Carton, University College Cork, Ireland

Narendra S. Chaudhari, Nanyang Technological University, Singapore

Petr Fiala, University of Economics Prague, Czech Republic

Dipak Laha, Jadavpur University, India

Purnendu Mandal, Lamar University, USA

Tshilidzi Marwala, University of Witwatersrand, South Africa

Carl Montano, Lamar University, USA

Jose D. Montero, Utah State University, USA

Reinaldo Moraga, Northern Illinois University, USA

Pradip Kumar Ray, Indian Institute of Technology Kharagpur, India

Mohan Rao, University of Texas Pan American, USA

Supriyo Roy, Management Institute of Durgapur, India

Joko Saputro, Lamar University, USA

Ravi Shankar, Indian Institute of Technology New Delhi, India

Sundar Srinivasan, University of Missouri-Rolla, USA

Petros Theodorou, Athens University of Economics and Business, Greece

Xian-Chuan Yu, Beijing Normal University, China

Section I
Computational Intelligence Methodologies

Chapter I
Heuristics and Metaheuristics for Solving Scheduling Problems

Dipak Laha
Jadavpur University, India

ABSTRACT

Manufacturing scheduling plays a very important function in successful operation of the production planning and control department of an organization. It also offers a great theoretical challenge to the researchers because of its combinatorial nature. Earlier, researchers emphasized classical optimization methods such as linear programming and branch-and-bound method to solve scheduling problems. However, these methods have the limitation of tackling only small-sized scheduling problems because of the consumption of high computational (CPU) time. As a result, heuristics as well as various efficient optimization methods based on the evolutionary computing paradigm such as genetic algorithms, simulated annealing, and artificial immune system have been applied to scheduling problems for obtaining near optimal solutions. These computational tools are currently being utilized successfully in various engineering and management fields. We briefly discuss the overview of these emerging heuristics and metaheuristics and their applications to the scheduling problems. Given the rise in attention by the researchers, more emphasis has been given to explore artificial immune system in details.

INTRODUCTION

Scheduling is concerned with the assignment of time to a set of jobs for processing through a group of machines (or their service sector equivalents) in order to best satisfy some criteria. A great deal of research has been carried out and will continue to be done on manufacturing scheduling problems (Baker, 1974). The reason is that scheduling offers a great theoretical challenge for researchers because of its combinatorial nature. Also, from the practical point of view, it plays a significant role in the successful operation of production, planning, and control department.

The general flowshop scheduling problem is known to be nondeterministic polynomial (NP)-complete (Gonzalez & Sahni, 1978). For solving scheduling problems, simple exact analytical methods such as integer programming (Sriker & Ghosh, 1986) or branch-and-bound (Lomnicki, 1965) have the limitation of dealing with only small-sized problems because of large computational effort. Heuristic polynomial-time algorithms (Campbell Dudek, & Smith, 1970; Johnson, 1954; Nawaz, Enscore, & Ham, 1983) probably are the most suitable means to solve large scheduling problems that are frequently encountered in many real-world situations. In general, heuristics provide good satisfactory (but not necessarily optimal) solutions in reasonable time and use problem-specific information.

The problems of manufacturing scheduling (Sarin & Lefoka, 1993) may be segregated based on (1) requirements, (2) complexity of the processes, and (3) scheduling objectives. Requirements may be produced either by open shop (customer orders) or closed shop (inventory replenishment). The complexity of the processes is primarily determined by the order in which the different machines appear in the operations of individual jobs. Broadly, manufacturing scheduling can be classified as flowshop scheduling and jobshop scheduling. In flowshop scheduling, it is generally assumed that all jobs must be processed on all machines in the same technological or machine order. In jobshop scheduling, the jobs may be processed following different machine orders. There is no common path of movement of jobs from machine to machine. Each machine is likely to appear for processing each operation of each job. The scheduling objectives are evaluated to determine the optimum schedule of jobs. Some of the objectives include makespan, total flow time, average job tardiness, and number of tardy jobs.

A variety of scheduling problems has been developed over the past years to address different production systems. The two commonly scheduling problems found in the scheduling literature of the past 50 years are flowshop scheduling and jobshop scheduling. Scheduling problems may be deterministic/stochastic and static/dynamic (Simons, 1992). The problem is deterministic or stochastic when the time required to process a task over respective machine takes a fixed or a random value. The scheduling problem is considered as static if ordering of jobs on each machine is determined once and will remain unchanged as opposed to the dynamic case that can accommodate changes of job ordering for accessing new jobs to the system.

A four-parameter notation (Conway, Maxwell, & Miller, 1967) is generally used to identify the individual scheduling problems, written as α / β / γ / δ.

α denotes the job-arrival process. For dynamic problems, α will denote the probability distribution of the times between arrivals. For static problems, it is assumed that they arrive simultaneously unless stated otherwise.

β describes the number of machines (m) used in the scheduling problem.

γ refers to the flow pattern of jobs through machines in the shop. The principal symbols are F for flowshop scheduling, R for randomly routed jobshop problem, and G for completely general or arbitrary flow pattern of jobs.

δ describes the criterion by which a schedule of jobs will be determined. The symbols to represent the scheduling criterion are F_{max} (minimize the maximum flow-time or makespan). As an example of this notation, Johnson's (1954) problem is described as n / 2/ F/ F_{max} which means flowshop scheduling with n jobs and 2 machines so as to minimize the maximum flow time or makespan. Similarly, for a generalized flowshop problem, the notation will be n/ m/ F/ F_{max}.

OBJECTIVES OF SCHEDULING

Most scheduling research has considered optimizing a single objective. The different performance measures or objectives include makespan, total flowtime, and job tardiness. Makespan of a schedule of jobs is the completion time of the last job in that schedule (it is assumed that the schedule starts at zero time). The total flow time of a schedule of jobs is the sum of completion times of all jobs in that schedule. Job tardiness indicates the lateness of the job with respect to its due date. Minimization of makespan results in maximization of overall resource utilization, whereas total flow time aims at minimizing work-in-process inventory and minimum tardiness yields minimum penalty.

A number of assumptions for flowshop or jobshop scheduling are considered. They are primarily considered for simplicity of the structure of the problems, but at the same time they help to build the generalized model. Most of the different applications using these models require relaxing one or several of these assumptions, so that they are not entirely realistic models for the applications. Some of the assumptions include availability of jobs, noninterference of machines, nonpassing of jobs, and so forth. Dudek and Teuton (1964) provide a complete list of these assumptions in their paper.

FORMULATION OF FLOWSHOP AND JOBSHOP SCHEDULING PROBLEM

In the flowshop scheduling problem, *n* jobs are to be processed on m machines. The order of the machines is fixed. We assume that a machine processes one job at a time and a job is processed on one machine at a time without preemption. Let $t_p(i, j)$ denote the processing time of job *j* on machine *i*, and $t_c(i, j)$ denote the completion time of job *j* on machine *i*. Let J_j denote the *j*-th job and M_i be the *i*-th machine. The completion times of the jobs are obtained as follows:

For $i = 1,2, \dots, m$ and $j = 1, 2, \dots, n$

$t_c(M_1, J_1) = t_p(M_1, J_1)$

$t_c(M_i, J_1) = t_c(M_{i-1}, J_1)+t_p(M_i,J_1)$

$t_c(M_1, J_j) = t_c(M_1, J_{j-1})+t_p(M_1,J_j)$

$t_c(M_i, J_j) = \max\{t_c(M_{i-1}, J_j), t_c(M_i,J_{j-1})\}+t_p(M_i, J_j)$

Total flowtime *is* defined as the sum of completion time of all jobs in a schedule, that is, total flow time is given by $\sum_{j=1}^{n} t_c(M_m, J_j)$. Makespan of a schedule of jobs is represented as $t_c\ (M_m, J_n)$.

Similarly, in the jobshop scheduling problem, let there be a set X of n jobs, a set Y of m machines, and a set Z of N operations. Each job has a sequence of operations which are to be performed in an uninterrupted manner on a set of machines. A schedule is laying out the operations of each job in time order on respective machines.

The problem can be stated as:

Minimize $\max_{j\in Z} \quad s_j + p_j$

Subject to:

$s_j \geq 0$ for all $j \in Z$

$s_k - s_j \geq p_j$ if j precedes k; $j, k \in Z$

and $s_k - s_j \geq p_j$ or $s_j - s_k \geq p_k$ if $m_j = m_k$

where, j, k = operations belonging to the set Z

s_j = start time of operation j

p_j = processing time of operation j

m_j = machine on which operation j is processed

In order to illustrate the flowshop scheduling problem, let us consider the following example:

Let there be two jobs each of which to be processed on three machines M_1, M_2, and M_3 in the order M_1, M_2, M_3. The workflow diagram in a pure flowshop is shown in Figure 1. The processing time matrix (in minutes) is given in Table 1.

Now, for the above two jobs and three machines flowshop problem, the schedule of jobs can be either $J_1 - J_2$ or $J_2 - J_1$.

Table 1. Processing time matrix

Job	Machine		
	M_1	M_2	M_3
J_1	2	6	5
J_2	4	3	1

Figure 1: Workflow in a pure flowshop

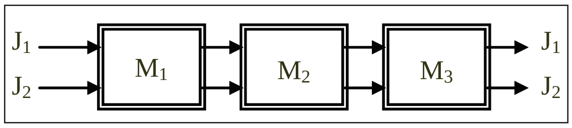
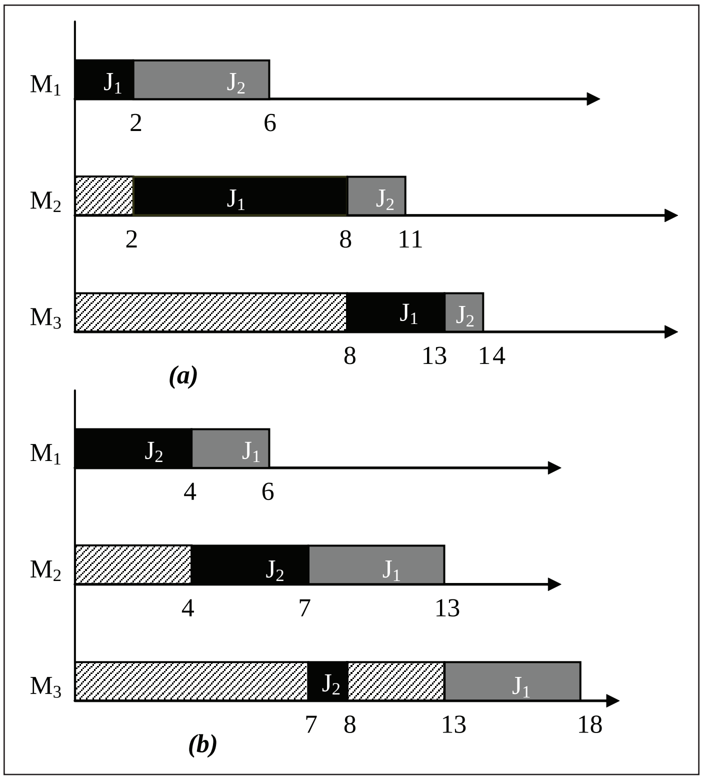


Figure 2. Gantt chart for schedule J1–J2 in (*a*) *and schedule J2–J1 in* (*b*) *in a flowshop*



The value of the objectives such as makespan (completion time of the last job in a schedule) and total flow time (sum of completion times of all jobs in a schedule) for two different schedules can be obtained from Gantt chart as shown in Figure 2.

Therefore, the makespans for schedules $J_1 – J_2$ and $J_2 – J_1$ are 14 and 18 respectively. Also, the total flow time of jobs in the $J_1 – J_2$ schedule is 13 + 14 = 27 and the same for the $J_2 – J_1$ schedule is 8 + 18 = 26. So, minimum makespan results in the $J_1 – J_2$ schedule whereas minimum total flow time is obtained in the $J_2 – J_1$ schedule.

Similarly, let us consider an example of job-shop scheduling problem consisting of three jobs whose three operations O_1, O_2, and O_3 are to be performed on three machines M_1, M_2, and M_3. Jobs J_1, J_2, and J_3 follow sequence of operations $O_1 – O_2$

Table 2. Processing time matrix

Job	Operation		
	M_1	M_2	M_3
J_1	2	3	4
J_2	1	3	3
J_3	4	3	2

Figure 3. Workflow in a jobshop

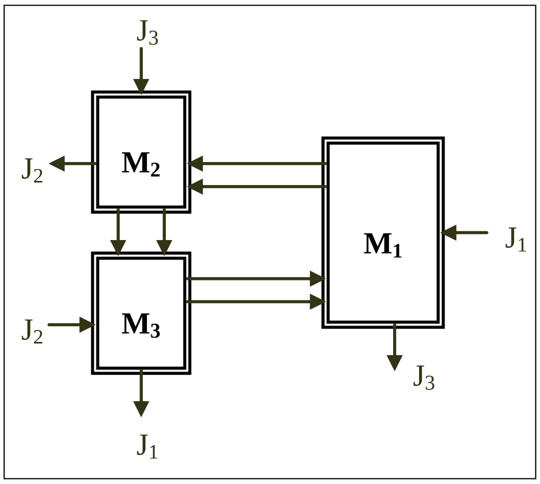
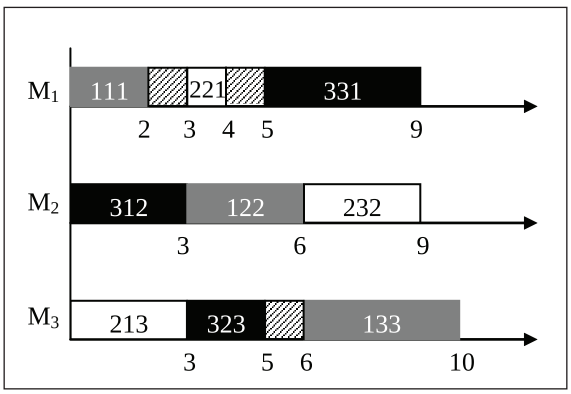


Figure 4. Gantt chart for a feasible schedule in a jobshop



– O_3, $O_3 – O_1 – O_2$, and $O_2 – O_3 – O_1$, respectively. The routing of these jobs is shown in Figure 3. The processing time matrix (in minutes) for this problem is given in Table 2.

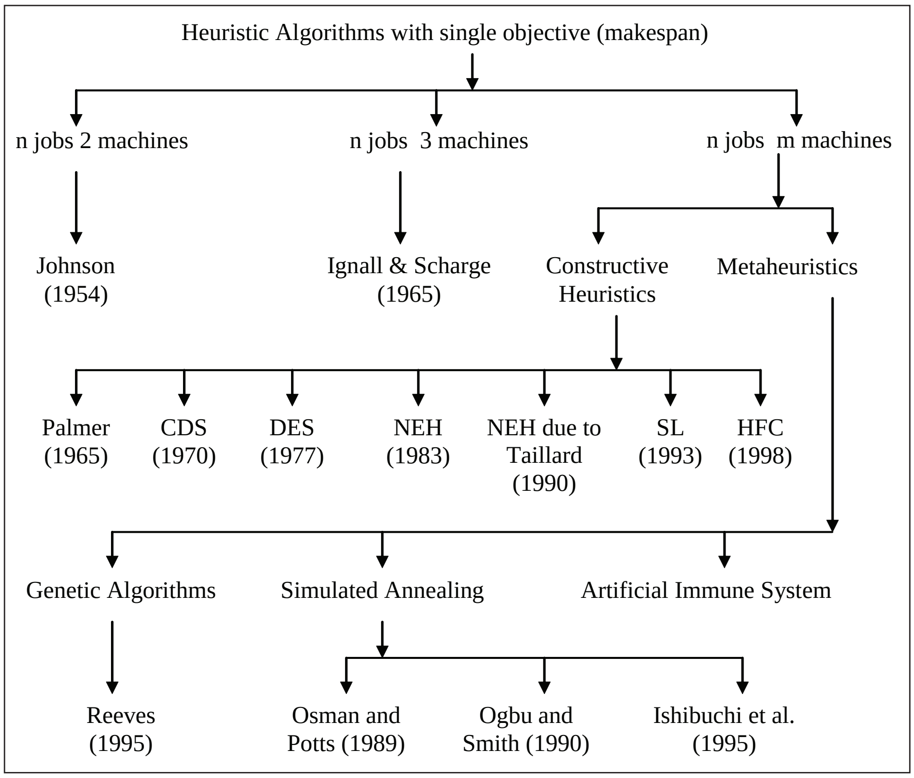
In the jobshop situation it is better to describe an operation with a triplet (i,j,k), where operation j of job i is processed on machine k. A feasible jobshop schedule is shown in the following Gantt chart (Figure 4). The makespan of this schedule of jobs is 10.

OVERVIEW OF SCHEDULING ON MAKESPAN CRITERION

It has been observed that the flowshop as well as jobshop scheduling problems, with few exceptions, belong to the class of combinatorial problems, which are termed as NP-complete for which no efficient polynomial time algorithm is available (Gonzalez & Sahni, 1978). Simple exact analytical methods such as brand-and-bound have been developed by Lomnicki (1965), Brown and Lomnicki (1966), and Bestwick and Hastings (1976). Although it is the best optimizing method available for solving NP-complete scheduling problems, it requires high central processing unit (CPU) time to solve large scheduling problems. So, the heuristic algorithms probably are the only means to solve especially large-sized scheduling problems that are frequently encountered in many real-life situations. These heuristics guarantee good solutions that are satisfactory though they may not be globally optimal.

For the past 50 years, flowshop scheduling has been one of the most important area in the scheduling literature. The scheduling heuristic approach generally cited as the foundation technique is the one developed by Johnson (1954). He presented a simple, well-known constructive heuristic algorithm to minimize makespan for the n-job, 2-machine scheduling problem. Due to the simplicity of Johnson's algorithm and its guarantee for giving optimal solutions, many researchers were encouraged to extend this idea to the general n-job, m-machine case, but without much success. Since then most of the efforts have been directed at finding optimal solution with m-machine scheduling problems. Ignall and Schrage (1965) developed an optimization algorithm using the branch-and-bound method for three-machine flowshop problems. Efficient heuristics that yield optimal solutions are desirable for generalized n-job, m-machine flowshop problems. Some of the noteworthy heuristics on the makespan criterion have been developed by Palmer (1965), King and Spachis (1980), Dannnenbring (1977), Campbell, Dudek, and Smith (1970) (called CDS), Nawaz, Enscore, and Ham (1983) (called NEH), Koulamas (1998), Widmer and Hertz (1989), Taillard (1990), Sarin and Lefoka (1993) (called SL), Osman and Potts (1989), and Ogbu and Smith (1990). These heuristics can be broadly divided into two categories: constructive heuristics and improvement heuristics. A constructive heuristic generates a schedule of jobs so that once a decision is taken it cannot be changed for improvement. The heuristics of Campbell et al. (1970), Nawaz et al. (1983), and Koulamas (1998) are of the constructive type. An improvement heuristic starts with an initial sequence of jobs and an attempt is made to improve the objective function by changing the job positions in the sequence. Some improvement heuristics are due to Ben-Daya (1998), Taillard (1990), Osman and Potts (1989), and Ogbu and Smith (1990). These heuristics are also called the metaheuristics. The classification of these heuristic algorithms based on makespan criterion for flowshop scheduling is shown in Figure 5.

Figure 5. Classification of heuristics with makespan objective for flowshop scheduling



Johnson's 2-machine algorithm gives the optimal solution with a view to minimizing the makespan but it fails to generalize to m-machine problems. Comparing CDS, NEH, and other heuristics, it is observed from Park's (1981) study that NEH is least biased and best operated of the heuristics and the CDS algorithm comes next. It is also proposed in the paper of Nawaz et al. that it would continue to perform better than CDS for problems with large numbers of machines and jobs (m, n $\geq$ 100). But Park, during his study, omitted Dannengbring's heuristic "rapid access with extensive search" (RAES), which has been found superior to CDS as pointed out by Turner and Booth (1987). Turner and Booth also observed that NEH proved to be more efficient than RAES based on both makespan and CPU time as measures of performance. So, NEH is clearly an improvement over the other published heuristics and RAES comes next.

To compare between NEH and SL, Sarin and Lefoka have shown that NEH is less effective than SL for scheduling problems with large number of machines. The effectiveness of NEH tends to improve as the number of jobs increases. Sarin and Lefoka also noted that the SL heuristic produces inferior solution compared to NEH for small and medium number of machines (m$\leq$100) and outperforms the NEH heuristic consistently for m $\geq$ 150 regardless the number of jobs. Also, the CPU time of the SL heuristic is very small compared to that of NEH. NEH requires more computational time because the work involved in computing makespan is a function of the number of jobs and each partial sequence is also a function of the number of machines.

Later, Koulamas (1998) in his paper proposed an effective constructive heuristic (called HFC for "heuristic flowshop scheduling with C_{max} objective") for flowshop scheduling problem with makespan objective. Computational results indicate that HFC performs as well as NEH on scheduling problems where a permutation schedule is expected to be optimal. However, HFC shows superiority over NEH on problems where a nonpermutation schedule may be optimal.

OVERVIEW OF SCHEDULING ON TOTAL FLOWTIME CRITERION

Apart from the heuristics on makespan criterion, there are some significant heuristics, which are either total-flow-time criterion based (Rajendran & Chaudhuri, 1991) or multiple criteria based (Rajendran, 1994). A survey of the flowshop scheduling literature has revealed that very little significant research work has been done on multiobjective criteria (considering more than two objectives) simultaneously.

Some noteworthy heuristics on total flow time criterion have been developed by Gupta (1971), Ho and Chang (1991), Rajendran and Chaudhuri (1991), Rajendran (1993), Ho (1995), Woo and Yim (1998), Liu and Reeves (2001), Allahverdi and Aldowaisan (2002), and Framinan and Leisten (2003). Framinan et al. (2005) present two new composite heuristics and the subsequent computational results show these heuristics to be efficient for the flowtime minimization in flowshops.

Rajendran and Chaudhuri (1991) propose three heuristics and compare with those of Gupta (1971), Miyazaki et al. (1978), and Ho and Chang (1991). The results reveal that their heuristics perform superior results in terms of both quality of the solution and computational time. Rajendran (1993) develops a new heuristic, which is better than that of Rajendran and Chaudhuri (1991), but at the expense of large computational effort.

Ho (1995) proposes an improvement heuristic based on finding local solution by adjacent pair wise interchange method, and later improves the solution by the insertion method. This heuristic performs better than the previous heuristics, but it consumes much higher CPU time for large problem sizes (Framinan & Leisten, 2003).

Framinan and Leisten (2003) propose a new heuristic based on the idea of optimizing partial schedules, already presented in the heuristic by Nawaz et al. (1983). The computational results show that their heuristic is currently the best for total flow time minimization in flowshops. It is compared with that of Woo and Yim (1998) having the same time-complexity of O (n^4m).

It is revealed from the survey of scheduling literature that the three heuristics of Rajendran and Chaudhuri (1991) yield consistently near optional solutions and require smaller CPU time. However, the heuristic by Framinan and Leisten (2003) outperforms the current best heuristic but its only disadvantage is that it requires higher computational effort.

SCHEDULING USING GENETIC ALGORITHMS

Recently, some efficient optimization methods based on the evolutionary computing paradigm such as genetic algorithms (GAs) (Davis, 1991; Goldberg, 1989; Holland, 1975; Pal & Wang, 1996) and simulated annealing (SA) (Aarts & Korst, 1989; Krikpatrick, Gelett, & Vecchi, 1983; Van Laarhoven & Aarts, 1987) have emerged. They have been developed for obtaining near optimal solutions from large, complex search spaces even in the presence of high dimensionality, multimodality, discontinuity, and noise. Characteristics of both these tools are recently being utilized for developing efficient optimization methods for different engineering problems. Both GAs and SA are superior to gradient descent or random search techniques as the search process is not biased to local optimal solutions. GAs are randomized search and optimization algorithms guided by the principles of evolution and natural genetics. They are efficient, adaptive, and robust search processes, producing near optimal solutions and a large amount of implicit parallelism. The GA approach, first developed by John Holland (1975), seeks to mimic the behavior of nature in the evolution of species, that is, is based on the principles of evolution and genetics to guide the search which results in the "survival of the fittest." It requires the specification of the candidate solutions in the form of a binary or nonbinary string. These strings of artificial genetic systems are analogous to chromosomes in nature. A chromosome is a candidate solution represented by a sequence of binary digits or integers (or floating-point numbers). A chromosome in turn consists of genes, each of which describes a unique feature of the organism. The value of the feature associated with a particular gene is called its allele. A collection of chromosomes is called a population. A genetic operator, called crossover, combines two chromosomes to create offspring (new candidates) that inherit the genetic material of the parent chromosomes. In each generation the selection of chromosomes to participate in the creation of new candidate solutions is based on their ability to survive in the competitive environment. A genetic operator, called mutation, is used to (re)introduce new genetic material into the population (typically with a very small probability). The simple genetic algorithm can be outlined as follows:

```
t = 0;
initialize population(t);
evaluate candidate points in population(t);
while predetermined termination condition not satisfied
{
t = t + 1;
select population(t) from population(t-1);
apply crossover and mutation to candidate points in population(t);
evaluate candidate points in population(t);
}
```

Reeves (1995) proposed a new genetic algorithm on the makespan criterion for flowshop sequencing. He used a new fitness function (proposed by Auckley [1987]) and shift mutation. The superiority of the proposed genetic algorithm over NEH and simulated annealing method was established using different problem sizes. Note that Reeves used the NEH heuristic to generate initial sequence in the population. Sridhar and Rajendran (1996) have proposed a genetic algorithm for the problem of scheduling in flowshop and flowline-based cellular manufacturing systems. The proposed genetic algorithm has been evaluated and is found to yield much better solutions than those given by the existing multiple criteria heuristic (Ho & Chang, 1991) in flowshop scheduling.

SCHEDULING USING SIMULATED ANNEALING

Simulated annealing (SA) simulates the annealing of physical systems for solving optimization problems. It also leads to near optimal solution through the process of probabilistic state transition. It attempts to overcome the disadvantage of the descent method. SA has its origin in statistical mechanics where the process of cooling solids until they reach a low energy level is called annealing. It is based on the work of Metropolis, Rosenbluth, Rosenbluth, Teller, and Teller (1956) who simulate the energy levels in cooling solids by producing a sequence of states. Kirkpatrick et al. (1983) pointed out the relevance of simulated annealing in combinatorial optimization problems. Van Larhoven and Aaarts (1987) have reviewed a wide variety of applications. The steps of the standard simulated annealing are as follows:

1. Initialize Max-iterations, Temp-start.
   Set Count = 1, T = Temp-start.
   Let the current sequence be x_c. Compute makespan(x_c).
2. Randomly generate a neighboring sequence by using some neighborhood schemes (interchange neighborhood or shift neighborhood). Let the neighboring sequence be called the adjacent sequence, x_a. Compute makespan (x_a).
3. If makespan(x_a) < makespan(x_c)
   then set $x_c = x_a$;
   else
   - Set Δ = makespan (x_a) – makespan (x_c);
   - Set T = Temp-start / log(1+Count);
   - With probability $e^{-\Delta/T}$ set $x_c = x_a$.
4. Increment Count by 1;
   If Count < Max-iterations, go to Step 2.
5. Output the current best sequence as the final solution.

Osman and Potts (1989) proposed a simulated annealing for permutation, flowshop problem with the objective of completion time (or makespan). The proposed heuristic shows superior results compared to the best known constructive heuristic (NEH heuristic). Osman and Potts in their heuristic used shift neighbourhood and a cooling schedule due to Lundy and Mees (1986). Ogbu and Smith (1990) also proposed different simulated annealing heuristics for the flowshop scheduling problem. But the simulated annealing result of Osman and Potts is slightly better than the Ogbu and Smith heuristic. Ishibuchi, Misaki, and Tanaka (1995) proposed two modified simulated annealing algorithms for the flowshop sequencing problem with the objective of minimizing the makespan. It was shown that their proposed algorithms perform as well as the simulated annealing of Osman and Potts on the average. Chakravarthy and Rajendran (1999) proposed a simulated annealing heuristic for scheduling in a flowshop with the objective of minimzing the makespan and maximum tardiness of a job.

SCHEDULING USING ARTIFICIAL IMMUNE SYSTEMS

Artificial immune systems (AISs) can be defined as the distribution and adaptive computational systems inspired by the human immune system and can be applied to solve specific problem. It is relatively a new area developed by theoretical immunologists (Jerne 1974; Perelson, 1989). Recently, it is considered as one of the important emerging computational tools that can be used in different areas of science, engineering, and management. The main interest of developing the AIS is not the modelling of immune system but to extract or glean from it some useful mechanisms that can be used as metaphors to consider as a computational tool for solving particular problems.

The literature survey reveals that AIS has been applied to various fields ranging from network security to optimization. In this connection, Timmis, Knight, de Castro, and Hart (2004) have narrated an in-depth overview of AIS. The early work in connection with AIS was performed in the area of fault diagnosis. Later, other works applied different AIS metaphors to the areas of computer security, optimization, and scheduling. However, it seems that there is no niche area for AIS. In this regard, it can be thought of as a novel soft computing paradigm with features of flexibility and robustness similar to many biologically inspired techniques such as neural networks and genetic algorithms that are suitable for various applications. An increasing number of conferences such as IEEE International Conference on Systems, Man, and Cybernetics (IEEE SMC) and International Conference on Artificial Immune Systems (ICARIS) are held due to the growing interest in developing AIS to the researchers. The first international conference on artificial immune system at the University of Kent in 2002 had a great success.

AIS can be considered as an important robust computational tool in the field of computing and engineering because it possesses a good number of features, such as:

- Learning
- Adaptability
- Highly distributed
- Self-organizing
- Diversity
- Recognition
- Memory

Learning: A process known as affinity maturation guarantees that the immune system becomes increasingly better to recognize the patterns.

Adaptation: The cells in the immune system are created sufficiently as when required to combat invading antigens, thereby replacing the older, ineffective, and dead ones for the survival of the organism.

Distribution: Each cell in the immune system having the inherent distribution feature responds and recognizes a foreign antigen that can invade the organism in any location.

Self-Organization: The immune system has inherent self-organizing mechanism that helps to control its population by local interactions in order to maintain a steady state within the whole system of the organism.

Diversity: There are two processes responsible for generating and maintenance of diversity in the immune system. The first process involves producing an almost infinite number of different types of receptor molecules by recombining the genes from a finite set. The second process helps in reproducing immune receptors within themselves, which is known as somatic hypermutaion.

Recognition: The immune system can recognize, identify, and respond to different cells. Additionally, the immune system has the ability to distinguish between self-cells and nonself cells.

Memory: A process known as maturation of the immune response permits some sets of cells and molecules a longer life span so that the immune system can respond to future infections caused by the same or similar antigens. Vaccination procedures in medicine and immunotherapy utilize this principle.

The Vertebrate Immune System

The purpose of the immune system is to protect our body system from infectious diseases caused by agents such as viruses, bacteria, fungi, and other parasites. It is made up of a set of cells and molecules that function with other bodily systems to maintain a steady state within the organism. Antigens belong to the surface of these agents that identify invading agents (pathogens) by the immune cells and molecules, resulting in an immune response. There are two types of immune systems: innate and adaptive. The innate immune system plays an important role in the initiation and regulation of the immune responses. The adaptive immune system mainly comprises lymphocytes or white blood cells, or B- and T-cells. These cells recognize and destroy specific antigens. Antigens are not the invading parasites themselves; they are the substances such as toxins and enzymes in the parasites considered as foreign agents by the immune system. There are two types of immune responses in the immune system, such as primary and secondary responses. Primary response allows the immune system to fight against an antigen for the first time. However, after a certain time period, the B-cells and antibodies begin to decay, until the antigen again faces encountering. After the primary response, some B-cells remain active in the immune system, acting as memory cells. The second response, known as the secondary response, is transferred to memory cells so that when the antigen is encountered, the new antibodies need not be produced to fight against them but the memory cells already existing in the system will help to eliminate them. B-cells clone and mutate to generate large numbers of antibodies to encounter the antigens from the infectious body. The antibodies are the specific protein that recognize and combine with other proteins. Each antibody consists of two paratopes and two epitopes that are specific protein to identify other molecules. Binding between the antibodies and antigens means how well the paratope of the antibody matches the epitope of the antigen. T-cells, the part of the immune system, affect B-cells during the immune response process.

It is revealed from the literature that two theories have been identified, namely clone selection and immune network, to explain how the immune memory is achieved and maintained.

- **Clonal selection theory:** The primary immune response to encounter an antigen is governed by few small clones of B-cells, each creating a diverse set of antibodies with different affinity. The effectiveness of secondary immune response to the antigen greatly depends upon the large clone with high affinity remaining after first encounter, named memory cells, in the immune system.
- **Immune network theory:** According to immune network theory, B-cells stimulate each other with the help of their receptor molecules in an attempt to produce mimic antigens. In this way, a network of B-cells is formed and B-cells with higher stimulation survive and less stimulated B-cells are eliminated from the system. The immune network describes an effective method for achieving the memory and stimulating a dynamic system. It can be thought of representing the network in the form of layered network that comprises representations, affinity, and immune algorithms.

Artificial Immune System

De Castro and Timmis (2002) proposed the idea of a framework for artificial immune system (AIS). They compared AIS with other biologically inspired methodologies such as artificial neural networks (ANN) and genetic algorithms (GAs) in order to gather similar ideas regarding understanding and development of such systems. AIS, ANN, and GA are all evolutionary and biologically inspired algorithms. In case of ANN, it is made up of a set of artificial neurons which can be arranged to construct a network. So, a framework to design an ANN consists of artificial neurons, a network of representing these neurons, and a learning algorithm in order to acquire knowledge. Similarly, genetic algorithms consider a set of artificial chromosomes representing a population of individuals which undergo a set of processes like reproduction, selection, crossover, and mutation. So, in order to design a framework of any biologically inspired algorithm, it has at least the following basic components:

- A pattern for representing the elements of the system.
- A set of evaluation processes so that the elements can interact with each other and the environment.
- An adaptive process for controlling the dynamics of the system.

Based on this approach, the basis for framework for AIS requires a representation to build an artificial model of immune organs, cells, and molecules, a set of functions, namely affinity functions, and a set of algorithms to govern the dynamics of the system.

The procedural steps of AIS can be summarized as follows:

1. A set of antibodies (called population) is created.
2. Decode the antibodies in the antibody population. Determine the affinity (makespan) of antibodies.
3. Calculate the selection probabilities and generate copies of antibodies (cloning).
4. Perform mutation (inverse or pairwise or both) to generate new clones.
5. Select the best antibodies by replacing some worse antibodies.
6. Repeat the process until the predetermined number of generations is satisfied.

In the context of scheduling problems, the affinity calculation helps to achieve the embodiment of diversity. The affinity value of a schedule of jobs (antibody) refers to its makespan value. It is computed as follows:

$$\text{Affinity (y)} = \frac{1}{\text{makespan(y)}}$$

where, y is the schedule (antibody)

Table 3. Comparison between GA and AIS and their analogous to scheduling problem

GA	AIS	Analogous to Scheduling Problem
1. Chromosomes: components of representation of the system	1. Antibodies: components of representation of the system	1. Schedules of jobs
2. Gene: an element of chromosome	2. Receptor molecules	2. Job in a schedule
3. Fitness function-evaluating each individual	3. Affinity function	3. Reciprocal of objective functions such as makespan
4. Crossover and mutation: mechanisms for creating new individuals	4. Cloning and mutation: mechanism for creating new individuals	4. Creating alternate feasible schedules.

According to clonal selection theory, the effectiveness of immune response depends on a clone with a higher affinity to encounter antigens in the immune system. Since the objective of these scheduling problems is to minimize the makespan, the affinity of an antibody is indirectly proportional to makespan. Engin and Doyen (2004), and Alisantoso, Khoo, and Jiang (2003) discussed the details of the above steps and the associated control parameters of AIS. Table 3 shows the comparison between the parameters of AIS and GA with reference to the scheduling problems.

Application of AIS to Scheduling Problems

In the area of optimization, Mori, Tsukiyma, and Fukuda (1998) employ somatic mutation of AIS combining with operators of standard genetic algorithm. They claim that their algorithm has the ability to possess higher diversity of candidate solutions compared to a standard GA. De Castro and Von Zuben (2000) propose the algorithm using clonal selection, affinity maturation process of an adaptive immune response to different applications, like optimization, machine learning, and pattern recognition. The different mechanisms of AIS (clonal selection, affinity mutation, memory, etc.) make the system very useful for solving different scheduling problems. In the context of scheduling problems, Mori et al. (1998) propose an immune system algorithm based on immunological metaphors of somatic hypermutaion and immune network principle in a dynamic scheduling environment. They determined the batch sizes and sequences of job orders with a view to optimizing the objective functions. In their work, antigens are considered analogous to input data in the scheduling problem, and antibodies analogous to generated schedules of jobs. The production of antibodies is governed by the metaphor of immune network and the effect of T-cells is not considered in the algorithm. Hart, Ross, and Nelson (1998) developed an adaptive method that can produce a diverse set of schedules, but not necessarily an optimal solution with respect to changing environment. They use the metaphor of antibodies and antigens as a single schedule and possible changes to the schedule in their system. They have generated new antibodies (a set schedules) using GA from the set of random data. Also, the schedules of jobs corresponding to the antigens in the set can be retrieved in their system. In another work, Mori, Tsukiyma, and Fukuda (1997) propose a methodology based on AIS to control a semiconductor production line. A set of agents was selected to control the production line. Each agent interacted with other agents as well as the production line. Costa, Vargas, Von Zuben, and Franca (2002) present a scheduling algorithm based on AIS with a view to minimizing makespan on parallel processors. The performance of their technique shows improvement results when compared to other heuristics, such as longest processing time, local search, and simulated annealing.

DISCUSSION

Other soft computing paradigms, such as SA, GAs, and fuzzy systems have identified themselves to have a well-described set of components and mechanisms in which the algorithms are made. These computing methods have a well-defined general framework for designing them properly. Although a great deal of research on genetic algorithms and simulated annealing has been done, the intricate relationship between different control parameters with which to design the algorithm are to be explored in details in different areas of engineering and management. On the contrary, AIS is lacking the development of a general framework in connection with extracting the in-depth ideas of its useful mechanisms. Although recently a framework has been proposed by De Castro et al. (2002) regarding modeling of AIS, much work remains in terms of formalization from a mathematical point of view, and development of new algorithms in the other areas of immunology as yet unexplored. Recently, the use of a new idea "danger theory" has been postulated by Acklien and Cayzer (2002). This theory has a wide scope to offer AIS in terms of a paradigm shift in thinking. In the computation, AIS has a tremendous potential to explore its rich novel area of research by in-depth interacting between biology and computer science.

CONCLUSION

In this chapter, the problem of manufacturing scheduling has been considered as scheduling is not only treated as an important module in the shop floor control system but also offers a great theoretical challenge for researchers because of its combinatorial nature. Due to its NP-completeness, heuristics and metaheuristics are probably most suitable means to solve large-scale scheduling problems. A comparative overview of various types of heuristics and metaheuristics including genetic algorithms, simulated annealing, and artificial immune system has been made. Theses computational tools are very effective in solving scheduling problems and other industrial problems.

FUTURE RESEARCH DIRECTIONS

Future research work could include the following:

1. Different types of scheduling problems like continuous flowshop scheduling, stochastic scheduling, and dynamic scheduling are to be considered.
2. It should include the explicit treatment and examination of online scheduling problems.
3. Recent optimization methods based on ant colony algorithm, particle swarm optimization, and so forth, should be suitably applied to scheduling problems.
4. More research in the direction of more general situation of combining constructive and search heuristic could be promising. It would be interesting how the modern heuristic such as simulated annealing, taboo search, and genetic algorithms could use their initial solutions generated by constructive heuristic.
5. Simulated annealing, artificial immune systems, and genetic algorithms provide a variety of options of the mechanisms and parameter settings, which have to be fully explored in the context of solving large and complex scheduling problems.
6. Better learning strategies using artificial neural networks and performances of sensitivity analysis to determine good parameter values in the context of scheduling problems appear to be promising.

REFERENCES

Aarts, E. H. L., & Korst, J. H. M. (1989). *Simulated annealing and boltzman machines.* Chichester: Wiley.

Ackley, D. H. (1987). An empirical study of bit vector function optimization. In L. Davis (Ed.), *Genetic algorithms and simulated annealing* (pp. 170-204). London: Pitman.

Acklien, U., & Cayzer, S. (2002). The danger theory and its application to artificial immune systems. In *Proceedings of the 1st International Conference on Artificial Immune Systems* (*ICARIS*), Canterbury, United Kingdom (pp. 141-148).

Alisantoso, D., Khoo, L. P., & Jiang, P. Y. (2003). An immune algorithm approach to scheduling of a flexible PCB flowshop. *International Journal* *of Advanced Manufacturing Technology, 22,* 819-827.

Allahverdi, A. & Aldowaisan, T. (2002). New heuristics to minimize total completion time in m-machine flowshops. *International Journal of Production Economics, 77,* 71-83.

Baker, K. R. (1974). *Introduction to sequencing and scheduling.* New York: Wiley.

Ben-Daya, M., & Al-Fewzan, M. (1998). A tabu search approach for the flow shop scheduling problem. *European Journal of Operational Research, 109,* 88-95.

Bestwick, P. F., & Hastings, N. A. J. (1976). A new bound for machine scheduling. *Operations Research Quarterly, 27,* 479.

Brown, A. P. G., & Lomnicki, Z. A. (1966). Some applications of the branch and bound algorithm to machine scheduling problem. *Operations Research Quarterly, 17,* 173-186.

Campbell, H. G., Dudek, R. A., & Smith, M. L. (1970). A heuristic algorithm for the n job m machine sequencing problem. *Management Science*, *16*(B), 630-637.

Chakravarthy, K. & Rajendran, C. (1999). A heuristic for scheduling in a flowshop with the bicriteria of makespan and maximum tardiness minimization. *Production Planning & Control, 10*, 707-714.

Conway, R. W., Maxwell, W. L., & Miller, L. W. (1967). *Theory of scheduling.* Reading, MA: Addison-Wesley.

Costa, A. M., Vargas, P. A., Von Zuben, F. J., & Franca, P. M. (2002). *Makespan minimization on parallel processors: an immune based approach*. Paper presented at The Congress on Evolutionary Computation, IEEE World Congress on Computational Intelligence (pp. 115-123). Hawaii.

Dannenbring, D. G. (1977). An evaluation of flow-shop sequencing heuristic. *Management Science, 23*, 1174-1182.

Davis, L. (1991). *Handbook of genetic algorithms.* Van Nostrand Reinhold.

De Castro, L. N., & Timmis, J. (2002). *Artificial immune systems: A new computational intelligence approach*. London: Springer-Verlag.

De Castro, L. N., & Von Zuben, F. J. (2000). The clonal selection algorithm with engineering application. In *Proceeding of the Workshop on GECCO*, Las Vegas (pp. 36-37).

Dudek, R. A., & Teuton, O. F. (1964). Development of M stage decision rule for scheduling ‘n’ jobs through ‘m’ machines. *Operations Research, 12*(3), 471-497.

Engin, O., & Doyen, A. (2004). A new approach to solve hybrid flowshop scheduling problems by artificial immune system. *Future Generations Computer Systems, 20,* 1083-1095.

Framinan, J.M. & Leisten, R. (2003). An efficient constructive heuristic for flowtime minimisation in permutation flow shops. *Omega*, *31,* 311-317.

Framinan, J.M., Leisten, R., & Ruiz-Usano, R. (2005). Comparison of heuristics for flow time minimization in permutation flow shop. *Computers & Operations Research, 32,* 1237-1254.

Goldberg, D. E. (1989). *Genetic algorithms in search optimization and machine learning.* Reading, MA: Addison-Wesley.

Gonzalez, T., & Sahni, S. (1978). Flow shop and job shop scheduling: Complexity & approximation. *Operations Research, 26,* 36-52.

Gupta, J. N. D. (1971). A functional heuristic algorithm for the flowshop scheduling problem. *Operational Research Quarterly, 22*(1), 39-47.

Hart, E., Ross, P., & Nelson, J. (1998). Producing robust schedules via an artificial immune system. In *Proceedings of the ICEC*, Anchorage (pp. 464-469).

Ho, J.C. (1995). Flowshop sequencing with mean flowtime objective. *European Journal of Operational Research, 81,* 571-578.

Ho, J. C., & Chang, Y. L. (1991). A new heuristic for the n -Job m-machine flow-shop problem. *European Journal of Operations Research, 52*, 194-202.

Holland, J. H. (1975). *Adaptation in natural and artificial systems*. Ann Arbor, MI: The University of Michigan Press.

Ignall, E., & Scharge, L. E. (1965). Application of the branch and bound technique to some flowshop scheduling problems. *Operations Research, 13*, 400-412.

Ishibuchi, H., Misaki, S., & Tanaka, H. (1995). Modified simulated annealing algorithms for the flowshop sequencing problem. *European Journal of Operational Research, 81,* 388-398.

Jerne, N. K. (1974). Towards a network theory of the immune system. *Annual Review of Immunology* (*Inst. Pasteur*), *125*(C), 373-389.

Johnson, S. M. (1954). *Optimal* two and three stage production schedule with set-up times included. *Naval Research Logistics Quarterly, 1*, 61-68.

King, J. R., & Spachis, A. S. (1980). Heuristics for flow-shop scheduling. *International Journal of Production Research, 18*(3), 345-357.

Kirkpatrick, S., Gelett, C. D., & Vecchi, M. P. (1983). Optimization by simulated annealing. *Science, 220*, 621-680.

Koulamas, C. (1998). A new constructive heuristic for the flowshop scheduling problem. *European Journal of Operational Research, 105*, 66-71.

Liu, J., & Reeves, C. R. (2001). Constructive and composite heuristic solution to the $P||\Sigma C_i$ scheduling problem. *European Journal of Operations Research, 132*, 439-452.

Lomnicki, Z. A. (1965). A branch-and-bound algorithm for the exact solution of the three machine scheduling problem. *Operational Research Quarterly, 16*, 89-100.

Lundy, M., & Mees, A. (1986). Convergence of an annealing algorithm. *Mathematical Programming, 34* (1), 111-124.

Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., & Teller, E. (1956). Equation of state calculations by fast computing machines. *Journal of Chemical Physics, 21*, 1087-1092.

Miyazaki, S., Nishiyama, N., & Hashimoto, F. (1978). An adjacent pairwise approach to the mean flowtime scheduling problem. *Journal of the Operations Research Scoeity of Japan, 21,* 287-299.

Mori, M., Tsukiyma, M., & Fukuda, T. (1997). Artificial immune based management system for a semiconductor production line. In *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics* (pp. 851-855).

Mori, M., Tsukiyma, M., & Fukuda, T. (1998). *Adaptive* scheduling system inspired by immune system. In *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, San Diego, California (pp. 3833-3837).

Nawaz, M. E., Enscore, E., & Ham, I. (1983). A heuristic algorithm for the m- machine n- Job flowshop sequencing problem. *Omega, 11*, 91-95.

Ogbu, F. A., & Smith, D. K. (1990). The application of the simulated annealing algorithm to the solution of the n/m/Cmax flowshop problem. *Computers & Operations Research, 17*, 243-253.

Osman, I. H., & Potts, C. N. (1989). Simulated annealing for permutation flowshop scheduling. *Omega, 17*, 551-557.

Pal, S. K., & Wang, P. P. (1986). *Genetic algorithms for pattern recognition*. Boca Raton, FL: CRC Press.

Palmer, D. S. (1965). Sequencing jobs through a multistage process in the minimum total time: A quick method of obtaining a near optimum. *Operations Research Q, 16*, 101-107.

Park, Y. B. (1981). *A simulation study and an analysis for evaluation of performance – Effectiveness of flow shop sequencing heuristics: A static and a dynamic flowshop model*. Unpublished master's thesis, Pennsylvania State University.

Perelson, A. S. (1989). Immune network theory. *Immunological Review, 110*, 5-36.

Rajendran, C. (1994). A heuristic for flowshop scheduling in flowshop and flowline based manufacturing cell with multicriteria. *International Journal of Production Research, 32*(11), 2541-2558.

Rajendran, C. (1993). Heuristic algorithm for scheduling in a flowshop to minimize total flowtime. *International Journal of Production Economics, 29*, 65-73.

Rajendran, C., & Chaudhuri, D. (1991). An efficient heuristic approach to the scheduling of jobs in a flowshop. *European Journal of Operational Research, 61*, 318-325.

Reeves, C. R. (1996). *A* genetic algorithm for flowshop sequencing. *Computes and Operations Research, 22*(1), 5-13.

Sarin, S., & Lefoka, M. (1993). Scheduling heuristic for the n-job m-machine flow shop. *Omega, 21*(2), 229-234.

Simons, J. V., Jr. (1992). Heuristics in flow shop scheduling with sequence dependent setup times. *Omega, 20*(2), 215-225.

Sridhar, J., & Rajendran, C. (1996). Scheduling in flowshop and cellular manufacturing systems with multiple objectives: A genetic algorithmic approach. *Production Planning and Control, 7*(4), 374-382.

Sriker, B. N., & Ghosh, S. (1986). A MILP model for the n-job, m-stage flowshop with sequence dependent set-up times. *International Journal of Production Research, 24*, 1459-1474.

Taillard, E. (1990). Some efficient heuristic methods for the flow shop sequencing problem. *European Journal of Operational Research, 47*, 64-74.

Timmis, J., Knight, T., de Castro, L. N., & Hart, E. (2004). An overview of artificial immune systems. *Natural Computation Series*, 51-86. New York: Springer.

Turner, S., & Booth, D. (1987). Comparison of heuristics for flowshop sequencing. *Omega, 15*(1), 75-85.

Van Laarhoven, P. J. M., & Aarts, E. H. L. (1987). *Simulated annealing: Theory and applications.* Dordrecht, The Netherlands: Reidel.

Widmer, M., & Hertz, A. (1989). A new heuristic method for the flow shop sequencing problem. *European Journal of Operational Research, 41*, 186-193.

Woo, D.S. & Yim, H.S. (1998). A heuristic algorithm for mean flowtime objective in flowshop scheduling. *Computers and Operations Research, 25*, 175-182.

ADDITIONAL READINGS

Aldowaisan, T., & Allahverdi, A. (2003). *New* heuristics for no-wait flowshops to minimize makespan. *Computers & Operations Research, 30*, 1219-1231.

Aldowaisan, T., & Allahverdi, A. (2004). A new heuristic for m-machine no-wait flowshop to minimize total completion time. *The International Journal of Management Science, Omega, 32*, 345-352.

Bertolissi, E. (2000). Heuristic algorithm for scheduling in the no-wait flow-shop. *Journal of Materials Processing Technology, 107*, 459-465.

Chandrasekaran, M., Asokan, P., Kumanan, S., Balamurugan, T., & Nickolas, S. (2005). Solving job shop scheduling problems using artificial immune system. *International Journal Advanced Manufacturing Technology, 31*(5-6), 580-593.

Davoud Pour, H. (2001). A new heuristic for the n-machine flowshop problem. *Production Planning and Control, 12*, 648-653.

Fink, A., & Vob, S. (2003). *Solving* the continuous flowshop scheduling problem by metaheuristic. *European Journal of Operations Research, 151*, 400-414.

Grabowski, J., & Pempera, J. (2007). The permutation flow shop problem with blocking. *Omega, 35*, 302-311.

Kumar, A., Prakash, A., Shankar, R., & Tiwari, M.K. (2005). Psycho-Clonal algorithm based approach to solve continuous flowshop scheduling problem. *Expert Systems with Applications, 31*(3), 504-514.

Laha, D., & Chakraborty, U. K. (2007). An efficient stochastic hybrid for flowshop scheduling. *Engineering Applications of Artificial Intelligence, 20*, 851-856.

Pinedo, M. (2002). *Scheduling, theory, algorithms and systems.* NJ: Prentice Hall.

Rajendran, C., & Ziegler, H. (2004). Ant-colony algorithms for permutation flowshop scheduling to minimize makespan/total flowtime of jobs. *European Journal of Operations Research, 155*, 426-438.

Ruiz, R., & Moroto, C. (2005). A comprehensive review and evaluation of permutation flowshop heuristics. *European Journal of Operations Research, 165*, 479-494.

Ruiz, R., & Stutzle, T. (2007). A simple and effective iterated greedy algorithm for the permutation flowshop scheduling problem. *European Journal of Operations Research, 177*(3), 2033-2049.

Schuster, C. J., & Framinan, J. M. (2003). Approximate procedures for no-wait job shop scheduling. *Operations Research Letters, 31*, 308-318.

Wang, L., & Zheng, D .Z. (2003). An effective hybrid heuristic for flowshop sequencing. *The International Journal of Advanced Manufacturing Technology, 21*, 707-722.

Chapter II
Solving Machine Loading Problem of FMS:
An Artificial Intelligence (AI) Based Random Search Optimization Approach

Anoop Prakash
University of Cincinnati, USA

Nagesh Shukla
University of Warwick, UK

Ravi Shankar
Indian Institute of Technology, India

Manoj Kumar Tiwari
Indian Institute of Technology, India

ABSTRACT

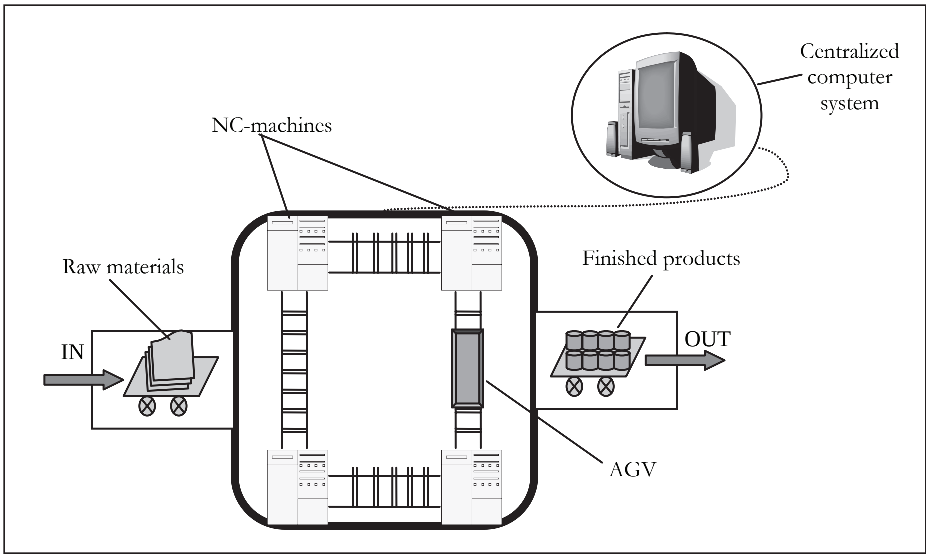
Artificial intelligence (AI) refers to intelligence artificially realized through computation. AI has emerged as one of the promising computer science discipline originated in mid-1950. Over the past few decades, AI based random search algorithms, namely, genetic algorithm, ant colony optimization, and so forth have found their applicability in solving various real-world problems of complex nature. This chapter is mainly concerned with the application of some AI based random search algorithms, namely, genetic algorithm (GA), ant colony optimization (ACO), simulated annealing (SA), artificial immune system (AIS), and tabu search (TS), to solve the machine loading problem in flexible manufacturing system. Performance evaluation of the aforementioned search algorithms have been tested over standard benchmark dataset. In addition, the results obtained from them are compared with the results of some of the best heuristic procedures in the literature. The objectives of the present chapter is to make the readers fully aware about the intricate solutions existing in the machine loading problem of flexible manufacturing systems (FMS) to exemplify the generic procedure of various AI based random search algorithms. Also, the present chapter describes the step-wise implementation of search algorithms over machine loading problem.

MACHINE LOADING PROBLEM IN FMS

Conceiving incomputable thrust in the market, and unassailable pressure from the competitors, it becomes worthy to the managers of modern manufacturing systems to analyze their systems critically. Time effective production with the least involvement of the cost has become a key issue in the automated manufacturing systems to strive hard in the market. Loading decisions play crucial role in inducing such behavior by processing the job in a feasible sequencing schedule. Effective loading decisions are particularly important in the large and complex manufacturing systems. In this chapter, we present an analytical study of the loading decisions of one of the most widespread domain of manufacturing systems, termed as flexible manufacturing systems, abbreviated as FMS. According to Stecke (1983), machine loading problem is one of six post release decisions of a flexible manufacturing system that is known for its computational complexity and high variability. A typical FMS consists of a set of highly automated *numeric control* (*NC*)-machines along with a material handling device both under the supervision of a centralized computer system (see figure 1). The highly sophisticated machine tools render the FMS to have wide range of manufacturing operations and allow simultaneous production of multiple part types maintaining a high degree of machine utilization (Tiwari, Hazarika, Vidyarthi, Jaggi, & Mukopadhyay, 1997). Machine-loading problems in flexible manufacturing deal with the assignment of various resources (fixture, pallets, tools, automated guided vehicles (AGVs), machine, etc.) to the operation of different part types planned for production in a given planning period. Loading problems in manufacturing deal with the assignment of various resources (machines, tools, fixtures, pallets, etc.) to the operations of different part types that are already planned for production in a given planning horizon. Decisions pertaining to loading problems have been viewed as tactical level planning decisions that receive their inputs from the preceding decision levels; namely, grouping of resources, selection of part mixes, aggregate planning that generates the inputs to the succeeding decisions of scheduling resources, and dynamic operations planning and control (Rai, Kameshwaran, & Tiwari, 2002). Therefore, it can be construed that loading decisions are acting as an important link between strategic and operation level decisions in manufacturing. Van Loovern, Gelders, and Van Wassenhove (1986), Kusiak (1985), Singhal (1978), and Whitney and Gaul (1984) discuss the interrelationships of various decisions and their hierarchies in a flexible manufacturing environment. Liang and Dutta (1993) address the part selection and machine loading problem simultaneously, which was separately treated by previous researchers. A number of researchers have addressed the part selection problem using various methods (Hwang, 1986; Kusiak, 1985; Stecke, 1986; Sawik, 1990), whereas, formulations and solution methodologies for various scenarios and combinations of parameters relating to the loading problem in an FMS have attracted the attention of numerous researchers including Liang and Dutta (1993), Tiwari et al. (1997), Laskari, Dutta, and Padhye (1987), Mukhopadhyay, Midha, and Krishna (1992), Chen and Askin (1990), Shanker and Tzen (1985), Shanker and Srinivasulu (1989), Ram, Sarin, and Chen (1990), Chen and Chung (1996), and so forth. Part selection, machine loading, and tool configurations are three different but interlinked problems that are connected by common restrictions such as tool magazine capacity, job-tool machine compatibility, and available machining time. However, for the clarity of the problem, most researchers have treated part selection, machine loading, and tool configuration in a discrete manner.

More generally, an FMS loading problem can be defined as:

Figure 1. A hypothetical layout of FMS



Let there be J jobs each needing assignment to a set of machines. Each job is characterized by a set of operations that can be performed on a definite set of alternate machines. Let there be M machines in the system, each of which has a definite processing time and fixed tool slots. Every operation of a job requires a prespecified processing time, fixed tool slot and a set of alternate machines. The problem is to find a unique sequence of jobs so that system unbalance can be minimized and throughput can be maximized by satisfying various technological constraints such as available machining time and tool slots. (Prakash et al., 2007)

In the past, researchers have estimated the complexity exist in the FMS loading problem and found it more complex than the jobshop scheduling problem (Stecke, 1993). The following factors justify the complexity of machine loading problem over jobshop scheduling problem:

- The machines are more versatile and capable of carrying out different operations.
- Several part types can be processed concurrently.
- Each operation of the job can be processed by more than one operation route.

Complexity of FMS loading problem can be deduced from the fact that even to process just eight different part types, it requires 104,509,440 (i.e., $8! \times 2{,}592$) operation-machine allocation combinations (Tiwari et al., 2006). Some of these allocations become nonoptimal due to their inability to satisfy system constraints. Determining an optimal solution in such a large search space is computationally intractable and needs a heuristic based approach to arrive at some optimal/near optimal solutions. Thus, in order to minimize the complexity in analyzing the machine loading problem following set of assumptions have been made in the literature.

- For each job, the process detail including processing time, tools, machines, and so forth are initially available.

- Splitting of the jobs are not required, that is, each operation of a certain part type is to be assigned only on a single machine.
- Unique routing of the job is required, that is, a job that has been selected for processing must be completed for all of its operation, before considering a new job.
- Tools can not be transferred across the machines
- Tool magazines at different machines may have different number of tool slots of identical shape and size

Historical Background of the Loading Problem

The machine loading problem consist of *J* jobs each having *O* operations and each operation could be done by a subset of machines that may have different processing times for the same operation. The operations of each job are typically subject to precedence constraints and some operations must be done in certain time slots. Moreover, due to fixture, tolerance, and tooling constraints, some operations must be done on the same machine as a group. Due to the relatedness of machine loading problem with the real world applications, it is one of the most explored problems. Stecke and Solberg (1981) have made the first attempt to solve the machine loading problem in FMS with the objective of maximizing the throughput rate. Since then, several approaches have been proposed to solve the machine loading problem. The chronological literature survey of the machine loading problem is given as follows: Stecke and Solberg (1981) introduced *branch and backtrack* policy to maximize the assigned workload thereby maximizing the workload balance. Their work is improved by Berrada and Stecke (1983). In the same year, Stecke (1983) formulated a nonlinear mixed integer programming for the loading problem, and identified six reasons to balance the processing time on the machines. In 1985, Shanker and Tzen combined the loading and dispatching problem while considering the random environment of FMS. Later, Shanker and Srinivasulu (1989) tackle the machine loading problem by giving analytical solution methodologies. Kumar, Singh, and Tiwari (1990) have done a multicriteria analysis of machine loading problem by applying *max-min* approach. In the same year, Chen and Askin (1990) propose a multiobjective heuristic to address the same problem. Tool selection and tool allocation problem have been integrated with the machine loading problem by many researchers (Chen & Askin, 1990; Kumar & Shanker, 2002; Sarin & Chen, 1987; Stecke, 1983). In order to efficiently reduce the system imbalance, Mukhopadhyay et al. (1992) have proposed an *essentiality ratio* based heuristic procedure to solve the machine loading problem. Further, Liang and Dutta (1993) combine the job-selection problem along with the load sharing problem to address the machine loading problem in a hybrid manufacturing system. A new *branch and bound* algorithm was developed by Kim and Yano (1994) to solve the loading problem in FMS environment. Tiwari et al. (1997) propose a heuristic solution approach to solve the loading problem in FMS, where, they analyzed the performance of different deterministic job sequencing rules. Mukhopadhyay, Singh, and Srivastava (1998) apply *simulated annealing (SA)* based approach to solve the FMS loading problem. Atmani and Lashkari (1998) have developed a model for simultaneous tool and operation allocation on the machine in FMS. The combination of part type selection and machine loading problem has been frequently addressed by many researchers (Gurrero, Lozano, Koltai, & Larraneta, 1999; Nayak & Acharya, 1998). A *genetic algorithm (GA)* based heuristic was developed for the machine loading problem by Tiwari and Vidhyarthi, (2000); here, system balance and throughput were considered as the objectives. A *tabu search* based heuristic to solve the machine loading problem was proposed by Sarma, Kant, Rai, and Tiwari (2002). They compare their tabu search with the various deterministic job sequencing rules. Further, Tiwari and Swarnkar (2004) propose a hybrid tabu search and simulated annealing heuristic to improve the solution quality and computational efficiency. In recent years, Tiwari et al. (2004b) proposed a modified version of genetic algorithm and simulated annealing to test the performance on the machine loading problem. Kim and Kim (2005) propose a multilevel symbiotic evolutionary algorithm to tackle the machine loading problem in FMS. Further, Prakash Khilwani, Tiwari, and Cohen (2007) propose an artificial immune system (AIS) based modified immune algorithm to resolve the traditional machine loading problem of FMS and compared the objective measures results with some known optimization techniques and heuristic procedures.

Mathematical Modeling

Machine loading problem comes under the broad range of flexible manufacturing system that consists of several NC machines and a material handling device, all under the supervision of a centrally located computer system. Until now, various objective measures have been considered by the researchers in order to expose the diversity of the machine loading problem. A description of the objectives considered in the literature pertaining to the loading problem can be found in the work of Tiwari et al. (2004b). The objectives discussed in this chapter are: (1) maximizing throughput, and (2) minimizing system's imbalance (this is equivalent to maximizing system's balance). Advantages of adopting these two objectives are as follows:

1. High system utilization can be achieved by minimizing the system imbalance or the idle time on machines.
2. Total system output would be augmented by considering maximum throughput, and this is one of the most desired objectives by manufacturers.
3. Tardiness can be limited by throughput maximization.

Perfect balanced is achieved if a system's imbalance is zero (the minimum imbalance value). For a given timeframe, the upper bound value of throughput is the sum of all the possible products that could be produced. Therefore, the upper bound throughput for the time horizon is equal to the sum of the batch sizes of all the jobs needed to be assigned. For calculating the system imbalance, both over utilized and under utilized times are considered. Given below are the notations that have been used in this chapter to formulate the objective functions, namely minimization of System unbalance and maximization of system's throughput.

Subscripts

j=index of job; $1 \leq j \leq J$; m=index of machine; $1 \leq m \leq M$; o = index of operation; $1 \leq o \leq O$;

Parameters

SU = system unbalance;
TH = throughput
H = planning horizon
$SU_{\max}$ = maximum system unbalance ($=M \times H$)
O_j = set of operations for the job j
M_{oj} = set of machines for performing o^th operation of the job j
bsz_j = batch size of the job j
MT_m^o = over utilized time on machine m
MT_m^u = under utilized time on machine m
MT_{jom}^a = time available on machine m for performing o^th operation of job j
$MT_{jom}^{r'}$ = time required by machine m for performing o^th operation of job j
$MT_{jom}^{r''}$ = time remaining on machine m after performing o^th operation of job j
MS_{jom}^a = tool slot available on machine m for performing o^th operation of job j

$MS^{r'}_{jom}$ = tool slot required by machine m for performing o[th] operation of job j

$MS^{r''}_{jom}$ = tool slot remaining on machine m after performing o[th] operation of job j

Decision Variables

$$\lambda_j = \begin{cases} 1 \ if \ \ job \ \ j \ \ is \ \ selected, \\ 0 \ otherwise. \end{cases}$$

$$\delta_{jom} = \begin{cases} 1 \ if \ \ operation \ o \ of \ \ job \ \ j \ \ is \\ \quad performed \ \ on \ \ machine \ \ m, \\ 0 \ otherwise. \end{cases}$$

Objective Function and Constraints

$$Maximize \ \ F_1 = \frac{SU_{\max} - \sum_{m=1}^{M} (MT^o_m + MT^u_m)}{SU_{\max}} \qquad (1)$$

$$Maximize \ \ F_2 = \frac{\sum_{j=1}^{J} bsz_j \times \lambda_j}{\sum_{j=1}^{J} bsz_j} \qquad (2)$$

$$Maximize \ \ F = \frac{(W_1 \times F_1) + (W_2 \times F_2)}{(W_1 + W_2)}$$

$$\text{Or} = \frac{\left(W_1 \times \frac{SU_{\max} - \sum_{m=1}^{M} (MT^o_m + MT^u_m)}{SU_{\max}} \right) + \left(W_2 \times \frac{\sum_{j=1}^{J} (bsz_j \times \lambda_j)}{\sum_{j=1}^{J} bsz_j} \right)}{(W_1 + W_2)} \qquad (3)$$

Where, W_1 and W_2 are the corresponding weights associated with the system unbalance and throughput, respectively. For the present case, equal weights have been assigned on the both objective (i.e., $W_1 = W_2 = 1$.).

The aforementioned objectives are subjected to following set of constraints:

$$\sum_{m=1}^{M} (MT^o_m + MT^u_m) \geq 0 \qquad (4)$$

$$\sum_{0=1}^{O_j} MS^{r'}_{jom} \delta_{jom} \leq MS^a_m \qquad (5)$$

$$\sum_{0=1}^{O_j} MT^{r'}_{jom} \delta_{jom} \leq MT^a_m \qquad (6)$$

$$\sum_{m \in M_{jo}} \delta_{jom} \leq 1 \qquad (7)$$

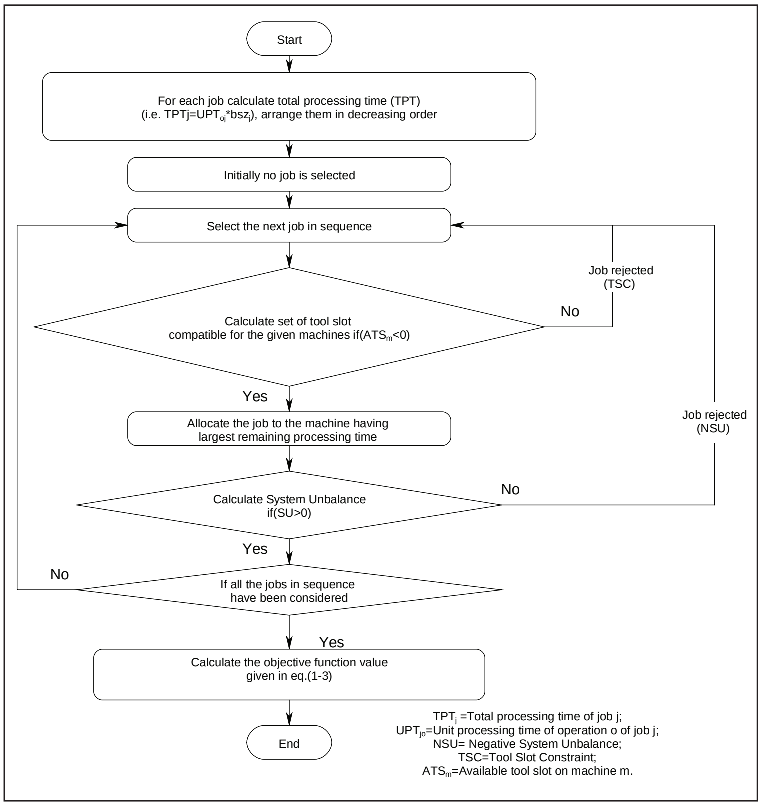
$$\sum_{m=1}^{M} \sum_{o=1}^{O_j} \delta_{jom} = \lambda_j \times O_j \qquad (8)$$

$$MS^{r''}_{jom} \delta_{jom} \geq 0 \qquad (9)$$

$$\sum_{m=1}^{M} MT^{r''}_{jom} \delta_{jom} \geq 0 \qquad (10)$$

The objective (equation (1)) is a measure of system unbalance, which is sum of both over utilized and under utilized time on the all machines used and objective. Equation (2) denotes the throughput that is the sum of batch size of the selected jobs. The objective in equation (3) is the combination of objectives defined in equations (1) and (2), respectively. The constraint in equation (4) shows that the value of system unbalance must be a positive quantity; zero system unbalance is the optimal case. The constraint given in equation (5) is based on the tool slot requirement of a job that explains the required tool slots by any machine for performing the operation a job must be less than the tool slot available on that machine. Constraint defined in equation (6) explains the same as of constraint in equation (5); only tool slot is

Figure 2. Step-wise procedure to evaluate the objectives of machine loading problem



replaced by the machining time. For the constraint shown in equation (7), in case of optional operation (where more than one machine is given for performing the operation), only single machine is required to perform the operation of a job. This process is termed as unique job routing. Constraint (equation (8)) describes the nonsplitting property of a job, where all the operation of a particular job is to be completed before considering a new job. The constraint shown in equation (9) depicts that the tool slot remaining on any machine after the completion of any operation of a job must be greater than or equal to zero. At last, the constraint define in equation (10) explains that the sum of remaining time on the all the machines after the completion of particular operation of a job must be a positive or zero quantity.

Procedure to evaluate the objectives (detailed earlier) utilized in machine loading problem is given in the Figure 2.

AI BASED RANDOM SEARCH ALGORITHMS

In the recent past, researchers have applied various AI based heuristics to resolve the complexities of the loading problem, thereby optimizing the concerned objectives. In this section, details of the concerned algorithms along with their generic procedure are provided.

Genetic Algorithm

A genetic algorithm is an "intelligent" probabilistic search algorithm that simulates the process of evolution by taking a population of solutions and applying genetic operators in each reproduction. Each solution in the population is evaluated according to some fitness measure. Fitter solutions in the population are used for reproduction. New offspring' solutions are generated and unfit solutions in the population are replaced. The cycle of evaluation-selection-reproduction is continued until a satisfactory solution is found (Goldberg, 1989; Michalewicz, 1992; Goldberg, Korb, & Deb, 1991). Holland (1975) introduced genetic algorithms which, later on, were applied to a wide variety of problems. Some of the typical applications of GAs includes the traveling salesman problem (Grefenstette, Gopal, Rormatia, & Vangucht, 1985), scheduling problem (Clevland & Smith, 1989; Davis, 1985a; Davis, 1985b; Davis, 1991), very large scale integration (VLSI) circuit layout design problem (Fourmann, 1985), computer aided gas pipeline operation problem (Goldberg, 1987a, 1987b), communication network control problem (Cox, Davis, & Qiu, 1991), real time control problem in manufacturing systems (Grefenstette, 1989, Lee, Piramuthu, & Tsai, 1997), cellular manufacturing (Gupta, Gupta, Kumar, & Sundram, 1996), and pattern classification (Bandyopadhyay, Murthy, & Pal, 1995; Bandyopadhyay & Pal, 1998). A generic procedure of genetic algorithm can be given as (Szumlanski, Wu, & Hughes, 2006):

Procedure GA
{
    Initialize population;
    While **termination condition** not satisfied
    {
        Select parents from population;
        Apply **genetic operators**;
        Evaluate offspring using **fitness function**;
        Replace parent population with offspring;
    }
}

Ant Colony Optimization

Introduced by Dorigo and Di Caro (1999), ant colony optimization (ACO) is a metaheuristic inspired by the foraging behavior of ant colonies. Ethnologists found that real ants, although blind, construct the shortest path from their colony to the feeding source through the use of pheromone trails. The process is imitated in ACO through the use of a set of simple agents (i.e., artificial ants) which were allocated with computational resources and they exploit stigmergic communication (Stutzle & Hoos, 2002), that is, a form of indirect communication mediated by the environment to find the solution to the problem at hand. During their motion, ants drop certain amount of pheromone, a chemical substance used by ants to communicate and exchange information about which course to follow, on their chosen path thus marking it with a trail of this substance. The coming ant senses the pheromone on different paths and probabilistically selects a path in accordance with the probability that is directly proportional to the amount of pheromone on it. During *NC* cycle for the k^{th} ant on node *i*, the selection probability of next node *j* to follow is given by:

$$\phi_{ij}^{k} = \begin{cases} \dfrac{[\tau_{ij}(NC)]^{\alpha}[\eta_{ij}(NC)]^{\beta}}{\sum_{k \in feas_k}[\tau_{ik}(NC)]^{\alpha}[\eta_{ik}(NC)]^{\beta}} & j \in feas_k \\ 0 & otherwise \end{cases} \qquad (11)$$

where, η_{ij} is a heuristic value called visibility on edge (*i*, *j*) and τ_{ij} is the amount of trail laid on edge (*i*,*j*). Thus it is an autocatalytic process characterized by a positive feedback loop where the probability of selection of a path improves with the number of ants that choose the same path. In order to ensure that the ants visit all the nodes, a special data structure called as *tabu* list is associated with each ant. The tabu list saves the node already visited and thereby forbids the ants to visit them again. Once all the ants complete their tours (which is called a cycle), the tabu list is utilized to calculate the ant's solution value and thereby update the pheromone count on each edge. The pheromone strength on each path is also decreased due to evaporation (ρ) and is calculated using a pheromone updation rule given as:

$$\tau_{ij}(NC+1) = (1-\rho)\,\tau_{ij} + \Delta\,\tau_{ij} \tag{12}$$

where

$$\Delta\tau_{ij} = \sum_{k=1}^{N} \Delta\tau_{ij}^{k} \tag{13}$$

and

$$\Delta\tau_{ij}^{k} = \begin{cases} \dfrac{Q}{L_k} & arc(i,j) \in tabu_k \\ 0 & otherwise \end{cases} \tag{14}$$

During the start of the new cycle, a *tabu* list is emptied and the updated values serve to guide the future path of ants. This process is iteratively repeated for a designated maximum number of cycles, specified central processing unit (CPU) time limit, or maximum number of cycles between two improvements of the global best solution. The ACO has successfully been applied to solve many complex NP-hard combinatorial optimization problems, such as traveling salesman problem (TSP) (Dorigo, Caro, & Gambardella, 1997), quadratic assignment problem (Maniezzo, Colorni, & Dorigo, 1999), and vehicle routing problems (Bell & McMulen, 2004) to name a few. The basic algorithm of the ACO introduced by Dorigo and Di Caro (1999) and Maniezzo, Dorigo, and Colorni (1996) is outlined as follows:

```
Initialize
    Represent the underlying problem by a weighted connected graph.
    Set initial pheromone for every edge.
Repeat
    For each ant do
        Randomly select a starting node.
    Repeat
        Move to the next node according to node transition rules.
            Until a complete tour is fulfilled.
    For each edge do
        Update the pheromone intensity using pheromone-updating rules.
        Until the stopping criterion is satisfied.
Output the global best result
```

Simulated Annealing

Simulated annealing (SA), originally proposed by Kirkpatrick, Gelatte, and Vecchi (1983), is a random search technique that is able to escape local optima using a probability function. SA draws its inspiration from physical annealing of solids, where a metal is brought to its lowest energy state by first heating it to a very high temperature (usually recrystallization temperature of metal) and then cooling at a very slow rate to a very low temperature. If the cooling is not slow enough, it may result in quenching, which is not desirable. Based on iterative improvement, the SA algorithm is a heuristic method with the basic idea of generating random displacement from any feasible solution. This process accepts not only the generated solutions, which improve the objective function, but also those which do not improve it with the probability expression of ($-\Delta f/T$), a parameter depending on the objective function and decreasing temperature.

This algorithm has two important features: perturbation scheme for generating a new solution and an annealing schedule that includes an initial and final temperature and rate of cooling.

A generic SA algorithm is as follows (Tiwari et al., 2006):

Get an initial solution S
Get an initial temperature T > 0
While not frozen, Do:
    Perform the following loop n times
        Pick a random neighbor S^* of S
        Let, $\Delta = f(S^*) - f(S)$
        If $\Delta<0$ (downhill move) then $S = S^*$
        If $\Delta \geq 0$ (uphill move), then $S = S^*$
        With probability $P(\Delta, T)$
        If $f(S) \geq f(S_{best})$ then Sbest= S
    T=New value of temperature
S_{best} is the best solution.

Artificial Immune System

People have always been fascinated by nature and it has inspired many theories to be applied to various areas. In a similar manner, the biological immune system present in the living body inspired the research community to develop an artificial immune system (AIS) to address the complex optimization problems. AIS can be defined as abstract or metamorphic computational system using ideas gleaned from theories and components of immunology (De Castro & Von Zuben, 2000).

It aims to solve a wide range of tasks related to complex computational and engineering problems such as pattern recognition, machine learning, and combinatorial optimization. In the literature, research on AIS has captured the attention of various researchers attempting to solve general combinatorial problems (Dasgupta, 1999; De Castro & Von Zuben, 2000; Endoh, Toma, & Yamada, 1998; Gaspar & Collard, 2000). All living organisms exhibits some sort of defense mechanism against foreign attack of disease or harmful cells. The main function of the immune system is to recognize all the cells in the body and categorize them into self and nonself cells. The immune system of a living being consists of a variety of molecules, cells, and organs spread throughout the body. There is nothing like central controlling organ for immune actions. There are several elements either in transition (e.g., blood stream) or in different stationary cells (e.g., thymus) performing their complementary roles. The immune system recognizes the malfunctioning and disease causing elements. Any molecules that can be recognized by the immune system are known as antigens. There are two types of antigens: self and nonself. Self-antigens originally belong to our own body and are harmless in their functioning, whereas nonself antigens are disease-causing elements. Some receptor molecules are there on the surface of immune cells, which are capable of recognizing almost limitless range of antigenic patterns. There are two major types of immune cells: B- and T-cells. These cells are similar in nature, but differ in the way they recognize antigens. Antigens free in solution (e.g., blood stream) can be recognized by B-cells, whereas T-cells recognize antigens present in other accessory cells.

When an animal is exposed to an antigen, some subpopulation of its bone marrow derived cells (B-lymphocyte) responds by producing antibodies. Antibodies are molecules present on the surface of B-cells which recognize and bind to an antigen. Thus, there is no distinction between B-cells and its receptor antibody. Recognition of an antigen is necessary for the immune system to become active and perform subsequent response. The recognition can only be activated if the cell receptor recognizes an antigen with an affinity greater than affinity-threshold. After T-cells are generated, they migrate into the thymus, an organ located behind the breastbone where they are matured. During maturation all self-antigens, which are recognized by T-cells, are excluded from the population of T-cells. This process is termed as negative selection (Nossal, 1994). If a B-cell encounters a nonself antigen with a suitable affinity-threshold, it proliferates and differentiates into memory and effecter cells, a process called clonal selection (Ada & Nossal, 1987). In contrast, if a B-cell recognizes a self-antigen, it may lead to suppression, as recommended by the immune network (Jerne, 1974). The process of adaptive biological evolution and the production of antibodies are strikingly similar considering that the two central processes involved in the production of antibodies, genetic recombination, and mutation are the same ones responsible for the biological evolution of sexually reproducing species (De Castro & Von Zuben, 2000). The functioning of the immune system is quite powerful as it can store memories of past experiences in terms of the strengths of the interactions of its constituent cells. It can also generate responses to new antigens patterns. The generic code of AIS is given as (Tiwari et al., 2006c):

Representation and Initialization POP

Define problem specific objective function

Maintain randomly generated initial POP of antibody.

Exposure of initial population to the threats posed by antigens.

Randomly choose an antigen ($Agj \in Ag_m$) and expose it to all antibody

Determine the affinity f_k of the entire antibody in relation to Ag_j.

Interaction between antibodies

Select n highest affinity antibody.

Make a new set of selected antibody.

Cloning of selected antibody proportionally to their antigenic affinity, $C_{k.}$

Hypermutation

New population C_{k*} of matured clones is generated by doing hypermutation. (Higher the affinity, lower the mutation rate.)

Determine the affinity f_{k*} of the matured clones. Now all the antibodies are selected to compose the memory.

Choose the best antibodies Ab_d by replacing d lowest-affinity antibody.

Repeat the process until termination criteria is not satisfied.

Tabu Search

Tabu search is regarded as a “high level” iterative improvement procedure that is used for solving various computationally complex problems (Sarma et al., 2002). Many successful implementations of tabu search, for obtaining near-optimal/optimal solutions of the problems pertaining to process planning, scheduling, set partitioning, and so forth are reported by Glover (1990) and Taillerd (1990).

Tabu search process begins with an initial feasible solution and attempts to find a better solution by investigation among a large pool of neighborhood solutions. This method is characterized by inherent simplicity, high adaptability, and a short tem memory via a tabu list. This process avoids recycling and subsequently also allows the backtracking to previous solutions. Features of tabu list and aspiration make tabu search a powerful optimization tools for solving any combinatorial optimization problem. Aspiration criterion is a checking condition for the acceptance of a solution. In general, application of tabu search technique can be characterized by (1) generation of initial feasible solution, (2) neighborhood generation, (3) tabu list size, (4) aspiration level, and (5) stopping criterion. On the basis of the abovementioned characteristics features of tabu search, a generic pseudo code is given as:

procedure tabusearch
begin
select a current point, currentNode, at random
update NodesGenerated
bestNode ← currentNode
print
repeat
select a new node, newNode, that has the lowest distance in the neighborhood of currentNode that is not on the tabuList, using the two Interchange method

```
  update NodesGenerated for all nodes checked in
neighborhood of currentNode
      currentNode ← newNode
      if evaluation(currentNode) <
      evaluation(bestNode)
         bestNode ← currentNode
         print
  until some counter reaches limit
end
```

TEST BED

In the present section, we aim to analyze the performance of AI based random search algorithms over the machine loading problem. In order to analyze the effectiveness of the same, 10 benchmark dataset of the machine loading problem have been considered (Mukhopadhyay et al., 1992). Description of one such problem is shown in Table 1. Performance of all the evolutionary algorithms have been analyzed on the objectives of minimization of system unbalance, maximization of throughput, and the combination of both. These optimization techniques have been compared with the best results obtained by previously applied heuristic procedures. Computational results showing their comparison is given in Table 2. It can be visualized from the table that all AI-based tools have shown a great improvement in the results as compared to existing heuristic procedures on all problem instances. Further, a graph showing the objective function values over the 10 benchmark dataset has been plotted to reveal the performance of considered AI techniques (see Figure 2). It the figure, almost all the algorithms have shown equivalent performance in many problem instances, however, among them ant colony algorithm outperformed others by showing the highest objective function value in all problem instances.

In order to make the readers fully aware to the implementation aspect of considered algorithms over machine loading problem, the step-wise implementation details has been provided in the Appendix section.

Figure 3. Comparison of search algorithms on the objective function values (problem no. 1-10)

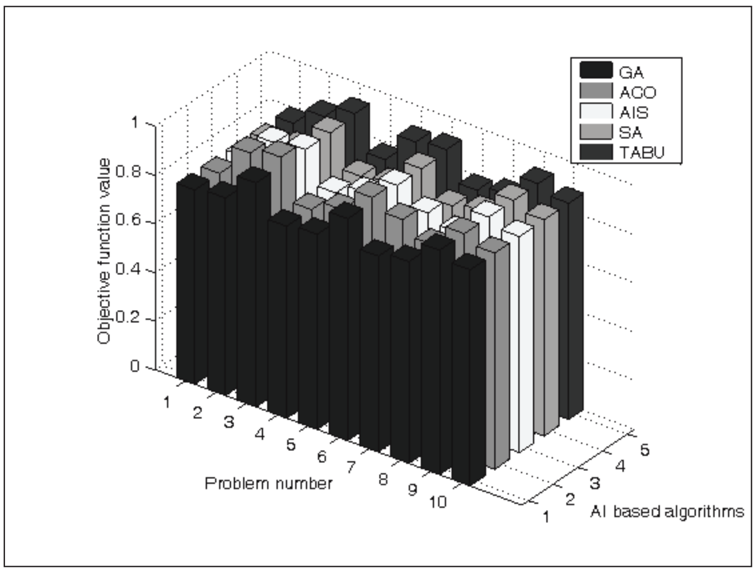


Table 1. Description of problem number 1 (Adopted from Shanker and Srinivasulu, 1989)

Job no.	**Batch size**	**Operation 1 Unit processing time/ available machine /tool slot required**	**Operation 2 Unit processing time/ available machine / tool slot required**	**Operation 3 Unit processing time/ available machine / tool slot required**
J_1	8	18/M_3/1	-	-
J_2	9	25/M_1/1	24/M_4/1	22/M_2/1
		25/M_4/1	-	-
J_3	13	26/M_4/2	11/M_3/3	-
		26/M_1/2	-	-
J_4	6	14/M_3/1	19/M_4/1	-
J_5	9	22/M_2/2	25/M_2/1	-
		22/M_3/2	-	-
J_6	10	16/M_4/1	7/M_4/1	21/M_2/1
		-	7/M_2/1	21/M_1/1
		-	7/M_3/1	-
J_7	12	19/M_3/1	13/M_2/1	23/M_4/3
		19/M_2/1	13/M_3/1	-
		19/M_4/1	13/M_1/1	-
J_8	13	25/M_1/1	7/M_2/1	24/M_1/3
		25/M_2/1	7/M_1/1	-
		25/M_3/1	-	-

CONCLUSION

Over the past few decades, artificial intelligence has produced a number of powerful tools. This chapter has reviewed five of those random search optimization tools namely, genetic algorithm, ant colony optimization, artificial immune system, simulated annealing, and tabu search. The applicability and robustness of such tools have facilitated the researchers and scientists in their analysis and experimentations. Here, we have analyzed the suitability of the aforementioned AI tools to solve one of combinatorial optimization problems, that is, machine loading in manufacturing domain. Results obtained from them are compared with the results of best heuristics applied previously to solve the machine loading problem. For all the benchmark dataset, AI tools have shown a great improvement in the results as compared to existing heuristics. The appropriate contribution of these AI based random search algorithms will continue contributing to the creation of more competitive engineering systems.

FUTURE RESEARCH DIRECTIONS

In the present chapter, authors have demonstrated the applicability of some well-known AI based random search optimization techniques to solve

Table 2. Comparison of the results obtained by AI techniques with best heuristic procedures

Pr.no	B.Sz	GA		ACO		AIS		SA		TABU		1		2		3	
		SU	TH	SU	TH	SU	TH	SU	TH	SU	TH	SU	TH	SU	TH	SU	TH
1	8	14	48	14	48	14	48	14	48	14	48	253	39	122	42	76	42
2	6	18	46	22	63	124	63	18	46	234	63	388	51	202	63	234	63
3	5	72	69	25	73	72	69	72	69	128	73	288	63	286	79	152	69
4	5	819	51	819	51	819	51	819	51	819	51	819	51	819	51	819	51
5	6	187	53	264	61	264	61	187	53	364	76	467	62	364	76	264	61
6	6	28	61	37	64	28	61	28	61	69	64	548	51	365	62	314	63
7	6	165	54	147	66	231	63	165	54	177	54	189	54	147	66	996	48
8	7	63	48	63	48	13	44	63	48	63	48	459	36	459	36	158	43
9	7	309	88	309	88	309	88	309	88	309	88	462	79	315	88	309	88
10	6	122	56	122	56	122	56	122	56	122	56	518	44	320	56	122	56

1— Shanker and Srinivasulu, (1989); 2—Mukhopadhyay et al., (1992) 3—Tiwari et al., (1997).

the machine loading problem in FMS. The scope of the present chapter can be widened by incorporating other loading devices such as pallets, fixtures and AGVs, and so forth. into the objective function. One can also include some other related objectives such as minimization of part movement and tool changeover time to generalize the model. Also, the present work can be extended by considering some dynamicity related to machine breakdowns, random processing time of the jobs, and so forth in the flexible manufacturing system. Thus, the future concerns of the present chapter will be to develop the stochastic modeling of the machine loading problem by incorporating machine and job disruptions related issues. Based upon the severity of the breakdown, the reliability of the FMS can be distinguished into three categories: (1) completely reliable FMS with no disruptions; (2) partially reliable FMS with known disruptions; and (3) completely unreliable FMS with random disruptions. In the presence of disruptions, an efficient technique is required that can easily map the situations and can provide some measures to improve the availability of the system. Since much of the information related to such dynamic situations is not known in the advance, thus, a "fuzzy" based approach would be suggested; yet another technique to resolve the complexity of machine loading problem in flexible manufacturing system when the information is vague and imprecise. Most of the AI based techniques dealt here have some limitations or drawbacks related to efficient exploration and exploitation of the search space. Therefore, some modification in the control operators of these techniques can be done to improve the robustness of the existing random search techniques

REFERENCES

Ada, G.L., & Nossal, G.J.V. (1987). The clonal selection theory. *Scientific American, 257,* 50-57.

Anand, R.B., Tiwari, M.K., & Shankar, R. (2006). A self-organized neural network metamodelling and clonal selection optimization-based approach for the design of a manufacturing system. *International Journal of Production Research*, *44*(6), 1147-1170.

Atmani, A., & Laskhari, R.S., (1998). A model of machine tool selection and operation alocation in FMS. *International Journal of Production Research*, *36*(5), 1339-1349.

Bandyopadhyay, S., Murthy, C. A., & Pal, S. K. (1995). Pattern classification using genetic algorithms. *Pattern Recognition Letters, 16,* 801-808.

Bandyopadhyay, S., & Pal, S. K. (1998). Incorporating chromosome differentiation in genetic algorithms. *Information Science, 104,* 293-319.

Bell, J.E., & McMullen, P.R. (2004). Ant colony optimization techniques for the vehicle routing problem. *Advanced Engineering Informatics*, *18,* 41-48.

Berrada, M., & Stecke, K. (1983). *A branch and bound approach for machine loading in flexible manufacturing system* (Working Paper No. 329). Ann Arbor, MI: Division of Research Graduates School of Business Administration, University of Michigan.

Chen, Y.J., & Askin, R.G. (1990). A multiobjective evaluation of flexible manufacturing system loading heuristics. *International Journal of Production Research*, *28*, 895-911.

Chen, I.J., & Chung, C.H. (1996) An examination of flexibility measurements and performance of flexible manufacturing systems. *International Journal of Production Research*, *34*(2), 379-394.

Clevland, G. A., & Smith, S. F. (1989). Using genetic algorithms to scheduling flow shop releases. In *Proceedings of the Third International Conference on Genetic Algorithm and Their*

Applications Palo Alto, California (pp. 160-169). Morgan Kaufmann.

Cox, L. A., Davis, L., & Qiu, Y. (1991). Dynamic anticipatory routing in circuit-switched telecommunications networks. *Handbook of genetic algorithms.* New York: Van Nostrand Reinhold.

Dasgupta, D. (1999). An overview of artificial immune systems and their applications. In D. Dasgupta (Ed.), *Artificial immune systems and their applications.* Berlin: Springer.

Davis, L. (1985a). Job shop scheduling with genetic algorithms. In *Proceedings of the International Conference on Genetic Algorithms and Their Applications,* Carnegie Mellon University, Pittsburgh, Pennsylvania (pp. 136-140).

Davis, L. (1985b). Applying adaptive algorithms to epistatic domains. In *Proceedings of the International Joint Conference on Artificial Intelligence* (pp. 162-164).

Davis, L. (1991). *Handbook of genetic algorithms.* New York: VanNostrand Reinhold.

De Castro, L.N., & Von Zuben, F.J. (2000). The clonal selection algorithm with engineering applications. In *Proceedings of Genetic and Evolutionary Computation Conference, Workshop on Artificial Immune Systems and Their Applications,* Las Vegas (pp. 36-37).

Dorigo, M., Caro, G.D., & Gambardella, L.M. (1997). Ant colony system: A cooperative learning approach to the travelling salesman problem. *IEEE Transactions on Evolutionary Computation, 1*(1), 53-66.

Dorigo, M., & Di Caro, G. (1999). The ant colony optimization meta-heuristic. In D. Corne, M. Dorigo, & G. Glover (Eds.), *New ideas in optimization* (pp. 11-32).

Endoh, S., Toma, N., & Yamada, K. (1998). Immune algorithm for n-TSP. In *Proceedings of the 1998 IEEE International Conference on Systems Man and Cybernetics,* San Diego, California (pp. 3844-3849).

Fourmann, M.P. (1985). Comparison of symbolic layout using genetic algorithm. In *Proceedings of the First International Conference on Genetic Algorithms and their Applications* (pp. 141-155). Hillsdale, NJ: Lawrence Erlbaum Associate.

Gaspar, A., & Collard, P. (2000). Two models of immunization for time dependent optimization. In *Proceeding of the 2000 IEEE International Conference on Systems, Man, and Cybernetics,* Nashville (pp. 113-118).

Glover, F. (1990). Tabu search: A tutorial. *Interfaces, 20*(4), 74-94.

Goldberg, D.E. (1987a). Computer aided gas pipeline operation using genetic algorithms and rule learning, Part 1: Genetic algorithms in pipeline optimization. *Engineering with Computers, 3*(1) 35-45.

Goldberg, D.E. (1987b). Computer aided gas pipeline operation using genetic algorithms and rule learning, Part 2: Rule learning control of a pipeline under normal and abnormal conditions. *Engineering with Computers, 3*(1) 47-58.

Goldberg, D.E. (1989). *Genetic algorithms in search, optimization and machine learning.* Reading, MA: Addison-Wesley.

Goldberg, D.E., Korb, B., & Deb, K. (1989). Messy genetic algorithm: Motivation analysis and first results. Complex System. *Complex System Publications,* 493-530.

Grefenstette, J. J., Gopal, R., Rormatia, B., & Vangucht, D. (1985). Genetic algorithms for travelling salesman problem In *Proceedings of the First International Conference on Genetic Algorithms and their Applications.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Guerreore, F., Lozano, S., Koltai, T., & Larraneta, J. (1999). Machine loading and part type selection in flexible manufacturing system. *International Journal of Production Research*, *37*(6), 1303-1317.

Gupta, M., Gupta, Y., Kumar, A., & Sundram, C. (1996). A genetic algorithm-based approach to cell composition and layout design problems. *International Journal of Production Research*, *34*(2), 447-482.

Holland, H.H. (1975). *Adaptation in natural and artificial systems*. Detroit: University of Michigan Press.

Hwang, S.S. (1986). *Models for production planning in flexible manufacturing system*. Unpublished doctoral thesis, University of California, Berkley.

Jerne, N.K. (1974). Towards a network theory of the immune system. *Annual Immunology (Inst. Pasteur)*, *125C*, 373-389.

Kim, J.Y., & Kim, Y.K. (2005). Multileveled symbiotic evolutionary algorithm: Application to FMS loading problems. *Artificial Intelligence*, *22*, 233-249.

Kim, Y.D., & Yano, C.M. (1994). A new branch and bound algorithm for loading problems in flexible manufacturing systems. *International journal of flexible manufacturing systems*, *6*, 361-382.

Kirkpatrick, F., Gelatte, C.D., & Vecchi, M.P. (Eds.). (1983). Optimization by simulated annealing. *Science*, *220*(4598), 671-680.

Kumar, A., Prakash, A., Tiwari, M.K., Shankar, R., & Baveja, A. (2006). Solving machine-loading problem of a flexible manufacturing system with constraint-based genetic algorithm. *European Journal of Operational Research*, *175*(2), 1043-1069.

Kumar, N., & Shanker, K. (2000). A genetic algorithm for FMS part type selection and machine loading. *International Journal of Production Research*, *16*(38), 3861-3887.

Kumar, P., Singh, N., & Tiwari, N.K. (1990). Multicriteria analysis of the loading problem in flexible manufacturing system using max-min approach. *International Journal of Production Research*, *2*(2), 13-23.

Kusiak, A. (1985). Flexible manufacturing systems: A structural approach. *International Journal of Production Research*, *23*(6), 1057-1073.

Lashkari, R.S., Dutta, S.P., & Padhye, A.M. (1987). A new formulation of operation allocation problem in flexible 1258 manufacturing systems: mathematical modeling and computational experience. *International Journal of Production Research*, *25*(9), 1267-1287.

Lee, C.Y., Piramuthu, S., & Tsai, Y.K. (1997). Job shop scheduling with genetic algorithm and machine learning. *International Journal of Production Research*, *35*(4), 1171-1191.

Liang, M., & Dutta, S.P. (1993). An integrated approach to part- selection and machine-loading problem in class of flexible manufacturing systems. *European Journal of Operational Research*, *67*(3), 387-404.

Maniezzo, V., Colorni, A., & Dorigo, M. (1999). The ant system applied to the quadratic assignment problem. *IEEE Transactions on Knowledge and Data Engg*, *11*(50), 769-778.

Maniezzo, V., Dorigo, M., & Colorni, A. (1996). Ant system: Optimization by a colony of cooperating agents. *IEEE Transaction on Systems, Man and Cybernatics*, *26*(1), 29-41.

Michaelwicz, Z. (1992). *Genetic algorithm + data structure = evaluation programs*. New York: Springer Verlag.

Mukopadhyay, S.K., Midha, S., & Krishna, V.A. (1992). A heuristic procedure for loading problem in flexible manufacturing system. *International journal of production research*, *30*(9), 2213-2228.

Mukhopadhyay, S.K., Singh, M.K., & Srivastava, R. (1998). FMS loading: A simulated annealing approach. *International Journal of Production Research, 36*(6), 1629-1647.

Nayak, G.K., & Acharya, D. (1998). Part type selection and machine-loading and part type volume determination problem in FMS planning. *International Journal of Production Research, 36*, 1801-1824.

Nossal, G.J.V. (1994). Negative selection of lymphocytes. *Cell, 76*, 229-239.

Prakash, A., Khilwani, N., Tiwari, M.K., & Cohen, Y. (in press). Modified immune algorithm for job selection and operation allocation problem in flexible manufacturing system. *Advances in Engineering Software.*

Rai, R., Kameshwaran, S., & Tiwari, M.K. (2002). Machine-tool selection and operation allocation in FMS: Solving a fuzzy goal-programming model using a genetic algorithm. *International Journal of Production Research*, *40*(3), 241-265.

Ram, B., Sarin, S., & Chen, C.S. (1990). A model and solution approach for the machine loading and tool allocation problem in FMS. *International Journal of Production Research*, *28*(4), 637-645.

Sarin, S.C., & Chen, C.S. (1987). The machine loading and tool allocation problem in a flexible manufacturing system. *International Journal of Production Research, 25*(7), 1081-1094.

Sarma, U.M.B.S., Kant, S., Rai, R., & Tiwari, M.K. (2002). Modeling the machine loading problem of FMSs and its solution using a tabu-search-based heuristic. *International Journal of Computer Integrated Manufacturing, 15*(4), 285-295.

Sawik, T. (1990). Modeling and scheduling of flexible manufacturing systems. *European Journal of Operational Research*, *45*(2-3), 177-190.

Shanker, K., & Srinivasulu, A. (1989). Some solution methodologies for a loading problem in random FMS. *International Journal of Production Research*, *27*(6), 1019-1034.

Shanker, K., & Tzen, Y.J. (1985). A loading and dispatching problem in a random flexible manufacturing system. *International Journal of Production Research*, *16*(5), 383-393.

Singhal, K. (1978). Integrating production decisions. *International Journal of Production Research*, *16*(5), 383-393.

Stecke, K. (1983). Formulation and solution of non-linear integer production planning problems for flexible manufacturing system. *Management Sciences*, *29*, 272-288.

Stecke, K.E. (1986). A hierarchical approach to solving grouping and loading problems of flexible manufacturing systems. *European Journal of Operational Research*, *24*(3), 369-378.

Stecke, K.E. (1993). Planning and scheduling approaches to operate a particular FMS. *European Journal of Operational Research*, *61*(3), 273-291.

Stecke, K., & Solberg, J. (1981). Loading and control policies for a flexible manufacturing system. *International Journal of Production Research, 19*, 481-490.

Stuetzle, T., & Hoos, H.H. (2000). Max-min ant system. *Future Generation Computer Systems, 16*, 889-914.

Szumlanski, S.R., Wu, A.S., & Hughes, C.E. (2006, July 16-20). Conflict resolution and a framework for collaborative interactive evolution. *American Association for Artificial Intelligence (AAAI).* Boston.

Taillerd, E. (1990). Some efficient heuristic methods for flow shop sequencing problems. *European Journal of Operational Research*, *47,* 65-74.

Tiwari, M.K., Hazarika, B., Vidyarthi, N.K., Jaggi, P., & Mukopadhyay, S.K. (1997). A heuristic solution approach to the machine loading problem of FMS and its petri net model. *International Journal of Production Research*, *35*(8), 2269-2284.

Tiwari, M.K., Kumar, Sa., Kumar Sh., Prakash, A., & Shankar, R. (2006). Solving part type selection and operation allocation problems in a FMS: An approach using consraint based fast simulated annealing algorithm. *IEEE Transaction on Systems, Man, and Cybernatics Part-A, 36*(6), 1170-1184.

Tiwari, M.K., & Swarnkar, R. (2004). Modeling machine loading problem of FMSs and its solution methodology using a hybrid tabu search and simulated annealing-based heuristic approach. *Robotics and Computer-Integrated Manufacturing, 20*, 199-209.

Tiwari, M.K., & Vidyarthi, N.K. (2000). Solving machine loading problems in a flexible manufacturing system using a genetic algorithm based heuristic approach. *International Journal of Production Research*, *38*(14), 3357-3384.

Van Looveren, A.J., Gelders, J.L.F., & Van Wassenhove, L.N. (1986). A review of FMS planning models. In A. Kusiak (Ed.), *Modeling and design of flexible manufacturing systems* (pp. 3-31). Amsterdam: Elsevier.

Whitney, C.K., & Gaul, T.S. (1984). Sequential decision procedure for batching and balancing in FMSs. *In Proceedings of the First ORSA/TIMS Conference on FMS* (pp. 243-248). University of Michigan.

ADDITIONAL READINGS

A more complete discussion on the flexible manufacturing systems (FMS) and its related issues can be found in:

Askin, R.G. (Ed.) *Modeling and analysis of manufacturing systems.* John Wiley & Sons Inc.

Groover, M.P. (Ed.). (2001). *Automation, production systems, and computer integrated manufacturing* (2nd ed.). Pearson Education Inc.

Kusiak, A. (Ed.) (1990). *Intelligent manufacturing systems.* Englewood Cliffs, NJ: Prentice-Hall.

Ratna, J., & Tchijov, I. (1990). Economics and success factors of flexible manufacturing systems: The conventional explanation revisited. *International Journal of Flexible Manufacturing Systems*, *2*, 169-190.

Reingold, E.M., Nievergelt, & Deo, N. (Eds.). (1977). *Combinatorial algorithms: Theory and practice*. Englewood Cliffs, NJ: Prentice Hall.

Viswanadham, N., & Narahari, Y. (Eds.). (1998). *Performance modeling of automated manufacturing systems.* India: Prentice-Hall Inc.

For the clear comprehension of the various pseudo-codes shown in this chapter and elsewhere, interested readers may refer to:

Belegundu, A.D., & Chandrupatla, T.R. (Eds.). (2002). *Optimization concepts and applications in engineering.* Pearson Education Inc.

Even, S. (Ed.). (1973). *Algorithmic combinatorics.* New York: Macmillan Company.

Horowitz, E., & Sahani, S. (Eds.). (1993). *Fundamentals of computer algorithms*. Galgotia Publications Inc.

Michalewicz, Z. (Ed.). (1994). *Genetic algorithm + data structure = evolution programs.* New York: Springer-Verlag.

Narahari, Y. (Ed.). (2000). *Data structures and algorithms* (Electronic Lecture Notes).

For the basic concepts of the various AI based random search optimization techniques and their implementation over some well known problem set, interested readers may refer to:

Gen, M., & Cheng, R. (Eds.). (1997). *Genetic algorithm and engineering design*. New York: John Wiley & Sons.

Gen, M., & Cheng, R. (Eds.). (1999). *Genetic algorithm and engineering optimization*. New York: John Wiley & Sons.

Wang, J., & Kusiak, A. (Eds.). (2001). *Computational intelligence in manufacturing.* London: CRC Press.

To cover the issues related to future concerns of the present chapter, refer to:

Aytug, H., Lawley, M.A., McKay, K., Mohan, S., & Uzsoy, R. (2003). Executing production schedules in the face of uncertainties: A review and some future directions. *European Journal of Operation Research*, *161*(1), 86-110.

Behle, U., & Kusiak, A. (1997). Dynamic scheduling of design activities with resource constraints. *IEEE Transactions on Systems, Man and Cybernatics-A*, *27*(1), 105-111.

Cai, X., Wu, X., & Zhou, X. (2005). Dynamically optimal policies for stochastic scheduling subject to preemptive-repeat machine breakdowns. *IEEE Transactions on Automation and Engineering*, *2*, 158-172.

Mehta, S.V., & Uzsoy, R.M. (1998). Predictable scheduling of a job shop subject to breakdowns. *IEEE Transactions on Robotics and Automation*, *14*(3), 365-378.

Savsar, M. (2005). Performance analysis of FMS operating under different failure rates and maintenance policies. *International Journal of Flexible Manufacturing System*, *16*, 229-249.

Tenenbaum, A., & Seidmann, A., (1989). Dynamic load control policies for a flexible manufacturing system with stochastic processing rates. *International Journal of Flexible Manufacturing Systems*, *2*, 93-120.

Yamamoto, M., & Nof, S.Y. (1985). Scheduling/rescheduling in the manufacturing operating system environment. *International Journal of Production Research*, *23*, 705-722.

APPENDIX

Step-wise implementation details of AI-techniques with sample illustration to solve machine loading problem is given as:

1. Genetic Algorithm

Step 1: Initialize the Control Parameters

Based on the preruns, GA control parameters have been initialized as population size (Pop=4), crossover probability (XPr=0.4), mutation probability (MPr=0.1), and maximum number of generations (Max_gen=30).

Step 2: Initialization

As per the GA procedure (given in the respective section), initialize the population (each population is equivalent to the sequence of the jobs). Let the four populations be initialized as:
Pop1= [5 7 6 8 1 4 3 2]; Pop2= [2 7 5 4 6 3 1 8]; Pop3= [1 5 2 3 6 4 8 7]; Pop4= [7 5 8 4 2 3 1 6]

Step 3: Evaluation

Similar to the procedure given in section 2.3. Let, we select pop1= [5,7,6,8,1,4,3,2]; Then overall evaluation procedure will be: {Job 5 (op 1, m 3, t 2), (op 2, m 2, t 1)}; {Job 7 (op 1, m 4, t 1) (op 2, m 1, t 1) (op 3, m 4, t 1)}; {Job 6 (op 1, m 4, t 1) (op 2, m 2, t 1) (op 3, m 1, t 1)}{Job 8 (op 3, m 1, t 1 (**job rejected due to tool slot constraint TSC**))} {Job 1 (op 1, m 3, t 1)} {Job 4 (op 1, m 3, t 1) (op 2, m 4, t 1)} {Job 3 (op 2, m 3, t 1 (**Job rejected due to negative system unbalance NSU**))} {Job 2 (op 2, m 4, t 1 (**Job rejected due to negative system unbalance NSU**))}. SU = 55 and TH = 45 Obj. fun value 0.96

Step 4: Crossover

Here, two-point crossover is applied on the two parents randomly selected, that is, Parent1= [5 4 3 7 1 6 8 2]; Parent2 = [1 5 2 3 6 4 8 7]. The children produced by it namely child1= [4 5 7 3 1 6 8 2]; Child2= [1 4 2 7 6 5 8 3]. Similarly this process will be repeated for rest of the parent population.

Step 5: Mutation

Let parent1= [2 7 5 4 6 3 1 8] be selected for the mutation process, thus resulting in Child1 = [2 7 5 4 6 3 8 1].

Step 6: Evaluation and Selection

With the crossover and mutation operator, the overall population size will become twice of the population size; the overall population is sorted in decreasing order of their fitness value. Among which the best candidate is selected.

Step 7: Stopping Criteria and Results

Previous steps are repeated until the stopping criterion (i.e., maximum number of generations) is not satisfied. After the completion of final generation the population with best fitness value is found.

2. Ant Colony Optimization

Step 1: Initialize the Control Parameters

ACO control parameters have initialized as total number of ants (N=total number of operations), relative importance of trail (α=1.0), relative importance of visibility (β=1.0), and pheromone evaporation rate (ρ=0.5).

Step 2: Initialization

a. Problem representation using a weighted directed graph (Vertices – Jobs, Edge- Difference between total processing times of the jobs).
b. Randomly distribute ants on the node
c. Set t =0 (time counter)
d. Set NC =0 (NC is the number of counter)
e. Set $tabu^k$ =0 (tabu represent the list of nodes traversed by ant k)
f. Select increase in trail level equal to zero

If NC>NC_max go to Step 8, otherwise proceed.

Step 3

If k>k_max, go to Step 7, otherwise proceed. (k_max is the maximum number of ants)

Step 4

If tabu list of ant k is full (i.e., $tabu^k$>=$tabu^k$_max) go to Step 6, otherwise proceed.

Step 5: Node selection

a. Generate random number p (0<=p<=1).
b. If p>=p_0, proceed, otherwise go to Step 5(d).

c. Compare the probability of possible outgoing nodes. (From equation 11)
d. Choose the node (i.e., job) having the highest probability.
e. Add the node to $tabu^k$ and remove it from further consideration

Step 6

k= k+1, go to Step 3

Step 7: Updating

a. Find p^+_{iter}; (p^+_{iter} is the objective function of the best solution found)
b. If $p^+_{iter} < p^+_{best}$ then $p^+_{iter} = p^+_{best}$ (p^+_{iter} is the best objective function value)
c. Update Pheromone: $\tau_{il}(t)$
d. Empty all tabu lists
e. NC=NC+1 (increase the counter number)
f. $p_o = \log(NC) / \log_e(N\_\max)$, go to Step 2

Step 8: Output:

p^+_{best} (best objective function value)

3. Artificial Immune Algorithm

Step 1: Initialization

Problem specific objective function is defined (see Figure 2).
A set of randomly generated initial population of antibody is maintained (encoding scheme is same as was used is GA).

Step 2: Fitness Evaluation

Here initial population is exposed to threats posed by antigens.

- Randomly choose an antigen (Ag_j Є Ag_m) and expose it to all antibody.
- Determine the affinity f_k (i.e., objective function, see figure 2) of the entire antibody in relation to Ag_j.

Step 3: Proliferation

Here interaction between antibodies is carried out for analyzing the relation between them.

- Select n highest affinity antibody. Make a new set of selected antibody in relation to Ag_j.
- Selected antibody will be cloned independently and proportionally to their antigenic affinity, C_k.

Step 4: Hypermutation

It generates the population C_k* of matured clones. (Higher the affinity, lower the mutation rate.)

- Determine the affinity f_k* of the matured clones. Now all the antibodies are selected to compose the memory.

Step 5: Selection

Choose the best antibodies Ab_d by replacing *d* lowest-affinity antibody.
Repeat the process for pre defined number of generation.

4. Simulated Annealing

Step 1: Initialization

Encode the Initial solution as defined in GA

- Set initial temperature $t = T_{max}$.
- Set final temperature $T = 0$.
- Randomly generate the initial solution (see the implementation steps for GA).
- Calculate the fitness value for initial solution (see Figure 2).
- Let the solution = S_1.

Step 2: Exploration

Diversification of solution space using Gaussian random variable

- Generate new solution using Gaussian random variable.
- Let the solution with fitness value S_2.

Step 3: Exploitation

Calculate $\Delta S = S_1 - S_2$
- If $\Delta S \geq 0$ (maximization case), expect the new solution for next generation.
- If $\Delta S < 0$, select the solution with a probability.
- Reduce the temperature as $t = \lambda * t,\ \lambda \in [0,1]$

Repeat the above procedure (from **Step 2**) for predefined number of generations.

5. Tabu Search

Step 1: Generation of Initial Feasible Solution

a. Number of iterations (i= 0)
b. Tabu list (TL={Φ, Φ, Φ}) (Φ is showing null component)
c. Best job sequence obtained is equal to best job sequence in candidate solution (S_b= S_c).
d. Best job sequence in candidate solution is equal to best part sequence of initial solution (S_c=S_o).
e. Objective function of current solution (F_1), best objective function from the candidate solution (F_c) and objective function of initial feasible solution (F_o) are set equal to each other. (F_1=F_c=F_o).

Step 2: Neighborhood Generation

a. Generate neighborhood of best job sequence in the candidate solution. (In the underlying problem, neighborhood is generated by replacing the unassigned jobs with assigned jobs in the sequence)
b. For each set of generated neighborhood, evaluate the objective function of job sequence F(S).
c. Compare the objective function of job sequence F(S) with objective function of initial feasible solution F_1.
d. If F(S) > F_1 go to Step 2(f), otherwise proceed
e. Check whether the sequence belongs to tabu list (TL) that is, (S ε TL), if yes proceed, otherwise go to step
f. Compare the objective function of job sequence in candidate solution F(Sc) with aspiration level of job sequence encountered so far A(S).
g. Check for the remaining solution, if yes go to Step 3(b), otherwise proceed.

Step 3: Aspiration Level

a. Include job sequence of current solution in tabu list (TL) and set aspiration level of current solution equal to the objective function of job-sequence in candidate solution. A(S_1)= F(Sc).
b. Compare objective function of current solution (F_1) with objective function of best job sequence encountered so far F(Sc). That is, if (F(S_b)> F_1), then set job sequence of best solution equal to job sequence of current solution (S_b= S_1).

Step 4: Stopping Criteria

a. Check for the stopping criteria, if satisfied proceed, otherwise set job sequence of current solution (S_1) equal to the job sequence of candidate solution (Sc), increase the iteration counter (i=i+1) and go to Step 2(a).
b. Evaluate the objective function (procedure given in figure 2) corresponding to best found sequence (S_b) in the search.

Chapter III
Computational Intelligence in the Financial Functions of Industrial Firms

Petros Theodorou
Athens University of Economics and Business, Greece

Dimitrios Karyampas
University of York, UK

ABSTRACT

Information technology has been proved to be a strategic weapon in the business armory for the creation and sustention of competitive advantage, especially, when it is aligned with the needs of the internal and external environment. Solutions are provided from the operational level up to strategic planning and are capable to support every choice in the strategy portfolio, from cost to quality and flexibility. IT systems in the manufacturing and operational level were analyzed extensively in literature: ERP systems, computer aided design/computer aided manufacturing (CAD/CAM), and so forth. According to Wong, Bo, Bodnovich, and Selvi (1997), 53.5% of the reviewed literature in artificial intelligence refers to applications in production and operations management. Nevertheless, the second most important area for advanced IT applications is that of finance (25.4%). This research will be focused on the common set of the two previously mentioned areas: production management and the necessary financial tools. Production and operation management requires specific financial tools in order to accomplish the functions of production planning, costing, investment appraisal, and so forth. Computational intelligence in those financial functions is mostly needed for the production operation department and for the production operation strategy. Specifically, the weight will be put on information technology automation of financial functions adopted by production departments: forecasting production needs, production planning and control, profit volume analysis, cost analysis, investment appraisal analysis, and so forth. An attempt will be made to classify the various quantitative and qualitative techniques in relation to various fi-

nancial aspects. Specifically, advances of neural networks, expert systems, advanced statistical analysis and operational research methods, and various hybrid techniques will be presented in relation to financial models applied in production. Financial applications will be analyzed according to their modules and their outputs in a strategic alignment concept. Finally, a strategic alignment model will be derived for the adoption of financial applications in businesses.

INTRODUCTION

A tremendous progress in production methods happened in the last decade. The new production models customer and financially oriented incorporate new quantitative and qualitative techniques integrated with the known production and operations management models. The "black processing box" of this new financially oriented model incorporates advanced computational intelligence techniques. Production is not restricted on the shop floor management; instead a market oriented approach along with financial functions for the increase of financial performance is a prerequisite for the strategic survival. Computational intelligence employed in those financial models includes techniques of advanced statistics (mainly time series with exceptions, like discriminant analysis for the credit risk evaluation), simulation of stochastic processes, and artificial and neural network models. Logit-Probit models, multivariate discriminant analysis, simulation techniques (Monte-Carlo), weighted moving average (WMA), autoregressive conditional heteroskedasticity-generalized autoregressive conditional heteroskedasticity (ARCH-GARCH), and so forth are some of the techniques included in the statistics toolbox. Moreover, techniques of artificial intelligence and neural networks include case base reasoning, genetic algorithms, genetic programming, heuristic methods of linear programming and neural optimization, and so forth. Machine learning techniques are applied in portfolio optimization and derivatives pricing. Furthermore, those systems try to estimate risks in order to predict bankruptcy and rate credibility. The financial oriented production model targets to optimum allocation of funds among production activities, and to further hedge operational risks with financial impact. The diversification in demand and the variability in the external environment increased the need for rapid response with diversification in product and production. Therefore, computational intelligence must be incorporated in order to facilitate flexibility (Theodorou, 1996).

Manufacturing companies organize their systems in order to manage their operations on the spot and hedge the risks in the secondary markets. The operation in those markets requires extensive computational intelligence. The integration of financial information systems (FIS) with the production management systems is increased in order to gain competitive advantage. The performance of the advanced FIS should be measured in relation to the strategic factors of quality, flexibility, dependability, and valuate the benefits of scope economies and lead-time shortenings under the strategic alignment perspective. In the following paragraphs the generic FIS will be presented along with the literature review concerning its components. Specific attention will be given on the basic quantitative processing techniques which are based on statistics, artificial intelligence and neural networks. Finally, an attempt will be made to integrate the generic FIS within the strategic alignment model for future research.

LITERATURE REVIEW

Financial information systems (FIS) are usually found as a subtopic of the accounting information systems, but they must be separated due to differences in principles and practices. Especially, the quantitative character of finance demands a completely different approach than that of the accounting information systems, even though forecasting financial statements and budgeting analysis uses inputs from accounting information systems (AIS) and management accounting information systems (MAIS) (Covaleski & Dirsmith, 1986). AIS feed the systems with data and information for financial statements forecasting, and MAIS with information regarding the cost categories (Abernethy & Brownell, 1997; Abernethy & Stoelwinder, 1996; Bruggeman & Slagmulder, 1995; Chenhall & Langfield-Smith, 1998; Chenhall & Morris, 1986; Choe, 2004; Cooper, Hayes, & Wolf, 1981; Cooper & Kaplan, 1992, 1991; Cooper & Suver, 1994; Feltham & Xie, 1994; Flamholtz & Das, 1985; Foster & Swenson, 1997; Govindarajan, 1984; Ittner, Lanen, & Larcker, 2002; Johnson, 1992; Kaplan & Norton, 1992; Karmarkar, Lederer, & Zimmerman, 1990; Krumwiede, 1998; Lawrence, 1990; McNair, 1990; Miller, 1992; Preston, 1992; Shank & Govindarajan, 1993; Simons, 1987). Those systems must work close but each is separate from the other in scope and practice. The black box of FIS includes quantitative analysis, fundamental analysis, and technical and fixed income analysis. Stochastic processes and portfolio analysis make extensive use of artificial intelligence and neural networks where extensive literature can be found. Financial information systems use various quantitative techniques that are employed to model credit ratings and sovereign ratings, to evaluate credit risk, to forecast failure, bankruptcy and financial risk, to model stock selection, and so forth. All of these models are employing statistical and artificial intelligence techniques. Some of the statistical techniques employed are linear and nonlinear regression, probit regression, logit analysis, linear or quadratic multivariate discriminant analysis, multidimensional scaling, simulation techniques, logistic regression and linear programming, WMA, ARCH-GARCH, and various transformations (Altman, 1968; Baker & Wurgler, 2004; Belkaoi, 1980; Black, Jensen, & Scholes, 1972; Cambell & Shiller, 1988; Ederington, 1985; Fama & French, 1998, 2001; Hodrick, 1992; Horrigan, 1996; Jobson, 1982; Keim & Stambaugh, 1986; Kothari & Shanken, 1997; Kothari & Warner, 2001; Lewellen, 2004; Mar, Apellaniz, & Cinca, 1996; Mcinish & Wood, 1992; Taffler, 1983; Trevino & Thomas, 2000a, 200b). Artificial intelligence techniques consists of neural networks, case base reasoning, genetic algorithms, genetic programming ,and so forth (Ahn, Cho, & Kim, 2000; Bennell, Crabbe, Thomas, & Gwilym, 2006; Bennell & Sutcliffe, 2004; Bentz & Merunka, 2000; Brown, Coakley, & Phillips, 1995; Chen & Leung, 2005; De Freitas, Niranjan, Gee, & Doucet, 2000; Geigle & Aronson, 1999; Gemela, 2001; Healey, Dixon, Read, & Cai, 2002; Kelly, 1994; Kim, 2006; Mahfoud & Mani, 1996, 2002; Malliaris & Salchenberger, 1993a, 1993b; Odom & Sharda, 1990; Qi & Maddala, 1996; Tsakonas, Dounias, Doumpos, & Zopounidis, 2006). Risk management and financial forecasting make use of neural networks and expert systems advances as well as advanced statistics and operational research (Zhang, Patuwo, & Hu, 1998). In the coming sections a more detailed analysis will be presented. Fundamental analysis is basically based on financial statements and ratio analysis. From the basic outputs of accounting information systems, expost budgeting analysis forecast financial statements for the coming years (proforma). Technical analysis is trying to describe and forecast the movement of the prices on the markets using techniques such as trendlines, channels, candlesticks, point and figure, indicators and oscillators, various moving average methods (simple, linear, exponential), ARCH, GARCH, and so forth. Moreover, the state of efficiency according to efficient market hypotheses determines the type of analysis as well as the random walk hypotheses. A more detailed presentation will follow in the coming sections regarding the quantitative methods and its use in financial modeling. All these models should be examined in relation to the level of detail, frequency of reporting, and so forth under the alignment perspective (Figure 3). In the following paragraphs a generic FIS presentation will be made and its components will be discussed in detail along with relevant literature and their application field.

THE GENERIC MODEL

The generic model in an abstracting mode can be presented in the following diagram.

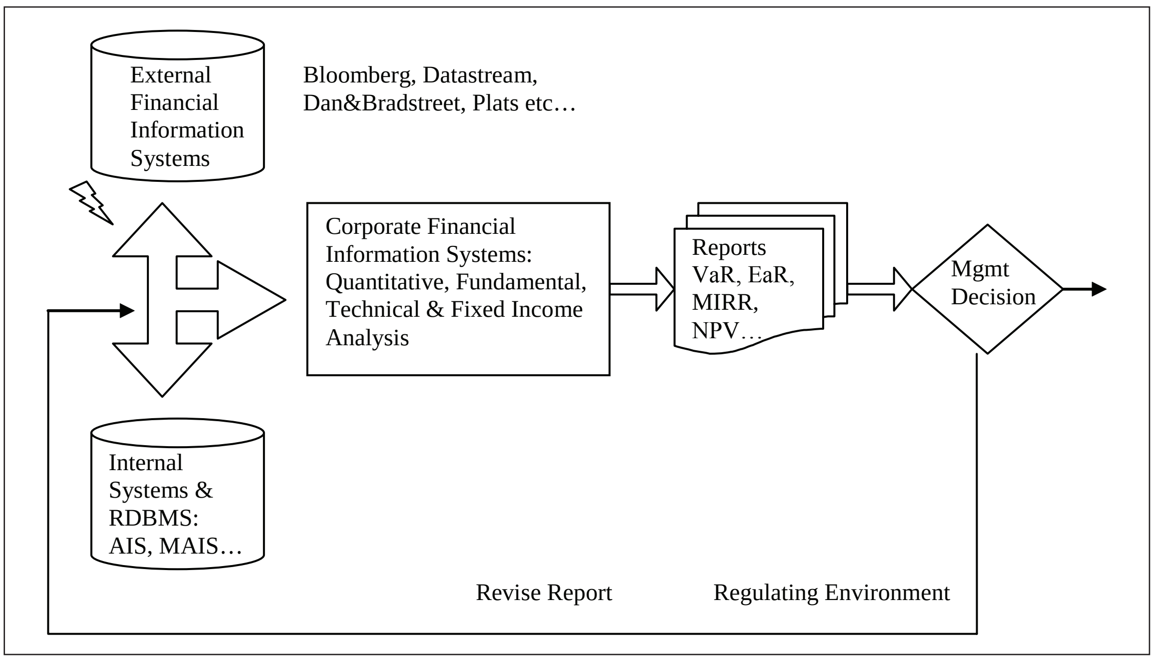
In this model a low level of detail is kept for the sake of discussion upon the generic topics. Further detail can be provided in future research for specific applications in specific industrial environment. This model should be adapted in the strategic alignment perspective (Figure 3) and should be discussed by taking in account relevant contingencies (Theodorou, 2003, 2004, 2005). The components of the generic FIS will be discussed in the following paragraphs.

The Internal RDBMS Systems

The internal relational database management system (RDBMS) include all available information and data generated and monitored within the firm by the firm's operations. Those systems comprise the internal pool of information separated conceptually (based on different economic principles) but interdependent and sometimes integrated from the software point of view. Those systems are the accounting information systems (AIS) and the costing or management accounting information systems (MAIS).

The accounting information systems generally include modules of accounts receivable-payable,

Figure 1. Generic FIS



generally ledger, payroll (sometimes separated module), order entry, billing, fixed assets accounting, and income tax preparation. The basic outputs of AIS used as inputs to the financial information systems are the income statements, the balance sheets, the cash flow statement, and the statement of changes in equity (based on the international financial reporting standards (IFRS)). Accounts receivable and payable record invoice and billing and helps schedule payments and issuing checks. Produce aging reports for cash collection that is also found in order entry applications. Payroll systems include pay rate, vacations hours tax, and other deductions and tax reports. Fixed asserts systems monitor and report the purchase of buildings, vehicles, and equipment for depreciation calculations and tax treatment, the gain or loss on assets sale, and so forth. Some systems are integrated with customer relationship management, e-commerce, project management, sales force automation, and work order management (*http://www.findaccounting-software.com/software/search/form.aspx?submittedfor=1&industry=Utilities&x=62&y=7* and *http://en.wikipedia.org/wiki/Comparison_of_accounting_software*).

The management accounting Information system (MAIS) is mainly dedicated in cost and inventory control and planning. The system collects, classify, and reports information that assists in financial planning and control of production activities (Bruggeman & Slagmulder, 1995; Flamholtz & Das, 1985). The system provides expost data for performance evaluation against predetermined goals and standards by the financial system. The financial part of MAIS can be grouped into traditional cost control information systems (TCCI) and advanced cost control information systems (ACCI) (Abernethy & Brownell, 1997; Chenhall et al., 1998; Govindarajan, 1984). One form of the advanced cost control information system is the design of activity based costing (ABC) (Foster & Swenson, 1997; Ittner et al., 2002; Krumwiede, 1998). TCCI systems are managing cost by means of standards, variances, and other metrics based at the individual level (Miller, 1992). In TCCI organizational performance is increased by maximizing individual efficiency (McNair, 1990) in relation to ACCI systems where interrelationships among function and group cooperation is of higher importance for performance (Choe, 2004). The critical attributes of those systems are the level of detail, the ability to disaggregate costs according to behavior, the frequency with which information is recorded, and the extent of calculation of variances. The more functional the type of MAIS the greater detail it provides, the better behavior classification it provides, as well as frequent reporting, and calculation of variances. The level of detail is analyzed by Chenhall and Morris (1986), Felthan (1977), Kaplan and Norton (1992), and Karmarkar et al. (1990). The system, in order to supply detail, must separate and classify cost according to behavior in fixed and variable, direct and indirect, controllable and noncontrollable (Feltham & Xie, 1994), and so forth.The greater the detail of MAIS, the greater the flexibility that it generates to analyze the cost for different purposes (Shank & Govindarajan, 1993). Feltham (1977) as well as Chenhall and Morris (1986) point out that decisions that are based on more detailed information are capable for performance increase in relation to decisions which are based on more aggregated information due to accuracy. Aggregations can be made in relation to the procedures, where direct costs are traced to the procedure and indirect, fixed, and variable costs allocated to the procedures. This requires a meaningful classification of cost according to behavior and activities (activity based costing system) (Cooper & Kaplan, 1992). The correct identification of cost behavior ensures the accurate information at all levels of detail (Cooper & Kaplan, 1991; McGown, 1998).

Cost reporting frequency enables the early recognition of problems in order for management to take corrective actions and accomplish fit with the environmental changes (i.e., takes advantage of opportunities and avoids threats). Chenhall and Morris (1986) found that frequency has a positive impact on performance for firms that operate in uncertain environment. Frequent reporting of the MAIS to the financial modeling system has found to have positive impact on financial performance. Hilton (1979) modeled the value of an MAIS in a cost volume profit decision setting where higher frequency of reporting increased the performance of the system.

The cost system trait, through variance analysis, determines the gap among expost and exante financial proforma estimations in order to resolve the causes of difference (Cooper & Suver, 1994; Feltham & Xie, 1994; Hilton, 1979; Johnson, 1992; Karmarkar et al., 1990; Khandwalla,1972; Simons, 1987). Variance analysis feed the financial information system with appropriate data in order to achieve budgetary control and help the administrative monitoring. Furthermore, inelastic pricing and contracts increase the need for variance analysis as they force firms to adopt risk of unexpected cost and utilization. Finally, variance analysis is extremely useful for the appropriate allocation of resources through measuring performance in relation to predetermined targets (Abernethy & Stoelwinder, 1996; Cooper et al., 1981; Covaleski et al., 1993; Lawrence, 1990; Ouchi, 1979; Preston, 1992).

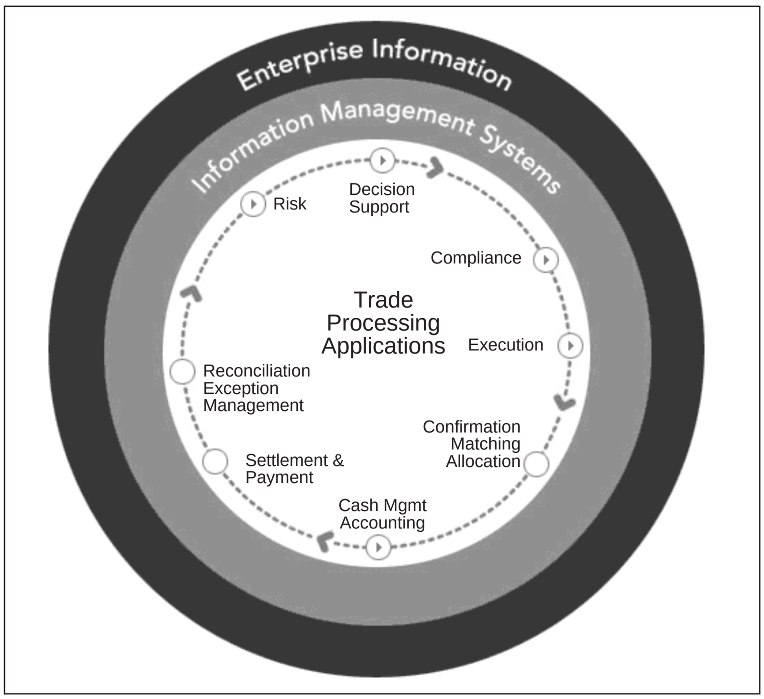
The External Financial Information Systems

Market data that will feed the financial information system are obtained through external database management system (DBMS). Those external financial information systems keep track of market data, like prices, indices, interest rates, and other macroeconomic data as well as published information of financial accounts of other firms. Some of those databases need registration while others are free of charge. Examples are Bloomberg, Datastream, Hoovers, Yahoofinance, Plats, Edgar, Ibbotson, Reuters, and so forth. In this category we can refer interbank systems like Society for Worldwide Interbank Financial Telecommunication (SWIFT) as well as stock exchange information systems provided from various brokers. Those systems monitor and report the results of the stock exchange markets. The Electronic Data Gathering Analysis and Retrieval System (EDGAR) automates the collection, validation, indexing, and forwarding of data for companies required by the law with Securities and Exchange Commission (SEC). EDGAR, among others, delivers fundamental data, global annual reports, aggregated information, SEC fillings plus comparison, and searching and screening tools. Moreover, EDGAR with I-Metrix Professional offers analytical tools for fundamental, market data, and earnings forecasting, peer analysis, benchmarking, valuation, and so forth. Ibbotson, founded in 1977, offers tools and information for asset allocation, investment management, forecasting, education, and National Association of Securities Dealers (NASD)-reviewed presentation material. Presentation material is used to explain and demonstrate asset allocation strategies and investment concepts. Ibbotson offers software and data for investment planning, analysis, and asset allocation. EnCorr Analyzer accesses a central database to perform historical capital market data analysis, while EnCorr Input Generator refines and generates the necessary forecasting and optimization analysis. EnCorr also has an optimizer to test and analyze portfolios on frontiers, created using resampling methodology. EnCorr Attribution is used to analyze management style and attribute performance to investor decisions. EnCorr Allocator is used to implement asset allocation policy and Ibbotson Scenario builder is used toanalyze what-if scenarios. These database offers access to 3500 domestic and international macroeconomic indices and performance data series with monthly and quarterly Internet updates since 1926. Investment Planning Software and Data includes the Portfolio Strategist and Analyst, the Security Classifier, and the Investment Planning Data module.

The software determines the optimal asset mix with the higher return and minimum risk. Find portfolios with the highest chance to obtain desired returns. Account the impact of taxes and create comparisons of multiple portfolio allocations on the efficient frontier. Moreover, examine the effect of changing assumptions (like taxes, interest rates, inflation, etc.). Classify security holdings to recommend an asset allocation to implement the plant with mutual funds and look at historical behavior. With historical analysis, rank performance and examine risk/return trade offs. The Security Classifier defines the allocation of the portfolio and maintains a database of over 21,000 mutual funds, annuity subaccounts, and stocks. In the Ibbotson database, more than 5,800 mutual funds, 7,200 annuity subaccounts and 8,350 individual securities have been classified. Finally, service extends to security classifier, risk assessment, mean variance optimizer, historical calculations, wealth forecasting using straightforward analytical model and Monte Carlo simulation, and fund optimizer that determines the portfolio of mutual funds and subaccounts that most closely matches a target asset allocation.

The Reuter systems deliver financial information generated by exchanges, over-the-counter markets, price contributors, research services, and news. Also included are real-time prices, price histories and news, statistics, broker research and company fundamental data, and estimates. The financial information system of Reuter enables the market analysis and trading and investing opportunities identification, risk assessment of different strategies, ability to communicate with other market participants, direct trading, and access to executable prices and trading tools. Reuter premium financial information system incorporates trading functionality of the equity, fixed income, foreign exchange, and commodities from

Figure 2. The Reuters enterprise financial information systems



the desktop system of the company. Regarding the front office, Reuter offers pretrade analysis, limits checking, real time positioning, pricing analytics, tactical risk management, and profit and loss reporting. Regarding the middle office, it offers limits management, market risk and credit risk management, back testing, enterprise wide risk management, regulatory reporting, double validation, and market conformity check. In the back office, prevalidation is offered along with back validation, settlement, cash management, accounting, reporting, and messaging.

Bloomberg provides real-time and archived financial, market data, pricing, trading news, and communications tools. Platts is a division of McGraw-Hill Companies dedicated to energy financial information for oil, electricity, gas, coal, nuclear, petrochemicals, and metals. Platts offers energy benchmark pricing, forward curves, oilgrams, and real-time market news and industry analysis.

All the data and information collected from internal and external FIS are processed further in the processing box in order to advise and support decision making. In the following paragraph we will present the processing box operations along with the quantitative methods used to solve specific problems.

CORPORATE FINANCIAL INFORMATION SYSTEM: PROCESSING BOX AND QUANTITATIVE METHODS OF ANALYSIS

External databases provide the financial information system with inputs such as: YTM of the long-term bond (risk free rate), level of inflation, GDP growth, exchange rates, interbank rates and spreads (cost of debt), prices of commodities (metals, fuels, electricity, etc.) that determine the manufacturing cost, and so forth. The internal database and mainly the MAIS will provide FIS with the inputs relative to the unit cost classified according to behavior and activities. AIS will provide the relevant inputs for the financial statement forecasting and for the estimation of pro-forma financial statements. Sales and marketing information systems will provide the relevant forecasts for the sales budget. In the processing box, budgeting analysis will determine future needs for sales and capacity additions as well as investment needs. Future capacity investments will be selected according to the required returns (internal rate of return (IRR), net present value (NPV), return on invested capital (ROIC)). Material needs and their costs will be determined in conjunction with material requirements planning I (MRPI) systems. Sales budget, production budget, and inventory budget along with former financial statements (provided by AIS) will determine the forecasted pro-forma financial statements upon which further analysis will be contacted (i.e., ratio analysis, determination of additional required funds, etc.).

Portfolio management, financial, and specifically credit risk analysis will also be used for working capital and current assets management. Trading department will bargain and insure the price of both materials and end products on the spot and secondary markets wherever are organized otherwise in over-the-counter. FIS will determine metrics like earnings-at-risk (EaR), value-at-risk (VaR), and so forth to impose trading limits and define trading and hedging strategies. The operations will be accomplished by the coordinative effort of two functions: the front/middle office and the back office. Fundamental, technical (trends, support and resistance levels, reversal and continuation patterns, head and shoulders, triangles, pennants, flags, candlesticks, point and figure charts, etc.) and quantitative analysis (statistical, artificial intelligence and simulation techniques like Monte Carlo, and relevant software like @ Risk and Crystal Ball, EaR, PaR, CaR, etc.) will be accomplished by the front/middle office functions, while the operation of back office will be only supportive and mainly focused on matters of accounting. In the front and middle office the operations of market analysis, risk hedging, and trading monitoring will be accomplished while back office operations will include, risk measurement and reporting, accounting handling, and credit risk control. Various quantitative techniques are employed to model credit ratings using artificial neural networks (ANN) and sovereign credit ratings to model credit risk evaluation, to forecast financial risk, to model sock selection, to model failure and bankruptcy prediction, and so forth. All of these models employ statistical and artificial intelligence techniques to solve different financial problems. In the first case techniques such as linear and nonlinear regression, probit regression, logit analysis, linear or quadratic multivariate discriminant analysis, multidimensional scaling, simulation techniques, logistic regression and linear programming, WMA, ARCH-GARCH, and so forth can be seen. The category of artificial intelligence consists of artificial neural networks, case base reasoning, genetic algorithms, and genetic programming. Each of the above techniques has been used so as to provide solutions in different financial topics and will be presented in the following paragraphs along with their usage in financial modeling.

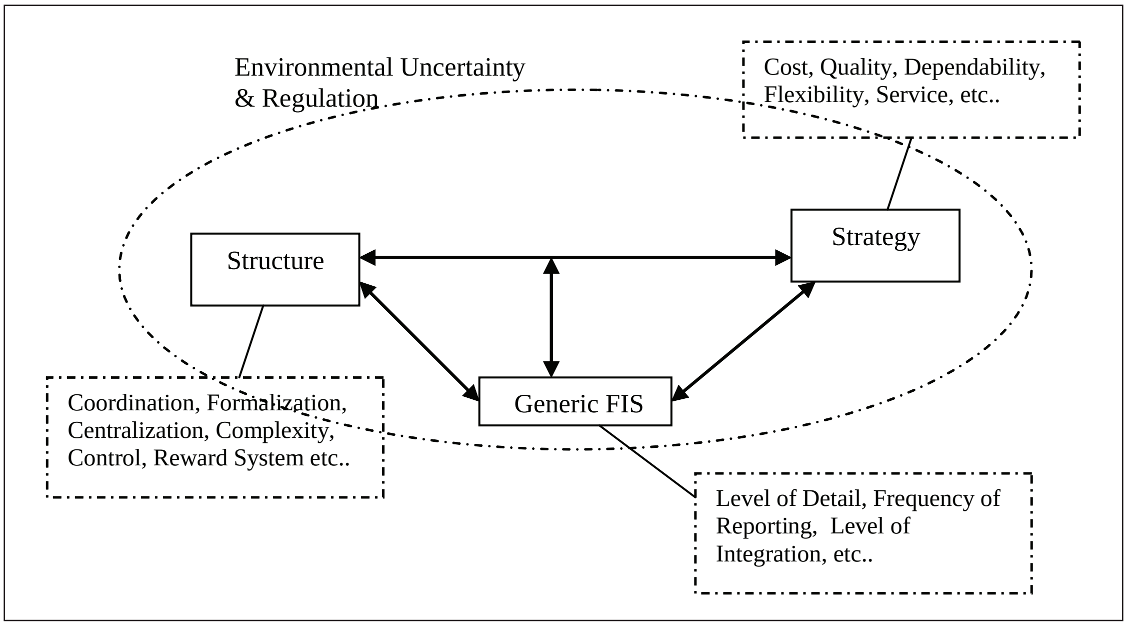
The linear regression technique, although trivial, can efficiently solve many forecasting problems. Multivariate linear regression was used by J.D. Jobson (1982) to test the arbitrage pricing theory and by Horrigan (1996) to find a link between the utility of the accounting data and the long-term credit administration. Schwartz (1988) identified four determinants of bid-ask spread (activity, risk information, and competition) and McInish and Wood (1992) used multivariate linear regression to demonstrate the relation among spreads and those factors. Prediction regressions and autoregression models have been applied by Keim and Stambaugh (1986) to predict the excess returns by some lagged variables: (a) the difference between the yield on long-term BAA underrated corporate bonds and the short-term Treasury bill rate; (b) the level of the S&P 500 index; and (c) the level of the small firm stock index. Using the same technique, Fama and French (1998) estimated the effect of the lagged dividend-price ratio on the stock returns. Additionally, Campell and Shiller (1988) found that the lagged dividend-price ratio and lagged dividend growth rate have a significant relation with stock returns. The same technique was applied by Hodrick (1992) to predict stock returns by the dividend-price ratio. Lewellen (2004) used the book-to-market ratio as a predictor while Kothari and Shanken (1997) have used the lagged book-to-market ratio. Predictive regression also was used by French, Schwert, and Stambaugh (1987) (predictor was the return variance obtained from an autoregressive integrated moving average (ARIMA) model). Linear regression finally could be found in popular studies of the capital asset pricing model (CAPM). Black et al. (1972) and Fama and McBeth (1973) have used the two-stage cross-sectional regression method. Finally, we can refer the work of Cantor and Packer (1996) and Trevino and Thomas (2000a, 2000b). Another type of regression which is also used in FIS is that of probit and logit. This technique has been used to examine the decreasing trend of firms to pay dividends (Baker & Wurgler, 2004; Ederington, 1985; Fama & French, 2001; Trevino & Tomas, 2000-2001). Linear multivariate discriminant analysis was used by Altman (1968), Belkaoi (1980), and Taffler(1983), and quadratic multivariate discriminant analysis was applied by Prinches & Mingo (1975). Mar et al. (1996) applied multidimensional scaling for bond ratings. Logistic regression was used by Wiginton (1980) to compare logit and discriminant models of consumer credit behavior. Linear programming techniques have been used by Mangasarian(1965), Freed & Glover (1981), and Hand (1981). Finally, simulation techniques have been used in economic analysis, decision making, and operational research. Simulation techniques have been applied in financial based computer systems, management games, for decision making, and operational research. Simulation procedures have been used by Kothari and Warner (2001) for the study of empirical properties of performance measures for mutual funds.

Genetic algorithm as a search algorithm is used to find solutions in complex optimization problems. Genetic algorithm offer advantages over the usual optimization techniques. Mahfoud and Mani (1996) say that the traditional optimization techniques "are of no use" if the problem that we want to solve is nondifferentiable. Genetic algorithms do not require gradient information, so there will be no problem in such optimization. Moreover, genetic algorithms are more likely to find a global optimum, as they search many points at once. The global optimum is more difficult to be found by other optimization techniques which do not search as genetic algorithms. Finally, genetic programming is a methodology that executes a procedure which guide automatically to the solution of the defined problem. Neural network applications can be found in financial topics such as financial and economic forecasting, data mining and target marketing, bankruptcy prediction, credit evaluation, portfolio selection/diversification, simulation of market behaviour, corporate loan portfolio risk evaluation, risk assessment, and pricing financial derivatives. Brown et al. (1995) talk about FALCON, a financial information systems that use ANNs by large credit card companies to screen transactions for potential fraud and credit advisor that accord credit for automobile loans. Furthermore, extensive literature can be found concerning the topic of pricing options using artificial neural networks. Malliaris and Salchenberger (1993a, 1993b) compare the performance of the Black-Scholes model and an ANN showing that the ANN model was superior for out-of-the-money options and the Black-Scholes model better for in-the-money options. Hutchinson, Lo, and Poggio (1994) compare three ANNs with Black-Scholes formula in pricing American call options on S&P100 futures. All three networks were better than the Black-Scholes model. Qi and Maddala (1996) compare the performance of an ANN in pricing European call options and found that the ANN produces better results. Alike to Qi and Maddala, Geigle and Aronson (1999) examine the performance of ANNs in pricing options. They have indicated that ANNs were better than Black-Scholes formula. The superiority of ANNs is shown also by De Freitas et al. (2000) who applied ANNs and Black-Scholes model to price Financial Times Stock Exchange (FTSE) 100 (United Kingdom) call and put options. Kelly (1994) shows that the ANN was superior than the binomial model in a study to price American-style put options on four U.S. firms. Healey et al. (2002) show that ANNs are well enough in pricing options. Bennell and Sutcliffe (2004) compare the performance of Black-Scholes model with an ANN in pricing European-style call options on the FTSE100 index. They provide an extensive study on the performance of ANNs in pricing UK options. The results have shown that ANNs were superior and they mentioned that this superiority "suggests that ANNs may have an important role to play in pricing other options." Another application of such networks is the use of Bayesian networks in financial analysis. Gemela (2001) studied the use of probabilistic Bayesian networks in fundamental financial analysis. They present the construction of a Bayesian network for the financial analysis of specific firms and sectors of the Czech economy. A survey of ANN applications in forecasting mentions that extensive literature can be found in financial applications of ANNs. ANNs have been used for forecasting bankruptcy and business failure (Tsakonas et al. 2006). Zhang, Hu, Patuwo, and Indro (1999) use neural networks for modelling bankruptcy and they linked to the Bayesian classification theory. Ahn, Cho & Kim (2000) presented a combination of rough sets and neural networks. Parag & Pendharkar (2004) proposed a threshold-varying artificial neural network for binary classification. Comparisons of results with other techniques such as inductive machine learning and genetic algorithms are also performed. In bankruptcy prediction the work of Odom and Sharda (1990) Tam and Kiang (1992), and Fletcher and Goss (1993) can be referred to. ANNs have also been used as a prediction and forecasting tool for exchange rates, stock prices, and volatility. The work of Casdagli and Eubank (1992), Hlupic, Walker, and Irani (1998), Refenes (1993) and Wu (1995) explores exchange rate forecasting. White (1998), Schoneburg (1990), Bergerson and Wunsch (1991) apply ANNs for stock price prediction. Kaastra and Boyd (1994) provide a procedure to design a neural network forecasting model for financial and economic time series. Kim (2006) proposes a genetic algorithm approach in ANNs for financial data mining. Hamid and Iqbal (2004) use ANNs to forecast the volatility of Standard & Poor's (S&P) 500 index futures prices. They present that the volatility forecasts from neural networks with realized volatility are about the same. Finally, Chen and Leung (2005) suggest that "neural networks are a better tool to forecast the exchange rate correlation." They compared the forecasting performance of neural networks with that of implied volatility and random walk models. Applications of ANNs could be found on predicting sovereign credit ratings. Bennell et al. (2006) believe that neural networks are superior to ordered probit models which have been used for credit rating. They found that ANNs, which are used by major rating agencies (e.g., Standard & Poor's and Moody's), can be used to predict sovereign credit rating. ANNs have also been applied in marketing problems. Bentz and Merunka (2000) showed that the neural network can be used as a generalization of the multinomial logit model which is used for brand choice modelling. Techniques such as genetic algorithm have also been used for financial forecasting. Mahfoud and Mani (2002) presented a genetic algorithm system to predict future performances of stocks. They tested the genetic algorithm system and a neural network system on 5,000 stocks where genetic algorithm systems produced better final results. Adaptive genetic algorithms have been used for construction finance decisions. Lam, Hu, NG, Yuen, Lo, and Wong (2001) propose adaptive genetic algorithms as an alternative method for modelling financial decisions. In this field, many optimization techniques have been used, such as heuristic method, linear programming, and neural optimization networks. The technique of genetic programming has been used in financial topics, such as option pricing. Chen, Lee, and Yeh (1999) use data from S&P 500 index options and distinguish them in in-the-money and out-of-the-money categories. They compare the genetic programming tree with the Black-Scholes model in its capability to hedge. Finally, we could find applications of machine learning techniques in financial topics, such as portfolio optimization problems. Ince and Trafalis (2006) developed a short-term management model based on the volatility around the earning announcements.

CONCLUSION AND FUTURE TRENDS: MANAGERIAL DECISION

The performance of a FIS as any other advanced IT system should be examined under the strategic alignment perspective (Theodorou, 2005, 2003, 1996; Theodorou & Giannoula, 2008). Otherwise performance comparisons and benchmarking practices will lead to wrong conclusions. Specifically, FIS should be aligned with the organizations' structure and strategy which is determined by the environment's uncertainty and volatility where the business operates under certain regulations (Chenhall, 2003; Theodorou, 2003). Performance achievements should be judged in relation to an alignment model in control of contingencies such as firm size, age, and so forth. Theodorou (2004) presents in detail the alignment model and its' variables with the relevant literature. The frequency of FIS reporting, the level of FIS detail, the level of integration among internal, and external financial information systems should be decided in relation

Figure 3. Alignment of FIS



to the strategic alignment mechanism (Figure 3) in control of contingencies such as size and age of the firm. The more volatile the environment where the firm operates the greater the need for the FIS to generate frequent reporting and higher level of detail in order for management to take corrective actions (Theodorou, 1996). Thus, decision making needs to do frequent corrective actions in a volatile environment. Rigid structure for firms working in a stable environment directs competitive strategy toward cost. Level of detail and frequency of FIS were kept at low level along with low level of integration among FIS and internal and external financial information systems.

The FIS model previously presented should be examined on the previous characteristics but this must be done under the "umbrella" of strategic alignment theory. Decision making in a more detailed FIS model is also another topic that has to be examined using the criterion of performance. Sometimes potential benefits of frequent reporting, greater detail, and integration do not exceed the cost of the system. Implementing a sophisticated FIS system entails costs of consulting, training, and maintenance, but not the cost of investing. For example, if the firm has no significant number of counterparts than the cost of a credit risk system may exceed the benefits realized. Finally, further research can be conducted regarding the commonly adopted quantitative methods by successor firms.

FUTURE RESEARCH DIRECTIONS

Future research for the design of a financial information system must make use of the alignment framework (Theodorou, 2005, 2003, 1996) and the contingency approach (Figure 3), especially for the processing box and the quantitative methods. The components of the financial information system (described in the previous sections) should be chosen according to the environment where the firm operates and its existing structure. Increased performance and competitive advantage by the system will be expected only if modifications of business structure and the systems design are aligned with the environment. For example, a business with a flexible structure designed around risk management can better operate in a volatile environment. The financial information system of this business should incorporate modules of risk management (calculations of VaR, EaR, etc.) and hedging techniques, to help management eliminate the variability. In such an environment where nonstationarity of the mean exist, the computational averaging techniques (simple/centered/doubled/weighted moving averages, exponential smoothing, ARIMA, etc.) may be obsolete. More advanced computational methods must be applied, such as random walk models (RWM), stochastic volatility (SV) models, models that use stochastic processes, and computational intelligence of nonlinear artificial neural networks (S-E-TAR ANNs). Moreover, the computational techniques of autoregressive and generalized autoregressive conditional heteroscedasticity (ARCH-integrated-GARCH, fractionally integrated-GARCH, etc.) should also be used in these FIS in order to better describe volatility clustering, excess kurtosis, and fat tailedness. Finally, there is no need to incorporate this computational intelligence and the structural processes of risk management in the financial information systems which operate in a stable and predictable environment. If we do so, than the return on the additional not needed invested capital will be decreased as well as the overall performance.

That is why future research must take into account the alignment framework (Theodorou, 2005, 2003, 1996) in order to decide about the design of an FIS as well as the computational intelligent which has to be employed by taking into account behavioral characteristics.

Finally, until now there is no research indicating which parts of the previously mentioned generic FIS model are adopted by the firms. Also there is nothing indicating how the effect of frequency of reporting, level of detail, and integration of the FIS model with enterprise resource planning systems (ERP) impacts business performance.

REFERENCES

Abernethy, M., & Brownell, P. (1997). Management control systems in research and development organizations: The role of accounting, behavior and personnel controls. *Accounting, Organizations and Society, 22*(3-4), 233-248.

Abernethy, M. A., & Stoelwinder, J. Y. (1996). The role of professional control in the management of complex organizations. *Accounting, Organizations and Society, 20,* 1-17.

Ahn, B. S., Cho, S. S., & Kim, C. Y. (2000). The integrated methodology of rough set theory and artificial neural network for business failure prediction. *Expert Systems with Applications, 18,* 65-74.

Altman, E. L. (1968). Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *Journal of Finance, 23*(3), 589-609.

Baker, M., & Wurgler, J. (2004). Appearing and disappearing dividends: The link to catering incentives. *Journal of Financial Economics, 73,* 271-288.

Belkaoi, A. (1980). Industrial bond ratings: A new look. *Financial Management, Autumn,* 44-51.

Bennell, J., Crabbe, D., Thomas, S., & Gwilym, O. (2006). Modelling sovereign credit ratings: Neural networks versus ordered probit. *Expert Systems with Applications, 30,* 415- 425.

Bennell, J., & Sutcliffe, C. (2004). Black-Scholes versus artificial neural networks in pricing FTSE 100 options. *Intelligent Systems in Accounting, Finance and Management, 12,* 243-260.

Bentz, Y., & Merunka, D. (2000). Neural networks and the multinomial logit for brand choice modelling: A hybrid approach. *Journal of Forecasting, 19,* 177-200.

Bergerson, K., & Wunsch, D. C. (1991). A commodity trading model based on a neural network-expert system hybrid. In *Proceedings of the IEEE International Conference on Neural Networks*, Seattle, Washington (pp. 1289-1293).

Black, F., Jensen, M. C., & Scholes, M. (Ed.). (1972). The capital asset pricing model: Some empirical tests. *Studies in the theory of capital markets.* New York: Praeger.

Brown, C. E., Coakley, J. R., & Phillips, M. E. (1995, May). Neural networks enter the world of management accounting. *Management Accounting, 11,* 51-57.

Bruggeman, W., & Slagmulder, R. (1995). The impact of technological change on management accounting. *Management Accounting Research, 6*, 241-252.

Campbell, J. Y., & Shiller, R. J. (1988). The dividend-price ratio and expectations of future dividends and discount factors. *Review of Financial Studies, 1,* 195-228.

Cantor, R., & Packer, F. (1996). Determinants and impact of sovereign credit ratings. *Federal Reserve Bank of New York Economic Policy Review, 2*.

Casdagli, M., & Eubank, S. (1992). Predicting sunspots and exchange rates with connectionist networks. In A. S. Weigend, B. A. Huberman, & D. E. Rumelhart (Eds.), *Nonlinear modeling and forecasting* (pp. 395-432). Redwood City, CA: Addison-Wesley.

Chen, S. H., Lee, W. C., & Yeh, C. H. (1999). Hedging derivative securities with genetic programming. *International Journal of intelligent Systems in Accounting, Finance and Management, 8,* 237-251.

Chen, A., & Leung, M. (2005). Performance evaluation of neural network architectures: The case of predicting foreign exchange correlations. *Journal of Forecasting, 24,* 403-420.

Chenhall, R. K. (2003). Management control systems design within its organizational context: Findings from contingency-based research and directions for the future. *Accounting, Organizations and Society, 28,* 127-168.

Chenhall, R., & Langfield-Smith, K. (1998). The relationship between strategic priorities, management techniques and management accounting: An empirical investigation using a system approach. *Accounting, Organizations and Society 23*(3), 243-264.

Chenhall, R. H., & Morris, D. (1986). The impact of structure, environment, and interdependence on the perceived usefulness of management accounting systems. *The Accounting Review, 61,* 16-35.

Choe, J. (2004). The consideration of cultural differences in the design of information systems. *Information & Management, 41,* 669-684.

Cooper, D., Hayes, D., & Wolf, F. (1981). Accounting in organized anarchies. *Accounting, Organizations, and Society, 6,* 175-191.

Cooper, R., & Kaplan, R. (1991). *The design of cost management systems: Text, cases, and readings.* Englewood Cliffs, NJ: Prentice Hall.

Cooper, R., & Kaplan, R. (1992). Activity-based systems: Measuring the costs of resource usage. *Accounting Horizons, 6,* 1-13.

Cooper, J., & Suver, J. (1994). Variance analysis refines overhead cost control. *Healthcare Financial Management, 48,* 40-52.

Covaleski, M. A., & Dirsmith, M. W. (1986). The budgetary process of power and politics. *Accounting, Organizations and Society, 11,* 193-214.

De Freitas, J. F. G., Niranjan, M., Gee, A. H., & Doucet, A. (2000). Sequential Monte Carlo methods on train neural network models. *Neural Computation*, *12*(4), 955-993.

Ederington, L. H. (1985). Classification models and bond ratings. *The Financial Review, 20*(4), 237-262.

Fama, E. F., & French, K. R. (1988). Dividend yields and expected stock returns. *Journal of Financial Economics, 22,* 3-25.

Fama, E. F., & French, K. R. (2001). Disappearing dividends: Changing firm characteristics or lower propensity to pay? *Journal of Financial Economics, 60,* 3-44.

Fama, E. F., & MacBeth, J. D. (1973). Risk, return and equilibrium: Empirical tests. *Journal of Political Economy, 81,* 607-636.

Feltham, G. A. (1977). Cost aggregation: An information economic analysis. J*ournal of Accounting Research, 15,* 42-70.

Feltham, G. A., & Xie, J. (1994). Performance measure congruity and diversity in multi-task principal/agent relations. *The Accounting Review, 69,* 429-453.

Flamholtz, E., & Das, T. (1985). Toward an integrative framework of organizational control. *Accounting, Organizations and Society, 10*(1), 35-50.

Fletcher, D., & Goss, E. (1993). Forecasting with neural networks: An application using bankruptcy data. *Information & Management, 24,* 159-167.

Foster, G. F., & Swenson, D. (1997). Measuring the success of activity-based cost management and its determinants. *Journal of Managerial Accounting Research, 9,* 109-141.

Freed, N., & Glover, F. (1981). A linear programming approach to the discriminant problem. *Decision Sciences, 12,* 68-74.

French, K. R., Schwert, G. W., & Stambaugh, R. F. (1987). Expected stock returns and volatility. *Journal of Financial Economics, 19,* 3-29.

Geigle, D.S., & Aronson, J.E. (1999). An artificial neural network approach to the valuation of options and forecasting of volatility. *Journal of Computational Intelligence in Finance, 7*(6), 19-25.

Gemela, J. (2001). Financial analysis using Bayesian networks. *Applied Stochastic Models in Business and Industry, 17,* 57-67.

Govindarajan, V. (1984). Appropriateness of accounting data in performance evaluation: An empirical examination of environmental uncertainty as an intervening variable. *Accounting, Organizations and Society, 9*(2), 125-135.

Hamid, S., & Iqbal, Z. (2004). Using neural networks for forecasting volatility of S&P 500 Index futures prices. *Journal of Business Research, 57,* 1116-1125.

Hand, D. J. (1981). *Discrimination and classification.* Chichester, UK: Wiley.

Healy, J., Dixon, M., Read, B., & Cai, F. F. (2002). A data-centric approach to understanding the pricing of financial options. *European Physical Journal B, 27*(2), 219-227.

Hilton, R. (1979). The determinants of cost information value: An illustrative analysis. *Journal of Accounting Research, 17,* 411-435.

Hlupic, V., Walker, P.A., & Irani, Z. (1998). Predicting movements in foreign currency rates using simulation modelling. *Management Decision, 36/7,* 465-472.

Hodrick, R. (1992). Dividend yields and expected stock returns: Alternative procedures for inference and measurement. *Review of Financial Studies, 5,* 357-386.

Horrigan, J. O. (1966). The determination of long term credit standing with financial ratios. *Journal of Accounting Research, Supplement, 4*(3), 44-62.

Hutchinson, J. M., Lo, A. W., & Poggio, T. (1994). A nonparametric approach to pricing and hedging derivative securities via learning networks. *Journal of Finance, 49*(3), 851-889.

Ince, H., & Trafalis, T.B. (2006). Kernel methods for short-term portfolio management. *Expert Systems with Applications, 30,* 535-542.

Ittner, C. D., Lanen, W. N., & Larcker, D. F. (2002). The association between activity-based costing and manufacturing performance. *Journal of Accounting Research, 40,* 711-725.

Jobson, J. D. (1982). A multivariate linear regression test for the arbitrage pricing theory. *The Journal of Finance, September,* 4.

Johnson, H. (1992). It's time to stop overselling activity-based concepts. *Management Accounting, 74,* 26-35.

Kaastra, I., & Boyd, M.S. (1995). Forecasting futures trading volume using neural networks. *The Journal of Futures Markets, 15*(8), 953-970.

Kaplan, R. S., & Norton, D. P. (1992). The balanced scorecard measures that drive performance. *Harvard Business Review, 70,* 71-79.

Karmarkar, U. S., Lederer, P. J., & Zimmerman, J. L. (Eds.). (1990). Choosing manufacturing production control and cost accounting systems. In R. Kaplan (Ed.), *Measures for manufacturing excellence*. Boston: Harvard Business School Press.

Keim, D. B., & Stambaugh, R. F. (1986). Predicting returns in the stock and bond market. *Journal of Financial Economics, 17,* 357-396.

Kelly, D. L. (1994, December). *Valuing and hedging American put options using neural networks* (Working Paper). University of California.

Khandwalla, P. (1972). The effect of different types of competition on the use of management controls. *Journal of Accounting Research, 10,* 275-285.

Kim, K. (2006). Artificial neural networks with evolutionary instance selection for financial forecasting. *Expert Systems with Applications, 30,* 519-526.

Kothari, S. P., & Shanken, J. (1997). Book-to-market, dividend yield, and expected market returns: A time-series analysis. *Journal of Financial Economics, 18,* 169-203.

Kothari, S.P., & Warner, J.B. (2001). Evaluating mutual fund performance. *The Journal of Finance, October*(5).

Krumwiede, K. R. (1998). The implementation stages of activity-based costing and the impact of contextual organizational factors. *Journal of Management Accounting Research, 10,* 239-277.

Lam, K. C., Hu, T., NG, T., Yuen, R. K. K., Lo, S. M., & Wong, C. (2001). Using an adaptive genetic algorithm to improve construction finance decisions. *Engineering, Construction and Architectural Management, 1,* 31-45.

Lawrence, C. H. (1990). The effect of ownership structure and accounting system type on hospital costs. *Research in Governmental and Nonprofit Accounting, 6,* 35-60.

Lewellen, J. (2004). Predicting returns with financial ratios. *Journal of Financial Economics, 74*(2), 209-23.

Mahfoud, S., & Mani, G. (1996). Financial forecasting using genetic algorithms. *Applied Artificial Intelligence, 10,* 543-565.

Malliaris, M., & Salchenberger, L. (1993a). A neural network model for estimating option prices. *Journal of Applied Intelligence, 3*(3), 193-206.

Malliaris, M., & Salchenberger, L. (1993b). Beating the best: a neural network challenges the Black–Scholes formula. *In Proceedings of the Ninth Conference on Artificial Intelligence for Applications* (pp. 445-449). IEEE

Mangasarian, O. L. (1965). Linear and nonlinear separation of patterns by linear programming. *Operations Research, 13,* 444-452.

Mar, M. C., Apellaniz, C. G., & Cinca, C. S. (1996). A multivariate study of Spanish bond ratings. *Omega, 24*(4), 451-462.

McInich, T. H., & Wood, R. A. (1992). An analysis of intraday patterns in bid/ask spreads for NYSE stocks. *The Journal of Finance, June*(2).

McNair, C. (1990). Interdependence and control: Traditional versus activity based responsibility accounting. *Journal of Cost Management, 4,* 15-24.

Miller, J. (1992). Designing and implementing a new cost management system. *Journal of Cost Management, 6,* 41-53.

Odom, M., & Sharda, R. (1990, Jannuary). A neural network model for bankruptcy prediction. In *Proceedings of the International Joint Conference on Neural Networks,* Washington, DC, (pp. 163-168, Vol. 2). INNS.

Ouchi, W. G. (1979). A conceptual framework for the design of organizational control mechanisms. *Management Science, 79,* 833-848.

Pinches, E., & Mingo, K. A. (1975). The role of subordination and industrial bond ratings. *Journal of Finance, March,* 201-206.

Preston, A. M. (1992). The birth of clinical accounting: A study of the emergence and transformations of discourses on costs and practices of accounting in U.S. hospitals. *Accounting, Organizations, and Society, 17,* 63-100.

Qi, M., & Maddala, G. S. (Ed.). (1996). Option pricing using artificial neural networks: The case of S&P 500 index call options. In Y. Abu-Mostafa, J. Moody, A. N. Refenes, & A. Weigend (Eds.), *Neural Networks in financial engineering: Proceedings of the third international conference on neural networks in the capital markets* (pp. 78-91). New York: World Scientific.

Refenes, A.N. (Ed.). (1993). Constructive learning and its application to currency exchange rate forecasting. In R.R. Trippi & E. Turban (Eds.), *Neural networks in finance and investing: Using artificial intelligence to improve real-world performance*. Chicago: Probus Publishing Company.

Schoneburg, E. (1990). Stock price prediction using neural networks: A project report. *Neurocomputing, 2,* 17-27.

Schwartz, R. (1988). *Equity markets.* New York: Harper and Row.

Shank, J., & Govindarajan, V. (1993). *Strategic cost management: The new tool for competitive advantage*. New York: The Free Press.

Simons, R. (1987). Accounting control system and business strategy: And empirical analysis. *Accounting, Organizations and Society, 12,* 357-374.

Taffler, R. (1983). The assessment of company solvency and performance using a statistical model. *Accounting and Business Research, 13,* 295-307.

Tam, K., & Kiang, M. (1992, July). Managerial applications of neural networks: The case of bank failure predictions. *Management Science, 38*(7), 926-947.

Theodorou, P. (1996, February 1). Business restructuring and flexibility strategy in liberalized markets. *Management of small and medium sized enterprises.* University Publications of Creta.

Theodorou, P. (1996). *Lean production and the correlation between variability, flexibility and I.T.* Paper presented at ALPS Euroconference, Bologna, Italy. Retrieved July 22, 2007, from http://www.fe.up.pt/si_uk/publs_pesquisa.FormView?P_ID=3133; http://www.bnet.co.uk/enimm/alps2.htm

Theodorou, P. (2003). Strategic information systems' the concept of alignment. In L. Antonio (Ed.), *IT-based management: Challenges and solutions.* Hershey, PA: IGI Global.

Theodorou, P. (2004). A DSS model that aligns business strategy and business structure with advanced information technology: A case study. *Annals of Cases on Information Technology, Volume 6/2004,* pp. 157-176.

Theodorou, P. (2005). Business strategy, structure and information technology alignment. *Encyclopedia of information science and technology* (Vol. 5). Hershey, PA: IGI Global

Theodorou, P., & Giannoula, F. (in press). Manufacturing strategies and financial performance the effect of advanced information technology. *International Journal of Management Science (OMEGA).*

Trevino, L., & Thomas, S. (2000a). *Systematic differences in the determinants of foreign currency sovereign ratings by rating agency* (Discussion papers in accounting and management science 00-153), University of Southampton, UK, School of Management.

Trevino, L., & Thomas S. (2000b). *The statistical determinants of local currency sovereign ratings* (Discussion papers in accounting and management science 00-158), University of Southampton, UK, School of Management.

Trevino, L., & Thomas, S. (2001). Local versus foreign currency ratings: What determines sovereign transfer risk? *Journal of Fixed Income, 11*(1), 65-76.

Tsakonas, A., Dounias, G., Doumpos, M., & Zopounidis, C. (2006). Bankruptcy prediction with neural logic networks by means of grammar-guided genetic programming. *Expert Systems with Applications, 30,* 449-461.

White, H. (1998). Economic prediction using neural networks. The case of IBM daily stock returns. *Proceedings of the IEEE International Conference on Neural Networks, 2,* 451-458.

Wiginton, J. C. (1980). A note on the comparison of logit and discriminant models of consumer credit behavior. *Journal of Financial and Quantitative Analysis, 15,* 757-770.

Wong, Bo., K., Bodnovich, T., A. & Selvi, Y. (1997). Neural network applications in business: A review and analysis of the literature (1988-95). *Decision Support Systems, 19,* 301-320.

Wu, B. (1995). Model-free forecasting for nonlinear time series (with application to exchange rates. *Computational Statistics and Data Analysis, 19,* 433-459.

Zhang, G., Hu, M. Y., Patuwo, B. E., & Indro, D. C. (1999). Artificial neural networks in bankruptcy prediction: General framework and cross-validation analysis. *European Journal of Operational Research, 116*(1), 16-32.

Zhang, G., Patuwo, B. E., & Hu, M.Y. (1998). Forecasting with artificial neural networks: The state of the art. *International Journal of Forecasting, 14,* 35-62.

ADDITIONAL READING

Cuthbertson, K., & Nitzsche, D. (2004). *Quantitative financial economics*. Wiley.

Frances, P. H., & Dijk, D. V. (2006). *Non linear time series models in empirical finance.* Cambridge University Press.

Galitz, L. C. (1995). *Financial engineering: Tools and techniques to manage financial risk.* Irwin, IL: Burr Ridge.

Golub, B. W., & Ortega, J. M. (1989). *Scientific computing and differential equations: An introduction to numerical methods.* New York: Academic.

Hull, J., & White, A. (1990). Valuing derivative securities using the explicit finite difference method. *Journal of Financial and Quantitative Analysis, 25*(1), 87-100.

Koop, G. (2006). *Analysis of financial data.* Wiley.

Law, A. M., & Kelton, W. D. (1991). *Simulation modeling and analysis* (2nd ed.). New York: McGraw-Hill.

Lo, A. W. (1994). Neural networks and other nonparametric techniques in economics and finance. In H. R. Fogler (Ed.) *Blending Quantitative and Traditional Equity Analysis.* Charlottesville, VA: Association for investment management and research.

Lyuu, Y. D. (2002). *Financial engineering and computation: Principles, mathematics, algorithms.* Cambridge University Press.

Marshall, J. F., & Bansal, V. K. (1992). *Financial engineering: A complete guide to financial innovation.* New York: New York Institute of Finance.

Mason, S. P., Merton, R. C., Perold, A. F., & Tufano, P. (1995). *Cases in financial engineering: Applied studies of financial innovation.* Prentice-Hall, NJ: Englewood Cliffs.

Phillips, D. T., Ravindran, A., & Solberg, J. J. (1976). *Operations research: Principles and practice.* New York: Wiley.

Streeter, L. A., Kraut, R. E., Lucas, H. C., Jr., & Caby, L. (1996). How open data networks influence business performance and market structure. *Communications of the ACM, 39*(7), 62-73.

Theodorou, P. (2003). Strategic information systems' the concept of alignment. In L. Antonio (Ed.), *IT-based management: Challenges and solutions.* Hershey PA, IGI Global.

Theodorou, P., & Giannoula, F. (in press). Manufacturing strategies and financial performance. *The International Journal of Management Science (OMEGA).*

Watsham, T.J., & Parramore, K. (1997). *Quantitative methods in finance.* Thomson.

Chapter IV
Fuzzy Sets and Analytical Hierarchical Process for Manufacturing Process Choice

M. Reza Abdi
Bradford University School of Management, UK

ABSTRACT

The aim of this chapter is to investigate the decision process of manufacturing systems (MSs) under uncertain conditions. The decision process needs a systematic approach to structure the system requirements and highlight the management preferences while considering vague criteria. In order to establish a suitable empirical approach for the decision process compatible with the current/future requirements, the analytical hierarchical process (AHP) is employed for structuring the criteria influencing the process choice. The application of the proposed AHP model for the selection of manufacturing process is demonstrated using numerical examples. In addition, due to dealing with vague data in the decision process, the uncertain criteria are characterised by typical fuzzy sets. The integrated fuzzy AHP is then analysed within the boundary conditions of the fuzzy criteria using the Expert Choice software. The proposed model is intended to be generic in structure and applicable to many firms.

INTRODUCTION

To survive in today's competitive market, customisation of manufacturing processes is the main requirement. Customers' requirements and their irregular behaviour are relatively unpredictable to the manufacturers. Therefore, manufacturers need to find solutions for fast adaptation of their resources to the dynamic market. In this way, an integration of process engineering and manufacturing management along with standardisation of products and processes are essential. Changing manufacturing processes provides the opportunity to grow in the different market situations. However, this might necessitate changes in the operations and resources that impose extra capital/operational cost to the system. As a result, the process choice must be selected by trading off between the technological elements and economical factors.

High quality products, high productive and low cost manufacturing are to be achieved by overcoming existing limits of technologies. These objectives can be obtained through activating the full potential of resources such as materials, manufacturing facilities, and operators' skills. In addition to these conventional objectives, manufacturers need to respond rapidly to changes in market demands. Consequently, responsiveness is becoming another objective to be considered for the manufacturing process selection. The responsiveness criterion is characterised by a fuzzy set in the AHP model and analysed within its fuzzy domain in order to evaluate the alternative manufacturing choices.

Different kinds of manufacturing processes will have different impacts on the system performance and customer satisfaction. Three different manufacturing choices are considered as the alternatives for the proposed model. A dedicated manufacturing is considered as one of the process alternative in which limited product types can be produced using transfer lines. This kind of manufacturing process is suitable and cost effective when there is a stable market for existing products. A responsive manufacturing is another system that its capacity and functionality are adjustable to fluctuations in product demands within the existing production range and also new products introduced to the system. The hybrid manufacturing choice is the other alternative which can cope with both situations in which limited products hold firm demands, but some other products have sharp changes in demands and/or new products with new manufacturing operations introduced to the system. The alternative processes are analysed with respect to the criteria and actors for the identified planning horizons, that is, long term, mid term, and short term through a case study.

BACKGROUND

Manufacturing processes need to be modelled in order to clarify the process requirements, better understand the system behaviour, and solve various problems of existing systems. Most related work in manufacturing system (MS) modelling has been involved with resource allocations, layout configurations, and job scheduling. Most existing analytical models attempt to simply minimise a static measure such as transportation cost, lead time, and/or work-in-process (WIP) subject to available resources. Reconfigurations and product variety have made the process of modelling notoriously difficult. This difficulty is due to a lack of analytical models, which are capable of evaluating the effects of product mix and resources reconfigurations on the manufacturing performance. In most conventional studies for MS modelling, the importance of changeover time, changeover cost, and product variants have been ignored. In addition, uncertainty caused by external/internal factors might lead to an ineffective decision; as a result, the conventional analytical models must be reformulated to obtain a comprehensive structure considering a variety of quantitative/qualitative parameters under uncertain conditions.

Advanced manufacturing processes need to be more flexible than ever before for their survival in the competitive environment. In the most studies, capital investment, operation cost, and WIP have been the major objectives of production planning to be minimised. However, considering new qualitative/quantitative requirements in manufacturing environments such as customer satisfaction, capacity changes, functionality requirements, and changeover cost and time, the decisions made by cost based models can no longer justify the investment for manufacturing processes. As a result, the top management and experienced engineers must participate in the decision process to jointly investigate the crucial factors affecting the system through a multicriteria evaluation approach. In addition, uncertainty of external/internal environments, which causes forthcoming changes in the decision structure, must be taken into account during the decision making process.

There have been several attempts to perform an analytical evaluation for justifying the appropriate MS design within manufacturing environments. For example, a previous survey on UK organisations (Willcocks & Lester, 1991) showed that about 60% of organisations used cost-benefit methods at first priority for the evaluation of the feasibility stage whereas around 40% used competitive advantage as their second priority. For evaluating manufacturing processes, a number of approaches have been found practical, which can assess manufacturing strategy, process design, and selection and operational modelling. Accordingly, appropriate decision support systems facilitate structuring and analysing decision making in different stages of product-process design. As shown in Figure 1, conventional evaluation approaches can be classified into three categories (Chan, Chan, & Chan, 2003):

1. The economic justification approaches
2. The strategic justification approaches
3. The analytic justification approaches

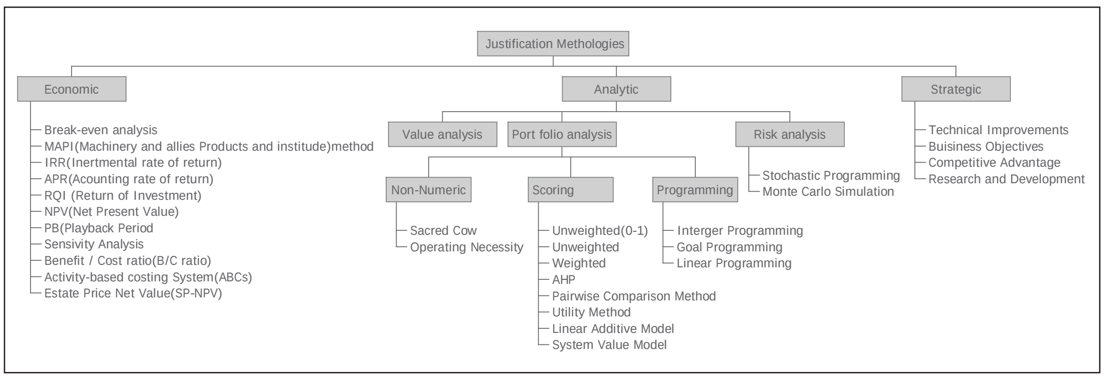
A justification approach for an MS design must consist of the following characteristics:

- Be simple in construct.
- Be flexible for a wide range of unconstructed criteria.
- Enable people to refine and improve judgments.
- Present the logical consistency of judgments.
- Accept both quantitative and qualitative attributes.
- Be able of sensitivity analysis of design parameters.

As it can be seen from Figure 1, the AHP is classified within the analytical scoring approaches to be employed for various manufacturing problems.

Simulation approaches based on queuing theory are suitable tools to measure the service level of a running MS with probabilistic attributes. A Semi-Markovian queue process is utilised to

Figure 1. Different evaluation methodologies (Chan et al., 2003)



model an appropriate performance measure of a reconfigurable MS (Xiaboo, Jiancai, & Zhenbi, 2001, pp. 750-755). However, such techniques are specifically suitable to evaluate a running MS at the post-design stage whereas product types with their corresponding manufacturing processes are predetermined. The lack of appropriate methodologies capable of assisting the evaluation and analysis of manufacturing processes before and during its detailed design must be taken into account, which may lead to the system failure.

Changes in market demand create a need for new designs of manufacturing processes. In order to sustain competitiveness in dynamic markets, manufacturing organisations should provide the sufficient flexibility to produce a variety of products on the same system (Chick, Olsen, Sethuraman, Stecke, & White, 2000). In this way, advanced manufacturing systems need to accurately consider economical aspects as well as engineering aspects; otherwise, they cannot obtain a reasonable share of competitive market to justify their investments.

Having determined the target products for manufacturing, the critical technological and economical factors must be identified and analysed prior to the determination of manufacturing elements such as machines, tools, and layout. Analysis of process selection for a MS design is of vital importance to the competitive strategy. One of the important issues in a manufacturing process design is to evaluate the feasibility of system configurations for different product types. Currently, there is no systematic method to evaluate the quality and productivity of systems with different configurations (Yang & Hu, 2000).

The AHP is a powerful tool for analysis of complex decision problems using weighted multiple choice criteria. The decision group involved with the decision process should divide the problem into the fundamental components of the hierarchical levels (Saaty, 1980). A few researchers have applied the AHP approach for the production planning and process choice. Oeltijenbruns, Kolarik, and Schenadt-Kirschner (1995) propose four fundamental steps of decision making for investment alternatives in order to support decisions for higher levels of automation/technology in manufacturing processes. These steps are as follows:

- Specification of investment alternatives and evaluation criteria such as quantitative and financial criteria.
- Pair-wise comparisons of criteria and categories.
- Rating of investment alternatives for each category.
- Overall ranking of investment alternatives for making decision.

The AHP can be applied for selecting plant layout configuration such as group technology, transfer lines, and functional layout with respect to the defined objectives, such as flexibility, volume, and cost (Abdul-Hamid, Kochhar, & Khan 1999). An AHP model can also be used for a continuous improvement process in industry with the decision elements defined by Labib and Shah (2001) as follows:

- **Scenarios:** Four possible combinations of two levels (low and high) of demand and supply.
- **Decision makers:** Managers.
- **Objectives:** Minimising changeover times, zero defects, zero waste, and zero breakdowns.
- **Options:** strategies such as speed losses

Maier-Speredelozzi and Hu. (2002) used the AHP for the selection of the most appropriate configuration of a manufacturing system with consideration of multiple performance criteria. The AHP model was structured based on four main performance objectives, that of productivity, quality, convertability, and scalability.

Due to the computational complexity for the analysis of the AHP models for complex problems, a number of packages have been marketed for decision analysis such as AutoMan and Expert choice. The user-friendly AHP package 'Expert Choice' (Expert Choice, 1999), has been found more applicable and accessible to engineers and managers by the author. This software is applied for demonstration of the proposed model through monitoring sensitivity analysis of fuzzy criteria. Recently another software called 'SuperDecision' is offered, which is capable of modelling and solving both the AHP and the analytical network process (ANP) models. The ANP is the general form of the AHP in which all criteria and alternatives are allowed to have an interrelationship rather than a hierarchy relationship. The ANP uses a network without the need to specify levels (Saaty, 2005).

Although the AHP has been broadly used for decision-making problems, the process is not without critics. The elements compared must be homogenous (relatively close) (Saaty, 1996). In addition, The AHP cannot deal with uncertainty (Saaty & Vergas, 1990). Due to dealing with vague data, fuzzy sets can be connected with the AHP model. In this manner, the criteria with vague weights are indicated by fuzzy numbers. A fuzzy number offers a continuous degree of membership between 0 and 1. Monitto, Papppalardo, and Toilo (2002) proposed a fuzzy AHP model for evaluating three kinds of manufacturing systems: flexible FMS, modified flexible production, and rigid transfer line while considering demand uncertainty through a case study. Similarly, Weck, Klocke, Schell, and Ruenauver (2002) added fuzzy logic to the classical AHP to evaluate different production cycles for an integrated product-process manufacturing system aimed at achieving an optimum degree of capacity utilisation and minimum environmental pollution. Extent analysis is an applicable method in group decision of the fuzzy AHP in which triangular fuzzy numbers are defined for pair-wise comparisons of the criteria at the same level. Consequently, triangular fuzzy weight vectors of the judgment matrix are normalised (Chang, 1996).

THE AHP AND FUZZY SETS: GENERAL DESCRIPTIONS

The most important feature of the AHP structure is the organisation of a complex problem through hierarchical levels. The first (highest) level defines a main goal of the decision problem and the last (lowest) level describes usually the decision alternatives or scenarios. The levels between the first and the last level can contain secondary goals, criteria ,and subcriteria of the decision problem. The number of the levels is not limited, but in the typical case it does not exceed four or five. Let us consider a simple three-level hierarchy that can represent a standard decision problem with finite set of alternatives: evaluation of n-alternatives A normal hierarchy is composed by: one global objective (goal), subobjective (strategic criteria), more specific subcriteria (technical criteria), and at the last level you can find the alternatives.

In order to evaluate relative importance of alternatives, pair-wise comparisons are performed for various criteria, which are difficult to quantify. Pair-wise weighing among *n* elements in each level leads to an approximation to the ratio of $a_{ij} = wi/wj$ which is the weight of element *i* to element *j*. The estimated weight vector *w* is found by solving the following eigenvector problem:

$$Aw = \lambda_{max} w \qquad (1)$$

where the matrix A consists of a_{ij}s, λ_{max} is the principal eigenvalue of A. If there is no inconsistency between any pairs of elements then $\lambda_{max\ j}$ is equal to n for any *i* and *j*, and we have:

$$A.w = n.w \qquad (2)$$

In reality, consistency does not usually take place and the Formulation (2.5) can be expressed as Aw = λ_{max}w =E, where E is the principal eigenvalue, a value around n (the total number of elements in the same level), and E is the eigenvalue. To estimate (E), each column of A is first normalised and then averaged over its rows. Eigenvector (E) is used to find the relative importance of each element with respect to the higher level of hierarchy. The inconsistency ratio (IR) is given as by $(\lambda_{max}-n) / (n-1)$ which is the variance of the error incurred in estimating matrix A. If an inconsistency becomes more than 10%, the problem and judgements must be investigated and revised (Saaty 1994).

To construct a fuzzy judgment matrix A we denote comparison quantitatively with the triangular fuzzy numbers as $a_{ij} = (l_{ij}, m_{ij}, u_{ij})$ in which m_{ij} is mid-value and can be valued from one to nine (1,2,..9) or in reverse (1/9, 1/8, ..1) as usually used in the AHP method (Chang, 1996, pp. 1-3). The synthesis judgement degree of a triangular fuzzy number located on the k^{th} layer can be derived from Formula (3).

$$S_i^k = \sum_{j=1}^{n} a_{ij}^k \otimes \left(\sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij}^k\right)^{-1} ;\ i = 1, 2, ..., n \qquad (3)$$

Under the fuzzy environment, the comparison ratios a_{ij} are described by membership functions. Fuzzy consistency can be described as the existence of relative weights within the fuzzy region (Leung & Cao, 2000, pp. 105-112).

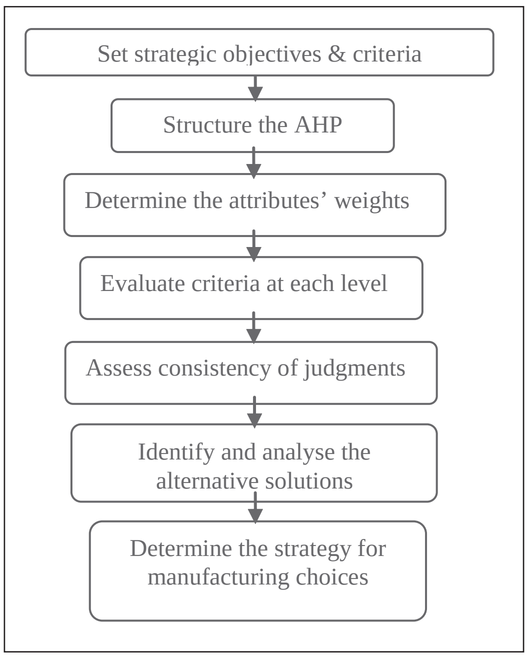
THE PROPOSED AHP MODEL

In this section, the AHP is employed as a multi-criteria strategic justification approach to select an MS type. The preference values are certain for crisp criteria, and the means for fuzzy criteria. In the model, the common design parameters for conventional MSs such as cost and quality plus new requirements such as responsiveness are taken into account. By trading off among all relevant objectives and criteria, process alternatives are evaluated. Figure 2 demonstrates a flowchart of the AHP steps.

In this section, an AHP model is proposed to systemise the process of MS choices, which reflects 'the preferred process technology' over planning horizons identified as the main goal (hierarchy level 0). Putting the decision elements altogether in a logical structure of a hierarchy can result in a structured framework as presented in Figure 2. The subsequent influencing parameters under the goal can be hierarchically categorised into five levels as follows:

- **Level 1:** Planning horizons.
- **Level 2:** Decision makers (actors).
- **Level 3:** Objectives of actors.
- **Level 4:** Criteria that satisfy those objectives.
- **Level 5:** Alternatives/decisions.

Figure 2. Flowchart of the AHP steps for process choice



The proposed model is generic and consists of technological/financial parameters that are valuable to many companies. In this respect, the hierarchy consists of general levels, that is, goal, criteria, and subcriteria as which can deal with wide range industrial of manufacturing companies. However, other levels of the hierarchy that include planning horizons, actors, and alternatives must be adapted according to the specific nature of the company under study. For example, feasible alternatives (manufacturing choices) may differ from a company to another because of influencing factors such as available technology, budget, and the volume and type of products to be manufactured. In the remainder of the section, the decision criteria shown in Figure 3 are discussed with their impacts on the required production process.

Impact of Planning Horizon (Level 1)

As shown in Figure 4, the design process of MSs can be hierarchically planned at three levels: strategic planning, tactical planning, and operational planning. The strategic design focuses on the economical part of the design process, whereas tactical level concerns the selection of parts, machines, pallets, and fixtures according to the master production plan. The operational planning deals with input-output at the real-time system in order to optimise the executive performance such as WIP and/or lead time.

One of the most important factors influencing the selection of a manufacturing process is the period of planning horizon. Planning horizon can be broken down to a number of time periods in order to reduce the uncertainty and risk caused over a long variable period. The first level of the AHP model deals with the three following planning periods for the process choice planning:

1. **Short Term (ST):** To redesign a MS to respond quickly to current demand variations (under two years).
2. **Mid Term (MT):** To (re)design a MS to demand changes while introducing a new

Figure 3. The AHP structure for the process choice

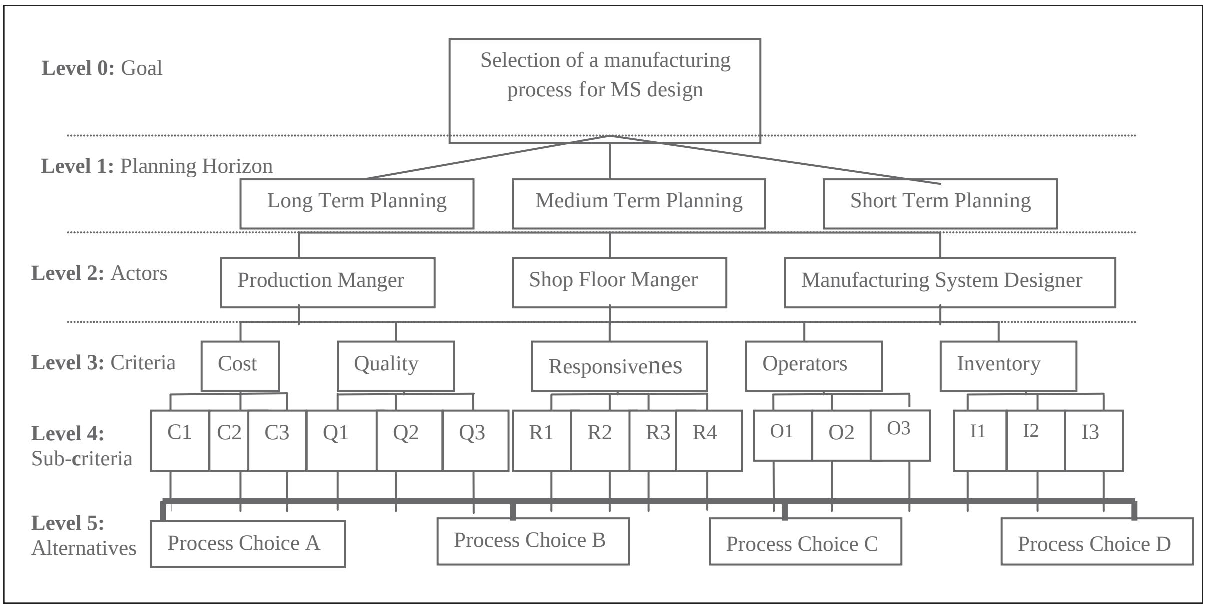
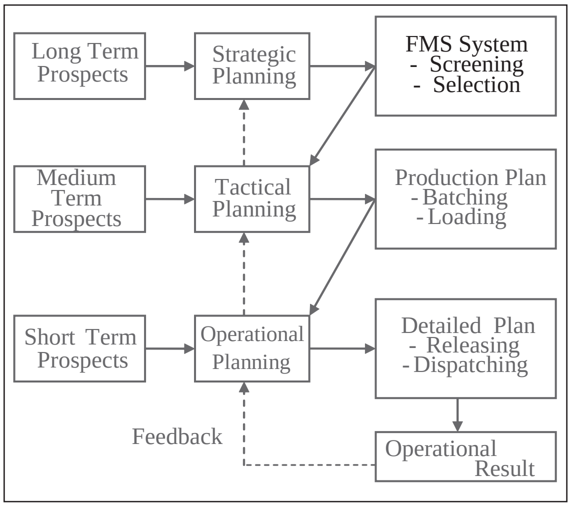


Figure 4. Hierarchical planning structure for flexible MSs (Kusiak, 1986)



product within a specified product family (between 2 and 5 years).

3. **Long Term (LT):** To design a MS to reflect any changes in products while introducing a new product family (over 5 years).

To synthesise the weights of the criteria for each time period, planning horizons themselves must be prioritised. The planning horizons' priorities can be elicited through an investigation of the manufacturing strategy of the plant by way of negotiation with all people deemed to be influencing the decision process such as managers, engineers, and system designers.

As an example, assume that three planning horizons with their priorities are given to compare for a decision over manufacturing choices. These are long term (LT), medium term (MD), and short term (ST) with absolute weights 5, 4, and 2 respectively. Therefore, the pair-wise matrix A can be formed whose rows give the ratios of the weight of each planning horizon with respect to all others as shown in Equation (4).

$$AW = \begin{array}{c c} & \begin{array}{ccc} LT & MT & ST \end{array} \\ \begin{array}{c} LT \\ MT \\ ST \end{array} & \begin{pmatrix} 5/5 & 5/4 & 5/2 \\ 4/5 & 4/4 & 4/4 \\ 2/5 & 2/4 & 2/2 \end{pmatrix} \end{array} \times \begin{pmatrix} 5 \\ 4 \\ 2 \end{pmatrix} = 3 \begin{pmatrix} 5 \\ 4 \\ 2 \end{pmatrix} \qquad (4)$$

To illustrate how to calculate Eigen vector (E), suppose that matrix A is given for the planning horizons with relative importance 5, 4, and 2 respectively, when there is an inconsistency (see section 2.5.3). As shown in Equation (5), E calculation concludes that the relative importance of impact of planning horizons for the design strategy of a MS are .687, .186, and .127 for LT, MT, and ST respectively.

$$AW = \begin{array}{c c} & \begin{array}{ccc} LT & MT & ST \end{array} \\ \begin{array}{c} LT \\ MT \\ ST \end{array} & \begin{pmatrix} 1 & 5 & 4 \\ 1/5 & 1 & 2 \\ 1/4 & 1/2 & 1 \end{pmatrix} \end{array} \times w = E =$$

$$\begin{pmatrix} \{1/(1+1/5+1/4) + 5/(5+1+1/2) + 4/(4+2+1)\}/3 \\ \{1/5(1+1/5+1/4) + 1/(5+1+1/2) + 2/(4+2+1)\}/3 \\ \{1/4(1+1/5+1/4) + 1/2/(5+1+1/2) + 1/(4+2+1)\}/3 \end{pmatrix}$$

$$= \begin{pmatrix} 0.687 \\ 0.186 \\ 0.12 \end{pmatrix} \qquad (5)$$

Decision-Makers (Level 2)

The AHP model emphasises the idea of team decision making by using input data from different groups within a company. This AHP model considers three actors for redesigning an existing MS toward a MS. An actor is an individual or a group, who plays a significant role in responding to forces that shape current events, and therefore future outcomes (Labib et al., 1996). Accordingly, the recommended actors influencing the decision making process in the proposed model are as follows:

- **Plant manager(s) (PM):** Top manager(s) of company who can evaluate the hierarchy of different criteria and provide judgement on the desirability of alternatives with respect to qualitative and intangible criteria.

- **Shop floor manager(s) (SM):** Top manager(s) of production system who can provide technological based performance data including feasibility and economical aspects of alternative manufacturing choices.
- **Manufacturing designer(s) (MD):** Top manager(s) of manufacturing design group who can support the decision process through evaluating and analysing entire hierarchy including impact of planning horizons, actors, objectives, and subobjectives on alternatives. MD can also provide technical information to evaluate feasibility of a MS choice and perform decision analysis and economic and risk analysis in order to validate the final decision. It is important to note that each of the actors above can be either a single manger and/or a group of experienced people working in the relevant departments of the plant.

Objectives and Criteria (Levels 3 and 4)

The strategic objectives towards designing MSs are identified as responsiveness (R), cost (C), quality (Q), inventory (I), and operators' skill (S). In order to facilitate an accurate decision analysis, all strategic objectives are broken into relevant criteria.

Product Cost (C)

Product cost can be decomposed into the following criteria:

- Raw material (C1), which includes all direct material used in manufacturing products.
- Process (C2), which includes the cost of capital investment on manufacturing equipment such as machines, tools, and material handling. Reusability can reduce extra investment for system reconfiguration.
- The operating cost consists of machine utilisation, operators running machines, and workers in the shop floor responsible for other tasks such as maintenance, transportation, quality control, and cleaning.
- Indirect cost (C3) consists of energy, engineers, and personnel officers at production planning, accounting, and so on.

Product Quality (Q)

Quality control of product can be undertaken in three steps:

- Raw material (Q1) is concerned with input quality of purchased material.
- Process (Q2) is concerned with quality of parts in the manufacturing, routes.
- Finished products (Q3), which is concerned with total quality of ordered products for delivery.

Responsiveness (R)

Manufacturing responsiveness is related to the ability of manufacturing systems to utilise its existing resources to make a rapid and balanced response to the predictable and unpredictable changes (Gindy & Saad, 1998). Obviously, different types of manufacturing choices have different levels of responsiveness. The proposed model gives great attention to this objective as a new manufacturing system requirement. This objective is then compared with other strategic objectives such as product quality and cost. Reusability as an economic/strategic factor significantly contributes to the rapid responsiveness (R). Four subobjectives (criteria) under the umbrella of responsiveness are considered in the hierarchy to evaluate the importance of responsiveness (R) over the MS alternatives. Those are:

- Wide variety of products (R1), which represents the ability of the plant to manufacture a range of products with different processing requirements.
- New product introduction (R2), which represents the ability of the plant to accept new design of products.
- Rapidly response to changes of product families using existing facilities (R3), which represents the ability of the plant to change its capacity and functionality with maximum reusability against demand fluctuations.
- Reduction of lead-time for product development (R4), which represents the ability of the plant to change tools for a given mix of products within a family with low ramp-up and set-up times. This criterion will be more important when batch sizes of product types within a family are very small and therefore set-up time of retooling machines must be short.

Operators Skills (O)

- Motivation (O1) encourages operators to activate extra effort for reconfiguring the system.
- Training (O2) facilitates the learning process for the changes of tasks when reconfigurations take place.
- Facilities type (O3) affects the required skill, for example, using dedicated and flexible machines need different levels of expertise.

Inventory (I)

- Raw material (I1) is the inventory in warehouse of the part to send to the system.
- Work In Progress (I2), (W2) is the inventory of parts in process before the manufacturing process is completed.
- Final product (I3) is the inventory of product before delivering to customers.

Alternatives (Level 5)

The last level of the hierarchy involves the specific manufacturing choices of a company for its process plan. The alternatives may differ from one company to another depending on the exiting system and feasible alternatives. Those choices might be classical systems such as dedicated system (DMS), and/or conventional systems such as cellular system (CMS) or flexible systems (FMS), and/or an advanced system such as reconfigurable MS (RMS). The latter can be characterised as a modular adjustable MS. The hybrid system (HMS) can also be defied for a possible combination of two MSs such as RMS and DMS. This means that some production lines are dedicated to the specific product lines and no changes are required for processing, and the other products are processed by an RMS.

FUZZY SETS AND THE AHP MODEL

According to the operations sequences required to process the products in the production range, quantitative and qualitative factors affecting the process choices are explored. The criteria for the process choice are not usually certain and the values may vary over time according to changes in the company strategy or external environments. In addition; the different actors may have different understanding of the criteria and therefore may prioritise them differently. As a result, the evaluation necessitates tolerating vague data. In this manner, consideration of fuzzy values along with the crisp values within an integrated AHP model can facilitate the determination of the suitable manufacturing process. As shown in Figure 5, the fuzzy multi-criteria decision approach consists of the following five steps:

1. Structuring the AHP model.
2. Identification of the fuzzy/crisp factors affecting performance of manufacturing process choices.
3. Quantification of the crisp factors and fuzzification of the fuzzy factors in comparison to the other factors at the same hierarchy level.
4. Defuzzification of the fuzzy attributes
5. Overall assessment of each alternative process.
6. Sensitivity analysis of the criteria within the fuzzy domain.

The Fuzzy Preference Scale

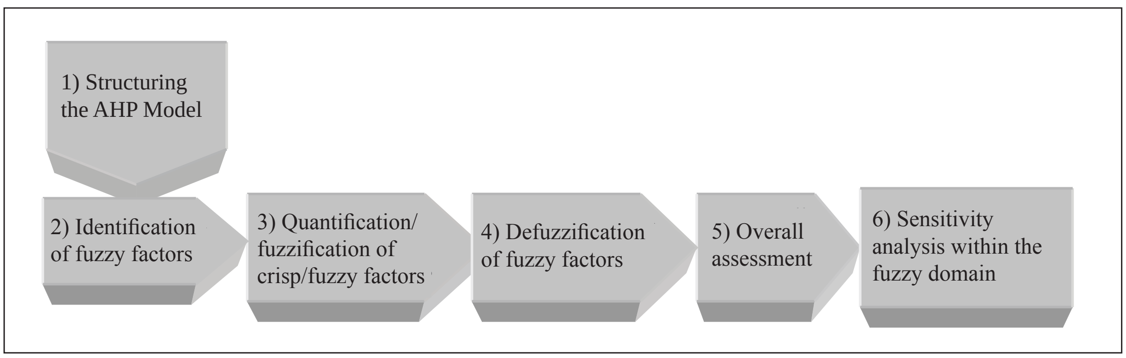
Due to uncertainty of the importance weights of pair-wise comparisons, the fuzzy elements of matrix A (see equation (1)) can be characterised by the fuzzy membership functions. The next step is to identify the relative importance of each pair factors in the same hierarchy level. By using fuzzy membership functions such as triangular fuzzy numbers, pair-wise comparisons can be undertaken. In this way, fuzzy evaluation matrix A of $\hat{a}_{ij}$ elements is constructed, in which $\hat{a}_{ij}$ = (l, m, u) is the importance of element *i* over element *j* under a certain criterion with lower (l), mean (m), and higher (h) values respectively. The value (u - l) represents a fuzzy degree of judgment. The greater (u- l), the fuzzier the degree; when u - l = 0 the judgment is a nonfuzzy number with *m* importance value. Fuzzy consistency can be described as the existence of relative weights within the region (Leung & Cao, 2000). Subject to the fuzzy consistency, $\hat{a}_{ij}^{-1}$ = (1/u, 1/m, 1/l) represents the fuzzy importance of element *j* over element *i* with lower value (1/h), mean (1/m), and higher value (1/l). As a result, the fuzzification increases the complexity of computational operations for synthesis judgments, which are basically performed on the fuzzy elements ($\hat{a}_{ij}$ s).

The fuzzy values are standardised into a single-pattern **fuzzy set** dealing with both linguistic and quantifiable criteria. Therefore, the fuzzy model can be simply evaluated through the typical pair-wise comparisons.

For the case study, the important weights are defined with five triangular fuzzy sets $\hat{1}, \hat{3}, \hat{5}, \hat{7}, \hat{9}$ with their corresponding lower, mean, and upper values defined in Equation (6).

$$\hat{a}_{ij} = \begin{cases} \hat{1} & ; \ \in (1,1,3) \\ \hat{x} & ; \ \in (x-2,x,x+2);\ 1 < x < 9 \\ \hat{9} & ; \ \in (7,9,9) \end{cases} \tag{6}$$

Figure 5. The fuzzy AHP decision approach



As shown in Figure 6, the fuzzy range of ($\hat{1}$, $\hat{3}$, $\hat{5}$, $\hat{7}$, $\hat{9}$) is used to express linguistic priorities for both quantitative and qualitative criteria. These model criteria can be compared and measured by using the fuzzy linguistic priorities in terms of equal (EQ), low (L), medium (M), high (H), and very high (VH).

The pair-wise comparison can then be undertaken by using the preference bar as outlined in Table 1. For example, the quantified preference values of EQ and VH for an element to another element are distinguished by 1 and 9 respectively. For example, if value 5 is assigned to criterion (c_j) at the right side of the bar, the criterion c_j will be more important than c_i with a moderate degree. Similarly, if value 5 is assigned to criterion (c_i) at the left side of the bar, the criterion c_i will be more important than c_j with a moderate degree. The corresponding values of the preferences can also appear between any two fuzzy linguistic preferences. In addition, if one of the nonzero numbers is assigned to criterion (c_i) when compared with criterion (c_j), then c_j has the reverse preference value when compared with c_i. For example, if functionality retains moderate importance value 3 comparing with capacity with respect to manufacturing process, then capacity posses the relative importance value 1/3 to functionality. Similarly, the reverse fuzzy value for the linguistic score $\hat{3}$ can be represented by $\hat{3}^{-1}$ or (1/$\hat{3}$).

The synthesis judgement degree of the triangular fuzzy numbers located on the k[th] layer can be derived by using Formula (3). The decision makers would like to determine S_i^k in order to find how much each alternative process can contribute to each objective and/or criterion. The following example illustrates a modification of fuzzy judgment for choosing a manufacturing system choice

Figure 6. Fuzzy linguistic priorities

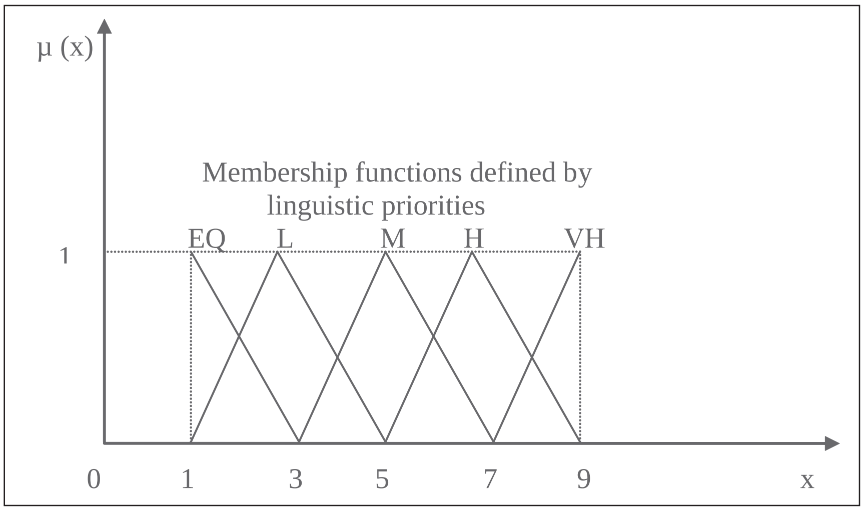


Table 1. Linguistic priorities and the quantified values between two criteria i and j

Criterion i	VH	H	M	L	EQ	L	M	H	VH	Criterion j
	9	7	5	3	1	3	5	7	9	

Table 2. The fuzzy pair-wise comparison matrix of the alternative process with respect to R

Alternative \ Alternative	DMS	HMS	RMS
DMS	1	$\hat{3}$	$\hat{1}$
HMS	$1/\hat{3}$	1	$\hat{5}$
RMS	$1/\hat{1}$	$1/\hat{5}$	1

among the three alternatives: DMS, HMS, and RMS. The alternatives are pair-wise compared with respect to manufacturing responsiveness (R) as shown in Table 2.

A CASE STUDY

In this section, a case study is undertaken to demonstrate the synthesis of criteria using the preference values for the crisp factors and the mean values for fuzzy factors within the classical AHP. All the factors including quantitative/quantitative ones and crisp/fuzzy ones are considered within the same assessment structure. However, for the fuzzy factors which seem to be curtail to the actors such as planning horizon or responsiveness, a sensitivity analysis will be undertaken within the fuzzy domain using the software.

Synthesis of Criteria Using Mean Values

The assessment process starts with pair-wise comparisons between planning horizons by managers in order to prioritise impact of each period on the design strategy. An investigation on the strategic and tactical plans of the firm under study is necessary to prioritise its planning horizons. As presented in Table 1, the importance of LT, MT, and ST with respect to goal is sorted as 0.687, 0.186, and 0.127 respectively. The E's values are given in the last column and inconsistency ratio (IR) is equal to 0.09. Needless to say that the number on the diagonal of the matrix will be '1' which shows that there is no priority between a parameter with itself.

Similarly, pair-wise comparisons between actors (PM, SM, MD) with respect to each planning horizon are presented in Tables 3 through 6.

The synthesis of the actors' comparison matrix with respect to planning horizons results in a likelihood matrix. As a result, a likelihood matrix can be obtained from matrices above as presented in Table 7. This matrix will determine the expected weight for each actor over each planning horizon. The expected weight of each actor is multiplied by the weight of each planning horizon (as presented in parentheses), and then added up over the corresponding row for obtaining the global importance of the actor as presented in the last column. For example, the global importance of PM equals to 0.54(0.371)+0.117(0.019)+ 0.678(0.086)= 0.476.

The next hierarchy process is to compare objectives and criteria by actors. As shown in Tables 8 through 13, pair-wise comparisons matrices of objectives and subobjectives are filled by plant managers to identify the priority of each objective against the other from their points of view. The same pair-wise comparisons tables must also be fulfilled by the other actors. E and IR values are given in the last column and the last row of each table respectively. For example, the criteria evaluation by plant manager with respect to LT results in the objectives' priorities as illustrated in Table 8, where IR is equal to 0.06.

According to data gathered from the above matrices, total combined weight of attributes with respect to PM and LT are calculated and presented in Table 14. It would also be interesting to see the value of f2 (skills of dedicated or multipurpose facilities) with respect to S with the maximum global priority having value 0.186 among all criteria.

Table 3. A comparison matrix of planning horizons

Impact of planning horizons	LT	MT	ST	EV
LT	1	5	4	0.687
MT	1/5	1	2	0.186
ST	1/4	1/2	1	0.127
Inconsistency rate = 0.09				

Table 4. A comparison matrix of decision makers with respect to LT

LT	PM	SM	MD	EV
PM	1	3	2	0.540
SM	1/3	1	2	0.163
MD	½	1/2	1	0.297
Inconsistency rate = 0.01				

Table 5. A comparison matrix of decision makers with respect to MT

MT	PM	SM	MD	EV
PM	1	4	3	0.117
SM	1/4	1	2	0.268
MD	1/3	1/2	1	0.614
Inconsistency rate = 0.03				

Table 6. A comparison matrix of decision makers with respect to ST

ST	PM	SM	MD	Eigenvector
PM	1	3	4	0.678
SM	1/3	1	3	0.101
MD	¼	1/3	1	0.226
Inconsistency rate = 0.08				

Table 7. A Likelihood matrix of decision makers

Goal	LT (.687)	MT (.168)	ST (.127)	Global importance of actors
PM	.540 (.371)	.117 (.019)	.678 (.086)	0.476
SM	.163 (112)	.268 (.045)	.101 (.013)	0.175
MD	.297 (204)	.614 (.103)	.226 (.029)	0.336

Table 8. A comparison matrix of objectives by PM with respect to LT

			Strengths of Objectives Plant Manager (PM)		
MS design	**R**	**C**	**Q**	**I**	**S**	**E**
R	1	1	3	3	3	0.137
C	1	1	1	4	3	0.184
Q	1/3	1/3	1	4	1	0.279
O	1/3	1/4	1/4	1	3	0.064
I	1/3	1/3	1	1/3	1	0.336
Inconsistency rate = 0.06						

Table 9. A comparison matrix of criteria by actors with respect to C

		Objective: Product Cost		
C	**C1**	**C2**	**C3**	**EV**
C1	1	4	1	0.167
C2	1/4	1	4	0.667
C3	1	1/4	1	0.167
Inconsistency rate = 0.0				

Table 10. A comparison matrix of criteria by actors with respect to Q

		Objective: Product quality		
Q	**Q1**	**Q2**	**Q3**	**EV**
Q1	1	1	1	0.337
Q2	1	1	1	0.333
Q3	1	1	1	0.333
Inconsistency rate = 0.0				

Table 11. A comparison matrix of criteria by actors with respect to R

		Objective: Responsiveness			
R	**R1**	**R2**	**R3**	**R4**	**EV**
R1	1	4	4	4	0.073
R2	1/4	1	1	3	0.214
R3	1/4	¼	1	3	0.214
R4	1/4	1/3	1/3	1	0.499
Inconsistency rate = 0.06					

Table 12. A comparison matrix of criteria by actors respected to R

		Objective: Operator's Skill		
O	**O1**	**O2**	**O3**	**EV**
O1	1	1	3	0.210
O2	1	1	2	0.240
O3	1/3	1/2	1	0.550
Inconsistency rate = 0.02				

Table 13. A comparison matrix of criteria by actors with respect to I

		Objective: Inventory		
I	**I1**	**I2**	**I3**	**EV**
I1	1	3	2	0.320
I2	1/3	1	4	0.122
I3	1/2	1/4	1	0.558
Inconsistency rate = 0.02				

Table 14. A combined priority matrix of criteria by actor PM

Objective	**criteria**	**Weights**	**Combined weights**
C (.184)	C1 C2 C3	0.167 0.667 0.167	0.030 0.123 0.030
Q (.279)	Q1 Q2 Q3	0.337 0.333 0.333	0.093 0.093 0.093
R (.137)	R1 R2 R3 R4	0.073 0.214 0.214 0.499	0.010 0.029 0.029 0.068
O (.336)	O1 O2 O3	0.210 0.240 0.550	0.070 0.080 0.186
I (.064)	I1 I2 I3	0.320 0.122 0.558	0.020 0.008 0.036

Sensitivity Analysis Inside the Fuzzy Domain

A case study is undertaken to demonstrate the model and analyse the decision problem of manufacturing choice. The AHP inputs must be derived from the data gathered from manufacturing company A with feasible alternatives DMS, HMS, and RMS defined for the planning horizons. Appropriate software is then required to derive actual data for the comparison of defined criteria and feasible alternatives. The software facilitates structuring, documenting, and analysing the decisions and provides quick restructuring of the model along with the sensitivity analysis. The systematic approach of the model assists managers to better understand the process that they require for future planning for investing in a new manufacturing process when needed.

Process Choice Assessment

As shown in Figure 7, HMS is placed as the best alternative for the mean value of the fuzzy long term planning criterion (LT) with IR= 0.07. The slope of the line verifies its choice in comparison with the two other options, DMS and RMS, in the long term. There are no changes of the preferences between the two fuzzy domain presented by the two dashed vertical lines.

Surface of decision is the area that the decision makers can evaluate the alternatives within the fuzzy domain. Accordingly, decisive managers would suggest maintaining some production lines as before and a reconfigurable process (RMS) for a few existing/new products. Although changes in the importance degree of each actor and/or criteria result in changes of the alternative priorities, increase in the importance of actor PM has not changed the overall solution of HMS>RMS>DMS. In the short term planning, the overall solution will result in the same solution with IR=0.08.

Process Choice with Respect to Responsiveness

Responsiveness (R) is one of the key characteristics for advanced manufacturing systems which reflect the level of market demand changes and

Figure 7. Criteria assessment in the long term

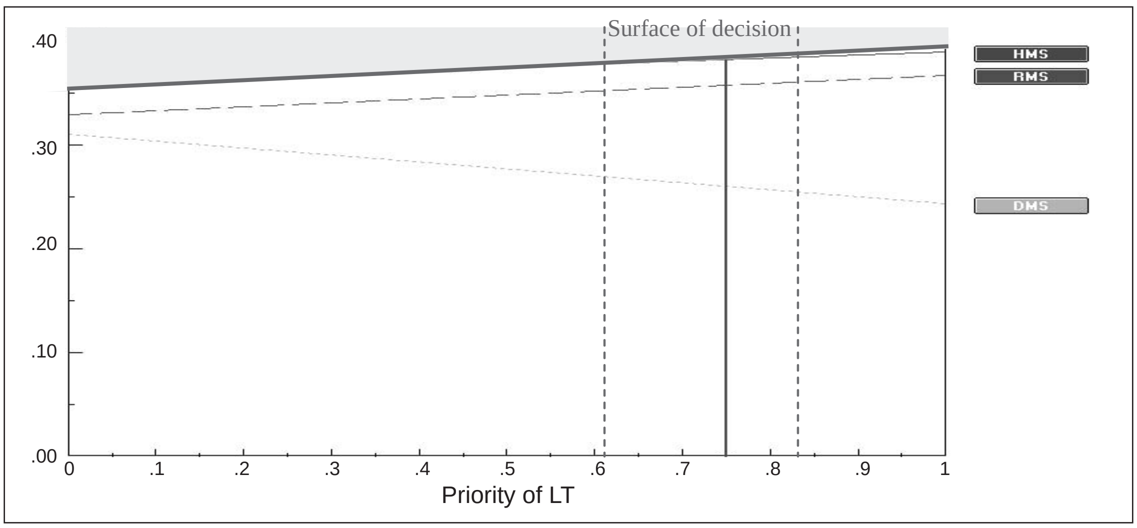
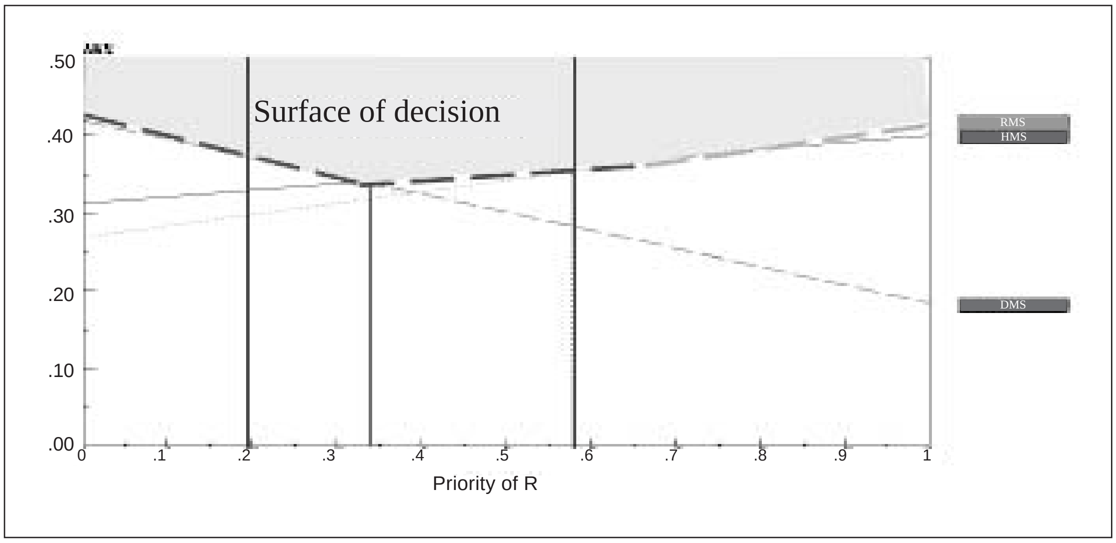


Figure 8. Criteria assessment over manufacturing responsiveness



customer satisfaction. As shown in Figure 8, the horizontal axis represents the priority of R. For the current priority mean (the bold vertical line), the solution remains the same as before with HMS = DMS>RMS at the intersection point. Surface of decision is located between the two dashed vertical lines representing the lower and upper values of fuzzy R. This is the area that the decision makers can evaluate the alternative process within the fuzzy domain of R. By increasing the priority of R via shifting the vertical line to the right, the solutions will change to HMS>RMS= DMS and then RMS>HMS>DMS respectively. Similarly, by decreasing the priority of R towards the lower fuzzy value the solution will change to DMS>HMS>RMS. Having assessed the results for the planning horizons, it can be found that the responsiveness degree (R) significantly affects the manufacturing solutions particularly in the short/medium term planning.

CONCLUSION

The chapter develops a fuzzy AHP model for the selection of the most appropriate manufacturing choice. The proposed model takes both quantitative and qualitative criteria with either crisp or fuzzy preference values into account. Manufacturing responsiveness, cost, quality, and operators' skills are considered as the main objectives. The triangular fuzzy linguistic scale is then utilised for adjusting common preference values for pairwise comparisons. The alternatives are compared and analyzed via monitoring sensitivity analysis within their fuzzy domains of the defined criteria through a case study.

The proposed model is intended to be generic and consists of parameters that are valuable for process selection in many firms. However, the criteria and alternatives can differ from a company to another because of differences in product types, process requirements, available technologies, and economical circumstances.

The mean values for the fuzzy criteria are used for the preliminary judgment over the process choices. The alternatives are then compared and analyzed via sensitivity analysis within the fuzzy range of the defined criteria. This seriously moderates the computational time and effort required to achieve the fuzzy analysis. Although the proposed fuzzy AHP facilitated the utilisation of the classical AHP with fuzzy values, the sensitivity analysis of a criterion is carried out within the fuzzy range while assuming other criteria are maintained certain. This is not a precise assumption for different situations in dynamic manufacturing environments. In addition, complexity in calculating inconsistency decreases the model validation.

FUTURE RESEARCH DIRECTIONS

Customer orientation and fast adaptation of new technologies in processes continuously change the structure of manufacturing. Future factories require new and innovative processes and flexible configurations of products and processes. The adaptability will significantly depend on the system architecture and standardisation of products and processes over planning horizons. Accordingly, fast production planning will be one of distinguishing features for the future factories. The identification of feasible manufacturing processes, renovation of manufacturing requirements, updating market requirements, and the systematic evaluation of critical factors can facilitate the high speed decision process for the selection of appropriate MS choice for each planning period.

The proposed model is a structured dynamic decision by introducing planning horizon as the elements in the structure. This is a classical way without using judgments as mathematical function of time. The other analytical way to deal with decisions over time is to functionally involve the time in the judgment. In this regard, each variable criterion can be a function of time and therefore, the eigenvector will be time dependent. This approach generates complex equations in the pair-wise matrix to be solved. The complexity of synthetic sharply increases when the number of matrix elements increases.

The initial hierarchy could be extended to the ANP as to be more flexible by allowing interactions among the high level elements and the lower level ones. The ANP model could deal with all kinds of interdependencies among parameters defined in the AHP model using the super matrix, which provides an integrated structure of the pair-wise comparisons.

REFERENCES

Abdul-Hamid, Y.T., Kochhar, A.K., & Khan, M.K. (1999). An analytic hierarchy process approach to the choice of manufacturing plant layout. *IMechE, Proceeding of Institute of Mechanical Engineers, 213*(Part B), 397-406.

Chan F.T.S., Chan, H.K., & Chan, M.H. (2003). An integrated fuzzy decision support system for multi-criteria decision-making problems. *IMechE, Proceeding of Institute of Mechanical Engineering, Journal of engineering Manufacture, 217*(Part B), 11-27.

Chang, D.Y. (1996). Theory and methodology: Application of the extent analysis method on fuzzy AHP. *European Journal of Operational Research*, (95), 649-655.

Chick, S.E., Olsen, T.L., Sethuraman, K., Stecke, K.E., & White, C.C. (2000). A descriptive multi-attribute model for reconfigurable machining system selection examining buyer-supplier relationships. *International Journal of Agile Management Systems, 2*(1), 33-48.

Expert Choice Software. (1999). *Expert choice inc*. Retrieved October 5, 2000, from http://www.expertchoice.com/

Labib, A.W., & Shah, J.(2001). Management decisions for a continuous improvement process in industry using analytical hierarchy process. *Work Study a Journal of Productivity Science*, *50*(4-5), 189-193.

Leung, L.V., & Cao, D. (2000). Theory and methodology on consistency and ranking of alternatives in fuzzy AHP. *International Journal of Operational Research*, (24), 102-113.

Maier-Speredelozzi, V., & Hu, S.J. (2002). *Selecting manufacturing system configurations based on performance using AHP*. Paper presented at the Transactions of NAMRI (pp. 637-644), West Lafayette, Indiana.

Monitto, M., Papppalardo, T., & Toilo, T. (2002). A new fuzzy AHP method for the evaluation of automated manufacturing systems. *Annals of the CIRP, 51*(1), 395-398.

Oeltjenbruns, H., Kolarik, W.J., & Schenadt-Kirschner, R. (1995). Strategic management in manufacturing systems-AHP application to an equipment replacement decision. *International Journal of Production Economics*, *38*, 189-197.

Saaty, T.L. (1980). *The analytic hierarchy process*. New York: McGraw Hill.

Saaty, T.L. (1996). *Decision making with dependence and feedback: The analytic network process*. Pittsburgh: RWS Publications.

Saaty, T.L. (2005). *Theory and applications of the analytical network*. Pittsburgh: RWS Publications.

Saaty, T.L., & Vergas, L.G. (1990). *Multi-criteria decision making analytical hierarchical process*. Pittsburgh: RWS Publications.

SuperDecision Software. Retrieved October 7, 2006, from http://www.superdecisions.com/

Weck, M., Klocke, F., Schell, H., & Ruenauver, E. (1997). Evaluating alternative production cycles using the extended fuzzy AHP method. *International Journal of Operational Research*, *100*, 351-366.

Willcocks L., & Lester, S. (1991). Information systems investments: Evaluation at the feasibility stage of projects. *Technovation, 11*, 283-302.

Xiaobo, Z., Jiancai, W., & Zhenbi, L. (2001). A stochastic model of a reconfigurable manufacturing system, part 3: Optimal selection policy. *International Journal of Production Research, 39*(4), 747-758.

Yang, S., & Hu, S.J. (2000). Productivity analysis of a six CNC machine manufacturing system with different configurations. In *Proceeding of the 2000 Japan-USA Flexible Automation Conference,* Ann Arbor, Michigan, (pp. 499-505).

ADDITIONAL READING

Abdi, M.R., & Labib, A.W. (2003). A design strategy for reconfigurable manufacturing systems (RMSs) using the analytical hierarchical process (AHP): A case study. *International Journal of Production Research, 41*(10), 2273-2299.

Abdi, M.R., & Labib, A.W. (2006). Chapter 23: Product design for transformable production and reconfigurable manufacturing. *Reconfigurable manufacturing systems and transformable factories* (pp 460-478). New York: Springer. ISBN-3-540-29391-4.

Hafeez, K., Zhang, Y-B., & Malak, N. (2002). Determining key capabilities of a firm using analytical hierarchical process. *International Journal of Production Economics, 76*, 39-51.

Kahraman, C., Tolga, E., & Ulukan, Z. (2000). Justification of manufacturing technologies using fuzzy benefit/cost ratio analysis. *International Journal of Production Economics*, *66*, 45-52.

Karmarkar, U., & Kekre, S. (1989). Manufacturing configuration, capacity and mix decisions considering operational costs. *Journal of Manufacturing Systems, 6*(4), 315-324.

Kulak, O., Durmusoglu, M.B., & Kahraman, C. (2005). Fuzzy multi-attribute equipment selection based on information axiom. *Journal of Materials Processing Technology, 169*(3), 337-345.

Labib, A.W., & Shah, J. (2001). Management decisions for a continuous improvement process in industry using analytical hierarchy process. *Work Study a Journal of Productivity Science, 50*(4-5), 189-93.

Mehrabi, M.G., Ulsoy, A.G., & Koren, Y. (2000). Reconfigurable manufacturing systems: Key to future manufacturing. *Journal of Intelligent Manufacturing, 11,* 413-419.

Vanegas, L.V., & Labib, A.W. (2001). Application of new fuzzy-weighted average (NFWA) method to engineering design evaluation. *International Journal of Production Research, 39*(6), 1147-1162.

Vanegas, L.V., & Labib, A.W. (2005). Fuzzy approaches to evaluation in engineering design. *ASME Journal of Mechanical Design, 127*(1), 4509-4537.

Zadeh, L.A. (1965). Fuzzy sets. *Information and Control, 8*, 338-353.

Chapter V
Computational Intelligence Approach on a Deterministic Production-Inventory Control Model with Shortages

Supriyo Roy
Bengal Engineering and Science University, India

S. Mukhopadhyay
Burdwan University, India

P. P. Sengupta
National Institute of Technology, India

ABSTRACT

Here, an attempt has been made to determine an optimal solution of a deterministic production-inventory model that consists of single deteriorating items and a constant rate of deterioration. In this proposed production-inventory model, lead time is taken to be negligible and demand rate is a ramp type function of time. Shortages are allowed and partially backlogged. During this shortage period, the backlogging rate is a variable which depends on the length of the waiting time over the replenishment period. Mathematical formulation of the problem highlighted the model as a complex nonlinear constrained optimization problem. Considering the complexities towards solution, modified real-coded genetic algorithms (elitist modified real coded genetic algorithm [MRCGA]) with ranking selection, whole arithmetic crossover, and nonuniform mutation on the age of the population has been developed. The proposed production-inventory model has been solved via MRCGA and simulated annealing and as well as standard optimization methods. Finally, the results are embedded with numerical example and sensitivity analysis of the optimal solution with respect to the different parameters of the system is carried out.

INTRODUCTION

In production-inventory management model, it is observed by the management scientists that the deterioration of an item plays a pertinent role in mathematical modeling and computation. With fluctuating nature of demand of end users', stock of items continuously decreases due to deterioration effect. Many commonly used items which are either damaged or decayed are not in a perfect up-to-date condition to satisfy the time varying and conflicting nature of demand of customer. This deterioration effect depends on preserving facility and environmental condition of warehouse/storage. As a result, this effect cannot be directly ignored from comprehensive analysis of the production-inventory model.

Recently, a number of mathematical models have been highlighted in the literature for estimating order quantity for deteriorating items (Ghare & Schrader, 1963). They first pointed out that the inventory is depleted not only by demand but also by decay (direct spoilage or physical depletion). They formulated an inventory model for exponentially decreasing inventory and constant rate of demand. Shah and Jaiswal (1997) developed an order-level inventory model for deteriorating items with a constant rate of deterioration. After that, Aggarwal (1978) formulated the same by correcting and modifying the errors in Shah and Jaiswal's analysis in calculating the average inventory holding cost.

In most of the inventory models, the demand rate is considered to be either constant or time-dependent but independent of the stock status. However, in the present market scenario, customers are influenced by the marketing policies-like attractive display of units in the market/business firm. Items like fruits, vegetables, fashionable commodities, and so forth and displays of those units in huge numbers has a motivational effect on people to buy more; demand is influenced by stock status, termed as stock dependent demand.

Again, inventory problems with deterministic time dependent demand patterns gain serious attention to several researchers. Demand of this type directly affects the sales volume in different phases of product life cycle. Again, demand for inventory items shows a conflicting in nature: it increases with time in growth phase and decreases in the decline phase. Donaldson (1977) first developed an inventory model with a linearly increasing time dependent demand rate over a finite planning horizon. After Donaldson, several researchers, such as Goyal and Aggarwal (1981), Giri, Goswami, and Chaudhuri (1996), Bhunia and Maiti (2001), and many more have developed this type of model by incorporating a time varying demand rate into their models for deteriorating and nondeteriorating items with or without shortages. But dealing with this type of complex analysis, management scientists are not always successfully implementing the traditional optimization methodologies for solving such a model; advanced methods are required to search for solving these complicated mathematical models.

In management science, dealings with decision making problems, generally with the traditional direct and gradient-based optimization methods, (conjugate gradient methods like Fletcher-Reeves, Polak-Ribiere, etc., and Quasi-Newton methods like Davidson-Fletcher-Powell, Broyden-Fletcher-Goldfarb-Shanno, etc.) are used for computation to get "optimal" or "near optimal" solution. But these methods do not process towards achieving reasonable solutions for difficult combinatorial mathematical optimization problems. Again, these methods are:

1. Dependent only on initial solution, are not universal; rather problem dependent.
2. Are getting trapped at a "local" optimum,
3. are not judiciously reducing the search space.
4. Are not amenable to parallel processing, that is, investigation of different solution sequences can not be done in parallel.

5. Do not have anymore adjustable parameters that permit flexibility.

To overcome the above limitations, there has been a growing interest in optimization algorithms and methods, popularly termed as computational intelligence (CI). This is a recently emerging area of fundamental and applied research exploiting a number of advanced information processing technologies (Koner, 2005). It is basically the study of adaptive mechanism to enable or facilitate intelligent behavior in complex and changing environment. This is also related with artificial intelligence, the science of creating a nonhuman intelligence with machines and/or computers. Although fairly widespread, there is no commonly accepted definition of the term computational intelligence. Attempts to define or at least to circumscribe this usually falls in one or more of the following categories:

1. Conceptual treatment of key notions and their roles in computational intelligence.
2. Relative definition comparing computational intelligence to artificial intelligence.

Computational intelligence is, therefore, a name for the combined and collective fields of evolutionary and neural computing, DNA computing, and quantum computing.

Again, evolutionary computing (EC) and GA are based on the principle of natural (Darwinian) evaluation for generating and evaluating a population of possible solution to a problem. As they are based on evolutionary learning, they come under the arena of computational intelligence. Such genetic systems have some selection process based on fitness of individuals and some genetic operations. Recently, several types of advanced computational intelligence methods, such as genetic algorithms (GA), simulated annealing, tabu search, and so forth have been used by the management scientists for parameter optimization, classification, and learning in a wide range of applications of managerial decision making problems. According to Goldberg (1989), Davis (1991), Michalewicz (1992), Mitchell (2004), and Sakawa (2002), genetic algorithms are adaptive computational procedures which are modeled as the mechanics of natural genetic systems. They express their abilities by efficiently exploiting the historical information to speculate on new offspring with expected improved performance. These algorithms are executed iteratively on a set of randomly generated coded solutions called population. In each iteration, three basic genetic operations, that is, selection, crossover, and mutation are performed. Professor J. H. Holland of University of Michigan envisaged the fundamental concept of this algorithm in the mid sixties and published his seminal work (Holland, 1975). Since a genetic algorithm works simultaneously on a set of randomly generated coded solutions, it has very little chance to get stuck at “local” minima. In the recent year, a number of researchers have contributed their work and tried to implement this methodology in several implicational fields like traveling salesman problems by Forrest (1993), scheduling problem by Davis (1991), mumerical optimization by Michalewicz (1992), and many more. Other developmental research work on computational intelligence based optimization refers to aircraft landing strut weight optimization by Minga (1986), combinatorial optimization by Papadimitriou and Steiglitz (1982), communication network link size optimization by Davis and Coombs (1987), recursive adaptive filter designing by Etter, Hicks, and Cho (1982), selection of detectors for binary pattern recognition by Cavicchio (1972), fitting of potential surfaces via GA by Schaffer (1985), calibration of population migration model via GA by Smith and Jong (1981), and many more. But, until now, only a very few researchers have applied computational intelligence methodologies in the field of managerial decision making problems like manufacturing and production-inventory control model. Extensive research work in this field includes work by

Sarkar and Lutfar (1999), Sarkar and Newton (2002) and Roy, Supriyo, Bhunia, and Mukhopadhyay (2005).

The present work is concerned basically with the famous work of Kun-Shan (2000) who formulated an inventory model for deteriorating items in the prescribed cycle length by assuming the rate of demand as a ramp type function of time. Demand pattern of this type is generally refers as to any new brand of consumer goods or pharmaceutical drugs, and so forth introduced in the competitive market. For such a new brand, the demand rate generally increases with time up to certain time and then it becomes constant. The time for which the demand becomes constant falls under either stock-in or stock-out period. Two possible cases may arise according to the relative size of stock-in period. Kun-Shan has considered only one case.

Here we have developed a production-inventory model for a single item that deteriorates at a constant rate. The demand rate is a ramp type function of time. Shortages are allowed and partially backlogged. Backlogging rate is totally dependent on the waiting time. To analyze such a situation, we take two possible cases. The mathematical formulation of the problem indicates that the model is a nonlinear constrained optimization problem. Considering the complexities of mathematical modeling, we explore the solution methodology by modified real coded genetic algorithm (MRCGA) with ranking selection, nonuniform mutation, and considering the property of elitism. This problem is again solved by another advanced search procedure via simulated annealing (SA). Finally, numerical example is evaluated to illustrate all the results and also the significant features. The chapter ends with a numerical presentation of highlighting the effects of changes in the parameters on the optimal profit, highest stock level, shortage level, and time-period.

MODEL DESCRIPTION

We follow up the basic assumptions and notations by following Roy et al. (2005) with some modifications.

Assumptions

1. Replenishment rate is infinite but replenishment size is finite. Nature of replenishment is instantaneous without lead time.
2. The inventory system deals with only one item and one stocking point.
3. Only a single order will be placed at the beginning of each cycle and the entire lot is delivered in one lot.
4. Lead-time is considered as zero.
5. There is no repair or replacement of deteriorated units in the given cycle.
6. Discounted quantity is negligible.
7. The demand rate $D(t)$ is assumed to be a ramped type function of time. μ is the time period where demand is time dependent up to $t = \mu$.
8. Shortages, if any, are allowed and partially backlogged.
9. Inspection, if any, is nondestructive and error-free.
10. The units in inventory deteriorate at a rate θ $(0 \leq \theta \leq 1)$ during the cycle time.

Notations

1. T is the total time period/planning horizon.
2. S is the highest stock level in the beginning of the cycle after fulfillment of the backlogged quantity, if any, and R is the highest shortage level.
3. Holding cost is C_1 / unit / unit time, shortage cost is C_2 / unit / unit time, purchasing

cost is C_3 / unit, and ordering cost is C_4 / per order.

4. Inspection cost is C_5 / unit and selling price s is known and constant.
5. C_{hold} and C_{short} denote the total inventory carrying cost and shortage cost respectively.
12. $I(t)$ is the on-hand inventory at time t over the planning horizon.

According to the above assumption the demand rate $D(t)$ takes the following form:

$$D(t) = D_0[t - (t - \mu)H(t - \mu)];\ D_0 \succ 0$$

where $H(t - \mu)$ is a well known *Heavisides' function* defined as follows:

$$H(t - \mu) = 1;\ t \geq \mu$$
$$= 0;\ t \prec \mu$$

MATHEMATICAL MODEL FORMULATION AND SOLUTION

Initially, a company starts with purchasing an amount of $(S + R)$ units of an item. Production begins at time $t = 0$, continues up-to time $t = t_1$, and during this time period, after fulfilling the backorder quantities, initial stock level reaches to a level S. After that, inventory level decreases due to both end users demand and deterioration until it becomes zero at time $t = T$. During the time interval (t_1, T), shortages, if any, are accumulated and all the demand in that interval is backlogged with backlogging rate $[1+\delta(T-t)]^{-1}$.

Following the assumptions of rate of demand $D(t)$, demand of an item is dependent on μ. We, therefore, have two types of equations:

$$D(t) = D_0 t;\ 0 \leq t \leq \mu$$
$$D(t) = D_0 \mu;\ \mu \leq t \leq \mathrm{T}$$

Based on these equations, two cases may occur:

Case 1: $t_1 \leq \mu$ and **Case 2:** $t_1 \geq \mu$

The pictorial diagram may be shown in Figure 1 and Figure 2 for both the cases.

Case 1: $t_1 \leq \mu$

Here, order quantity is $Q = (S + R)$. The differential equations describing the instantaneous states of the inventory level $I(t)$ over the planning horizon of this system is given by:

Figure 1. The inventory situation of Case 1

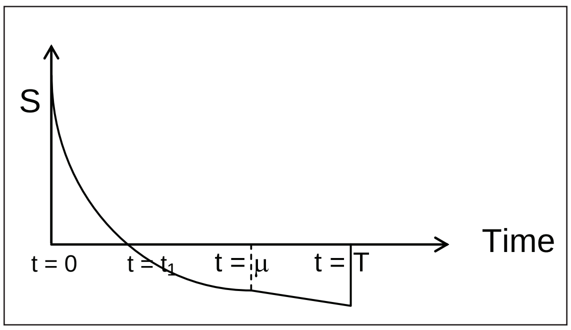
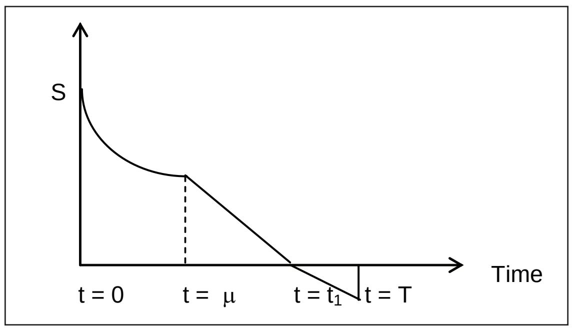


Figure 2. Inventory situation of Case 2



$$\frac{dI(t)}{dt} + \theta I(t) = -D_0 t; 0 \le t \le t_1 \tag{1}$$

$$\frac{dI(t)}{dt} = \frac{-D_0 t}{[1+\delta(T-t)]}; t_1 \le t \le \mu \tag{2}$$

$$\frac{dI(t)}{dt} = \frac{-D_0 \mu}{[1+\delta(T-t)]}; \mu \le t \le T \tag{3}$$

with boundary conditions:

$I(t) = S$ at $t = 0$
$I(t) = O$ at $t = t_1$
$I(t) = R$ at $t = T$
and $I(t)$ is continuous at $t = \mu$. (4)

Using the above boundary conditions, solutions of the differential Equations (1) through (3) follow:

$$I(t) = [-D_0(\theta t - 1) + (\theta^2 S - D_0)e^{-\theta t}]/\theta^2 ;$$
$$0 \le t \le t_1 \tag{5}$$

$$= \frac{D_0}{\delta}[(t-t_1) + (1+\delta T).\ln\left|\frac{1+\delta(T-t)}{1+\delta(T-t_1)}\right|/\delta] ;$$
$$t_1 \le t \le \mu \tag{6}$$

$$= \frac{D_0\mu}{\delta}.\ln|1+\delta(T-t)| - R ;$$
$$\mu \le t \le T \tag{7}$$

To calculate **highest stock level** S as $I(t) = 0$ at $t = t_1$, hence from Equation (5), we have

$$I(t) = [-D_0(\theta t - 1) + (\theta^2 S - D_0)e^{-\theta t}]/\theta^2 = 0 ;$$

simplifying, we have:

$$\Rightarrow \theta^2 S = D_0[1 + (\theta t_1 - 1)e^{\theta t_1}$$
$$\Rightarrow S = D_0[1 + (\theta t_1 - 1)e^{\theta t_1}]/\theta^2 \tag{8}$$

Again, for **highest shortage level** R, since $I(t)$ is continuous at $t = \mu$, using Equations (6) and (7), we have:

$$\frac{D_0}{\delta}[(t-t_1) + (1+\delta T).\ln\left|\frac{1+\delta(T-t)}{1+\delta(T-t_1)}\right|/\delta]$$
$$= \frac{D_0\mu}{\delta}.\ln|1+\delta(T-t)| - R$$

After rearranging R can be simplified as:

$$R = \frac{D_0\mu}{\delta}.\ln|1+\delta(T-\mu)|$$
$$-\frac{D_0}{\delta}[(\mu-t_1) + (1+\delta T).\ln\left|\frac{1+\delta(T-t)}{1+\delta(T-t_1)}\right|/\delta] \tag{9}$$

The **number of deteriorated units** is given by:

$$N' = S - \int_0^{t_1} R(t)dt = (S - \frac{D_0 t_1^2}{2})$$

Now, the **total inventory holding cost** in the cycle is given by:

$$C_{hold} = C_1\int_0^{t_1} I(t)dt$$
$$= \frac{C_1}{\theta^2}\int_0^{t_1}[-D_0(\theta t - 1) + (\theta^2 S - D_0)e^{-\theta t}]dt ;$$

after simplification comes out as:

$$= C_1[D_0(t_1 - \frac{\theta t_1^2}{2}) + (\theta^2 S - D_0)(1 - e^{-\theta t_1})/\theta]/\theta^2 \tag{10}$$

Total shortage cost in the cycle is given by:

$$C_{short} = C_2\int_{t_1}^{T}(-I(t))dt$$
$$= C_2[\frac{D_0}{\delta}\int_{t_1}^{\mu}\left\{(t_1-t)-\frac{(1+\delta T)}{\delta}.\ln\left|(1+\delta T)\right|\right.$$
$$\left.+\frac{(1+\delta T)}{\delta}.\ln\left|1+\delta\,(T-t_1)\right|\right\}dt$$
$$+R(T-\mu)-\frac{D_0\mu}{\delta}\int_{\mu}^{T}\ln\left|1+\delta\ (T-t)\right|dt];$$

after simplification comes out as equation (11), shown in Box 1.

This expression of shortage factor is highly complex and inclusion of this factor makes the profit function highly complicated. But considering this present business scenario, we consider this factor to highlight the perplexing nature of end users' in purchasing.

The **net profit** for this (Case 1) production-inventory system is, therefore, given by:

NP = | *Sales Revenue* | - | *Purchase Cost* |
- | *Ordering Cost* | - | *Holding Cost* |
- | *Shortage Cost* | - | *Inspection Cost* |

$=[s(S + R - N') - C_3(S + R)$
$- C_4 - C_{\text{hold}} - C_{\text{short}} - C_5S]$

Therefore the average profit/time is given by:

$$\zeta_1(t_1,T) = NP/T$$
$$=[s(S + R - N') - C_3(S + R) - C_4 - C_{\text{hold}} - C_{\text{short}} - C_5S] / T \qquad (12)$$

Hence our problem in hand is to find the optimal values of t_1, T, S, R which maximizing the profit function $\zeta_1(t_1, T)$, where t_1, T are continuous variables.

Optimization Problem

Maximize $\zeta_1(t_1, T)$
Subject to $t_1 \leq \mu$ (13)

This model may be solved either with generalized reduced gradient (GRG) or any advanced methodologies like the computational intelligence method. Considering the complexities, we now try to develop MRCGA with elitism and also SA.

Case 2: $[t_1 \geq \mu]$

Proceeding in a similar way, the **net profit** for Case 2 is again given by:

Box 1. Equation 11, simplified

$$= C_2[D_0\{\frac{-(t_1-\mu)^2}{2}-\frac{(1+\delta T)}{\delta}[-\mu+t_1-(1+\delta\,(T-\mu)).\ln\left|1+\delta\,(T-\mu)\right|/\,\delta$$
$$+(1+\delta\,(T-t_1)).\ln\left|1+\delta\ (T-t_1)\right|/\delta\ ]+\frac{(1+\delta T)}{\delta}.\ln\left|1+\delta\,(T-t_1\right|(\mu-t_1)/\delta\,\}$$
$$+R(T-\mu)+D_0\mu\{T-\mu-(1+\delta\ \ (T-\mu)).\ln|\,1+\delta\,(T-\mu\,)\,/\,\delta\}\,/\,\delta\,] \qquad (11)$$

***NP* = | *Sales Revenue*| – | *Purchase cost*| – | *Ordering cost*| – | *Holding cost*| | *Shortage cost*| – | *Inspection cost*|**

$$=[s(S+R-N') - C_3(S+R) - C_4 - C_{\text{hold}} - C_{\text{short}} - C_5 S] \tag{14}$$

Profit function $\zeta_2(t_1, T)$ is, therefore, follows:

$$\zeta_2(t_1, T) = [s(S+R-N') - C_3(S+R) - C_4 - C_{\text{hold}} - C_{\text{short}} - C_5 S] / T \tag{15}$$

where the **total holding cost** during the cycle length is given by:

$$C_{hold} = C_1[\{\theta^2 S - D_0\}\{1-\exp(-\theta\mu)\}/\theta - D_0(\theta\mu^2 - 2\mu)/2\}/\theta^2 + D_0\mu\{(\exp(\theta(t_1-\mu)) - 1)/\theta + \mu - t_1\}/\theta \tag{16}$$

The **total Shortage cost** during the entire cycle is given by:

$$C_{short} = C_2 D_0 \mu[(T-t_1)\ln|1+\delta(T-t_1)| + (T-t_1)(1+\delta(T-t_1) * \ln|1+\delta(T-t_1)|/\delta]/\delta \tag{17}$$

Optimization Problem

Maximize $\zeta_2(t_1, T)$
Subject to $t_1 \geq \mu$ (18)

This model may be solved either GRG or any advanced methodologies like computational intelligence method. Considering the complexities, we now try to develop MRCGA with elitism and also SA.

COMPUTATIONAL INTELLIGENCE APPROACH

Computational intelligence is the study of adaptive mechanisms to enable or facilitate intelligent behavior in complex and changing environments. The first published definition by J. C. Bezdek (1989) about computational intelligence on manufacturing and production management states that, "(strictly) computational systems depend on numerical data supplied by manufacturing sensors and do not rely upon 'knowledge'". Later, in 1994, Bezdek differentiated CI as "low-level computation in the style of the mind", with AI which states that "mid-level computation in the style of the mind."

Recently, Poole et al. (1996) identified computational intelligence in a broad way:

Computational Intelligence is the study of the design of intelligent agents . . . An intelligent agent is a system that acts intelligently: what it does is appropriate for its circumstances and its goals, it is flexible to changing environments and changing goals, it learns from experience, and it makes appropriate choices given perceptual limitations and finite computation.

The most elaborate concept regarding computational intelligence is given by Eberhart (1998):

CI is defined as a methodology involving computing that exhibits an ability to learn and / or deal with new situations such that the system is perceived to possess one or more attributes of reason, such as generalization, discovering, association and abstraction. The output of a computationally intelligent system often includes predictions and / or decisions. Put another way, CI comprises practical adaptation concepts, paradigms, algorithms and implementations that enables or facilitate appropriate actions (intelligent behavior) in complex and changing environments.

There are also other fields whose goal is to build machines that act intelligently. Two of these fields are control engineering and operations research.

These areas starts from different points than CI, namely in the use of continuous mathematics. As building real agents involves continuous control and CI type reasoning, these disciplines should be seen as symbiotic with CI. In modern complex industrial scenario, there exists a very real need to assist operators in their decision making, particularly in abnormal situations in which they often are bombarded with conflicting goals. The advent of computational intelligence and unconventional control free operators of many of the tedious and complex chores of monitoring and controlling a production unit assuring fast and consistent support in their decision making. Thus science of CI could be described as "synthetic psychology," "experimental philosophy," or "computational epistemology; epistemology is the study of knowledge." Now proceeding with computational intelligence approach, we develop an algorithm for determining the optimal value of t_1 and T with the average profit of the proposed production-inventory system by a real coded genetic algorithm for two continuous variables. Here we have developed both modified real coded genetic algorithm (considering ranking selection and elitism property) and simulated annealing.

Genetic Algorithms

A genetic algorithm (GA) is a parallel and evolutionary search algorithm based on the Darwinian Theory (Goldberg, 1989). It is used to search large, nonlinear solution space where expert knowledge is lacking or difficult to encode. Moreover, it requires no gradient information; it evolves from one population to another and produces multiple optima rather than a single local one. In most cases, GA is conducting its search so that it may optimize some known fitness function. The way in which the function itself is "known" is of limited theoretical, but much practical importance. At the extreme, the fitness of a proposed solution could be tested in the physical world. Once, an initial population has been formed, "selection," "crossover," and "mutation" operations are repeatedly performed until the fitness number of evolving population converges to an optimal fitness value. Alternately, the GA may run for a user defined number of iterations to evolve a strategy of building a good solution instead of finding the solution directly.

Why Genetic Algorithms in Production-Inventory Control Model?

The operations of a genetic algorithm for controlling inventories are conceptually very simple and straight worthy. The developer provides a cost algorithm (fitness function, possibly a simulation, etc.) based on standard inventory models and the GA develops a series of candidate policies $p_1, p_2, \ldots . p_n$ which are tested for fitness by the cost algorithm and the results returned to the GA. The "best" of these is passed for implementation, containing a set of fixed policies and resource assignments.

General characteristic of the production-inventory optimization process is that once "fairly" good solutions have been formed, their features will be carried forward into "better" solution and lead ultimately to optimal solution. On the other hand, it is the nature of GA that new solution are formed from the features of known good solution. Again, comparing with other optimization methods, GAs are suitable for traversing large search space since they can do this relatively rapidly and as the mutation operator diverts the method away from "local" minima which will tend to become more common as the search space increases in size. Being suitable for large spaces is a useful advantage when dealing with schedules of increasing size since the solution space will grow very rapidly, especially when this is compounded by features. It is important that these large search spaces are traversed as rapidly as possible to enable the practical and useful implementation of inventory optimization. For these, GAs are particularly attractive for handling the complex problems in production-inventory control by the business intelligence group/management scientists.

Development of Genetic Algorithms

The stepwise procedure of genetic algorithms is shown as follows:

- **Step 1:** Initialize the parameters of genetic algorithm, bounds of variables and different parameters of the proposed inventory system.
- **Step 2:** $t=0$ (t represents the number of current generation)
- **Step 3:** Initialize $P(t)$; $P(t)$ represents the population at t - th generation
- **Step 4:** Evaluate $P(t)$
- **Step 5:** Find optimal result from $P(t)$
- **Step 6:** $t = t + 1$
- **Step 7:** If (t > maximum generation number) go to Step 14
- **Step 8:** Select $P(t)$ from $P(t-1)$ by standard selection process
- **Step 9:** Alter $P(t)$ by crossover and mutation operation
- **Step 10:** Evaluate $P(t)$
- **Step 11:** Find optimal result from $P(t)$
- **Step 12:** Compare optimal results of $P(t)$ and $P(t-1)$ and store the best one
- **Step 13:** Go to Step 6
- **Step 14:** Print optimal result
- **Step 15:** Stop

Implementation of the above algorithms for solving our constraint maximization problem requires the following basic components:

- Parameterization
- Chromosome representation
- Initialization
- Evaluation function
- Selection process
- Genetic operators
- Elitism

Parameters of Genetic Algorithm

Genetic algorithm depends on different parameters like population size (POPSIZE), probability of crossover (PCROS), probability of mutation (PMUTE), and maximum number of generation (MAXGEN). Population size in genetic algorithm is problem dependent. If the population is too large, there arise some difficulties in storing of the data for large population size. However, if the size of it is small, the crossover operation can not be implemented accordingly. Again, according to genetics, probability of crossover is always greater than mutation. Generally, the probabilities of crossover and mutation are taken as 0.85 to 0.95 and 0.05 to 0.15, respectively. In our present inventory analysis, we have taken the values of these parameters as follows:

$POPSIZE = 200$, $PCROS = 0.91$, $PMUTE = 0.1$, $MAXGEN = 200$.

Chromosome Representation

The main problem in applying genetic algorithm is to design an appropriate chromosome representation of solutions of the problem with genetic operators. Traditional binary vectors used to represent the chromosomes are not effective in many highly physical nonlinear problems. As our proposed model is highly nonlinear; to overcome this difficulty, a real number representation is implemented here. A real row matrix $V_j = [V_{j1}, V_{j2}]$ is used to represent a chromosome where V_{j1} and V_{j2} represent the variables t_1, the no-shortage period, and T, the cycle length.

Initialization

To initialize a population in genetic algorithm, generally POPSIZE number of chromosomes $V_1, V_2 \ldots, V_{popsize}$ are generated randomly. However, it is very difficult for complex optimization problems to produce feasible chromosome explicitly. Generally, for each chromosome V_i, every gene is randomly generated within the desired domain in such a way that it should be feasible in nature.

Evaluation Function

Evaluation function plays exactly the same role in GA as the environment plays in natural evolution. Generally, evaluation function $EVAL(X)$ for the chromosome X is taken equivalent for the objective function $f(X)$. After getting a population, our objective is to search a chromosome that gives better value of the objective function. For this, we have to calculate the fitness for each chromosome. The value of the objective function due to the chromosome V_i is taken as the fitness value of V_i and it is denoted by $eval(V_i)$.

Selection

Every kind of selection must prefer individuals within a good fitness to such one with a worse. Weak individuals ought to get a little chance to pass their alleles to the next generation yet. This is very important for the spread out of the individuals over the parameter space; otherwise the evolution will make a premature convergence to local optimum. Thus the purpose of selection is to emphasize the better individuals in the population for recombination in hopes that their offspring will in turn have even higher fitness. Selection has to be balanced with variation from crossover and mutation; too strong selection means that suboptimal highly fit individuals will take over the population by reducing the diversity needed for further change and progress and too weak selection will result in too slow evolution.

Several standard techniques are used for selection, such as deterministic sampling method, roulette wheel selection (also called stochastic sampling with replacement and without selection), (N, μ) selection, stochastic tournament (Wetzel ranking) or ranking selection method, and so forth. In this ranking method, selection probabilities are calculated normally and successive pairs of individuals are drawn by using roulette wheel selection. Here, we have modified our work by using ranking selection over roulette wheel where fitness is taken as the value of the objective function.

The algorithm we use for this probabilistic selection process is:

- **Step 1:** Compute all the fitness $f_i \ \forall i = 1, 2, \ldots, n$.
- **Step 2:** Sort all $f_1, f_2, \ldots, f_n$ in descending order for maximization problem and identify all the sorted $f_i \ \forall i = 1, 2, \ldots, n$ by marking numbers $(1, 2, \ldots, n)$.
- **Step 3:** Generate a random real number r_1 in [0,1].
- **Step 4:** Calculate the probability $p_i \ \forall i = 1, 2, \ldots, n$ for each chromosome V_i by using the formula $p_i = r_1(1 - r_1)^{i-1}$
- **Step 5:** Compute the cumulative probability P_i for each chromosome V_i by using the formula:

$$P_i = \sum_{j=1}^{i} p_j$$

- **Step 6:** Generate a random real number r_2 in [0,1].
- **Step 7:** Obtain the minimal k such that $P_k \succ r_2$
- **Step 8:** Select the k-th individual.
- **Step 9:** Repeat Steps 6 through 8 until the number of selected individuals coincide with population size.

Crossover Operation

The exploration and exploitation of the solution space is made possible by exchanging genetic information of the current chromosomes. Crossover operator operates on two parent solutions (chromosomes) at a time and generates offspring by combining both parent solutions features. For this operation, expected N (the integral value of *PCROS*POPSIZE*) number of solutions will take part. Hence, in order to perform the crossover operation, *PCROS*POPSIZE* number of chromosomes are to be selected. For this purpose, we adopt the random stochastic sampling scheme (without replacement). After selection of chromosomes, the crossover operation is applied. In this chapter, we use nonlinear whole arithmetical crossover.

Different Steps of Crossover Operation

- **Step 1:** Assign N as the integral value of *PCROSS* POPSIZE*
- **Step 2:** Generate a random real number r_3 in [0,1].
- **Step 3:** Select the chromosome V_k and V_i robustly from the population for crossover operation if $r_3 \prec PCROS$
- **Step 4:** Generate a proper fraction λ by the formula:

  $$\lambda = {p_{\max}} \Big/ {(p_{\max} + p_{\min})}, \text{ where}$$

  $p_{\max} = \max[p_j \,\forall j = 1, 2, \ldots, POPSIZE]$ and
  $p_{\min} = \min\,[p_j \,\forall j = 1, 2, \ldots, POPSIZE]$

- **Step 5:** Produce two offsprings $V_k^{/}\ \& V_1^{/}$ by:

  $$V_k^{/} = \lambda V_k + (1-\lambda)V_1$$
  $$V_1^{/} = \lambda V_1 + (1-\lambda)V_k$$

- **Step 10:** Repeat Step 2 and Step 5 for $N/2$ times.

Mutation Operation

Mutation introduces random variations in the population and is used to prevent the search process from converging to local optima rapidly. Mainly, this operation is responsible for fine tuning of the system and is applied to a single chromosome. After introducing genetic diversity and turning the population gently into a slightly better converge way, mutation operation performed with a low probability by defeating the order building as generated via selection and crossover. Here, we shall use nonuniform mutation whose action is dependent on the age of the population. If the element V_{ik} of chromosome V_i is selected for mutation and domain of V_{ik} is $[l_{ik}, u_{ik}]$, then the reduced value of V_{ik} is given by:

$$V_{ik}^{/} = V_{ik} + \Delta(t, u_{ik} - V_{ik}); \text{ if a random digit is 0,}$$
$$= V_{ik} - \Delta(t, V_{ik} - l_{ik}); \text{ if a random digit is 1.}$$

where k ∈ { 1,2 } and the function Δ(t, y) returns a value in the range [0, y] such that the value of Δ(t, y) being close to 0 as t increases. This property causes this operator to search the space uniformly initially (when t is small) and vary locally at later stages.

Here, we have taken:

$$\Delta(t, y) = y[1 - r^{(1-t/T)^b}]$$

where, r is a random number from [0, 1], T = MAXGEN, t represents the current generation and b (which is called the nonuniform mutation parameter) is constant.

The algorithm of mutation operation is as follows:

- **Step 1:** $i = 1$
- **Step 2:** Generate a random number r from [0, 1].
- **Step 3:** If $r \prec PMUTE$, then select the chromosome V_i and go to Step 5.

- **Step 4:** $i = i + 1$
- **Step 5:** Select a particular gene V_{ik} of selected chromosome V_i
- **Step 6:** Create new gene corresponding to the selected gene V_{ik} by mutation operation.
- **Step 7:** Repeat the Steps 1 through 6 for *PMUTE*POPSIZE* times.

Elitism

As modified real coded GA technique is a stochastic optimization technique, sometimes the best chromosome may be lost when a new population is created by crossover and mutation. In this sequel, to increase the performance of MRCGA rapidly, one more highly fitted chromosomes (individuals) are considered in the new population to prevent the loss of best-found solution. This property is called elitism and we have taken this factor in this model, over our previous work.

Termination

If number of iteration is less than or equal to MAXGEN, then the process is going on; otherwise it terminates.

Thus the performance of genetic algorithms differs from traditional optimization algorithms in four important respects:

1. Their performance is based on using encoding of the control variables, rather than the variables themselves.
2. They search from one population of solutions to another, rather than from individual to individual.
3. They use only objective function information, not derivatives.
4. They use probabilistic, not deterministic transition rules.

Of course, genetic algorithms share the last two attributes with simulated annealing.

Simulated Annealing

Simulated annealing (SA) is a related global optimization technique that traverses the search space by testing random mutations on an individual solution (Ingber, 1989). A mutation with increasing fitness is always accepted. A mutation that lowers fitness is accepted probabilistically based on the difference in fitness and a decreasing temperature parameter. SA has a "cooling mechanism" (referred to as the 'temperature') which initially allows moves to less fit solutions. The effect of cooling on the simulation of annealing is that the probability of following an unfavorable move is reduced. This initially allows the search to move away from local optima in which the search might be trapped. When minimizing, the objective function is usually referred to as a cost function; but for maximizing, it is usually referred to as fitness function. In SA parlance, one pinpoints of seeking the lowest energy instead of the maximum fitness. SA can also be used within a standard GA by starting with a relatively high rate of mutation and decreasing it over time along a given schedule (Otten & Van Ginneken, 1989).

Simulated annealing has been applied to numerous problems in OR/MS as cell formation, as explored by Adil, Rajamni, and Strong (1997), Lot-sizing Kuik and Solaman (1999), and so forth. However, none of these papers discuss the problem of production-inventory with ramp type of demand nature. It is well-known that simulated annealing works well for certain type of minimization problems, but it is not suitable for all type. Simulated annealing algorithm used in this chapter is very robust in the sense of cooling schedule. It is interesting to investigate whether SA can be applied efficiently and effectively to manufacturing and production problems. Active research work in this area is necessary.

Stepwise Procedure of Simulated Annealing

From an initial solution, SA repeatedly generates a neighbor of the current solution and transfers to it according to some strategy with the aim of improving the objective function value (as we are trying to maximize the objective function). During this process, SA has the possibility to visit worse neighbors in order to escape from local optima. Specifically, a parameter, called temperature T is used to control the possibility of moving to a worse neighbor solution. The algorithm, starting from a high temperature, repeatedly decreases the temperature in a strategic manner (called cooling schedule) until the temperature is low enough or some other stooping criteria is satisfied. Algorithm accepts all "good" moves and some of the "bad" moves according to the Metropolis probability, defined by $\exp(-\delta/T)$ where δ is the decrease in the objective function value. Further, for running simulated annealing, one of the prime tasks is to define a suitable neighborhood function so that this generation function guarantees every feasible solution within the search space. We then fix an initial solution and there are also many control parameters that need to be set-up (e.g., initial temperature, cooling ratio, and a stopping criterion).

In our present optimization problem, the proposed algorithm works as follows:

Define an objective function *f*, and a set of heuristics H.
Define a cooling schedule: starting temperature $t_s \succ 0$, a temperature reduction function ϕ, a number of iterations for each temperature *ntemp*, and termination criterion.
Select an initial solution s_0;
Repeat
  Randomly select a heuristic $h \in H$;
  iteration_count = 0;
  Repeat
    iteration_count ++ ;
    applying h to s_0, get a new solution s_1;
    $\delta = f(s_1) - f(s_0)$
    if $(\delta \geq 0)$ then $s_0 = s_1$;
    else
      generate a random x uniformly in the range (0, 1);
      if $x \prec \exp(\delta/T)$ then $s_0 = s_1$;
  Until iteration_count = *ntemp*;
  Set $t = \phi(t)$;

Until the stopping criteria = true

Performance of simulated annealing in performing an optimization problem is based on the following criterion:

Initial Solution: Generated uniformly at random within a user defined range.

Initial Temperature: This was generated by using the following procedure: (a) generate specific number of solution at random; (b) find the solutions with the two optimized cost; and (c) initial temperature was set at twice the difference between the two optimized costs. In each run, initial temperature is allowed to be adjusted by the user.

Cooling Schedule: In the developmental phase of our proposed algorithm, we use different types of cooling schedule to find the global optimum. Popular use of this type is as $T = \zeta * T$ where ζ is a parameter that may be adjusted by the user. Optimal setting of ζ is highly problem-dependent robust in nature. The common practice to search a near-optimal ζ is to start with a relatively smaller value and then gradually turns it to a large value close to 1.0 (say 0.98 / 0.99). This trial-and-error process is very tedious. The second type of cooling schedule which one may use is:

$$T = T/(1+\theta T)$$

where, θ is the parameter adjusted by the user and is strictly problem dependent. Again, there is no universally optimal θ which is best for all problems; it is strictly problem dependent. In our analysis, we go on by using first type of cooling schedule.

Stopping Criterion: The inner loop criterion was fixed by user by number of iterations for which the temperature T was kept the same. The outer loop criterion was determined by the maximum number of iterations set by user and also the lowest temperature. This implies that algorithm of simulated annealing stops when its temperature is below this lowest value.

NUMERICAL ILLUSTRATION

To illustrate the developed model, an example has been considered. As there is no real-world data available due to commercial confidentially and neither is there any benchmark data available from the literature, the values considered here are feasible and may be suitable to any real life decision making problem. All algorithms were coded in Microsoft Visual C++ version 6.0 and all experiments were run on a PC Pentium IV 1.8 GHZ with 256 RAM running Microsoft Windows 2000 professional Version 5. All algorithms started from a solution produced by a greedy heuristic and allowed fairly computation time for a better comparison.

Let, $C_1 = 1$, $C_2 = 10$, $C_3 = 10$, $C_4 = 100$, $C_5 = 0.5$, $D_0 = 150$, $s = 15$, $\theta = 0.08$, $\mu = 2$ in appropriate units.

Solving the nonlinear optimization problems by MRCGA, SA and GRG methods, the optimal values of t_1, T, S, R and the average profit for different values of the backlogging rate parameter δ are obtained. The corresponding results are shown in Tables 1 through 6, respectively.

From the tabular value, (Tables 1 through 3) it is seen that Case 1 gives the better result than

Table 1. Case 1: [$t_1 \le \mu$]. *Optimum results for different values of* δ *by using computational intelligence methodology (via elitist MRCGA). Status of Optimizer: Global Optimum*

δ	**0.5**	**1.0**	**10**	**20**	**50**
t_1	1.8452	1.8414	1.9013	1.9269	1.9427
T	2.2027	2.1835	2.0930	2.0615	2.0249
S	281.9341	280.7013	300.2468	308.8260	312.1234
R	97.1094	86.8151	31.8319	19.4629	12.8648
Profit	**381.8904**	**373.6248**	**322.2299**	**300.6084**	**296.4526**

Table 2. Case 1: [$t_1 \le \mu$]. *Optimum results for different values of* δ *by using standard optimization methodology (via generalized reduced gradient). Status of Optimizer: Local Optimum*

δ	0.5	1.0	10	20	50
t_1	1.9822	1.9697	1.9358	1.9411	1.9723
T	2.3341	2.3051	2.1265	2.0724	2.0595
S	327.7595	323.4480	311.8191	313.6237	317.4657
R	97.2337	86.7106	31.8964	19.2406	15.4526
Profit	**377.9514**	**370.4410**	**321.8222**	**307.8612**	**295.1245**

Table 3. Case 1: [$t_1 \leq \mu$]. Optimum results for different values of δ by using computational intelligence methodology (via simulated annealing). Status of Optimizer: Global Optimum

δ	0.5	1.0	10	20	50
t_1	1.8359	1.8399	1.8527	1.8649	1.8925
T	2.2210	2.1945	2.1142	2.0692	2.0242
S	284.5011	281.4529	289.4012	292.0641	301.1242
R	95.1120	91.1024	54.5926	49.2674	39.7526
Profit	**389.4211**	**382.5979**	**359.4029**	**311.1111**	**301.9112**

Table 4. Case 2: [$t_1 \geq \mu$]. Optimum results for different values of δ by using computational intelligence methodology (via elitist MRCGA). Status of Optimizer: Global Optimum

δ	0.5	1.0	10	20	50
t_1	2.0001	2.0000	2.0002	2.0003	2.0006
T	2.2997	2.3001	2.3020	2.2003	2.2000
S	334.0094	334.0089	334.0011	333.9011	333.9112
R	83.7701	76.5597	39.4521	27.3512	12.0261
Profit	**377.7951**	**365.1020**	**318.1122**	**297.1266**	**287.4500**

Table 5. Case 2: [$t_1 \geq \mu$]. Optimum results for different values of δ by using standard optimization methodology (via generalized reduced gradient). Status of Optimizer: Local Optimum

δ	0.5	1.0	10	20	50
t_1	2.0000	2.0000	2.0000	2.0000	2.0000
T	2.3501	2.3017	2.2015	2.1211	2.0720
S	333.9947	333.9947	333.9947	333.9947	333.9947
R	96.4501	85.7111	32.9218	22.5649	7.4211
Profit	**376.4503**	**368.7111**	**320.0001**	**305.4692**	**292.4443**

Table 6. Case 2: [$t_1 \geq \mu$]. Optimum results for different values of δ by using computational intelligence methodology (via simulated annealing). Status of Optimizer: Global Optimum

δ	0.5	1.0	10	20	50
t_1	2.0000	2.0001	2.0001	2.0002	2.0000
T	2.2727	2.2820	2.2900	2.2801	2.3001
S	335.0061	334.0101	333.9021	333.9921	334.0087
R	82.4200	79.5060	56.5941	293520	12.1212
Profit	**376.4010**	**369.2020**	**325.2511**	**322.0047**	**301.1021**

Case 2. Also, MRCGA and SA give the better result than GRG method for Case 1 whereas it is reverse for Case 2. But, for MRCGA and SA method, all results are "global" optimum and for GRG method it is "local" optima only.

SENSITIVITY ANALYSIS

For the above numerical example mentioned earlier, sensitivity analysis is performed to study the effect of under or over estimation of various parameters on stock-in period, cycle length, maximum inventory level, maximum shortage level, and the average profit. This analysis is performed by changing (increasing and decreasing) the system parameters from -25 % to 25 %, taking one parameter at a time and keeping the others at their original values by applying modified real coding genetic algorithm coding. The results of this analysis are shown in Table 7 and Table 8.

CONCLUSION

The challenge in inventory control is not only to pare inventory to the bone to reduce costs or to have plenty around to fulfill all demand, but to have the right amount available to achieve the competitive priorities for the business most efficiently. In this chapter, we have extended and modified the existing inventory model for deteriorating items with ramp type demand. This is the first time the model is solved simultaneously

Table 7. Effect of % changes in D_0 ,θ , μ and s on t_1, T, S, R, and average profit

Para-meters	% changes in parameter	% changes				
		t_1	T	S	R	Average Profit
D_0	-25	0.31	0.45	-27.6	-25.01	-26.77
	-10	-0.01	-0.11	-10.01	-10.47	-11.19
	10	-0.55	-0.31	8.72	10.77	11.23
	25	-1.33	-1.34	17.55	19.54	27.44
θ	-25	6.77	5.11	12.01	-3.95	9.62
	-10	2.01	1.21	3.21	-2.09	3.69
	10	-3.36	-2.26	-6.02	1.65	-3.49
	25	-6.27	-4.12	-11.01	5.79	-8.02
μ	-25	-14.01	-12.59	-30.01	-25.11	-8.12
	-10	-13.29	-11.36	-25.82	-12.44	-4.31
	10	4.47	4.72	9.62	12.92	2.13
	25	11.01	12.77	22.11	28.71	3.01
s	-25	-12.59	-11.97	-25.01	-6.38	-118.05
	-10	-10.06	-9.16	-19.92	-10.65	-59.53
	10	8.39	8.81	18.46	11.84	65.52
	25	9.01	9.97	17.49	21.01	134.60

Table 8. Effect of % changes in C_1, C_2, C_3, C_4 and C_5 on t_1, T, S, R, and average profit

Para-Meters	% changes in parameters	% changes				
		t_1	*T*	*S*	*R*	Average Profit
C_1	-25	5.47	3.81	11.84	-3.08	8.78
	-10	2.78	1.89	5.93	-1.66	4.05
	10	-3.02	-2.23	-6.24	0.38	-3.78
	25	-5.54	-3.67	-11.27	3.55	-8.01
C_2	-25	-1.99	1.69	-4.12	18.51	4.79
	-10	-1.33	0.62	-2.77	9.42	2.15
	10	0.13	-1.03	0.27	-6.58	-1.84
	25	0.49	-1.99	0.98	-13.89	-3.99
C_3	-25	-4.89	-5.59	-9.86	-11.55	-35.86
	-10	-2.78	-3.14	-5.74	-6.00	-17.58
	10	1.91	2.66	4.14	6.76	17.94
	25	3.94	4.21	8.99	10.39	36.98
C_4	-25	-1.56	-1.37	-2.81	-1.97	2.49
	-10	-0.39	-0.60	-0.82	-1.76	1.19
	10	0.01	0.00	0.02	0.00	-1.21
	25	0.02	0.01	0.00	0.00	-2.39
C_5	-25	-0.01	-0.15	-0.02	-0.53	1.77
	-10	-0.01	-0.15	-0.02	-0.53	0.67
	10	-0.55	-0.33	-1.16	0.75	-0.66
	25	-0.79	-0.58	-1.55	1.17	-1.35

by GRG, MRCGA, and SA method. Successful application of computational intelligence methodology in this type of production-inventory management model is very complex in nature; attempts have been made to modify the exiting inventory model by incorporating different new approaches like ranking selection and elitism over conventional GA. As the exact solution cannot be translated into practical term because of some unforeseen constraints, the alternate solution by GA may provide significant guideline for actual implementation.

Again, Metaheuristic algorithms are such algorithms, which, in order to escape from local optima, drive some basic heuristic, both a constructive starting point from a null solution and adding elements to build a good one. The metaheuristic part allows low-level heuristic to achieve solutions better than those it could have achieved alone, even if iterated by randomizing the set of local neighbor solutions to consider at local search (as in the case of simulated annealing).

It has been demonstrated that computational intelligence methodology gives better and global optimum values than the conventional nonlinear optimization method. Again modification of real coded GA gives even better results than earlier. This present production-inventory problem may further be developed for multiple items, and finite time horizon considering discount policies in different types of environments like fuzzy and fuzzy-stochastic.

FUTURE RESEARCH DIRECTIONS

This chapter is particularly directed to business intelligence providers (analysts as well as managers) with the advanced/recent use of mathematical modeling as an aid to computational intelligence methodology in manufacturing and production management. Again, this problem may further be developed for multiple items, and finite time horizon considering discount policies in different environments like fuzzy and fuzzy-stochastic environments. Further, advantage of GA and SA are summarized in the line of robustness, problem independence, high parallel working environment, and successful parametric optimization domains, but lack in final tuning capabilities and time complexities. For this, we may explore the possibility of using advanced computational intelligence methodologies like interactive genetic algorithm or memetic algorithms (MA), also called hybrid genetic algorithm among others, which is a relatively new evolutionary method where local search is applied during the evolutionary cycle. The idea of memetic algorithm comes from memes, which unlike genes, can adapt themselves. In some manufacturing and production management problems, memetic GA may be more efficient than traditional evolutionary algorithms. Further research is envisaged.

Again, ant colony optimization (ACO) is a novel and very promising research field lying at the crossing between computational intelligence and operational research. This (also known as ant algorithms) is a paradigm for performing metaheuristic algorithms for combinatorial optimization (Dorigo, Caro, & Gamberdella, 1999). Ant algorithms are biologically inspired from the behavior of colonies of real ants, and in particular how they forage for food. The essential trait of ACO algorithms is the combination of a priori information about the structure of a promising solution with a posteriori information about the structure of previously obtained good solution. Development of ACO applications in optimization area refers to the well known symmetric as well as asymmetric traveling salesman problem (Dorigo et al., 1997a), quadratic assignment problem (Stutzle & Hoos, 2000), jobshop scheduling (Fang et al., 1994), vehicle routing with time window constraints (Gamberdella et al., 1999), and so forth. Again, this advanced stochastic optimization method is diversified in different complex fields like sequential ordering (Gamberdella & Dorigo, 2000), graph coloring, assembly line balancing (Bautista & Pereira, 2002), image processing and network routing (Di Caro & Dorigo, 1998c), and many more. But, until now, no one can apply this methodology in case of production-inventory model. Active research directions for characterizing and understanding and also effective parallelization of ACO algorithms for handling present type of various complex problems in production-inventory are envisaged.

REFERENCES

Adil, G. K., Rajamni, D., & Strong, D. (1997). Assignment allocation and simulated annealing algorithms for cell formulation. *IIE Transactions*, *29*, 53-67.

Aggarwal, S. P. (1978). A note on an order: Level model for a system with constant rate of deterioration. *Opsearch*, 15, 184 -187.

Bhunia, A. K., & Maiti, M. (2001). An inventory model of deteriorating items with lot: Size dependent replenishment cost and a linear trend in demand. *Applied Mathematical Modeling*, *23*, 301-308.

Cavicchio, D. J. (1972). Reproductive adaptive plans. In *Proceedings of the ACM* (pp. 1-11).

Davis, L. (Ed.). (1991). *Handbook of genetic algorithms*. New York: Van Nostrand Reinhold.

Davis, L., & Coombs, S. (1987). Genetic algorithms and communication link spread design: Theoretical consideration, genetic algorithms and their applications. In *Proceeding of the Second International Conference on Genetic Algorithms* (pp. 252-256).

Deb, K. (2001). *Multi-objective optimization using evolutionary algorithms*. Chichester, UK: Wiley.

Di Caro, G. & Dorigo, M. (1998). AntNet: Distributed stigmergetic control for communication networks. *Journal of Artificial Intelligence Research, 9*, 317-365.

Donaldson, W. A. (1977). Inventory replenishment policy for a linear trend in demand: An analytical solution. *Operational Research Quarterly, 28*, 663-670.

Dorigo, M., Stutzle, T. & Di Caro, G. (Eds.). (2000). Special Issue on Ant Algorithms. *Future Generation Computer Systems*, 16, 851-956.

Dorigo, M. & Gambardella, L.M. Ant colonies for traveling salesman problem. *Bio Systems, 43*(2), 73-81.

Etter, D. M., Hicks, M. J., & Cho, K. H. (1982). Recursive adaptive filter design using an adaptive genetic algorithm. *Proceeding of IEEE International conference on Acoustics, Speech and Signal Processing*, *2* (pp. 635-638).

Forrest, S. (Ed.). (1993). Genetic algorithms. In *Proceeding of Fifth International Conference on Genetic Algorithms*, California. Morgan Kaufman.

Ghare, P. M., &Schrader, G. P. (1963). A model for exponentially decaying inventory. *Journal of Industrial Engineering*, *14*, 238-243.

Giri, B. C., Goswami, A., & Chaudhuri, K. S. (1996). An EOQ model for deteriorating items with time varying demand and costs. *Journal of the Operational Research Society, 47*, 1398-1405.

Goel, V. P., & Aggarwal, S.P. (1981). Order level inventory system with power demand pattern for deteriorating items. In *Proceeding of the All India Seminar on Operational Research and Decision Making*, University of Delhi, (pp. 19 -34).

Goldberg, D. E. (1989). *Genetic algorithms: Search, optimization and machine learning.* Boston: Addison Wesley.

Goyal, S. K. & Aggarwal, S. P. (1981). Order level inventory system with power demand pattern for deteriorating items. In *Proceedings of All India Seminar on OR & Decision Making* (pp. 19-34).

Holland, J. H. (1975). *Adaption in natural and artificial systems*. Ann Arbor: The University of Michigan Press.

Ingber, L. (1989). Very fast simulated re-annealing. *Mathematical and Computer Modeling, 12*(8), 967-973.

Ingber, L., & Rosen, B.E. (1992). Genetic algorithms and very fast simulated re-annealing. *Mathematical and Computer Modeling, 16*(11), 87-100.

Koner, A. (2005). *Computational intelligence*. Springer-Verlag.

Kuik, R., & Saloman, M. (1999). Multi-level lot sizing problems: Evaluation of a simulation annealing heuristics. *European Journal of Operational Research*, *45*, 25-37.

Kun-Shan, W. U. (2000). An EOQ inventory model for items with Weibull distribution deterioration, ramp type demand rate and partial backlogging. *Production Planning & Control*, *12*(8), 787-793.

Lundy, M., & Mees, A. (1986). Convergence of the annealing algorithm. *Mathematical Programming, 34*, 111-124.

Michalewicz, Z. (1992). *Genetic algorithms + data structure = evolution programs.* Berlin: Springer Verlag.

Minga, A. K. (1986). *Genetic algorithms in aerospace design*. Paper presented at the AIAA Southeastern Regional Student Conference, Huntsville, Alabama.

Mitchell, M. (2004). *An introduction to genetic algorithms.* India: Prentice- Hall.

Otten, R. H. J. M., & Van Ginneken, L. P. P. (1989). *The annealing algorithm*. Kluwer Academic Publishers.

Papadimitriou, C., & Steiglitz, K. (1982). *Combinatorial optimization: Algorithms and complexity*. Upper Saddle River, NJ: Prentice Hall.

Roy, S., Bhunia, A.K., & Mukhopadhyay, S. (2005). A genetic algorithmic approach on a deterministic inventory model for deteriorating items with shortages. *Information Systems and Operational Research (INFOR), 43*(3), 271-282.

Sakawa, M. (2002). *Genetic algorithms and fuzzy multi-objective optimization*. Kluwer Academic Publishers.

Sarkar, R., & Lutfar, R. K. (1999). An optimal batch size for a production system operating under periodic delivery policy. *Computers and Industrial Engineering, 37,* 711-730.

Sarkar, R., & Newton, C.(2002). A genetic algorithm for solving economic lot size scheduling problem. *Computers and Industrial Engineering, 42*, 189-198.

Schaffer, J. D. (1985b). Multiple objective optimizations with vector evaluated genetic algorithms. In *Proceeding of the International Conference on Genetic Algorithm and their Applications* (pp. 93-100).

Schwefel, H. P. (1981). *Numerical optimization of computer models.* Chichester, UK: Wiley.

Shah, Y. K., & Jaiswal, M. C. (1997). An order level inventory model for a system with constant rate of deterioration. *Opsearch, 14,* 174-184.

Shi, Y. & Eberhart, R. (1998). A modified particle swarm optimizer. In *Proceedings of the 1998 IEEE INternational Conference on Evolutionary Computing* (pp. 69-73). Retrieved from http://ieeexplore.ieee.org/

Smith, T., & De Jong, K. A. (1981). Genetic algorithms applied to the calibration of information driven models of US migration patterns. In *Proceedings of the 12th Annual Pittsburgh Conference on Modeling and Simulation* (pp. 955-959).

ADDITIONAL READING

Adina, D. & Pham, D. T. (Eds.). (2007). *Computational Intelligence for Engineering and Manufacturing.* New York: Springer.

Bhunia, A. K., & Maiti, M. (1997). Deterministic inventory models for variable production. *Journal of Operations Research, 48,* 221-224.

Bhunia, A.K., & Maiti, M. (1997) An inventory model for decaying items with selling price, frequency of advertisement and linearly time-dependent demand with shortages. *Journal of the Indian Association for Productivity, Quality & Reliability, 22,* 41-49.

Bonabeau, E., Dorigo, M., & Theraulaz, G. (1999). *Swarm intelligence: From natural to artificial systems.* New York: Oxford University Press.

Cantu-Paz, E. (2000). *Efficient and accurate parallel genetic algorithms.* Boston: Kluwer Academic Publishers.

Chambers, L. (1995). *Practical handbook of genetic algorithms.* New Frontiers: CRC Press Publication.

Chem Zhengxin. (2000). *Computational intelligence for decision support.* Boca Raton, FL: CRC Press.

Coley, D. A. (1999). *An introduction to genetic algorithms for scientists and engineers.* World Scientific Publishing Co. Pvt. Ltd.

Corne, D., Dorigo, M., & Glover, F. (Eds.). (1999). *New ideas in optimization.* London: McGraw Hill.

Covert, R. T., & Philip, G. C. (1973). An EOQ model for items with Weibull distribution deterioration. *A.I.T.E. Transactions, 5,* 323-326.

Deb, K. (1995). *Optimization for Engineering Design Algorithms and Examples.* New Delhi: Prentice Hall.

Deb, K. (2001). *Multi-objective optimization using evolutionary algorithms.* New York: John Wiley & Sons.

Dorigo, M., Di Caro, G., & Gamberdella, L. M. (1999). Ant algorithms for discrete optimization. *Artificial Life, 5*(2), 137-172.

Fang, H. L., Ross, P., & Corne, D. (1994). A promising hybrid genetic algorithms/heuristic approach for open shop schedule problems. In A. G. Cohn (Ed.), *Proceeding of the 11th European Conference on Artificial Intelligence* (pp. 590-594). Chichester: John Wiley & Sons.

Fonseca, C. M, and Fleming, P. J. (1993). Genetic algorithms for multiobjective optimization: Formulation, discussion and generalization. In S. Forrest (Ed.), *Proceeding of the 5th International Conference on Genetic Algorithms*, 416-423.

Giffer, B., & Thomson, G. T. (1960). Algorithms for solving production scheduling problems. *Operations Research, 18,* 487-503.

Giri, S, C., & Goyal, S. K. (2001). Recent trends in modeling of deteriorating inventory. *European Journal of Operational Research, 134,* 1-16.

Goldberg, D. E., Deb. K., & Horn, J. (1992) Massive multimodality, deception, and genetic algorithms, In R. Manner & B. Manderick (Eds.), *Parallel Problem Solving from Nature II* (pp. 37-46). Amsterdam.

Grefenstette, J. J. (Ed.). (1985) *Proceedings of an International Conference on Genetic Algorithms and Their Applications.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Grefenstette, J. J., (Ed.). (1987) *Proceedings of Second International Conference on Genetic Algorithms.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Hadley, G., & T. M. Whitin (1963) *Analysis of Inventory Systems.* Englewood Cliffs, NJ: Prentice-Hall.

Hajela, P. (1990) Genetic search: An approach to the non-convex optimization problem, *AAAI Journal*, 28, 1205-1210.

Hall, R.W (1983) *Zero Inventories.* Homewood, Dow Jones, Irwin.

Hromkorie, J. (2003). *Algorithmics for hard problems* (2nd ed.). Berlin: Springer-Verlag.

Ishibuchi, H., T. Yoshida, and T. Murata (2003). Balance between genetic search and local search in memetic algorithms for multiobjective permutation flowshop scheduling. *IEEE Transaction on Evolutionary Computation, 7,* 204-223.

Karaboga, D., & Pham, D. T. (2000). *Intelligent optimization techniques.* Berlin: Springer-Verlag.

Kirkpatrick, S., Gelatt, C. D., & Vecchi, M. P. (1983). Optimization by simulated annealing. *Science*, *220*, 671-680.

Liaw, C. F. (2000). A hybrid genetic algorithm for the open shop scheduling problem. *Euro-*

pean Journal of Operational Research, *124*(1), 28-42.

Martin, O., & Otto, S. W. (1996). Combining simulated annealing with local search heuristics. *Annals of Operations Research*, *63*, 57-75.

Mitchell, M. (1996). *An introduction to genetic algorithms.* Cambridge, MA: MIT Press.

Nahmias, S. (1997). *Production and Operations Analysis* (3rd ed.). McGraw Hill International Editions.

Osman, I. H., & Kelly, J. P. (Eds.). (1996). *Meta-heuristics: Theory and applications.* Boston: Kluwer Academic Publishers.

Pelikan, M., Sastry, K., & Cantu-Paz, E. (2006). *Scalable optimization via probabilistic Modeling: From algorithms to applications (Studies in computational intelligence).* Berlin: Springer Verlag.

Rao, S. S. (1984). *Optimization Theory and Application*. New Delhi: Wiley Eastern.

Rawlins, G. J. E. (Ed.). (1991). *Foundations of Genetic Algorithms*. San Francisco: Morgan Kaufmann.

Reinfled, N.V. (1987). *Handbook of Production and Inventory Control*. NJ: Prentice Hall.

Reklaitis, G. V., Ravindran, A., & Ragsdell, K. M. (1983). *Engineering Optimization Methods and Applications*. New York: Wiley.

Romeo, F., & Sangiovanni-Vincentelli, A. (1991). A theoretical framework of simulated annealing. *Algorithmica*, *6*(3), 302-345.

Silver, E. A. (1976). Establishing the order quantity when the amount received is uncertain. *INFOR*, *14*, 32-37.

Srinivas, N. and Deb, K. (1995). Multiobjective function optimization using nondominated sorting genetic algorithms. *Evolutionary Computation*, *2*(3), 221-248.

Starr, M. K. & Miller, D. W. (1990). *Inventory Control: Theory and Practice*. Englewood Cliffs, NJ: Prentice Hall.

Thierens, D. and Goldberg, D. E. (1993). Mixing in genetic algorithms, In S. Forrest (Ed.), *op. cit.* (pp. 38-45).

Whitley, D. (Ed.). (1993). *Foundations of Genetic Algorithms II.*, San Francisco: Morgan Kaufmann.

Wright, A. (1991). Genetic algorithms for real parameter optimization. In G. J. E. Rawlins (Ed.), *op. cit.* (pp. 205-220).

Chapter VI
Condition Monitoring Using Computational Intelligence

Tshilidzi Marwala
University of the Witwatersrand, South Africa

Christina Busisiwe Vilakazi
University of the Witwatersrand, South Africa

ABSTRACT

Condition monitoring techniques are described in this chapter. Two aspects of condition monitoring process are considered: (1) feature extraction; and (2) condition classification. Feature extraction methods described and implemented are fractals, kurtosis, and Mel-frequency cepstral coefficients. Classification methods described and implemented are support vector machines (SVM), hidden Markov models (HMM), Gaussian mixture models (GMM), and extension neural networks (ENN). The effectiveness of these features was tested using SVM, HMM, GMM, and ENN on condition monitoring of bearings and are found to give good results.

INTRODUCTION

Condition monitoring of machines is gaining importance in industry due to the need to increase machine reliability and decrease the possible loss of production due to machine breakdown. By definition, condition monitoring is performed when it is necessary to access the state of a machine and to determine whether it is malfunctioning through reason and observation (William, Davies, & Drake, 1992). Condition monitoring can also be defined as a technique or process of monitoring the operating characteristics of a machine so that changes and trends of the monitored signal can be used to predict the need for maintenance before a breakdown or serious deterioration occurs, or to estimate the current condition of a machine. Condition monitoring has become increasingly important, for example, in manufacturing companies due to an increase in the need for normal undisturbed operation of equipment in manufacturing. An unexpected fault or shutdown can result in a serious accident and financial loss for a company. Manufacturing companies must find ways to avoid failures, minimize downtime, reduce maintenance costs, and lengthen the lifetime of their equipments. With reliable condition monitoring process, machines can be utilized in a more optimal fashion. Time-based maintenance follows a schedule to decide when maintenance is to be conducted. This leads to inefficiencies because either the maintenance may be conducted needlessly early or a failure may happen before scheduled maintenance takes place. Condition monitoring can therefore be used for condition based maintenance, or predictive maintenance.

Rotating machines are used in various industrial applications. One of the most common components in modern rotating machinery is the rolling element bearing. Most machine failures are linked to bearing failure (Lou & Loparo, 2004), which often result in lengthy downtime that have economic consequences. As a result, an increasing volume of condition monitoring data are captured and presented to engineers. This leads to these key problems, such as, the data volume is too large for engineers to deal with and the relationship between the plant item, its health, and the data generated is not always well understood. Therefore, the extraction of meaningful information from data is difficult. Hence, a reliable, fast, and automated diagnostic technique allowing relatively unskilled operators to make important decisions without the need for a condition monitoring specialist to examine the data and diagnose problems is required. The most commonly used condition monitoring system is vibration-based condition monitoring. Vibration monitoring is based on the principle that all systems produce vibration. When a machine is operating properly, vibration is small and constant; however, when faults develop and some of the dynamic processes in the machine change, the vibration spectrum also changes (Marwala, 2001).

The success of a classification system depends very much on the effectiveness of the extracted features. Another crucial step is to establish a reliable and effective condition monitoring classification system. The objective of this chapter is to give a review of three feature selection techniques that have been used recently for bearing fault diagnosis (Nelwamondo, Marwala, & Mahola, 2006) These techniques are: 1) Mel-frequency cepstral coefficients (MFCC), which is a time-frequency domain analysis technique that has been used extensively in speech recognition; 2) kurtosis, which is the time domain analysis method; and 3) Fractal dimension analysis method which is a time domain analysis method that has been applied to problems in image processing. This chapter also evaluates the effectiveness of the extracted feature for bearing fault diagnosis with the support vector machine (SVM), hidden Markov model (HMM), Gaussian mixture model (GMM) and extension neural network (ENN) classifiers. SVM was chosen as it has been applied successfully in many fault diagnosis applications. The inspiration for the use of GMM and HMM is their success in speech recognition, and ENN was chosen because of its success for pattern recognition of partial discharges.

BACKGROUND

As mentioned earlier, the success of a classification system depends on the effectiveness of the extracted observation sequence to represent a particular machine state or condition. During the past decades, considerable research effort has been put into the development of various feature extraction techniques and condition monitoring systems. Feature extraction techniques can be classified into three domains, namely; frequency domain analysis, time-frequency domain analysis, and time domain analysis (Ericsson, Johansson, Persson, Sjöberg, & Strömberg, 2004). The frequency domain methods often involve frequency analysis of the vibration signals and look at the periodicity of high frequency transients. In the processes, the frequency domain methods search for a train of repetitions occurring at any of the characteristic defect frequencies (Ocak & Loparo, 2004). This procedure gets complicated considering the fact that the periodicity of the signal may be suppressed. These frequency domain techniques include the frequency averaging technique, adaptive noise cancellation, and the high frequency resonance technique (HFRT) amongst others. The HFTR is the most popular for bearing fault detection and diagnosis (Ocak & Loparo, 2004). The disadvantage of the HFTR technique is that it requires several impact tests to determine the bearing resonance frequency; hence, it becomes computationally expensive (Ocak & Loparo, 2004). McFadden and Smith (1984) present the envelope analysis which is another commonly used frequency domain technique for detection and diagnosis of bearing faults. The main disadvantage of the frequency domain analysis is that it tends to average out transient vibrations and therefore becomes more sensitive to background noise. To overcome this problem, the time-frequency domain analysis is used which shows how the frequency contents of the signal changes with time. The examples of such analyses are short time Fourier transform (STFT), the Wigner-Ville distribution (WVD), and most commonly the wavelet transform (WT). These techniques are studied in detail by Li et al. (2000). The last category of the feature extraction is the time domain analysis. Time domain methods usually involve indices that are sensitive to impulsive oscillations, such as peak level, root mean square (RMS) value, crest factor analysis, kurtosis analysis, shock pulse counting, time series averaging method, signal enveloping method, and many more (Li, Chow, Tipsuwan, & Hung, 2000; Ocak & Loparo, 2004). Ericsson et al. (2004) and Li et al. (2000) show that unlike the frequency domain analysis, the time-domain analysis is less sensitive to suppressions of the periodicity.

Various feature extraction techniques have been implemented successfully for vibration-based condition monitoring. Ocak and Loparo (2004) and Nikolaou and Antoniadis (2002) have used wavelet transforms to detect and classify different faults in bearings while Rojas and Nandi (2006) use spectral and statistical features for the classification of bearing faults. Peng, Tse, and Chu (2005) compare the Hilbert-Huang transform with wavelet transform in bearing fault diagnosis. Junsheng, Dejie, and Yu (2006) propose a feature extraction method based on empirical mode decomposition method and autoregressive model for the roller bearings. Baillie and Mathew (1996) implemented an autoregressive modeling that does not only classify, but also provides a one-step-ahead prediction of the vibration signal using the previous outputs. Yang et al. (2006) apply the basis pursuit and obtained better results than wavelet transforms. Altman and Mathew (2001) also use the discrete wavelet packet analysis to enhance the detection and diagnostic of rolling element bearing faults. Prabhakar, Mohanty, and Sekhar (2002) show that DWT can be used to improve the detection of bearing faults.

The second crucial decision in machine condition monitoring is to choose an effective classification system. Condition classification includes the identification of the operating status of the machine and type of failure by interpreting the representative system condition. The classification system can be classified into two main groups: knowledge-based and data-based models. Knowledge-based models rely on human-like knowledge of the process and its faults. Knowledge-based models like expert systems or decision trees apply human-like knowledge of the process for fault diagnosis. In fault diagnostics, the human expert could be a person who operates the diagnosed machine or process and who is very well aware of different kinds of faults occurring in it. Building the knowledge base can be done by interviewing the human operator on fault occurrences in the diagnosed machine and on their symptoms. Expert systems are usually suitable for problems, where a human expert can linguistically describe the solution. Typical human knowledge is vague and inexact, and handling this kind of information has often been a problem with traditional expert systems. For example, the limit when the temperature in a sauna is too high is vague in human mind. In practice, it is very difficult to obtain adequate representations of the complex and highly nonlinear behavior of faulty plants using quantitative models. Knowledge-based models may be utilized together with a simple signal-based diagnostics, if the expert knowledge of the process is available. However, it is often impossible even for a human expert to distinguish faults from the healthy operation, and also multiple information sources may need to be used for trustworthy decision making. Thus, the data-based models are the most flexible approach to automated condition monitoring.

Data-based models are applied when the process model is not known in the analytical form and expert knowledge of the process performance under faults is not available. The data-based models can be created in numerous ways. During the last years artificial neural network based models like multilayer perceptron (MLP) and radial basis function (RBF) have been used extensively for bearing condition monitoring. Samanta and Al-Bushi (2003) use artificial neural network with time-domain features for rolling element bearing detection. Yang, Han, and An (2004) apply the ART-KOHONEN to the problem of fault diagnosis of rotating machinery. Lately, kernel-based classifiers such as support vector machine (SVM) have been used for bearing fault diagnosis. Rojas and Nandi (2006) use SVM for the detection and classification of rolling element bearing faults. Samanta (2004) use both artificial neural networks (ANN) and SVM with genetic algorithm for bearing fault detection. Yang et al. (2006) use multiclass SVM for fault diagnosis of rotating machinery. However, data-based statistical approaches have achieved considerable success in speech recognition and have been recently used for condition monitoring. Ertunc, Loparo, and Ocak (2004) use HMM to determine wear status of the drill bits in a drilling process. Ocak and Loparo (2004), Purushotham, Narayanana, and Prasadb (2005), Miao and Makis (2006), and Nelwamondo et al. (2006) use HMM for bearing fault detection and diagnosis.

FEATURE EXTRACTION

Various features that can be extracted from vibration signals of bearing elements have been investigated. This section presents a brief discussion of fractal analysis, Mel-frequency cepstral coefficients, and kurtosis.

Fractal Dimension

Most of the vibrations are periodic movements with some degree of turbulence. To detect different bearing faults, these nonlinear turbulence features must be extracted. The nonlinear turbulence features of the vibration signal are quantified using the fractal model (Maragos & Potamianos, 1999). To define the fractal dimension, let the continuous real-valued function, $s(t), 0 \le t \le T$, represents a short-time vibration signal. Furthermore, let the compact planar set (Maragos & Potamianos, 1999):

$$F = \{(t, s(t)) \in R^2 : 0 \le t \le T\} \tag{1}$$

The fractal dimension of a compact planar set F is called the Hausdorff dimension and its value generally lies between 1 and 2 (Maragos & Potamianos, 1999). The problem with this dimension is that it is only a mathematical concept and is therefore extremely hard to compute. Due to this, other methods are used to approximate this dimension such as the Minkowski-Bouligand dimension and box-counting dimension (Maragos & Potamianos, 1999). In this study, fractal dimension is approximated using the box-counting dimension, which is discussed further in the next section.

Box-Counting Dimension

The box-counting dimension (D_B) of, F, is obtained by partitioning the plane with a squares grids of side ε, and $N(\varepsilon)$ number of squares that intersect the plane and is defined as (Falconer, 1952):

$$D_B(F) = \lim_{\varepsilon \to 0} \frac{\ln N(\varepsilon)}{\ln(1/\varepsilon)} \tag{2}$$

Assuming a discrete bearing vibration signal, $s_1, s_2, ..., s_T$ then D_B is given by (Falconer, 1952):

$$D_B(F) = \left\{ J.\left(\sum_{j=1}^{J} \ln(1/\varepsilon_j).\ln(N(\varepsilon)) \right) - \left(\sum_{j=1}^{J} \ln(1/\varepsilon_j) \right).\left(\sum_{j=1}^{J} \ln N(\varepsilon) \right) \right\} / \left\{ J.\sum_{j=1}^{J} (\ln(1/\varepsilon_j))^2 - \left(\sum_{j=1}^{J} \ln(1/\varepsilon) \right)^2 \right\} \tag{3}$$

where J is the computation resolutions and $\varepsilon_{\min} \le \varepsilon_j \le \varepsilon_{\max}$ with $\varepsilon_{\max}$ and $\varepsilon_{\min}$ represent the maximum and minimum resolutions of computation. In Equation 3, D_B is equal to the slope obtained by fitting a line using the least squares method (Maragos & Potamianos, 1999).

Multiscale Fractal Dimension (MFD)

It should be noted that the fractal dimension discussed in the last section is a global measure and therefore does not represent all the fractal characteristics of the vibration signal (Falconer, 1952). To overcome this problem, the multiscale fractal dimension (MFD) set is created. The MFD ($D(s,t)$) of the vibration signal, s, is obtained by computing the dimensions over a small time window. This MFD set is obtained by dividing the bearing vibration signal into K frames, then K maximum computation resolutions are set as (Falconer, 1952):

$$\varepsilon_k^{\max} = k\varepsilon_{\min} (1 \le k \le K) \tag{4}$$

where $\varepsilon_{\min}$ is the same as before, which is the minimum valid resolution of the computation. The box-counting dimension expression can then be written as:

$$D^k(F) = \left\{ k.\left(\sum_{j=1}^{k} \ln(1/j\varepsilon_{\min}).\ln(N(j\varepsilon_{\min})) \right) - \left(\sum_{j=1}^{k} \ln(1/j\varepsilon_{\min}) \right).\left(\sum_{j=1}^{k} \ln N(j\varepsilon_{\min}) \right) \right\} / \left\{ k.\sum_{j=1}^{k} (\ln(1/j\varepsilon_{\min}))^2 - \left(\sum_{j=1}^{k} \ln(1/j\varepsilon_{\min}) \right)^2 \right\} \tag{5}$$

Finally, the corresponding MFD of the vibration signal is given by:

$$MFD(s) = \{D^1(s), D^2(s), ..., D^K(s)\} \tag{6}$$

where, $D^k(s)$ is the fractal dimension of the k^{th} frame and this is called the fractogram (Maragos & Potamianos, 1999).

Mel-Frequency Cepstral Coefficients (MFCCs)

MFCCs have been widely used in the field of speech recognition and are able to represent the dynamic features of a signal as they extract both linear and nonlinear properties. A MFCC can be a useful tool of feature extraction in vibration signals as vibrations contain both linear and nonlinear features. The Mel-frequency cepstral coefficients (MFCC) is a type of wavelet in which frequency scales are placed on a linear scale for frequencies less than 1 kHz and on a log scale for frequencies above 1 kHz (Wang, Wang, & Weng, 2002). The complex cepstral coefficients obtained from this scale are called the MFCC (Wang et al., 2002). The MFCC contain both time and frequency information of the signal and this makes them useful for feature extraction. The following steps are involved in MFCC computations.

1. Transform input signal *x(n)* from time domain to frequency domain by applying fast Fourier transform (FFT) using (Wang et al., 2002):

$$Y(m) = \frac{1}{F}\sum_{n=0}^{F-1} x(n)w(n)e^{-j\frac{2\pi}{F}nm} \tag{7}$$

   where F is the number of frames, $0 \le m \le F-1$ and $w(n)$ is the Hamming window function given by:

$$w(n) = \beta\left(0.5 - 0.5\cos\frac{2\pi n}{F-1}\right) \tag{8}$$

   where $0 \le n \le F-1$ and β is the normalization factor defined such that the root mean square of the window is unity (Wang et al., 2002).

2. Mel-frequency wrapping is performed by changing the frequency to the Mel using the following equation (Wang et al., 2002):

$$mel = 2595 \times \log_{10}(1 + \frac{f_{Hz}}{700}) \tag{9}$$

   Mel-frequency warping uses a filter bank, spaced uniformly on the Mel scale. The filter bank has a triangular band pass frequency response, whose spacing and magnitude are determined by a constant Mel-frequency interval.

3. The final step converts the logarithmic Mel spectrum back to the time domain. The result of this step is what is called the Mel-frequency cepstral coefficients. This conversion is achieved by taking the discrete cosine transform of the spectrum as:

$$C_m^i = \sum_{n=0}^{F-1}\cos\left(m\frac{\pi}{F}(n+0.5)\right)\log_{10}(H_n) \tag{10}$$

   where $0 \le m \le L-1$ and L is the number of MFCC extracted form the i^{th} frame of the signal. H_n is the transfer function of the n^{th} filter on the filter bank. These MFCC are then used as a representation of the signal.

Kurtosis

There is a need to deal with the occasional spiking of vibration data, which is caused by some types of faults and to achieve this task kurtosis is used. Kurtosis features of vibration data have also been successfully used in tool condition monitoring by El-Wardany (1996). The success of kurtosis in vibration signals is based on the fact that vibration signals of a system under stress or having defects differ from those of a normal system. The sharpness or spiking of the vibration signal changes when there are defects in the system. Kurtosis is a measure of the sharpness of the peak and is defined as the normalized fourth-order central moment of the signal (Wang, Willett, DeAguiar, & Webster, 2001). The kurtosis value is useful in identifying transients and spontaneous events within vibration signals (Wang et al., 2001) and is one of the accepted criteria in fault detection. The calculated kurtosis value is typically normalized by the square of the second moment, as shown in Equation 11. A high value of kurtosis implies a sharp distribution peak and indicates that the signal is impulsive in nature (Altman & Matthew, 2002).

$$K = \frac{1}{N}\sum_{i=1}^{N}\frac{(x_i - \overline{x})^4}{\sigma^4} \tag{11}$$

where $\overline{x}$ the mean and σ is the variance.

CLASSIFICATION SYSTEM

Once the relevant features are extracted from vibration signals, these features are used for automatic bearing fault detection and diagnosis by applying them to a nonlinear classifier. SVM has outperformed ANN for a number of classification problems especially in condition monitoring. SVM has been successfully used for bearing fault detection and diagnosis (Samanta, 2004). Nonlinear classifiers like Gaussian mixture model (GMM) and hidden Markov model (HMM) have shown to outperform ANN in a number of classification problems but mostly in speech related problems. Only recently, researchers like Purushotham et al. (2005) have applied speech pattern classifiers such as HMM to fault detection of mechanical systems due to their success in speech recognition. ENN is a fairly new nonlinear classifier and it has shown to outperform ANN for partial discharge pattern recognition (Wang, 2005). A brief background of SVM, GMM, HMM, and ENN is presented below.

Support Vector Machine (SVM)

SVM is a powerful widely used technique for solving supervised classification problems due to its generalization ability. In essence, SVM classifiers maximize the margin between the training data and the decision boundary, which can be formulated as a quadratic optimization problem in the feature space. The subsets of patterns that are closest to the decision boundary are called support vectors. For a linearly separable binary classification problem, the construction of a hyperplane, $w^t x + b = 0$, so that the margin between the hyperplane and the nearest point is maximized can be posed as the following quadratic optimization problem:

$$\min_{w} \frac{1}{2}(w^T w) \tag{12}$$

subject to,

$$d^i((w^T x^j) + b) \geq 1 \tag{13}$$

where $d^i \in \{-1, 1\}$ stands for i^{th} desired output, and $x^i \in R^P$ stands for the i^{th} input sample of the training data set $\{x^i, d^i\}_{i-1}^{N}$. Equation 13 forces a rescaling on (w,b) so that the point closest to the hyperplane has a distance of $1/\|w\|$. Maximizing the margin corresponds to minimizing the Euclidean norm of the weight vector. Often in practice, a separating hyperplane does not exist. Hence the constraint is relaxed by introducing slack variable $\xi_{i \geq o,} i = 1, ..., N$. The optimization problem for penalty constant C becomes

$$\min_{w,\xi} \frac{1}{2}(w^T w) + C\sum_{i=1}^{N}\xi_i \tag{14}$$

subject to

$$d^i((w^T x^j) + b) \geq 1 - \xi_i \tag{15}$$

where,

$$\xi_i \geq 0 \quad with \quad i = 1, ..., N$$

The C controls the tradeoff between the robustness of the machine and the number nonseparable points. By introducing the Langrange multiplier α_i and using the Karush-Kuhn-Tucker theorem of optimization theory (Vapnik, 2005), the decision function for the vector *x* then becomes (Bishop, 1995):

$$f(x) = \mathrm{sgn}\left(\sum_{i=1}^{N} d^i \alpha_i \left\langle x, x^i \right\rangle + b \right) \tag{16}$$

By replacing the inner product $\left\langle x, x^i \right\rangle = (x^T)(x^i)$ with a kernel function $k(x, x^i)$, the input data are mapped to a higher dimensional space. It is then in this higher dimensional space that a separating hyperplane is constructed to maximize the margin. In the lower dimensional data space, this hyperplane becomes a nonlinear separating function. The typical kernel functions are the polynomial kernel $k(x, x^i) = (x \times x^i + 1)^d$ and the Gaussian kernel $k(x, x^i) = \exp(-(x - x^i)^2 / \delta^2)$, where d is the degree of the polynomial and δ^2 is the bandwidth of the Gaussian kernel.

Hidden Markov Model (HMM)

HMM has recently been applied to various conditions monitoring application machine tool monitoring (Owsley, Atlas, & Bernard, 1997), and bearing fault detection (Lou & Loparo, 2004; Purushotham et al., 2005). HMM is a stochastic signal model and is referred to as Markov sources or probabilistic functions of Markov chains (Rabiner, 1989). A Markov chain is a random process of discrete-valued variables that involves a number of states. These states are linked by possible transitions, each with an associated probability and each state has an associated observation. The state transition is only dependent on the current state and not on past states. The actual sequence of states is not observable and hence the name *hidden Markov* (Ertunc et al., 2001). The compact representation for an HMM with a discrete output probability distribution is given by:

$$\lambda = \{A, B, \pi\} \tag{17}$$

where λ is the model, $A = \{a_{ij}\}$, $B = \{b_{ij}(k)\}$, and $\pi = \{\pi_i\}$ is a transition probability distribution, the observation probability distribution and initial state distribution, respectively. These parameters of a given state, S_i, are defined as (Rabiner, 1989):

$$a_{ij} = P(q_{t+1} = S_j \mid q_t = S_i), \quad 1 \le i, j \le N \tag{18}$$

$$\begin{aligned} &b_{ij}(k) = P(o_k \mid q_t = S_i), \\ &1 \le j \le N, 1 \le k \le M \end{aligned} \tag{19}$$

$$\pi_i = P(q_1 = S_i)), \qquad 1 \le i \le N \tag{20}$$

where q_t the state at time *t* and *N* is denotes the number of states. Furthermore, o_k is the k^{th} observation and *M* is the number of distinct observation. HMM is thus a finite-state machine which changes state every time unit. There are three basic HMM problems to be solved. First, evaluation, which finds the probability of the observation sequence, $O = o_1, o_2, ..., o_T$, of visible states generated by the model λ. Using the model in Equation 21, the probability is computed as (Rabiner, 1989):

$$P(O, \lambda) = \sum_{allS} \pi_{S_0} \prod_{t=0}^{T=1} a_{S_t S_{t+1}} b_{S_{t+1}}(o_{S_{t+1}}) \tag{21}$$

Second, decoding, which finds a state sequence that maximizes the probability of observation sequence, is realized by the so-called Viterbi algorithm (Viterbi, 1967). Lastly, training which adjusts model parameters to maximize probability of observed sequence.

Gaussian Mixture Models (GMM)

Gaussian mixture models have been a reliable classification tool in many applications of pattern recognition, particularly in speech and face recognition. GMM have proved to perform better than hidden Markov models in text independent speaker recognition (Reynolds, Quatieri, & Dunn, 2003). The success of GMM in classification of dynamic signals has also been demonstrated by many researchers such as Cardinaux, Cardinaux, and Cardinaux et al. (2003) who compared GMM and MLP in face recognition and found that the GMM approach easily outperforms the MLP approach for high resolution faces and is significantly more robust in imperfectly located faces. Other advantages of using GMM are that it is computationally inexpensive and is based on well understood statistical models (Reynolds et al., 2003). GMM works by creating a model of each fault which is written as (Reynolds et al., 2003):

$$\lambda = (w, \mu, \Sigma) \tag{22}$$

where λ is the model, w represents the weights assigned to the Gaussian means, μ is the diagonal variance of the features used to model the fault, and Σ is the covariance matrix. GMM contains a probability density function of the observation consisting of a sum of normal observations. A weighted sum of Gaussians normally provides an accurate model of the data. Each Gaussian comprises a mean and a covariance, hence, a mixture of components. The Gaussian probability density function is given by (Reynolds et al., 2003):

$$p(x) = \frac{1}{\sigma\sqrt{2x}} e^{\frac{-(x-\mu)^2}{2\sigma^2}} \tag{23}$$

where μ is the mean and σ is the standard deviation of the distribution of a variable x. For a case where x is a vector of features, Equation 23 becomes (Reynolds et al., 2003):

$$p(x) = \frac{1}{\sqrt{(2\pi)^n |\Sigma|}} e^{-\frac{1}{2}\left[(x-\mu)\Sigma^{-1}(x-\mu)\right]} \tag{24}$$

where n is the size of the vector feature, x. The log-likelihood is then computed as (Reynolds et al., 2003):

$$\hat{s} = \boldsymbol{\arg\max_{1 \le f \le F}} \sum_{k=1}^{K} \log \boldsymbol{p(x_k | \lambda_f)} \tag{25}$$

where f represents the index of the type of fault, whereas F is the total number of known fault conditions and $x = \{x_1, x_2, ..., x_K\}$ is the unknown fault vibration segment. $P(x_k|\lambda_f)$ is the mixture density function. An arbitrary probability density of a sample vector x can be approximated by a mixture of Gaussian densities (Bishop, 1995) as:

$$p(x|\lambda) = \sum_{i=1}^{M} w_i p(x) \tag{26}$$

where all mixtures weights, w_i, are adjusted to satisfy the constrains (Meng, Frei, Osorio, Strang, & Nguyen, 2005) and $0 \le w_i \le 1$. Training GMM is a fast and straightforward process which estimates the mean and covariance parameters from the training data (Meng et al., 2005). The training procedure estimates the model parameters from a set of observations using the expectation maximization (EM) algorithm. The EM algorithm tries to increase the expected log-likelihood of the complete data x given the partially observed data (Meng et al., 2005) and finds the optimum model parameters by iteratively refining GMM parameters for a given bearing fault feature vector.

Extension Neural Network (ENN)

ENN is a new pattern classification system based on concepts from neural networks and extension theory (Wang & Hung, 2003). The extension theory uses a novel distance measurement for classification processes, and the neural network can embed the salient features of parallel computation and learning capability. The classifier is well suited to classification problems where there exist patterns with a wide range of continuous inputs and a discrete output indicating which class the pattern belongs to. ENN comprises of an input layer and an output layer. The input layer nodes receive an input feature pattern and use a set of weighted parameters to generate an image of the input pattern. There are two connection weights between input nodes and output nodes; one connection represents the lower bound for this classical domain of features and the other represents the upper bound. The entire network is thus represented by a matrix of weights for the upper and lower limits of the features for each class W_U and W_L. A third matrix representing the cluster centers is also defined as:

$$z = \frac{W_u + W_l}{2} \tag{27}$$

ENN uses supervised learning, which tunes the weights of the ENN to achieve a good clustering performance or to minimize the clustering error. The network is trained by adjusting the network weights and recalculating the network centers for each training pattern depending on the extension distance (ED) of that pattern to its labeled cluster. Each training pattern adjusts the network weights and the centers by amounts that depend on the learning rate. In general, the weight update for a variable x_i is:

$$w^{new} = w^{old} - \eta(x_i - w^{old}) \tag{28}$$

where η is the learning rate and w can either be the upper or the lower weight matrices of the network centers. It can be shown that for t training patterns for a particular class C, the weight is given by (Mohamed, Tettey, & Marwala, 2006):

$$w^c(t) = (1-\eta)w^c(0) - \eta\sum(1-\eta)^{t-1}x_i^c \tag{29}$$

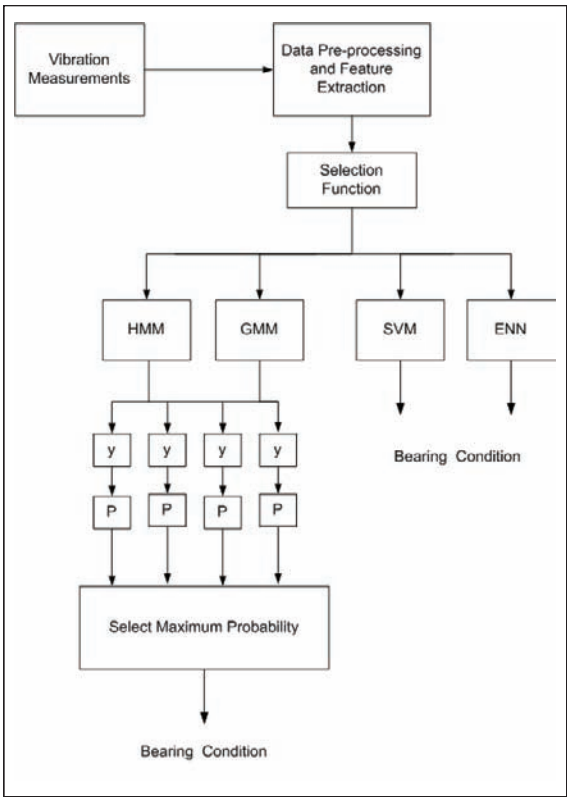
This equation demonstrates how each training pattern reinforces the learning in the network by having the most recent signal determine only a fraction of the current value of z_c^t. The equation indicates that there is no convergence of the weight values since the learning process is adaptive and reinforcing. Equation 29 also highlights the importance of the learning rate, η. Small values of η require many training epochs, whereas large values may results in oscillatory behavior of the network weights, resulting in poor classification performance.

PROPOSED ARCHITECTURE

The architecture of the proposed framework is shown in Figure 1. As shown in this figure the framework consists of two major stages, namely, data preprocessing with feature extraction and classification stage.

The vibration signal is preprocessed by first dividing the vibration signals into segments of equal lengths. Preprocessing is followed by extraction of features of each window using feature extraction techniques as explained earlier. The vibration signal was first broken into segments, each being five revolutions long. Firstly, MFCC features were extracted from each frame of the signal. During the extraction of MFCC features, each segment of the signal was further broken into 14 frames of equal duration. The number of MFCC features was varied from 9 to 16 in order to obtain the optimal value. Secondly, fractal features were extracted in the same format with different MFD size. The MFD size was also

Figure 1. The proposed block diagram for the condition monitoring



varied from 2 to 20 in order to obtain the MFD size that gives best results. Lastly, one kurtosis feature from each segment was extracted. When relevant features are extracted, reference models that will be used to classify faults are built and are used to classify the fault conditions. SVM, GMM, HMM, and ENN classifiers were implemented. Figure 1 also shows that both GMM and HMM build models for all possible machine condition and the normal condition. For GMM and HMM, diagnosis of the bearing fault is achieved by calculating the probability of the feature vector, given the entire previously constructed fault model. GMM or HMM with maximum probability then determines the bearing condition.

RESULTS AND DISCUSSION

Vibration Data

The investigation in this chapter is based on the data obtained from the Case Western Reserve University Web site (Loparo, 1998). The experimental setup is comprised of a Reliance Electric 2HP IQPreAlert connected to a dynamometer. Faults of size 0.007, 0.014, 0.021, and 0.028 inches were introduced into the drive-end bearing of a motor using the electric discharge machining (EDM) method. These faults were introduced separately at the inner raceway, rolling element, and outer raceway. An impulsive force was applied to the motor shaft and the resulting vibration was measured using two accelerometers, one mounted on the motor housing and the other on the outer race of the drive-end bearing. All signals were recorded at a sampling frequency of 12 kHz.

Results

Figure 2 shows samples bearing vibration signals for the four bearing conditions. In order to diagnose faults, features need to be extracted and be classified.

The optimal classifier parameters were found using exhaustive search. The optimum SVM architecture used polynomial kernel function with degree of 5. The optimum HMM architecture used for experimentation is a two state model with diagonal covariance matrix that contains 10 Gaussian mixtures. GMM architecture also used diagonal covariance matrix with three centers. The main advantage of using the diagonal covariance matrix for GMM and HMM is that this decorrelates the feature vectors. This is necessary since fractal dimensions are highly correlated values. ENN architecture with an optimal learning rate of 0.219 was used.

The first set of experiments measures the effectiveness of the time domain fractal dimension based feature extraction using vibration signal.

Figure 2. Vibration signals on the bearing under normal condition, inner race fault, outer race fault and ball fault conditions

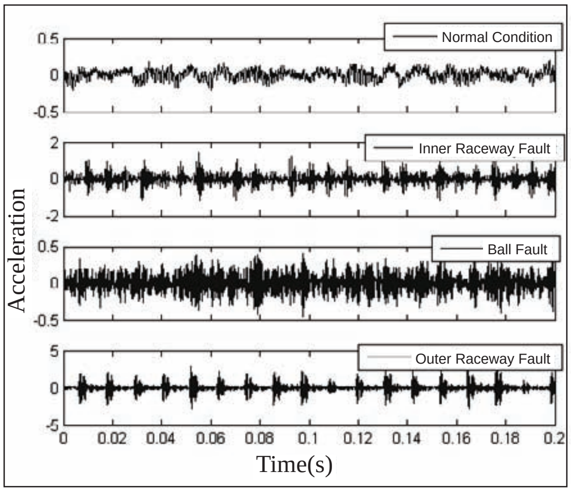


Figure 3. MFD feature extraction comparison for the normal, inner, outer and ball fault for the 1s vibration signal

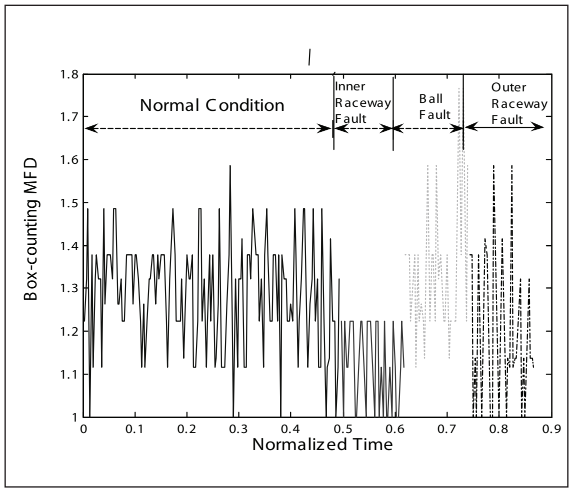


Figure 4. The graph of the change classification rate with change in MFD size

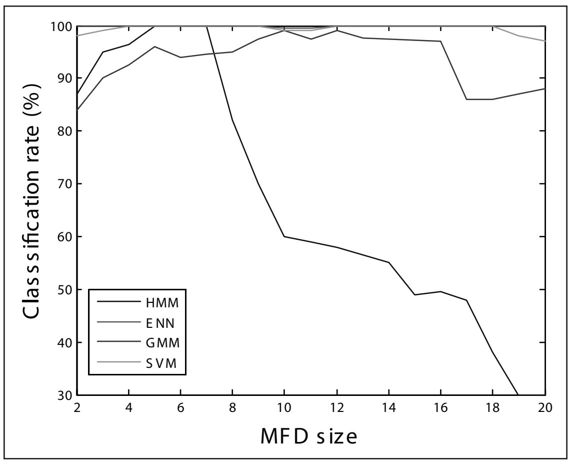


Figure 5. Change in classification rate with change in the number of MFCC

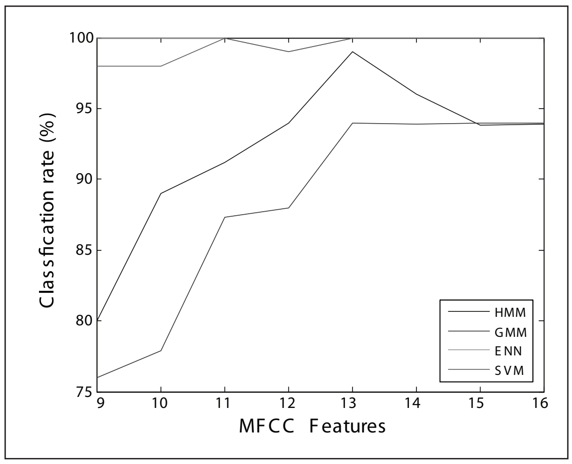


Figure 3 shows the MFD feature vector which extracts the bearing fault specific information, plotted for the first second of the vibration signal. Figure 3 shows that the proposed feature extraction technique does extract fault specific features which are used to classify different bearing conditions. However, the optimum size of the MFD must be found.

Figure 4 shows the change of the system accuracy with the change of the MFD size. This figure shows that the size of MFD does not affect the classification accuracy of SVM and

Table 1. The confusion matrix for the SVM, HMM, GMM, and ENN classifier used with fractal features

	SVM				HMM			
	Normal	Inner	Outer	Ball	Normal	Inner	Outer	Ball
Normal	100	0	0	0	100	0	0	0
Inner	0	100	0	0	0	100	0	0
Outer	0	0	100	0	0	0	100	0
Ball	0	0	0	100	0	0	0	100
	GMM				ENN			
	Normal	Inner	Outer	Ball	Normal	Inner	Outer	Ball
Normal	100	0	0	0	100	0	0	0
Inner	0	100	0	0	0	100	0	0
Outer	0	0	100	0	0	0	100	0
Ball	1.8	0	0	98.2	0	0	0	100

Figure 6. MFCC values corresponding to different fault conditions

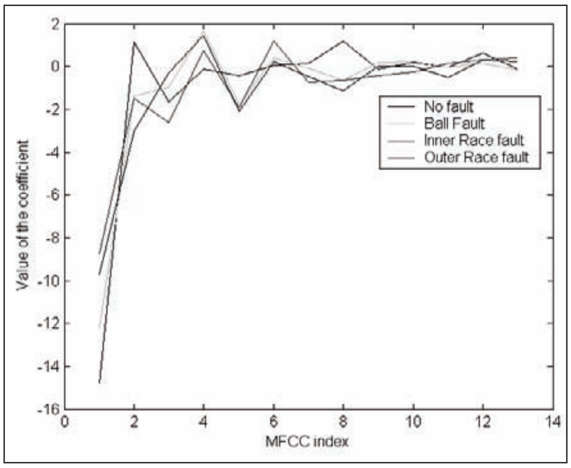
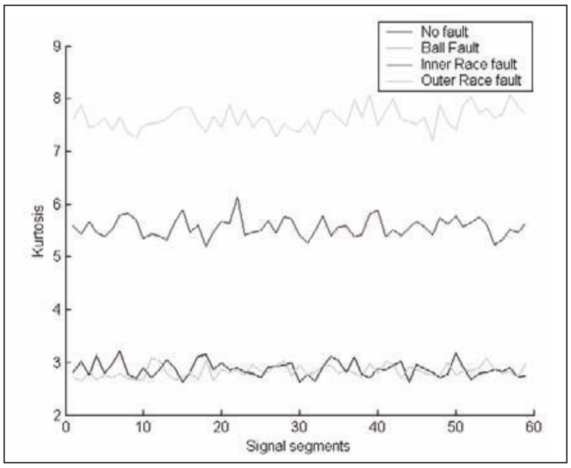


Figure 7. Kurtosis values corresponding to different fault conditions



ENN. It also shows that the GMM generally has a large optimum MFD size of 13 compared to 5 for HMM.

Using optimum SVM, HMM, GMM, and ENN architecture together with MFD, the confusion matrix obtained for different bearing faults are presented for HMM, GMM, and ENN in Table 1.

We further investigate the use of MFCC with the GMM, HMM, and ENN. We first observe the effect of varying the number of MFCC on the classification rate for the classifiers.

The figure shows that the number of MFCC does not affect the classification accuracy of SVM and ENN. Figure 5 shows that 13 MFCC give optimal results for both HMM and GMM and will

Table 2. The confusion matrix for the SVM, GMM, HMM, and ENN classifiers used with MFCC features

	SVM				HMM			
	Normal	Inner	Outer	Ball	Normal	Inner	Outer	Ball
Normal	100	0	0	0	100	0	0	0
Inner	0	100	0	0	0	100	0	0
Outer	0	0	100	0	0	0	100	0
Ball	0	0	0	100	0	0	0	100
	GMM				ENN			
	Normal	Inner	Outer	Ball	Normal	Inner	Outer	Ball
Normal	86.6	0	0	13	100	0	0	0
Inner	0	96.6	3.3	0	0	100	0	0
Outer	0	0	100	0	0	0	100	0
Ball	3.7	1.8	0	94	0	0	0	100

Table 3. Summary of classification results

	SVM (%)	***HMM (%)***	***GMM (%)***	***ENN (%)***
Fractal Features	100	100	99	100
MFCC	100	99	94	100
MFCC + Kurtosis	100	100	99	100

therefore be used. It is observed that increasing the number of MFCC above 13 does not improve the classification results.

The results obtained with MFCC features are shown in the confusion matrix in Table 2. On trying to improve the performance of the MFCC with GMM, we now add the kurtosis features to the MFCC. The kurtosis was computed in time for each segment of the signal. Figure 6 shows how kurtosis of the vibration signals varies with different fault conditions for 60 segments of the vibration data.

When kurtosis values are added to the MFCC, an improvement of about 5% is obtained for the GMM. The good classification accuracy of classifier is due to the fact that features seem to be more unique for different fault conditions. Overall results are summarized in Table 3.

CONCLUSION

The chapter discussed the two crucial requirements for an automated condition monitoring system. The first requirement is a feature extraction technique that can effectively and accurately extract the condition specific features. The second requirement is a classification system that can effectively classify the machine conditions. This chapter gives a review of three feature extraction techniques that are used for condition monitoring. These techniques are fractal analysis, Mel frequency cepstral coefficients and kurtosis. The effectiveness of the extracted feature was tested using four classifiers, namely, support vector machines, hidden Markov models, Gaussian mixture models, and the extension neural network. As vibration-based condition monitoring is the most popular condition monitoring system in machine, the proposed condition monitoring system is tested on bearings. The proposed system gives very good results for bearing fault diagnosis. It should be noted, however, that the proposed condition monitoring system can be applied to various condition monitoring.

FUTURE RESEARCH DIRECTION

The proposed system gave very good results for bearing condition monitoring, and it should be noted that this system is not limited to bearings only. The general trend is that condition monitoring systems require knowledge and understanding of the links between actual faults and the data generated as a result of these faults. When such understanding does not exist, often detailed data mining and analysis activities are undertaken to derive the required knowledge, data patterns, and relationships. Due to the need for automated condition monitoring system, the latter approach has been applied extensively in various condition monitoring systems. However, many new condition monitoring systems are implemented when there is no underlying knowledge of the data expected. Neither is there any historical data from which patterns and trends can be derived. These systems use online learning algorithm that enable them to incrementally learn the system as data become available. Furthermore, it is now widely recognized that problems due to the complexity of condition monitoring can be overcome with architectures that contain many dynamically interacting intelligent distributed modules, called intelligent agents.

REFERENCES

Altman, J., & Mathew, J. (2001). Multiple band-pass autoregressive demodulation for rolling element bearing fault diagnosis. *Mechanical Systems and Signal Processing, 15*(5), 963-997.

Baillie, D. C., & Mathew, J. (1996). A comparison of autoregressive modeling techniques for fault diagnosis of rolling element bearings. *Mechanical Systems and Signal Processing*, *10*(1), 1-17.

Bishop, C. M. (1995). *Neural networks for pattern recognition*. Oxford University Press.

Cardinaux, F., Cardinaux, C., & Marcel, S. (2003) Comparison of MLP and GMM classifiers for face verification on XM2VTS. *Springer-Verlag GmbH*, *2688*, 911-920.

El-Wardany, T. I., Gao D., & Elbestawi, M. A. (1996). Tool condition monitoring in drilling using vibration signature analysis. *International Journal of Machine Tools & Manufacture*, *36*(6), 687-711.

Ericsson, S., Grip, N., Johansson, E., Persson, L. E., Sjöberg, R., & Strömberg, J.O. (2004). Towards automatic detection of local bearing defects in rotating machines. *Mechanical Systems and Signal Processing*, *19*(3), 509-535.

Ertunc, H. M., Loparo, K. A., & Ocak, H. (2001). Tool wear condition monitoring in drilling operations using hidden Markov models. *International Journal of Machine Tools & Manufacture*, *41*, 1363-1384.

Falconer, K. (1952). *Fractal geometry, mathematical foundations and application*. New York: John Wiley.

Hu, Q., He, Z., Zhang, Z., & Zi, Y. (2007). Fault diagnosis of rotating machine based on improved wavelet package transform and SVMs ensemble. *Mechanical Systems and Signal Processing*, *21*(2), 688-705.

Junsheng, C., Dejie, Y., & Yu, Y. (2006). A fault diagnosis approach for roller bearings based on EMD method and AR model. *Mechanical Systems and Signal Processing*, *20*(2), 350-362.

Li, B., Chow, M. Y., Tipsuwan, Y., & Hung, J. C. (2000). Neural-network-based motor rolling bearing fault diagnosis. *IEEE Transactions on Industrial Electronics*, *47*(5), 1060-1068.

Loparo, K. A. (1998). Bearing data centre. *Case Western Reserve University*. Retrieved November 19, 2005, from, http://www.eecs.cwru.edu/laboratory/bearing

Lou, X., & Loparo, K. A. (2004). Bearing fault diagnosis based on wavelet transform and fuzzy inference. *Mechanical Systems and Signal Processing*, *18*(5), 1077-1095.

Maragos, P., & Potamianos, A. (1999). Fractal dimensions of speech sounds: computation and application to automatic speech recognition. *Journal of the Acoustical Society of America*, *105*(3), 1925-1932.

Marwala, T. (2001). *Fault identification using neural network and vibration data*. Unpublished doctoral thesis, University of Cambridge, Cambridge, UK.

McFadden, P. D., & Smith, J. D. (1984). Vibration monitoring of rolling element bearings by high frequency resonance technique: A review. *Tribology International*, *77*, 3-10.

Meng, L., Frei, M. G., Osorio, I., Strang G., & Nguyen, T. Q. (2004). Gaussian mixture models of ECoG signal features for improved detection of epileptic seizures. *Medical Engineering & Physics*, *26*(5), 379-393.

Miao Q., & Makis, V. (2007). Condition monitoring and classification of rotating machinery using wavelets and hidden Markov models. *Mechanical System and Signal Processing*, *21*(2), 840-855.

Mohamed, S., Tettey, T., & Marwala, T. (2006). An extension neural network and genetic algorithm for bearing fault classification. In *Proceedings of the International Joint Conference on Neural Network* (pp. 7673-7679).

Nelwamondo, F. V., Marwala, T., & Mahola, U. (2006). Early classifications of bearing faults using hidden Markov models, Gaussian mixture models, Mel-frequency Cepstral coefficients and fractals. *International Journal of Innovative Computing, Information and Control*, *2*(6), 281-1299.

Nikolaou, N. G., & Antoniadis, L. A. (2002). Rolling element bearing fault diagnosis using wavelet packets. *NDT&E International*, *35*(3), 197-205.

Ocak, H., & Loparo, K. A. (2004). Estimation of the running speed and bearing defect frequencies of an induction motor from vibration data. *Mechanical Systems and Signal Processing*, *18*(3), 515-533.

Owsley, L. M. D., Atlas, L. E., & Bernard, G. D. (1997). Self-organizing feature maps and hidden Markov models for machine tool monitoring. *IEEE Transactions on Signal Processing*, *45*(11), 2787-2798.

Peng, Z. K., Tse, P. W., & Chu, F. L. (2005). A comparison study of improved Hilbert–Huang transform and wavelet transform: Application to fault diagnosis for rolling bearing. *Mechanical Systems and Signal Processing*, *19*(5), 974-988.

Prabhakar, S., Mohanty A. R., & Sekhar, A. S. (2002). Application of discrete wavelet transform for detection of ball bearing race faults. *Tribology International*, *35*(12), 793-800.

Purushotham, V., Narayanana, S., & Prasadb, S. A. N. (2005). Multi-fault diagnosis of rolling bearing elements using wavelet analysis and hidden Markov model based fault recognition. *NDT&E International*, *38*(8), 654-664.

Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of IEEE*, *77*(2), 257-286.

Randall, R. B. (2004). Detection and diagnosis of incipient bearing failure in helicopter gearboxes. *Engineering Failure Analysis*, *11*(2), 177-190.

Reynolds, D. A., Quatieri, T. F., & Dunn, R. B. (2000). Speaker verification using adapted Gaussian mixture models. *Digital Signal Processing, 10*(1), 19-41.

Rojas, A., & Nandi, A. K. (2006). Practical scheme for fast detection and classification of rolling-element bearing faults using support vector machines. *Mechanical Systems and Signal Processing*, *20*(7), 1523-1536.

Samanta, B. (2004). Gear fault detection using artificial neural network and vector machines with genetic algorithms. *Mechanical Systems and Signal Processing, 18*(3), 625-644.

Samanta, B., & Al-Bushi, K. R. (2003). Artificial neural network based fault diagnostic of rolling elements bearing using time-domain features. *Mechanical Systems and Signal Processing, 17*(2), 317-328.

Viterbi, A. J. (1967). Error bounds for convolutional codes and an asymptotically optimal decoding algorithm. *IEEE Transaction on Information Theory*, *13*(2), 260-269.

Wang, M. H. (2005). Partial discharge pattern recognition of current transformers using an ENN. *IEEE Transactions on Power Delivery, 20*(3), 1984-1990.

Wang, M. H., & Hung, C. P. (2003). Extension neural network and its application. *Neural Network*, *16*(5-6), 779-784.

Wang, J., Wang, J., & Weng, Y. (2002). Chip design of MFCC extraction for speech recognition, *Integration. The VLSI Journal*, *32*(1-3), 111-131.

Wang, Z., Willett, P., DeAguiar, P. R., & Webster, Y. (2001). Neural network detection of grinding burn from acoustic emission. *International Journal of Machine Tools & Manufacture, 41*(2), 283-309.

Williams, J. H., Davies, A., & Drake, P. R. (1992). *Condition-based maintenance and machine diagnostics.* Chapman & Hall.

Yang, B. S., Han, T., & An, J. L. (2004). ART-KOHONEN neural network for fault diagnosis of rotating machinery. *Mechanical Systems and Signal Processing, 18*, 645-657.

Yang, B. S., Han, T., & Hwang, W. W. (2005). Fault diagnosis of rotating machinery based on multi-class support vector machines. *Journal of Mechanical Science and Technology*, *19*(3), 846-859.

Yang, H., Mathew J., & Ma, L. (2005). Fault diagnosis of rolling element bearings using basis pursuit. *Mechanical Systems and Signal Processing*, *19*(2), 341-356.

Zhang, Y., Li, X. R, & Zhou, K. (1998). A fault detection and diagnosis approach based on hidden Markov chain model. In *Proceedings of the American Control Conference*, Philadelphia (pp. 2012-2016).

ADDITIONAL READINGS

Billington, S. A. (1997). *Sensor and machine condition effects in roller bearing diagnostics.* Unpublished master's thesis, Georgia Institute of Technology, Atlanta.

Brahim-Belhouari, S., Bermak, A., Shi, M., & Chan, P. C. H. (2005). Fast and robust gas identification system using an integrated gas sensor technology and Gaussian mixture models. *IEEE Sensors Journal*, *5*(6), 1433-1444.

Burges, C. C. (1998). A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery, 2,* 955-974.

Carden, E. P., & Fanning, P. (2003). Vibration based condition monitoring: A review. *Structural Health Monitoring*, *3*(4), 355-377.

Cristianini, N., & Shawe-Taylor, J. (2000). *A introduction to support vector machines and other kernel-based learning methods.* Cambridge, UK: Cambridge University Press.

Davies, A. (Ed.). (1998). *Handbook of condition monitoring: techniques and methodology*. London: Chapman & Hall.

Gunn, S. R. (1998). *Support vector machines for classification and regression* (Technical Report). University of Southampton.

Haykin, S. (1999). *Neural networks.* New Jersey: Prentice-Hall.

Heck, L. P., & McClellan, J. H. (1991). Mechanical system monitoring using HMMs. *IEEE International Conference on Acoustics, Speech, and Signal Processing, 3,* 1697-1700.

Kiral, Z., & Karagu¨lle, H. (2003). Simulation and analysis of vibration signals generated by rolling element bearing with defects. *Tribology International*, *36*, 667-678.

Knapp, G. M., & Wang, H. P. (1992). Machine fault classification: A neural network approach. *International Journal of Production Research*, *30*(4), 811-823.

Lou, X. (2000). *Fault detection and diagnosis for rolling element bearing.* Unpublished doctoral thesis, Case Western Reserve University.

Marwala, T. (2000). Damage identification using committee of neural networks. *Journal of Engineering Mechanics, 126*(1), 43-50.

Marwala, T., & Hunt, H. E. M. (1999). Fault identification using finite element models and neural networks. *Mechanical Systems and Signal Processing*, *13*(3), 475-490.

McCormick, A. C., Nandi, A. K., & Jack, L. B. (1998). Digital signal processing algorithms in condition monitoring. *International Journal of COMADEM, 1,* 5-14.

Nabney, I. (2001). *Netlab: Algorithms for pattern recognition*. UK: Springer-Verlag.

Orhan, S., Aktürk, N., & Çelik, V. (2006). Vibration monitoring for defect diagnosis of rolling element bearings as a predictive maintenance tool: Comprehensive case studies. *NDT & E International*, *39*(4), 293-298.

Peng, Z. K., & Chu, F. L. (2004). Application of the wavelet transform in machine condition monitoring and fault diagnostics: A review with bibliography. *Mechanical Systems and Signal Processing*, *18*, 199-221.

Rao, B. K. N. (1996). Handbook of condition monitoring. *Advanced technology.* Oxford: Elsevier.

Sandy, J. (1988). Monitoring and diagnostics for rolling element bearings. *Sound and Vibration*, *22*(6), 16-20.

Tandon, N., & Choudhury, A. (1999). A review of the vibration and acoustic measurement methods for detection of defects in rolling element bearings. *Tribology International*, *32*(8), 469-480.

Theodoridis, S., & Koutroumbas, K. (2006). *Pattern recognition*. New York: Academic Press.

Wang, K. (2003). *Intelligent condition monitoring and diagnosis system*. IOS Press.

Wang, L., & Gaio, R. X. (Eds.). (2006). *Condition monitoring and control for intelligent manufacturing*. UK: Springer-Verlag.

Wang, L., Mostaga, G. M., & Kannatey-Asibu, E. (2002). Hidden Markov model-based tool wear monitoring in turning. *Journal of Manufacturing Scientific Engineering, 124*, 651-658.

Chapter VII
Demand Forecasting of Short Life Span Products:
Issues, Challenges, and Use of Soft Computing Techniques

Narendra S. Chaudhari
Nanyang Technological University, Singapore

Xue-Ming Yuan
Singapore Institute of Manufacturing Technology, Singapore

ABSTRACT

This chapter briefly reviews forecasting features of typical data mining software, and then presents the salient features of SIMForecaster, a forecasting system developed at the Singapore Institute of Manufacturing Technology. SIMForecaster has successfully been used for many important forecasting problems in industry. Demand forecasting of short life span products involves unique issues and challenges that cannot be fully tackled in existing software systems like SIMForecaster. To introduce these problems, we give three case studies for short life span products, and identify the issues and problems for demand forecasting of short life span products. We identify specific soft computing techniques, namely small world theory, memes theory, and neural networks with special structures, such as binary neural networks (BNNs), bidirectional segmented memory (BSM) recurrent neural networks, and long-short-term-memory (LSTM) networks for solving these problems. We suggest that, in addition to these neural network techniques, integrated demand forecasting systems for handling optimization problems involved in short life span products would also need some techniques in evolutionary computing as well as genetic algorithms.

INTRODUCTION

Sophisticated machines like manufacturing robots, aircrafts, and so forth have many electronic components that have shorter life span as compared to their parent system. Such mismatch of lifecycles results in unique demand forecasting problems. During the manufacturing of such sophisticated products, their relatively longer manufacturing time, certification time, and so forth add to unique demand forecasting problems (Foucher, Kennedy, Kelkar, Ranade, Govind, Blake, W., et al. 1998). Existing forecasting tools based on time series tools need large data and are not suitable for short life-span products. Some studies have been reported to develop the models for this problem (Henke & Lai, 1997; Solomon, Sandborn, & Pecht, 2000); however, the problem has increased complexity due to the fact that the short life span products do not have adequate data for the construction of suitable model for forecasting. To enable us to model the decision making and to effectively tackle the data inadequacies, in this chapter, we propose the adaptation of the general framework of soft computing.

Professpr Lofti Zadeh introduced an approach of soft computing in 1965; this approach has since then found a variety of applications in diverse application domains (Zimmermann, 1999) like design of controllers, natural language processing, and so forth. In the following section, we review salient features a typical data mining software as we take SAS Enterprise Miner (Walsh, 2005) as an example to illustrate these features. Further, we give main features of one demand forecasting tool developed by researchers at the Singapore Institute of Manufacturing Technology. This tool, SIMForcaster, mainly uses standard forecasting tools based on statistical forecasting methods like time-series, but enhances them by combining various optimization algorithms (Yuan, 2004a, 2004b). An attractive feature of SIMForecaster is that it has an estimation module to automatically estimate and select the optimization algorithm resulting in better performance. Although SIMForecaster has these advanced features, it has limitations for its use to products like short life span products. To develop and appreciation of these limitations, in the third section we give three case studies of short life span products, with emphasis on demand forecasting. We also point out the problems for the deployment of existing tool like existing softwares (e.g., SIMForecaster) for demand forecasting of short life span products. In the forth sectionl we give the recommendations regarding the adaptation of selected soft computing technologies in *SIMForecaster.* In the last section, we briefly give concluding remarks as well as a short sketch of future trends.

EXISTING TOOLS

Typical Data Mining Software

To highlight the importance of domain knowledge in data mining activity, it is useful to point out a well-known quote by Herb Edelstein (Beck, 1997):

If you got terabytes of data and you are relying on data mining to find interesting things for you, you have lost before you have even begun. …. You really need people who understand what it is they are looking for—and what they can do with it once they find it.

Successful data mining involves a combination of statistical methods, visualization tools, and database techniques. Data mining requires various expertise, of which the first important expertise is the domain knowledge, second expertise is about handling of data, and third expertise is understanding and judicious use of various analytical methods.

There are many data mining software packages available commercially that are useful tools for data mining, and they include forecasting capabilities. As an illustration, we give some features in SAS

Enterprise Miner (available commercially from SAS Institute Inc, Cary, North Carolina, USA). Predictive mining is available in SAS Enterprise Miner in the form of supervised classification, and collection of many model construction tools such as (i) regression, (ii) decision tree, (iii) neural network models, (iv) memory based reasoning, (v) rule induction, and so forth are available as ready-to-use techniques (Walsh, 2005). The package called SAS forecasting technologies includes capabilities for time series analysis and forecasting, econometrics and systems modeling, and financial analysis and reporting; further it includes "what-if" simulation analyses. However, many problems in forecasting need additional models, and hence the team of researchers in Singapore Institute of Manufacturing Technology (SIMTech) has developed a tool called SIMForecaster. We describe main features of SIMForecaster (Yuan, 2004a, 2004b) below.

SIMForecaster: Existing Forecasting Tool

To adapt the manufacturing processes for the uncertain and fluctuating customer demand, we need to predict many issues, including inventory levels to be maintained. There are many aspects of the uncertainty in customer demand; it is fluctuating and uncertain. It is difficult to model and predict how many customer orders need to be served in near future, say on the next day, next week, or next month. Further, one important aspect of the customer order is that, even after the manufacturing company receives the customer order, it may be cancelled due to some contingent events happening to the customer. Thus, it is very difficult to determine how much inventory should be kept for tomorrow's customer demand. If we keep higher inventory, it results in the higher holding cost. If we keep lower inventory, we may not be able to satisfy the customers' demands, resulting in loss of profit and/or poor customer service level. These challenges lead to difficulties in efficiently managing inventory and significantly reducing inventory costs. The main feature of the SIMForecaster is that it combines optimization techniques and statistical forecasting algorithms so it has wider applications.

Integration of Technologies

There are many software packages based on statistical and time series analysis for prediction; however, unlike the existing software packages, SIMForecaster is a powerful algorithm-intensive forecasting tool with simplified user interfaces and optimized forecasting algorithms. There are more than 20 forecasting methods and algorithms developed for SIMForecaster. Importantly, SIMForecaster incorporates optimization techniques into statistical forecasting methods for optimally determining the values of various parameters. By selecting the BestFit algorithm, SIMForecaster automatically searches and evaluates the forecasting methods and algorithms for each individual part/product, and provides the most accurate forecasting results for the part/product. Another unique feature of SIMForecaater is that it can compute the forecasting accuracy based on the user's preference. There are more than 10 accuracy indices for the user to choose. For example, if the user is interested in knowing the excess on hand inventory (EOH) and inventory shortage (STG), the user can select EOH/STG as the accuracy index. SIMForecaster is useful in accurately forecasting customer demands and significantly reducing inventory costs, which caters for any dynamic business environment with uncertain customer demands.

Key Capabilities

Key capabilities of SIMForecaster include:

- **Forecast:** Customer demand, product sales, and material/component purchase.

- **Plan:** Production resources, facilities/personnel allocation, and inventory.
- **Schedule:** Purchase, production, shipment, and delivery.
- **Execute:** Optimized plan and schedule for purchase, production ,and delivery.
- **Track:** Demand patterns, sales patterns, and purchase patterns.
- **Deliver:** On time delivery and order fulfillment.

Some Success Stories

By using SIMForecater, SIMTech team has successfully completed the industry projects with several MNCs, and actively provided consultancy to a number of SMEs and PLEs over the past two years. On average, SIMForcaster was able to deliver up to 94.83% reduction in excess on hand inventory (EOH) and up to 54.67% reduction in inventory shortage (STG). We give these results in Figure 1 and Figure 2 below. It has promis-

Figure 1. Typical observed reduction in excess on hand inventory (EOH) using SIMForecaster

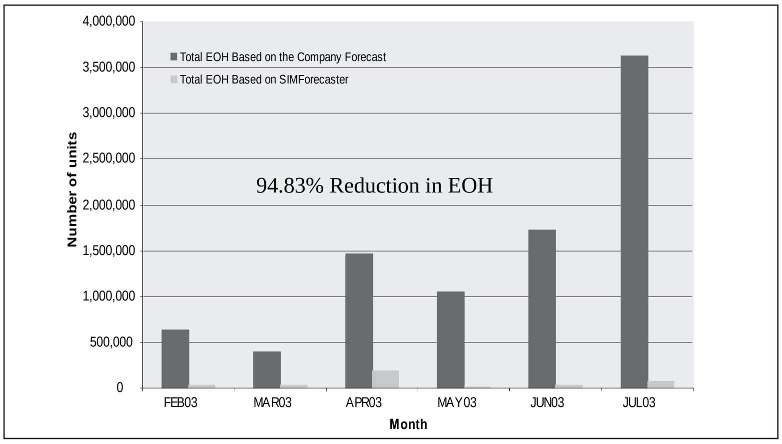
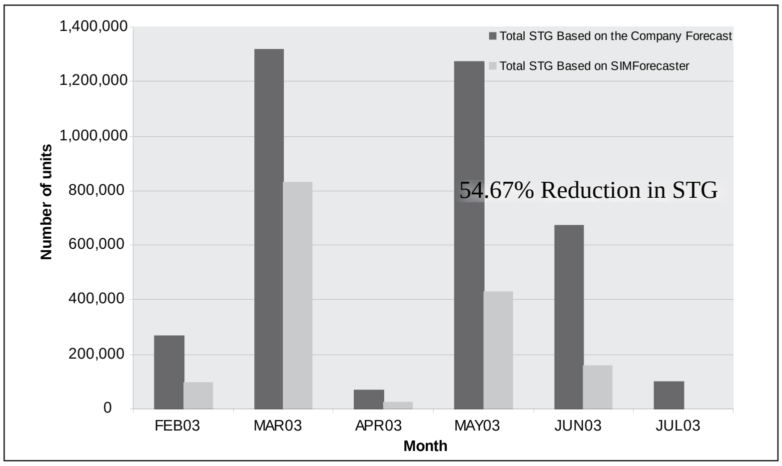


Figure 2. Typical observed reduction in inventory shortage (STG) using SIMForecaster



ing applications in many industry clusters, and it has effectively led to significant reduction in inventory costs.

CASE STUDIES

We now give three case studies to identify the issues and problems in short life span product demand forecasting. Our case studies are drawn from various diverse areas like electronics components industry, perishable goods industry, and entertainment and gaming industry.

Electronics Components Industry

One Singapore-listed company provides electronics manufacturing services to original equipment manufacturers, primarily in the telecommunications, instrumentation and control, and medical and bioscience industries. Its products have application in high-end industries and commercial ventures such as wireless telecommunication, infrastructure equipment, test instrumentation, network-management systems, and medical devices with more than 9,000 low-volume/high-mix line products. The lifecycles of its most products are very short, less than one year; for even some of the products, their lifecycles are only a few months. Because customer demands for its short lifecycle products fluctuate and uncertainties about order quantities and fulfillment times exist, the company faces $14 to $15 million in inventory costs and suffers from low inventory turns. At the same time, great material shortages occur because very few common materials and parts can be used in different products and long supply lead times (30% of materials have more than three months' supply lead times) persist, mainly from Europe.

One of the products is characterized as intensive technique/capital, short lifecycle, and greatly fluctuating in market demand. The customer demand for the product nowadays tends to be more diversified than before, so the company has to deal with cost increase due to high inventory and great customization. It needs to deal with inaccurate forecast, high obsolesce, and large stock-out. More than half of its key parts have to be imported from overseas; therefore, any decision made may influence the whole supply chain. Questions of vital importance to the company are: How to produce accurate forecast for customer demand for the product? How to keep minimal inventory for the product without stock-out? Should the product be assembled in Singapore or at oversea locations near the markets?

In the above case study, import from Europe is needed for many key parts; before enough data can be generated for modeling this behavior through SIMForecaster, the parts become obsolete. The statistical and computational techniques in existing SIMForecaster are unable to develop a model to tackle questions like (i) accurate forecast of customer demand and (ii) minimal inventory. The management decision problem of whether the product should be assembled in Singapore or overseas is another important decision for which existing models in SIMForecaster are inadequate.

Perishable Goods Industry

The major challenge for grocery retailers is dealing with perishable goods like food. Even before the expected date of expiry, around 10-15% of perishable goods require disposal. How to accurately forecast customer demands becomes very crucial in food industry.

One of the leading food companies in the world has its presence in Singapore to offer worldwide brands and customized solutions to independent and chain customers around the world. Recently the company has just started a development of refrigerated prepared food products which they call chill products. These products, such as sandwiches and salads, are distributed through various vending channels, for instance, a cold vending machine or a refrigerator placed in a convenience store or a supermarket. The development of this new product is in the preliminary stage and the company desires to explore the possibilities to expand and grow this particular product category. The current existing product that has been launched in the market is a sandwich product. As has been understood, one of the unique characteristics of this kind of product is its life span characteristic. These products have a very short life span compared to other products that have been developed successfully by this company. The expected life span in this case is only few days, while most of the existing products of the company have a longer life span, say months or even years. Therefore, the development of this kind of product is a new challenge for the company. This new challenge means that the company should be more careful in studying the opportunity in this new business, thus more careful in strategic decisions such as the investment cost or in operational decisions (e.g., the supply chain process design). Those strategic decisions should ensure the achievement of the objective to bring the product into a big success like other products of the company. Considering that forecast is always involved as the underlying factor in making any decision in planning, the need of a "good" forecast is inevitable.

Looking back to the reason of new experience in handling such a short lifecycle type of product, the company wishes to have a good forecasting that could be used to support decision making in the operational level of supply chain, for example, to support the company in a decision of production quantity. This is driven by the fact that the tolerance of excess stocks for this type of product is smaller than for a nonperishable product. Besides the obvious risk of financial lost that is involved due to an overstock condition, a problem related to an environmental issue may be raised. Meanwhile, out of stock condition has never been an option in the company's policy for any type of its product at all. Especially in a new development stage, out of stock would harm the product penetration both to the customers and consumers. As a consequence, a good planning with respect to supply chain operation is highly desired in order to manage this new product development into success.

However, to have good forecasting is not an easy task. First of all, the hard data and information regarding this matter is very limited, even unavailable. This creates a big challenge to forecast. Moreover, the uncertainty that is always involved in forecasting, for instance, people's behavior in refrigerated prepared food consumption, competitors' activity, and price elasticity, increases the difficulties in forecasting these new products.

Entertainment and Gaming Industry

Entertainment products such as DVDs and games have remarkably short lifecycles. Their customer demands usually peak upon product launch and the majority of sales occur within a few weeks following the introduction. Forecasting customer demands for such a product is particularly challenging. Unlike products with longer lifecycles, the planners do not have much data to assess. It leads to difficulties in forecasting such short lifecycle products.

The DVD market has received little attention in academic research compared to the attention given to the theatrical movie market. But DVDs currently bring far greater revenues to studios than theatrical films do and have been growing much faster than box-office revenues. In the U.S. market alone, in 2004, DVDs accounted for $15 billion in sales, whereas the box office revenue totaled only $9 billion. Further, the 2004 sales of DVDs represented a 33% growth over the 2003 sales, while box-office sales were essentially stable. The DVD market has become the most important revenue stream for major studios.

In general, forecasting is particularly difficult in the context of entertainment products such as DVDs and games because such products are unique by nature. Even if many variables such as genre, stars, and production costs could be used for forecasting, there is still substantial variation that is unexplained due to each movie's idiosyncrasies.

RECOMMENDATIONS: USE OF SOFT COMPUTING

The main issue in modeling of demand forecasting of short life span products is the nonavailability of extensive data for construction of model. This means that such forecasting should not be purely left to statistical and population based methods for model construction, and we need to have the models with their ability to deal with "missing" and "inadequate" data to construct the gracefully degraded model. Due to such special problems involved, we need to look for nontraditional and interesting models like, (a) small world theory, and (b) the theory of memes, that are also part of broad area of soft-computing. To adapt these models for demand forecasting of short life span products, we need to combine them with more traditional soft-computing models like neural networks, genetic algorithms, and evolutionary computing techniques. Hence, in this section, we give short description of these techniques.

Neural Networks

The origins of a neural network can be traced in the attempts to model neurons and their interconnection structures in the human brain (Pitts & McCullough, 1947). In last two decades, neural networks have evolved as a specialized field, and many theoretically significant results are now known; further, they have extensively been applied to many diverse areas.

Neural network stores its knowledge in the form of learning within interneuron connection strengths known as synaptic weights. These networks have shown themselves to be adept at solving pattern classification, prediction, and so forth. The most common neural network model is the multilayer perceptron (MLP). A given feed forward neural network can be trained using a learning algorithm by which it adapts to the network's connection weights so as to minimize the error over the pattern training sets. This success is hampered by the difficulties of initially defining the network's structure and training parameters and the other problems caused in the weight space. There are other factors which are to be decided, which include network architecture (e.g., number of hidden layer, number of hidden nodes, initial weights, etc.).

The back-propagation algorithm is a well known method for training a multilayer feed forward network but it follows a gradient descent method which has a tendency to get stuck in a local minimum and thus the global minimum is not reached. The performance of the algorithm depends on the selection of the initial weights and on the parameter used.

The MLP and other neural network models can be trained using a learning algorithm such as (error) back-propagation, steepest descent, least square error, genetic algorithm, evolutionary computation, and so forth. Using one of these algorithms, the weights are determined and the network is said to be trained for a set of data.

Three specialized neural network structures of importance in demand forecasting of short lifecycle products are (i) binary neural networks (BNNs) (Kim & Park, 1995; Wang & Chaudhari, 2004), (ii) bidirectional segmented memory (BSM) recurrent neural networks (RNNs) (Chen & Chaudhari, 2006), and (iii) long-short-term-memory (LSTM) neural networks (Hochreiter & Schmidhuber 1997). Binary neural networks are especially interesting due to the fact that their construction methods do not require us to specify the structure of neural network, and construction methods exist to systematically add more number of neurons for BNN construction. The BSM and LSTM neural networks are recurrent neural network (RNN) structures that are attractive from the point of view of their capabilities of modeling and capturing various interesting interactions involved in forecasting demands related to short lifecycle products.

Kim and Park (1995) have given a method based on the concept of geometric expansion for the construction of binary neural networks (BNNs). This method called expand and truncate learning (ETL). It works for two class problems. Wang and Chaudhari (2006) have proposed a more efficient method named fast covering learning algorithm for BNN construction, employing one hidden layer which works for two class problems. The adaptation of this method for multiclass problems is needed for using these techniques for modeling and extracting important features for demand forecasting of short life span products.

Binary Neural Networks (BNNs)

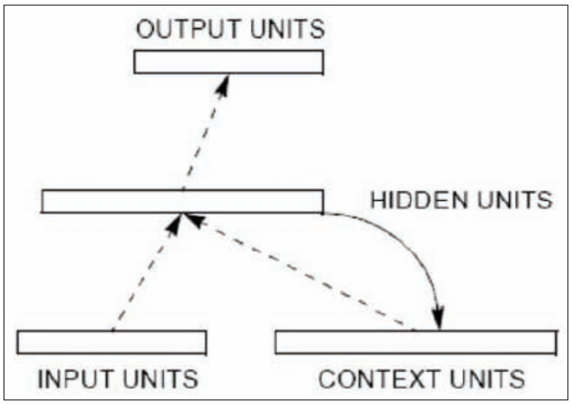
Current training algorithms for nonbinary neural networks are based on the gradient descent technique, which causes iterative computing and long training time. Popular example of neural network construction is error back-propagation (EBP) technique; it uses gradient descent and needs a large number of iterations for adjusting the weights. Such construction methods fix the neural network structure (the number of hidden layers and the number of hidden neurons in each hidden layer) and the information is stored in weights. Hence, for such networks, weights need to have arbitrary precision (value range), and it is difficult to determine a proper network structure in advance which can both guarantee convergence and avoid over fitting. Hence many researchers have adopted constructive learning (Campbell, 1997; Kim & Park, 1995; Kwok &Yeung, 1997). Constructive learning begins with a minimal (or empty) structure, and dramatically increases the network by adding hidden neurons until a satisfactory solution is found. The BNN construction methods begin with an empty hidden layer, and add new hidden neurons when necessary. This constructive learning avoids blind selection of neural network structure.

Input variables are of two types: ordinal variables and nonordinal variables. BNNs are preferred models to solve problems with nonordinal variables, especially with Boolean (true or false) variables. To use BNNs with ordinal variables, we need to use the binary representation of ordinal values and construct a BNN for each bit value. This feature makes use of BNNs difficult. However, for ordinal variables, well-known and popular approaches include fuzzification to reduce the values to a few discrete values; discrete values can nicely be handled by BNNs. For ordinal variables involving real values, many other models like recurrent neural networks (RNNs) structures are useful. For important applications, specialized RNNs have been studied by various researchers, and we give below an account of a few such important special RNNs.

Elman's Simple Recurrent Neural Network

Elman (1991) designed the simple recurrent network (SRN) (Elman's network) for symbolic sequence prediction. Elman's network is a three-layer feed-forward network, but it also has a context layer parallel to the input layer (Figure 3). The hidden unit activation pattern at time t is copied back to the context unit verbatim through weight connections set equal to 1. Then the context activations are processed through a weight matrix and returned to the hidden layer in parallel with the next input pattern at time t+1. The SRN is fed with one symbol each cycle and trained to predict the next symbol. In other words, given a sequence $X_1, X_2, \ldots, X_T$, after the first t symbols $X_1, X_2, \ldots, X_t$ have been presented, the desired output is the t + 1th symbol X_{t+1}. The dynamics of Elman's network is described by a next-state function f :

Figure 3. Elman's network



$$x_i[t] = f_i(x[t-1], u[t]) = g\left(\sum_{j=1}^{n_X} W_{ij}^{xx} x_j[t-1] + \sum_{j=1}^{n_U} W_{ji}^{xu} u_j[t] + W_i^x\right) \quad (1)$$

with u the input vector, x the state vector, n_U the number of input units, and n_X the number of state units.

The recurrent connections permit the Elman's network to encode arbitrary temporal dependencies, which makes it possible for Elman's network to learn languages. The Elman's network has been used for several small natural language problems. Lawrence, Giles, and Fong (2000) examined the inductive inference of a more complex grammar with recurrent neural networks.

Bidirectional Recurrent Neural Networks (BRNNs)

BNNs and many models in machine learning focus on machines with reactive behavior. Recurrent neural networks (RNNs), however, are more general sequence processors inspired by human brains. They have adaptive feedback connections and are in principle as powerful as any computer (Siegelmann & Sontag, 1991). Earlier RNNs, however, could not learn to look far back into the past. To alleviate this limitation, many variants have been proposed. One variant, called bidirectional RNN (BRNN), was originally proposed for speech recognition. It is bidirectional in the sense that current output (say, at instance t) is dependent on dynamics captured by input at the previous instance (t-1), current input instance (say, t), as well as next input instance (say, t+1). This bidirectional dynamics modeled as follows:

$$\begin{aligned} F_t &= \varphi(F_{t-1}, I_t) \\ B_t &= \beta(B_{t+1}, I_t) \\ O_t &= \eta(F_t, B_t, I_t) \end{aligned} \quad (2)$$

The functions $\varphi()$, $\beta()$, and $\eta()$ are realized by the three subnetworks, namely, the forward RNN, the backward RNN, and the feed-forward network. Figure 4 gives this neural network structure and Figure 5 illustrates its working.

Pollastri's Bidirectional Recurrent Neural Networks (PBRNNs)

In his doctoral thesis work, Pollastri (2003) proposed a bidirectional recurrent neural network (PBRNN) for PSS prediction. We abbreviate Pollastri's BRNN as PBRNN. The PBRNN is a noncausal dynamic system that captures both upstream and downstream contextual information (Figure 6).

The contextual information is contained in the vectors F_T and B_T that are defined by the following recurrent bidirectional equations:

$$\begin{aligned} F_t &= \varphi_1(FH_t) = \varphi_1(\varphi_2(F_{t-1}, I_t)) = \varphi(F_{t-1}, I_t) \\ B_t &= \beta_1(BH_t) = \beta_1(\beta_2(B_{t+1}, I_t)) = \beta(B_{t+1}, I_t) \end{aligned} \quad (3)$$

where $\varphi()$ and $\beta()$ are adaptive nonlinear transition functions. They are realized by two recurrent neural networks, namely the left and right subnetworks of PBRNN. The output at residue t is a vector of posterior probabilities of secondary

Figure 4. Structure of BRNN architecture

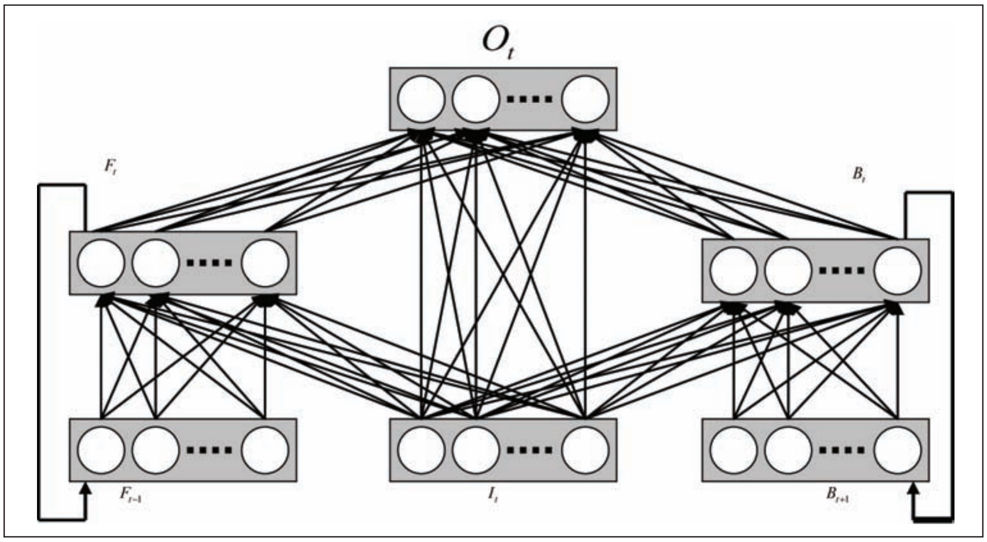
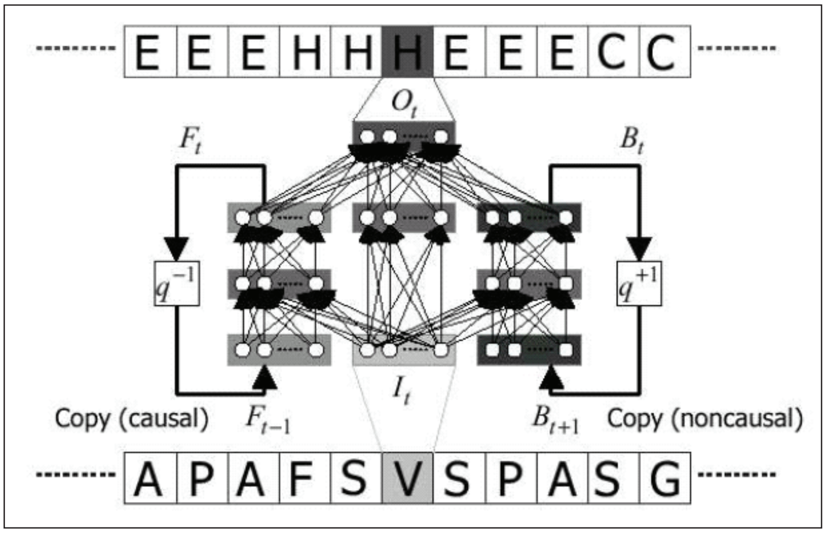


Figure 5. Working of BRNN architecture

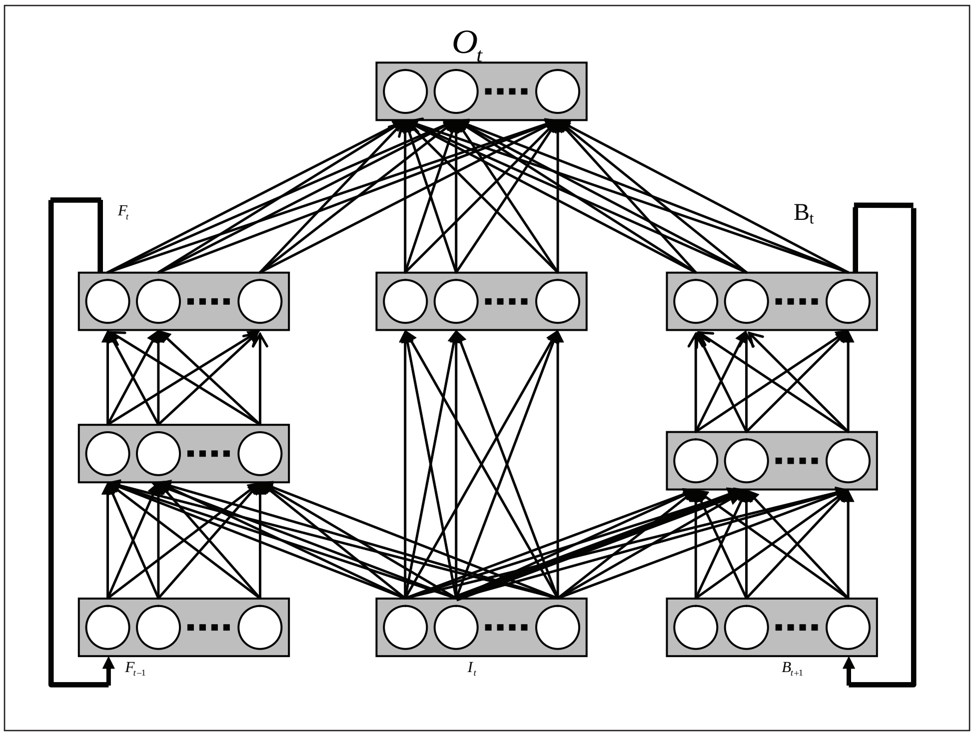


structure classes for sequence position t given the protein's sequence of amino acids. In this model, the output is computed as:

$$O_t = \eta_1(F_t, B_t, H_t) = \eta_1(F_t, B_t, \eta_2(I_t)) \\ = \eta(F_t, B_t, I_t) \tag{4}$$

where $\eta()$ is realized by an MLP, that is, the top subnetwork of PBRNN. Pollastri (2003) reports the successful application of PBRNNs for improving prediction of protein secondary structure.

Figure 6. Pollastri's bidirectional recurrent neural network.



Segmented Memory Recurrent Neural Networks (SMRNNs)

The memory of recurrent neural network is limited. If the network reads in a long sequence at a time, it tends to forget the head part of the sequence. It seems that we as human beings also have similar difficulty memorizing long sequences. When we want to remember some long numbers or long sentences, we tend to do so in segments; for instance, each time we try to memorize 10 digits or subject-predicate-object in the sentence sequentially. We break long sequence into a few segments, memorize each segment first and then cascade them to form the final sequence. To model this process in the domain of RNNs, researchers have introduced segmented memory neural networks (SMRNNs).

The SMRNN has hidden layer H_1 and hidden layer H_2 representing symbol-level state and segment-level state respectively. Both H_1 and H_2 have recurrent connections among themselves. The states of H_1 and H_2 at previous cycle are copied back verbatim through weight connections which are set equal to 1 and stored in context layer S_1 and context layer S_2 respectively. Context layers S_1 and context layer S_2 store contextual information at symbol level and at segment level respectively. Most importantly, we introduce a new attribute, *interval*, into the network, which denotes the length of each segment. The architecture of SMRNN is illustrated in Figure 7.

Given a sequence of T symbols $X_1, X_2, \ldots, X_T$, the network reads one symbol per cycle. The state of H_1 (dashed arrows in Figure 8, labeled with odd numbers 1,3,5,7, and 9) is updated at the

Figure 7. Segmented-memory recurrent neural network with interval=d

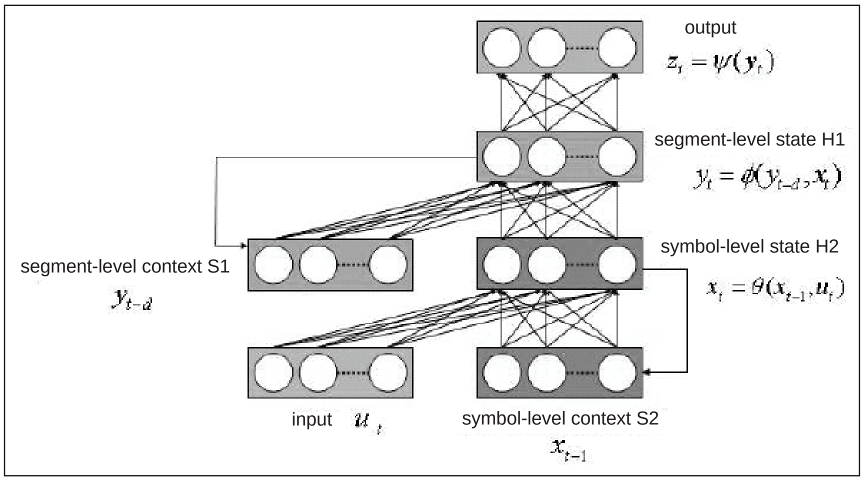
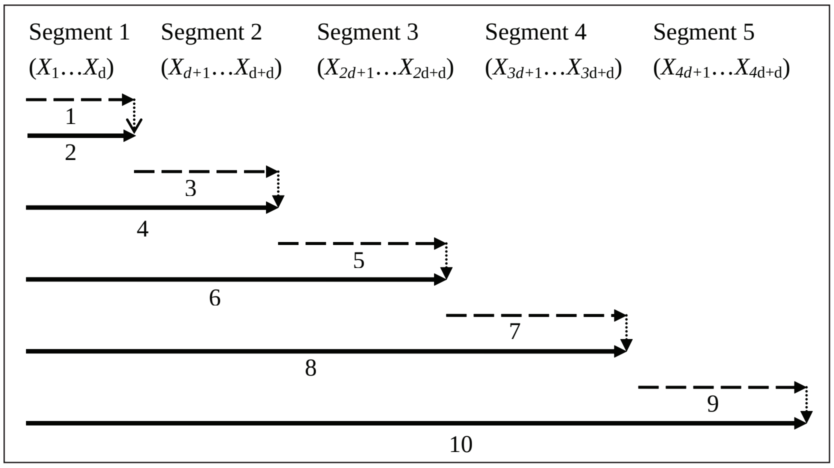


Figure 8. Segmented memory with interval d



coming of each single symbol, while the state of H_2 (continuous arrows in Figure 8, labeled with even numbers 2, 4,6,8, and 10) is updated only after reading an entire segment. As illustrated in Figure 8, the memorizing process of SMRNN is like a ladder: the substrings in parentheses represent segments of equal length *d*; blue arrows indicate the update of contextual information associated to symbols; arrows indicate the update of contextual information associated to segments; and numbers under the arrows indicate the order of memorization.

In order to implement the segmented memory illustrated in Figure 8, we give the dynamics of SMRNN below.

For $k = 1, \ldots, n_Z$,

$$z_k^t = g(\sum_{j=1}^{n_Y} W_{kj}^{zy} y_j^t) \quad (5)$$

For k =, 1, …, n_Y if t mod d = 0, or $t = T$

$$y_k^t = g(\sum_{j=1}^{n_Y} W_{kj}^{yy} y_j^{t-1} + \sum_{i=1}^{n_X} W_{ki}^{yx} x_i^t) \quad (6)$$

otherwise

$$y_k^t = y_k^{t-1} \quad (7)$$

For k = 1, …, n_Y

$$x_k^t = g(\sum_{j=1}^{n_X} W_{kj}^{xx} x_j^{t-1} + \sum_{i=1}^{n_U} W_{ki}^{xu} u_i^t) \quad (8)$$

The variables in the above equations have the following meanings:

- y_j^{t-1} denotes the previous state of hidden layer H_2, which is stored in context layer S_2.
- x_j^{t-1} denotes the previous state of hidden layer H_1, which is stored in context layer S_1.
- n_Z, n_Y, n_X, *and* n_U and denote the numbers of neurons at output layer, hidden layer H_2(context layer S_2), hidden layer H_1(context layer S_1), and input layer respectively.
- W_{kj}^{zy} denotes the connection between the kth neuron at output layer and the jth neuron at hidden layer S_2.
- W_{kj}^{yy} denotes the connection between the kth neuron at hidden layer H_2 and the jth neuron at context layer S_2.
- W_{ki}^{yx} denotes the connection between the kth neuron at hidden layer H_2 and the ith neuron at hidden layer H_1.
- W_{ki}^{xx} denotes the connection between the kth neuron at hidden layer H1 and the jth neuron at context layer S_1.
- W_{ki}^{xu} denotes the connection between the kth neuron at hidden layer H1 and the ith neuron at input layer.
- d denotes the length of interval.

The nonlinearity $g(x)$ used in Equations (5), (6), and (8) can be chosen as follows.

$$g(x) = 1{=}(1 + exp(¡x)) \quad (9)$$

Usually the number of input neurons is equal to the size of the input symbol set. Symbols in the sequence are presented to the input layer with one-hot coding. When a_l (the l-th symbol in the input symbol set) is read, the kth element of the input vector is $I_k = \delta(k,l)$ (where $\delta(k,l)$ is 1 if $k = l$ and 0 otherwise). Normally the synaptic input equals the sum of inputs multiplied by weights. Hence if the inputs are zero, the weights would not be updated. Thus for practical applications involving zero input, it is possible to use a contractive mapping $f(x) = \varepsilon + (1–2\varepsilon)\, x$, with ε a small positive number, to input units. The number of output units can be chosen to suit different applications.

Long Term Dependencies: Limitations of RNNs

The popular learning algorithms for recurrent neural networks are based on computing the gradients of a cost function with respect to the weights of the network. Unfortunately, the gradient based algorithms face an increasingly difficult problem as the duration of the dependencies to be captured increases. In such gradient based learning, error signals flowing backwards in time either tend to (i) blow up, or (ii) decay. Case (i) may cause oscillating weights, while in Case (ii) learning to store contextual information over extended time intervals takes a prohibitive amount of training time or does not work at all.

Bengio, Simard, and Frasconi (1994) explain the essential reason for the difficulty in using gradient descent to learn long-term dependencies in the input/output sequences. The long-term storage of definite bits of information into the state variables is referred to as *information latching.* For many practical applications, the goal of recurrent networks is to robustly latch information.

For simplicity, let us consider an autonomous discrete-time dynamical system:

$$a_t = M(a_{t-1}) \tag{10}$$

where a_t is n-dimensional vector representing the system state at time t, and M is the nonlinear map from state at time t–1 to t.

In order to latch one bit of state information one wants to restrict the system state a_t to a subset S of its domain. To make sure that a_t remains in such a region, the system dynamics can be chosen such that this region is the basin of attraction of an attractor. Bengio, Simard, and Frasconi (1994) use *hyperbolic attractors* to latch bits of information. However, it has been analytically and experimentally proven that the necessary conditions of robust information latching bring a problem of vanishing gradients. In other words, when the robust latching conditions are satisfied, the vanishing gradient effect occurs. The consequence is that short-term influences completely dominate long-term influences, making the task of capturing distant information very hard. Several approaches have been proposed to cope with the long-term dependency problem, some considering alternative optimization algorithms, others trying alternative network architectures (Lin, Horne, Tino, & Giles, 1996).

Bengio, Simard, and Frasconi (1994) investigate several alternative algorithms such as simulated annealing, multigrid random search, time-weighted pseudo-Newton optimization, and discrete error propagation.

Adding connections with time delays to the RNNs helps gradient descent algorithms in learning long-term dependencies. Two typical examples are *nonlinear auto-regressive models with eXogenous* (NARX) recurrent neural networks (Lin, Horne, Tino, & Giles, 1996) and *hierarchical recurrent neural networks* (Hihi & Benjio, 1996). The time-delayed connections in NARX provide a more direct link between the activity of a neuron and the output of the network at a later moment. Experiments indicated that NARX can retain information for two to three times as long as conventional recurrent neural networks.

Hochreiter and Schmidhuber (1997) introduced a recurrent network architecture called *long short-term memory* (LSTM) in 1997. LSTM was designed to overcome the problem of gradient vanishing. By enforcing constant error flow through internal states of special units, LSTM can learn to bridge long time lags.

Long Short-Term Memory (LSTM) Networks

Long short-term memory (LSTM) (Hochreiter & Schmidhuber, 1997) networks are able to solve numerous tasks not solvable by previous learning algorithms for recurrent neural networks (RNNs). Gers, Schimidhuber, and Cummins (2000) identify a weakness of LSTM networks processing continual input streams that are not a priori segmented into subsequences with explicitly marked ends at which the network's internal state could be reset. Without resets, the state may grow indefinitely and eventually cause the network to break down. To alleviate this problem, Gers, Schimidhuber, and Cummins (2000) introduce a novel, adaptive "forget gate" that enables an LSTM cell to learn to reset itself at appropriate times, thus releasing internal resources.

Small World Theory

Some interesting results in the theory of graphs have been viewed by Duncan J. Watts (2002) from the point of view of modeling real-world phenomena. The resulting models are interesting for modeling the dynamics of phenomena involved. In his popular book named *Six Degrees: The Science of Connected Age*, Dunkan J. Watts (2003), the promoter of the small world theory, points out that we have many examples of small worlds. The examples that he points out includes the brain as a network of neurons. Further, he points out that market is a network of interacting buyers and sellers. Networks of markets form bigger market, which we may call as "regional" market. Further, interconnections of markets for different categories of commodities, like perishable commodities, electronic goods, gold and jewellery, real estates and properties, share markets, and so forth form regional or national economies. To take the concept further, global economies includes interactions of national economies. Yet another example from other domains includes organizations as networks of persons. This modeling gives rise to novel problem solving technique for problems arising in the context of many areas like food webs, ecosystems, Internet, and even in the topics in a conversation, words in a language.

Important and interesting contributions to small world theory establish that, after certain conditions are satisfied, the local actions in such networks have global consequences, and the relationship between local and global dynamics depends critically on the network's structure. Klienberg (2000) investigates not just the existence of short paths, but also deals with the problem of finding them. Strogatz (2003) illustrates the use of these and related concepts for creation of order from such phenomena. Successful applications of these results have been demonstrated for many situations, like the spread of infectious disease through a structured population, the evolution of cooperation in game theory, the computational capacity of cellular automata, and the synchronization of coupled phase-oscillators (Strogaz 2003). This leads us to investigate these models further for modeling the problems of short lifecycle demand forecasting as pointed out in the third section using small world theory.

Evolutionary Computing Techniques

In the last decade, evolutionary computation (EC) has become a standard term to denote a very broad group of algorithms and techniques that are based on the principles of natural processes involving biological evolution. Evolutionary algorithms (EAs) are mainly metaheuristic and optimization methods that share some generic concepts borrowed from the natural process of biological evolution. Research in this area has mainly been focused on solving the problems which can be formulated as an exhaustive search over the space of all possible solutions. Using evolutionary computing frameworks, many approaches have been proposed in the last decade. Some approaches for global optimization algorithms include the approaches based on the evolution of species (Davis, Vose, & Whitley, 1999), the immune system (Castro & Timmis, 2002), social behavior of ants (Bonabeau, Dorigou, & Theraulaz, 2000), and so forth. Many variants of ECs are also studied by various researchers. For example, Boettecher and Percus (2001) propose a new optimization algorithm that is based on the principles of natural selection, but it does not follow the basic GA framework for population reproduction. Their approach is one step towards integrating different models like principles of self-organized criticality of Bak and Sneppen (1993) in broad EC framework.

The EC techniques are especially interesting for demand forecasting of short lifecycle products due to their special features of solving optimization problems and self-organization.

Genetic Algorithms

Genetic algorithms (Goldberg, 1989) are based on the Darwinian-type survival of the fittest strategy with sexual reproduction, and Mendel's theory of genetics as the basis of biological inheritance. In these theories, stronger individuals in the population have a higher chance of creating offspring. Each individual in the population represents a potential solution to the problem to be solved. Genetic algorithms do not work with a single point on the problem space but use a set or population of points to conduct a search. This gives GAs the power to search multimodal spaces littered with local optimum points.

GAs can be used to train a multilayer perceptron in which weights form a parameter space. While GAs have the advantage of not getting stuck in local optima, they have other problems. When the search space is very large, then GA methods generally take a long time to converge to good quality solutions. The length of the search is due to the optimal generalization of the training process with no a-priori knowledge about the parameter space.

Theory of Memes

The German biologist Richard Semon, in his work *Die Mnemische Empfindungen in ihren Beziehungen zu den Originalenempfindungen*, formally introduced the concept of Mneme in 1904. In fact, he work is translated in English in 1921 as *The Mneme*. Richard Dawkins (1976) coined the slightly different term "meme" to describe a unit of human cultural evolution similar to the gene in 1970's. The main theme on which Dawkins promoted is that replication also happens in culture. He points out that myths, inventions, language, and political systems as structures are made of memes. The modeling of these structures in terms of memes theory is in the sense of units of ideas, habits, skills, stories, customs, and beliefs that are passed from one person to another by imitation or teaching.

The theory of memes is one of the controversial ways of thinking about cultural evolution. Dawkins argues that the meme is a unit of information residing in the brain is the main theme of Dawkin's book (Dawkin, 1976). He further argues that, for in human cultural evolution, memes is good model for the mutating replicator. This created great debate among sociologists, biologists, and scientists of other disciplines, because Dawkins himself did not provide a sufficient explanation of how the replication of units of information in the brain controls human behavior and ultimately culture. The theory of memes, however, became fashionable (and controversial) since then, and its variations have been investigated by many researchers. Since the 1970's, the theory of memes has variants, like Richard Brodi's (1996) *The New Science of Mind*, and the *Meme Machine* by Susan Blackmore (1999).

Meme has two important properties: (i) meme propagates in the neighborhood, and (ii) meme influences its surroundings.

Memes theory has given rise to new computational models, where the above phenomena are studied in an evolutionary framework by many researchers. This form of search algorithm may be regarded as a marriage between population-based global search and the local improvement made by each of the individuals. It has the potential of exploiting the complementary advantages of EAs (generality, robustness, and global search efficiency) and problem-specific local search (exploiting application-specific problem structure and rapid convergence toward local minima). Such combinations of optimizers are commonly known as hybrid methods. In diverse contexts, hybrid EAs are also commonly known as memetic algorithms (MAs), Baldwinian EAs, Lamarckian EAs, cultural algorithms, or genetic local search. Such methods are demonstrated to converge to high quality solutions more efficiently than their conventional counterparts. In the context of applying this theory to demand forecasting of short lifecycle products, we identify following core issues:

- How often the local search should be applied, that is, local search frequency?
- The issue of selecting appropriate individuals among the EA population that should undergo local search?
- How long should the local learning be executed, that is, local search intensity or local search computational time budget?
- Which local learning procedure or local search or meme to use?

FUTURE RESEARCH TRENDS

In this chapter, we first reviewed salient features of SIMForecaster, the existing forecasting system. SIMForecaster has successfully been deployed for many important forecasting problems in industry. Later, we gave three case studies for short life span products, and identified the issues and problems for demand forecasting of short life span products. Subsequently, we gave a brief discussion about some soft computing techniques, and identified a few specific soft computing techniques, namely, neural networks with special structures, such as binary neural networks (BNNs), bidirectional segmented memory (BSM) recurrent neural networks, and long-short-term-memory (LSTM) networks. In the near future, with careful combination of applications of such models, we expect that there would be demonstrable and useful results for live data sets from industry.

The modeling of demand forecasting of short life span products remained a complex problem and to our knowledge, it has not been researched extensively. Due to its increasing importance, we expect that the innovative techniques would be developed for tackling this problem.

The short life span products do not have adequate data for the construction of suitable model for forecasting and failure prediction within their environments. Many researchers have used innovative approaches to tackle short lifecycle analysis, for example, loop dominance based analysis is given by Kamath and Roy (2005); for uncertain supply chain management, adaptation of graph coloring based heuristics is reported by Lim and Wang (2005).

We argue that the field of demand forecasting of short life span products needs a more powerful set of innovative techniques to tackle this problem successfully. In this chapter, we identified a few nontraditional soft-computing models to tackle this problem, and they include memes theory, small world theory, and specialized neural network structures like binary neural networks and long-short-term-memory (LSTM) neural networks. We briefly gave description of these techniques in this chapter.

We expect that in the near future, this problem would attract the attention of many researchers as well as developers, and demonstrable and useful results would be available for practical (even live) data sets available from industry by demonstrating the synergy of traditional forecasting and soft computing techniques. We expect that such demonstrable results would be possible by combining the soft computing techniques for evolving an important set of tools for tackling this problem.

REFERENCES

Bak, P., & Sneppen, K. (1993). Punctuated equilibrium and criticality in a simple model of evolution. *Physical Review Letters*, *7*(24), 4083-4086.

Beck, A. (1997, October). Herb Edelstein discusses the usefulness of data mining. *DS *: The online Executive Journal for Data Intensive Decision Support, 1*(2). Retrieved September 15, 2006, from http://www.taborcommunications.com/dsstar/971014/ 971014.html

Bengio, Y., Simard, P., & Frasconi, P. (1994). Learning long-term dependencies with gradient descent is difficult. *IEEE Transactions on Neural Networks*, *5*(2), 157-166.

Blackmore, S. (1999).*The Meme machine*. Oxford University Press.

Boettcher, S., & Persus, A.G. (2001). Optimization with extremal dynamics. *Physical Review Letters*, *86*, 5211-5214.

Bonabeau, E., Dorigou, M., & Theraulaz, G. (2000). Inspiration for optimization from social insect behaviour. *Nature*, *406*, 39-42.

Brodie, R. (1996). *Virus of the mind: The new science of Meme*. Seattle: Integral Press.

Campbell, C. (1997). Constructive learning techniques for designing neural network systems. In C.T. Leondes (Ed.), *Neural network systems, techniques and applications* (pp. 1-54). San Diego: Academic Press.

Castro, L.N., & Timmis, J. (2002, May). An artificial immune network for multimodal function optimization. In *Proceedings of the IEEE Congress on Evolutionary Computation (CEC'02)*, Hawaii (Vol. 1, pp. 669-674).

Chen, J., & Chaudhari, N.S. (in press). Cascaded bidirectional recurrent neural networks for protein secondary structure prediction. *IEEE Transactions on Computational Biology and Bioinformatics.*

Davis, L.D, Vose, D.J.K., & Whitley, L.D. (Eds.). (1999). Evolutionary algorithms. *The IMA volumes in mathematics and applications* (Vol. 111). Berlin: Springer Verlag.

Dawkins, R. (1976). *The selfish gene*. Oxford University Press.

Elman, J.L. (1991). Distributed representations, simple recurrent networks, and grammatical structure. *Machine Learning 7*(2/3), 195-226.

Foucher, B., Kennedy, R., Kelkar, N., Ranade, Y., Govind, A., Blake, W., et al. (1998). Why a new parts selection and management program? *IEEE Transactions on Components, Packaging and Manufacturing Technology, Part A*, *21*(2), 375-382.

Gers, F.A., Schimidhuber, J., & Cummins, F. (2000). Learning to forget: Continual prediction with LSTM. *Neural Computation*, *12*, 2451-2471.

Goldberg, M. (1989). *Genetic algorithms for search, optimization and machine learning.* Reading, MA: Addison-Wesley.

Henke, A.L., & Lai, S. (1997). Automated parts obsolescence prediction. In *Proceedings of the DMSMS Conference*, San Antonio, Texas. Retrieved September 25, 2006, from http://smaplab.ri.uah.edu/dmsms/laip.pdf

Hihi, S.E., & Bengio, Y. (1996) Hierarchical recurrent neural networks for long-term dependencies. In M. Perrone, M. Mozer, & D.D. Touretzky (Eds.), *Advances in neural information processing systems* (pp. 493-499). Cambridge, MA: MIT Press.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, *9*(8), 1735-1780.

Kamath, B., & Roy, N.R. (2005). Supply chain structure design for a short lifecycle product: A loop dominance based analysis. In *Proceedings of the 38th Annual Hawaii International Conference on System Sciences (HICSS'05)*, Hawaii (p. 78a-Track 3). IEEE.

Kim, J.H., & Park, S.K. (1995). The geometrical learning of binary neural networks. *IEEE Transactions on Neural Networks*, *6*(1), 237-247.

Kleinberg, J. (2000). The small-world phenomenon: An algorithmic perspective. In *Proceedings of the 32nd Annual ACM Symposium on Theory of Computing*, New York (pp. 163-170). ACM.

Kwok, T.Y., & Yeung D.Y. (1997). Constructive algorithms for structure learning in feedforward neural networks for regression problems. *IEEE Transaction on Neural Networks*, *8*(3), 630-645.

Lawrence, S., Giles, C.L., & Fong, S. (2000). Natural language grammatical inference with recurrent neural networks. *IEEE Transactions on Knowledge and Data Engineering*, *12*(1), 126-140.

Lim, A., & Wang, F. (2005). Robust graph coloring for uncertain supply chain management. In *Proceedings of the 38th Annual Hawaii International Conference on System Sciences (HICSS'05)*, Hawaii, (Track 3). IEEE.

Lin, T., Horne, B.G., Tino, P., & Giles, C.L. (1996). Learning long-term dependencies in NARX recurrent neural networks. *IEEE Trans. on Neural Networks, 7,* 1329-1337.

Pitts, W., & McCullough, W. (1947). How we know universals: The perception of auditory and visual forms. *Bulletin of Mathematical Biophysics, 9*, 127-147.

Pollastri, G. (2003). *A machine learning approach to protein structure prediction*. Unpublished doctoral thesis, University of California, Department of Information and Computer Science.

Siegelmann, H.T., & Sontag, E.D. (1991). Turing computability with neural nets. *Applied Mathematics Letters*, *4*(6), 77-80.

Solomon, R., Sandborn, P., & Pecht, M. (2000, December). Electronic part life cycle concepts and obsolescence forecasting. *IEEE Transactions on Components and Packaging Technologies*, *23*(4), 707-713.

Strogaz, S.H. (2003). *Sync: The emerging science of spontaneous order.* Los Angeles: Hyperion.

Walsh, S. (2005). *Applying data mining Techniques using SAS Enterprise Planner (course notes).* Cary, North Carolina: SAS Institute, Inc.

Wang, D., & Chaudhari, N.S. (2004, March). An approach for construction of Boolean neural networks based on geometrical expansion. *Neurocomputing, 57*(4), 455-461.

Wang, D., & Chaudhari, N.S. (2006, December). A fast modified constructive covering algorithm for binary multi-layer neural networks. *Neurocomputing*, *70*(1-3), 445-461.

Watts, D.J. (2003). *Six degrees: The science of connected age*. New York: W.W. Norton & Co.

Watts, D.J., Dodds, P.S., & Newman, M.E.J. (2002). Identity and search in social networks. *Science*, *296*, 1302-1305.

Yuan, X.M. (2004a). Optimal inventory logistics networks boost competitive edge. *Innovation, 4,* 17-19.

Yuan, X.M. (2004b). Reducing inventory costs through accurate demand forecasting. *Cutting Edge*, *4*, 2.

Zadeh, L.A. (1965). Fuzzy sets. *Information and Control*, *8*(3), 338-353.

Zimmermann, H.-J. (1999). *Practical applications of fuzzy technologies.* Berlin: Springer Verlag.

ADDITIONAL READING

Abraham, B., & Ledolter, J. (1983). *Statistical methods for forecasting.* New York: Wiley.

Abraham, B., & Ledolter, J. (1986). Forecast functions implied by autoregressive integrated moving average models and other related forecast procedures. *International Statistics Review, 54,* 51-66.

Andrews, R.L. (1994). Forecasting performance of structural time series models. *Journal Business and Economic Statistics*, *12*, 129-133.

Aoki, M. (1990). *State space modeling of time series* (2nd ed.). Berlin: Springer-Verlag.

Armstrong, J.S. (2001). *Principles of forecasting: A handbook for researchers and practitioners.* Norwell, MA: Kluwer.

Barnett, G., Kohn, R., & Sheather, S. (1996). Estimation of an autoregressive model using Markov chain Monte Carlo. *Journal Econometrics*, *74*, 237-254.

Berliner, L.M. (1991). Likelihood and Bayesian prediction of chaotic systems. J*ournal of the American Statistics Association*, *86*, 938-952.

Bernado, J.M., & Smith, A.F.M. (1994). *Bayesian theory.* Chichester: Wiley.

Bishop, C.M. (1995). *Neural networks for pattern recognition*. Oxford: Clarendon Press.

Brockwell, P.J., & Davis, R.A. (1996). *Introduction to time series and forecasting.* New York: Springer-Verlag.

Butler, R., & Rothman, E.D. (1980). Predictive intervals based on reuse of the sample. *Journal of the American Statistics Association*, *75*, 881-889.

Caudill, M., & Butler, C. (1994). *Understanding neural networks: Computer explorations*. Cambridge, MA: The MIT Press.

Callen, J.L., Kwan, C.C.Y., Yip, P.C.Y., & Yuan, Y. (1996). Neural network forecasting of quarterly accounting earnings. *International Journal of Forecasting, 12,* 475-482.

Clemen, R.T. (1989). Combining forecasts: A review and annotated bibliography. *International Journal of Forecasting*, *5*, 559-583.

Collopy, F., & Armstrong, J.S. (1992). Rule-based forecasting: Development and validation of an expert systems approach to combining time series extrapolations. *Man. Sci*., *38*, 1394-1414.

DeLurgio, S. A. (1998). *Forecasting principles and applications* (1st ed.). Boston: McGraw-Hill.

Draper, D. (1995). Assessment and propagation of model uncertainty. *Journal of the Royal Statistical Soicety, Series B*, *57,* 45-97.

Faraway, J. (1998). Data splitting strategies for reducing the effect of model selection on inference. *Computer Science and Statistics*, *30,* 332-341.

Faraway, J., & Chatfield, C. (1998). Time series forecasting with neural networks: A comparative study using the airline data. *Journal of Applied Statistics, 47,* 231-250.

Franses, P.H. (1996). *Periodicity and stochastic trends in economic time series.* Oxford University Press.

Frean, M. (1990). The upstart algorithm: A method for constructing and training feedforward neural networks. *Neural Computation, 2,* 198-209.

Funahashi, K. (1989). On the approximation realization of continuous mappings by neural networks. *Neural Networks*, *2*, 183-192.

Geisser, S. (1993). *Predictive inference: An introduction*. New York: Chapman and Hall.

Hill, T., O 'Connor, M., & Remus, W. (1996). Neural network models for time series forecasts. *Man*. *Sci*., *42*, 1082-1092.

Hoptroff, R. (1993). The principles and practice of time series forecasting and business modelling using neural nets. *Neur.Comput.Applic*., *1*, 59-66.

Kuo, C., & Reitsch, A. (1995). Neural networks verus conventional methods of forecasting. *Journal of Business Forecasting*, *14,* 17-22.

Luna, X. (2000). Prediction intervals based on autoregression forecasts. *The Statistician*, *49*, 1-29.

Makridakis, S. (1988). Metaforecasting. *International Journal of Forecasting*, *4*, 467-491.

Meade, N., & Islam, T. (1995). Prediction intervals for growth curve forecasts. *Journal of Forecasting*, *14*, 413-430.

Chapter VIII
Introduction to Data Mining and its Applications to Manufacturing

Jose D. Montero
Management Information System Department, Utah State University, USA

ABSTRACT

This chapter provides a brief introduction to data mining, the data mining process, and its applications to manufacturing. Several examples are provided to illustrate how data mining, a key area of computational intelligence, offers a great promise to manufacturing companies. It also covers a brief overview of data warehousing as a strategic resource for quality improvement and as a major enabler for data mining applications. Although data mining has been used extensively in several industries, in manufacturing its use is more limited and new. The examples published in the literature of using data mining in manufacturing promise a bright future for a broader expansion of data mining and business intelligence in general into manufacturing. The author believes that data mining will become a main stream application in manufacturing and it will enhance the analytical capabilities in the organization beyond what is offered and used today from statistical methods.

INTRODUCTION

There are many driven factors leading manufacturing companies to embrace manufacturing intelligent solutions. Among them are the rigorous customer specifications for designing, developing, and testing of products, the complexity of products, the need for rapid market response, the pressures of global competition, and the strict regulatory requirements. Competitiveness, productivity, and efficiency in the global economy will be affected by how manufacturing companies utilize their computational resources for decision support processes beyond the traditional uses of those resources for just operation processes.

Manufacturing companies are notorious for accumulating mountains of operational data from design, laboratory testing, engineering and research, manufacturing processes, and testing of final products. Some data bases collect operational and historical data of many years. How can manufacturing companies overcome the data barriers presented from these large and in many situations disparate and nonintegrated data bases? How can they leverage the large amounts of data to extract knowledge to support the decision making process? The answers to these questions include the adoption of data integration and data mining. This last subject is the main focus of this chapter.

It has been said as a cliché that "today's high-tech world is drowning in data but is starved for knowledge". That could also be applicable to manufacturing, as well as to other areas of science and engineering. The gap between data and their analysis is growing, and opportunities for extracting valuable knowledge from those data are being lost. This is well expressed by NASA Ames, in the Web site of their Intelligent Data Understanding project.

Many scientists are deluged with data, and the gap between data collection and analysis is growing. Data can be archived for later use, but at a cost (including lost knowledge about the data). There are still LANDSAT data sets from the 70's and 80's that have not been analyzed. (NASA AMES, 2005)

To launch the new millennium, the January/February 2001 issue of *Technology Review*, MIT's magazine of innovation, hailed data mining as one of the top 10 technologies that will change the world in the 21st century (MIT, 2001; Van der Werff, T. J., 2001). These 10 emerging areas of technology have been predicted to have a profound impact on the economy and how we live and work. Data mining, since its most prominent emergence in late 1990s, has revolutionized everything from how companies monitor customers' online purchasing habits and how supermarkets place products in their shelves, to how the federal government practices counterterrorism in the post 9/11 world.

Data mining has been applied successfully in a wide range of businesses in the last decade. It is primarily used in retail, insurance, finance, banking, communications, and direct marketing (Braha, 2001; Han & Kamber, 2001). However, in manufacturing, the application of data mining has started only a few years ago. Factors attributed for slow start of data mining in manufacturing include academic researchers in data mining not being familiar with manufacturing, the majority of researchers in manufacturing are not familiar with data mining, and the restricted access or limitations to publish proprietary and sensitive enterprise data.

The available volumes of data in manufacturing provide many opportunities for knowledge extraction with data mining. Some areas of engineering and manufacturing where data mining has been used with great success, as reported in the literature, are fault diagnosis, process and quality control, process analysis, maintenance interval prediction, production and research process optimization, resource management, and process modeling.

In just a few short years the application of data mining methods in manufacturing has proven very effective in improving both the quality of products and the quality of decision making. It has also provided substantial savings in time and money. One example of this is the Motorola Laboratories' experience in using data mining (Gardener & Bieker, 2000). On the other hand, there is no free lunch. Data mining projects may require a substantial investment. This is particularly true in organizations without well integrated data, or with data of poor quality, or in general with data not ready to be mined. This is summarized by Collier and Held (2000):

While data mining can generate high returns, it requires a substantial investment. Effective data mining requires well-defined objectives, high quality data in a form ready to be mined, and generally some amount of data pre-processing and manipulation. This technology is not a fully automated process. (p. 8)

The main objective of this chapter is to provide a brief overview of what data mining is and some examples of its applications in manufacturing. We will also cover, though in a brief and succinct fashion, data warehousing as a valuable and strategic resource for quality improvement and as a major enabler for data mining applications.

The remaining of this chapter is organized as follows:

- Data mining definitions and the data mining process
- Main categories of data mining tasks
- Examples of applications of data mining in manufacturing
- Data mining and data warehousing
- Future trends
- Conclusions

DATA MINING DEFINITIONS AND THE DATA MINING PROCESS

Goal of Data Mining

In recent decades we have witnessed the exponential growth and explosion of accumulated data in all disciplines of human endeavor. The easiness offered by tools for data collection and the maturity of data base management systems have contributed to the tremendous amounts of data accumulated in data bases and in data warehouses.

Beyond the *known* data, as defined in data bases, there is also *hidden* information of strategic importance for the enterprise. We are now interested and challenged on extracting *novel* and *unsuspected* information from these data. This is a goal of the nascent and rapidly emerging discipline of data mining.

Data Mining Definitions

The simpler definition of data mining *is the extraction of knowledge from large amounts of data*. In this regards, data mining is also referred to as knowledge discovery in data bases (KDD). Others see data mining as the key step in the KDD process. A more formal and working definition of data mining follows:

Data mining is the analysis of (often large) observational data sets to find unsuspected relationships and to summarize the data in novel ways that are both understandable and useful to the data owner. (Hand, Mannila, & Smyth, 2001, p. 1)

The term data mining is really a misnomer, we really do not mine data. Jiawei Han and Micheline Kamber (2001) make the analogy that we do not call sand or rock mining when referring to mining gold from rocks or sand. They say that "data mining should have been more appropriately named *knowledge mining from data*" (p. 5). Therefore using the term knowledge discovery in databases (KDD) could be more appropriate.

It is also useful, particularly in the context of manufacturing companies new to the use of data mining, to define what data mining is not. Complex *data digging* from multiple data sources, even when complex queries are generated, is not data mining. This author has read reports from respected manufacturing companies with a title emphasizing the words *data mining*, even when not a single method from data mining has been used in the analysis. No matter how complex the queries are for accessing those data bases, we are still extracting data that we *know* exist. As we have said before data mining aims to the *discovery* of unknown, novel, and interested knowledge hidden in huge amounts of data.

Data mining is neither a magical wonder. We can not expect to pour data and miraculously find interesting patterns or relationships, or get magical predictions. As Dorian Pyle (1999) warns us:

[Data mining] has been described as 'a method of automatically extracting unexpected hidden patterns from data.' […]As a statement, it makes data mining appear to exist in a world where things happened by themselves. This leads to 'the expectation of magic' from data mining. (p. 12)

For practical purposes a more pragmatic definition is required: "Data mining is a process, not just a series of statistical analyses" (SAS Institute, 2004). Even more, the overall data mining process is inherently an interactive and an iterative process (Fayyad, Piatetsky-Shapiro, Smyth, & Uthurusamy, 1996). I offer therefore, a pragmatic definition that encompasses these elements.

Data mining is an ***interactive** and **iterative** process* that uses a variety of analytical tools to *discover* nontrivial, implicit, previously unknown, and potentially useful and understandable patterns and relationships in data.

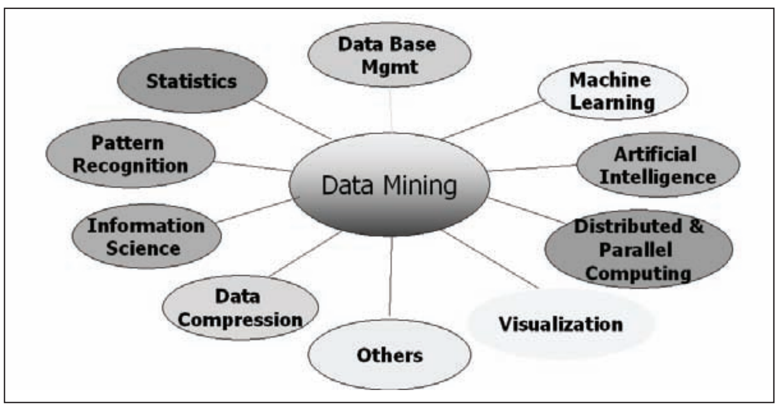
Data mining as an *interactive* and an *iterative* process requires a substantial team effort. In addition to the domain experts and data mining experts, data analysts, and statisticians may also be key players in a successful enterprise data mining team. Collier and Held (2000) hit the right note on this matter:

Data mining assumes a combination of knowledge about the business/production processes and the advanced analytical skills required to ask the right questions and interpret the validity of the answers. Typically data mining is done as a team effort to assemble the necessary skills. (p. 8)

Disciplines Contributing to Data Mining

The term *data mining* existed for years in the statistical data analysis literature before its current definition in computer science. Statistical tools are used extensively in the knowledge discovery process of data mining. Statistics continues to be a major contributor to data mining. Other enablers to the surge of data mining are parallel computer systems, data base management systems, machine

Figure 1. Disciplines contributing to data mining



learning, artificial intelligence, information science, data compression, and visualization. Data mining is in fact a confluence from several disciplines as illustrated in Figure 1.

The Data Mining Process

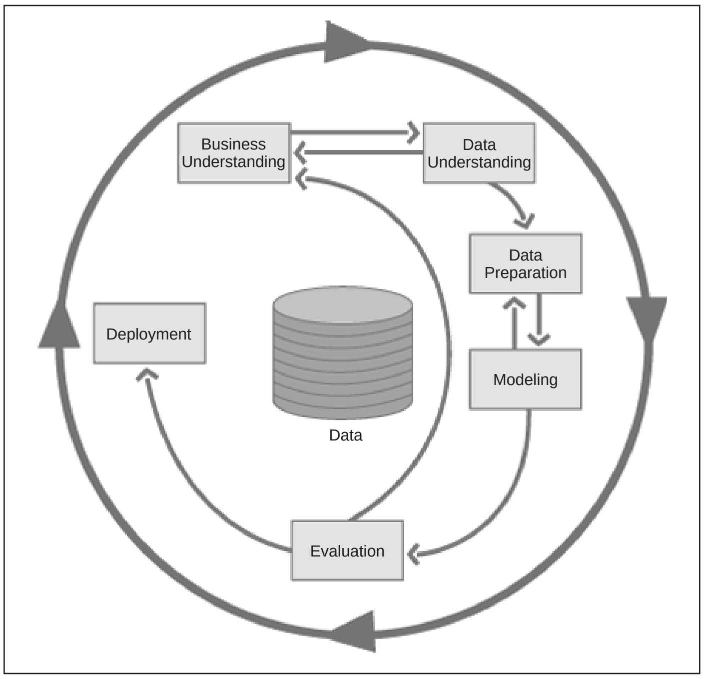
The number and name of the stages in the data mining process varies with the different proposals available. Data mining vendors usually provide their own methodology. SAS, for example, calls their process SEMMA that stands for sample, explore, modify, model and assess. Fayyad et al. (1996) propose the following stages: 1) domain understanding; 2) data selection; 3) cleaning and preprocessing; 4) discovering patterns; 5) interpretation, and 6) reporting and using discovered knowledge. These steps are captured in the cross-industry standard process for data mining (CRISP-DM).

The CRISP-DM is the industry standard methodology for data mining and predictive analytics and it is described in the project Web site *http://www.crisp-dm.org/*. The CRISP-DM project as an industry-and-tool-neutral independent data mining process model is applicable in diverse industry sectors. There is an effort currently underway, at the writing of this chapter, to update this model. Figure 2 illustrates the CRISP-DM model and it reflects the iterative nature of the data mining process. Table 1 summarizes the major tasks at each step of the CRISP-DM process.

Table 1. Major tasks in the data mining process

Step in the DM process	Main tasks
Business understanding	• Assessing and understanding the problem • Defining project objectives, requirements • Translating business or research objective into data mining statements • Creating a project plan.
Data understanding	• Data collection • Describing data and evaluating data quality • Discover first insights into the data or detect interesting subsets to form hypotheses for hidden information.
Data preparation	• Cleaning up data (dealing with bad data, missing values, and inconsistent data) • Integrating data (dealing with data with different formats, inconsistencies, etc.) • Exploring data (identify measures of central tendency and dispersion; deal with outliers, deal with extreme values) • Sampling data • Transforming data (standardizing data, combining variables, eliminating noncontributing variables) • Preparing data in format needed for data mining model
Modeling	• Selecting modeling technique(s) • Generating data for training, cross validation & testing • Building and implementing the model(s) • Typically, there are several techniques for the same data mining problem type.
Evaluation	• Evaluating the model's results
Deployment	• Produce final report • Implement discovered knowledge

Figure 2. The CRISP-DM process model (Source: http://www.crisp-dm.org/)



The Need for Data Integration

Data mining algorithms need data that are ready to be mined. The data understanding and data preparation steps are time consuming. Among the tasks involved we have the data collection from disparate and nonintegrated data bases, the cleaning up and integration of data, and the transformation of data into a format amenable to the data mining model.

The tasks of the data understanding and data preparation steps alone may count for the largest duration of a data mining project. Pyle (1999, p. 11), attributes 75% of the time in the data mining process only to these two steps. Others quote a similar percentage. Gonzalez and Kamrani say that "as much as 80% of KDD is about preparing data" (Gonzalez & Kamrani, 2001, p. 44). Even vendors of data mining solutions, such as SPSS, say that "preparing data often accounts for 80-90% of the typical data mining process" (SPSS, 2004).

This is one of the reasons why we will explore briefly in the fifth section the role of data warehousing as an enabler for data mining. The data in a data warehouse are already integrated, cleaned, and ready for data mining.

Data Mining Data Models

In the previous section we described briefly the CRISP-DM model. This model deals with the identification of the steps and tasks to perform in the iterative process of data mining. Other standards have been proposed to define the input data for the diverse set of analytical data mining algorithms, the parameters used for the analytical methods, and how to store the data and metadata supporting data mining results.

Microsoft has attempted to add a data mining application process interface (API) to its SQL database called OLEDB DM (OLEDB for Data Mining). In this case a data mining model resembles a table in MS-SQL.

The University of Waikato, in New Zealand, has an interesting and very helpful project called Weka: Data Mining Software in Java (*http://www.cs.waikato.ac.nz/ml/weka/)*. Witten and Frank (2005) describe in detail the Weka project and provide many examples on how to use it. Weka is a collection of machine learning algorithms for data mining tasks. The algorithms can either be applied directly to a dataset or called from your own Java code. Weka is open source software issued under the GNU General Public License. Weka uses its own format to model the data and to specify the parameters of the used data mining algorithms. Weka's data format is called Attribute-Relation File Format (ARFF).

The data mining query language (DMQL) adopts an SQL-like syntax. It was proposed by Han and Kamber (2001). Another model is the Java API for data mining, called Java data mining (JDM). The detailed description of this proposed Java API can be found in the Java Community Process Web site, Java Specification Request 73 (JSR-73) (*http://jcp.org/en/jsr/detail?id=73*).

Perhaps the most promising data model, given by the number and importance of its supporters, is the predictive model markup language (PMML) developed by the Data Mining Group (*http://www.dmg.org*). PMML is an XML based language that provides a quick and easy way to define predictive models. The Data Mining Group (DMG) is an independent, vendor led group that has the support from the major providers of data mining solutions. Among the sponsors of the DMG, listed at the writing of this chapter, are IBM, Microsoft, MicroStrategy, National Center for Data Mining—University of Illinois at Chicago, Oracle Corporation, SAS, SPSS, and StatSoft.

MAIN CATEGORIES OF DATA MINING TASKS

This section presents a very high level description of the main categories of tasks targeted by data mining. The number of categories and algorithms available for each task is expanding continually due to the very active academic and industrial research on data mining.

The main categories of data mining tasks, as presented by Han and Kamber (2001), are concept description, association rule mining, classification and prediction, and cluster analysis.

Concept Description

It generates descriptions for characterization and comparisons of the data. "The first and simplest analytical step in data mining is to describe the data - summarize its statistical attributes" (Two Crows Corporation, 1999, p. 5).

Examples of methods: attribute-oriented induction, attribute relevance analysis, and use of visualization methods to present descriptive statistics.

Association Rule Mining

It searches for interesting associations or correlation relationships among a large set of items. The association rules (correlation and causality) are defined in the format:

"X → Y [support, confidence]",
example A → B [50%, 75%]
where Support(X, Y) = Pr(X, Y) (joint probability)
Confidence(X, Y) = Pr(Y|X) = Pr(X, Y)/ Pr(X) (conditional probability).

Examples of algorithms: The Apriori algorithm and FP-growth for finding frequent itemsets.

Classification and Prediction

It deals with the construction of models for classification and prediction. Classification predicts categorical labels while prediction models continuous functions. Since the class labels are known, these methods are also called supervised learning algorithms.

There is an extensive family of proposed data mining algorithms for classification and prediction. It is not possible to describe them here due to the extent of this chapter. Among these methods are multivariable statistical regression methods, decision tree induction methods, Bayesian classifiers, and neural networks.

We will describe briefly these last three methods in a following section, as we will refer to them in the examples provided later. These three families of algorithms come from computational intelligence and their use in data mining is extensive and it is expanding.

Cluster Analysis

Clustering is the process of grouping the data into classes or clusters so that objects within a cluster have a high degree of similarity (Han & Kamber, 2001, p. 335). The main goal of these methods is to maximize intraclass similarity and to minimize interclass similarity. These methods are also called unsupervised learning algorithms.

Examples of algorithms: partitioning methods (k-means, k-medoids methods, and their variations), hierarchical methods (AGNES, DIANA, BIRCH, CURE), density -based methods (DBSCAN, OPTICS, DENCLUE), neural-networks-based methods (competitive learning and self-organizing feature maps).

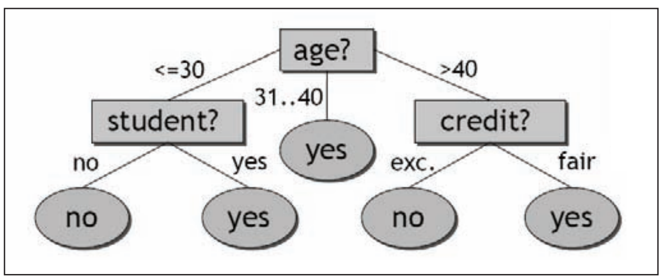
A Brief Overview of Decision Trees, Neural Networks and Bayesian Networks

Decision Trees

Decision trees methods come from machine learning. These are among the most popular classification algorithms in current use in data mining and machine learning. There are several algorithms proposed. Some of them are classification and regression trees (CART), Chi-squared automatic interaction detection (CHAID), C4.5 and C5.0, rain forest, random forest.

Decision tree algorithms develop in a recursive manner a tree-shape model. The top node is called the root node. It is the attribute that best organize the data for the classification in mind. In a recursive matter the subsequent nodes are the root node of the sub-trees. Several methods have been proposed for selecting the attribute that best separates the input data samples data into the individual classes. A common method

Figure 3. A simple decision tree



uses an entropy-based measure known as information gain.

Figure 3 illustrates a simple decision tree. This example is based on the one given by Han and Kamber (2001, p. 284). This decision tree models the concept of *buying a computer*. Each internal node represents a test in an attribute. Each leaf node represents a class: *buy* and *not-buy*. In this simple example, we have three attributes: age, role (student or not a student), and credit history ranking.

The decision tree model found that for the sample data provided as input, the age was the most significant attribute for organizing the data into the two classes: *buy* or *not buy*. The age attribute split the data into three subtrees depending on the age. For each subtree the algorithm finds, on the remaining attributes, the one that best organize the data. For example, if age is greater than 40 then the best attribute for explaining the data is credit. However, if age is less or equal than 30, the best attribute to organize the data is the role. If age is between 31 and 40, that input sample is classified as "buys computer."

One advantage of decision trees is the easiness to derive if-then-else decision rules from the constructed tree. For example, if age <=30 and role = 'student then' "buys computer."

Neural Networks

An artificial neural network (ANN) is a computer model that provides a very rudimentary imitation of their biological counterparts. It is "a massively parallel system of simple processing units that are interconnected via trainable weights" (Yang, 1999, p. 3).

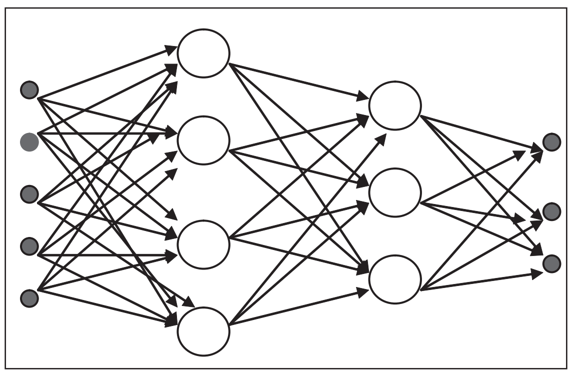
The first proposition of an artificial neural network is attributed to McCulloch and Pitts in 1943. Though, for several decades the limitations of the initial model did not contribute with significant applications of ANNs. The renaissance of neural networks started in mid 1980s with the application in 1986 by Rumelhart, Hinton, and Williams of the back-propagation algorithm.

ANNs are configured for a specific application, such as pattern recognition or data classification, through a learning process. Many different models of neural networks have been proposed (Hagan, Demuth, & Beale, 1996; Haykin, 1994; Yang, 1999). The most common architecture of neural networks is the multilayered feed forward network (Benitez, Castro, & Requena, 1997, p. 1156). In this architecture, the input data flow forward thru one or multiple layers of neurons to generate one or multiple outputs. The intermediate layers are known as hidden layers.

Figure 4 illustrates a feed forward ANN with two layers of neurons. One additional feature that is usually present in the network is the dedication of an input as a bias. The bias serves to alter the net input used as argument to the activation function. In mathematical terms, bias serves to translate the net input provided by along an axis representing net input, thereby altering the location of the activation function in net input space. For simplification of the illustration the bias input vectors are not illustrated in Figure 4.

An artificial neuron is a simple model of its biological counterpart that inspired this computational metaphor. Figure 5 illustrates the common elements of an artificial neuron.

Figure 4. Simple multilayer feed-forward neural network



There are several activation functions proposed. Two popular sigmoidal-shaped functions used as activation functions are the logistic and the hyperbolic tangent (tanh) (Brieley & Batty, 1996, p. 243):

Logistic function $f(x) = 1 / (1 + e^{-x})$
Hyperbolic tangent $\tanh(x) = (e^{x} - e^{-x}) / (e^{x} + e^{-x})$

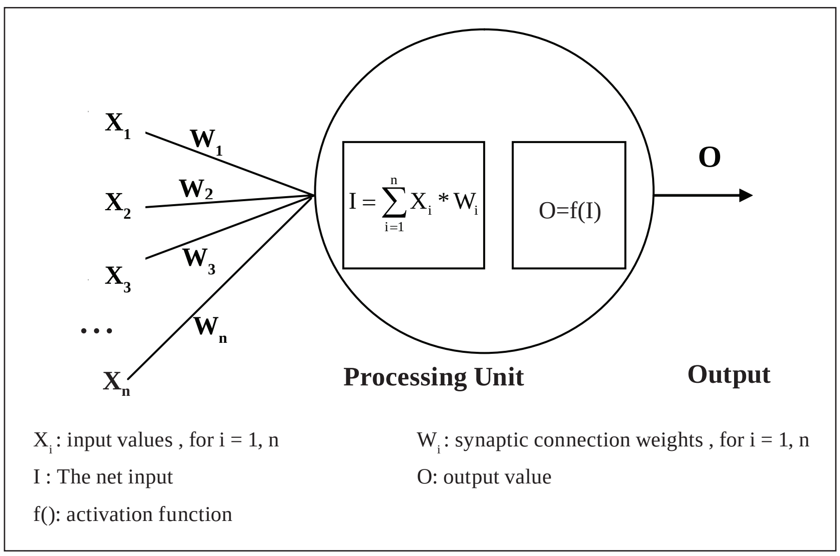
Several advantages and reasons for using neural networks have been proposed, some of them are:

- ANNs have the potential for noise and fault tolerance. A larger number of correct inputs can outweigh a smaller number of incorrect inputs. Similarly, the effect of a few faulty neurons can be nullified by a larger number of properly functioning neurons (Yang, 1999).
- Neural networks can be trained on historical data and used to predict the outcome for new combination of inputs (Brieley & Batty, 1996).
- The parallel nature of ANNs can take advantage of advances in parallel processing.
- Neural networks give statistically better predictive accuracy than the decision trees for many tested problems. (Setiono, 2001, p. 2867).
- Neural networks are universal approximators. They can approximate to any desired degree of accuracy any real-valued continuous function or one with a countable number of discontinuities between two compact sets (Benitez, Castro, & Requena, 1997, p. 1157).

Artificial neural networks have proven to be efficient computing models. As classifiers they have in general been as accurate as or better than Bayesian networks and decision trees. As expressed by Embrechts (1997) "many statisticians admit that the classification power of neural networks is up to par with many statistical methods" (p. 744).

Some drawbacks have also been mentioned on neural networks. They are more computa-

Figure 5. Components of an artificial neuron



tional expensive than decision trees or Bayesian networks. Still, for many businesses, engineering, or scientific applications, the extra cost of computation could be less important when the accuracy is paramount.

Neural networks have been criticized and its use limited as data mining tools due to its "black-box" approach. This is a significant weakness. Without the ability to produce comprehensible decision rules, as with decision trees, it is hard to trust the reliability of neural networks. Jiawei Han says "this factor has motivated research in extracting the knowledge embedded in trained neural networks and in representing that knowledge symbolically" (Han & Kamber 2001, p. 310). Tickle, Andrews, Golea, and Diederich (1998) provide a summary of the advances and challenges of rule mining from neural networks. Even though the challenges for rule extraction from ANNs are big, still there is a bright and promising future:

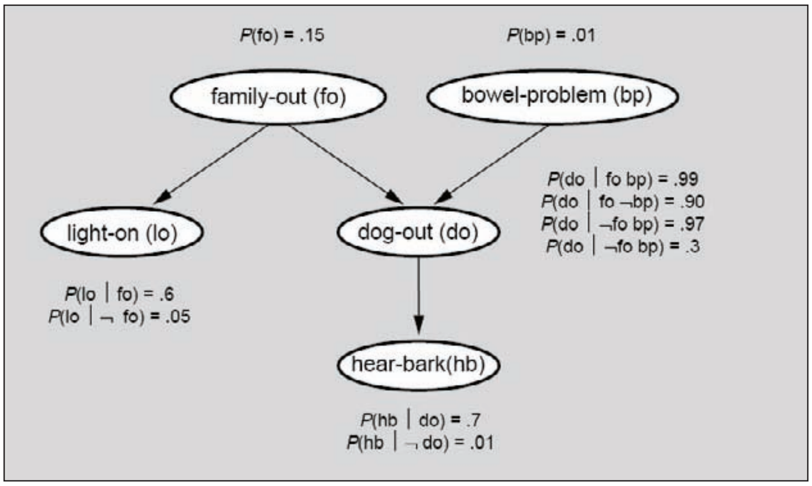
Rule discovery and rule formulation with neural networks are at its infancy, but from what is already reported in the literature there is a good hope that neural network based rule inference systems might actually lead to elegant rule systems that are better, more natural, and more compact than current traditional data mining counter parts such as decision trees and Bayesian networks. (Embrechts, 1997, p. 74)

Bayesian Networks

Bayesian networks are a graphical representation of uncertain knowledge. While many techniques of knowledge discovery rely solely on data, and in the other hand, expert systems rely usually solely on the domain expert's expertise, Bayesian networks have the best of both worlds (Heckerman, 1996, p. 274).

A Bayesian network model uses conditional probabilities for constructing iteratively a network based on the training data. Also known as Bayesian belief networks, knowledge maps, and probabilistic causal networks. Bayesian methods have blossomed after Pearl's introduction of "belief networks" in 1988. Heckerman and others used Bayesian networks on the Pathfinder project 1992. This is a program to diagnose diseases of the lymph node (Heckerman, 1992).

Figure 6. Bayesian network for family out problem (Charniak, 1991)



Some important properties of Bayesian networks are their ability to diagnose multiple simultaneous root causes, the combination of which may have never been anticipated by the experts, and their ability to submit and retract evidence and to recompute its effect in any order during the diagnostic process. There are several examples of applications of Bayesian networks used for diagnostics in several commercial implementations in companies such as Microsoft, General Electric, and Intel.

Charniak (1991) provides a brief and clear introduction to the basics of Bayesian networks. From his introduction we have the illustration in Figure 6. This is a simple Bayesian network to represent *the family is out* scenario. When we return home, states such as *the dog is out* or *the light is on* may be indicators that *the family is out.* The nodes in this network denote states of affairs, and the arcs can be interpreted as causal connections. The conditional probabilities of the different scenarios are provided.

EXAMPLES OF DATA MINING APPLICATIONS IN MANUFACTURING

The list of data mining applications is extensive. Some of the areas where data mining has been used for years are market analysis and management, risk analysis and management, DNA and bio-data analysis, finance planning and asset, resource planning, fraud detection, and detection of outliers. New areas of applications of data mining continue to emerge. Two more recent areas where data mining has been applied with great success in more recent years are in science and in manufacturing.

The available volumes of data in manufacturing provide many opportunities for knowledge extraction with data mining. However, as mentioned earlier, the use of data mining in manufacturing started only a few years ago. Today the research and interest on applying data mining in manufacturing has increased considerably. It is evidenced by many recent publications and well-documented case studies.

Some of the applications of data mining in manufacturing mentioned in recent publications are fault diagnosis, process and quality control, process analysis, maintenance interval prediction, production and research process optimization, resource management, and process modeling.

Fault diagnosis is considered the area where data mining has been applied the most in manufacturing (Buchner, Anand, & Hughes, 1997). Data mining methods have been applied successfully to compliment, facilitate, and to reduce the time and cost on the design of experiments (DOE) for identifying the root causes of problems. The number of DOEs increase exponentially with the number of possible causal input variables considered. In some manufacturing scenarios, the number of variables and their interaction may run in the thousands. The selection of the contributing variables to explain the problem may be really daunting. In addition, the physical experiments and testing could be time consuming and very expensive. This is an area where the feature selection methods available in data mining can prove to be a great help in the identification of root causes of failures (Collier & Held, 2000; Gardener & Bieker, 2000).

In process control, data mining has been successfully used to expand and compliments the use of statistical process control (SPC) methods (Collier & Held, 2000).

In the last two decades the interest in neural networks has increased dramatically. There are many documented applications of them in industrial environments. Japan, since the renaissance of neural networks, has made great in roads into the application of them to their industrial processes (Asakawa & Takagi, 1994).

The number of applications and areas of use of neural networks keeps expanding rapidly. The discovery and application of new and faster learning algorithms, and the availability of more powerful and faster computers, have contributed to the continued growth of applications of neural networks. Referring to the expansion of neural networks, Hagan says:

The applications are expanding because neural networks are good at solving problems, not just in engineering, science and mathematics, but in medicine, business, finance and literature as well. Their application to a wide variety of problems in many fields makes them very attractive. Also, faster computers and faster algorithms have made it possible to use neural networks to solve complex industrial problems that formerly required too much computation. (Hagan, Demuth, & Beale 1996, p. 16)

Neural networks have been used in recent years in important applications in manufacturing. Taylor (1996) presents applications of neural networks in areas such as automotive and aerospace industries. One example is on estimating helicopter strain. Jain (1999) provides more examples of neural networks to industrial applications. The book by Leondes (1998) is also dedicated exclusively to the applications of neural networks to industrial and manufacturing systems.

These examples are evidence of the vigor and strength of neural networks in solving industrial problems. Some specific applications of neural networks in manufacturing mentioned on those references are:

- Predictions of failure in continuous casting (Futjitsu and Nippon Steel)
- Inspection and classification of welding defects (Komatsu and Futjitsu)
- Semiconductor fabrication (Kopin and AT&T)
- Fault detection in semiconductor fabrication (Siemens)
- Modeling pup production (Siemens and Cellulose de Caima, Portugal)
- Modeling the quality of paper in a paper machine (Enso Gutzeit, Finland)
- Electric arc furnace control (Neural Applications and Siemens)

The combination of data mining methods such as self-organizing map (SOM) neural networks and induction decision trees have been applied to solve very complex problems in the semiconductor industry with substantial savings in time and money. The three cases described by Gardener and Bieker (2000) are excellent examples of the effective use of those data mining methods at Motorola Laboratories. The reported return on investment in terms of time saving, customer satisfaction, and cost reductions was substantial.

In other applications of data mining, association rules mining have been proposed for standardizing manufacturing components and processes (Agard & Kusiak, 2004).

These are just few examples to illustrate how data mining can help in solving manufacturing problems dealing with multifactor and nonlinear interactions, intermittent problems, dynamically changing processes, multiple products, and increasing volumes of data. The models generated by data mining algorithms can help in the design of controlled experiments and process changes.

DATA MINING AND DATA WAREHOUSING

In the definition of a strategy for introducing or expansion of the use of data mining in manufacturing, we must consider the role that data warehouses and its relationship to data mining. A common question asked in this regard is: *Is a data warehouse needed for data mining?*

Operational Transaction Needs vs. Analytical Needs

A major obstacle for the analytical and decision-making process in many companies is data accessibility. Data is spread in multiple transactional data bases designed for the specific operational needs of the system that they support. Most IT organizations in manufacturing companies place most of their efforts and resources to the operational needs of the organization and much less in support of the decision-making processes.

Traditional data bases are created mainly to conduct the daily operations of the enterprise and to answer well-defined and usually repetitive questions. They focus mostly on detail data**.** On the other hand, data warehouses main purpose is to support the analytical and decision-making process. They aim to answer queries of aggregated and historical data. They contain both detailed and summarized data.

A data warehouse is not strictly needed for data mining. We can apply data mining methods to data that have been extracted from diverse data sources, not necessarily from a data warehouse. However, as described in the second section, a key and time consuming step in the data mining process is the data gathering and data preprocessing. We mentioned that up to 80% of the data mining process can be spent in the data gathering and data preprocessing. The existence of a data warehouses will facilitate a lot the data mining process. As expressed by Sethi (2001), "although the existence of a data warehouse is not a prerequisite for data mining, in practice the task of data mining *is made a lot easier* by having access to a data warehouse" (p. 4). We have voices such as Inmmon (1996b) expressing that "data warehousing is absolutely essential for effective data mining."

The data in a data warehouse are already integrated, cleaned, and ready for data mining. The existence of historical data in a data warehouse facilitates the data mining algorithms to find patterns and relationships. Data mining methods will benefit also from the presence of summarized data in the data warehouse.

A Definition of a Data Warehouse

"A data warehouse is a *subject-oriented*, *integrated*, *time-variant* and *non volatile* collection of data in support of management's *decision making process*" (Inmmon, 1996a, emphasis added).

Subject-Oriented: A data warehouse is organized around major subjects, such as customer, product, and sales. It provides a simple and concise view around particular subject issues by excluding data that are not useful in the decision support process.

Integrated: A data warehouse is constructed by integrating multiple, heterogeneous data sources. Data cleaning and data integration techniques are applied.

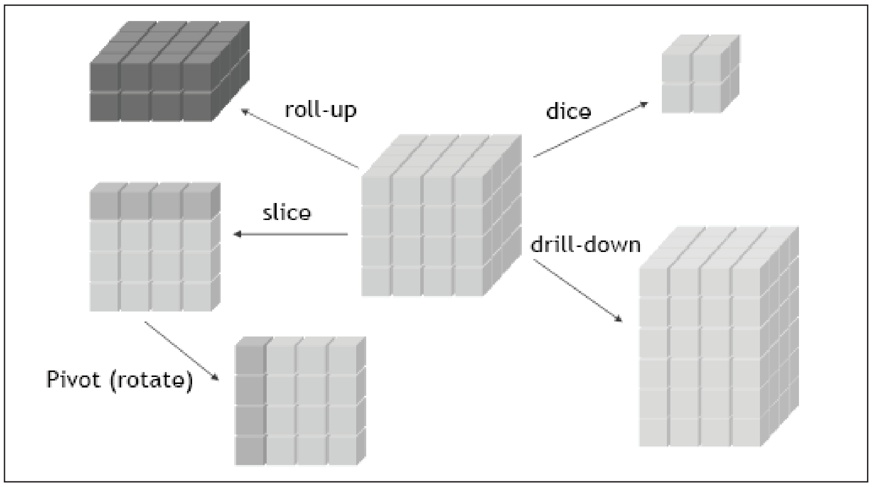
Time Variant: The time horizon for the data warehouse is significantly longer than that of operational systems. While the operational database contains current value data, the data warehouse provides information from a historical perspective (e.g., past 5-10 years).

Non-Volatile: A data warehouse is a physically separate store of data transformed from the operational environment. Operational update of data does not occur in the data warehouse environment.

Data Mining and OLAP

The existence of a data warehouse allows the use of online analytical processing (OLAP) tools that go beyond the traditional reports and queries. OLAP allows users to interactively apply operations to the defined data cubes. These operations are used for summaries, comparisons, analysis, and forecasts. Some of the typical OLAP operations, illustrated in Figure 7 are:

Figure 7. OLAP operations (Source Lim, 2004)



- Slice-and-dice (project and select).
- Roll up (summarize data by climbing up hierarchy or by dimension reduction).
- Drill-down (reverse of roll-up from higher level summary to lower level summary or detailed data, or introducing new dimensions).

Traditional decision support techniques provide descriptive answers to complex queries. They usually assume some explicit knowledge of the *existing* data in the data base or data warehouse, as well as of the factors affecting the decision problem at hand. OLAP operations can help to describe where the problem occurs. Data mining can go beyond in helping to identify which combination of causes are responsible for the problem. As expressed by Oracle (2004), queries, reports, and OLAP usually report on the history of the data:

Traditional business intelligence (BI) tools such as reports, interactive query and reporting, and Online Analytical Processing (OLAP), only report on what has happened in the past. [...] Data mining [...] enables you to discover new insights, segments and associations, make more accurate predictions, find the variables that most influence your business, and in general, extract more information from your data. (p. 4)

The Evolution of Quality Management in Manufacturing

Collier and Held (2000) summarize the evolution of quality management in manufacturing as following:

First Generation

It focuses on controlling individual processes. The quality control approach typically uses statistical control processes (SPC).

Second Generation

It takes the enterprise-wide quality solution approach. Quality management is focused in the whole enterprise. It involves the creation of a manufacturing quality data warehouse to collect historical quality data to address questions of a more predictive nature. "A Data Warehouse will facilitate conducting extensive quantitative analyses and understanding the factors affecting quality" (Klenz & Fulenwider, 1999, p. 12).

From an IT perspective, the quality initiative is now required to build up quality data warehouses, which cover the whole production process including other quality-related data in an analysis-ready form. A quality data warehouse links supplier data, process data, data from other manufacturing plants, and human resource data to address questions such as comparing quality across products, manufacturing lines, or plants, linking warranty problems to internal process data, or predicting product quality before the product reaches the customer. (Collier & Held, 2000, pp. 6-7)

Third Generation

It involves the adoption of analytical modeling and data mining tools. Historical quality data are used to answer questions of a more predictive nature. Data mining methods expand and compliment the use of traditional statistical analysis methods or OLAP operations. These "traditional online transaction (OLTP) and OLAP are insufficient for identifying root causes of field failure" (Collier & Held, 2000, p. 12).

Klenz and Fulenwider (1999) provide a list of the typical computer systems contributing data to a manufacturing quality data warehouse:

Enterprise quality improvement requires collecting information from many departments within an organization, such as production, quality assurance, engineering, customer service, and purchasing. Typically, these groups collect large amounts of data from disparate and disconnected systems. The systems related to quality improvement include: statistical process control (SPC) systems, laboratory information management systems (LIMS), manufacturing execution systems (MES), enterprise resource planning (ERP) systems, manufacturing resource planning (MRP) systems. (p. 3)

FUTURE TRENDS

The Growing Interest in Data Mining

The interest in data mining and the application of powerful algorithms, many of them coming from computational intelligence, will continue in the manufacturing industry. Part of this motivation comes from the need to extract knowledge from the large accumulated operational data sources, typical in manufacturing.

Data mining is emerging as an area of computational intelligence that offers new theories, techniques, and tools for the data-based spaces. It has gained considerable attention among practitioners and researchers. The growing volume of data available in digital form spurs this accelerated interest. (Kusiak, 2000, p. 498)

We have described before, though in a very brief manner, three families of computational intelligence algorithms that have been applied extensively in data mining: neural networks, decision trees, and Bayesian networks. Other areas of computational intelligence, such as fuzzy logic and evolutionary algorithms, have also been used in data mining methods.

Computational intelligence tools will continue to be applied in data mining in a synergetic manner through most of the steps of the knowledge discovery process (Abonyi & Feil, 2005). For example, in the research of methods for training neural networks and for extracting rules from them, we will continue to see the application of evolutionary, genetic algorithms, decision trees, and fussy logic algorithms (Jan & Sun, 1993; Sethi, 2000; Setiono & Leow, 2001)

Data mining will be more an integral part of the research and development (R&D) efforts of manufacturing organizations. Even more, data mining will be thought as R&D initiatives. Some results from research applying data mining will be deployable into business. At the end, data mining methods will become an intrinsic part of the analytical tools used in manufacturing as it is today the use of statistical process control and other statistical analytical methods. Manufacturing companies will follow a similar pattern as the retail counterpart on the use of data mining. Not too long ago, association rule mining as applied to basket analysis was an interesting and perhaps an esoteric concept. Today it is an integral part of the retail industry, and it has been incorporated in most customer relationship management (CRM) products.

The retail, banking, and financial industries have already recognized the importance of quality data in building strong customer relations. They have learned that is not enough to place the emphasis only on obtaining and analyzing the output of sophisticated, high-end analytics systems. For manufacturing companies it is also applicable the old adage, "*Garbage In, Garbage Out*." Some data spread across disparate and nonintegrated systems will push manufacturing companies to invest in data integration and data quality initiatives as well.

We anticipate in manufacturing a more collaborative integration of analytical tools: traditional statistical analysis, online analytical processing (OLAP) provided by data warehousing tools, data mining, and text mining. We have not talked in this chapter about text mining, as some consider it a separate specialized area of data mining. Text mining analyzes unstructured textual data by finding and discovering the patterns and relationships within thousands of documents, such as e-mails, call reports, Web sites, and other information sources.

Data Mining Vendors

As the interest and applications of data mining continue to expand across a growing spectrum of industries, so are the offerings of data mining tool sets by a diverse variety of vendors. Four major groups of data mining tool providers could be identified. The list of vendors presented in each category here is not intended to be exhaustive.

Data Mining Tools Integrated With Data Base Management Systems

Here Oracle, IBM, and Microsoft are in the front line. These tools offer in-database mining platform. IBM has the DB2 Intelligent Miner family of product (*http://www-306.ibm.com/software/data/iminer/*). Oracle's data mining is offered as an option to Oracle Database 10*g* Enterprise Edition (*http://www.oracle.com/technology/products/bi/odm/index.html*). Microsoft offering of data mining is part of the services offered by Microsoft SQL Server 2005 Analysis Services (*http://www.microsoft.com/sql/technologies/dm/default.mspx*).

Data Mining Tools Integrated With Statistical Analysis Tools

We have here SAS with their Enterprise Miner product **(***http://www.sas.com/technologies/analytics/datamining/miner/***),** SPSS Clementine 8.5 (*http://www.spss.com/clementine/*), and StatSoft with Data Miner (*http://www.statsoft.com/products/dataminer.html*).

SAS and SPSS are considered leaders among vendors of data mining tools. The integration of the analytical and predictive methods of data mining with their offering of statistical analysis methods and visualization tools place them in an advantage position over other vendors.

Data Mining Tools Embedded Into Business Applications

Business application vendors are increasingly adding data mining tools embedded into their applications. Examples here are SAP and Seibel.

Specialized Data Mining Vendors

There are also vendors that specialize in the offering of specific tools for data mining. For example, NeuralWare specializes in offering a variety of neural networks solutions (*http://www.neuralware.com/products.jsp*).

In its September 2004 report on the evaluation of data mining tools, Meta Group (now part of Gartner), provided a glimpse of offerings from data mining vendors. It contains the following recommendation:

Data mining tools should be considered part of an organization's IT portfolio for driving new business opportunities and making informed business decisions, due to their predictive analytic capabilities. (Meta Group, 2004, p. 3)

In its 2004 report the Meta Group had SAS and SPSS the clear leaders among data mining vendors. In the most recent report by Gartner of the Magic Quadrant for Customer Data Mining, 1Q06, Herschel (2006), it has again SAS and SPSS as the leading vendors in data mining.

CONCLUSION

Data mining is becoming a key enterprise application and a must-have business process. The set of available methods in data mining is very rich and growing fast. They will complement and expand the analytical capabilities offered by statistical methods and by online analytical processing (OLAP) operations. Traditional online transaction (OLTP) and OLAP are insufficient for identifying root causes of field failure.

Data mining tools should be part of the organization's IT portfolio. They will be of strategic value for generating new business opportunities, for making informed business decisions, for improving quality, and for reducing costs. Data mining promises to generate high returns even though the process is not free.

The availability of very good tool sets of data mining algorithms, offered by a variety of vendors, is good news. The acquisition of a data mining tool set should be considered as an important investment. It will accelerate and facilitate the application of data mining.

Even though a data warehouse is not needed for data mining, a data warehouse could be a valuable and strategic resource for quality improvement and a major enabler for data mining.

Finally, data mining promises high returns but it is not a magical wonder. The *interactive* and *iterative* nature of the process requires a substantial team effort. The team composition could change as the needs of specific research projects may demand.

FUTURE RESEARCH DIRECTIONS

The field of applications of data mining to manufacturing is open with many possibilities for research. On the application side, there are numerous opportunities for exploring the application or adaptation of existing data mining algorithms to specific areas of manufacturing.

There are good examples of the applications of neural networks to manufacturing problems for classification and for clustering applications. We mentioned only some of them and did it in a very succinct way given the scope and the space of this chapter. Related with this area of research is the exploration of new architectures of neural networks and applying them to manufacturing problems. One limitation of the applications of neural networks to manufacturing, as expressed in this chapter, is the "black box" nature of neural networks. For many scientific and engineering problems the precision of the predicting model may not be sufficient. We may need to have an understanding of how the neural network got those results. There is an active research field to create methods for the extraction of decision rules from trained neural networks. This is a challenging research field but one with great promises.

The area of applying Bayesian belief networks to manufacturing is new. There a tremendous potential for the use of Bayesian belief networks for diagnosis and prognosis applications in manufacturing.

Another area for research is how to create an analysis framework where the analytical capabilities provided by data mining be integrated with the statistical process control (SPC) to expand the capabilities of root cause analysis. The great majority of manufacturing companies still rely almost exclusively on SPC which is limited. SPC may point to the existence of a problem but data mining methods can go beyond that to identify the root causes of the problem. The author mentioned briefly in the chapter examples of using decision trees for root cause analysis.

It is important to recognize the great opportunities that exist in manufacturing for the applications of other areas of business intelligence (BI), in addition to data mining. For example, there are good opportunities for devising frameworks and models for data integration/data quality. The future trend in BI is towards what is called now operational BI. This means that in addition to the traditional role of BI with data warehouses and its online analytical processes (OLAP) for analyzing historical data, the new data warehouses are been used for both historical, decision making data with operational data. One area where we will see this expansion in the next five years is for example on the proliferation of dash boards. This is a broad field with many opportunities for new research.

In a summary, the field of data mining, and if we want to expand it to the field of business intelligence, provides new and extensive opportunities for research per se and in particular when those technologies are applied to the large range of tasks involved in manufacturing.

REFERENCES

Abonyi, J., & Feil, B. (2005). Computational intelligence and data mining. *Informatica, 29*, 3-12.

Agard, B., & Kusiak, A. (2004, August 1-4). *Standardization of components, products and processes with data mining.* Paper presented at the International Conference on Production Research Americas 2004, Santiago, Chile.

Asakawa, K., & Takagi, H. (1994, March). Neural networks in Japan. *Communications of the ACM, 37*(3).

Benitez, J.M., Castro, J.L., & Requena, I. (1997, September). Are artificial neural networks black boxes? *IEEE Transactions on Neural Networks, 8*(5), 1156-1164.

Braha, D. (2001). *Data mining for design and manufacturing.* Norwell, MA: Kluwer Academic Publishers.

Brieley, P., & Batty, B. (1996). Data mining with neural networks: An applied example in understanding electricity consumption patterns. In M. A. Bramer (Ed.), *Knowledge discovery and data mining.* London: Short Run Press.

Buchner, A.G., Anand, S.S., & Hughes, J.G. (1997). Data mining in manufacturing environments: Goals, techniques and applications. *Studies in Informatics and Control, 6*(4), 319-328.

Charniak, E. (1991). Bayesian networks without tears. *Artificial Intelligence Magazine, 12*, 50-63. Retrieved November 14, 2005, from http://www.aaai.org/Library/Magazine/Vol12/12-04/Papers/AIMag12-04-007.pdf

Collier, K., & Held, G. (2000, October). *Data mining: Quality in manufacturing data - Best practices approach to the manufacturing industry* (SAS Institute Report). Retrieved December 10, 2004, from http://www.sas.com/apps/whitepapers/whitepaper.jsp

Embrechts, M.J. (1997). Neural networks for data mining. In C.H. Dagli, M. Akay, O. Ersoy, B.R Fernández,, & A. Smith (Eds.), *Intelligent engineering systems through artificial neural networks* (Vol. 7, pp. 741-746). New York: Asme Press.

Fayyad, U.M., Piatetsky-Shapiro, G., Smyth, P., & Uthurusamy, R. (1996). *Advances in knowledge discovery and data mining.* Cambridge, MA: AAAI Press/MIT Press.

Gardener, M., & Bieker, J. (2000, August). Data mining solves though semiconductor manufacturing problems. In *Proceedings of the Sixth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 376-383). New York: ACM Press.

Gonzalez, R., & Kamrani, A. (2001). A survey of methodologies and techniques for data mining and intelligent data discovery. In D. Braha (Ed.), *Data mining for design and manufacturing.* Norwell, MA: Kluwer Academic Publishers.

Hagan, M., Demuth, H., & Beale, M. (1996). *Neural network design*. New York: Brooks Cole.

Han, J., & Kamber, M. (2001). *Data mining: Concepts and techniques.* San Francisco: Academic Press.

Hand, D., Mannila H., & Smyth, P. (2001). *Principles of data mining.* Cambridge, MA: The MIT Press.

Haykin, S. (1994). *Neural networks, a comprehensive foundation*. New York: Macmillan College Publishing Company.

Heckerman, D.B. (1992). An evaluation of the diagnostic accuracy of Pathfinder. *Computer and Biomedical Research*, *25*, 56-74.

Heckerman, D. (1996). Bayesian networks for knowledge discovery. In U.M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, & R.Uthurusamy (Eds.), *Advances in knowledge discovery and data mining.* Cambridge, MA: AAAI Press/MIT.

Herschel, G. (2006, January 18). *Magic quadrant for customer data mining, 1Q06*. (Gartner RAS Core Research Note G00132466). R1659 01232007.

Inmmon, W.H. (1996a). *Building the data warehouse*. New York: John Wiley & Sons.

Inmmon, W.H. (1996b, November). The data warehouse and data mining. *Communications of the ACM*, *39*(11), 49-50.

Jain, L.C. (1999). *Industrial applications of neural networks.* New York: CRC Press.

Jan, J., & Sun, T. (1993). Functional equivalences between radial basis function networks and fuzzy inference systems. *IEEE Transactions on Neural Networks*, *4*, 156-159.

Klenz, B., & Fulenwider, D. (1999). *The quality data warehouse: Solving problems for the enterprise*. Retrieved December 10, 2004, from http://support.sas.com/rnd/app/papers/qualitydw.pdf

Kusiak, A. (2000). *Computational intelligence in design and manufacturing.* New York: John Wiley and Sons.

Leondes, C. T. (1998). Industrial and manufacturing systems. *Neural network systems: Techniques and applications* (Vol. 4). New York: Academic Press.

Lim, S. (2004, Fall). Lectures on data mining: Introduction. *Utah State University*. Retrieved August 13, 2004, from http://otherland.cs.usu.edu/~lim/dm/dm-intro_4.pdf

Meta Group. (2004, September). Data mining tools, METAspectrum evaluation. *Meta group.* Retrieved December 21, 2004, from http://www.metagroup.com/webhost/ONLINE/477658/60.1marketsummary.pdf

McCulloch, W., & Pitts, W. (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics, 5*, 115-133.

MIT. (2001, January/February). The technology review ten. *Technology Review*. Retrieved March 2005, from www.technologyreview.com/magazine/jan01/TR10_fayyad.asp

NASA AMES. (2005). *On intelligent data understanding (IDU) project and knowledge discovery*. Retrieved September 10, 2005, from http://is.arc.nasa.gov/IDU/KD.html

Oracle. (2004, October). *Oracle data mining know more, do more, spend less*. October 2004. Retrieved December 20, 2004, from http://www.oracle.com/technology/products/bi/odm/pdf/bwp_db_odm_10gr1_1004.pdf

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: Networks of plausible inference.* San Mateo, CA: Morgan-Kaufman.

Pyle, D. (1999). *Data preparation for data mining.* San Francisco: Morgan Kaufmann Publishers.

Rumelhart, D., Hinton, G., & Williams, R. (1986). Learning internal representations by error propagation. In D.E. Rumelhart & J.L. McClelland (Eds.), *Parallel processing: Explorations in the microstructure of cognition*. Cambridge, MA: MIT Press.

SAS Institute. (2004). *Finding the solution to data mining: Exploring the features and components of enterprise miner*. Retrieved December 10, 2004, from http://www.sas.com/apps/whitepapers/whitepaper.jsp

Sethi, L. (1990). Entropy nets: From decision trees to neural network. *Proceedings of IEEE, 78*, 1605-1613.

Sethi, I. (2001). Data mining: An introduction. In D. Braha (Ed.), *Data mining for design and manufacturing.* Norwell, MA: Kluwer Academic Publishers.

Setiono, R. (2001). Feedforward neural network construction using cross validation. *Neural Computation, 13*, 2865-2877.

Setiono, R., & Leow, W. (2001). On mapping decision trees and neural networks. *Knowledge Based Systems*, *13*, 95-99.

SPSS. (2004). *Improve IBM® DB2® data mining productivity.* Retrieved December 24, 2004, from http://www.spss.com/clementine/ibmdb2.htm

Taylor, J.G. (1996). *Neural networks and their applications.* New York: John Wiley and Sons.

Tickle, A.B., Andrews, R., Golea, M., & Diederich, J. (1998, November). The truth will come to light: Directions and challenges in extracting the knowledge embedded within trained artificial neural networks. *IEEE Transactions on Neural Networks, 9*(6), 1057-1068

Two Crows Corporation. (1999). *Introduction to data mining and knowledge discovery* (3rd ed.). Retrieved from July 7, 2004, from http://www.twocrows.com/intro-dm.pdf

Van der Werff, T. J. (2001). 10 emerging technologies that will change the world. *Global future report.* Retrieved January, 5, 2005 from http://www.globalfuture.com/mit-trends2001.htm

Witten, I., & Frank, E. (2005). *Data mining: Practical machine learning tools and techniques* (2nd ed.). San Francisco: Morgan Kaufmann.

Yang, J. (1999). *Adaptive agents for information retrieval and data-driven knowledge discovery.* Unpublished doctoral thesis, Iowa State University.

ADDITIONAL READING

The list of references provided in this chapter is extensive and is an excellent source for additional readings. Some of the publications deserve to be highlighted.

On the learning of data mining and the algorithms developed in data mining, Han and Kamber (2001) provide a good introduction to the concepts of and techniques of data mining. Hand, Mannila, and Smyth (2001) deal more on the scientific principles of data mining. Given the wide variety of topics covered by data mining, from computer science and statistics among other sciences, this book focuses on topics that are mode fundamental on data mining.

Witten and Frank (2005) present a good introduction to data mining. They explain a wide variety of machine learning methods. Some explanations are more pedagogically motivated to explain how the basic concepts work. This book is a reference book also on the use of the Waikato Environment for Knowledge Analysis, known as WEKA. This is a comprehensive software resource for data mining written in Java. The software is available via the World Wide Web at http://www.cs.waikato.ac.nz/ml/weka/. WEKA is open source software issued under the GNU General Public License.

The book by Dorian Pyle (1999) on data preparation for data mining deserves very special attention for any one interested in the application of data mining. As expressed in this chapter, the largest percentage of time of a data mining project is data preparation. The tasks of the data understanding and data preparation steps alone may count for the largest duration of a data mining project, even to 80% (some even say up to 90%) of the data mining process may be just for these two tasks. Dorian Pyle presents a treatise on techniques for these areas that is both practical and technical.

For those of you looking for reading on more detailed examples of applications of data mining to manufacturing, three publications listed in the references are full with exciting examples. These books are written by Kusiak (2000), Braha (2001), and Leondes (1998). Leondes's book is dedicated exclusively to the applications of neural networks to industrial and manufacturing systems. For applications of data mining in other areas, Bramer's work (1999) is a good source.

Other readings may be according to the interest of the researcher. For example, the field of neural networks is per se an extensive area of artificial intelligence. Haykin (1994) provides a comprehensive foundation on neural networks. The following are suggested readings on neural networks.

Andrews, R., Diederich, J., & Tickle, A. (1995). A survey and critique of techniques for extracting rules from trained artificial neural networks. *Knowledge Based Systems, 8*, 373-389.

Gibb, J., & Hamey, L. (1995). *A comparison of back propagation implementations* (Tech. Rep. C/TR95-06). Macquarie University, Sydney Australia, Department of Computing.

Hagan, M., Demuth, H., & Beale, M. (1996). *Neural network design*. New York: Brooks Cole.

Jacobsson, H. (2004). *Rule extraction from recurrent neural networks: A taxonomy and review* (Tech. Rep: HS-IKI-TR-04-002). University of Skovde, Sweden, School of Humanities and Informatics.

Kirpatrick, S., C.D., Gelatt, Jr., & Vecchi, M.P. (1983). Optimization by simulated annealing. *Science, 220*, 671-680.

Kolman, E., & Margaliot, M. (2004). Are artificial neural networks white boxes? *Tel Aviv University, Tel Aviv, Israel, Department of Electrical Engineering Systems.* Retrieved October 28, 2004, from www.eng.tau.ac.il/~michaelm/eyal.ps

Rumelhart, D., Hinton, G., & Williams, R. (1986). Learning internal representations by error propagation. In D.E. Rumelhart & J.L. McClelland (Eds.), *Parallel processing: Explorations in the microstructure of cognition*. Cambridge, MA: MIT Press.

Taylor, J.G. (1996). *Neural networks and their applications*. New York: John Wiley and Sons.

Tickle, A.B., Andrews, R., Golea, M., & Diederich, J. (1998, November). The truth will come to light: directions and challenges in extracting the knowledge embedded within trained artificial neural networks. *IEEE Transactions on Neural Networks*, *9*(6), 1057-1068.

Werbos, P. (1974). *Beyond regression: New tools for prediction and analysis in behavioral sciences*. Unpublished doctoral thesis, Harvard University, Cambridge, MA.

Witten, I. H., & Frank, E. (2000). *Data mining: Practical machine learning tools and techniques with Java implementations*. San Francisco: Morgan Kaufmann Publishers.

Yang, J. (1999). *Adaptive agents for information retrieval and data-driven knowledge discovery*. Unpublished doctoral thesis, Iowa State University.

Chapter IX
Evolutionary Computing in Engineering Design

Rajkumar Roy
Cranfield University, UK

Ashutosh Tiwari
Cranfield University, UK

Yoseph Tafasse Azene
Cranfield University, UK

Gokop Goteng
Cranfield University, UK

ABSTRACT

This chapter presents an overview of the application of evolutionary computing for engineering design. An optimal design may be defined as the one that most economically meets its performance requirements. Optimisation and search methods can assist the designer at all stages of the design process. The past decade has seen a rapid growth of interest in stochastic search algorithms, particularly those inspired by natural processes in physics and biology. Impressive results have been demonstrated on complex practical optimisation of several schools of evolutionary computation. Evolutionary computing unlike conventional technique, have the robustness for producing variety of optimal solutions in a single simulation run, giving wider options for engineering design practitioners to choose from. Despite limitations, the act of finding the optimal solution for optimisation problems has shown a substantial improvement in terms of reducing optimisation process time and cost as well as increasing accuracy. The chapter aims to provide an overview of the application of evolutionary computing techniques for engineering design optimisation and the rational behind why industries and researchers are in favor of using it. It also presents the techniques application trend rise in the past decade.

INTRODUCTION

The use of evolutionary computing (EC) to solve many optimisation problems in engineering design is becoming more and more popular among computational and data intensive scientific and engineering application areas. Examples of such computational and data intensive applications are engineering design optimisation, bioinformatics, pharmaceutical and particle physics simulations, and many others. Additionally, since EC techniques mimic nature through natural selection, knowledge-driven problem solving environments (PSEs) are ideal frameworks for optimisation problems using EC.

This chapter will focus on how evolutionary computing techniques can be used for engineering design within a large range of engineering disciplines.

The concept of design was born the first time an individual created an object to serve human needs (Hernandez & Fontan, 2002). Today design is still the ultimate expression of the art and science of engineering. From the early days of engineering, the goal has been to improve design so as to achieve the best way of satisfying the original need, within the available needs.

The design process can be described in many ways; however, there are certain elements in the process that any description must contain: recognition of need, an act of creation, and a selection of alternatives. Traditionally, the selection of the "best" alternative is the phase of design optimisation (Chinyere, 2000). Optimisation is the act of obtaining the best result under given circumstances (Jun, Tischler, & Venkayya, 2002). In design, construction, and maintenance of any engineering system, engineers have to take many technological and managerial decisions at several stages. The ultimate goal of all such decisions is either to minimise the effort required or to maximise the desired benefit. Since the effort required or the benefit desired in any practical situation can be expressed as a function of certain decision variables, *optimisation* can be defined as the process of finding the conditions that give the maximum or minimum value of a function (Singiresu, Xiong, & Ying, 2000).Today, however, the complexity of this process is alleviated by the introduction of *artificial intelligence* and advancement of computing so as the availability of more powerful optimisation techniques. It also gives the opportunity to designers to choose a technique that is relevant to the particular design problems in finding the "best" alternative.

TAXONOMY OF ENGINEERING DESIGN OPTIMISATION

It is imperative to understand the organisational structure or *taxonomy of engineering design optimisation*. This orderly classification allows engineers to understand design optimisation problems according to certain characteristics and problem solving philosophy and facilitates organisation and reuse of knowledge in the design process. It also enhances the representation of knowledge and capturing of the reasoning schemes behind designs. Taxonomies help engineers to compare different design methods and tools and come up with suggestions on how best to use computer aided systems in designs. Ullman and D'Ambrosio (1995) propose four classifications starting with structure of engineering design optimisation, problem focus, and range of independence and level of support. Structure of engineering design optimisation consists of four sub classes, namely, decision space, which consists of problem completeness, abstraction level, and determinism, preference model which consists of objective function, preference model which consists of consistency and completeness of engineering design, and belief model which is concerned with dimension and belief completeness of the design.

Tiwari (2001) proposes a number of classification schemes for engineering design optimisation problems. A summary of these classification schemes is given in Table 1. These classifications enable the categorisation of problems based on their prominent features. This facilitates the choice of a suitable algorithm for a given engineering design optimisation problem. The classification schemes proposed by Tiwari are described below.

Based on number of variables: The engineering design optimisation problems can be classified as single and multidimensional based on the number of variables involved in the problem.

Based on existence of constraints: An engineering design optimisation problem can be classified as constrained or unconstrained depending on whether constraints exist in the problem.

Based on number of objective functions: Depending upon the number of objective functions in the engineering design optimisation problem, it can be classified as single-objective and multiobjective.

Based on nature of objective functions: The objective functions involved in an engineering design optimisation problem may be either *quantitative or qualitative* in nature. Qualitative objective functions involve issues like manufacturability and designers' special preferences. Based on the nature of objective functions, the optimisation problems can be classified as quantitative, qualitative, or hybrid.

Based on separability of functions: A function is said to be separable if it can be expressed as the sum of single-variable functions. An alternative definition of separability relaxes the above definition to include decomposition into functions that involve groups of variables rather than just a single variable. Inseparability manifests itself as cross-product terms, and makes the effect of a variable on the function dependent on the values of other variables in the function. The engineering design optimisation problems can be classified as separable and inseparable based on the separability of objective functions. Inseparability causes difficulties for an optimisation algorithm by requiring it to update all decision variables in a unique way in order to converge to an optimum solution.

Based on dependence among variables: Variable dependence occurs when the variables are functions of each other, and hence cannot be varied independently. Here, the change in one variable has an impact on the value of the other. This causes additional problems for an optimisation algorithm due to the requirement that all dependency relationships need to be satisfied while searching for an optimum solution. This has an effect of constraining the search space.

Based on nature of search space: The nature of search space also defines an important classification of engineering design optimisation problems. Based on this, the two categories that are identified are known search space and unknown search space optimisation problems. The real-life optimisation problems in which the designers lack prior knowledge about the shape of search space and about the location and performance of optimal points are classified as unknown search space optimisation problems. As opposed to this, most theoretical problems (test cases), being lab-designed, are classified as known search space optimisation problems. The nature of search space also classifies the engineering design optimisation problems as unimodal and multimodal based on the number of optimal solutions that the problem has.

Based on nature of equations involved: This classification is based on the nature of expressions that represent the *objective functions* in the optimisation problem. According to this classification, the engineering design optimisation problems can be classified as linear, nonlinear, geometric, and quadratic. Based on this criterion, the engineering design optimisation problems can also be classified as continuous and discontinuous depending on whether the equations involved in the problem have any discontinuities.

Based on nature of design variables: Based on the nature of design variables, the engineering design optimisation problems can be classified as static and dynamic. In parameter or static optimisation problems, the design variables are independent of each other whereas in trajectory or dynamic optimisation problems, the design variables are all continuous functions of some other variable(s). Another perspective of this classification is provided by Schwefel (1995) based on time-dependence of the optimisation problems.

Based on permissible values of design variables: Depending on the values permitted for design variables, the engineering design optimisation problems can be classified as integer-valued, real-valued, and hybrid (that involve both integer and real variables).

EVOLUTIONARY COMPUTING

Evolutionary computing uses techniques such as *genetic algorithms* (GAs), genetic programming (GP), evolutionary programming (EP), and many others. These techniques work on the principles of natural selection based on Darwin's theory of evolution. This principle works on the composition of genetic traits called chromosomes, in which

Table 1. Classification schemes for engineering design optimisation problems

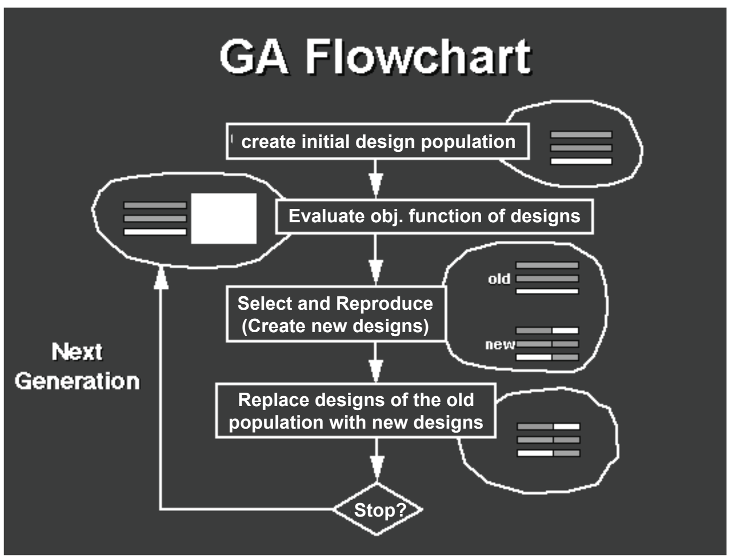
Classification Schemes	Categories
Based on Number of Parameters	• Single-dimensional • Multidimensional
Based on Existence of Constraints	• Constrained • Unconstrained
Based on Number of Objective Functions	• Single-objective • Multiobjective
Based on Nature of Objective Functions	• Quantitative • Qualitative • Hybrid
Based on Separability of Functions (for Quantitative and Hybrid Problems)	• Separable • Inseparable
Based on Dependence among Variables	• Independent-variable • Dependent-variable
Based on Nature of Search Space	• Known Search Space • Unknown Search Space
	• Unimodal • Multimodal
Based on Nature of Equations Involved (for quantitative and hybrid problems)	• Linear • Nonlinear • Geometric • Quadratic
	• Continuous • Discontinuous
Based on Nature of Design Variables	• Parameter or Static • Trajectory or Dynamic
Based on Permissible Values of Design Variables	• Integer-valued • Real-valued • Hybrid

successive operations through crossover or mutation give rise to better performing off-springs (population) due to successive refinement of these hereditary traits. In the same way, evolutionary computing techniques mimic this phenomenon of natural selection to improve upon classical methods of optimisation (Deb, 2001). Such as preference-based and generating methods that work well only on single-objective optimisation problems. With evolutionary computing techniques, multiobjective optimisation problems can be solved, producing multiple optimal solutions in a single simulation run. This is not possible with classical optimisation methods. However, evolutionary computation can be computationally expensive for complex optimisation problems.

Evolutionary computing techniques for multidisciplinary design optimisation which involves multiobjective functions may require the collaboration of distributed experts located in different geographical locations around the world. Computational steering allows each expert to make changes to particular parameters in the course of the optimisation to observe interesting patterns in the optimal solutions. Visualisation models located at the distributed sites enhances the understanding and decision making during this process. Visualisation systems can display the graphical representation of the fitness functions and convergence points during optimisation. Interrupting the optimisation process for GA to observe patterns in the solutions using EC algorithms can help experts to pinpoint which parameter or parameters have the greater impact on the process. Advanced visualisation capabilities integrated with simulation based EC algorithms and design processes provide not only valuable design insight and design directions but also facilitate collaboration in design (Kodiyalam, Yang, & Gu, 2003). Close designer interaction with the design process supports exploration involving off-line processing and visualisation of initial results which leads to a redefinition of the solution space by dynamic evolutionary computing techniques (Parmee et al., 2002).

Knowledge based engineering (KBE) systems create intelligent master model (IMM). This IMM contains the 'what,' the 'why,' and the 'how' of a

Figure 1. Genetic algorithm (GA) flowchart



design using EC algorithms to produce a range of global optimal solutions (Rohl, Kolonay, Irani, Sobolewski, Kao, & Bailey, 2000).Evolutionary programming, which was introduced by Fogel, is an attempt to create *artificial intelligence* (AI) using finite state machines (FSM) to predict future events on the basis of previous behaviours of elements during successive transitions (Back, Hammel, & Schwefel, 1997). FSM is an abstract machine which converts input symbols into sequence of output results depending on some finite set of states and finite state of transition rules. Other evolutionary algorithms such as GA and evolutionary strategies work in similar ways that may need knowledge based framework for intelligent application of evolutionary computing used in various real world problems and experimental simulations of random and optimised stimulus sequences over a range of design parameters (Wager & Nichols, 2003). Figure 1 shows a simple diagram of GA process.

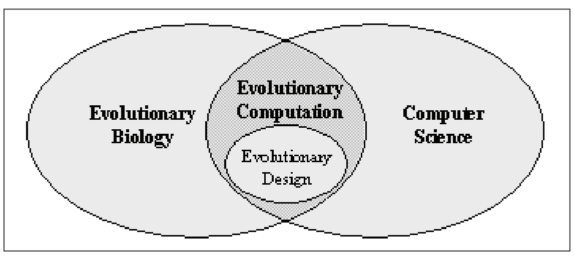
In the natural world, evolution has created an unimaginable diverse range of design, having much greater complexity than mankind could ever hope to achieve. Inspired by this, researches have started using the EC technique that uses the principles of evolution to guide the optimisation process. There are a number of benefits of evolutionary-based optimisation that justify the effort invested in this area. The most significant advantage lies in the gain of flexibility and adaptability to the task at hand. Unlike its classical counter part, where the technique use a point by point deterministic procedure for approaching the optimum solution and maintain a single best solution found so far, an evolutionary algorithm maintains a population of candidate solutions. Only one of these is 'best,' but the other members of the population are 'sample points' in other regions of the search space, where a better solution may later be found. The use of a population of solutions helps the evolutionary algorithm avoid becoming trapped at a local optimum, when an even better optimum may be found outside the vicinity of the current solution. Considering the complexity of real world engineering design optimisation problems, that are multi in nature, and the shortcomings that the classical optimisation technique has in dealing with this problems have led to the growth of research in the field of evolutionary techniques (Oduguwa & Roy, 2006).

OVERVIEW OF EVOLUTIONARY COMPUTING APPLICATIONS IN ENGINEERING DESIGN

The use of evolutionary computation to generate designs has taken place in many different disciplines over the last ten or fifteen years. Designers have optimised selected parts of their designs using evolution, artists have used evolution to generate aesthetically pleasing forms, architects have evolved new building plans from scratch, and computer scientists have evolved the morphologies and control systems of artificial life.

What is evolutionary computation? Briefly, evolutionary computation refers to a collection of stochastic search algorithms whose designs are gleaned from natural evolution: genetic inheritance and the Darwinian principle of the survival-of-the-fittest (natural selection). The several different styles of evolutionary algorithms each share a common feature, that is, modelling the search process by mimicking a biological evolu-

Figure 2. Evolutionary computation and evolutionary design have their roots in computer science and evolutionary biology.



tion process which is operated over the solution space. They are different mainly in the evolution operators involved and the representation of the solution space. For example, in genetic programming, each solution is represented by a computer program, and hence the evolution process is implemented on a society of computer programs. Today, evolutionary computation as a search principle can be extensively used in many disciplines. As the power of evolution gains increasingly widespread recognition evolutionary computing has been used to tackle a broad variety of problems in an extremely diverse array of fields; for example, in engineering or applied mathematics, evolutionary computation has been extensively applied to problems whose solution space is irregular, that is, too large and highly complex, so that it is difficult to employ conventional optimisation procedures to search for the global optimum. Some examples of fields where evolutionary computing is widely applied are shown below.

Evolutionary computing creates a population of solutions with repeated iterations. Thus, more data is produced which requires intelligent method of analysis to select the optimum solution using multiple trade-offs. This process of obtaining the optimum solution in the presence of multiple trade-offs is better achieved through an intelligent framework of collaborating individuals using ontology to explicitly specify conceptualisation where definitions associate concepts, functions, context, taxonomies, and relationships with high-level, human-readable text as well as low-level, machine-readable axioms (Yuhua & Zhengding, 2004). The high-level text serves as an intuitive inference point and knowledge reuse for collaborators can be achieved while the low-level text is important for communication among computational resources as well as among the collaborators. Schikuta and Weishaupl (2004) demonstrate a way of using intelligent systems in their N2Grid project which uses neural networks to exchange information and exploit available computational resources using a 3-layered architecture. The layers are knowledge layer, which uses 2-dimensional information for analysis and mining of data to provide problem solving mechanisms, information layer, which uses 1-dimensional information for Semantic interpretation, and data layer which provides data administration potential.

The concept of design optimisation deals with betterment and improvement (Oduguwa, Tiwari, & Roy, 2005). Over the years with the advancement of computer technology, optimisation and optimisation techniques have been receiving new attention from across the industry sectors. There are three main optimisation techniques known today that most industries rely on, evolutionary computing as a primary technique for optimisation. Unlike traditional methods, which are often employed to solve complex real world problems that tend to inhibit elaborate exploration of the search space, evolutionary computing is generating considerable interest for solving real world engineering problems as it is robust in delivering global optimal solutions and helping to resolve limitations encountered in traditional methods. Nevertheless, over the last decade, the hybrid method, in which EC plays a bigger part, also has matured considerably. When coupled with the marked advances in computing hardware, these hybrid methods become solution for complex industrial optimisation problems.

A major goal driving current design optimisation technique research is to significantly decrease the cost, since the optimal design is the one that most economically meets its performance requirements, time, and increase quality of products (Mill, Shen, Martinz, Noel, & Ram, 2001). Study shows that over the years the ever increasing competitiveness in the market forces industry to look for better methods and techniques that help them not only stay competitive in the market but also put them in a winning position. Among those techniques and methods, the subject of this chapter, design optimisation, is the one getting most attention from industries today.

In the past, we have seen extensive developments in computational applications in order to improve the efficiency of a design process, for example, FEA, CAD/CEM, and virtual modeling. In most resent years much research has been conducted in optimisation. The introduction of optimisation in design is revolutionary in that it aids both the efficiency and creativity of a designer, improving the quality of a design itself. Optimisation and search methods can assist the designer at all stages of the design process to reach to the final optimal product, the final product that meets the performance requirements most (Tiwari, Roy, Jarad, & Munaux, 2002). The designer can choose the optimisation and search methods by tacking in to consideration the nature and type of problems needed to be solved. Broadly speaking, there are three main optimisation techniques that the designer can choose from, namely, classical technique, evolutionary computing, and hybrid technique. Research shows that these three techniques are all popular in industry based on their ability each technique and methods has in terms of dealing with problems that need to be solved. Over the years however, due to the fact that current engineering deign problems are large scale, multiobjective, and complex in nature, industries are fevered more in using evolutionary computing techniques (Schonning, Richard, & Zarda, 2005). It is also learned that hybrid technique are getting equal consideration as a solution for current optimisation problems. Hybrid techniques develop by combining two or more technique together to create a tailor made technique capable to deal with optimisation problems. On the contrary, the classical methods of optimisation are useful in finding the optimum solution of continuous and differentiate functions (Simpson & Mistree, 2001). These methods are analytical and make use of the techniques of differential calculus in locating the optimum points. Since some of the practical problems involve objective functions that are not continuous and/or differentiate, the classical optimisation techniques have limited scope in practical applications.

Figure 3 shows a significant growth of interest in evolutionary computing techniques compare to classical methods and hybrid techniques.

EC in Aerospace Engineering

Obayashi, Daisuke, Yukihiro, and Naoki (2000) use a multiobjective *genetic algorithm* to design the wing shape for a supersonic aircraft. Three major considerations govern the wing's configuration: minimising aerodynamic drag at supersonic cruising speeds, minimising drag at subsonic speeds, and minimising aerodynamic load (the

Figure 3. Surveyed papers percentage share of techniques

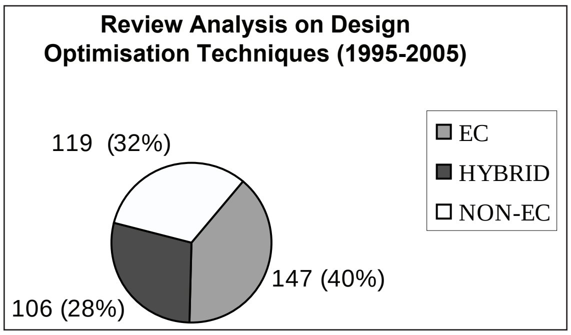
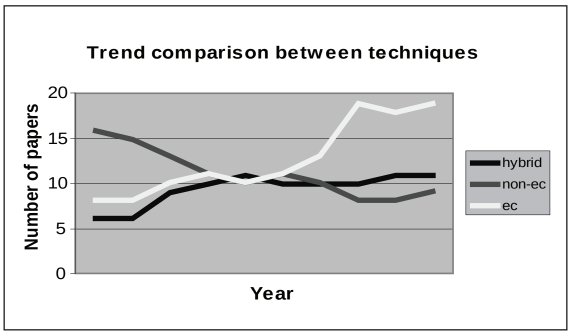


Figure 4. A comparative trends of techniques in the last 12 years



bending force on the wing). These objectives are mutually exclusive, and optimising them all simultaneously requires tradeoffs to be made.

Williams, Crossley, and Lang (2001) apply genetic algorithms to the task of spacing satellite orbits to minimise coverage blackouts. As telecommunications technology continues to improve, humans are increasingly dependent on Earth-orbiting satellites to perform many vital functions, and one of the problems engineers face is designing their orbital trajectories. Satellites in high Earth orbit, around 22,000 miles up, can see large sections of the planet at once and be in constant contact with ground stations, but these are far more expensive to launch and more vulnerable to cosmic radiation. It is more economical to put satellites in low orbits, as low as a few hundred miles in some cases, but because of the curvature of the Earth it is inevitable that these satellites will at times lose line-of-sight access to surface receivers and thus be useless. Even constellations of several satellites experience unavoidable blackouts and losses of coverage for this reason. The challenge is to arrange the satellites' orbits to minimise this downtime. This is a multiobjective problem, involving the minimisation of both the average blackout time for all locations and the maximum blackout time for any one location; in practice, these goals turn out to be mutually exclusive.

When the GA was applied to this problem, the evolved results for three, four, and five-satellite constellations were unusual, highly asymmetric orbit configurations with the satellites spaced by alternating large and small gaps rather than equal-sized gaps as conventional techniques would produce. However, this solution significantly reduced both average and maximum revisit times, in some cases by up to 90 minutes. In a news article about the results, Dr. William Crossley noted that "engineers with years of aerospace experience were surprised by the higher performance offered by the unconventional design."

EC in Chemical Engineering

High-powered, ultra short pulses of laser energy can split apart complex molecules into simpler molecules, a process with important applications to organic chemistry and microelectronics. The specific end products of such a reaction can be controlled by modulating the phase of the laser pulse. However, for large molecules, solving for the desired pulse shape analytically is too difficult; the calculations are too complex and the relevant characteristics (the potential energy surfaces of the molecules) are not known precisely enough.

Assion, Baumert, Bergt, Brixner, Kiefer, Seyfried, et al. (1998) solved this problem by using an evolutionary algorithm to design the pulse shape. Instead of inputting complex, problem-specific knowledge about the quantum characteristics of the input molecules to design the pulse to specifications, the EA fires a pulse, measures the proportions of the resulting product molecules, randomly mutates the beam characteristics with the hope of getting these proportions closer to the desired output, and the process repeats. (Rather than fine-tune any characteristics of the laser beam directly, the authors' GA represents individuals as a set of 128 numbers, each of which is a voltage value that controls the refractive index of one of the pixels in the laser light modulator. Again, no problem-specific knowledge about the properties of either the laser or the reaction products is needed.) The authors state that their algorithm, when applied to two sample substances, "automatically finds the best configuration . . . no matter how complicated the molecular response may be demonstratin" automated coherent control on products that are chemically different from each other and from the parent molecule.

Gillet (2002) also discusses the use of a multiobjective genetic algorithm for the product-based design of combinatorial libraries. In choosing the compounds that go into a particular library, qualities such as molecular diversity and weight, cost of supplies, toxicity, absorption, distribution, and metabolism are considered. In a related paper, Glen and Payne (1995) discuss the use of genetic algorithms to automatically design new molecules from scratch to fit a given set of specifications. Given an initial population either generated randomly or using the simple molecule ethane as a seed, the GA randomly adds, removes, and alters atoms and molecular fragments with the aim of generating molecules that fit the given constraints. The GA can simultaneously optimise a large number of objectives, including molecular weight, molecular volume, number of bonds, number of chiral centres, number of atoms, number of rota table bonds, polarisability, dipole moment, and more in order to produce candidate molecules with the desired properties. Based on experimental tests, including one difficult optimisation problem that involved generating molecules with properties similar to ribose (a sugar compound frequently mimicked in antiviral drugs), the authors conclude that the GA is an "excellent idea generator" that offers fast and powerful optimisation properties and can generate a diverse set of possible structures.

Materials Engineering

Giro, Cyrillo, and Galvão (2002) use genetic algorithms to design electrically conductive carbon-based polymers known as polyanilines. These polymers, a recently invented class of synthetic materials, have "large technological potential applications" and may open up windows onto "new fundamental physical phenomena." However, due to their high reactivity, carbon atoms can form a virtually infinite number of structures, making a systematic search for new molecules with interesting properties all but impossible. To tackle this problem Giro et al. propose a GA-based approach to the task of designing new molecules with prespecified properties, starting with a randomly generated population of initial candidates. They conclude that their methodology can be a "very effective tool" to guide experimentalists in the search for new compounds and is general enough to be extended to the design of novel materials belonging to virtually any class of molecules.

Weismann, Hammel, and Bäck (1998) apply evolutionary algorithms to a "nontrivial" industrial problem: the design of multilayer optical coatings used for filters that reflect, transmit, or absorb light of specified frequencies. These coatings are used in the manufacturing of sunglasses, for example, or compact discs. Their manufacture is a precise task. The layers must be laid down in a particular sequence and particular thicknesses to produce the desired result, and uncontrollable environmental variations in the manufacturing environment such as temperature, pollution, and humidity may affect the performance of the finished product. Many local optima are not robust against such variations, meaning that maximum product quality must be paid for with higher rates of undesirable deviation. The EA operated by varying the number of coating layers and the thickness of each, and produced designs that were "substantially more robust to parameter variation" and had higher average performance than traditional methods. The authors conclude that "evolutionary algorithms can compete with or even outperform traditional methods" of multilayer optical coating design without having to incorporate domain-specific knowledge into the search function and without having to seed the population with good initial designs.

Robin, Andrea, Esteban, Otte, and Werner (2003) use EC to design exposure patterns for an electron lithography beam used to etch sub micrometer-scale structures onto integrated circuits. Designing these patterns is a highly difficult task; it is cumbersome and wasteful to determine them experimentally, but the high dimensionality of the search space defeats most search algorithms. As many as 100 parameters must be optimised simultaneously to control the electron beam and prevent scattering and proximity effects that would otherwise ruin the fine structures being sculpted. The forward problem—determining the resulting structure as a function of the electron dose—is straightforward and easy to simulate, but the *inverse* problem of determining the electron dose to produce a given structure, which is what is being solved here, is far harder and no deterministic solution exists.

Genetic algorithms, which are known to be able to find good solutions to very complex problems of high dimensionality without needing to be supplied with domain-specific information on the topology of the search landscape, were applied successfully to the problem. The chapter's authors employed a steady-state GA with roulette-wheel selection in a computer simulation, which yielded 'very good optimised' exposure patterns.

Systems Engineering

Benini and Toffolo (2002) applied a *genetic algorithm* to the multiobjective task of designing wind turbines that used to generate electric power. This design is a complex procedure characterised by several trade-off decisions, where the constructing of the blades with the optimal configuration of characteristics such as tip speed, hub/tip ratio, and chord and twist distribution. In the end, the GA was able to find solutions competitive with commercial designs, as well as more clearly elucidate the margins by which annual energy production can be improved without producing overly expensive designs.

EVOLUTIONARY COMPUTING APPLICATION TRENDS IN THE LAST 12 YEARS

It is a documented fact that conventional numerical optimisation methods have the known advantage of their efficiency; however, they are very sensitive to the starting point selection and are very likely to stop at nonglobal optima. The search for algorithms that are capable of escaping from local optima has led to the development of stochastic optimisation techniques via the introduction of probabilistic factors in the search process that encourage global exploration. In addition, stochastic techniques, unlike conventional numerical optimisation methods, produce new design points that do not use information about the local slope of the objective function and are thus not prone to stalling at local optima. Further, they have shown considerable potential in the solution of optimisation problems characterised by nonconvex and disjoint or noisy solution spaces. Modern

Figure 5. EC technique application trend in the last decade

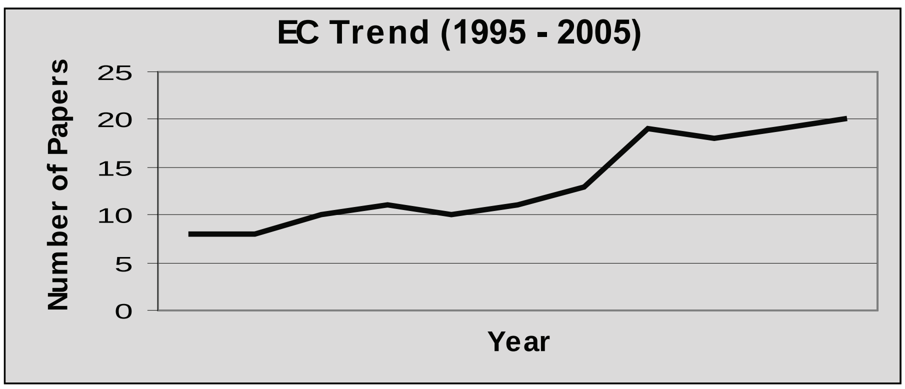
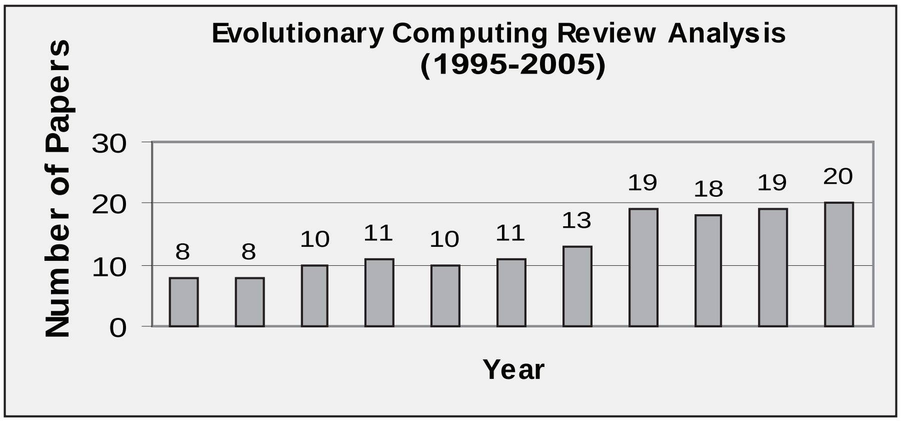


Figure 6. Year wise growth of EC technique in the last 12 years



stochastic optimisers which have attracted much attention in recent years include simulated annealing, tabu search, genetic algorithms, evolutionary programming, and evolution strategies (Back et al., 1991; Davis, 1991; Kirkpatrick, Gelatt, & Vecchi, 1983). These stochastic methods have been successfully applied to many engineering and other fields in solving complex and high disturbance optimisation problems. The survey conducted and analysed by this report auteur also confirmed that in recent years the application trend has shown a considerable increase compare to conventional and hybrid technique. As shown in Figures 3 and 4, the technique emerging is a new engineering computational paradigm, which may significantly change the present and future design optimisation practice. It is also learned from the survey that due to its robustness, powerful global optimisation nature for solving large scale problems and flexibility and adaptability to the task at hand makes the EC technique better suited to deal with current design optimisation problems as compared to its classical and hybrid counterpart. Figures 4, 5, and 6 show year-wise growth techniques in application in the last 12 years.

CHALLENGES IN DESIGN OPTIMISATION

Engineering design optimisation is a challenging discipline. The obvious challenge is in decision making. This is because many experts are usually involved in any engineering design and often times are experts geographically dispersed. Collaborative software systems are used to cater for collaboration among designers to solve this problem of geographical separation. However, there are also challenges in interoperability of hardware and software platforms used by different designers in different domain areas. Optimisation is also computationally and data intensive. This is also another challenge. Parallel distributed computing and recently grid computing technologies are used to address this problem. Security is an issue in these emerging technologies. Using evolutionary techniques such as genetic algorithm (GA), genetic programming (GP), and classifier strategies is also challenging. The different objective functions that need to be optimised can present great challenges and conflicting results during optimisation. The qualitative aspect of optimisation is subjective and presents challenges to design experts. Using soft computing techniques such as fuzzy-logic and neural networks to model qualitative parameters of design optimisation process should be considered.

LIMITATIONS OF EVOLUTIONARY COMPUTING

Although evolutionary computing has proven to be an efficient and powerful problem-solving strategy as shown in the example above, they are not a problem-free tool. EC do have certain limitations, in particular in the following areas.

The first, and most important, consideration in creating a genetic algorithm is defining a representation for the problem. The language used to specify candidate solutions must be robust; that is, it must be able to tolerate random changes such that fatal errors or nonsense do not consistently result.

The problem of how to write the fitness function must be carefully considered so that higher fitness is attainable and actually does equate to a better solution for the given problem. If the fitness function is chosen poorly or defined imprecisely, the genetic algorithm may be unable to find a solution to the problem, or may end up solving the wrong problem. (This latter situation is sometimes described as the tendency of a GA to 'cheat,' although in reality all that is happening is that the GA is doing what it was told to do, not what its creators intended it to do.)

In addition to making a good choice of fitness function, the other parameters of a GA–the size of the population, the rate of mutation and crossover, and the type and strength of selection–must be also chosen with care. If the population size is too small, the genetic algorithm may not explore enough of the solution space to consistently find good solutions. If the rate of genetic change is too high or the selection scheme is chosen poorly, beneficial schema may be disrupted and the population may enter error catastrophe, changing too fast for selection to ever bring about convergence.

One type of problem that EC have is that when genetic algorithms have difficulty in dealing with problems with 'deceptive' fitness functions, those where the locations of improved points give misleading information about where the global optimum are likely to be found (Mitchell, 1996). For example, imagine a problem where the search space consisted of all eight-character binary strings, and the fitness of an individual was directly proportional to the number of 1s in it; that is, 00000001 would be less fit than 00000011, which would be less fit than 00000111, and so on. But there are two exceptions: the string 11111111 turned out to have very low fitness, and the string 00000000 turned out to have very high fitness. In such a problem, a GA (as well as most other algorithms) would be no more likely to find the global optimum than random search.

One well-known problem that can occur with a GA is known as *premature convergence*. If an individual that is more fit than most of its competitors emerges early on in the course of the run, it may reproduce so abundantly that it drives down the population's diversity too soon, leading the algorithm to converge on the local optimum that that individual represents rather than searching the fitness landscape thoroughly enough to find the global optimum (Forrest, 1993; Mitchell, 1996). This is an especially common problem in small populations, where even chance variations in reproduction rate may cause one genotype to become dominant over others.

Finally, several researchers, including Holland (1992), Forrest (1993), and Haupt and Haupt (1998), advise against using genetic algorithms on analytically solvable problems. It is not that genetic algorithms cannot find good solutions to such problems; it is merely that traditional analytic methods take much less time and computational effort than GAs and, unlike GAs, are usually mathematically guaranteed to deliver the one exact solution. Of course, since there is no such thing as a mathematically perfect solution to any problem of biological adaptation, this issue does not arise in nature. In addition, a skied expert and a cost, initial set up and running, required implementing and integrating it with systems.

FUTURE RESEARCH DIRECTIONS

The field of evolutionary computing continues to grow rapidly and now it is the main technique to deal with the real-world engineering design problems, especially for multiobjective optimisation. As the study shows, in the last decade EC has gained considerable popularity among engineers and researchers. Traditional methods, although they are favoured to solve complex real world optimisation problems, have considerable drawbacks in efficiency, cost, and time. However, although evolutionary computation, unlike classical counterpart, aims to exploit design tolerance for imprecision, uncertainty, and partial truth to achieve tractability and robustness with relatively minimal cost, its solutions are more of independent rather than integrated manner. Therefore, as astonishing and counterintuitive as it may seem, evolutionary computation itself does still have certain limitation. Some of the limitations are as follows.

Studies are required to develop optimisation algorithms that can deal with various combinations such as complementary and noncooperative forms of integrated qualitative and quantitative (Q^T and Q^L) design information in a single framework.

Scalability of integrated qualitative and quantitative (Q^T and Q^L) design optimisation strategies to higher dimensional problems is an important success criterion for wider applications. This is influenced by the feature of the problem (large number of parameters) and the nature of the resulting search space (discontinuous).

An inherent feature of process optimisation problems is the hierarchical nature of the problem. Another important property of the problem is that since each process unit offers a unique behaviour, the optimal design path and convergence characteristics with which each stage search process unit responds to the overall system level optimal solution does vary. Since the optimal design path and convergence characteristics are key features for efficient search performance, it is necessary to develop techniques incorporating the hierarchical behaviour in process optimisation problems.

CONCLUSION

The field of evolutionary computing continues to grow rapidly and now it is the main technique to deal with the real world engineering design problems, especially for multiobjective optimisation. As the study shows, in the last decade EC has gained considerable popularity among engineers and researchers. Traditional methods, although they are favoured to solve complex real world optimisation problems, they have considerable drawbacks in efficiency, cost, and time. It has been learned that over the years the classical technique is showing a negative growth and becoming less favoured in application. On the other hand hybrid, in which EC considered as a vital part, is also a method that has matured considerably. Despite the drawbacks in some ways it may have, evolutionary computing is still considered by many as the solution for today's optimisation problems.

ACKNOWLEDGMENT

The researchers are grateful for the contributions from the members of the Decision Engineering Centre at Cranfield University and the Industrial Partners.

REFERENCES

Assion, A., Baumert, T., Bergt, M., Brixner, T., Kiefer B., Seyfried, V., et al. (1998). Control of chemical reactions by feedback-optimized phase-shaped femtosecond laser pulses. *Science*, *282*, 919-922.

Au, W. H., Chan, K., & Yao, X. (2003). A novel evolutionary data mining algorithm with applications to churn prediction. *IEEE Transactions on Evolutionary Computation*, *7*(6), 532-545.

Back, T., Hammel, U., & Schwefel, H-P. (1997). Evolutionary computation: Comments on the history and current state. *IEEE Transactions on Evolutionary Computation*, 13-17.

Benini, E., & Toffolo, A. Optimal design of horizontal-axis wind turbines using blade-element theory and evolutionary computation. *Journal of Solar Energy Engineering*, *124*(4), 357-363.

Booker, R. (1991). In *Proceedings of the Fourth International Conference on Genetic Algorithms* (pp. 2-9).

Burke, E.K., & Newall, J.P. (1999). A multistage evolutionary algorithm for the timetable problem. *IEEE Transactions on Evolutionary Computation*, *3*(1), 63-74.

Charbonneau, P. (1995).Genetic algorithms in astronomy and astrophysics. *The Astrophysical Journal Supplement Series*, *101*, 309-334.

Cvetkovic, D., & Parmee, I. (2002). Agent based support within an interactive evolutionary design system. *Adaptive computing in design and manufacture* (5th ed., pp. 355-367). London: Springer-Verlag London Limited.

Davis, L. (1991). *Handbook of genetic algorithms.* New York: Van Nostrand Reinhold.

Dirk, W., Hammel, U., & Back, T. (1998). Robust design of multilayer optical coatings by means of evolutionary algorithms. *IEEE Transactions on Evolutionary Computation*, *2*(4), 162-167.

Forrest, S. (1993). Genetic algorithms: Principles of natural selection applied to computation. *Science*, *261*, 872-878.

Gillet, V. (2002). Reactant- and product-based approaches to the design of combinatorial libraries. *Journal of Computer-Aided Molecular Design*, *16*, 371-380.

Giro, R., Cyrillo, M., & Galvão, D.S. (2002). Designing conducting polymers using genetic algorithms. *Chemical Physics Letters*, *366*(1-2), 170-175.

Glen, R.C., & Payne, A.W.R. (1995). A genetic algorithm for the automated generation of molecules within constraints. *Journal of Computer-Aided Molecular Design*, *9*, 181-202.

Graham-Rowe, D. (2002). Radio emerges from the electronic soup. *New Scientist*, *175*(2358), 19.

Haupt, R., & Haupt, S.E. (1998). *Practical genetic algorithms.* New York: John Wiley & Sons.

He, L., & Mort, N. (2000). Hybrid genetic algorithms for telecommunications network back-up routeing. *BT Technology Journal*, *18*(4), 42-50.

Hernandez, S., & Fontan, A. (2002). *Practical application of design optimisation.* Southampton, Boston: WIT press.

Holland, J. (1992). Genetic algorithms. *Scientific American*, 66-72.

Jun, S., Tischler, V.A., & Venkayya, V.B. (2002). Multidisciplinary design optimisation of a build-up wing structure with a tip missile. *43*, 22-25.

Kalyanmoy, D. (2001). *Multi-objective optimisation using evolutionary algorithms.* United Kingdom: John Wiley & Sons Ltd.

Kewley, R., & Embrechts, M. Computational military tactical planning system. *IEEE Trans-*

actions on Systems, Man and Cybernetics, Part C—Applications and Reviews, *32*(2), 161-171.

Kirkpatrick, S., Gelatt, C.D., & Vecchi, M.P. (1983). Optimization by simulated annealing. *Science, 220*(4598), 671-680.

Kodiyalam, S., Yang, R.J., & Gu, L. (2003). *High performance computing rapid visualization for design steering in multi design optimisation.* Paper presented at the 44th AIAA/ASME/ASCE/AHS Structures, Structural Dynamics and Materials Conference.

Li, Y., & Lu, Z. (2004). Ontology-based universal knowledge grid: Enabling knowledge discovery and integration on the grid. *In Proceedings of the 2004 IEEE International Conference on Services Computing (SCC'04).* IEEE Computer Society.

Mahfoud, S., & Ganesh, M. (1996). Financial forecasting using genetic algorithms. *Applied Artificial Intelligence*, *10*(6), 543-565.

Mitchell, M. (1996). *An introduction to genetic algorithms.* MIT Press.

Mill, F., Shen, W., Martinz, I., Noel, P.H., & Ram, M. (2001). Knowledge methodology for design decision. *Artificial Intelligence in Engineering, 15*(2), 144-164.

Obayashi, S., Daisuke, S., Yukihiro, T., & Naoki, H. (2000). Multi objective evolutionary computation for supersonic wing-shape optimization. *IEEE Transactions on Evolutionary Computation*, *4*(2), 182-187.

Oduguwa, V., & Roy, R. (2006). A review of rolling system design optimisation. *International Journal of Machine Tools and Manufacture*, *46*(7-8), 912-928.

Oduguwa, V., Tiwari, A., & Roy, R. (2005). Evolutionary computing in manufacturing industry: an overview of recent applications. *Applied Soft Computing*, *5*, 281-299.

Okechi, O., & Onwubiko, C. (2000). *Introduction to engineering design optimisation.* Upper Saddle River, NJ: Prentice Hall.

Park, J. W., Park, S. H., Moon, J. J., Yoon, Y., & Kim, S. J. (2005). *High-Fidelity simulation-based optimum design utilizing computing grid technology.* Paper presented at the 46th AIAA/ASME/ASCE//AHS/ASC Structures, Structural Dynamics & Materials Conference.

Rizki, M., Michael, Z., & Louis, T. (2002). Evolving pattern recognition systems. *IEEE Transactions on Evolutionary Computation*, *6*(6), 594-609.

Robin, F., Andrea, O., Esteban, M., Otte, H., & Werner, B. (2003). Simulation and evolutionary optimization of electron-beam lithography with genetic and simplex-downhill algorithms. *IEEE Transactions on Evolutionary Computation*, *7*(1), 69-82.

Rohl, P. J., Kolonay, R. M., Irani, R. K., Sobolewski, M., Kao, K., & Bailey, M. W. (2000). *A federated intelligent product environment.* Paper presented at the 8th Annual AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis and Optimization, Long Beach, California.

Sambridge, M., & Kerry, G. (1993). Earthquake hypocenter location using genetic algorithms. *Bulletin of the Seismological Society of America*, *83*(5), 1467-1491.

Schikuta, E., & Weishaupl, T. (2004). N2Grid: Neural networks in the grid. *IEEE,* 1409-1414.

Schonning, A., Richard, J., & Zarda, R. (2005). An integrated design and optimisation environment for industrial large scaled system. *16*, 86-95.

Schwefel, H.P. (1995). *Evolution and optimum seeking.* New York: John Wiley & Sons.

Simpson, T.W., & Mistree, F. (2001). Kriging models for global approximation in simulating-based multidisciplinary design optimisation. *39*(12), 2233-2241.

Singiresu, R.S., & Xiong, Y. (2005). Mixed discrete multi objective programming for engineering optimisation using hybrid genetic algorithm. *AIAA Journal, 43*(1), 1580-1590.

Tiwari, A. (2001). *Evolutionary computing techniques for handling variable interaction in engineering design optimisation*. Unpublished doctoral thesis, Cranfield University, Cranfield, UK.

Tiwari, A., Roy, R., Jarad, G., & Munaux, O. (2002). Evolutionary-based techniques for real-life optimisation development and testing. *1*, 301-329.

Wager, T.D., & Nichols, T.E. (2003). Optimisation of experimental design in fMRI: A general framework using a genetic algorithm. *IEEE NeuroImage Academic Press, 18*, 293-309.

Williams, E., William, C., & Thomas, L. (2001). Average and maximum revisit time trade studies for satellite constellations using a multi-objective genetic algorithm. *Journal of the Astronautically Sciences, 49*(3), 385-400.

ADDITIONAL READING

Azadivar, F., & Wang, J. (2000). Facility layout optimisation using simulation and genetic algorithms. *International Journal of Production Research, 38*(17), 4369-4383.

Bäck, T., (1996). *Evolutionary algorithms in theory and practice*. New York: Oxford University Press.

Bentley, P.J. (1997). The revolution of evolution for real-world applications. *Emerging technologies '97: Theory and application of evolutionary computation*. University College London.

Bentley, P.J., & Wakefield, J.P. (1997). *Generic evolutionary design*. Chawdhry, P.K.

Coello, C.A.C. (1999). A comprehensive survey of evolutionary-based multiobjective optimization techniques. *Knowledge and Information Systems, 1*(3), 269-308.

Coello, C. (2000). An updated survey of GA-based multiobjective optimization techniques. *ACM Computing Surveys*, *32*(2), 109-143.

Coit, D.W., Alice, E.S., & David M.T. (1996). Adaptive penalty methods for genetic optimization of constrained combinatorial problems. *INFORMS Journal on Computing*, *8*(2), 173-182.

Conrad, M. (1981). Algorithmic specification as a technique for computing with informal biological models. *Bios Stems, 13*, 303-320.

Deb, K. (1995).*Optimization for engineering design: Algorithms and examples*. New Delhi, India: Prentice-Hall.

Deb, K. (1999). Evolutionary algorithms for multi-criterion optimization in engineering design. In K. Miettinen, M.M. Makela, P. Neittaanmaki, & J. Periaux (Eds.), *Evolutionary algorithms in engineering and computer science*. UK: Wiley Press.

Fogel, D.B. (Ed.). (1998). *Evolutionary computation: The fossil record*. New York: IEEE Press.

Fogel, D.B., & Ghozeil, A. (1997). A note on representations and variation operators. *IEEE Transactions on Evolutionary Computation, 1*(2), 159-161.

Fogel, D.B., & Barricelli, N. (2006). Artificial life, coevolution, self-adaptation. *IEEE Computational Intelligence Magazine*, 41-45.

Fonseca, C.M., & Fleming, P.J. (1995). An overview of evolutionary algorithms in multiobjective optimization. *Evolutionary Computation, 3*(1), 1-16.

Gen, M., & Cheng, R. (1997). *Genetic algorithms and engineering design*. New York: John Wiley & Sons.

Goldberg, D. (1989). *Genetic algorithms in search, optimization, and machine learning.* Boston: Addison-Wesley.

Horn, J., Goldberg, D.E., & Deb, K. (1994). Implicit niching in a learning classifier system: Nature's way. *Evolutionary Computation*, *2*(1), 37-66.

Kerr, C.I.V., Roy, R., & Sackett, P.J. (2004a). *A product ontology for automotive seat specification.* Paper presented at the ASME International Design Engineering Technical Conferences & the Computer and Information in Engineering Conference and 30th Design Automation Conference, Salt Lake City.

Kerr, C.I.V., Roy, R., & Sackett, P.J. (2004b, September 7-9). *The assignment of product specification responsibilities for automotive system development.* Paper presented at The 2nd International Conference on Manufacturing Research—Encouraging Applied Research in Industry, Sheffield, UK

Oduguwa, V., Tiwari, A., & Roy, R. (2004, September 20-October 8). *An overview of soft computing techniques used in the drug discovery process.* Paper presented at the 9th Online World Conference on Soft Computing in Industrial Applications (WSC-9).

Parashar, M., Matossian, V., Bangerth, W., Klie, H., Rott, B., Kurc, T., et al. (2005). *Towards dynamic data-driven optimization of oil well placement* (pp. 656-663). Heidelberg, Berlin: Springer-Verlag.

Parmee, et al. (2002). Preferences and their application in evolutionary multi objective optimization. *IEEE Transactions on Evolutionary Computation, 6*(1).

Parmee, I.C., Abraham, J., Shackelford, M., Rana, O.F., & Shaikhali, A. (2005). Towards autonomous evolutionary design systems via grid-based technologies. In *Proceedings of ASCE International Conference on Computing in Civil Engineering,* Cancum, Mexico.

Song, W., Ong, Y.S., Ng, H.K., Keane, A., Cox, S., & Lee, B.S. (2004). *A service-oriented approach for aerodynamic shape optimisation across institutional boundaries.* Paper presented at the 8th International Conference on Control, Automation, Robotics and Vision, Kunmlng, China.

Wolpert, D.H., & Macready, W.G. (1997). No free lunch theorems for optimization. *IEEE Transactions on Evolutionary Computation*, *1*(1), 67-82.

Woyak, S., Kim, H., Mullins, J., & Sobieszczanski-Sobieski, J. (2004). *A Web centric architecture for deploying multi-disciplinary engineering design processes.* Paper presented at the 10th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference, Albany, New York. USA. AIAA Inc.

Section II
Supply Chain and Decision Support Systems

Chapter X
Towards a Methodology for Monitoring and Analyzing the Supply Chain Behavior

Reinaldo Moraga
Northern Illinois University, USA

Luis Rabelo
University of Central Florida, USA

Alfonso Sarmiento
University of Central Florida, USA

ABSTRACT

In this chapter, the authors present general steps towards a methodology that contributes to the advancement of prediction and mitigation of undesirable supply chain behavior within short- and long- term horizons by promoting a better understanding of the structure that determines the behavior modes. Through the integration of tools such as system dynamics, neural networks, eigenvalue analysis, and sensitivity analysis, this methodology (1) captures the dynamics of the supply chain, (2) detects changes and predicts the behavior based on these changes, and (3) defines needed modifications to mitigate the unwanted behaviors and performance. In the following sections, some background information is given from literature, the general steps of the proposed methodology are discussed, and finally a case study is briefly summarized.

INTRODUCTION

In today's business environment, supply chain management (SCM) plays a crucial role in modern companies endeavoring to uphold their competitive advantages. In the past, most of the managerial practices employed to control the supply chain relied mainly on monitoring data with respect to sales, demand, and inventory levels so as to react appropriately when needed. The implicit assumption was that demand and supply would remain predictable, which was valid in a market that was dominated by the supplier's perspective, not the consumer's. Today, an effective SCM involves the management of supply chain assets, products, information, and fund flows if a company is to maximize total supply chain profitability (Chopra & Meindl, 2007). The long-term goal of maximizing total profitability in SCM is multidimensional in nature and translates into cost minimization, increased levels of service, improved communication among partners, and increased flexibility in terms of delivery and response (Lancioni, Smith, & Oliva, 2000). All these dimensions represent conflicting objectives that make the decision making process throughout the supply chain more complex. In addition, a supply chain is a complex dynamic system consisting of a hierarchical nesting of both continuous and discrete dynamics. Therefore, modeling plays a fundamental role in both analysis and improving the supply chain.

In practice, due to the complexity observed in supply chain systems, diverse research has been required to design and analyze a supply chain system by using different modeling approaches. Most approaches fall into one out of four main categories of methodology: continuous time differential equation models, discrete time difference models, discrete event models, and classical operational research methods. In the latter case, researchers have studied the application of various operations research techniques to analyze supply chain problems (Anderson & Marklund, 2000; Bose & Pekny, 2000; Jayaraman & Pirkul, 2001; Lakhal, Martel, Kettani, & Oral, 2001; Riddalls, Bennett, & Tipi, 2000; Shapiro, 2001). Riddals et al. (2000), however, said that none of the operations research techniques is best, and none are worthless. These techniques may be of unique benefit at the tactical level, but they certainly fail to provide insights into the dynamic behavior of the supply chain as whole. However, simulation models based on dynamics of systems may be the only way to study phenomena like the demand amplification, where small fluctuations in demand at the retailer end of the supply chain are augmented as they advance through the supply chain. The interested reader is referred to works by Riddals et al. (2000) and Beamon (1998) for a more comprehensive review of the models and methods used in the literature.

In this chapter, general steps towards a methodology are introduced to monitor and analyze the supply chain behavior patterns. Additional information emphasizing other aspects of this methodology has been recently published by Rabelo, Helal, Lertpattarapong, Moraga, and Sarmiento (in press). The proposed methodology contributes to the advancement of prediction and mitigation of undesirable supply chain behavior within short- and long-term horizons by promoting a better understanding of the structure that determines the behavior modes through the integration of tools such as system dynamics, neural networks, eigenvalue analysis, and sensitivity analysis. It is imperative for manufacturing industries to equip themselves with tools to detect changes in the supply chain behavior due to external and/or internal factors and be prepared to counteract any undesirable consequences. This methodology (1) captures the dynamics of the supply chain, (2) detects changes and predicts the behavior based on these changes, and (3) defines needed modifications to avoid (or mitigate) the unwanted behaviors and performance. In the following sections, some background information is given from the literature, the general steps of the proposed methodology are presented, and finally a case study is briefly summarized.

LITERATURE BACKGROUND

The use of simulation techniques is having a significant impact in SCM for modeling and analysis. Several scenarios may be assessed through the use of a computer simulation model and those that result in best cost-effective metrics may be selected for implementation later in the supply chain. Literature distinguishes four types of simulation techniques:

1. Spreadsheet simulation(Ezingeard & Race, 1995; Smith, 2003; Sparling, 2002).
2. Discrete-event simulation (Persson & Olhager, 2002; Stratman, Roth, & Gilland, 2004).
3. System dynamics (Morecroft, 1983; Otto & Kotzab, 2003; Santos, Belton, & Howick, 2002).
4. Business games (Kleijnen & Smits, 2003).

The reader is referred to works by Kleijnen (2005) and Terzi and Cavalieri (2004) for excellent surveys on simulation in the supply chain context. For the purposes of this chapter we will focus on system dynamics (SD).

SD is a computer-aided technique that employs concepts from feedback control theory and continuous simulation to model and analyze the dynamic behavior of complex physical, biological, and social systems. Underneath the SD model, a mathematical description is carried out through the use of ordinary differential equations. Contrasting this to other simulation techniques like agent-based modeling which is considered a bottom-up approach, where individual agents interact with each other operating a set of rules and then the emergent behavior of the overall simulated interaction is analyzed, SD is labeled a top-down approach because variables and all relationships between them are defined in advance before analyzing the overall behavior (Richardson, 2003). SD has been widely applied to many areas such as healthcare systems (Dangerfield, 1999; Lane, Monefeldt, & Rosenhead, 2000), economics (Harvey, 2006), project management (Rodrigues & Williams, 1998), defense (Coyle, Exelby, & Holt, 1999), urban policy (Forrester, 1969), and other policy and managerial issues (Lyneis, 1980; Morecroft, Lane, & Viita, 1991; Sterman, 2000).

The use of SD modeling in SCM can be traced back to Forrester's (1961) pioneering work on industrial dynamics, but only over the last decade has SD become increasingly more popular given the observed complexity in supply chain systems. Different settings for factors, such as the uncertainties of demand, the number of suppliers, logistics, inventory methods, distribution channels, and so forth, may have a strong influence on the dynamic behavior of the supply chain. Since SD provides a high-level of understanding of large systems to its users in terms of system feedback loops, systemic delays, and unintended consequences (Affeldt, 1999), it is ideal for supply chain modeling. Bruniaux, Pierreval, and Caux (1999) explain the significance of simulating the supply chains for enterprises in order to understand its complex behavior. Once the behavior of the supply chain is comprehended, it can significantly help decision makers to formulate correct decisions.

The authors in this research chapter highlight the ability of SD to evaluate supply chains. Moreover, SD has been suitable to study demand amplification in supply chains or the so called "bullwhip" effect, through which the variability in orders placed by different customers downstream is amplified at each stage of the supply chain, ultimately causing strong demand oscillation patterns for the manufacturer (Towill, Naim, & Wikner, 1992; Wikner, Towill, & Naim, 1991). For instance, due to the variability in orders for replenishing stock, manufacturers make assumptions with respect to their future needs, which lead to undersupply or oversupply that can have serious economic impacts for the manufacturer. Using SD, Higuchi and Troutt (2004) study phenomena such as "bullwhip effect," "bust," and "boom," and their interaction with other factors; authors use scenario based dynamic simulations to study the short product life cycle case. One of their findings is that information distortion is the main cause of bust and boom and bullwhip effect. Dejonckheere, Dysney, Lambrecht, and Towill (2003) prove that the bullwhip effect is guaranteed in the order-up-to replenishment strategy irrespective of the forecasting method used. Sterman (2000) discusses various topics on systems thinking, simulation modeling, strategic thinking, operations, and various other topics related to business dynamics. He presents generic models that can be adapted for particular applications. Various case studies and interesting problems are proposed in this book in order to improve readers understanding of the SD field. Angerhofer and Angelides (2000) provide a concise review and taxonomy of SD applications to SCM within categories such as inventory management, demand amplification, supply chain design, and international SCM. Zhang and Dilts (2004) show there are significant effects of demand and network structural factors, and their interactions, on cost and fill rate performance in two-echelon and two supply-chain network organization models. As demand becomes dynamic, the cooperative interaction model, where supply chains cooperate to satisfy customer demand, is found to have better system performance than the competitive supply chain model. Spengler and Schröter (2003) use SD to model an integrated production and recovery system for supplying spare parts to evaluate possible strategies for meeting spare-parts demand for electronic equipment in the end-of-life service period. Lin, Tung and Huang (2006) apply SD to explore factors affecting the industrial cluster effect, which is crucial in determining national and industrial competitive advantage. The considered factors are manpower, technology, money, and market flows and a comprehensive causal loop is constructed. The authors conclude that all these factors have a positive influence on the industrial cluster effect and recognize the effectiveness of the SD approach with respect to other methodologies in this type of studies.

SD modeling not only specifies mathematical relationships among the objects that are part of the system under study, but also involves the model calibration, which is the process of estimating the model parameters to get a match between observed and simulated behavior, to proceed later with the generation of inferences and solutions. An iterative process to build confidence in these models is normally emphasized in literature and some tests that models should pass have been proposed (Barlas, 1994; Sterman, 2000), but Oliva (2003) argues that not much attention has been given to the iterative process itself. He says the model calibration represents a rigorous test for the dynamic hypothesis that relates structure and behavior and thus he proposes three heuristics to increase the power of automated calibration as testing tool. Hayden (2006) opens a flank of criticism against the adequacy of SD for representing institutional principles of hierarchy, feedback, and openness. He argues that the way these models are calibrated matching historical vs. simulated behavior makes SD models to mimic databases rather than validating a true explanation for the behavior of a social system. As recognized by Oliva (2003), it is certainly difficult to verify a hypothesis, but at least alternative explanations may be ruled out through a scientific and systematic approach and in this way the confidence in a stated hypothesis may be increased.

To get a better understanding of the model structure that determines the supply chain modes of behavior continues being crucial for enhanced performance in SCM. This claim “understanding model behavior” was also put by Richardson (1995) as first in a list of eight problem areas deserving the attention of SD practitioners. In this regard, Forrester (1982) made the earliest attempt in tracking the dynamics of feedback loop dominance in high-order nonlinear models through eigenvalue analysis and eigenvalue elasticity analysis. The idea consists of linearizing the SD model at any point in time, calculating its eigenvalues, and then observing the real and complex numerical components which are associated to particular behavior modes. This approach also lays out the basis for the eigenvalue elasticity analysis, which consists of calculating how the eigenvalues change as causal link gains change in the linearized model. During the last decade, the research in this area has been reactivated and several methods to understand causes of model behavior have been proposed (Güneralp, 2005; Gonçalves, Lertpattarapong, & Hines, 2000; Kampmann, 1996; Kampmann & Oliva, 2006; Mojtahedzadeh, 1997; Oliva, 2004; Saleh & Davidsen, 2001) The eigenvalue elasticity analysis requires recognizing all feedback loops and links associated to them in the model. Since this becomes a time consuming task, some proposed approaches attempt to diminish this effort by focusing on the most important structures within the SD model that best describe the behavior modes. For example, the pathway participation matrix (Mojtahedzadeh, 1997; Mojtahedzadeh, Richardson, & Andersen, 2004), and the independent loop set (Kampmann 1996) and its extensions (Oliva, 2004.) One interesting finding is that the behavior mode of any state variable is a function not only of the eigenvalues but also of the eigenvector (Gonçalves, 2006) and hence in short-term periods behavior may be more influenced by the eigenvector than eigenvalues. Although, most of these approaches and their extensions to understand model behavior are promising, the main limitation for a broadly adoption of these procedures is the automating effort required and the lack of clarity on how to interpret this analysis (Saleh, Davidsen, & Bayoumi, 2005.)

The use of neural networks (NN), which is the other technique employed in the proposed methodology, has been broadly applied to the SCM field. Bruzzone and Orsoni (2003) discuss the implementation of artificial intelligence (AI) and simulation techniques for the assessment of transportation logistics in terms of costs and expected production losses. To accomplish this, authors analyzed and compared three different modeling approaches with one of them based on NN architecture. Choy, Lee, and Lo (2003) found the use of AI and NN appropriate in the field of supplier relationship management (SRM) under a competitive economy. In a case study they proposed and exemplified the use of NN techniques for selecting and benchmarking potential suppliers. Bose and Pekny (2000) discuss the concepts of pattern recognition methods in agile supply chains. They talk about the supply chains for fast moving consumer goods (FMCG). Authors say that in order to maintain a substantial level of customer service, inter organizational modules cannot be isolated and then optimized, but instead the information and demand flows have to be taken into account, and then optimization has to be done on the basis of an emergent system. Leung (1995) argues that NN have been applied to various constituents of a supply chain, but seldom applied to an entire supply chain. This chapter discusses various ways in which NN can benefit supply chain management.

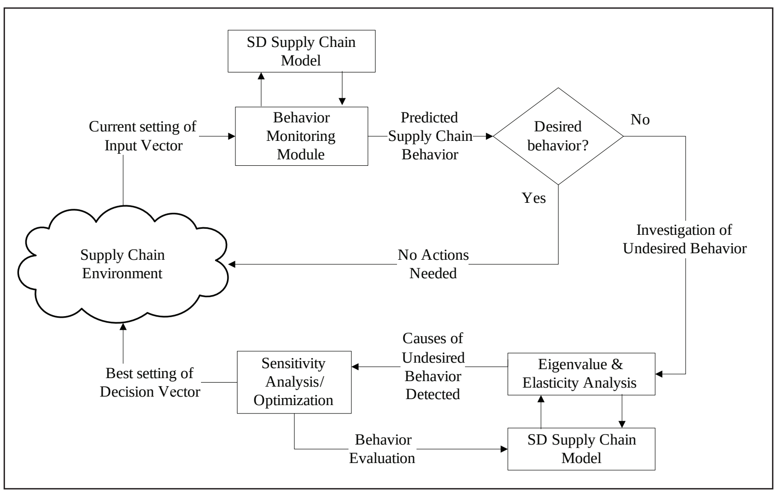
The proposed methodology takes advantage of the synergy produced by using complementary capabilities from SD modeling, NN, eigenvalue, and elasticity analysis of the supply chain. The value of this approach lies in the better use of the SD simulation results. Instead of only analyzing the SD outputs to adjust the system and achieve stability for a particular time, the use of NN accumulates and builds on the knowledge gained from the SD results. Given the system structure, the management can use the behavior categories made by NN to predict the consequences of proposed system changes or expected, as well as unexpected, market events. Intelligent agents based on NN can be developed to work in real time to populate the knowledge gained from SD and analyzed by NN to managers. In addition to that, the use of eigenvalue analysis allows better understanding of how the system evolves over time and the importance of the causal relationships in determining that evolution. Eigenvalues determine the expected behavior of the system, which depends on the position of the eigenvalue on the complex plane. This can support the detection and the analysis of the system behavior. In addition, because of the numerous direct and indirect causal relationships among the variables in the SD model, it is likely that the behavior of one variable will influence the behavior of several others. To determine whether such influences exist and their relative strengths, elasticities of the eigenvalues are computed. This measure helps identify the links or loops that contribute most significantly to the model behavior. Therefore, it measures how one variable is linked (even with time delays) to another one.

METHODOLOGY

The proposed methodology is a procedure that captures the dynamics of the supply chain, predicts and analyzes future behavior modes, and indicates potentials for modifications in the supply chain settings in order to avoid (or mitigate) the undesirable behaviors and performances. Figure 1 shows the different stages of the methodology and the general functioning is explained as follows.

The actual supply chain configuration is described through the input vector, which is fed to the behavior monitoring module (BMM). This module contains encapsulated knowledge of the supply chain dynamics previously captured by a SD model. The BMM processes the information from the input vector and anticipates the supply chain behavior for a future period of time. If the predicted behavior does not show oscillations, then no actions are needed to be carried out over the supply chain, otherwise the undesired behavior should be investigated. At this point, the causes for undesired behavior can be discovered by employing eigenvalue and elasticity analysis. To achieve this, the SD model is run with a current setting of decision and state variables. The analysis requires the computation of eigenvalues at certain points in time to figure out the structural pieces (links or loops) that have the greatest contribution to the oscillatory behavior mode. Once these structural pieces have been determined, the decision variables linked up with them are employed in the subsequent sensitivity analysis. This stage consists of systematically changing the values of decision variables connected with links or causal loops producing such undesirable behavior. The SD model is run to assess the behavior of the supply chain in the new setting. If the best setting of decision

Figure 1. General procedure of the methodology



variables is obtained, then it is implemented in the actual supply chain to ensure it is kept stable. In case the undesired behavior mode persists, the investigation continues by using eigenvalue and elasticity analysis and values of decision variables are changed. This loop continues to find the best setting for the supply chain.

Supply Chain Environment

This module represents the actual participants, structure, strategies, policies, objectives, variables, constraints, and parameters that configure different scenarios of the supply chain over time. All configurations require making different decisions that when implemented will produce changes in the behavior modes of the supply chain. The information that is entered to the BMM is a reading of the actual supply chain configuration in the form of variables that compose the input vector.

Behavior Monitoring Module

The BMM is a feedforward NN based agent that maps the current setting of the input vector and anticipates the behavior of the state vector. The latter contains more than one performance criterion of interest (i.e., assembly inventory, desired capacity, etc.) Any change in the setting of the decision variables may produce a behavioral change in the state variables of the supply chain, which is predicted and identified by the BMM. Behavior in the supply chain is referred to the observed patterns in the state variables or performance criteria of interest.

To construct the BMM module it is necessary to first capture the dynamics of the supply chain. This task can be accomplished by performing the following four steps: (1) the definition of decision, state, and input vectors; (2) the construction of a SD model of the supply chain; (3) the behavior mode classification; and finally (4) feedforward NN training.

Definition of Decision, State, and Input Vectors

The first task is to define the variables that compose the decision, state, and input vectors. The vector of decision variables (**d**) is the one that contains independent variables. Some of these variables may be of endogenous nature and hence the company has more control over them to make adjustments internally, that is, a change of the desired inventory level. Whereas other variables may be of exogenous nature, which means they are controlled mainly by the market and the competitiveness level in the supply chain environment. In this case, to produce an improvement in performance would require the company significant financial investments. On the other hand, the state vector(**s**) is defined by those variables of interest that represent an aggregate state of the supply chain. Normally, these variables correspond to criteria used to measure performance in the supply chain, that is, available capacity, desired capacity, and so forth. A modification to the values of the variables within the decision vector will generate different supply chain behavior modes in the variables within the state vector. In practice, these behavior modes may be perceived but often difficultly understood.

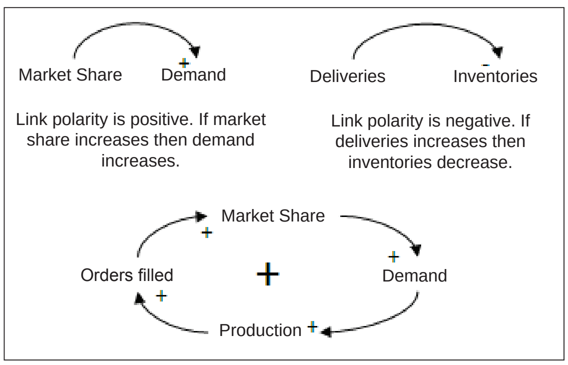
The input vector (**I**) is a composite vector used by BMM as input to predict the behavior mode of the target variables in the state vector. The composition of the input vector is formed by the decision vector (**d**), the current state vector (**s**), and the trend vector (**w**). The current state vector is the state vector but containing the current values for the state variables. The trend vector contains the values of the state variables from the last two periods. The number of periods selected for trends is not a minor issue because it influences the quality of the prediction.

SD Model of the Supply Chain

The next step in capturing the knowledge of the supply chain is to construct the SD based model. To accomplish this task there are various commercial off-the-shelf (COTS) simulation software available in the market such as Vensim, Stella, Powersim, Dinamo, and Anylogic. All of them offer some kind of good capabilities for modeling and analysis. For the purpose of this methodology, Vensim is the COTS software used because of its advanced data handling capabilities. This is a very good package for research, developing, analyzing, and packaging high quality dynamic feedback models. Models built with Vensim are easily plugged in to the Java based software package called AnalyzIt (Hines, 2001) for eigenvalue calculation and analysis.

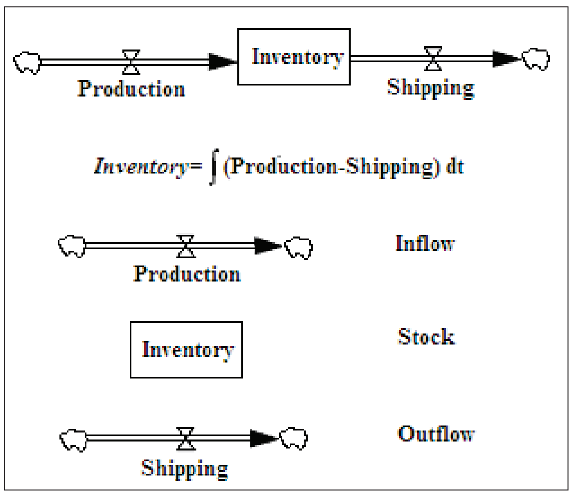
In system dynamics, the modeling is often an extensive and time consuming process. SD practitioners attempt to ease and accelerate the process by creating collections of templates or libraries of commonly used components, that is, Hines' Molecules (Hines, 2001) and Sterman's Business Dynamics templates (2000). However, those components are generic and they may not capture all the details of the supply chain environment. Therefore, if used, some of those components may need to be tailored to fit the supply chain model. The creation of a SD model of a system requires the identification of the causal and feedback loops that connect all the desired supply chain components. Feedback loops can be either negative or positive. A negative feedback loop is a series of causal relationships that tend to move the behavior toward a goal. In contrast, a positive feedback loop amplifies disturbances in the system to create even higher variations in behavior. A causal loop diagram consists of variables connected by arrows denoting the causal influence among the variables. Figure 2 shows a simple causal loop diagram and the negative and positive causal relationships.

Figure 2. Causal loop diagram and the negative and positive causal relationships



From these causal loops, the stock and flow graphical structure representation is developed (see Figure 3). Stocks are accumulations of information or physical materials that characterize the state of the system. They also create delays by accumulating the difference between the inflow and outflow of a process. Flows are rates that are added to or subtracted from a stock. This graphical description of the system based on stocks and flows is then mapped into a mathematical description of the real system. Stocks are mathematical integrations of the net difference between the inflows and outflows connected to them.

Figure 3. Stocks and flows representations in SD models

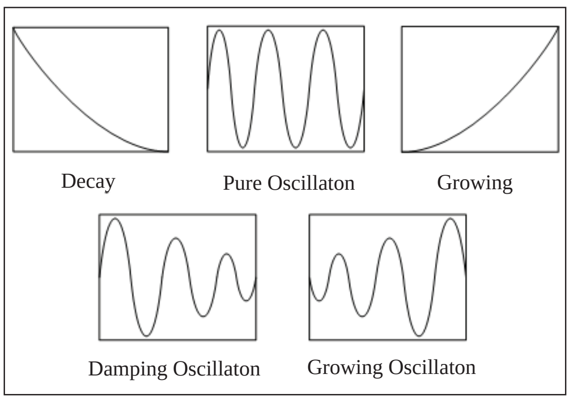


The inventory in Figure 3 represents a state variable (or stock) that accumulates the inflow of production and is reduced by the outflow of shipments. When decision variables configure different scenarios, a state variable might have different types of behavior modes. These overall behaviors can be described as a combination of elemental behavior modes such as decay, growth, pure oscillation, damping oscillation, and growth oscillation, as shown in Figure 4. Any behavior mode associated with an oscillatory pattern is an undesired behavior for any state variable in the supply chain. If it happens, actions for investigation of undesired behavior would be needed.

Once the supply chain knowledge is captured, the SD model is used to build scenarios that will allow it to train a neural network that predicts behavior modes of the state variables. An experimental design is set up to create these scenarios and obtain the behavior of the state variables with the SD model. This results in many data sets that will be used to construct the BMM.

In addition, the SD model is used as part of the methodology to assess changes in the supply chain configuration as needed. For example, if a new scenario has to be evaluated, then the new setting of the decision variables and current values of the state variables are entered to the SD model.

Figure 4. Elemental behavior modes



The model is run for a certain simulation time length and the new behavior of the state variables is obtained and assessed.

Behavior Mode Classification

Before training the feedforward NN for the BMM, the behavior modes of the state variables need to be observed and classified. The classification scheme can be made based on graphs—with particular shapes—over a future horizon in months for each state variable. The selection of this future horizon will depend on the situation under study, but a time length anywhere between 18 to 24 months is good enough as to observe behavior modes that are not initially evident and may appear several months later. The simulation of the SD model is started in equilibrium before proceeding to change the setting of the decision variables, and then the future behavior mode is observed in the target state variables during the fixed future horizon.

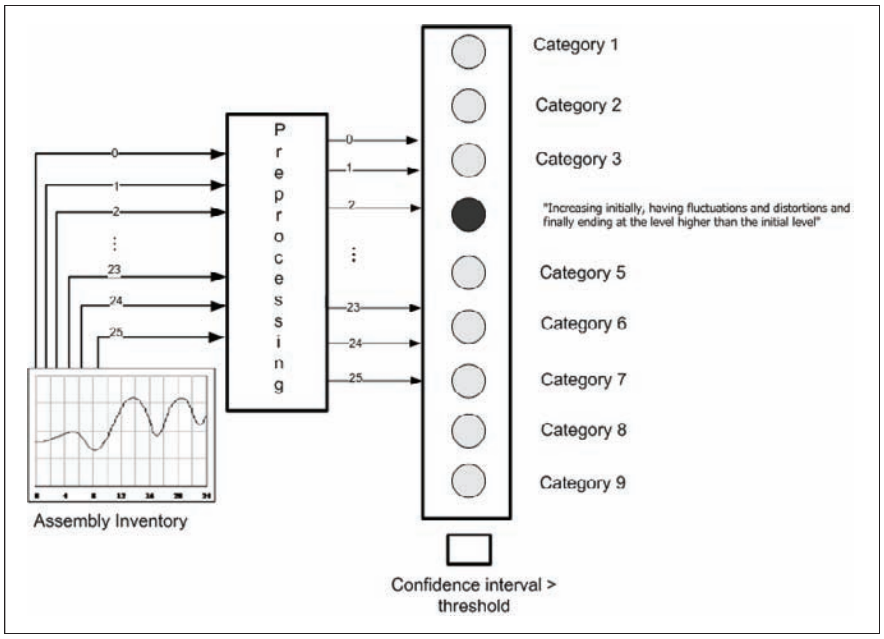
The graphs of these behavioral modes are studied by using clustering techniques based on NN. Fuzzy adaptive resonance theory (ART) NN is one popular class of algorithms due to its ability to uncover patterns (Carpenter, 1989; Carpenter, Grossberg, & Rosen, 1991.) The fuzzy ART architecture is used in applications where no output or dependent variable is known, as opposed to supervised applications—feedforward NN—where the goal is to adapt a neural network so that its actual outputs come close to some target outputs. The fuzzy ART NN is used to discover similarities in the graphs related to each state variable, which requires its own NN. As the fuzzy ART NN is exposed during training to the different graphs—or behavior modes—of different simulation scenarios—or settings of the decision vector—it begins to organize clusters of similar graphs devising a category scheme based on their shape and amplitudes. Two data sets are needed to accomplish this unsupervised training: a first set to build a classification scheme for the future behavior of the state variables, and a second set to validate the classification scheme created with the first set. Once the categories for all state variables are obtained and validated, verbalized descriptions of each category are also needed. Notice that the number of categories obtained for each variable in the state vector may vary since a different fuzzy ART NN needs to be set up.

Figure 5 shows a graph containing a behavior mode for the state variable "Assembly Inventory" that is being exposed to its respective fuzzy ART NN. The shape of this graph is classified as category 4, which is verbalized as "increasing initially, having distortions and finally ending at the level higher than the initial level."

Feedforward Neural Network Training

Since several categories for each state variable are known so far, the final step for constructing the BMM is to design and train a feedforward NN. This NN is of a supervised type of training, which means it adapts so that error between actual and target future behavior modes is minimized while different settings of the input vector are mapped to future behavior modes. To accomplish this task, three data sets are normally needed: one set is for training, a second set is for validation, and the third one is for testing. The methodology considers to select the right architecture (i.e., the number of hidden neurons) through empirical methods and to use the backpropagation training algorithm. Using this approach, NN architectures with different numbers of hidden neurons might be needed to test. Then the validation error is calculated, the architecture with minimum validation error is selected, and the number of hidden neurons is used to proceed with further evaluation and optimization using the training data set. It is known from theoretical developments and from empirical results that the generalization of the backpropagation algorithm depends on a balance between the information in the training examples and its architecture. To search for a suitable

Figure 5. Utilization of fuzzy ART for category formation using assembly inventory



learning algorithm, several algorithms such as those involving gradient descent optimization, regularization parameters, Bayesian, Levenberg-Marquardt (using second order derivatives), and conjugate gradient-based schemes may be necessary to compare.

Figure 6 schematically shows the feedforward NN architecture designed for the BMM. The input vector is composed by the decision vector, the current state vector, and the trend vector. Each input node in the figure represents a set of input nodes associated to the type of variables contained in the input vector. The layer of hidden nodes is where the process of summation and transformation of input values occurs to give as result an output or a future behavior mode for each state variable.

Eigenvalue and Elasticity Analysis

If the BMM is predicting undesired behavior for some of the state variables of interest when changing the input vector setting associated with the decision variables, then investigation is needed to identify the parameters, variables, and/or causal relationships that could be causing such undesirable behavior (oscillations.) Identifying the causes for undesired behavior modes allows how these effects can be mitigated to be established. Eigenvalue and elasticity analysis can support these investigations and are described as follows. The application of both analyses requires the use of the SD model as indicated in Figure 1.

Eigenvalue Analysis

This analysis allows identification of the elemental behavior modes that rule the SD model behavior, since each elemental behavior mode is characterized by a particular eigenvalue. The superposition of these elemental behavior modes gives rise to the observed behavior patterns of the state variables.

The eigenvalues (λ) of a square matrix $\boldsymbol{J}$ of dimension n describe the behavior modes of n state variables ($x_1, x_2, \ldots, x_n$) in a linear SD model and they are calculated as the roots of the characteristic polynomial q(λ), where q(λ)= det($\lambda \boldsymbol{I} - \boldsymbol{J}$) = 0. The eigenvalues ($\lambda_1, \lambda_2, \ldots, \lambda_n$) are scalars that belong to the set of complex numbers, which have the form $a \pm bj$, where j^2=1. The real part of the eigenvalue will determine the mode stability. A pure negative real part will cause decay or goal seeking modes (see Figure 4), whereas a pure positive eigenvalue will cause exponential growth. A pure imaginary eigenvalue will cause sustained oscillations with period equal to 2π divided by the eigenvalue. A complex eigenvalue will cause either damping or growth oscillations, depending on the sign of its real part. In all cases, the associated time constant will be the inverse of the eigenvalue's real part. Therefore, the eigenvalue analysis provides clues for understanding the oscillatory behavior and can tell us about the stability of the supply chain.

As we stated before, the eigenvalue analysis works with linear systems while supply chain models built in SD are expected to be nonlinear. A general nonlinear system can be written as $\dot{\boldsymbol{x}} = \boldsymbol{g}(\boldsymbol{x})$, where $\boldsymbol{g(x)} = (g_1(\boldsymbol{x}), g_2(\boldsymbol{x}), ..., g_n(\boldsymbol{x}))$ is a vector function, $\boldsymbol{x} = (x_1, x_2, ..., x_n)$ is the state vector, and $\dot{\boldsymbol{x}}$ its time derivative. Thus, linearization of the SD model plays an important role to analyze the system. This is done by using Taylor series approximation at the current settings of the SC. The slope of a nonlinear curve at a certain point is a good approximation of the curve over a small neighborhood of the point. Thus, a nonlinear system can be treated as linear for a small deviation about an operating point $\boldsymbol{x_0} = (x_{10}, x_{20}, ..., x_{n0})$. When the higher-order terms are neglected, the approximation for the i^{th} component of the vector $\dot{\boldsymbol{x}}$ gives:

$$\dot{x}_i = g_i(\boldsymbol{x_0}) + \frac{\partial g_i}{\partial x_1}\bigg|_{\boldsymbol{x}=\boldsymbol{x_0}} (x_1 - x_{10}) + \frac{\partial g_i}{\partial x_2}\bigg|_{\boldsymbol{x}=\boldsymbol{x_0}} (x_2 - x_{20}) + \cdots$$

$$(+\frac{\partial g_i}{\partial x_n}\bigg|_{\boldsymbol{x}=\boldsymbol{x_0}} \; x_n - x_{n0})$$

The linearized system written in matrix notation is expressed as $\dot{\boldsymbol{x}} = \boldsymbol{J}\boldsymbol{x} + \boldsymbol{B}$. Matrix $\boldsymbol{J} = \partial g_i / \partial x_j$ is a state transition matrix known as the Jacobian matrix and $\boldsymbol{B}$ is a constant vector.

The idea is to decompose the original SD model into several models at various points in time to linearize individual nonlinear static relationships. By plotting the eigenvalues of each linear model with respect to time, it is possible to identify shifts in the behavior modes of the state variables (behavior phases).

Now it is possible to extend the eigenvalue analysis that characterizes the behavior modes of the model, using the elasticity analysis to identify the structural components (links/loops) that most influence these modes of behavior.

Elasticity Analysis

At the lowest level, model relationships are defined as links between variables. To establish the relationship between the eigenvalues and the links/loops that constitute the structure of the model, it is necessary to define the units of analysis of structure as the gain of a link. The gain of the link between two variables is defined as the partial derivative of the output variable with respect to the input variable ($g_{ab} = \partial a / \partial b$). By definition, the gain of a loop is the product of the gains of the links that constitute the loop.

The eigenvalue elasticity (ε) measures how a specific eigenvalue (λ) changes with respect to changes in a specific link or loop gain (k), and it is defined as:

$$\varepsilon = \frac{\partial\lambda / \lambda}{\partial k / k} = \frac{\partial(\ln\lambda)}{\partial(\ln k)}$$

This measure helps identify the loops that contribute most significantly to the model behavior. Therefore, the eigenvalue and elasticity analysis permits to examine the structure-behavior relationship in a SD model to establish the causes for a particular behavior mode of state variables. Because behavior mode evolves as a consequence of the dynamic interactions among several feedback loops of the model, these analyses can give valuable insights into how to design or redesign the SC to obtain the desired behavior.

Sensitivity Analysis and Optimization

After knowing the structures in the SD model that are producing undesired behavior, the logical step is to perform sensitivity analysis. The main goal of the sensitivity analysis is to mitigate the effects of undesired behavior and to evaluate the robustness of the system. The sensitivity analysis is done by changing the values of the variables in the decision vector to assess thorough the SD model if the oscillatory modes can be avoided or reduced in the supply chain. The systematic experimentation with the decision variables is done by ranking them prior in nondecreasing order according to investment cost and implementation time required to execute the change needed in reality. For instance, changing the time to adjust inventory policies is less expensive than changing the manufacturing cycle time. Changing the policies of adjusting the inventory levels can be done by training and a series of meetings to coordinate among the different facilities and production personnel. On the other hand, changing the manufacturing cycle time can be obtained by investing in new equipment, or maybe implementing a Six-Sigma program, and so forth.

Advanced approaches, like optimization techniques, can be considered for the idea of searching several settings of decision variables to achieve robustness in the system. Optimization techniques aim to find the best setting of the decision variables to keep the supply chain robust over time. Since a small subset of loops may be sufficient to uniquely describe eigenvalues of a system dynamics model (Kampmann, 1996), decision variables that belong to this small subset can be used to search for system stability. The utilization of metaheuristic based searching algorithms may have a great potential, although the cost/benefit ratio for the automation of such algorithms linked to a SD model is something still not thoughtfully assessed.

CASE STUDY: LSMC'S SUPPLY CHAIN

The Environment

Leading semiconductor manufacturing company (LSMC) is an electronics manufacturing company that supplies products for personal computers (PCs) to original equipment manufacturers (OEMs) such as Dell, Gateway, and Hewlett-Packard. LSMC was facing a problem of persistent oscillations in its finished goods inventory and desired capacity. Since 1998, many OEMs have changed their strategies by adopting built-to-order (BTO) and just-in-time (JIT) processes. These changes in PCs in addition to their short life cycles have amplified the coordination problems in the company's supply chain, which in turn has caused excess inventories and sometimes difficulties to keep up with demand. Moreover, the competition has forced the company to introduce more product varieties at lower prices into the market to protect its existing and potential market share. Production capacity is another factor that adds to supply chain complexity because its long delays, huge investments, and new products with more complex manufacturing processes that previous generations. In addition, these complementary PC products are at the upstream of the supply chain for PCs and their resulting fluctuations are higher.

A brief summary is presented here with the analysis conducted for LSMC's supply chain using the proposed methodology to detect and mitigate oscillations due to an unexpected change in demand. For additional information on this case study, the reader is referred to Lertpattarapong (2002).

The Behavior Monitoring Module

Definition of the Input Vector

Several participants (at different levels of the managerial hierarchy) from various departments that have been involved in the supply chain at LSMC were interviewed to identify the relevant parameters and variables to the company's supply chain operations. Based on that information, the three components of the input vector were defined, that is, the decision vector, the state vector, and the trend vector. The decision vector **d** contains seventeen parameters, the state vector **s** has seven variables, and the trend vector **w** has fourteen variables. The complete definition of the vectors is shown next.

$$d = \begin{bmatrix} \text{Manufacturing Cycle Time} \\ \text{Minimum Order Processing Time} \\ \text{Time to Complete Assembly} \\ \text{Time to Adjust Backlog} \\ \text{Time to Perceive Present Demand} \\ \text{Capacity Acquisition Delay} \\ \text{Safety Stock Coverage} \\ \text{Forecast Horizon} \\ \text{Backlog Switch} \\ \text{Line Yield} \\ \text{Component per Lot Yield} \\ \text{Time to Adjust Assembly Inventory} \\ \text{Pre - assembly Inventory Adjustment Time} \\ \text{Time to Adjust Finished Goods Inventory} \\ \text{Time to Update Channel Orders} \\ \text{Competitor's Attractiveness} \\ \text{Channel Demand} \end{bmatrix}$$

$$\mathbf{w}(t) = \begin{bmatrix} \text{Historical Demand}(t-1) \\ \text{Available Capacity}(t-1) \\ \text{Desired Capacity}(t-1) \\ \text{Pre-assembly Inventory}(t-1) \\ \text{Assembly Inventory}(t-1) \\ \text{Finished Goods Inventory}(t-1) \\ \text{Channel Order Backlog}(t-1) \\ \text{Historical Demand}(t-2) \\ \text{Available Capacity}(t-2) \\ \text{Desired Capacity}(t-2) \\ \text{Pre-assembly Inventory}(t-2) \\ \text{Assembly Inventory}(t-2) \\ \text{Finished Goods Inventory}(t-2) \\ \text{Channel Order Backlog}(t-2) \end{bmatrix}$$

$$\mathbf{s}(t) = \begin{bmatrix} \text{Historical Demand}(t) \\ \text{Available Capacity}(t) \\ \text{Desired Capacity}(t) \\ \text{Pre-assembly Inventory}(t) \\ \text{Assembly Inventory}(t) \\ \text{Finished Goods Inventory}(t) \\ \text{Channel Order Backlog}(t) \end{bmatrix}$$

The current state vector and the trend vector were set to a specific point in time when the behavior mode was predicted with the feedforward NN.

LSMC's SD Supply Chain Model

The SD model for LSMC supply chain was built by using the Vensim simulation package. The first step was to create the causal loop diagram which consisted of several hypotheses (see Figure 7). A description of the principal hypotheses is provided below (Lertpattarapong, 2002).

- **Market share needs production capacity:** Capacity increase means more orders and this translates into more future demand. But if demand increases beyond capacity customers will become unsatisfied because of delivery delays and might shift to competitors.
- **Investments in production capacity depend on market share:** Higher revenues are realized with increases in market share and then more investments can be made on production capacity.
- **Competition increases with the increase in profits and vice versa:** Higher profits create an environment that attracts new entrants to the market and hence competition and lower profits create less motivation for new entrants. In addition, the market growth increases the competition.
- **Growth, new product development, and product life cycle:** Higher revenues are realized with increases in market share and then more investments can be made on new product developments. The increase in new product development could lead to the obsolescence of old products.

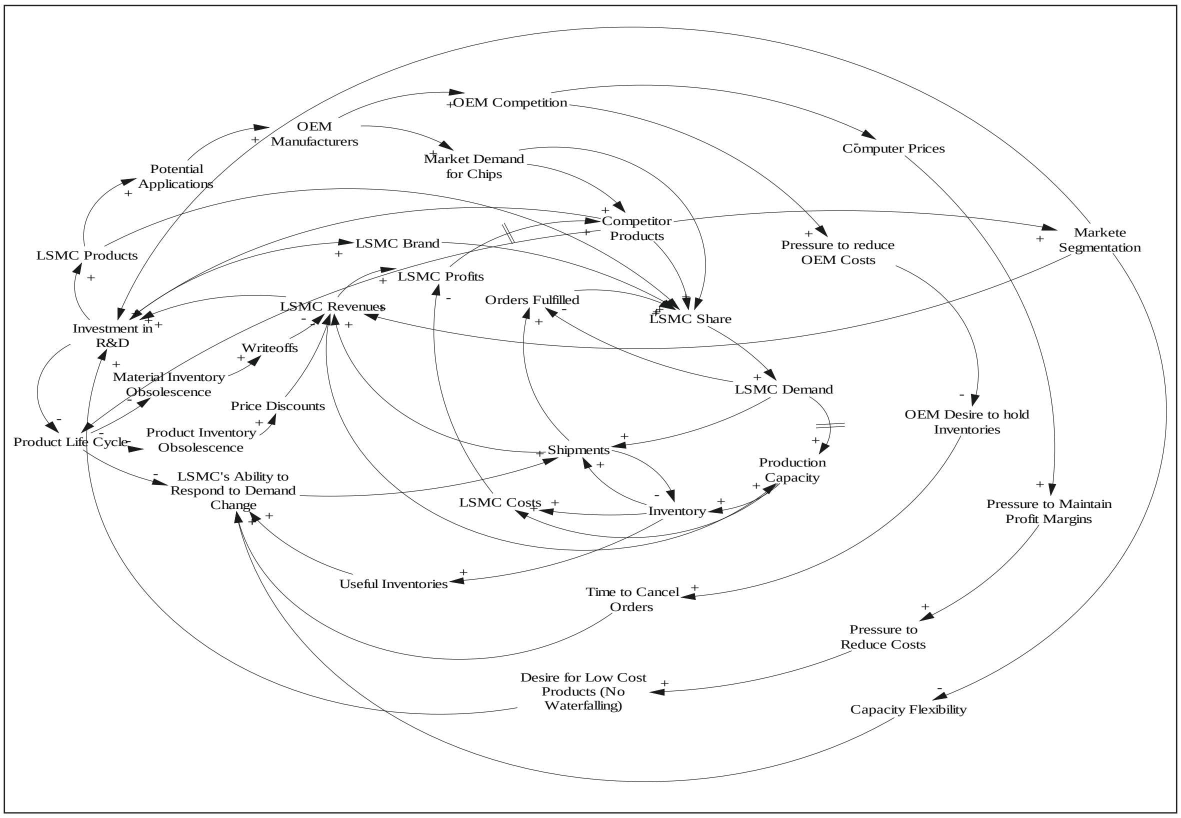
The following step was to convert the casual loop diagram into stock and flow diagrams and defining the mathematical formulation. The complete model has more than 91 equations, including differential and auxiliary, and it is comprised of three connected stock and flow submodels: (1) the production model, (2) the market share and shipment model, and (3) the demand forecast and capacity model.

Behavior Mode Classification

The SD model was run for 24 months generating five sets each with 800 different scenarios. The number of 800 different combinations was provided by following an estimate of the prediction risk as provided by Akaike's final prediction error (Akaike, 1970; Moody, 1994; Moody & Utans, 1994).

From the first set of simulation scenarios, the 800 graphs of each state variable were exposed to the fuzzy ART NNs. The first fuzzy ART NN is used for the Historical Demand, which was able

Figure 7. Causal loop diagram of LSMC's SC



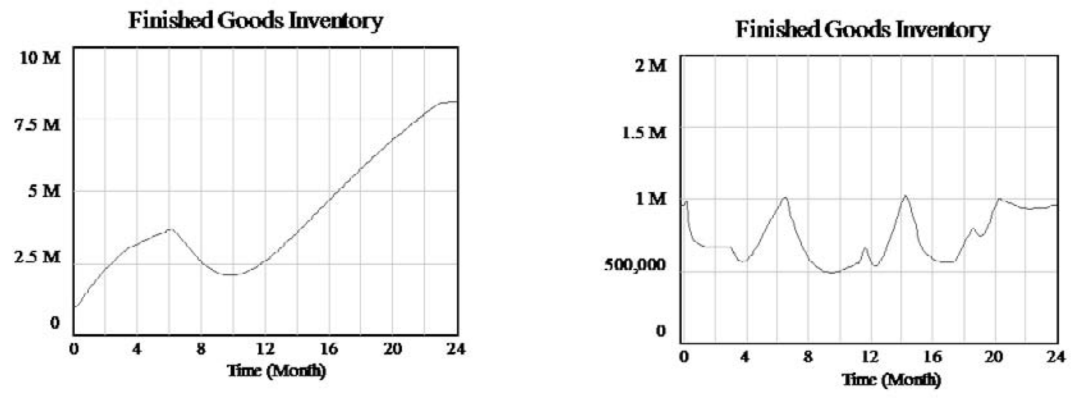
to develop nine stable and different categories of behavior. A second fuzzy ART NN was used for the available capacity, generating eleven stable and different categories. Similarly, four categories for the desired capacity, eight categories for the preassembly inventory, nine categories for the assembly inventory, six categories for the finished goods inventory, and six categories for the channel order backlog were generated.

The different categories of each state variable were validated using the second set of 800 samples. The validation of the categories using this data set was 100% correct. Figure 8 shows two possible behavior categories of the Finished Goods Inventory variable.

Feedforward Neural Network Training

Before starting the training process, the input data and behavior classifications (output) were properly scaled and preprocessed. The last three sets of 800 samples each were used for training, validation, and testing. The backpropagation NN was trained using different architectures and learning algorithms (as the ones mentioned in the methodology). The Levenberg-Marquardt algorithm, which provided the most reliable and fastest training option, was used to select the best architecture. NN architectures with 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 30, and 40 hidden neurons were evaluated. The architecture with five hidden neurons showed the minimum validation error and was selected for further analysis. This architecture was then tested using the testing data set and the final testing error was considerably smaller. The qualitative result based on the category selected, not the root mean

Figure 8. Behavior categories of finished goods inventory



Category 1: "Increasing initially, fluctuating, and ending above the initial level"

Category 2: "Decreasing initially, fluctuating, and ending up at the same initial level"

squared error, indicated a very high performance in testing (~99% accuracy).

Detecting Undesired Behavior

As an example, while the system was in equilibrium with no oscillations, the channel demand experienced a sudden increase of 10% of its value at the six month. The resulting behavior of the different inventories was oscillatory, as shown in Figure 9, for the Finished Goods Inventory, starting at the eighth month (two months after the 10% increase).

NNs were applied at the seventh month, setting the current state vector and the trend vector for t=7. These NNs were able to predict that the behavior of the supply chain would be of the "oscillatory with undesired amplitudes" category.

Eigenvalue & Elasticity Analysis of Supply Chain Behavior

The original model was decomposed into 11 nonlinear models, considering the state variables in Table 1.

The last three variables (the three versions of the perceived fraction order filled) in the list were required to represent the smooth function (the third order exponential smooth function) that was used to capture the perceptions of the electronics company for the fraction of orders filled. These models superposed in time reproduced exactly the behavior of the original model. However, because the emergent behavior in each of these models was still nonlinear, they were linearized by using Taylor expansion for the current settings of the supply chain. The analysis clearly shows the oscillations in the systems behavior due to its current settings.

Sensitivity Analysis/Optimization

In order to avoid or mitigate the oscillatory behavior, several experiments were performed by

Table 1. State variables for original model

1. Available Capacity	7. Pre-Assembly Inventory
2. Channel Order Backlog	8. Finished Goods Inventory
3. Expected Channel Demand	9. Perceived LSMC Fraction Orders Filled
4. Historical Demand	10. Perceived LSMC Fraction Orders Filled 1
5. Perceived Present Demand	11. Perceived LSMC Fraction Orders Filled 2
6. Available Inventory	

Figure 9. Finished goods inventory oscillations

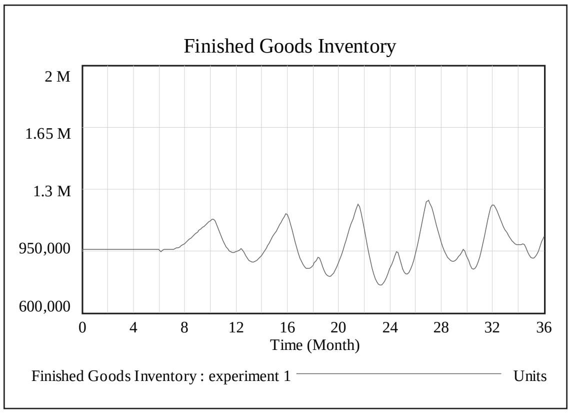
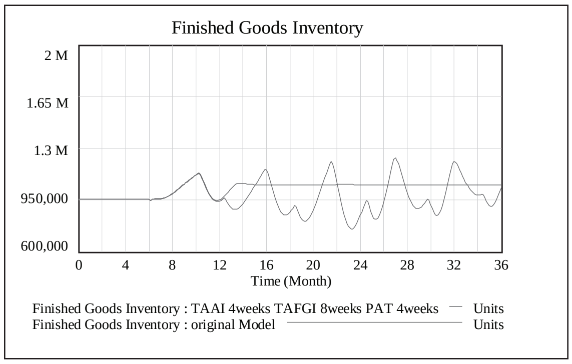


Figure 10. Finished goods inventory with and without oscillations



changing different parameters from the decision vector. In this case, just sensitivity analysis was used through experimental design, but optimization techniques can be also used. The following parameters were selected: time to adjust assembly inventory (TAAI), time to adjust finished goods inventory (TAFGI), and preassembly adjustment time (PAT). The best combination that mitigated the undesirable behavior was to set TAAI to four weeks, TAFGI to 8 weeks, and PAT to four weeks. These changes do not only mitigated the oscillations, but also made the system more robust to perturbations. Figure 10 shows that oscillations were eliminated in the finished goods inventory behavior by carefully setting these times.

It is concluded that the oscillatory behavior is endogenous. Internal actions, including adjusting the times to update production inventories such as PAT, TAAI, and TAFIG, could eliminate or minimize the oscillations. Lertpattarapong (2002) reports that this result was not expected by management, who believed that the main causes for their problems were exogenous and not under the company's control.

CONCLUSION

Companies are discovering that effective SCM is having a tremendous impact to increase profit and market share. The SCM solutions available today do not have the capability to detect changes taking place in the business environment and hence are not able to provide the companies with accurate predictions for the effects of these changes. This inefficient SCM can cause numerous problems, such as excessive inventory investment, poor customer service, lost revenues, erroneous capacity plans, and so forth, which cost companies millions of dollars.

Effectively managing a supply chain requires visibility to detect unexpected variations at an early stage. The proposed methodology uses SD to model dynamic behavior of the supply chain. The "systems thinking" approach provided by the SD discipline can improve managers' mental models of their organizations and allows them to make better decisions. Neural networks were used to capture the knowledge of the SD model and make it available to the enterprise in real-time fashion to detect changes and predict the future behavior of the supply chain. Since supply chain models are nonlinear, model decomposition and linearization is used to simplify the structure of the model prior to the eigenvalue analysis. The eigenvalue analysis could be used to make an in-depth analysis of the changes in the supply chain behavior.

The methodology suggested here can contribute to assist in implementing Six-Sigma programs, improved forecasts, and other management initiatives as well. Most important, it will allow the analysis of planning strategies to design robust supply chains that can effectively cope with significant changes and disturbances, with the corresponding cost savings to the companies.

The methodology introduced in this chapter contributes to the integration SD models, neural networks, optimization techniques, model-predictive control, and encapsulation of knowledge for a more effective and efficient SCM.

FUTURE RESEARCH DIRECTIONS

The proposed methodology contributes to the advancement of prediction and mitigation of undesirable supply chain behavior within short- and long-term horizons by promoting a better understanding of the structure that determines the behavior modes through the integration of tools such as system dynamics, neural networks, model-predictive control, and sensitivity analysis/optimization. However, great room for research lies ahead for an adequate integration of all these tools. The literature on NN shows that this approach has been applied to monitor and predict disturbances in engineering and business applications with great success. NN simulation metamodels have been incorporated recently in SCM to predict oscillatory behavior. Although they have demonstrated to be efficient capturing the behavior of a complex supply chain, still, there is a need for a generic methodology that can be applied to almost any supply chain. Therefore the question is which type of classification methods and NN topologies are the most appropriate to produce best results.

From other side, a necessity is evident for a more extensive test of the eigenvalue/elasticity analysis methodology in large-scale and highly nonlinear models, which is the case of actual supply chains. Even more, according to literature it is still missing a fully computerized implementation of eigenvalue/elasticity analysis routines. The achievement of this task might impulse the research on several techniques to identify causal structures responsible for oscillatory behavior. In addition, it is clear the lack of a methodology that connects the model optimization and the eigenvalue/elasticity analysis technique. Regarding this last point, the eigenvalue optimization problem consists of determining the optimal parameters of a linear or linearized dynamic system in such a way that the eigenvalues are moved towards the left side in the complex plane as far as possible without violating the given constraints of the problem. Literature shoes the lack of an optimization methodology based on simulation and evolutionary algorithms that use nonlinear optimal control theory to eliminate instability of the supply chain to produce robust policies. Preliminary results of using optimization techniques based on genetic algorithms to find the best setting of the supply chain parameters that minimize the oscillations is reported by Sarmiento, Rabelo, and Moraga

(2007). However, the complexity of generalizing the stability criterion for nonlinear dynamic systems makes it difficult to propose a generic methodology capable to optimize the behavior of any SD model.

Finally, it is authors' hope that the ideas stated in this chapter will encourage researchers to advance this fascinating field of monitoring and analyzing the supply chain behavior.

REFERENCES

Affeldt, J.F. (1999). The application of system dynamics (SD) simulation to enterprise management. In P.A. Farrington, H.B. Nembhard, D.T. Sturrock, & G.W. Evans (Eds.), *Proceedings of the 1999 Winter Simulation Conference* (pp. 1496-1500). NJ: Institute of Electronics and Electrical Engineers.

Akaike, A. (1970). Statistical predictor identification. *Annals of the Institute of Statistics and Mathematics*, *22*, 203-217.

Anderson, J., & Marklund, J. (2000). Decentralized inventory control in a two level distribution system. *European Journal of Operational Research*, *127*, 483-506.

Angerhofer, B., & Angelides, M. (2000). Systems dynamics modeling in supply chain management: research review. In J.A. Joines, R.R. Barton, K. Kang, & P.A. Fishwick (Eds.), *Proceedings of the 2000 Winter Simulation Conference* (pp. 342-351). NJ: Institute of Electronics and Electrical Engineers.

Barlas, Y. (1994). Model validation in system dynamics. In *Proceedings of the 1994 International System Dynamics Conference*. Albany, NY: System Dynamics Society.

Beamon, B.M. (1998). Supply chain design and analysis: models and methods. *International Journal of Production Economics*, *55*, 281-294.

Bose, S., & Pekny, J.F. (2000). A model predictive framework for planning and scheduling problems: A case study of consumer goods supply chain. *Computers and Chemical Engineering*, *24*, 329-335.

Bruniaux, R., Pierreval, H., & Caux, C. (1999). *Simulating supply chains with system dynamics.* Paper presented at the 1999 European Simulation Symposium (pp. 209-213). CA: The Society for Modeling and Simulation International.

Bruzzone, A., & Orsoni, A. (2003). AI and simulation-based techniques for the assessment of supply chain logistic performance. In *Proceedings of the 36th Annual Simulation Symposium*, Los Alamitos, California (pp. 151-161). IEEE Computer Society Press.

Carpenter, G. (1989). Neural network models for pattern recognition and associative memory. *Neural Networks*, *24*, 243-257.

Carpenter, G., Grossberg, S., & Rosen, D. (1991). *Fuzzy ART: An adaptive resonance algorithm for rapid, stable classification of analog patterns* (Tech. Rep. CAS/CNS-91-006). Paper presented at the International Joint Conference on Neural Networks (pp. 411-416). NY: IEEE Press.

Chopra, S., & Meindl, P. (2007). *Supply chain management: Strategy, planning, & operation* (3rd ed.). NJ: Prentice Hall.

Choy, K., Lee, W., & Lo, V. (2003). Design of an intelligent supplier relationship management system: A hybrid case based neural network approach. *Expert Systems with Applications*, *24*(2), 225-237.

Coyle, J.M., Exelby, D., & Holt, J. (1999). System dynamics in defense analysis: some case studies. *Journal of the Operational Research Society*, *50*, 372-382.

Dangerfield, B.C. (1999). System dynamics applications to European health care issues. *Journal of the Operational Research Society*, *50*, 345-353.

Dejonckheere, J., Dysney, S.M., Lambrecht, M.R., & Towill, D.R. (2003). Measuring and avoiding the bullwhip effect: A control theoretic approach. *European Journal of Operational Research*, *147*, 567-590.

Ezingeard, J., & Race, P. (1995). Spreadsheet simulation aid capacity management of batch chemical processing using JIT pull control. *International Journal of Operations & Production Management*, *15*(10), 82-88.

Forrester, J. W. (1961). *Industrial dynamics*. Cambridge, MA: Productivity Press.

Forrester, J. W. (1969). *Urban dynamics*. Cambridge, MA: Productivity Press.

Forrester, N. (1982). *A Dynamic synthesis of basic macroeconomic theory: Implications for stabilization policy analysis*. Unpublished doctoral dissertation, Massachusetts Institute of Technology.

Gonçalves, P. (2006). Eigenvalue and eigenvector analysis of dynamic systems. In *Proceedings of the 2006 International System Dynamics Conference*. Albany, NY: System Dynamic Society.

Gonçalves, P., Lertpattarapong, C., & Hines, J. (2000). Implementing formal model analysis. In *Proceedings of the 2000 International System Dynamics Conference*. Albany, NY: System Dynamic Society.

Güneralp, B. (2005). Towards coherent loop dominance analysis: Progress in eigenvalue elasticity analysis. In *Proceedings of the 2005 International System Dynamics Conference*. Albany, NY: System Dynamic Society.

Harvey, J.T. (2006), Modeling interest rate parity: A system dynamics approach. *Journal of Economic Issues*, *40*(2), 395-403.

Hayden, F.G. (2006). The inadequacy of Forrester system dynamics computer programs for institutional principles of hierarchy, feedback, and openness. *Journal of Economic Issues*, *40*(2), 527-535.

Higuchi, T., & Troutt, M.D. (2004). Dynamic simulation of the supply chain for a short life cycle product: Lessons from Tamagotchi case. *Computers and Operations Research*, *31*, 1097-1114.

Hines, J. (2001). Analyzit: Link Elasticity Software [computer software]. Cambridge, MA.

Jayaraman, V., & Pirkul, H. (2001). Planning and coordination of production and distribution facilities for multiple commodities. *European Journal of Operational Research*, *133*, 394-408.

Kampmann, C.E. (1996). Feedback loop gains and system behavior. In *Proceeding of the 1996 International System Dynamics Conference Boston*. Albany, NY: System Dynamic Society.

Kampmann, C.E., & Oliva, R. (2006). Loop eigenvalue elasticity analysis: Three case studies. *System Dynamics Review*, *22*(2), 141-162.

Kleijnen, J. (2005). Supply chain simulation tools and techniques: A survey. *International Journal of Simulation and Process Modeling*, *1*(1-2), 82-89.

Kleijnen, J. & Smits, M. (2003). Performance metrics in supply chain management. *Journal of the Operational Research Society*, *54*(5), 507-514.

Lakhal, S., Martel, A., Kettani, O., & Oral, M. (2001). On the optimization of supply chain networking decisions. *European Journal of Operational Research*, *129*, 259-270.

Lancioni, R., Smith, M., & Oliva, T. (2000). The role of Internet in supply chain management. *Industrial Marketing Management*, *29*, 45-46.

Lane, D.C., Monefeldt, C., & Rosenhead, J.V. (2000). Looking in the wrong place for healthcare improvements: A system dynamics study of an accident and emergency department. *Journal of the Operational Research Society*, *51*, 518-531.

Lertpattarapong, C. (2002). *Applying system dynamics approach to the supply chain management problem*. Unpublished master's thesis, Massachusetts Institute of Technology.

Leung, H. (1995). Neural networks in supply chain management. In *Proceedings of the 1995 International Engineering Management Conference* (pp. 347-352). NJ: Institute of Electronics and Electrical Engineers.

Lin, C.-H., Tung, C.-M., & Huang, C.-T. (2006). Elucidating the industrial cluster effect from a *system dynamics* perspective. *Technovation, 26*(4), 473-482.

Lyneis, J.M. (1980). *Corporate planning and policy design*. Cambridge, MA: Productivity Press.

Mojtahedzadeh, M.T. (1997). *A path taken: Computer assisted heuristics for understanding dynamic systems.* Unpublished doctoral dissertation, University at Albany.

Mojtahedzadeh, M.T., Andersen, D.F., & Richardson, G.P. (2004). Using Digest to implement the pathway participation method for detecting influential system structure. *System Dynamics Review*, *20*(1), 1-20.

Moody, J. (1994). *The Effective number of parameters: An analysis of generalization and regularization*. In J. Moody, S. Hanson, & R. Lippmann (Eds.), *Advances in neural information processing system* 4 (pp. 847-854). San Mateo, CA: Morgan Kaufman Publishers.

Moody, J., & Utans, J. (1994). Principled architecture selection for neural networks: application to corporate bond rating prediction. In J. Moody, S. Hanson, & R. Lippmann (Eds.), *Advances in neural information processing dystem* 4 (pp. 683-690). San Mateo, CA: Morgan Kaufman Publishers.

Morecroft, J. (1983). Managing product lines that share a common capacity base. *Journal of Operations Management*, *3*(2), 57-66.

Morecroft, J.D.W., Lane, D.C., & Viita, P.S. (1991). Modelling growth strategy in a biotechnology startup firm. *System Dynamics Review*, *7*(2), 93-116.

Oliva, R. (2003). Model calibration as a testing strategy for system dynamics models. *European Journal of Operational Research*, *151*, 552-568.

Oliva, R. (2004). Model structure analysis through graph theory: Partition heuristics and feedback structure decomposition. *System Dynamics Review*, *20*(4), 313-336.

Otto, A., & Kotzab, H. (2003). Does supply chain management really pay? Six perspectives to measure the performance of managing a supply chain. *European Journal of Operational Research*, *144*, 306-320.

Persson, F., & Olhager, J. (2002). Performance simulation of supply chain designs. *International Journal of Production Economics*, *77*, 231-245.

Rabelo, L., Helal, M., Lertpattarapong, C., Moraga, R., & Sarmiento, A. (in press). Using system dynamics, neural nets, and eigenvalues to analyse supply chain behaviour: A case study. *International Journal of Production Research*.

Richardson, G.P. (1995). Loop polarity, loop dominance, and the concept of dominant polarity. *System Dynamics Review*, *11*(1), 67-88.

Richardson, K. (2003). On the limits of bottom-up computer simulation: Towards a nonlinear modeling culture. In *Proceedings of the 36th Hawaiian International Conference on System Sciences*. Los Alamitos, CA: IEEE Computer Society Press.

Riddalls, C.E., Bennett, S., & Tipi, N.S. (2000). Modelling the dynamics of supply chains. *International Journal of Systems Science*, *31*(8), 969-976.

Rodrigues, A.G., & Williams, T.M. (1998). System dynamics in project management: Assessing the impacts of client behaviour on project performance. *Journal of the Operational Research Society*, *49*, 2-15.

Saleh, M., & Davidsen, P. (2001). The origins of behavior patterns. In *Proceedings of the 2001 International System Dynamics Conference*. Albany, NY: System Dynamic Society.

Saleh, M., Davidsen, P., & Bayoumi, K. (2005). A comprehensive eigenvalue analysis of system dynamics models. In *Proceedings of the 2005 International System Dynamics*. Albany, NY: System Dynamic Society.

Santos, S., Belton, V., & Howick, S. (2002). Adding value to performance measurement by using system dynamics and multicriteria analysis. *International Journal of Operations & Production Management*, *22*(11), 1246-1272.

Sarmiento, A., Rabelo, L., & Moraga, R. (2007). *A methodology for detecting and minimizing the oscillations in the supply chain* (Working Paper).

Shapiro, J.F. (2001). *Modeling the supply chain*. Pacific Grove, CA: Duxbury Press.

Smith, G. (2003). Using integrated spreadsheet modeling for supply chain analysis. *Supply Chain Management: An International Journal*, *8*(4), 25-290.

Sparling, D. (2002). Simulations and supply chains: Strategies for teaching supply chain management. *Supply Chain Management: An International Journal*, *7*(5), 334-342.

Spengler, T., & Schröter, M. (2003). Strategic management of spare parts in closed-loop supply chains: A system dynamics approach. *Interfaces*, *33*(6), 7-17.

Sterman, J.D. (1989). Misperceptions of feedback in dynamic decision making. *Organizational Behavior and Human Decision Processes*, *43*(3), 301-335.

Sterman, J.D. (2000). *Business dynamics: Systems thinking and modeling for a complex world*. Boston: McGraw-Hill.

Stratman, J., Roth, A., & Gilland, W. (2004). The deployment of temporary production workers in assembly operations: A case study of the hidden costs of learning and forgetting. *Journal of Operations Management*, *21*(6), 689-707.

Terzi, S., & Cavalieri, S. (2004). Simulation in the supply chain context: A survey. *Computers in Industry*, *53*, 3-16.

Towill, D.R., Naim, M.M., & Wikner, J. (1992). Industrial dynamics simulation models in the design of supply chains. *International Journal of Physical Distribution & Logistics Management*, *22*(5), 3-13.

Wikner J., Towill, D.R., & Naim, M. (1991). Smoothing supply chain dynamics. *International Journal of Production Economics*, *22*(3), 231-248.

Zhang, Y., & Dilts, D. (2004). System dynamics of supply chain network organization structure. *Information Systems and E-Business Management*, *2*(2-3), 187-206.

ADDITIONAL READING

Akkermans, H., Bogerd, P., & Vos, B. (1999). Virtuous and vicious cycles on the road towards international supply chain management. *International Journal of Operations and Production Management*, *19*(5-6), 565-581.

Akkermans, H., & Dellaert, N. (2005). The rediscovery of industrial dynamics: The contribution of system dynamics to supply chain management in a dynamic and fragmented world. *System Dy-*

namics Review, 21(3), 173-186.

Balakrishnan, A., Geunes, J., & Pangburn, M.S. (2004). Coordinating supply chains by controlling upstream variability propagation. *Manufacturing and Service Operations Management*, *6*, 163-183.

Eberlein, R.L. (1989). Simplification and understanding of models. *System Dynamics Review, 5*(1), 51-68.

Ford, D. (1999). A behavioral approach to feedback loop dominance analysis. *System Dynamics Review*, *15*(1), 3-36.

Forrester, J. (1971). *Principles of systems.* Cambridge, MA: Wright-Allen Press.

Geert, H., Slinger, J., Güneralp, B., & van Daalen, C. (2006). Investigating model behavioral analysis: A critical examination of two methods. In *Proceedings of the 2006 International System Dynamics Conference*. Albany, NY: System Dynamic Society.

Georgiadis, P., & Vlachos, D. (2004). The effect of environmental parameters on product recovery. *European Journal of Operational Research*, *157*(2), 449-464.

Gonçalves, P., Hines, J., & Sterman, J. (2005). The impact of endogenous demand on push-pull production systems. *System Dynamics Review, 21*(3), 187-216.

Güneralp, B., & Getner, G. (2006). Feedback loop dominance analysis of two tree mortality models. In *Proceedings of the 2006 International System Dynamics Conference.* Albany, NY: System Dynamic Society.

Hao, G., Shang, J.-S., & Vargas, L.G. (1995). A neural network model for on-line control of flexible manufacturing systems. *International Journal of Production Research*, *33*(10), 2835-2854.

Kilmer, R.A., Smith, A., & Shuman, L.J. (1997). An emergency department simulation and neural network metamodel. *Journal of the Society for Health Systems*, *5*(3), 63-79.

Lau, H.C.W., Hui, I.K., Chan, F.T.S., & Wong, C.W.Y. (2002). Monitoring the supply of products in a supply chain environment: a fuzzy neural approach. *Expert Systems*, *19*(4), 235-243.

Minegishi, S., & Thiel, D. (2000). System dynamics modeling and simulation of a particular food supply chain. *Simulation Practice and Theory*, *8*(5), 321-339.

Nasereddin, M., & Mollaghasemi, M. (1999). The development of a methodology for the use of neural networks and simulation modeling in system design. In P.A. Farrington, H.B. Nembhard, D.T. Sturrock, & G.W. Evans (Eds.), *Proceedings of the 1999 Winter Simulation Conference* (pp. 537-542). Phoenix, AZ: Institute of Electronics and Electrical Engineers.

Richardson, G.P. (1986). Dominant structure. *System Dynamics Review*, *2*(1), 68-75.

Saleh, M., Oliva, R., Davidsen, P., & Kampmann, C. (2006). Eigenvalue analysis of system dynamics models: Another approach. In *Proceedings of the 2006 International System Dynamics Conference*. Albany, NY: System Dynamic Society.

Sterman, J. (2006). Operational and behavioral causes of supply chain instability. In O. Carranza & F. Villegas (Eds.), *The bullwhip effect in supply chains* (pp. 17-56): Basingtoke: Palgrave Macmillan.

Chapter XI
Decision Support System for Project Selection

Prasanta Kumar Dey
Aston University, UK

ABSTRACT

The evaluation and selection of industrial projects before investment decision is customarily done using marketing, technical, and financial information. Subsequently, environmental impact assessment and social impact assessment are carried out mainly to satisfy the statutory agencies. Because of stricter environment regulations in developed and developing countries, quite often impact assessment suggests alternate sites, technologies, designs, and implementation methods as mitigating measures. This causes considerable delay to complete project feasibility analysis and selection as complete analysis requires to be taken up again and again until the statutory regulatory authority approves the project. Moreover, project analysis through the above process often results in suboptimal projects as financial analysis may eliminate better options as more environment friendly alternative will always be cost intensive. In this circumstance, this study proposes a decision support system which analyses projects with respect to market, technicalities, and social and environmental impact in an integrated framework using analytic hierarchy process, a multiple attribute decision-making technique. This not only reduces duration of project evaluation and selection, but also helps select an optimal project for the organization for sustainable development. The entire methodology has been applied to a cross-country oil pipeline project in India and its effectiveness has been demonstrated.

INTRODUCTION

Many industries are in a period of rapid change brought about by technological breakthrough. Improvements in communications and networking technologies have allowed many companies to expand their operations globally. Advances in computer technologies are changing the management philosophies of organizations. The explosive growth of the Internet and the World Wide Web is creating virtual organizations and changing the whole concepts of management. Customers increasingly expect products and services of higher quality at lower price and with quicker delivery.

Projects transform organizations' vision into reality. To remain competitive in this globalized business environment, organizations should select and implement right projects efficiently. Major projects can apply science and technology in a sustainable manner but in many instances adversely affect their environment. Socioeconomic impacts of projects can occur at all the four stages of project life: preconstruction (planning/policy development); construction (implementation); operation and maintenance; and decommissioning (abandonment). (Ramanathan & Geetha, 1998).

Customarily, the search for promising project ideas is the first step towards identifying promising projects. Identification of such opportunities requires imagination, sensitivity to environmental changes, and realistic assessment of what the organization can do. The strengths, weaknesses, opportunities, and threats analysis is used by many organizations for identifying projects. Then the projects are prioritized for investment decision. Subsequently, the market analysis for the project is taken up to decide the throughput for the project in line with projected supply and demand scenario. On the basis of the planned throughput, preliminary engineering and design is carried out, which forms the basis for technical analysis. Technical analysis identifies a few feasible project alternatives with respect to project sites, throughput, technology, materials usage, product/service mix, and implementation method depending on type of projects. Subsequent financial analysis determines the most optimum project for investment. While selecting the optimal project for investment, financial analysis also considers uncertainties of the project and suggests mitigating measures. Due to increasing concern of the project affected people (PAP) and statutory environmental regulatory authorities, all projects are require to get environmental and social clearances before receiving approval from competent authorities for implementation (Calvin & Dey, 2002). Accordingly, an environmental and social impact assessment study is required to be taken to determine the positive and negative impact of projects on environment and to develop measures for mitigating the negative impacts. The outcome of the project feasibility analysis is a feasibility report, which is the instrument for receiving approval from competent authorities.

The feasibility analysis of industrial projects using the above steps suffers from the following shortcomings:

- Due to stricter environment regulations in developed and developing countries, quite often impact assessment suggests alternate sites, technologies, designs, and implementation methods as mitigating measures. This causes considerable delay to complete project feasibility analysis and selection as complete analysis requires to be taken up again and again until the statutory regulatory authorities approve the project.
- Moreover, project analysis through the above processes results in suboptimal project as financial analysis eliminates better options (environment friendly alternatives are mostly cost intensive).

Hence, there is enough logic to take up environmental and social impact assessment together with technical analysis so as to select a project for sustainable development not only to ensure organization's competitiveness, but also to provide ecological balance of the earth. Accordingly, the objective of the study has been formulated to develop an integrated project evaluation and selection model for industrial organizations.

METHODOLOGY

This study adopts analytic hierarchy process (AHP), a multiple-attribute decision-making technique to develop a decision support system (DSS) for project evaluation and selection. The model has been developed with the active involvement of the project stakeholders' in-group decision-making framework. The entire methodology has been explained through a case study of cross-country petroleum pipelines project in India.

LITERATURE ON PROJECT SELECTION

There is a large literature dedicated to the project selection problem. It includes several approaches, which take into account various aspects of the problem. Strategic intent of the project, factors for project selection, and various qualitative and quantitative project selection models has been thoroughly discussed by Meredith and Mantle (2000). Danila (1989) and Shpak and Zaporojan (1996) survey some of the project selection methodologies. Various articles discuss application of operation research tools in project selection. Mehrez and Sinuany-Stern (1983) use utility function. Khorramshahgole and Steiner (1988) and Dey, Tabucanon, and Ogunlana (1996a) apply goal programming. Chu, Pin-Yu, Hsu Yeh-Liang, and Fehling (1996) and Coffin and Taylor (1996) demonstrate project selection process using fuzzy theory. Project selection decision and fund allocation problem using 0 – 1 mathematical modeling is discussed by Lockett and Stratford (1987) and Regan and Holtzman (1995). Ghasemzadeh, Archer, and Iyogun (1999) and Ghasemzadeh and Archer (2000) propose a 0 – 1 integer linear programming model for selecting ad scheduling an optimal project portfolio based on an organization's objectives and constraints. Analytic hierarchy process has been used by many authors to resolve decision-making issues in project selection (Dey & Gupta 2001; Mian & Christine, 1999). Project selection issues have been discussed in various management functions such as research and development (Loch & Kavadias, 2002), environmental management (Eugene & Dey, 2005), and quality management (Hariharan, Dey, Kumar, & Moseley, 2004). Projects are unique in nature. Hence, each model has its own pros and cons for various applications. This study adopts AHP as project selection model because of its advantage of using in-group decision-making framework with the involvement of the concerned stakeholders.

AN AHP-BASED APPROACH TO EVALUATE PROJECT

The analytic hierarchy process (AHP) developed by Saaty (1980) provides a flexible and easily understood way of analyzing complicated problems. It is a multiple-criteria decision-making technique that allows subjective as well as objective factors to be considered in decision-making process. The AHP allows the active participation of decision makers in reaching agreement, and gives managers a rational basis on which to make decisions. AHP is based on the following three principles: decomposition, comparative judgment, and synthesis of priorities.

The AHP is a theory of measurement for dealing with quantifiable and intangible criteria that has been applied to numerous areas, such as decision theory and conflict resolution (Vargas, 1990). AHP is a problem-solving framework and a systematic procedure for representing the elements of any problem (Saaty, 1983).

Formulating the decision problem in the form of a hierarchical structure is the first step of AHP. In a typical hierarchy, the top level reflects the overall objective (focus) of the decision problem. The elements affecting the decision are represented in intermediate levels. The lowest level comprises the decision options. Once a hierarchy is constructed, the decision maker begins a prioritization procedure to determine the relative importance of the elements in each level of the hierarchy. The elements in each level are compared as pairs with respect to their importance in making the decision under consideration. A verbal scale is used in AHP that enables the decision maker to incorporate subjectivity, experience, and knowledge in an intuitive and natural way. After comparison matrices are created, relative weights are derived for the various elements. The relative weights of the elements of each level with respect to an element in the adjacent upper level are computed as the components of the normalized eigenvector associated with the largest eigenvalue of their comparison matrix. Composite weights are then determined by aggregating the weights through the hierarchy. This is done by following a path from the top of the hierarchy to each alternative at the lowest level, and multiplying the weights along each segment of the path. The outcome of this aggregation is a normalized vector of the overall weights of the options. The mathematical basis for determining the weights was established by Saaty (1980).

Project evaluation is usually a team effort, and the AHP is one available method for forming a systematic framework for group interaction and group decision making (Saaty, 1982). Dyer and Forman (1992) describe the advantages of AHP in a group setting as follows: 1) both tangibles and intangibles, individual values, and shared values can be included in an AHP-based group decision process; 2) the discussion in a group can be focused on objectives rather than alternatives; 3) the discussion can be structured so that every factor relevant to the discussion is considered in turn; and 4) in a structured analysis, the discussion continues until all relevant information from each individual member in a group has been considered and a consensus choice of the decision alternative is achieved. A detailed discussion on conducting AHP-based group decision-making sessions including suggestions for assembling the group, constructing the hierarchy, getting the group to agree, inequalities of power, concealed or distorted preferences, and implementing the results is explored by Saaty (1982) and Golden, Wasli, and Harker (1989). For problems with using AHP in group decision making, see the article by Islei, Lockett, Cox, and Stratford (1991).

AHP was used for the simultaneous technical, environmental, and socioeconomic analysis. AHP was used for as a part of the project analysis model because:

- The factors that lead to project selection are both objective and subjective;
- The factors are conflicting, achieving of one factor may sacrifice others;
- Some objectivity should be reflected in assessing subjective factors;
- AHP can consider each factor in a manner that is flexible and easily understood, and allows consideration of both subjective and objective factors; and
- AHP requires the active participation of decision makers in reaching agreement, and gives decision makers a rational basis upon which to make their decision.

Researchers use AHP in various industrial applications. Partovi, Burton, and Banerjee (1990) use it for operations management decision making. Dey, Tabucanon, and Ogunlana (1994) use it in managing the risk of projects. Korpela and Tuominen (1996) and Dey (2002) use AHP for benchmarking logistic operations and project management respectively. Dey (2004) applies AHP for deriving risk-based maintenance method for oil pipelines.

In this study, an AHP-based approach to evaluate and select projects has been demonstrated.

APPLICATION

Cross-country petroleum pipelines are the most energy-efficient, safe, environmental friendly, and economical means for transporting hydrocarbons (gas, crude oil, and finished product) over long distances within a country and between countries. Today, a significant part of a nation's energy requirement is transported through pipelines. The economy of a country can be heavily dependent on smooth and uninterrupted operation of these lines. (Dey & Gupta, 2000)

Therefore, it is important to ensure safe and failure free operation of these pipelines. While pipelines are one of the safest means for transporting bulk energy, with failure rates much less than railroads, failures do occur and sometimes have catastrophic consequences. In addition, disruption of pipeline operations can lead to large losses in business. (Dey, Ogunlana, Gupta, & Tabucanon, 1998).

To avoid failures, pipeline operators choose optimal pipeline routes (Dey & Gupta, 1999) and consider the long-term profitability of each project (Dey, Tabucanon, & Ogunlana, 1996b). The study by Dey and Gupta (1999) suggest to select optimum route of pipelines with the consideration of technical factors like operability, maintainability, approachability, constructability, and environmental friendliness. Although environmental factors were considered in their analysis, they did not give enough emphasize on environmental and social aspects of oil pipelines during normal operations as well as during any failure situation. In the recent years, social impact assessment (SIA) and environmental impact assessment (EIA) have emerged to help ensure a project is both profitable and a contributing agent to the society. Their findings, however, are sometimes ignored by the project owner, causing conflict within surrounding populations. This conflict can cause damage either to the project or to the society.

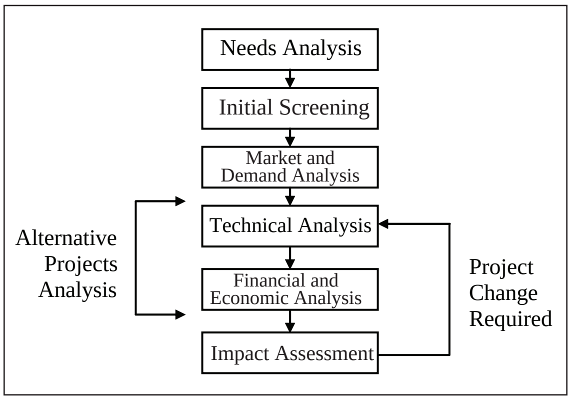
Project Description

The project under study is a cross-country petroleum pipeline project in western India. Its length is 1,300 kilometers plus a 123-kilometer branch line. The pipeline is designed carry 5 million metric tons per annum (MMTPA) of throughput. The project includes three pump stations, one pumping/delivery station, two scraper stations, four delivery stations, and two terminal stations. The project cost was estimated as 600 million US $. A detailed description of the project is available in Dey (1997).

Customary Project Evaluation Processes of Petroleum Pipelines

Figure 1 shows the customary cross-country petroleum pipeline feasibility analysis processes. Rapid industrial growth calls for the study of many potential pipeline projects, which are scrutinized to identify a few feasible projects for detailed analysis. Market and demand analysis determines the pipeline route and supply-demand points. The technical analysis assesses a few alternatives with respect to pipe diameter and the number of intermediate stations. The optimum alternative is selected on the basis of financial evaluation criteria

Figure 1. Customary project evaluation and selection processes



such as pay back period, net present value, and internal rate of return. The environmental and socioeconomic impact assessment is then conducted out on a single selected project to identify means to mitigate negative environmental impacts.

The pipeline planners who follow above the steps encounter the following problems:

1. A long study time frame because studies are completed sequentially.
2. Only one alternative is addressed during impact assessment, which is called upon to justify this alternative generated from financial analysis.
3. Impact assessment findings often demand alteration of the project site (pipeline route) and use of a different technology, necessitating revision of the technical and financial analysis.
4. Although sometimes projects get statutory approval from the regulatory authorities based on impact assessment reports, there is evidence of project abandonment at later stage or delayed because of public protest.
5. Project approval takes time because approving authorities often ask for additional information, necessitating further detailed analysis.
6. Sometimes, the selected projects prove to be not fully effective in the operations stage because of large operating and maintenance costs and lack of expansion opportunities.
7. The likelihood of failure and its impact is very high, as this is not emphasized in design stage.

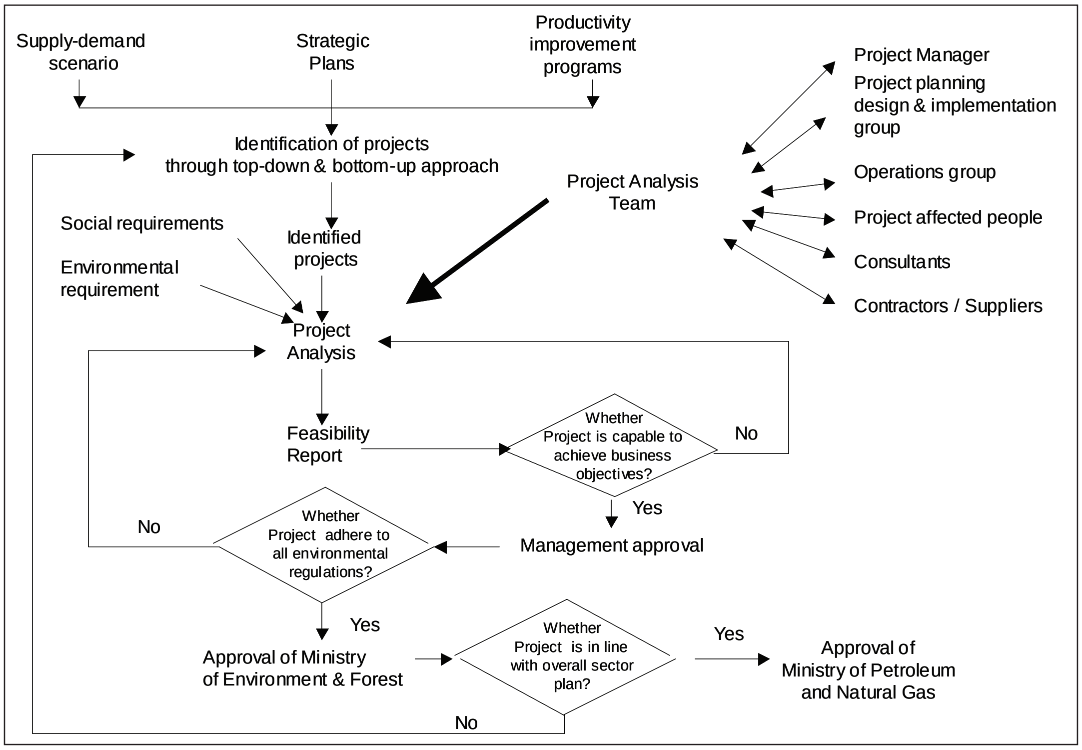
The above problems can be resolved by incorporating the feasibility analyses and impact assessments into an integrated framework with active involvement of all the stakeholders.

Potential projects are first identified through both top-down and bottom-up approaches that involve different levels of executives. They consider the supply and demand of petroleum products and crude, the organization's strategic plans, and productivity improvement. Brainstorming and/or the Delphi technique are employed to screening the feasible projects. Next, a project analysis team is formed. A project analysis team consists of representatives of a design (civil, electrical, mechanical, and telecommunication) group, a planning group, an implementation group, an operations group, and a finance group. They are selected based on their experience and past performance. They form the feasibility study's core working group. They identify the project stakeholders, determine their concerns, and involve them in the analysis.

The project analysis team establishes environmental and social impact assessment requirements, based in part on the results of interaction with environmental regulators and project-affected people. Project stakeholders are active during the identification of alternatives and project selection criteria. Stakeholders also take part in decision making, including the development of comparison matrices. A resulting feasibility report is used by the owner's management to decide whether a recommended project has potential for implementation and organization's growth.

The feasibility report is submitted to Ministry of Environment and Forest for environmental clearance. The ministry examines the project with respect to sustainable development and use of clean technology. Considering environmental requirements at this early stage permits quick approval from the ministry. A quick response can be made to the ministry's queries since an environmental analysis is complete and available. The Ministry of Petroleum and Natural Gas is the ultimate authority for approving the project in principle and allocating funds for implementation. The entire analysis and approval processes, along with communication network among project stakeholders, are shown in Figure 2.

Figure 2. Communications among project stakeholders during feasibility study and approval processes



Proposed Project Evaluation and Selection Model

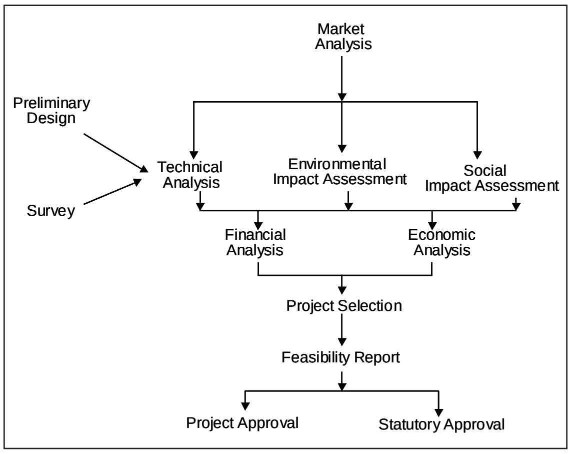
Figure 3 shows the model for feasibility analysis of a cross-country petroleum pipeline used for the pipeline project under study. The technical analysis (TA), the environmental impact assessment (EIA), and the socio-economic impact assessment (SEIA) is conducted concurrently. These studies solve site selection (pipeline route) problems, as well as a few technological considerations. The least cost option is then identified through a financial and economic analysis of a few feasible alternative projects.

For the project under study, the decision makers were petroleum executives having more than 15 years of working experience. They established a common consensus for the AHP hierarchy through group decision making. Disagreements were resolved by reasoning and collecting more information. Their hierarchy contained the details necessary to project selection. It gave insight into the whole process and a basis for selection to the approving authority. A joint meeting (decision makers and approving authority) could further facilitate the approval process. The sensitivity utility of AHP provided decision makers with an opportunity to understand the implications of their decision.

The following steps were adopted for selecting an optimal project:

- Identification of alternative pipeline routes and creation of a data base for each route

Figure 3. Integrated project analysis model



using a geographical information system (GIS) (Montemurro & Barnett, 1998).

- Identification of factors and subfactors needed to select an optimal project.
- Creation of the project selection model in an AHP framework, taking into account TA, EIA, and SEIA.
- Analyzing each factor and subfactor by comparing them in pairs and analyzing each alternative using available data with respect to each subfactor.
- Synthesizing the results across the hierarchy to identify the optimal project.

The following factors are considered for the analysis and the factors are described in the sections that follow.

Technical Factors

The technical factors important to selection of pipeline route include length, operability, maintainability, approachability, and constructability.

Pipeline Length

Pipeline length governs the capacity requirement of almost all equipment for the entire pipeline project, as pipeline head loss is directly proportional to the length of the pipeline. Hence, the shorter the length of a pipeline the lesser is the capital cost of the project and vice versa.

Operability

The hydraulic gradient is a major factor in selecting prime mover power for pipeline operations as negative hydraulic gradient demands for higher prime mover power. Similarly, more route diversion causes more friction loss, resulting in higher prime mover power for the same throughput. These cause more capital investment. A pipeline is designed for specific throughput in line with demand; a pipeline may need to be augmented in the future to cope with the demand for maximizing profit. Therefore, expansion/augmentation capability is one attribute of properly designed pipeline. In addition to improving the existing prime mover capacity, a pipeline can also be augmented by installing more pumping stations along the route and laying loop lines/parallel lines.

Maintainability

Though pipelines are designed with adequate safety factors; they are subjected to failure due to various reasons. Pipeline corrosion, pilferage, and third party activities are the factors that may create quantum throughput loss along with chances of disaster. Therefore, these factors should be carefully considered during the feasibility study. In a decision model, these factors may influence the selection of a specific route.

One of the major causes of pipeline failure is corrosion, an electrochemical process that changes metal back to ore. Corrosion generally takes place when there is a difference of potential between two areas having a path for the flow of current. Due to this flow, one of the areas loses metal.

External interference is another leading cause of pipeline failure. It can be malicious (sabotage or pilferage) or be caused by other agencies sharing the same utility corridor. The latter is known as third-party activity. In both cases, a pipeline can be damaged severely. External interference with malicious intent is more common in socio-economically-backward areas, while in regions with more industrial activity, third-party damage is common.

Poor construction, combined with inadequate inspections and low quality materials, also contributes to pipeline failure.

Other reasons include human and operational error and equipment malfunctions. Computerized control systems considerably reduce the chance of failure from these factors.

All activities, industrial or otherwise, are prone to natural calamities, but pipelines are especially vulnerable. A pipeline passes through all types of terrain, including geologically sensitive areas. Earthquakes, landslides, floods, and other natural disasters are common reasons for pipeline failures.

Approachability

Although a cross-country petroleum pipeline is buried underground, the right of way should allow uninterrupted construction activities as well as operation, inspection, and maintenance. The ideal pipeline route should be along a railway track or a major highway. This is not always possible due to the long length of pipelines, which may require river crossings and traveling through forests, deserts, and so forth. Therefore, a pipeline route with better approachability gets an edge over other routes.

Constructability

Laying pipeline across states/province or national boundaries requires permission from statutory government authorities. Stringent safety and environmental stipulations sometimes are hindrances to project activities. Mobilization is a major construction activity. One factor in pipeline routing is the provision for effective mobilization by the contractor. Distance to market, the availability of power and water, and the number of skilled and unskilled laborers are typical requirements for starting effective construction activities. Pipeline construction methods vary greatly with terrain conditions. For example, laying pipeline across a river requires horizontal direction drilling (HDD), while laying across rocky area requires rock trenching techniques. Therefore, location characteristics are a major cost component of pipeline construction. Inappropriate route selection can cause major time and cost overruns.

Environmental Factors

Pipelines handle hazardous petroleum products. Although pipelines are designed with safety features, failure is not uncommon. Sometimes failures result in a release of large quantities of petroleum products into the environment. If this should happen, a pipeline located in a remote area is less of a safety concern.

The following factors are to be considered to assess environmental impact:

- Effect on environment during failure of pipelines.
- Effect on environment during failure of pipeline stations.
- Effect on environment during normal pipelines operations.
- Effect on environment during normal station operations.

- Effect on environment during pipeline construction.

The above factors considerably affect the selection pipeline route.

Socio-Economic Factors

Planning Stage

At this stage, pipeline route is finalized. Owner requires acquiring land for pipe laying across various terrains. Acquisition of agricultural land for industrial purposes involves several issues. Some of the important ones are payment of compensation for the land, and provision of employment, alternative accommodation, and other rehabilitation measures to the PAP.

Construction Stage

The socioeconomic issues, which need to be addressed during the construction stage of a pipeline project, are mainly the effect of employment generation and a new construction activity leading to an additional burden on local infrastructure facilities. These are only short-term impacts lasting during the construction phase of the project.

Effect of Employment Generation

During the construction phase, the major positive socioeconomic impact will be in the sphere of generation of temporary employment of very substantial numbers. This additional employment generation may lead to an influx of people into the impact area.

Table 1. Pipelines database

Description	Route I	Route II	Route III	Route IV
Throughput (MMTPA*)	3	3	3	3
Length (Km)	780	1,000	750	800
No. of Stations **	3	3	3	3
Terrain detail (Km)				
a. Normal Terrain	430	785	570	770
b. Slushy Terrain	2	5	45	15
c. Rocky Terrain	---	1	3	2
d. Forest Terrain	---	5	7	2
e. River Crossing	3	4	5	1
f. Populated area	330	200	120	10
g. Coal belt area	15	---	---	---
Soil conditions	Less corrosive soil	Less corrosive soil	Corrosive soil for slushy terrain	Less corrosive soil
Third Party activity	More because of coal belt and populated area	More because of populated area	More because of populated area	---
Chances of Pilferage	Higher because of populated area	Higher because of populated area	Higher because of populated area	---

** MMTPA = million metric tons per annum.*
***1 – originating pumping station, 1 – intermediate pump station, 1 – Terminal delivery station.*

Figure 4. AHP model for project evaluation and selection

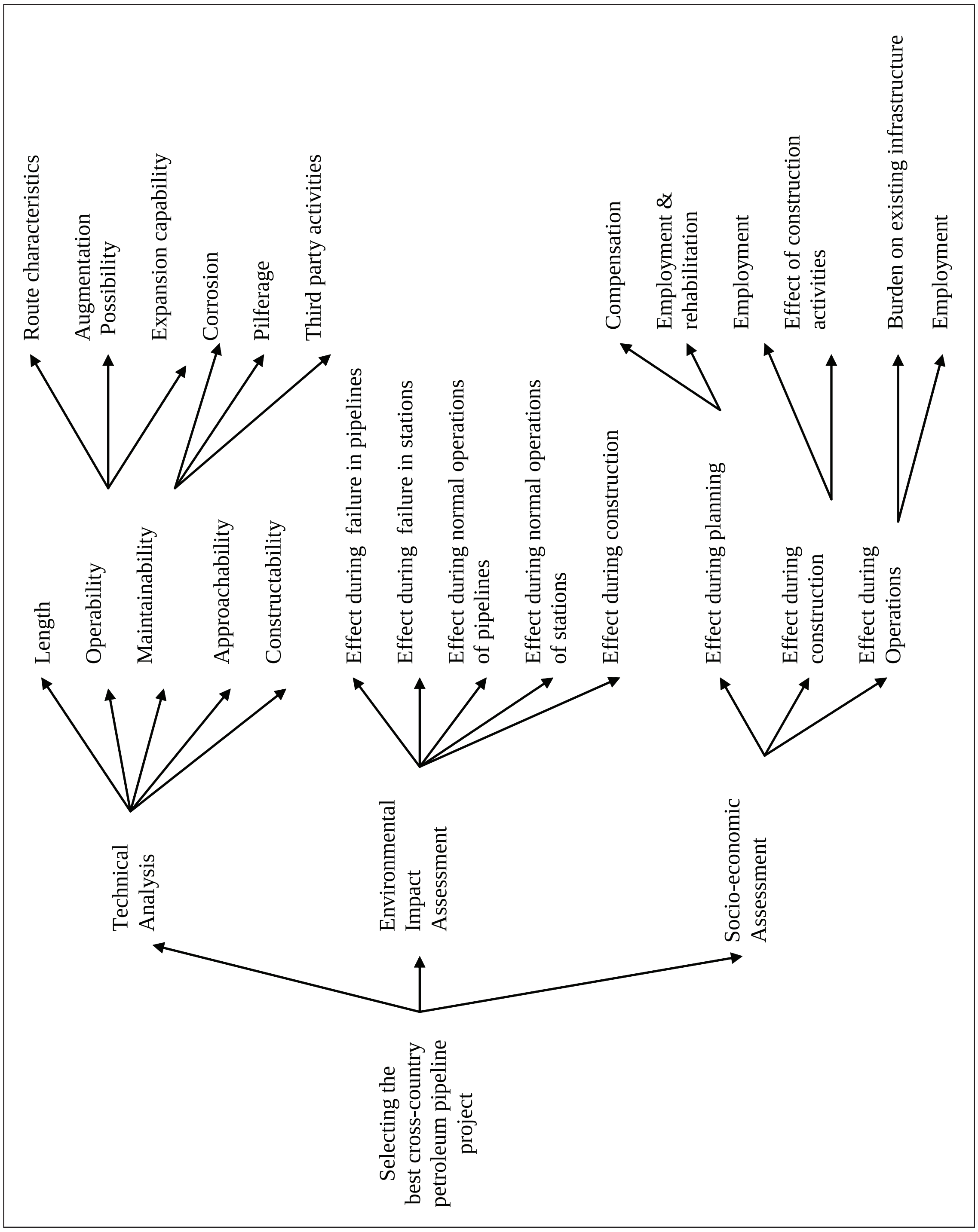


Table 2. Project selection data analysis

Factors (i)	Weights (ii)	Sub-factors (iii)	Weights (iv)	Sub-factors (v)	Weights (vi)	Normalize weights of sub-factors (vii)	Route 1 (weights) (viii)	Route 2 (weights) (ix)	Route 3 (weights) (x)	Route 4 (weights) (xi)
Technical Analysis	0.45	Length	0.31			0.1400	0.27	0.1	0.37	0.26
		Operability	0.20	Route characteristics	0.21	0.0190	0.2	0.22	0.3	0.28
				Augmentation possibility	0.44	0.0400	0.25	0.36	0.12	0.27
				Expansion capability	0.35	0.03100	0.26	0.37	0.08	0.29
		Maintainability	0.24	Corrosion	0.6	0.0650	0.23	0.3	0.15	0.32
				Pilferage	0.25	0.0270	0.21	0.24	0.25	0.3
				Third party activities	0.15	0.0160	0.21	0.28	0.25	0.26
		Approachability	0.1			0.0450	0.23	0.33	0.13	0.31
		Constructability	0.15			0.0675	0.21	0.28	0.17	0.34
Environmental Impact Assessment	0.25	During failure of pipelines	0.41			0.1020	0.23	0.3	0.15	0.32
		During failure of stations	0.33			0.0820	0.21	0.28	0.25	0.26
		During normal pipelines operations	0.07			0.0175	0.22	0.32	0.18	0.28
		During normal station operations	0.08			0.0200	0.18	0.32	0.15	0.35
		During pipeline construction	0.11			0.0270	0.20	0.28	0.22	0.3
Socio-economic Impact Assessment	0.30	Effect during planning	0.43	Compensation	0.7	0.0900	0.21	0.16	0.33	0.3
				Employment & rehabilitation	0.3	0.0387	0.14	0.24	0.28	0.34
		Effect during construction	0.36	Employment	0.5	0.0540	0.25	0.25	0.15	0.35
				Effect of construction activities	0.5	0.0540	0.12	0.18	0.27	0.43
		Effect during operations	0.21	Employment	0.2	0.0126	0.25	0.25	0.25	0.25
				Burden on existing infrastructure	0.8	0.0500	0.17	0.18	0.3	0.35
Overall weights							**0.218**	**0.241**	**0.232**	**0.309**
Ranking							**IV**	**II**	**III**	**I**

Effect of Construction Activity

Construction activity involves movement of heavy vehicles, leading to disruption of other agriculture activities. Pipeline construction sometimes leads to local transport disruption also.

Operation Stage

The operational stage of the project covers the entire lifespan of the pipelines. Hence, the impacts of the operational phase extend over a long period time. However, pipeline projects seldom generate employment opportunity in this stage and provide fewer burdens to existing infrastructure as the pipelines remain buried under the earth. However, agricultural activities remain restricted on right-of-way (ROW) throughout the lifespan of the pipelines.

Project Selection Model

Figure 4 shows the project evaluation and selection model in AHP framework. Level I is the goal of selecting the best cross-country petroleum pipeline project. Levels II and III are the factors and subfactors considered for selection. Level IV is the alternative projects of various feasible pipeline routes.

Table 1 shows the database used for each alternative route for the project under study. These data along with the experience of pipeline operators were utilized to apply the AHP model to select the best pipeline project. The data were analyzed by using Expert Choice software package.

Results and Findings

Table 2 shows the final analysis results and selection. The analysis of the data indicates that Alternative 4 is the best pipeline route for the project under study, although it is not the shortest route. Alternative 4 outranks the other alternatives with respect to its operability, maintainability, environment friendliness, and impact on society.

Validation of the Model

Table 3 shows the life cycle cost estimate of the project with the shortest route and with route

Table 3. Life cycle cost estimate for pipelines (millions of US$)

S N	Description	Shortest Route	Optimal Route
1.	Capital cost	37.7	39.5
2.	Operating cost: First 5 years Second 5 years Third 5 years Fourth 5 years	 1.1 each year* 1.5 each year* + 2.1** 2 each year* + 6** 2 each year*	 0.75 each year* 0.75 each year* 1 each year* + 1.1*** 1 each year*
3.	Net Present Value (NPV) MARR = 10% (assumed)	1.70#	5.05#

* Normal operation and maintenance cost.
** Major inspection and maintenance cost in subsequent five years that includes additional patrolling, special arrangement for failure, a water logged area, water pollution control and special coating/ CP Surveillance, intelligent pigging cost, including cost for loss of production for not being able to augment the pipeline.
*** Additional capital cost for augmentation.
\# For deriving NPV, positive cash flow has been determined with the following assumptions:
15% return on capital for first and last 5 years
20% return on capital for second and third 5 years
US $1 million cost to be incurred for abandoning the project

Table 4. Capital cost of pipelines construction (figures in millions US$)

Item Description	Optimal Route (800 km)	Shortest Route (750 km)
Pipes	3.25	3
Survey	1	0.9
Coating	2	1.8
Laying	15	14
Cathodic protection	1.25	1
Building	2.25	2.25
Pumping units	5.75	5.75
Telecommunication	8	8
Others	1	1
Total	39.5	37.7

Table 5. Operating cost (millions US$)

Item Description	Optimum Route				Shortest Route			
	0-5 years	5-10 years	10-15 years	15-20 years	0-5 years	5-10 years	10-15 years	15-20 years
Energy	0.20	0.20	0.25	0.25	0.05	0.10	0.25	0.25
Routine inspection	0.05	0.05	0.15	0.15	0.25	0.25	0.15	0.15
Overhead	0.25	0.25	0.30	0.30	0.25	0.25	0.30	0.30
Purchase of equipment spares and other sundry items	0.25	0.25	0.30	0.30	0.30	0.40	0.30	0.30
Additional inspection	-	-	-	-	0.25	0.30	0.40	0.40
Cost of failure	-	-	-	-	-	0.20	0.60	0.60
Total	0.75	0.75	1	1	1.1	1.5	2.0	2.0
Cost of augmentation	-	-	1.1		-	-	-	-

Alternative 4, the optimal route. Tables 4 and 5 show capital and operating costs respectively. The present value (PV) of the project with shortest route is much less than the selected project using AHP. Therefore, the life cycle costing (LCC) model also favors Alternative 4. Collecting the information for LCC, however, is time consuming and expensive. In addition, an estimate of LCC is generally based on many assumptions. On the other hand, the decision support system (DSS) using AHP provides a model for project selection that relies on the experience of project staff as well as concerned stakeholders. It also considers the project life cycle when selecting the project.

Financial Analysis

The financial analysis was then conducted, considering only a few alternative pipeline diameters.

The least cost option was selected. This report does not describe the financial analysis of various design options since this is a standard practice of all pipeline project planning.

SUMMARY AND CONCLUSION

Project selection is strategic to any organization. Today's project evaluation increasingly emphasizes on environmental and social factors along with technical factors while selecting the best alternative project. Therefore, project feasibility shall be established for entire products/services life using a holistic approach with consideration of market, technical, environmental, social, and financial factors. This calls for manipulating a large database and making intelligent assumptions to arrive into appropriate selection. Involvement of the stakeholders in project evaluation and selection not only helps to develop a model for project evaluation, but also helps to select the most appropriate project in an uncertain environment. This study shows an AHP-based framework for project evaluation and selection.

Evaluation of pipeline projects is presently conducted within a fragmented framework with many studies occurring prior to impact assessment. Because of stronger environmental laws and regulations, impact assessment quite often either suggests substantial changes in the project or abandonment of the project on environmental grounds. Such findings can result in substantial revisions to the project proposal, including new site studies, use of alternate technologies, and alternate implementation methodologies. This not only increases project feasibility study time considerably but also increases the cost and effort of the project's proponent.

This article presents an integrated framework of technical, environmental impact, and social impact assessment for project evaluation and selection. This model uses an AHP framework that considers both subjective and objective factors. The model has the following advantages:

- It allows incorporation of interactive input from the executives of related functional areas.
- It integrates technical, financial, and impacts assessment with benefits to both the project owner and affected populations.
- Its aids objective decision making by quantifying many subjective factors.
- Both tangible and intangible elements can be included in the AHP hierarchy. Qualitative judgment and quantitative data can be included in the priority setting process.
- AHP is an effective tool for conducting group planning sessions in an analytical and systematic manner.
- It demands collection of information that is ultimately of use during the detailed engineering stage.
- The sensitivity analysis utility of AHP provides the decision makers a sense of effects of their decisions.
- It improves communication among project stakeholders and allows consideration of the concerns of project stakeholders in a structured way.
- It not only helps in managing project effectively, but also helps develop a quality project with a potential for desired outputs.

The model suffers from the limitation of not completely removing subjectivity from the decision model. However, it is an improvement over the present practice.

Although the application of the model was explained through a representative cross-country petroleum pipeline project, it can be applied universally across various project selection problems. Considerable research is required, however, for each application.

FUTURE RESEARCH DIRECTIONS

Project selection is one of the first steps of managing projects. This study develops a framework of project section using analytic hierarchy process. It considers all the factors of project selection together to identify the best option. The proposed framework has been applied to an oil pipeline project. The same framework could be applied to any other projects. Moreover, other multiple attribute decision-making techniques could also be used to develop the decision support system for project selection. Additionally, a few hybrid models could also be formulated to develop group decision support system. An empirical study could be carried out to examine the effectiveness of various decision support systems across projects. An empirical research on whether the effectiveness of the proposed framework depends on management/leadership skill would also be very important in order to improve performance. Project selection is a strategic decision, which could be linked with the supply chain decision making on a strategic level with the consideration of efficiency and responsiveness trade off. Sustainability issues of any organization are directly linked with its project selection process. Sustainability issues are multifactorial. A study could be developed with the consideration of all the sustainability issues in relation to productivity and environmental performance and addressing those through appropriate selection of projects.

REFERENCES

Calvin, H., & Dey, P. K. (2002). Social impact assessment of a sewerage project in Barbados. *Impact Assessment and Project Appraisal, 20*(2), 215-223.

Chu, Pin-Yu V., Hsu Yeh-Liang, & Fehling, M. (1996). A decision support system for project portfolio selection. *Computer in Industry, 32*(2), 141-149.

Coffin, M. A., & Taylor, B. W. (1996). Multiple criteria R&D project selection and scheduling using fuzzy logic. *Computer & Operations Research, 23*, 207-221.

Danila, N. (1989). Strategic evaluation and solution of R&D projects. *R&D Management, 19*, 47-62.

Dey, P. K. (1997). *Symbiosis of organisational re-engineering and risk management for effective implementation of large-scale construction project.* Unpublished doctoral thesis, Jadavpur University.

Dey, P. K. (2002). Benchmarking project management practices of Caribbean organizations using analytic hierarchy process. *Benchmarking: An International Journal, 9*(3), 326-356.

Dey, P. K. (2004). Decision support system for inspection and maintenance of cross-country petroleum pipeline. *IEEE Transactions on Engineering Management, 51*(1), 47-56.

Dey, P. K., & Gupta, S. S. (1999). Decision support system for pipeline route selection. *Cost Engineering, 41*(10), 29-35.

Dey, P. K., & Gupta, S. S. (2000). Analytic hierarchy process boosts risk analysis objectivity. *Pipeline and Gas Industry Journal, 83*(9), 6-72.

Dey, P. K., & Gupta, S. S. (2001). Feasibility analysis of cross-country pipeline projects: A quantitative approach. *Project Management Journal, 32*(4), 50-58.

Dey, P. K., Ogunlana, S. O., Gupta, S. S., & Tabucanon, M. T. (1998). A risk based maintenance model for cross-country pipelines. *Cost Engineering, 40*(4), 24-31.

Dey, P. K., Tabucanon, M. T., & Ogunlana, S. O. (1994). Planning for project control through risk analysis: A case of petroleum pipeline laying project. *International Journal of Project Management*, *12*(1), 23-33.

Dey, P. K., Tabucanon, M. T., & Ogunlana, S. O. (1996a). Hierarchical approach to project planning. *Applied Math. Modelling*, *20*, 683-698.

Dey, P. K., Tabucanon, M. T., & Ogunlana, S. O. (1996b). Petroleum pipeline construction planning: A conceptual framework. *International Journal of Project Management*, *14*(4), 231-240.

Dyer, R. F., & Forman, E. H. (1992). Group decision support with the analytic hierarchy process. *Decision Support System*, *8*, 99-124.

Ghasemzadeh, F., & Archer, N. P. (2000). Project portfolio selection through decision support. *Decision Support Systems*, *29*, 73-88.

Ghasemzadeh, F., Archer, N. P., & Iyogun P, (1999). A zero-one model for project portfolio selection and scheduling. *Journal of Operational Research Society*, *50*(7), 745-755.

Golden, B. L., Wasli, E. A., & Harker, P. T. (1989). *The analytic hierarchy process: Applications and studies*. New York: Springer Verlag.

Hariharan, S., Dey, P. K., Kumar, A. Y., & Moseley, H. S. L. (2004). A new tool for measurement of process-based performance of multispecialty tertiary care hospital. *International Journal of Health Care Quality Assurance*, *17*(6), 302-312.

Islei, G., Lockett, G., Cox, B., & Stratford, M. (1991). A decision support system using judgmental modeling: A case of R&D in the pharmaceutical industry. *IEEE Transaction on Engineering Management*, *38*, 202-209.

Khorramshahgole, R., & Steiner, H. M. (1988). Resource analysis in project evaluation: A multicriteria Approach. *Journal of Operational Research Society*, *95*(2), 795-803.

Korpela, J., & Tuominen, M. (1996). Benchmarking logistic performance with an application of the analytic hierarchy process. *IEEE Transaction on Engineering Management*, *43*(3), 323-333.

Loch, C. H., & Kavadias, S. (2002). Dynamic portfolio selection of NPD programs using marginal returns. *Management Science*, *48*(10), 1227-1241.

Locket, G., & Stratford, M. (1987). Ranking of research projects, experiments with two methods. *Omega 15*, 395-400.

Mehrez, A., & Sinuany-Stern, Z. (1983). An interactive approach to project selection. *Journal of Operational Research Society*. *34*, 621-626.

Meredith, J. R., & Mantel, S. J. (2000). *Project management: A managerial approach* (4th ed.). New York: John Wiley & Sons.

Mian, S. A., & Christine, N. D. (1999). Decision-making over the project life cycle: An analytical hierarchy approach. *Project Management Journal*, *30*(1), 40-52.

Montemurro, D., & Barnett, S. (1998). GIS-based process helps trans Canada select best route for Expansion Line. *Oil and Gas Journal*, *96*(25), 63-69.

Partovi, F. Y., Burton, J., & Banerjee, A. (1990). Application analytic hierarchy process in operations management. *International Journal of Operations and Production Management*, *10*(3), 5-19.

Ramanathan, R., & Geetha, S. (1998). Socio-economic impact assessment of industrial projects in India. *Impact Assessment and Project Appraisal*, *16*(1), 27-32.

Ramcharan, E., & Dey, P. K. (2005). Environmental factors play a major role in site selection for limestone quarry expansion project of cement plant. *Impact Assessment and Project Appraisal*, *23*(2), 147-154.

Regan, P., & Holtzman, S. (1995). R&D decision advisor: An interactive approach to normative decision system model construction. *European Journal of Operational Research, 84*(1), 116-133.

Saaty, T. L. (1980). *The analytic hierarchy process.* New York: McGraw Hill.

Saaty, T. L. (1982). *Decision making for leaders.* New York: Lifetime Learning.

Saaty, T. L. (1983). Priority setting in complex problems. *IEEE Transaction on Engineering Management, 30*, 140-155.

Shpak, A., & Zaporojan, D. (1996). Working out R&D program via multicriteria analysis. *Computer Science of Moldova, 4*(2), 239-259.

Vargas, L. G. (1990). An overview of the analytic hierarchy process and its applications. *Europe Journal of Operational Research*, *48*(1), 2-8.

ADDITIONAL READING

Badria, M.A., Davisb, D., & Davisb, D. A. (2001). Comprehensive 0-1 goal programming model for project selection. *International Journal of Project Management*, *19*(4), 243-252.

Cheng, E. W. L., Cowan, E., & Li, H. (2005). Analytic network process applied to project selection. *Journal of Construction Engineering and Management*, *131*(4), 459-466.

Dey, P. K., & Hariharan, S. (2006). Integrated approach to healthcare quality management: A case study. *The TQM Magazine*, *18*(6), 583-605.

Dey, P. K., Hariharan, S., & Chen, D. (2007). Healthcare quality management in project management framework. *International Journal of Service Operations Management*, *3*(3), 261-278.

Dey, P. K., Hariharan, S., & Clegg, B. (2006). Measuring the operational performance of intensive care units using the analytical hierarchy process approach. *International Journal of Operations and Production Management, 26*(8), 849-865.

Dey, P. K., Kinch, J., & Ogunlana, S. O. (2007). Managing risk in software development projects. *Industrial Management and Data System, 107*(2), 284-303.

Henriksen, A. D., & Traynor, A. J. (1999) A practical R&D project-selection scoring tool. *IEEE Transactions on Engineering Management, 46*(2), 158-170.

Jiang, J. J., & Klein, G. (1999). Information system project-selection criteria variations within strategic classes. *IEEE Transactions on Engineering Management*, *46*(2), 171-176.

Kuanchin, C., & Gorla, N. (1998). Information system project selection using fuzzy logic. *IEEE Transactions on Systems, Man and Cybernetics, 28*(6), 849-855.

Meade, L. M., & Presley, A. (2002) R&D project selection using the analytic network process. *IEEE Transactions on Engineering Management*, *49*(1), 59-66.

Santhanam, R., & Kyparisis, J. (1995) A multiple criteria decision model for information system project selection. *Computers and Operations Research*, *22*(8), 807-818.

Chapter XII
Modeling and Coordination of Dynamic Supply Networks

Petr Fiala
University of Economics Prague, Czech Republic

ABSTRACT

The chapter is devoted to modeling and analysis of supply systems. Supply chain management is more and more affected by network and dynamic business environment. In supply chain behavior there are inefficiencies. Coordination and cooperation can significantly improve the efficiency of supply networks. There are some approaches to model and analyze the supply dynamics. Important features of this environment are established in the proposed approach. The combination of network structure modeling and simulation of dynamic behavior of units in supply network can be a powerful instrument of performance analysis of supply networks. The problem of coordination in dynamic supply networks involves multiple units with multiple goals. Multicriteria analysis of supply network performance includes such criteria as quantity, quality, time, cost, and profit. Simulation approach is an appropriate tool for prediction of real supply situation.

INTRODUCTION

The globalization and technology improvement has changed the business environment. It became more complex and dynamic, but one consequence is that organizations are making efforts to deal with the increasing challenges and to stay competitive. Production will need to find a new role in an extended form with supply and distribution networks (e.g., Vidal & Goetschalckx, 1997). It is necessary to redefine the boundaries of manufacturing and production management. Supply chain management is a philosophy that provides the tools and techniques enabling organizations to develop strategic focus and achieve sustainable competitive advantage. It presents management with a new focus and way of thinking about how their organization exists and operates within the wider business environment.

Computational intelligence refers to intelligence artificially realized through computation. It has produced a number of powerful tools, some of which are used in engineering to solve difficult problems normally requiring human intelligence. Manufacturing applications play a leading role in modern economics. There are recent advances on the computational intelligence applications in manufacturing, including system design, process planning, process monitoring control, product quality control, and equipment fault diagnosis.

The chapter is devoted to modeling and analysis of supply systems. Supply chain management is more and more affected by network and dynamic business environments. Networking transforms static and isolated resources in a pool of dynamic and connected resources. A network is the infrastructure that allows the information interchange between different points. In supply chain behavior there are inefficiencies. The relation between supply chain management and information and communication technology is key to the development of an integrated supply chain. Fundamental to this approach is the application of appropriate information technologies to facilitate the efficient and effective movement of goods, services, and information along the supply chain. Coordination and cooperation can significantly improve the efficiency of supply networks.

In this chapter, we present applications of computational intelligence to modeling and analysis of supply systems. There are some approaches to model and analyze the supply dynamics. Important features of this environment are established in the proposed approach. The combination of network structure modeling and simulation of dynamic behavior of units in supply network can be a powerful instrument of performance analysis of supply networks. The problem of coordination in dynamic supply networks involves multiple units with multiple goals. Multicriteria analysis of supply network performance includes such criteria as quantity, quality, time, cost, and profit. Simulation approach is an appropriate tool for prediction of real supply situation. Sourcing has come up as a strategic issue in the management of supply chain networks in the modern era of global competition. Modeling framework is demonstrated on solutions of supplier selection problems. The proposed approach can be completed with a three-layer framework for modeling of coordination process of units in the dynamic supply networks. The modeling framework is composed from three inter-related network structures: flow net, Petri net, and neural net.

DYNAMIC SUPPLY NETWORKS

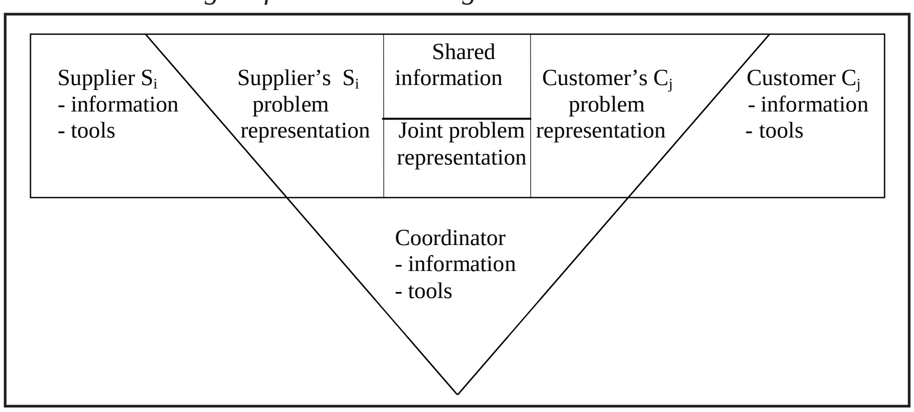
Supply chain management has generated a substantial amount of interest both by managers and by researchers. Supply chain management is now seen as a governing element in strategy and as an effective way of creating value for customers. There are many concepts and strategies applied in designing and managing supply chains (Simchi-Levi, Kaminsky, & Simchi-Levi, 1999). The expanding importance of supply chain integration presents a challenge for research to focus more attention on supply chain modeling (Tayur, Ganeshan, & Magazine, 1999). Supply chain management is more and more affected by network and dynamic business environments. The overall business environment is becoming increasingly dynamic. Demand and supply for custom products can be very dynamic. Supply chains operate in network environment as supply networks. Dynamic information and decision-making models are called to accommodate this new changes and uncertainties.

Supply networks are dynamic multilevel systems with sets of suppliers, manufacturers, distributors, retailers, and customers. The multiple decision makers are interconnected with dynamic structures and dynamic linkages by material, financial, information flows, and decision flows. Most supply networks are composed of independent units with individual preferences. Each unit will attempt to optimize a personal preference. Behavior that is locally efficient can be inefficient from a global point of view. In supply network behavior there are inefficiencies. An increasing number of companies in the world subscribe to the idea that developing long-term coordination and cooperation can significantly improve the efficiency of supply networks and provide a way to ensure competitive advantage. Once traditionally driven by pure competition, the supply chain for many successful firms has matured from an adversarial relationship to one of supply chain partnership (MacBeth & Ferguson, 1994; Thomas & Griffin, 1996). The relationship is created to increase the financial and operational performance of each chain member through reduction in total costs, reductions in inventories throughout the supply chain, and increased levels of shared information. The partners are looking to work cooperatively in providing improved service, technological innovation, and product design. Among the solutions, supply chain contracts, which have drawn much attention from the researchers recently, are used to provide some incentives to adjust the relationship of supply network partners to coordinate the supply chain. Contracts are evaluated by desirable features such as coordination of the supply chain, flexibility to allow any division of the supply chain's profit, and ease of use.

System dynamics is concerned with the modeling of complex economic systems. It allows policy experimentation in a time simulation environment with the help of causal models of the systems. Supply networks are dynamic multilevel systems with sets of suppliers, manufacturers, distributors, retailers, and customers. The multiple decision makers are interconnected with dynamic structures and dynamic linkages by material, financial, information flows, and decision flows. There are some approaches to model and analyze the supply chain dynamics.

The problem of coordination in dynamic supply networks involves multiple units with multiple goals. Goals can be divided into two types: goals that are mutual for all the agents and goals that are different and require cooperation of multiple agents to achieve a consensus. A cooperative decision making requires free communication among agents and gives synergic effects in a conflict resolution. The basic trend in the cooperative decision making is to transform a possible conflict to a joint problem. Coordination of actions can be provided through information sharing (Fiala, 2005) between suppliers and customers

Figure 1. Coordination through information sharing



(schematically see Figure 1). A supplier S_i and a customer C_j have information and analytical tools for their problem representations. A coordinator helps by communication through information sharing and by formulation of a joint problem representation.

The chapter presents an approach that respects these new specifications:

- Dynamic environment
- Network environment
- Uncertain environment
- Multiple decision makers
- Multiple criteria

The approach combines the modeling of system dynamics, dynamic network process, and the dynamic version of GROUP-ALOP (aspiration level oriented procedure). A three-layer model of cooperative problem solving framework for an integration process is proposed.

SYSTEM DYNAMICS

System dynamics is concerned with problem solving in living systems (Forrester, 1961). It links together hard control theory with soft system theory. System dynamics needs relevant tools from both ends of the systems spectrum. If the possible causal factors are identified and their respective contribution to the overall dynamics are quantitatively measured and benchmarked, then it would be conducive to performance improvement by eliminating or reducing the relevant dynamics. Systems of information feedback control are fundamental to all systems. Feedback theory explains how decisions, delays, and predictions can produce either good control or dramatically unstable operation.

The supply chain dynamics (de Souza, Song, & Chaoyang, 2000; Swaminathan, Smith, & Sadeh, 1998; Towil, 1996) lead to the increase in the cost of the units and the whole chain. A feedback control system causes a decision, which in turn affects the original environment. In supply chains, orders and inventory levels lead to manufacturing decisions that fill orders and correct inventories. As a consequence of using system dynamics in supply chain redesign, we are able to generate added insight into system dynamic behavior and particularly into underlying causal relationships. This new knowledge can be exploited in the improved design, robustness, and operating effectiveness of such systems.

The so-called bullwhip effect (Lee, Padmanabhan, & Whang, 1997), describing growing variation upstream in a supply chain, is probably the most famous demonstration of system dynamics in supply chains. The basic phenomenon is not new and has been recognized by Forrester. There are some known causes of the bullwhip effect: information asymmetry, demand forecasting, lead-times, batch ordering, supply shortages, and price variations. Information sharing of customer demand has a very important impact on the bullwhip effect.

The analyses of causes and suggestions for reducing the bullwhip effect in supply chains are challenges to modeling techniques. We consider a k-stages supply chain. The customer demands are independent and identically distributed random variables. The last stage observes customer demand D and places an order q to previous stage. All stages place orders to the previous stage in the chain. The orders are received with lead-times L_i between stages i and i+1. The stages use the moving average forecast model with p observations. To quantify increase in variability, it is necessary to determine the variance of orders q^k relative to the variance of demands D.

In the case of decentralized information the variance increase is multiplicative at each stage of the supply chain:

$$\frac{Var(q^k)}{Var(D)} \geq \prod_{i=1}^{k}\left(1+\frac{2L_i}{p}+\frac{2L_i^{\,2}}{p^2}\right)$$

In the case of centralized information, that is, the last stage provides every stage of the supply chain with complete information on customer demand, the variance increase is additive:

$$\frac{Var(q^k)}{Var(D)} \geq 1 + \frac{2(\sum_{i=1}^{k} L_i)}{p} + \frac{2(\sum_{i=1}^{k} L_i)^2}{p^2}$$

The centralized solution can be used as a benchmark, but the bullwhip effect is not completely eliminated.

The structure of supply chains and relations among units can be modeled by different types of networks. AND/OR networks can be applied for modeling flexible and dynamic supply chains (Zeng, 2001). The approach follows an activity on arc representation where each arc corresponds to a particular supply chain activity. Each activity has multiple performance criteria. Nodes represent completion of activities and establish precedent constraints among activities. The initial suppliers without predecessors and end customers without successors are represented by nodes displayed as circles. Two types of nodes are defined to specifying prior activities. AND nodes (displayed as triangles △) are nodes for which all the activities must be accomplished before the outgoing activities can begin. OR nodes (displayed as triangles ▽) require at least one of the incoming activities must be finished before the outgoing activities can begin.

Figure 2 illustrates a AND/OR supply network that consists of a structure of suppliers, different production modes, an assembly of components, and production of an end product to a customer. As an example of dynamic problem can be analyzed a stochastic inventory problem with the finite time horizon. We can describe the behavior of the network decision makers and propose a dynamic system that captures the adjustments of the commodity shipments and the prices over space and time.

The Structural Thinking Experimental Learning Laboratory with Animation (STELLA) software is one of several computer applications created to implement concepts of system dynamics (Ruth & Hannon, 1997). It combines together the strengths of an iconographic programming style and the speed and versatility of computers. The instrument is very appropriate to proposed modeling framework for dynamic multilevel supply network.

Figure 2. AND/OR supply network

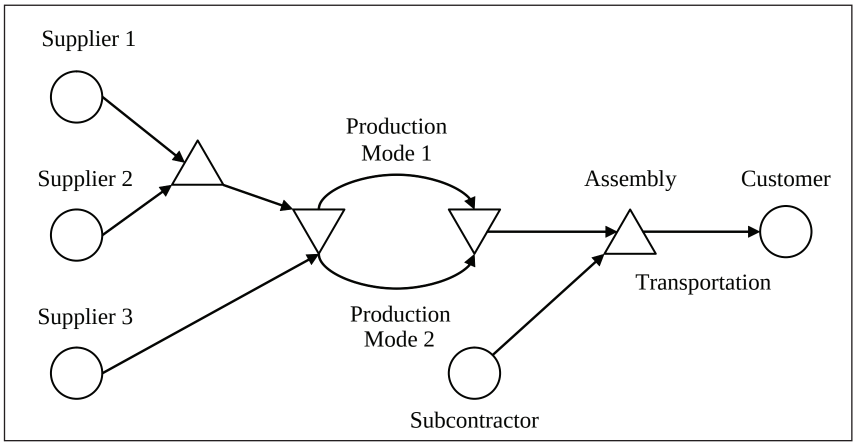
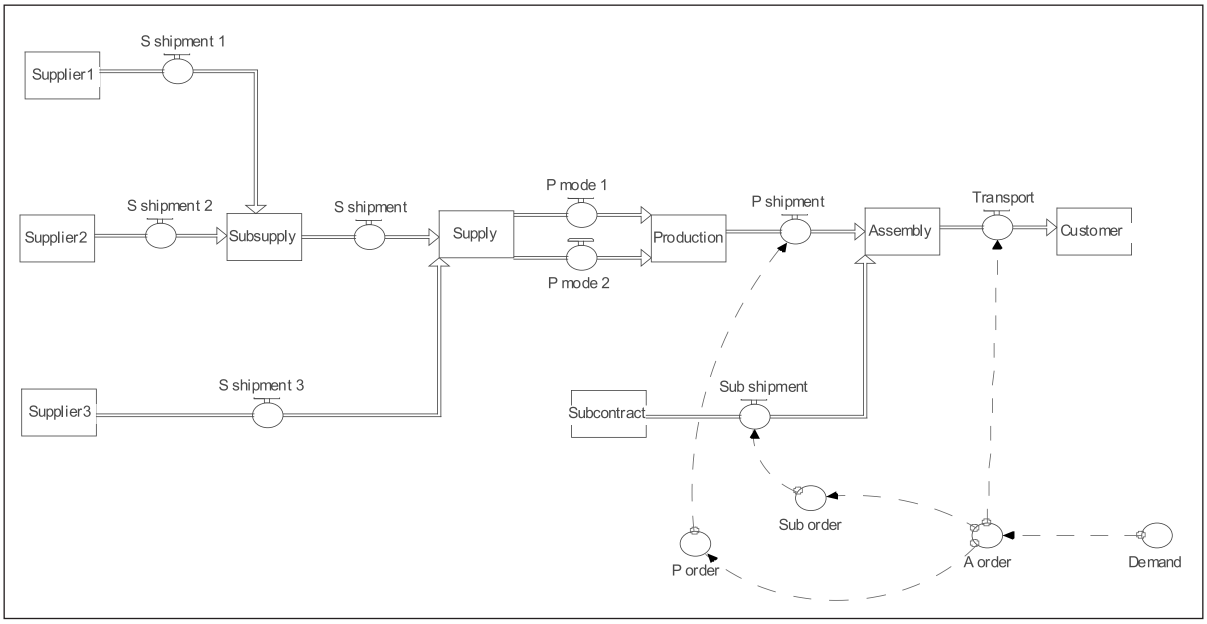


Figure 3. Supply network by STELLA software



The approach enables solving of a broad class of dynamic problems. Differential equations can be used for modeling of system dynamics. STELLA software offers the numerical techniques (Euler's method, Runge-Kutta-2, and Runge-Kutta-4 methods) to solve the model equations.

STELLA software contains many built-in functions that can facilitate dynamic modeling of supply networks. There are some examples of instruments for proposed modeling approach:

- AND/OR to modeling of AND/OR network environment.
- DELAY to modeling of led-times.
- DT time step.
- FORCST forecasts demand in stages of supply chain.
- RANDOM generates random customer demand.

As an example of dynamic problem, a stochastic inventory problem can be analyzed with the finite time horizon. AND/OR supply network consists of a structure of suppliers, different production modes, an assembly of components, and production of an end product to a customer. We can describe the behavior of the network decision makers and propose a dynamic system that captures the adjustments of the commodity shipments and the prices over space and time. The bullwhip effect can be demonstrated by comparison of random customer demand and orders in different stages of the supply network by decentralized information. Centralized information of customer demand can reduce the bullwhip effect.

A multilevel network model was proposed (Fiala, 2003). The model consists of the material network, the informational network, and the financial network. The multiple decision makers use multiple criteria as quantity, time, and cost. The efficient frontier of solutions can be identified. This network model is appropriate for analyzing of system dynamics. It can be formulated as a broad class of dynamic supply network problems.

Dynamic behavior of orders, inventories, prices, and costs at different stages of supply network can be analyzed.

We illustrate using of STELLA software on a simple supply network model. The AND/OR network from Figure 2 can be modeled by STELLA software (see Figure 3).

Figure 4 demonstrates the bullwhip effect by comparison of random customer demands (Demand) and orders in different stages of the supply network (A order, Sub order, and P order) by decentralized information. Centralized information of customer demand can reduce the bullwhip effect.

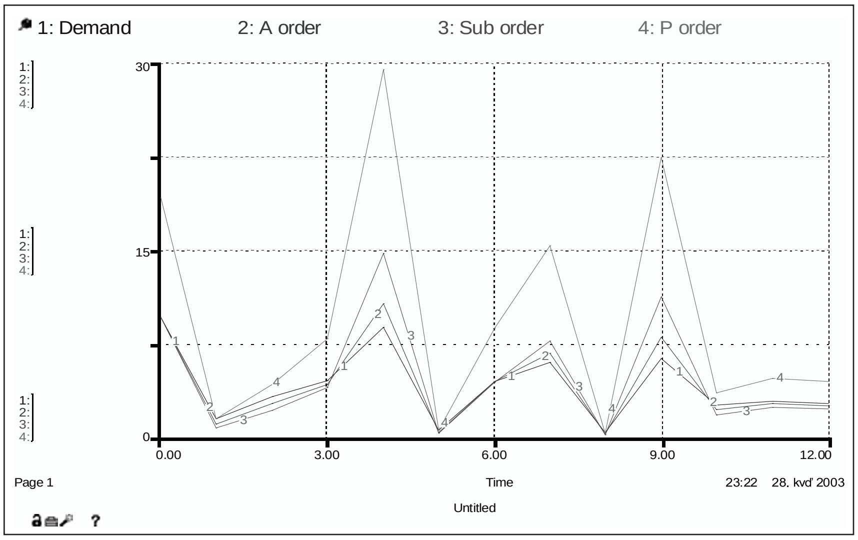
The combination of network structure modeling and simulation of dynamic behavior of units in supply chain can be a powerful instrument of performance analysis of supply chains. Simulation approach by STELLA software is an appropriate tool for prediction of real supply chain situation.

DYNAMIC NETWORK PROCESS

The analytic hierarchy process (AHP) is the method for setting priorities (Saaty, 1996). A priority scale based on reference is the AHP way to standardize nonunique scales in order to combine multiple performance measures. The AHP derives ratio scale priorities by making paired comparisons of elements on a common hierarchy level by using a 1 to 9 scale of absolute numbers. The absolute number from the scale is an approximation to the ratio w_j / w_k and then it is possible to derive values of w_j and w_k. The AHP method uses the general model for synthesis of the performance measures in the hierarchical structure:

$$u_i = \sum_{j=1}^{n} v_j w_{jk}$$

Figure 4 .Bullwhip effect

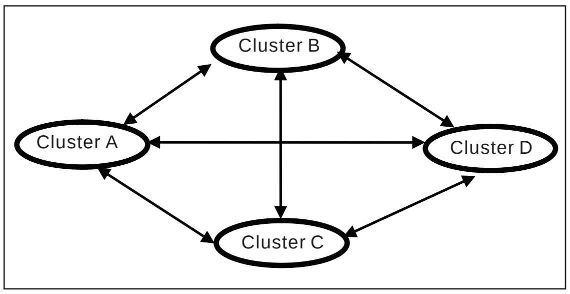


The analytic network process (ANP) is the method (Saaty, 2001) that makes it possible to deal systematically with all kinds of dependence and feedback in the system. The well-known AHP theory is a special case of the analytic network process that can be very useful for incorporating linkages in the system.

The structure of the ANP model is described by clusters of elements connected by their dependence on one another. A cluster groups elements that share a set of attributes. At least one element in each of these clusters is connected to some element in another cluster. These connections indicate the flow of influence between the elements (see Figure 5). The clusters in modeling of supply networks can be economic units, products, items, evaluating criteria, and so forth. The connections among members of supply networks are material, financial, and information flows.

Paired comparisons are inputs for computing a global performance of network systems. A supermatrix is a matrix of all elements by all elements. The weights from the paired comparisons are placed in the appropriate column of the supermatrix. The sum of each column corresponds to the number of comparison sets. The weights in the column corresponding to the cluster are multiplied by the weight of the cluster. Each column of the weighted supermatrix sums to one and the matrix is column stochastic. Its powers can stabilize after some iteration to limited supermatrix. The columns of each block of the matrix are identical and we can read off the global priority of units.

Figure 5. Flows of influence between the elements



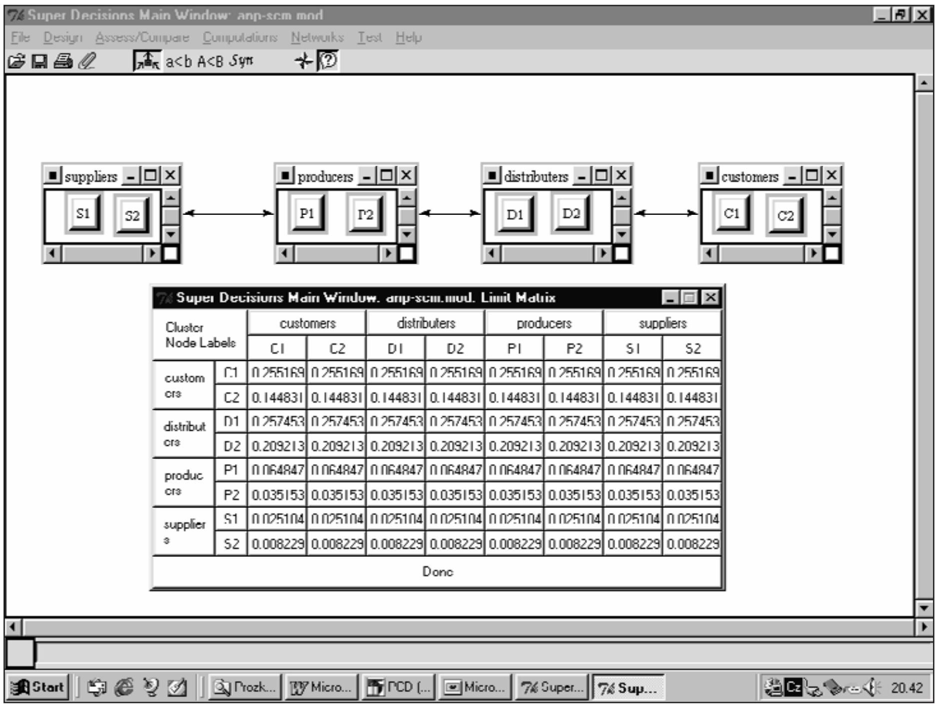
By ANP approach there are determined weights of elements in the network model. We made some experiments with evaluation of different supply network structures. For computation the priorities of units we use the software Super Decisions provided by Creative Decisions Foundation (see *www.creativedecisions.net*). We show a simple example of performance evaluation of units in supply network structure composed from two suppliers, two producers, two distributors, and two customers. The initial paired comparisons of units were implemented. On the Super Decisions main window (see Figure 6), the structure of the system and global priorities of the units are shown.

The AHP and ANP have been static but for today's world analyzing is very important in time dependent decision making. The dynamic hierarchy process and dynamic network process (DHP/DNP) methods were introduced (Saaty, 2003). There are two ways to study dynamic decisions: structural, by including scenarios, and functional, by explicitly involving time in the judgment process. For the functional dynamics there are analytic or numerical solutions. The basic idea with the numerical approach is to obtain the time dependent principal eigenvector by simulation. The DNP provides weights for time periods t =1, 2,…,T. The connections are time dependent. The importance of the economic units, criteria, aspiration levels, and so forth change. There are many time dependent situations in network economy (Shapiro & Varian, 1999). For example, the dynamics of new product adoption can be expressed by the S-curve (see Figure 7).

Dependencies and feedbacks between units, products, or items in networks are important:

- Substitution and complementarity
- Positive and negative feedback

Figure 6. Super Decisions



Cluster Node Labels		customers		distributers		producers		suppliers	
		C1	C2	D1	D2	P1	P2	S1	S2
customers	C1	0.255169	0.255169	0.255169	0.255169	0.255169	0.255169	0.255169	0.255169
	C2	0.144831	0.144831	0.144831	0.144831	0.144831	0.144831	0.144831	0.144831
distributers	D1	0.257453	0.257453	0.257453	0.257453	0.257453	0.257453	0.257453	0.257453
	D2	0.209213	0.209213	0.209213	0.209213	0.209213	0.209213	0.209213	0.209213
producers	P1	0.064847	0.064847	0.064847	0.064847	0.064847	0.064847	0.064847	0.064847
	P2	0.035153	0.035153	0.035153	0.035153	0.035153	0.035153	0.035153	0.035153
suppliers	S1	0.025104	0.025104	0.025104	0.025104	0.025104	0.025104	0.025104	0.025104
	S2	0.008229	0.008229	0.008229	0.008229	0.008229	0.008229	0.008229	0.008229

Figure 7. Adaptation dynamics

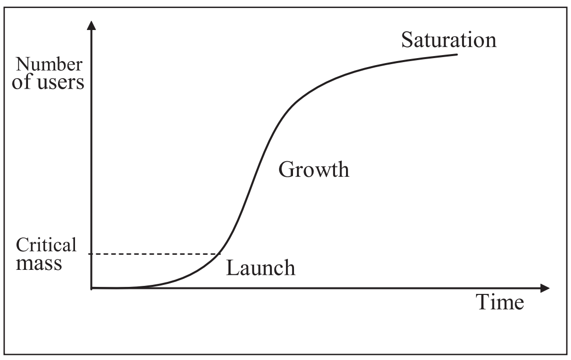
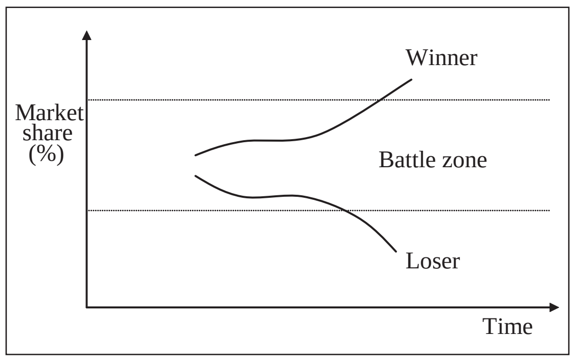


Figure 8. Positive feedback



Two elements A and B are complementary, if it holds for weights $w(\{A, B\}) > w(\{A\}) + w(\{B\})$.

Two elements A and B are substitute, if it holds for weights $w(\{A, B\}) < w(\{A\}) + w(\{B\})$.

The positive feedback can be expressed as the strong will be stronger and the weak will be weaker (see Figure 8).

Table 1. Dynamic comparisons

t	$a_{12}(t)$	$w_1(t)$	$w_2(t)$
0	1,13	0,53	0,47
0,1	1,66	0,62	0,38
0,2	2,38	0,7	0,3
0,3	3,26	0,77	0,23
0,4	4,27	0,81	0,19
0,5	5,29	0,84	0,16
0,6	6,24	0,86	0,14
0,7	7,04	0,87	0,13
0,8	7,65	0,88	0,12
0,9	8,10	0,89	0,11
1	8,41	0,9	0,1

Figure 9. Adaptation dynamics – example

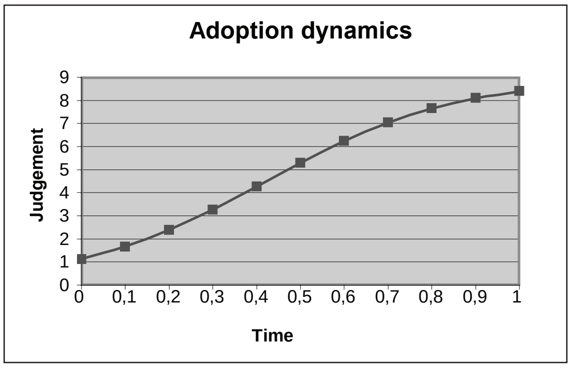
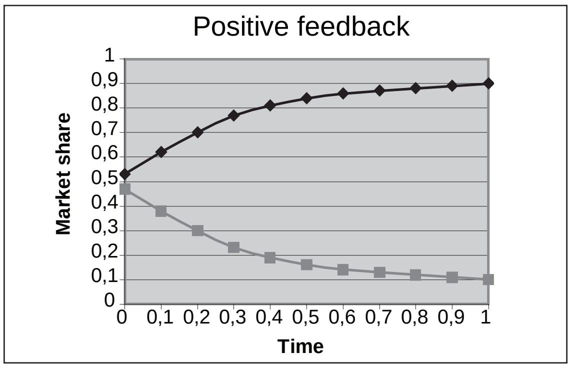


Figure 10. Positive feedback - example



Example 1

We use the DNP method for an illustration of positive feedback. The time dependent comparison of two products is expressed by the S-curve:

$$a_{12}(t) = \frac{9}{1+7\cdot 0{,}01^{t}}$$

and the paired comparison matrix:

$$\begin{bmatrix} 1 & a_{12}(t) \\ 1/a_{12}(t) & 1 \end{bmatrix}$$

The numerical data are shown in Table 1 and plotted in Figure 9 and Figure 10.

DYNAMIC GROUP-ALOP PROCEDURE

Cooperative problem solving is a process by which a joint decision is made by two or more parties. It is a communication designed to reach an agreement when the parties have some interests that are shared and others that are opposed. Both the communication and cooperation are required in the problem solving process. Some basic ideas of formal approaches of the problem solving can be introduced to cooperative decision making. There are two aspects of the problem solving: representation and searching. The state space representation introduces the concepts of states and operators. An operator transforms one state into another state. A solution could be obtained by a search process, then first applies operators to the initial state to produce new states and so on, until the goal state is produced.

The two phases' interactive approach for solving cooperative decision-making problems GROUP-ALOP is proposed (Fiala, 1997):

1. Finding the ideal solution for individual units.
2. Finding a consensus for all the units.

In the first phase every decision maker searches the ideal alternative by the assertivity principle. The general formulation of a multicriteria decision problem for an individual unit is expressed as:

$$z(x) = (z_1(x), z_2(x), \ldots, z_k(x)) \rightarrow \text{"max"}$$
$$x \in X,$$

where X is a decision space, x is a decision alternative and $z_1, z_2, \ldots, z_k$ are the criteria. The decision space is defined by objective restrictions and by mutual goals of all the decision makers in the aspiration level formulation. The decision alternative $\mathbf{x}$ is transformed by the criteria-to-criteria values $\mathbf{z} \in Z$, where Z is a criteria space. Every decision-making unit has its own criteria. People appear to satisfy rather than attempting to optimize. That means substituting goals of reaching specified aspiration levels for goals of maximizing.

We denote $\mathbf{y}^{(t)}$ aspiration levels of the criteria and $\Delta\mathbf{y}^{(t)}$ changes of aspiration levels in the step t. We search alternatives for which it holds:

$$\mathbf{z}(\mathbf{x}) \geq \mathbf{y}^{(t)}$$
$$\mathbf{x} \in X$$

According to heuristic information from results of the previous condition, the decision-making unit changes the aspiration levels of criteria for step t+1 with:

$$\mathbf{y}^{(t+1)} = \mathbf{y}^{(t)} + \Delta\mathbf{y}^{(t)}$$

We can formulate the multicriteria decision problem as a state space representation. The state space corresponds with the criteria space Z, where the states are the aspiration levels of the criteria $\mathbf{y}^{(t)}$, and the operators are changes of the aspiration levels $\Delta\mathbf{y}^{(t)}$. The start state is a vector of the initial aspiration levels and the goal state is a vector of the criteria levels for the best alternative. For finding the ideal alternative we use the depth-first search method with backtracking procedure. The heuristic information is the distance between an arbitrary state and the goal state.

We propose an interactive procedure, aspiration levels oriented procedure (ALOP), for multiobjective linear programming problems, where the decision space X is determined by linear constraints:

$$X = \{\mathbf{x} \in R^n;\ A\mathbf{x} \leq b\ ,\ \mathbf{x} \geq 0\}$$

and $z_i = \mathbf{c}_i\mathbf{x}$, i =1, 2, ... , k, are linear objective functions. Then $\mathbf{z}(\mathbf{x}) = C\mathbf{x}$ where C is a coefficient matrix of objectives.

The decision alternative $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ is a vector of n variables. The decision maker states aspiration levels $\mathbf{y}^{(t)}$ for the criteria values. There are three possibilities for aspiration levels $\mathbf{y}^{(t)}$. The problem can be feasible, infeasible, or the problem has a unique nondominated solution. We verify the three possibilities by solving the problem:

$$v = \sum_{i=1}^{k} w_i^+ \ d_i^+ \rightarrow \max$$

$C\mathbf{x} - \mathbf{d}^+ = \mathbf{y}^{(s)}$
$\mathbf{x} \in X$, $\mathbf{d}^+ \geq \mathbf{0}$

If it holds:

- $v > 0$, then the problem is feasible and d_i^+ are proposed changes $\mathbf{y}^{(t)}$ of aspiration levels which achieve a nondominated solution in the next step.
- $v = 0$, then we obtained a nondominated solution.
- The problem is infeasible, then we search the nearest solution to the aspiration levels by solving the goal programming problem:

$$v = \sum_{i=1}^{k} \frac{1}{\overline{z}_i} (d_i^+ + d_i^-) \rightarrow \min$$

$C\mathbf{x} - \mathbf{d}^+ + \mathbf{d}^- = \mathbf{y}^{(t)}$
$\mathbf{x} \in X$, $\mathbf{d}^+ \geq \mathbf{0}$, $\mathbf{d}^- \geq \mathbf{0}$

The solution of the problem is feasible with changes of the aspiration levels $\Delta\, \mathbf{y}^{(t)} = \mathbf{d}^+ - \mathbf{d}^-$. For small changes of nondominated solutions, the duality theory is applied. Dual variables to objective constraints in the problem are denoted as u_i, $i = 1, 2, \dots, k$.

If it holds:

$$\sum_{i=1}^{k} u_i \Delta y_i^{(t)} = 0 \,,$$

then for some changes $\Delta\, \mathbf{y}^{(t)}$ the value $v = 0$ is not changed and we obtained another nondominated solution. The decision maker can state k-1 small changes of the aspiration levels $\Delta y_i^{(t)}$, $i = 1, 2, \dots, k$, $i \neq r$, then the change of the aspiration level for criterion r is calculated from the previous equation.

The decision maker chooses a forward direction or backtracking. Results of the procedure ALOP are the path of tentative aspiration levels and the ideal solution.

In the second phase a consensus could be obtained by the search process and the principle of cooperativeness is applied. The heuristic information for the decision-making unit is the distance between the decision maker's proposal and the opponent's proposal. We assume that all the decision makers found their ideal alternatives. We propose an interactive procedure of GROUP-ALOP for searching a consensus.

For simplicity we assume the model with one supplier and one customer:

$z^1(\mathbf{x}) \rightarrow$ "max"
$z^2(\mathbf{x}) \rightarrow$ "max"
$\mathbf{x} \in X$

The decision-making units search a consensus on a common decision space X. The decision-making units change aspiration levels of the criteria y^1, y^2. The sets of feasible alternatives for the aspiration levels y^1 and y^2 are X^1 and X^2.

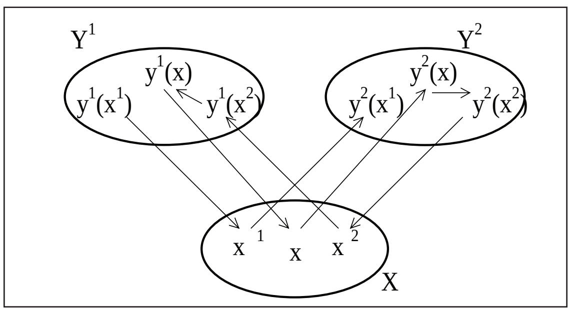
$z^1(\mathbf{x}) \geq y^1$	$z^2(\mathbf{x}) \geq y^2$
$\mathbf{x} \in X$	$\mathbf{x} \in X$

The consensus set S of the negotiations is the intersection of sets X^1 and X^2

$S = X^1 \cap X^2$

The negotiation process can be schematically represented by Figure 11.

Figure 11. Negotiation process



By changes of the aspiration levels the consensus set S is changed too. The decision makers search one element consensus set S by alternating of the consensus proposals. Independently on the meaning of local symbols, the cooperative decision-making model of negotiation is represented by the state vector $\mathbf{z}$. The image of a partner's proposal can be taken as aspiration levels in one's own criteria space. In searching for a consensus the distance between the proposals is heuristic information. The paths of the tentative aspiration levels can be used for the backtracking procedure. The forward directions can be directed by proposed new aspiration levels in step s+1:

$$\mathbf{y}^1(s+1) = (1-\alpha)\mathbf{y}^1(s) + \alpha\, \mathbf{z}^1(\mathbf{x}^2)\ ,$$
$$\mathbf{y}^2(s+1) = (1-\beta)\mathbf{y}^2(s) + \beta\, \mathbf{z}^2(\mathbf{x}^1)\ ,$$

where $\alpha, \beta \in <0,1>$ are the coefficients of cooperativeness.

Each decision maker applies cooperative strategy as long as the decision maker's partner does the same. If the partner exploits the decision maker on a particular step, the decision maker then applies the exploitative strategy on the next step and continues to do so until the partner switches back to the cooperative strategy. Under these conditions, the problem stabilizes with the decision makers pursuing the mutually cooperative strategy and receiving the consensus.

The current structure is a dynamic representation of the results of negotiation process among units. The proposed model is a discrete dynamic model and the cooperation of units is based on contracts and formal agreements achieved in negotiation process. The contracts are evaluated by multiple criteria as time, quality, and costs. There are different approaches to modeling multicriteria negotiation processes, such as utility concept, concept of pressure, and concept of coalitions. The set of modeling concepts can be a basis for developing negotiation support.

The combination of dynamic network process and dynamic version of GROUP-ALOP seems to be the appropriate method for the specific features of problems in network economy. The approach combines time dependent weights $\mathbf{w}(t)$ from DNP and time dependent aspiration levels $\mathbf{y}(t)$, t =1,2,…,T, from GROUP-ALOP.

The approach can be structured in the following phases:

1. The DNP is used for comparison of importance of the decision-making units, supplies, products, criteria, and so forth. There are dependencies among units. Inputs are objective and subjective information of units. Outputs are weights for time periods t =1,2,…,T.
2. The GROUP-ALOP approach is applied for negotiation process between decision-making units. For every time period in negotiation steps *s* the aspiration levels of criteria are changed to get a consensus. Inputs are the common decision space, criteria, and weights from the previous phase. Outputs are proposals for a consensus for time periods t =1,2,…,T.
3. Participants evaluate the proposals by own characteristics and make next proposals or determine the final values. Outputs are solutions for time periods t =1,2,…,T.

The proposed methodology can be used for solving various problems in dynamic supply networks. We will illustrate the approach on supplier selection problems.

SUPPLIER SELECTION PROBLEM

Supplier selection processes have received considerable attention in business (e.g., Simchi-Levi et al., 1999). The analysis and design of supply chains has been an active area of research (Tayur et al., 1999). Sourcing has come up as a very strategic issue in the management of supply chain networks in the modern era of global competition. Most production systems are organized as networks of units. Sourcing decisions have the capability of impacting the effectiveness of supply chain networks. Determining suitable suppliers in supply chain networks has become a key strategic issue. The nature of these decisions is usually complex and unstructured. The supplier selection problem is a multiple criteria problem. Many influencing factors such as price, quality, flexibility, and delivery performance must be considered to determine suitable suppliers. These influencing factors can be divided into quantitative and qualitative factors.

Generally, supplier selection is a multicriteria decision problem. The methods suggested in the related literatures can be classified into two categories:

- Weighting models.
- Mathematical programming models.

The weighting model, which focuses on commonly used evaluation criteria, includes:

- The linear scoring model (e.g., Timmerman, 1986).
- The analytic hierarchy process (AHP) model (e.g., Barbarosoglu, 1997).

The linear scoring model assigns weights and scores arbitrarily, for example, 1 for "unsatisfactory" and 5 for "outstanding." Hence, the model has an implicit and incorrect assumption; for example, "outstanding" is five times better than "unsatisfactory." The problem is avoided in the AHP model by converting the priorities into the ratings with regard to each criterion using pairwise comparisons.

Mathematical programming models are:

- The goal programming or multiobjective programming (e.g., Weber, 1993).
- The linear programming or mixed integer programming with the expression of multiple objectives as constraints (e.g., Rosenthal, Zydiak, & Chaudhry, 1995).

Objective function coefficients should be determined prior to making mathematical programming models. The drawback of goal programming and multiobjective programming is that they require arbitrary aspiration levels and cannot accommodate subjective criteria.

In recent years, research on supplier selection process has highlighted the relationships that exist between companies in supply chains. Supply strategies adopt the network approach to supplier selection and focus on the coordination and integration of different supply chains. Supplier-customer relationships are changing to a cooperative form. The impact of information sharing plays a crucial role. Supplier selection process becomes a multicriteria group cooperative decision-making problem. It is necessary to take into account the network and dynamic environment.

The decision-making process for supplier selection has some specifications:

- Multiple selection criteria
- Qualitative and quantitative criteria
- Group decision-making problem
- Cooperative behavior
- Incomplete information
- Networks
- Dynamic and uncertain environment

The scope of strategic fit refers to the functions and stages within a network system that coordinate strategy and target a common goal. Agile intercompany scope refers to a firm's ability to achieve strategic fit when partnering with network stages that change over time. A manufacturer may interface with a different set of suppliers depending on the product. The situation in reality is much more dynamic as product life cycles get shorter and companies try to satisfy the changing needs of individual customers. The level of agility becomes more important as the competitive environment becomes more dynamic.

The proposed model is a combination of advantages of traditional approaches with adding new approaches for new specifications of supplier selection problem. The approach combines the ANP and the GROUP-ALOP. The ANP provides weights **w** in the network model. The elements can be members of supply network, evaluating criteria, products, items, and so forth. By ANP, qualitative criteria can also be evaluated. The weights **w** are used in the GROUP-ALOP approach.

Today's world is dynamic. The proposed approach can be used for this dynamic environment. The AHP and ANP have been static but for today's world analyzing time dependent decision making is very important. The DHP/DNP methods were introduced (Saaty, 2003). There are two ways to study dynamic decisions: structural, by including scenarios, and functional, by explicitly involving time in the judgment process. For the functional dynamics there are analytic or numerical solutions. The basic idea with the numerical approach is to obtain the time dependent principal eigenvector by simulation. The DNP provides weights for time periods t =1,2,…,T.

The proposed model respects these specifications. The approach combines the analytic network process (ANP) and the aspiration level oriented procedure (ALOP). The ANP is a network generalization of AHP. The ALOP is based on goal programming approach. The GROUP-ALOP approach respects the supplier selection problem as a group decision-making problem. The proposed approach can be used for dynamic environment.

The approach is illustrated in Example 2. The approach is very flexible and a simple example can clarify basic insights only.

Example 2

Assume a manufacturer produces three products (P_i, i = 1,2,3) from two key parts (A, B). The product P_1 contains one piece of the part A, the product P_2 contains one piece of the part B, and the new product P_3 contains one piece of the part A and one piece of the part B. The manufacturer looks at three suppliers (S_j, j = 1,2,3) providing the two parts (A, B) and compares bids according the criteria of prices and reliability levels (p, r). The supplier S_1 produces parts A, the supplier S_2 produces parts B, and the supplier S_3 produces parts A and B. The supplier selection process is dynamic in time periods (t = 1,2,3).

The relative importance of criteria (p, r) for parts (A, B) changes dramatically in time periods. The criteria are dependent on each other and the parts are dependent on each other for the product P_3 also. The DNP method was used for weights calculation. For simplicity, we assume that weights are the same for the parts.

In every time period there will be in progress negotiation process with suppliers. The firm negotiates quantity, price, and reliability levels of parts A and B. The weights **w** are used in the GROUP-ALOP approach. In every negotiation step s aspiration levels are changed. Results of the negotiation process are price and reliability levels for time periods.

The decision set for the firm is restricted by forecasted demands $D_i(t)$, i = 1,2,3, t = 1,2,3, and capacities. Unit profits $c_i(t)$, i = 1,2,3, t = 1,2,3, are

Table 2. Weights of criteria

t	$w_p(t)$	$w_r(t)$
1	0.8	0.2
2	0.5	0.5
3	0.3	0.7

Table 3. Negotiated final price and reliability levels

t	$p_{1A}(t)$	$p_{2B}(t)$	$p_{3A}(t)$	$p_{3B}(t)$	$r_{1A}(t)$	$r_{2B}(t)$	$r_{3A}(t)$	$r_{3B}(t)$
1	2	3	3	4	0.7	0.7	0.7	0.7
2	2.5	3.5	2	3	0.75	0.75	0.8	0.8
3	3	4	1.5	2.5	0.8	0.8	0.9	0.9

Table 4. Forecasted demands and unit profits

t	$D_1(t)$	$D_2(t)$	$D_3(t)$	$c_1(t)$	$c_2(t)$	$c_3(t)$
1	50	80	10	5	6	4
2	30	60	30	4	5	7
3	10	30	100	3	4	10

Table 5. Production quantities and profits

t	$x_1(t)$	$x_2(t)$	$x_3(t)$	$z(t)$
1	20	80	0	580
2	10	60	30	550
3	0	0	100	1000

Table 6. Required supplies

t	$q_{1A}(t)$	$q_{2B}(t)$	$q_{3A}(t)$	$q_{3B}(t)$
1	20	80	0	0
2	10	60	30	30
3	0	0	100	100

dependent on price and reliability levels of parts A, B, and among others.

The production quantities $x_i(t)$, i = 1,2,3, t = 1,2,3 are bounded by forecasted demands:

$$x_i(t) \le D_i(t) \; i = 1,2,3,\; t = 1,2,3.$$

The firm capacity makes possible to produce 100 final products in every time period:

$$x_1(t) + x_2(t) + x_3(t) \le 100, \; t = 1,2,3.$$

For simplicity, assume the firm evaluates the negotiation position by expected profit $z(t)$, t = 1,2,3,

$$z(t) = c_1(t)x_1(t) + c_2(t)x_2(t) + c_3(t)x_3(t) \to \max$$

The solution of the decision problem is illustrated in Table 5.

The required supplies $q_{jA}(t)$, $q_{jB}(t)$, i = 1,2,3, t = 1,2,3, are calculated in Table 6.

THREE LAYER MODELING FRAMEWORK

The proposed approach can be completed with a three-layer framework for modeling of coordination process of units in the dynamic supply networks. The learning process is also involved in the framework. The process modeling of coordination of units in supply networks in general is a complex problem based on several kernel ideas. The framework of the proposed discrete dynamic model is separated to three parts (Fiala, 2003):

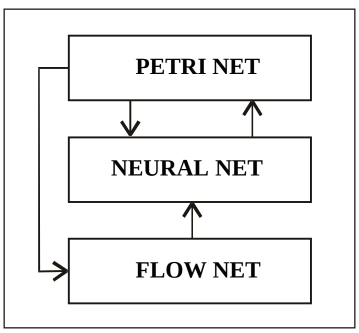
- Deterministic
- Logical
- Stochastic

According to these three parts, the modeling framework is composed from three inter-related network structures:

- Flow net
- Petri net
- Neural net

The supply network can be shown as a set of units interconnected by material, information, and financial flows from initial suppliers to ultimate customers. Petri net is used to coordinate asynchronous events of different units in the supply network and to model negotiation process. A neural net serves as an instrument for inductive learning of negotiation strategies.

Figure 12. Sandwich structure of three layer model



The resulting model (see Figure 12) consists of three layers where the soft neural layer is situated in the central part as in the real sandwich. The logical part of the model is realized as a Petri net with inhibitory arcs (Braunl, 1993) using only logical variable vector **x** for the state description. Its component x_i is true when the *i*-th node of Petri net is occupied by the token. The deterministic part of the model is called the basic model with the state represented by the real vector **z**. The stochastic part of the model is realized as an artificial neural network (ANN) (Haykin, 1994) with the state represented by probability vector **y.** The whole model is described in vector form by three difference equations of:

$$\mathbf{x}_{t+1} = \mathbf{f}(\mathbf{x}_t, \mathbf{y}_t)$$
$$\mathbf{y}_t = \mathbf{g}(\mathbf{x}_t, \mathbf{z}_t)$$
$$\mathbf{z}_{t+1} = \mathbf{h}(\mathbf{x}_t, \mathbf{z}_t)$$

The main characteristic of the model is the partial interconnectivity. Petri net cannot use the state of the flow net and the flow net cannot use the state of an artificial neural network. Then the Petri net changes its state by internal parallel process. The tokens can be set by the artificial neural network with probability **y**. The basic model has its own deterministic dynamics and uses the Petri net state as an input signal. The artificial neural network layer plays the role of a soft and learnable coordinator. Using the states of Petri net and the flow net, the artificial neural network changes its own state without any internal dynamics. According to state **y** the *i*-th token of Petri net is set with the probability y_i. The main advantage of the ANN is the ability of learning its own coordinator role. Using the weight vector **w**, the ANN layer is extended to the form:

$$\mathbf{y}_t = \mathbf{g}(\mathbf{x}_t, \mathbf{z}_t, \mathbf{w}_t)$$

In the period of learning the weight vector **w** varies in time by any paradigm (Haykin, 1994). The learning process is described as:

$$\mathbf{w}_{t+1} = \mathbf{g}(\mathbf{x}_t, \mathbf{z}_t, \mathbf{w}_t)$$

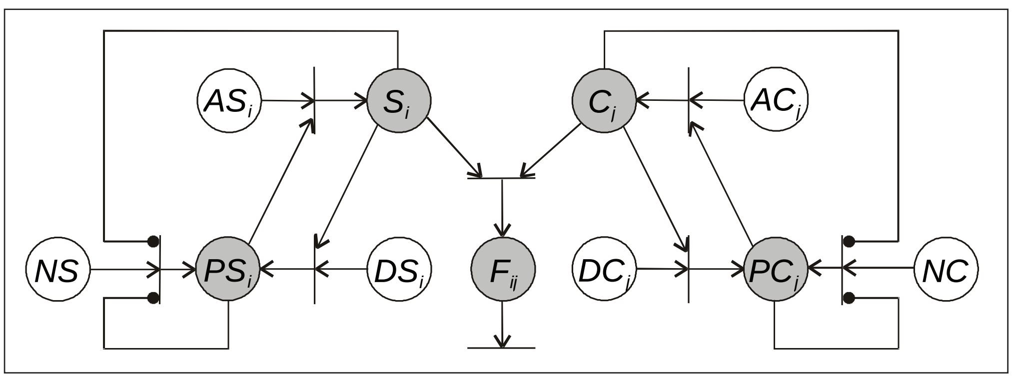
Petri net is used to coordinate asynchronous events of different units in the supply network and to model the negotiation process. The transaction is realized when both supplier and customer are prepared to make it. The *i*-th supplier and *j*-th customer negotiation is the basic element of a Petri net layer. The S_i, PS_i, C_j, PC_j, and F_{ij} nodes are read only while the remaining nodes can be set by artificial neural network. When the token is in the node S_i, the *i*-th supplier is prepared for the transaction. In the opposite case the token is in the node PS_i. The C_j and PC_j nodes represent the state of the *j*-th customer with the same meaning. The node F_{ij} contains a token just after transaction agreement. The remaining nodes are filled by the tokens from the ANN layer. The nodes AS_i and AC_j are enable to activate the *i*-th supplier and *j*-th customer. The nodes DS_i and DC_j are enable to deactivate the *i*-th supplier and *j*-th customer. The nodes NS and NC are enable to introduce new suppliers and customers. The Petri net layer of real negotiation model contains basic elements (see Figure 13) in complex structure according to the structure of flow network.

The artificial neural network layer of the model is able to learn the negotiation process strategy. The main advantage of the artificial neural network is the ability of learning its own coordinator role. The learning process is then equivalent to fitting the model parameters.

CONCLUSION

The new, very important features in supply systems are dynamic network structure and cooperative decision making by coordination. Information asymmetry is one of the most powerful sources of inefficiencies in supply networks. The proposed approach combines advantages of the traditional methods with new procedures. There is a combination of system dynamics modeling, dynamic network process and the dynamic version of GROUP-aspiration levels oriented procedure. The approach is very flexible. The aim is coordination of units, as well as managing supplier-customer relations. Building of different types of strategic partnerships and different type of contracts among participants can significantly reduce or eliminate inefficiency in supply networks. The

Figure 13. Petri net layer



expected result is a mutually beneficial, win-win partnership that creates a synergistic network in which the entire network is more effective than the sum of its individual units. Supply network partnership leads to increased information flows, reduced uncertainty, and a more profitable supply network. The ultimate customer will receive a higher quality, cost effective product in a shorter amount of time.

The proposed approach is illustrated on supplier selection problem. Supplier selection process is a very important strategic issue. The process is very complex. There are new trends in supply process. The new, very important features in supplier selection problem are network structure of suppliers and items, dynamic connections, and cooperative decision making. The proposed model captures important trends in supply process. The approach combines advantages of the traditional approaches for supplier selection problems, weighting models, and mathematical programming models and adds approaches for new specifications of supplier selection problem. The aim is not only a supplier selection but managing supplier-customer relations also.

FUTURE RESEARCH DIRECTIONS

Global competition and rapidly changing customer requirements are demanding increasing changes in manufacturing environments. Enterprises are required to constantly redesign their products and continuously reconfigure their manufacturing systems. Traditional approaches to manufacturing systems do not fully satisfy this new situation. Business models of the past are no longer adequate due to the evolution of the business environment.

The network economy is a term for today's global relationship among economic units characterized by massive connectivity. The central act of the new era is to connect everything to everything in deep Web networks at many levels of mutually interdependent relations, where resources and activities are shared, markets are enlarged, and costs and risk are reduced. Business processes can take advantage of networks in several ways; they have four properties:

- **Ubiquity:** It refers to the ever-present power of a network
- **Pooling and sharing:** Networks allow reciprocity; in this sense networks can became the platform for business interaction
- **Specialization and alliance:** When a firm becomes part of a network it can specialize in a specific task while other parts can specialize in complementary tasks
- **Intelligence:** Networks enable new value activities to create value with information

Important trends in organization are spread from collaborations to operational challenges to customer participation. These trends define a new set of requirements for the collaborating organizations at all levels in product life cycle. These formations of teams of collaborating trading partners will be dynamic in nature as customer requirements may change. Active customer intervention in all the decision making in the product realization process enables the manufacturing organization to be customer-centric and truly driven by the customers. Increasing uncertainty of supply networks, globalization of businesses, proliferation of product variety, and shortening of product life cycles have forced organizations to look beyond their walls for collaboration with supply chain partners. Developing improved control polices requires simulation of the physical, financial, decision, and information flows involved.

Manufacturing organizations are adopting the collaborative business paradigm in recent years. New Web-centric collaborative technologies help them to share mission critical product and manufacturing information with trading partners and value-added suppliers. It allows customers, suppliers, and business partners to form a virtual enterprise by way of collaboration to achieve their common business goals. The paradigm calls for new requirements in the way the product and team data are managed among collaborating design and manufacturing partners.

Many authors have proposed that computational intelligence will bring the flexibility and efficiency needed by manufacturing systems. New paradigms in computational intelligence as hybrid networks, connectionist systems, evolutionary algorithms, and fuzzy systems are being explored by designers, implementers, and managers to improve understanding, evaluation, and assessment of such systems. Modeling and simulation techniques provide fundamental support at all stages of the product life cycle.

Proposed three layer modeling framework of cooperative problem solving can be very helpful by the integration process of network units. Research work continues and testing on real applications is needed. Further developments will be devoted to improvements and modifications of this framework for solving of real problems. The main future research directions are oriented on:

- Networking.
- New trends in organizations.
- New trends in information and communication technologies.
- New paradigms in computational intelligence.

ACKNOWLEDGMENT

The research project was supported by Grant No. 402/05/0148 from the Grant Agency of Czech Republic.

REFERENCES

Barbarosoglu, G., & Yazgac, T. (1997). An application of the analytic hierarchy process to the supplier selection problem. *Production and Inventory Management Journal*, *38*(1), 14-21.

Braunl, T. (1993). *Parallel programming.* New York: Prentice Hall.

de Souza, R., Song, Z., & Chaoyang, L.(2000). Supply chain dynamics and optimization. *Integrated Manufacturing Systems*, *11*(5), 348-364.

Fiala, P. (1997). Models of cooperative decision making. In T. Gal & G. Fandel (Eds.), *Multiple criteria decision making* (pp. 128-136). Heidelberg: Springer.

Fiala, P. (2003). Modeling of relations in supply chains. *Vision: The Journal of Business Perspective (Special Issue on Supply Chain Management), 7*, 127-131.

Fiala, P. (2005). Information sharing in supply chains. *OMEGA: The International Journal of Management Science*, *33*, 419-423.

Forrester, J.W. (1961). *Industrial dynamics.* Cambridge, MA: MIT Press.

Haykin, S. (1994). *Neural networks.* New York: Macmillan.

Lee, H.L., Padmanabhan, P., & Whang, S. (1997). Information distortion in a supply chain: The bullwhip effect. *Management Science*, *43*, 546-558.

MacBeth, D.K., & Ferguson, N. (1994). *Partnership sourcing: An integrated supply chain management approach*. London: Pitman Publishing.

Rosenthal, E.C., Zydiak, J.L., & Chaudhry, S.S. (1995). Vendor selection with bundling. *Decision Sciences*, *26*(1), 35-48.

Ruth, M., & Hannon, B. (1997). *Modeling dynamic economic systems*. New York: Springer.

Saaty, T.L. (1996). *The analytic hierarchy process*. Pittsburgh: RWS Publications.

Saaty, T.L. (2001). *Decision making with dependence and feedback: The analytic network process*. Pittsburgh: RWS Publications.

Saaty, T.L. (2003). Time dependent decision-making: Dynamic priorities in AHP/ANP: Generalizing from points to functions and from real to complex variables. In *Proceedings of the 7th International Conference on the Analytic Hierarchy Process*, Bali, Indonesia, (pp.1-38).

Shapiro, C., & Varian, H. (1999). *Information rules: A strategic guide to the network economy*. Boston: Harvard Business School Press.

Simchi-Levi, D., Kaminsky, P., & Simchi-Levi, E. (1999). *Designing and managing the supply chain: Concepts, strategies and case studies*. Boston: Irwin/McGraw-Hill.

Swaminathan, J., Smith, S., & Sadeh, N. (1998). Modeling supply chain dynamics: A multiagent approach. *Decision Sciences*, *29*(3), 607-632.

Tayur, S., Ganeshan, R., & Magazine, M. (1999). *Quantitative models for supply chain management*. Boston: Kluwer.

Thomas, D.J., & Griffin, P.M. (1996). Coordinated supply chain management. *European Journal of Operational Research*, *94*, 1-15.

Timmerman, E. (1986). An approach to vendor performance evaluation. *Journal of Purchasing and Materials Management*, *26*(4), 2-8.

Towil, D.R. (1996). Industrial dynamics modelling of supply chains. *International Journal of Physical Distribution & Logistics Management*, *26*(2), 23-42.

Vidal, C.J., & Goetschalckx, M. (1997). Strategic production-distribution models: A critical review with emphasis on global supply chain models. *European Journal of Operational Research*, *98*, 1-18.

Weber, C.A., & Current, J.R. (1993). A multiobjective approach to vendor selection. *European Journal of Operational Research*, *68*, 173–184.

Zeng, D.D. (2001). Managing flexibility for interorganizational electronic commerce. *Electronic Commerce Research*, *1*(1-2), 33-51.

ADDITIONAL READING

Ahuja, R.K., Magnanti, T.L., & Orlin, J.B. (1993). *Network flows*. Englewood Cliffs: Prentice-Hall.

Aoki, M. (1998). *New approaches to macroeconomic modeling*. Cambridge University Press.

Brandenburger, A.M., & Nalebuff, B. (1996). *Co-opetition*. New York: Currency/Doubleday.

Economides, N. (1996). The economics of networks. *International Journal of Industrial Organization*, *14*(2).

Fingar, P., & Aronica, R. (2001). *The death of "e" and the birth of the real new economy*. Tampa: Meghan-Kiffer Press.

Frolund, S. (1996). *Coordinating distributed objects. An actor-based approach to synchronization*. Cambridge: MIT Press.

Hayes, R.H., Wheelwright, S.C., & Clark K.B. (1988). *Dynamic manufacturing*. New York: Free Press.

Ishida, T. (Ed.). (1998). *Community computing: Collaboration over global information networks*. New York: Wiley.

Kelly, K. (1998). *New rules for the new economy: 10 radical strategies for a connected world*. New York: Viking Press.

Kreps, D. (1991). *Game theory and economic modelling.* Oxford University Press.

Kusiak, A. (2000). *Computational intelligence in design and manufacturing.* New York: Wiley.

Lorenz, H. (1989). *Nonlinear dynamical economics and chaotic motion*. Berlin: Springer.

Macho-Stadler, I., & Pérez-Castrillo, J.D. (2001). *An introduction to the economics of information incentives and contracts*. Oxford University Press.

Mueller, J.P. (1996). *The design of intelligent agents.* Berlin: Springer.

Nagurney, A. (1999). *Network economics: A variational inequality approach*. Dordrecht: Kluwer Academic Publishers.

Nagurney, A. (Ed.). (2003). *Innovations in financial and economic networks.* Cheltenham: Edward Elgar Publishers

Nagurney, A., & Dong, J. (2002). *Supernetworks: Decision-making for the information age*. Cheltenham: Edward Elgar Publishers.

Nilsson, A.S., Tolis, C., & Nellborn, C. (Eds.). (1999). *Perspectives on business modelling: Understanding and changing organizations*. Berlin: Springer.

Oliver, J. (1997). Artificial agents learn policies for multi-issue negotiation. *International Journal of Electronic Commerce, 1*(4).

Ruth, M., & Hannon, B. (1997). *Modeling dynamic economic system*. New York: Springer.

Shy, O. (2001). *The economics of network industries.* Cambridge University Press.

Steuer, R.E. (1986). *Multiple criteria optimization: Theory, computation, and applications*. New York: Wiley.

Timmers, P. (1999). *Electronic commerce: Strategies and models for business-to-business trading.* Chichester: John Wiley.

Wang, J., & Kusiak, A. (Eds.). (2001). *Computational intelligence in manufacturing handbook*. Boca Raton: CRC Press.

Zeleny, M. (2005). *Human systems management: Integrating knowledge, management and systems.* New Jersey: World Scientific.

Chapter XIII
Modeling with System Archetypes: A Case Study

Mahendran Maliapen
IGate Consulting, Canada

ABSTRACT

This chapter examines the application of system archetypes as a systems development methodology to create simulation models. Rapid organizational change and need to adapt to new business models limits the lifespan of both the databases and software applications. With the information representation permitted by archetypes, diagnostic analysis and can help to evolve generic classes and models for representing the real world. Archetypes do not describe any one problem specifically. They describe families of problems generically. Their value comes from the insights they offer into the dynamic interaction of complex systems. The case of a healthcare system is discussed here. As part of a suite of tools, they are extremely valuable in developing broad understandings about the hospital and its environment, and contribute more effectively to understanding problems. They are highly effective tools for gaining insight into patterns of strategic behavior, themselves reflective of the underlying structure of the system being studied. Diagnostically, archetypes help hospital managers recognize patterns of behavior that are already present in their organizations. They serve as the means for gaining insight into the underlying systems structures from which the archetypal behavior emerges. In the casemix model of the hospital, the investigation team discovered that some of the phenomena as described by these generic archetypes could be represented. The application of system archetypes to the strategic business analysis of the hospital case reveals that it is possible to identify loop holes in management's strategic thinking processes and it is possible to defy these fallacies during policy implementation as illustrated by the results of the archetype simulation model. In this research study, hospital executives found that policy modification with slight variable changes helps to avoid such pitfalls in systems thinking and avoid potentially cost prohibitive learning had these policies been implemented in real life.

INTRODUCTION

The system development methodologies applied to information systems are based on the premise that the domain concepts with the use of information engineering tools represent business processes and data directly into software and database models (Jacobsen, 1992; Martin & Odell, 1992; Rumbaugh et al., 1991; Walden & Nerson, 1995). The approach allows rapid prototyping and quick testing.

However, rapid organizational change and need to adapt to new business models limits the lifespan of both the databases and software applications. The lifecycles of change to update these systems with the organic growth of organizations is costly and often requires rearchitecting the very foundational concepts these software systems are based on. In UML, the foundation model and classes are clear (Booch, 1994). The time lag between model concept, use cases, and skeleton software code by applying forward and reverse software engineering to incorporate base changes forces continual rebuilding, testing, and redeployment of systems. These methodologies do not support the representation of information in such a way that would enable system diagnosis, analysis, and simulation to assist and to gauge the "appropriateness" of its real world representation. If these diagnostic changes are not made, the systems suffer creeping obsolescence, and as a result, diminishing utility over time.

The term *archetype* is used to denote knowledge level models which define valid information structures. A system dynamics archetype can be defined as a molecular building block of stocks and flows for a model structure. An archetype is therefore an abstraction of a feedback structure that is known to generate a particular type of behavior. It describes a representative unique behavior, which is the characteristic for that thumbnail part of the greater structure within the model. It can be considered a "plug and play" module into the structure of any other model where such characteristic behavior is required. With the publication of *The Fifth Discipline* (Senge, 1990), there was dearth of interest in using the science of archetypes; explicit system modeling of complex issues can be achieved by examining the whole system. The goal is to understand how the feedback structure of a system contributes to its dynamic behavior. The stocks and flows, the polarities of feedback loops interconnecting them, and shifts in the significance or dominance of various loops in the structure help contribute to this understanding.

At a molecular level, there are two distinct types of causal loop structures, namely balancing and reinforcing loops. The behavior mode of a simulation at any given time is determined by the strongest feedback loop(s). The overall pattern of behavior over time can be related to changing relative strengths of feedback loops. A system with one balancing and one reinforcing loop produces S-shaped development if the reinforcing loop dominates in the first phase, and the balancing loop dominates in the second phase. This form of information representation is consistent and universal.

System archetypes are patterns in corporate structure or behavior that recur again and again. They are that part of an organization, which represents keys to "pattern recognition" activities, incorporated in the discipline of system thinking. Nine archetypes are generally acknowledged as forming the set of tools that reveal patterns of behavior in systems (Senge, 1990).

- Limits to growth (a.k.a. limits to success)
- Shifting the burden
- Eroding goals
- Escalation
- Success to the successful
- Tragedy of the commons
- Fixes that fail
- Growth and under investment
- Accidental adversaries

Archetypes are useful for gaining insight into the "nature" of the underlying problem and for offering a basic structure or foundation upon which the model can be further developed and constructed. The archetypes are rarely sufficient models in and of themselves. They are generic in nature and generally fail to reveal important variables that are part of the real system structure. Without an explicit awareness of these real variables, it is difficult for managers to pinpoint specific leverage points where changes in structure can achieve sustainable changes in system behavior.

The concept of using system archetypes as a diagnostic tool was applied to the healthcare industry with particular reference to hospitals to develop a decision support dashboard for senior hospital executives. These business models were developed to help hospital executives to future proof their long-term and strategic decisions. These management decisions were applied in an interactive simulation model which used a two year time horizon to compute and forecast scenarios. The systems development methodology was based on system archetypes which helped the simulation model developers to identify and test a combination of instrumental simulation parameters.

METHODOLOGY

In general, Kim and LannonKim (1994) suggest that in seeking any of the above system archetypes in causal maps, the following investigative steps should be applied in sequence:

- Identify the symptoms of the problem.
- Map current interventions and how they are expected to rectify the problem.
- Map any unintended consequences of the intervention.
- Identify the fundamental causes of the problem.
- Ensure that causes and effects are linked.
- Identify any high leverage interventions.

The system archetypes are patterns of behavior that emerge from the underlying system structure. They can be used diagnostically to reveal insights into the structure that already exists, or prospectively to anticipate potential problems and/or problem symptoms. Archetypes do not describe any one problem specifically. They describe families of problems generically. Their value comes from the insights they offer into the dynamic interaction of complex systems (Kim et al., 1994). Figure 1 illustrates the development of an archetype dynamic hypothesis focusing on systemic root causes and issues and leveraging the benefits of strategic hospital policy changes. In the dynamic hypothesis, arrows represent flows. The polarity of influences is specified as reinforcing (in the same direction with 'S' sign) or in an opposing direction (with a 'O' sign). If a change of the two variables is in the same direction, then the link is "S" type. On the other hand, if a change is in an opposite direction, then the link is "O" type. All the variables are linked, creating feedback loops. A detailed description of the causal loops, links, and secondary effects that were developed through focus group sessions with hospital executives and clinicians is available from the author.

In this cycle of the systems development, semantics play an important part in the accurate labeling of the factors. It is usually best to use terms that are the same as or similar to those used by hospital executives. Nevertheless, Hayakawa (1990) distinguishes between the language of reports and the language of affection: the former is for communicating information, the latter for communicating emotions. The "ladder of abstraction" is an important semantic tool. It made hospital executives conscious of the remoteness of variable names from concrete facts, direct observable data, and the descriptions provided. System dynamics models are generally constructed at a high level,

Figure 1. Dynamic hypothesis: System archetypes approach

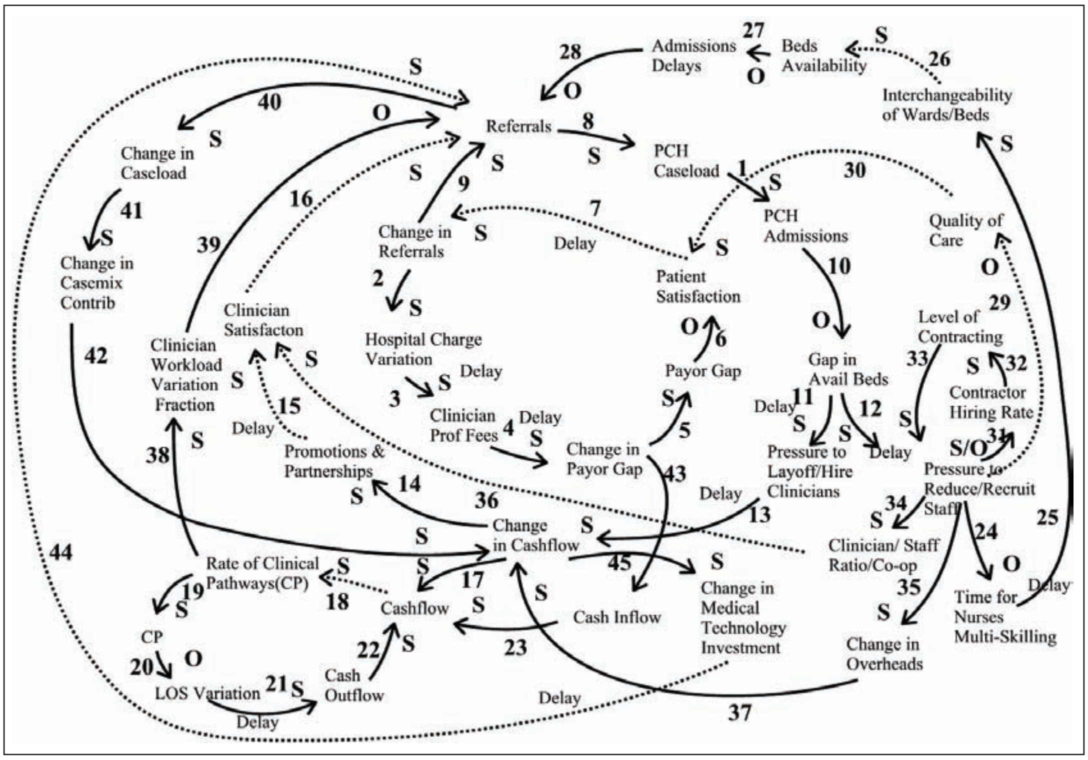
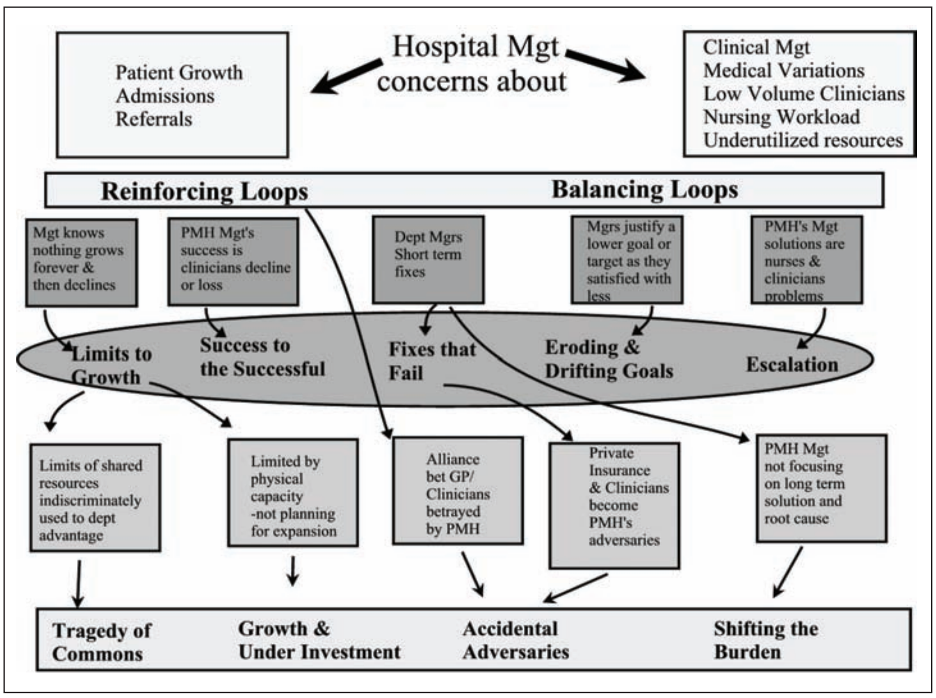


Figure 2. Taxonomy of system archetypes



while most people's day-to-day experiences are rooted at a much lower level. The ladder of abstraction can be useful as walking down the ladder can bring participants to a level at which the dynamics makes more business sense to them.

Hospital service and government performance benchmarks require senior hospital executives to measure performance against key targets for cashflow, bed utilization, available beds, quality of care, patient satisfaction, and clinician utilization. Several strategic policies such outsourcing, clinical pathways development, and market growth policies are central to the development of the hypothesis. Hospital and clinician professional fees/costs are assumed at an aggregate level.

Archetypes are characterized by a combination of balancing and reinforcing loops presented in the analysis at *http://www.systems-thinking.org/arch/arch.htm* and by Wolstenholme (2003). A summary of the potential combinations of balancing and reinforcing loops for the hospital business environment is presented in Figure 2.

CAUSAL LOOP ANALYSIS AND RESULTS: APPLYING THE SYSTEM ARCHETYPES

A total of 11 causal loops are labeled as B1 to B7 in Figure 3, B8 and B9 in Figure 4, and B10 in Figure 5 (B10 doubles as a reinforcing and balancing loop). These were identified with the system archetype approach from the dynamic hypothesis in Figure 1 by applying causal loop analysis. These interconnecting feedback loops are the basic structural elements in systems that generate dynamic behavior (Forrester, 1968; Goodman, 1974).

A complex simulation model based on this analysis was constructed and validated against two years of historical casemix patient episodic data. The output of the results of the simulation model runs are captured in the graphs that follow and the interactions of the variables is described to explain the behavior of the models. A description of selected archetypes identified together

Figure 3. Archetype model: Causal loop analysis

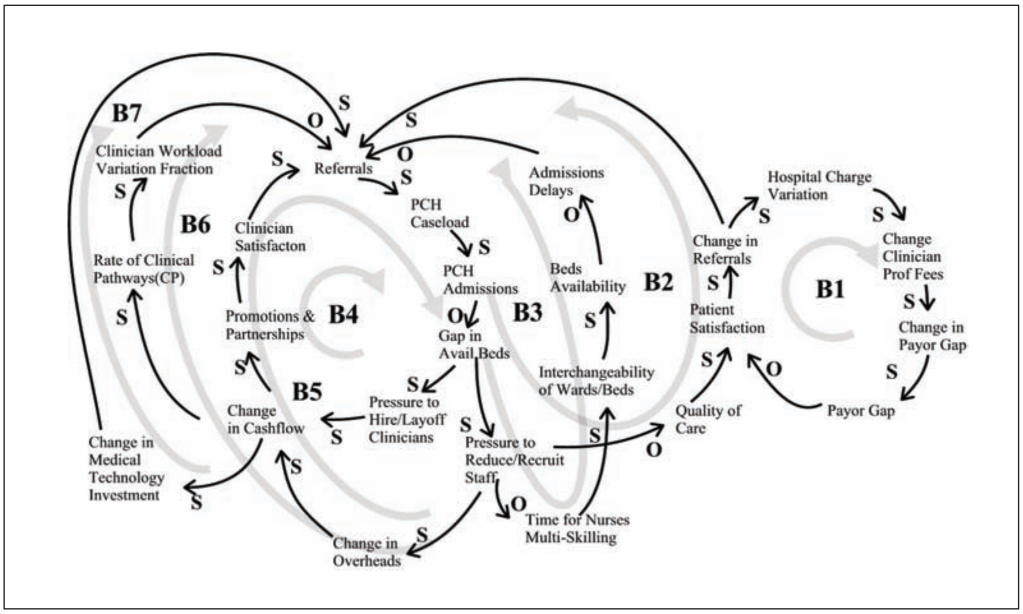


Figure 4. Eroding goals archetype with Loops B9 and B8

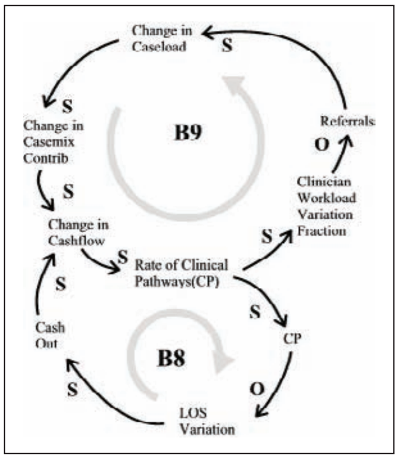
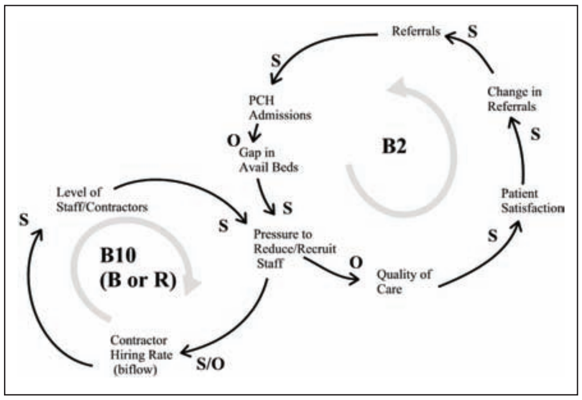


Figure 5. Drifting goals archetype with LoopsB10 and B2



with the simulation graphs from the model is provided below.

CHANGES IN VOLUME OF PATIENT REFERRALS

Balancing Loops B1 and B2 are opposing each other as shown in Figure 3. B1 describes the hospital charge variations for clinicians' use of hospital facilities, which can be reduced by the hospital's head office (Graph 5 in Figure 7) in a bid to entice more clinicians to reduce their professional fees. A potential reduction in their fees would result in a change in the payor gap of the patient (Graph 5 in Figure 6) and increase patient satisfaction (Graph 4 in Figure 6). This induces more patients with the desire for treatment by clinicians thereby causing a change in referrals. So, the B1 loop reflects the hospital's head office's strategic action of lowering hospital charges to achieve the goal of increased referrals (Graph 3 in Figure 6).

In Loop B2, the pressure on departmental managers to reduce medical and nursing staff (including part time contractors) arising from the unused capacity of hospital beds is shown. This in turn raises alarm and staff's reaction surfaces from delays to patient's requests and administration of medication with lower service levels.

The archetypal behaviors over time are tested and validated in the simulation model and shown in Figures 6 and 7. The figures show the comparative results between localized management action (departmental action) and the centralized actions of corporate office of hospital (head office action).

The reduction of this departmental pressure in Loop B2 relaxes the situation and maintains quality of patient care and patient satisfaction. The reduction in referrals is observable when the pressure to reduce/recruit staff is relaxed (Graph 3 on Figures 6 and 7) and is a consequence of the lowering of departmental targets when the pressure is off. It is important to recognize that even if staff is recruited, although there are training delays, the level of quality care is still maintained. So, B2 is the counteracting loop to B1 as B2 results in lowering the goal of higher referrals. In terms of a generic archetype, it is indicative of a drifting goals archetype.

Figure 6. Interactions of Loops B2 and B1: Head office action

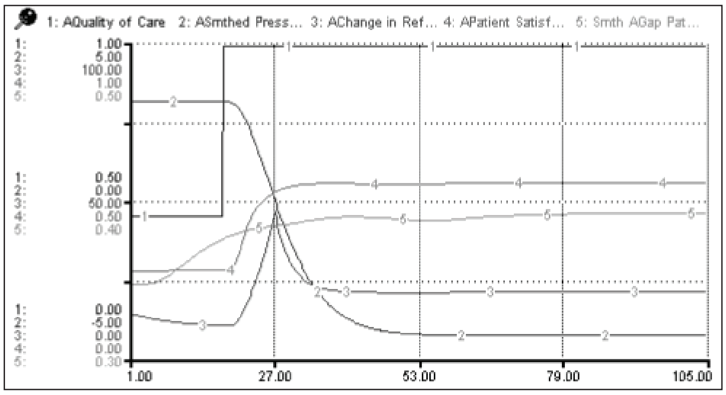
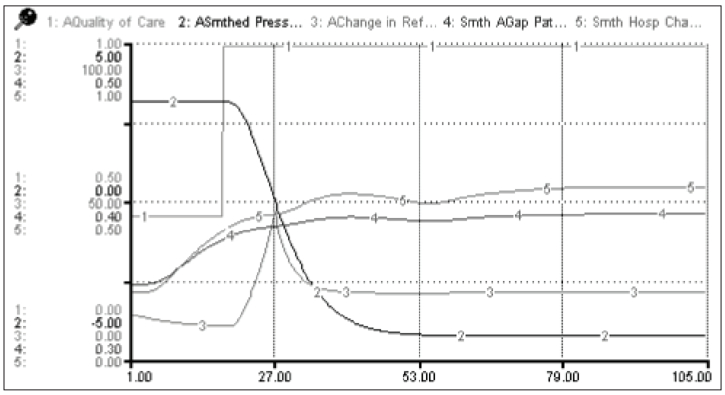


Figure 7. Interactions of Loops B2 and B1: Departmental action



GAPS IN AVAILABILITY OF HOSPITAL BEDS

Similarly, in Figure 3, balancing Loops B2 and B4 interact in opposition to each other to produce an eroding goals archetype. In Loop B4, the opposing effect of management pressure increases cashflow (Graph 1 in Figure 9 shows the negative cashflow decreases) by reducing the number of practicing clinicians with low patient volume. The increased cashflow provides the justification for increased expenditure on improving provider partnerships/ alliances (Graph 1 in Figure 8 increases in Week 76) with consortiums.

It also supports the hospital's marketing promotions/efforts to improve patient census. The net effect is that clinician satisfaction improves and thereby increases the volume of referrals to fill the gap in available beds at the hospital (Graph 3 in Figure 8 shows the reduction in the available bed days gap). This is a classic eroding goals archetype with B2 reducing the gap in available beds by lowering departmental goals and B4 increasing clinician satisfaction with more referrals to fill the gap. Over the simulation, balancing Loop B4 shows dominance and the bed utilization increases as the gap of available bed days decreases (Graph 3 in Figure 9).

Figure 8. Interactions of Loops B4 and B2—drifting goals

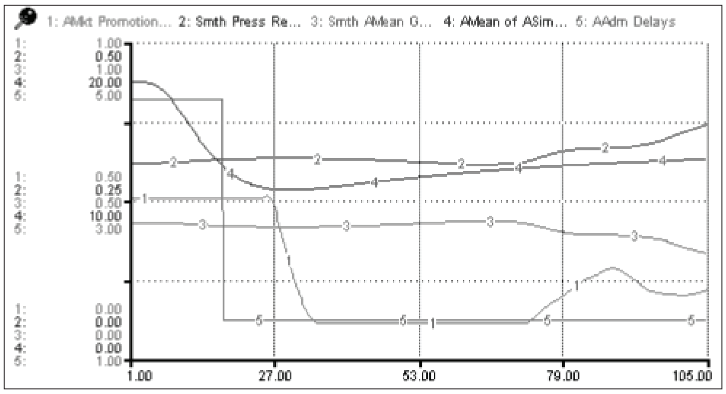


Figure 9. Interactions of Loops B4 and B3—drifting goals

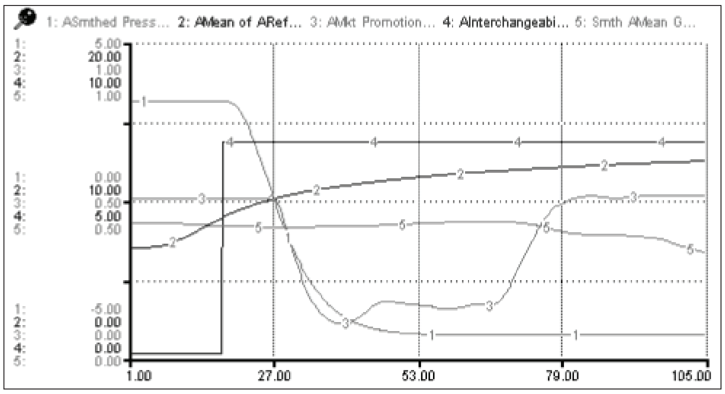
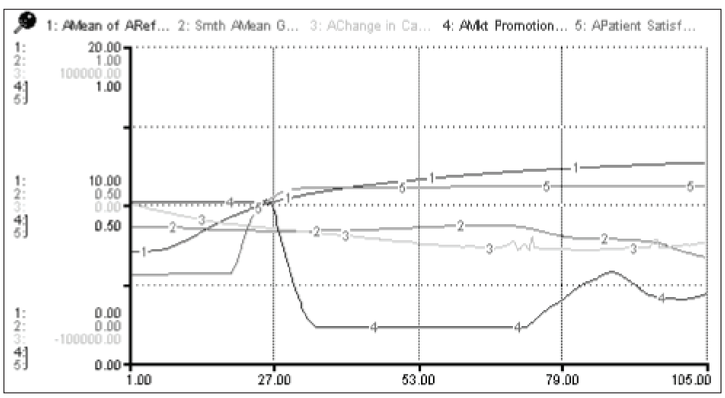


Figure 10. Interaction of loops B5 and B2



However, in B5, the pressure to reduce nursing staff also produces the effect of increasing net cashflow which provides management with the financial arsenal (Graph 3 in Figure 10) to continue their efforts to increase the strength of partnership and downstream GP network marketing to increase patient referrals (Graph 1 in Figure 10) and hospital admissions. Therefore, opposing balancing Loops B5 and B2 creates another eroding goals archetype.

Table 1 summarizes the key loop interactions and details of hospital managements' strategic actions identified to leverage the archetypes discovered in the dynamic hypothesis.

Table 1. Summary of loop interactions and archetype leverages

Loop(s)	**Type of Loop(s) /Archetype**	**Effects / Leverage**
R11	Reinforcing (B2 & B6) Escalation Archetype	Loops B6 and B2 together form a reinforcing loop in Figure 3 where the action of changes in referrals reinforces referrals and changes in overheads and hence cashflow and clinical pathways (CP). The rate of CP influences the changes in clinician variation fraction, which reinforces it effect on patient satisfaction. So the net effect is that the escalation archetype either continuously reduces or increases patient referrals depending on hospital management implementation of CP. So, the leveraging point is to control and monitor the implementation of CP.
R11 & B1	Limits to Growth Archetype	The limiting action, which is the leveraging action, comes from the balancing effect of reduced referrals from the change in payor gap, which limits the referrals as a result of reduced patient satisfaction.
B1 & B2	Drifting Goals Archetype	Opposing effects of B1 and B2 leads to the drift of the goal to increase patient referrals. The leveraging action is to set benchmarks and anchor the goal to staff incentives and identify the procedures and centralize the policies for staff hiring/layoffs.
B2 & B4	Drifting Goals Archetype	The opposing management actions in B4 and B2 resulting from the gap in unused bed capacity result in a drift of performance. The leverage here is to set some external standards to measure the cost implications of unused bed capacity and even though clinician-hiring policies are centralized, staff and contractor recruitment needs to be controlled globally in the hospital.
B2 & B5	Eroding Goals Archetype	The goal drifts as B2 and B5 oppose each other's actions to reduce unused bed capacity. The leveraging point is to control and monitor the effects of clinician partnerships/promotions and forecast patient growth more rigorously with commitment of numbers of patient admissions by clinicians.
B3 & B5	Drifting Goals Archetype	Performance shifts from the target with B3 attempts to increase staff and bed interchangeability opposed by B5 with reduced spending on promoting referrals thorough alliances with clinicians. Leveraging action here would be to centralize staff training activities and seek global measures as targets for admission delays. This prevents suboptimization at a departmental level.
B3 & B6	Eroding Goals Archetype	B6 attempting to implement CP and reducing clinician workload and hence decrease hospital caseload while B3 results in the ease of pressure to reduce staff and increasing bed availability and reducing delays. The leveraging action is again to control the level of implementation of CP and centralize staff training activities.
B2 & B7	Drifting Goals Archetype	The effect of balancing Loops B3 and B2 is to reduce the admissions by lowering the gap of available beds. The leveraging action for management is to set global targets for unused bed capacity.
B3 & B7	Drifting Goals Archetype	The interaction of B7 with B3 and likewise of B7 with B3 results in the drifting goals archetype.

continued on following page

Table 1. continued

Loop(s)	Type of Loop(s) /Archetype	Effects / Leverage
B8 & B9	Eroding Goals Archetype	These two balancing Loops B9 and B8 act in opposing directions to create a eroding goals archetype with fluctuating impact on CP implementation. The management action to reduce or recruit clinicians and thereby affecting hospital caseload variations should be balanced with management's desire to extend the implementation of CP. The leveraging action is to sensitively balance these two loops by interacting with clinicians to commit these to CP practices and targets.
B10 & B2	Drifting Goals Archetype	B10 interacts with B2 to form a drifting goals archetype as Loop B2 attempts to lower the goal of change in patient referrals and effectively increasing the pressure to reduce staff and the B10 loop attempts to decrease the pressure to reduce staff with increased hire of staff/contractors in the hospital system. The leveraging action for this archetype is to centralize the hire and layoff at a corporate level rather than at a departmental level.
B10 & B3	Drifting Goals Archetype	It therefore interacts with B3 to form a drifting goal archetype as Loop B3 attempts to increase multiskilling of staff and reduce the gap in beds availability while the B10 loop attempts to decrease the pressure to reduce staff with increased hire of staff/contractors in the hospital system.
B10 & B5	Drifting Goals Archetype	The drift in performance between B10 and B5 can be regulated if the management action to increase referrals through alliances with clinicians is balanced with departmental action to hire contractors. This is considered the best leveraging action to offset the effects of the drifting goals archetype.
R10 & B5	Limits to Growth	The limiting action, which is the leveraging action, comes from the balancing effect of increased referrals, which limits action of R10 to layoff staff/contractors and the availability of staff as a result of pressure to reduce staff.
R10 & B2	Limits to Growth	The limiting action, which is the leveraging action, comes from the balancing effect of changes in quality of care, which limits the referrals as a result of reduced patient satisfaction.
B2 &B4 & R10	Limits to Growth and Under Investment	Loops B4 and B2 interact with the contractor reinforcing Loop R10 forming a growth and under-investment archetype an extension of the limits to growth archetype. The leveraging action for management is to maintain the long-term requirement to continue to keep its capabilities and core competencies at a level that ensures its competitive advantage.

ARCHETYPE MODEL REFERENCE BEHAVIOR

The descriptors of hospital performance are defined as the following quantitative and qualitative measures and are used as the simulation model metrics for output assessments:

- Hospital bed occupancy.
- Gap in available beds comparing national average length of stay (NLOS) and length of stay (LOS) for the patient admissions.
- Deviations between the NLOS and the hospital's LOS by diagnostic related group (DRG) for patient admissions.
- Average marginal costs per patient.
- Cashflow in hospital.
- Net cash balance.
- Clinician satisfaction with hospital management.
- Patient satisfaction with hospital and clinician services.
- Quality of care.

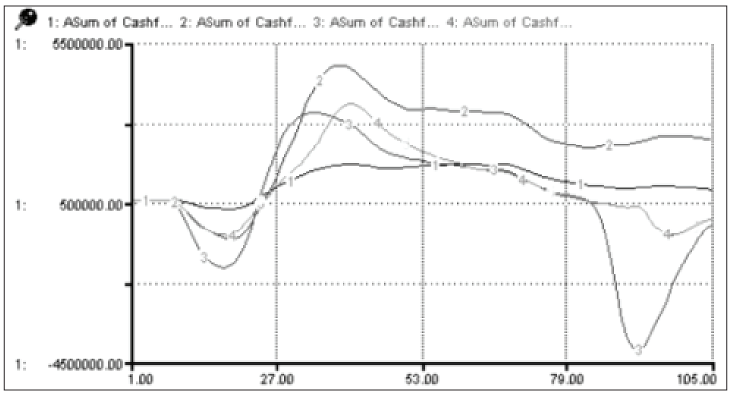
SCENARIO ANALYSIS

Hospital executives decided that there was a need to examine the crossover and interactions of strategic policy implementations with respect to market growth, outsourcing, and clinical pathway policies. The following scenarios are explored to examine the combinatorial effect of the policies for future strategic planning.

The best case "Nightingale" scenario (Graph 2) demonstrates that there is clearly a bold need for hospital management to increase staff strength (Figure 12) commensurate with market growth, which further improves cashflow, which then declines slightly due to threat of the limits to growth archetype in Figure 11. The strategy is best envisioned with the understanding that such limits do exist and to anticipate them as in the Nightingale scenario in order to sustain cashflow by improving clinician satisfaction and modifying the rates of clinical pathways implementation in specific DRGs.

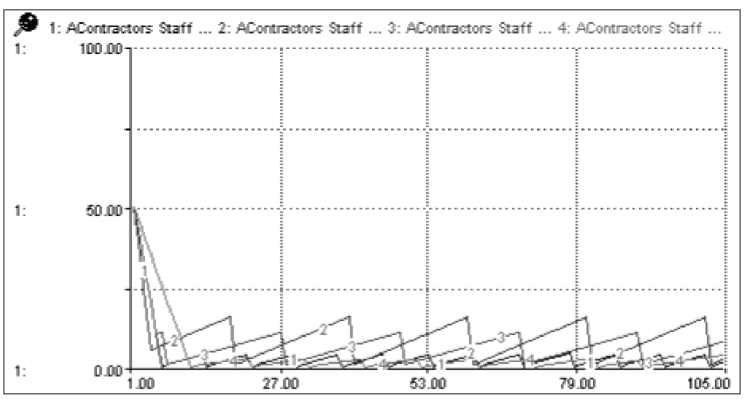
There is a need to be vigilant about bed capacity with either improving staff multiskilling and bed interchangeability or with capacity and facilities expansion; improvements can sustain long-term market growth. Even with these caveats, the Nightingale scenario improves the hospital's cashflow position to a weekly average of \$2.1M and a cash balance of \$7.2M compared to \$1.1M and \$5.2M of the base case (see Table 2).

Figure 11. Cashflow for archetype scenarios 1-4



(Graph 1 - Base Case; Graph 2 - Nightingale; Graph 3 - MOR; Graph 4 - Doldrums)

Figure 12. Contractor levels for scenarios 1-4



(Graph 1 - Base Case; Graph 2 - Nightingale; Graph 3 - MOR; Graph 4 - Doldrums)

Table 2. Comparison of performance indicators for archetype scenarios 1-4

Performance Indicator	Archetype Base Case – (Graph 1)	Archetype Best Case “Nightingale” – (Graph 2)	Archetype Middle of the Road - (Graph 3)	Archetype Worst Case ‘Doldrums’ - (Graph 4)
Bed Occupancy (max 1)	0.33	0.51	0.49	0.52
NLOS Available Bed Days Gap	-103.13	-69.53	-74.59	-68.41
Deviation from NLOS (days)	-1.98	-0.65	-0.06	-0.60
Average Marginal Costs per Patient	$1448.23	$961.96	$1032.58	$1053.49
Sum of Cashflow	$1101541.20	$2108048.71	$1978321.14	$1068916.81
Average of Cash Balances	$5196897.74	$7273164.56	$6939464.77	$7166100.30
Clinician Satisfaction	0.75	0.75	0.75	0.76
Patient Satisfaction	0.44	0.29	0.29	0.48
Quality of Care	0.65	0.45	0.45	0.77

IMPLICATIONS AND CONCLUSION

The application of system archetypes to the strategic analysis of the hospital case reveals that it is possible to identify loop holes in management’s strategic thinking processes and it is possible to defy these fallacies during policy implementation as illustrated by the results of the archetype simulation model.

Hospital executives discovered that policy modification with slight variable changes helps to avoid such pitfalls in systems thinking and avoid potentially cost prohibitive learning had these policies been implemented in real life. The application of the system archetype tools improved the manager’s understanding of the key underlying causes of hospital performance issues and thus enabled the development of a more comprehensive software model.

In the casemix model of the hospital, the team discovered that some of the phenomena as described by these archetypes could be represented in the software. Although in some cases it required some amount of modification of the models to demonstrate similar behavior. With the application of these incremental approaches to model decision support processes and to architect the software to simulate strategic behavior, it was possible to “future proof” investment decisions, hedge risks, and avoid management decisions failures. The experiences gained in the modeling and development process were used to develop and model the decision support systems.

FUTURE DIRECTIONS

The lessons learned by our investigations in this study include that people can be taught to see the flaws in their mental models through “actionable knowledge” or “system archetypes.” Mental models are deeply ingrained assumptions, generalizations, or even pictures or images that influence how one understands the world and takes action. The discipline of working with mental models begins by turning the mirror inward, learning to unearth one’s internal pictures of the world, bring them to the surface, and hold them up to rigorous scrutiny. The penchant for the “quick fix” or the “low hanging fruit” combined with the unwillingness in long-term solutions requires researchers to seriously find new ways to evangelize the key concepts. Some suggested future directions for this messaging would be to link archetypes to build more complete models. Another aspect of future research direction would be the creation of simulations with new datasets for other healthcare organizations. The development of an executive toolkit to highly leverage interventions and provide early warning systems of failure patterns and thus providing an improvement in strategic thinking is also a key area for future studies.

REFERENCES

Bardoel, A.E., & Haslett, T. (2004). Success to the successful: The use of systems thinking tools in teaching OB. *Organization Management Journal Teaching & Learning, 1*(2), 112-124.

Bellinger, G. (2003). *Archetypes.* Retrieved August 4, 2007, from http://www.systems-thinking.org/arch/arch.htm#archss

Braun,W. (2001a). The system archetypes. *The systems modeling workbook.*

Coyle, R. G. (1996). *System dynamics modeling: A practical approach*. London: Chapman & Hall.

Forrester, J.W. (1968). *Principles of systems* (pp 1-7). Cambridge, MA: Productivity Press.

Goodman, M. (1974). *Study notes in system dynamics.* Cambridge, MA: MIT Press.

Hayakawa, S. I. (1990). *Language in thought and action* (5th ed.). New York: Harcourt Brace Jovanovich.

Kim, D., & LannonKim, C. (1994, May). A pocket guide to using the archetypes. *The Systems Thinker.*

Kornbluh, M., & Little, D. (1976). The nature of computer simulation model. *Technological Forecasting and Social Change*, *9*, 3-26.

Lyneis, J.M. (2003, January). System dynamics for business strategy: A phased approach. *System Dynamics Review, 15*(1), 37-70.

Senge, P. (1990). *The fifth discipline.* New York: Doubleday Currency.

Stevenson, R. (1993). Strategic business process engineering: A system thinking approach using IThink. In K. Spurr, P.Layzell, L. Jennison, & N. Richards (Eds.), *Software assistance for business re-engineering* (pp. 99-118). John Wiley & Sons Ltd.

Wolstenholme, E.F. (1994). *System enquiry: A system dynamics approach*. Chichester: John Wiley & Sons.

Wolstenholme, E.F. (2003, January). Towards the definition and use of a core set of archetypal structures in system dynamics. *System Dynamics Review, 19*(1), 7-26. John Wiley & Sons.

ADDITIONAL READING

Some of these scholarly works have been quoted as references in the chapter, but as a guiding principle the whole book indicated below should be read by those intending to take a deeper view of the system archetype constructs.

Bardoel, A.E., & Haslett, T. (2004). Success to the successful: The use of systems thinking tools in teaching OB. *Organization Management Journal Teaching & Learning, 1*(2), 112-124.

Bellinger, G. (2003). *Archetypes.* Retrieved August 4, 2007, from http://www.systems-thinking.org/arch/arch.htm#archss

Braun, W. (2001a). The system archetypes. *The systems modeling workbook.*

Coyle, R.G. (1996). *System dynamics modeling: A practical approach*. London: Chapman & Hall.

Goodman, M., & Kleiner, A. (1993, December 1– 994, January). Using the archetype family tree as a diagnostic tool. *The Systems Thinker.*

Kim, D. (1995, June/July). Notes on dynamic theories of the archetypes are taken from, 'archetypes as dynamic theories.' *The Systems Thinker.*

Kim, D., & Anderson, V. (1998). *System archetype basics: From story to structure.*

Kim, D., & LannonKim, C. (1994, May). A pocket guide to using the archetypes.

Meadows, D., Meadows, D., Randers, J., & Behrens, W. (1972). *Limits to growth*. New York: New American Library.

Senge, P. (1990). *The fifth discipline* (pp. 142, 15055, 156, 226). New York: Doubleday Currency.

Senge, P., et al. (1994). *The fifth discipline fieldbook* (pp. 149-150). New York: Doubleday Currency.

Shapiro, A., & Lorenz, C. (2000, February). Large scale projects as complex systems: Managing 'scope creep.' *The Systems Thinker, 11*(1). Pegasus Communications.

Stevenson, R. (1993). Strategic business process engineering: A system thinking approach using IThink. In K. Spurr, P. Layzell, L. Jennison, & N. Richards (Eds.), *Software assistance for business re-engineering* (pp. 99-118). John Wiley & Sons Ltd.

Wolstenholme, E.F. (1994). *System enquiry: A system dynamics approach*. Chichester: John Wiley & Sons.

Wolstenholme, E.F. (2003, January). Towards the definition and use of a core set of archetypal structures in system dynamics. *System Dynamics Review, 19*(1), 7-26.

Chapter XIV
Integrated Manufacturing Applications and Management Decision Making:
Putting the P Back into ERP

Fergal Carton
University College Cork, Ireland

Frédéric Adam
University College Cork, Ireland

ABSTRACT

The provision of timely, accurate, relevant, and concise information for managerial decision making has traditionally represented a challenge to information systems designers. The mass adoption of enterprise resource planning (ERP) systems has multiplied the amount of data being recorded about the movement of inventory in the supply chain. However, this online information requires much off-line manipulation in order for it to be meaningful to managers. In addition, these data are based on physical structures and business models that evolve over time, and thus inevitably a gap opens between the virtual enterprise and reality. Despite the benefits of inventory visibility and expenditure control afforded by ERP systems, managers still require data from other, nonintegrated systems. In this chapter the authors present their research on decision-making support in two manufacturing organisations, with the objective of understanding how these integrated applications support the manager in achieving his or her goals.

INTRODUCTION

Enterprise resource planning (ERP) systems have created a virtual infrastructure for the management of physical resources in manufacturing environments. In addition to their primary function of centralising business **transaction processing**, ERP systems provide a virtual audit trail of the physical activities and movements that make up everyday production activity (e.g., consumption and replenishment of inventory, work order control, warehouse management, sales order allocation to finished goods, packing, and shipping). These highly integrated software packages have become a de facto standard for transaction processing in the modern manufacturing company, and indeed a basis for its interfaces with supply chain partners and customers.

In streamlining the transaction processing for production activity, ERP systems aim to reduce the potential for error in information handling as inventory is bought, transformed, and delivered to customers. By incorporating efficient data capture technology with powerful networking capacity, the vital transactional information is captured at source by users as an integral part of their work processes, and then made available instantaneously for management. This theoretical ideal makes the concept of ERP an enormously attractive one for large firms faced with on-going struggles to reduce costs and report detailed results to shareholders.

ERP systems are designed and implemented with efficient transaction processing, not management decision making, in mind. The time saved by the organisation in automating administrative processes does not necessarily imply faster decision processes. This is primarily because management decisions are generally based on the ability to compare actual data (often provided by the ERP system) with planned data (future performance scenarios in the form of targets). Also, quite simply, because the systems architecture and user interface that these diverging goals require are totally different. When selecting and implementing an ERP package, companies are looking at a good fit with operations, not with the more abstract decision processes. ERP vendor hype tends to build the perception among potential customers that if the transaction processing system works, managers will benefit from the increased visibility of operational data, independently of any study of how they might exploit this abundance of information.

Management is also highly concerned with the efficient use of human resources, but the data model of ERP applications is inventory centric, rather than user centric. People are simply users from the application perspective, not resources with associated costs to be managed. For example, ERP systems are not designed to analyze the time taken by users to process a complete transaction (although transaction time and date stamping is an inherent part ERP). Neither are they designed to record the difficulty the user experiences attempting to reconcile a transaction screen with information related to a physical transaction that does not quite match the software template for data entry.

As with any system design, the implementation of a standard ERP package in a particular company, with its specific requirements, involves a certain number of assumptions about that company and how it operates. These assumptions may change over time due to the organisation changing what it does. Thus, even at go-live, there may be a gap between the physical processes of purchasing, manufacturing, and order fulfilment and their corresponding virtual processes in the ERP system.

This gap may manifest itself in an ever increasing requirement from managers for reports and queries to interpret the operational data. **Business intelligence (BI)** tools allow managers to interpret the underlying ERP source data in a more flexible manner, even to the extent of correcting idiosyncrasies in the source information with respect to the desired interpretation. The attribution of sales transactions to different regions or profit centers based on how the company wants to report its results is a case in point.

Relying on in-depth field work in two manufacturing organisations, this research aims to evaluate the significance of the gap between the reality of conducting business in a dynamic marketplace and the "hardwiring" of transactional support systems. In rendering the gap visible, it is hoped that a more meaningful discussion can take place regarding the resources required to maintain a close fit between information systems and an evolving business model in the specific context of ERP systems.

In this chapter the authors present the background to the use of ERP systems to support management decision making, before presenting the research approach and the main findings from the field work, and end with the conclusions.

BACKGROUND

Enterprise resource planning (ERP) systems are commercial software packages that enable the **integration** of transaction-oriented data and business processes throughout an organisation (Markus, Axline, Petrie, & Tanis, 2003). Implemented as a suite of integrated software modules, ERP systems are used to administer the physical movement of inventory through the supply chain and the allocation of sales orders to finished goods in the **demand** fulfilment cycle.

ERP software is a semifinished product with tables and parameters that user organisations and their implementation partners configure to their business needs (Shang & Seddon, 2000). It is the complete set of **configuration** options (often called the template) selected by the customer implementing the software that define how a system will work.

ERP systems use relational database technology and workflow principles to link dynamic information (transactional data) to static information (master data), thereby providing a system of record for the basic business transactions that make up daily activity across a range of functions in the organisation.

A cornerstone of the ERP concept is the notion of "integration," which means that data which are required by different parts of the organisation are captured once at source, and then served to the different business functions in the relevant format. This integrated data model minimises data redundancy and facilitates data maintenance.

From our review of the literature related to the support managers may obtain from their ERP applications, we developed themes that fall into three categories:

- Use of management support systems and ERP
- Organisation design configuration and ERP
- Decision making and ERP

Management Support Systems and ERP

As Ackoff (1967) suggests, it is important in looking at the impact of information systems on decision making to differentiate between the decision process itself and the information required to support that process. It was Ackoff's contention, well before the age of global ERP systems, that most managers suffer not from a lack of relevant information, but rather from an over-abundance of irrelevant information. Gorry (1971) decries the tendency to assume that improved decisions will result from increasing the information provided. This warning was echoed by Benjamin and Blunt (1992), suggesting that "managers and workers are in danger of dying from a surfeit of communication." The power of information technology to *informate*, such that "events, objects and processes become visible, knowable and shareable in a new way," is dependent upon a recipient's ability to interpret the information provided (Zuboff, 1988).

Gorry and Scott Morton (1971) describe the characteristics of the information required by operational, management and strategic levels in the firm as significantly different. Operational control activities require information that is detailed, real-time, and based on the actual use of internal resources. Managerial control, on the other hand, requires more summary information, not necessarily real-time, and includes external sources of information.

Although it is tempting to believe that improved management control should stem from mastery of the detail contained in operational systems (and certainly the language used by ERP vendors would encourage this perception), Gorry and Scott Morton (1971) would argue that these are two distinct levels of activity, with different information characteristics and, therefore, requirements.

As early as 1958, Leavitt and Whisler observe that there was increasing pressure on businesses to radically reorganise to deal with a "complicating, speeding world." Furthermore, the more uncertainty surrounding a given decision, the more the manager will ask for information to help illuminate a decision, "playing it safe" in the words of Ackoff (1967). The system designer, by turn playing it safe, provides even more information. "One cannot specify what information is required for decision making until an explanatory model of the decision process and the system involved has been constructed and tested" (p. 150).

Moreover, the "nimbleness" of large integrated systems has been questioned in the light of the rapidly changing context of managerial decision processes. Dearden (1972) feels that the notion of one "completely integrated supersystem" was "absurd." Eisenhardt (2002) adds, "Complicated intertwined activity systems or elaborately planned leveraging of core competencies make sense in slower and more linear situations" (p. 89).

Davenport (1998) explores the paradoxical impact of ERP on a company's organisation and culture. On the one hand, by providing universal, real-time access to operating and financial data, ERP systems allow companies to streamline their management structures, creating flatter, more flexible, and more democratic organisations. On the other hand they also involve the centralisation of control over information and the standardisation of processes, which are qualities more consistent with hierarchical, command, and control organisations with uniform cultures.

There is an absence of published research on the relationship between enterprise computing and **decision support** (Holsapple & Sena, 2001). It is of significant interest to senior management of organisations, IS practitioners, and IS academic researchers to know more about the post implementation period of ERP systems, the business benefits that occur during the period, and how and why these consequences occurred (Staehr, Shanks, & Seddon, 2004).

In the next section we look at organisation design configuration and how ERP systems interact with these parameters.

Organisation Design Configuration and ERP

The problem of organisation design is to create mechanisms that permit coordinated action across large numbers of interdependent roles (Galbraith, 1983). So the organisation adopts integrating mechanisms which increase its information processing capabilities.

An ERP system may be considered as an "integrating mechanism" in that data that are required by different departments are structured in a common database. Furthermore, processes that require this data are constructed such that different users can intervene at different stages of the same process; user profiles define responsibilities and unique transaction numbers preserve the integrity of process components.

All information systems have an impact on the shape of the organisation, automating, as they do, the communication patterns and tasks which have traditionally defined roles and responsibilities. The potential impact of ERP systems, because of their cross functional nature, is far reaching, particularly where changes to the physical organisation are resisted, and, the resulting compromise between the legacy organisation and the new administration processes may generate more friction than benefit.

Davenport (1998) states that an enterprise system is a generic solution. Its design reflects a series of assumptions about the way companies operate in general. Thus, it may well be that some of the firm's traditional management decision processes, with their own periodicity, **granularity**, and level of **aggregation**, are not supported in the standard transactional template or **configuration**. Reporting (the provision of aggregated and analytical information to assist in decision making) is typically given little, if any, focus during the long implementation phase of most ERP systems.

On the other hand, Koch (2001) proposes that configuring an ERP system means setting the basic features of the manufacturing control tool, and in so doing, implies "designing the tools for direct operational decision support". In other words, we do not design ERP systems around management decision processes, but their implementation will set the scene for decision making, at least at an operational level.

Koch (2001) makes the point that the complexity of implementing ERP solutions has left many companies content to merely substitute legacy systems, rather than explore the potential of ERP as a decision support tool. Koch (2001) acknowledges that decision making in operational management areas such as financial control and production planning are subject to constant change, and that the management of these changes with respect to the ERP configuration is itself a considerable task.

Decision Making and ERP

Mason (1969) points out that much information system design is based on the activity of the organisation, rather than the decision-making processes, and it could be said that more than 30 years later, ERP systems enshrine *the administrative process for transactional activity* as the most important driver in the running of the business, and therefore still failing to address the decision-making processes as a design parameter.

Research into how managers obtain and use information for decision making has proved inconclusive, with a large part of the manager's capacity to make rapid decisions on a broad range of issues seemingly based on intuition, or the ability to quickly aggregate and internalise the meaning behind the data.

Much research in decision making during the last century was focused on the difficulty of defining a rational model for an ever-changing process that also allows for the irrational or contextual factors that influence the decisions made by management in organisations.

Research would suggest that there might be a myriad number of decision processes undertaken by managers, almost on a jobshop basis (Mintzberg, 1989). Key to the design of this research is the tactic of asking participants to prioritise decisions themselves based on their own organisational goals.

Thus, we used a longitudinal pilot study of an initial site (hereafter referred to as Firm A), to develop an approach to evaluating the impact of **ERP** systems by collecting data on management perception of what the company's core competence was prior to the ERP implementation. This revealed that the impact must be understood in terms of a number of dimensions. These dimensions were identified by the survey participants *themselves* as important strategic attributes of the company in our pilot study of Firm A.

In short, managers were asked to identify the organisational goals to which they worked. The assumption was then made that decisions pertaining to these goals would be de facto *critical* management decisions. Only then were managers asked to identify the information requirements they have in achieving those goals.

Langley et al. (1995) identify three aspects of decision making which render it a difficult subject for empirical research:

- Many decisions do not imply distinct identifiable choices, and are difficult to pin down, in time or in place.
- Decision-making processes do not necessarily proceed as a linear sequence of steps, rather they are driven by the emotion, imagination, and memories of the decision makers, punctuated by sudden crystallisations of thought.
- It is difficult to isolate decision processes, as decisions typically become intertwined with other decisions.

Pfeffer (1992) discusses the selective use of information in management to rationalise decision processes, and how, under conditions of uncertainty, individuals would prefer to use data and decision-making processes "with which they are comfortable."

Interestingly, the implementation of an ERP system will only serve to exacerbate the issue of information "usefulness." Firstly, each ERP package uses different process models as an underlying framework and these models can differ in terms of how they operate. Both Oracle™ and SAP™ are based on the principle of "work orders," for example, which correspond to unique production jobs. However the manner in which they consume inventory and tie back to sales orders is different from one package to the other.

Secondly, managers may not initially understand the reasoning behind some of the configuration options embodied in the business template as implemented by the ERP project team. Once implemented, users will usually be dissuaded from any course of action which implies changes to these decisions.

Thirdly, there is a wealth of information important for decision making, which lies outside the traditional ERP boundaries (Stefanou, 2001). Information generated by computer based systems does not include much of the information that is most important to management, especially important qualitative information (Dearden, 1972). For example, information from external sources, such as published statistics, market data, and experts' opinions, are not easily accommodated within the ERP environment. Legacy systems may contain years of historic data that can be crucial in determining trends and patterns.

If anything, the **latency** inherent in gathering and collating of information might be seen as a hindrance to managers who need to be able to react on the spot to crises, both real and perceived. It would appear that ERP systems contribute to drowning the manager in ever increasing volumes of low level data, thus the contribution to decision making is unproven.

RESEARCH APPROACH

As already indicated, our study followed a two stage strategy, including a pilot case in Firm A and a main case study in Firm B. The two companies studied for this research are presented in Table 1.

For the purposes of this research, we selected a sample group of managers with roles that fall broadly under the umbrella of manufacturing:

- **Level:** Corporate and site (head office or manufacturing site).
- **Function:** Manufacturing, materials, and distribution.

The research involved 39 interviews of one hour, which were taped and subsequently transcribed. The breakdown of interviews over the different manufacturing functions was as illustrated in Table 2.

The next section reports on the empirical data gathered in both pilot (Firm A) and in-depth (Firm

Table 1. Presentation of the case study companies

	Firm A (pilot case)	**Firm B (in-depth case)**
Industry	**Pharmaceutical**	**Data management**
Turnover 05 ($bn)	38.72	9.66
Employees	100,000	26,500
WW operations	119	52
Manufacturing sites	80	3
Head Office	London, UK	Boston, USA
ERP System	SAP R/3	Oracle 11.03
Architecture	Single instance	Single instance
Server location	Pennsylvania, USA	Boston, USA
Go-live	Phased 2004	Big-bang 2001

Table 2. Number of interviews across different manufacturing functions

	# Interviews	
Function	**Firm A**	**Firm B**
Materials/Planning	8	8
Manufacturing	6	6
Distribution	2	9
Total	16	23

B) case studies and draws some conclusions from the findings.

ISSUES, CONTROVERSIES, AND PROBLEMS

The cases studied allowed us to develop several key themes concerning the relationship between ERP and decision making. These are presented in synthetic form in this section under three headings:

- Data
- Organisation and applications
- Key management decisions and role of ERP

Data

Information that is usable by managers in support of decision making infers a certain level of aggregation, typically this involves losing some of the detail captured at transaction level. In both cases studied this involves jettisoning the data structures native to the ERP application in favour of separate metadata definitions housed in a **data warehouse** application.

Granularity of Data and Metadata Logic

The transactional history of the ERP system allows managers highly granular data on each and every sales order. When aggregated, this data can be "sliced and diced" according to different criteria, depending on what message is being researched.

Although ERP vendors are now adding BI functionality to their offerings, both cases studied have taken the approach of employing third party query and broadcast tools. In general, third party tools are perceived to offer a wider range of interfaces to non-ERP data sources, such as production planning and budgeting applications.

In Firm B, on the other hand, although transacting backlog orders (sales orders that have been approved and are ready for shipment) is part of the product allocation and shipment process, distribution managers base decisions on backlog in terms of its total value, as this constitutes the potential additional revenue that can be made for that period. As these business targets are themselves defined in terms of sales parameters (revenue stream, region, country, sales channel, district, rep, etc.), managers use dynamic pivot tables to 'slice and dice' backlog orders such that it can be viewed in the light of these different business goals. Access to this information must be real-time and user friendly, as managers will require frequent refreshes to the backlog view as end of quarter approaches.

Thus the data structures native to an ERP system support manufacturing and distribution inventory transactions, but the decisions these managers are required to make are based on logic that stems from a sales operations view of the world.

Data Latency and the Technical Infrastructure

In Firm B, pressure builds towards the end of the quarter, as late sales orders are entered into the system in order to be shipped before quarter end. Because the system is integrated, and single instance, this pressure results in a slow down in general system processing, such that not only do approved sales orders have trouble getting pushed through the system such that they appear in backlog and can be shipped, in addition, orders that have physically shipped are still showing up on backlog. The latency of information in the system is, in this sense, hampering the physical process.

This becomes a critical issue as quarter end approaches, and target revenue goals are at stake:

People start making decisions off of what they see on the screen.

Latency in the ERP based order processing and fulfilment modules has given rise to drastic measures to economise on the processing load in the critical end of quarter period:

I mean they shut off Self-Service Procurement on the 31st March, to try and get the orders booking and all that.

But as the system falls behind reality, managers are left looking at data that are out of synch, but what they do not know is how badly out of synch the data are, and at what stage they are likely to catch up with the physical movements of products.

The tough thing is, now they tell you, it's 3 hours behind, well what do you do?

In this instance, as with nearly all management reports, data are being viewed through a mirrored data warehouse, not through the ERP system itself. The ERP system handles the integrated transactions, and the data warehouse handles the logic with which the transactional data are interpreted before presentation to managers. The additional layer of infrastructure creates latency that hampers rapid response from management.

ORGANISATION AND APPLICATIONS

Goal Focus and Key Performance Indicators (KPIs)

The primary goal in terms of materials management is to ensure availability of sufficient raw material to satisfy the planned level of finished goods (availability). Doing this while keeping supply costs under control and keeping inventory to a minimum is the challenge for managers in both cases studied.

A common feature of both cases was the use of centrally defined functional goals to drive performance in a dispersed global organisation. Managers are evaluated on their ability to achieve organisation goals. These goals cascade down through the functional hierarchies from corporate headquarters. Organic differences in priorities mean that some functions are not aligned and in some cases this can result in "goal conflict." There is evidence that too narrow a focus on functional goals may encourage inappropriate behaviour from the point of view of the organisation as a whole.

In Firm A, it was questioned whether the "did I meet my plan?" KPI alone would drive the right behaviour, as a given node in the supply network could "meet the plan and yet sub-optimise the supply chain." In this case the introduction of a second metric "did I keep the next node in stock?" was much more conducive to "creating the correct behaviour and the correct discussions between the 2 nodes."

In Firm B, the manufacturing mindset is focused on achieving the build plan in terms of finished goods, and moving the product out the door. Although distribution is responsible for matching the finished goods to an approved sales order, and its subsequent release, picking, and shipment, manufacturing is driven by the internal build plan (planned number of units), not the sales orders themselves. Distribution, on the other hand, is driven by a revenue goal that is achievable only through satisfying actual customer demand, and aim to finish the quarter with no backlog. There is an inherent conflict here between the two functions, which is accepted as a necessary evil by managers on both sides. Although this has little to do with the choices made in implementing an ERP system, managers could be expected to want their information systems to assist in rendering this organic conflict visible by providing the right information at the right time.

Another finding was that KPIs with standard tolerance levels designed for comparison across sister manufacturing plants (for example, internal delivery performance, or the ability to deliver to customers on time) ignore local conditions and arrangements. If the customer has a replenishment agreement with the supplier, and inventory is kept within certain 'bands,' it matters much less when the delivery takes place, and "the supplier has a degree of freedom to move deliveries around if they require to." Therefore an internal delivery performance metric loses its relevance in this context.

In Firm B, customer satisfaction measures include the ability to deliver product in one single shipment (complete shipment metric) on or before commit date, regardless of where the components of that order originate. These measures are also quoted as the **key performance indicator** for the manufacturing operation. Currently there is no ERP capability to compile the "complete shipment" statistic, as a single order may give rise to several shipment transactions that do not necessarily happen at the same time. Currently this KPI is compiled by manually collating all shipment transactions relating to the same sales order. This is hardly scalable, and will become more onerous as the business model moves towards multiple different points of material supply for one order.

Information Latency and the Complexity of the Business Model

Both firms struggle with supply decisions because of the variability of demand. In Firm A, the true indicator of business performance would be "whether we are on the shelf" in either the supermarket or pharmacy. Even if this information were available, it would have little relevance back at the point of primary supply, as the manufacturing of active ingredient must, by definition, follow a longer range forecast. Faced with this dilemma of cascading demand information back through the supply chain, managers use performance "proxies" such as external shipment performance to drive manufacturing. Customers (of the manufacturing plant) are replenished systematically within defined service levels, and plants will be measured on their ability to maintain these levels.

In Firm A, the first level of issue management in the supply chain is at the execution level, where something breaks down (e.g., a batch failure, a technical problem with production, etc.) or where an event is picked up at the other end of the supply chain (e.g., a competitor goes off the market). The challenge at this level is to pick the event up quickly, and then quickly assess what the response should be, in order that sales performance and stock availability performance do not suffer. A typical reaction to this type of event would be to ration product across commercial entities. So the demand requirements planning (DRP) system (in this case Manugistix) provides an accurate view of inventory levels across the nodes of the supply network. A defined escalation process exists, such that once the event has been picked up, managers know how to escalate the issue to an above site organisation who will in turn estimate the impact across the supply chain. However, this relies on human beings identifying that something has just happened which could threaten the ability to supply in the short and medium term. However, managers have yet to be able to come up with "any semi-automated signals to allow the triggering of a response." Although Manugistix can identify that there is a demand and supply mismatch, it is entirely reliant on the quality of the data that are in the local ERP and DRP systems. If local planners have not updated their systems the process breaks down.

Furthemore, latency builds up in the supply chain simply because of the number of nodes and time taken to process information back through the nodes to allow a decision to be made at an upstream manufacturing site:

There's got to be at least a 1 week and up to 3-week time-lag before, as you roll through the various nodes' planning systems before the mismatch becomes evident, so that's an issue that we haven't got satisfactory resolution to in terms of a data driven trigger for events as opposed to human intelligence just saying, we've got a problem here.

There is thus an inherent business constraint in how far IT can support decision making in the demand fulfilment processes observed.

KEY MANAGEMENT DECISIONS AND ROLE OF ERP

The manufacturing sites of Firm B have for a long time organised themselves along the principles of the following activities: *material planning, buying* (procurement), *making* (assembly and test), and *delivering* (distribution). This 'Plan, Buy, Make, Deliver' model had served the business well during its rapid growth phase during the 90s where manufacturing concentrated on one key product range.

Categorising the management decisions observed in the field interviews according to the "Plan, Buy, Make, Deliver" model, yields the results presented in Table 3.

Decisions D^0 to D^6 are decisions pertaining to the execution of a build plan and the fulfilment

Table 3. Key management decisions for manufacturing (Materials, Manufacturing and Distribution) for Firm B

Decision type	Management Decisions (Site)
D^0 What to build?	How to deal with build plan changes?
	Mfg build plan less than sales for end of life product
	Many changes to plan are not systemised
	Without a sales order, you make an intellectual decision to build it
D^1 What to buy?	How to deal material supply changes?
	Estimating usability of material in repair at vendor (RMA) is difficult
	once MRP released, we run planning detail reports, and net receipts manually
D^4 What to stop building?	Component issues and false starts
	What is the impact of a purge?
	Decisions to override a purge on a ship-to customer
D^6 What to allocate & ship?	Allocation of orders to finished goods
	Massive changes to allocation at quarter end
	Separation of raw data and their meaning
	Quarter end decisions made on the fly
	Planning and distribution meet before going to Sales
Mgt report	What is material total cost of ownership, including failure and returns
	Capacity and staffing
Process decisions	Pricing was local, now global
	Best inventory process (KanBan, lean, etc.)
	Decision on electronic data interchange (EDI) interface delayed

of customer orders. These could be considered to be the “bread and butter” decisions of a manufacturing operation. The data bear witness to a high degree of intuition being brought to bear in each of the decision types mentioned above (D^0, D^1, D^4, and D^6).

For example, for decision type D^0 (What to build) it is common practice for production planners to “second guess” sales forecast numbers because they have first hand experience of previous forecast accuracy. Conversely the material purchasing process cannot be automated to the point of true MRP, as the demand picture changes so frequently that planners would continuously be updating purchase orders.

Looking more closely at the information requirements for each of these decision types and specifically at the underlying systems that provide the data (Table 4) is revelatory of the role of ERP in the management decision making.

It can be seen that although the ERP system is the repository for key procurement information (supplier details, bills of materials (BOMs), etc.), inventory status and sales order progress, it is inevitably coupled with another tool for presentation (e.g., BI tool) or for access to work-in-progress status (manufacturing execution system (MES)).

Another reason for the multiplicity of systems to support these key decisions is that any manage-

Table 4. Information and underlying systems support for key management decisions in Firm B

Decision	Information	System
D^0 What to build?	Forecast units and $	Sales forecast
D^1 What to buy?	MRP	Planning tool & ERP & Excel
D^2 What has been built?	Finished goods inventory (FGI)	Process control & ERP
D^4 What has been approved?	Bookings	ERP & BI tool
D^5 What to stop building	Purge instructions	Process control
D^6 What has been delivered?	Billings	ERP & Bi tool
D^7 What to allocate and ship?	Backlog	ERP & BI tool & Excel

ment decision requires two fundamental pieces of information, the planned target and the actual figure. In both cases, although the ERP system is the system of record for actual data (real time picture of inventory levels, cash, and assets), a different system is used for the planned figures (for both production planning and financial budgets). This implies that BI functionality is required to extract actual data from ERP, planned data from the other systems, and to then collate and compare this information in order to evaluate performance.

D^0 What to Build?

Production planners receive corporate sales targets at the beginning of the quarter, and their objective is to set in motion both the material procurement process and the production planning process (or starts plan) that will correspond to achieving the stated availability goals. A custom application was built alongside ERP to facilitate the task of deriving and communicating these corporate targets.

One of the principle constraints in building this custom application was the ability to plan at a "model" level, one step above actual configured final product. This means that senior managers can set target sales by model type, without specifying in greater detail the component configuration. Conversely, the ERP system can understand planning at a BOM level, for actual configured products, but it is not useful for decisions at this higher level of aggregation.

In addition, senior executives discussing sales targets need to be able to create and manipulate scenarios before committing to them, especially in the context of external shareholder driven targets for revenue, margin, and financial results. An ERP system will execute a stated build plan, but is not well suited to the iterative modelling capability implied here.

Firm B does not use ERP for production planning, this is maintained on spreadsheets, and buyers and planners can refer to these spreadsheets for an update on supply chain demands. This is a manual process, and relies critically on a good relationship between the materials managers (planners and buyers) and shop floor production managers. This communication gives materials its most critical indicator of supply chain performance, that there are sufficient materials available for current production plans.

D^1 What to Buy?

MRP does not work for Firm B for several reasons, the principle one being the pervasive "load and chase" mentality:

We say load it and we'll make it happen. So we might pay premiums, we might go visit vendors, *we might beat 'em up or flex our muscle or pay for more overtime shifts, or whatever.*

MRP works well in an environment which is predictable, and where lead times drive purchasing and production schedules. In Firm B, supplier lead time is considered another variable to be played with in order to meet the material availability plan:

And, trust me, the procurement group that's set up, our core competency is figuring out how to get the stuff with no lead times.

MRP also prefers a stable demand picture in order to be of benefit. In this company, demand forecasting is anything but predictable:

Maybe a dartboard would be better.

D2 What Has Been Built?

The ERP system has no visibility of the production process once assembly begins. A separate process control system (APC) keeps track of the physical movement of inventory through the various stages of assembly and test. A critical production status meeting takes place with increasing frequency as the quarter end approaches. The contribution of ERP to the production status report is merely informational, providing customer details and pick release status. There is no transactional element in ERP deriving from the production status meeting.

D4 What to Stop Building?

Purge decisions are necessary when a quality issue has arisen with a component or subassembly. Purge decisions take into account many factors, especially the customer commit date, the cost of the recall, the availability of alternative parts, the test plan, and so forth. Almost all these decisions are unique:

A delayed installation can be a bigger impact on customer than a 'just in case' change.

The quality of the information is, on the other hand, critical in providing senior colleagues with the basis they need to make the decision:

In making big decisions, we need to provide the VP with right information.

D5 What has Been Delivered?

Standard ERP functionality triggers an accounts receivable (AR) transaction (a billing) upon shipment of a product, this is a signal that the customer can be invoiced. In this case, the most reliable source of information is the shipping transactions in the ERP system, and the subsequent AR interface entries. The key constraint raised in the field work in terms of reporting on billings is partial shipments and how ERP handles the occurrence of a single sales order that is shipped partially because the product is not available, or because the shipment was keyed at different points in time.

D6 What to Allocate and Ship?

Field observations demonstrate the duality of the information required in managing unit shipments to a financial revenue target. Manufacturing managers understand numbers of physical units, sales, and distribution must deal in dollar values of orders, in order to be able to judge performance against a predefined quarter target. The two information streams, though derived simultaneously at a strategic level at the start of quarter from the same demand forecast, will never converge at an operational level for two reasons:

- The revenue forecast is based on an average selling price per unit, not an actual price.
- The build plan does not refer to actual customer orders, but to a specific produc-

tion goal that itself was based on a forecast demand for the different models.

Thus manufacturing is working to a build plan for *x* units per model regardless of what actual demand is doing, and sales operations is working towards overall revenue and profitability goals without prioritising the match between physical configuration and sales order specification.

Again this scenario highlights the discrepancy between the transactional rigour (necessary to expedite sales orders) imposed by the ERP system, and the more intuitive nature of the managerial decisions surrounding straightforward manufacturing decisions concerning production schedule and fulfilment.

LESSONS LEARNED FROM FIELD WORK

The empirical data show that the reality of ERP implementations, or rather the reality of changes to the business model post go-live, mean that the threshold between structured and unstructured decisions, rather than moving, as was originally suggested by Gorry and Scott Morton (1971), towards the more unstructured information, it has, in fact, moved back towards the structured data. When ERP systems (an inherently *structuring* force on management decision making) run up against the dynamism and uncertainty of the real world, instead of being able to programme *more* decisions, we actually have to 'unzip' the tight coupling between physical actions and their virtual mirror in the ERP system.

This decoupling is inevitably accomplished by extracting out of the core ERP transactional database the raw data in order to load it into the more flexible and user friendly BI tools of today. So what was originally conceived as structuring has become a handicap, and it is retained merely as a legacy system providing source data, and thereby becoming a source of rigidity and inertia.

On the evidence of the cases observed, the complexity and cost of building and maintaining managerial reporting infrastructures on the foundation of ERP systems is comparable in scale to the implementation of the ERP system itself.

The manner in which goals and performance indicators are managed within the company is central to performance. The old adage "what gets measured, gets done" seems to apply strongly in both cases. Managerial goals are highly quantified in nature, and connected directly to operational targets such as material availability, on-time customer shipment, and sales revenue.

In Firm B, goals are set at senior management level according to market expectations, and are then translated down through the organisation into unit and revenue targets for each manufacturing unit.

Both firms studied employ performance indicators that relate to the ability of the manufacturing plant to deliver to customers on time. In one case this is called internal delivery performance (the plant's customers are within the corporation), in the other it is called performance to commit. It is also clear in both cases that these performance measures are transactional in nature; they require individual transaction level detail. It is interesting that they are nonetheless compiled manually.

In Firm B we also have seen the distribution model change radically towards electronic distribution for software. Firm A reacts to very sudden crises such as pandemics which can throw a distribution model into complete disarray overnight.

ERP systems, despite being the transactional engine for key resources such as sales orders and inventory, do not assist managers directly in decisions concerning how best to meet demand with supply. The majority of managers require some additional data warehousing or reporting tools in order to aggregate the detail to a higher level of abstraction using metadata, and in order to incorporate nontransactional information such as forecasts, targets, and budgets.

CONCLUSION

This chapter has looked at the research on the impact of ERP systems on the organisation from three angles: informational, organisation, and decisional.

It is clear that the findings have bearings on the design and ongoing maintenance of large enterprise systems. Goal setting is influenced greatly by the ability to measure, but the focus on measurement and internal competition can lead to undesirable behaviour and detract from the overall goal of satisfying customer requirements.

The role of a transaction processing system is to support execution, and in so doing, to support managerial decision making related to efficient and effective use of resources. Execution models evolve and change, and it is clear that ERP systems incorporate a degree of process rigidity that eventually impedes managers in exercising decisions on process improvements.

Decision making among manufacturing managers is tightly aligned to organisational goals, highly quantified, and characterised by a 'just do it' mentality. On the other hand, the tools available to managers to support decisions related to the attainment of their goals are well behind the tools available for transaction processing.

Organisations and their staff face a constantly changing IT landscape largely shaped by the divergence between information models embedded in their systems and the reality of physical movement of resources. ERP systems are superimposed on this changing landscape as a one-time solution to the transaction processing requirements, but are going out of synch from the day they go-live. Neither are they designed to support management decision-making processes, in anything more than promising a list of static reports akin to what was available prior to go-live. In no way are they implemented with the organisational goals in mind, embedding, as they do, generic processes and relationships that discourage, rather than promote, process flexibility and adaptability.

With ERP systems, managers and users inherit assumptions about relationships between the real world and the data recorded. There are two problems with this. First they may not understand the inherited assumptions, and the system itself takes on the aspect of another organisational actor ("Oracle wants this," "SAP does it this way," etc.). Secondly, the assumptions are based on organisational goals that worked for particular organisations at a point in time, but these assumptions are not necessarily the right goals for any company in a dynamic marketplace.

At go-live, it can be hoped that the match between reality and the virtual administration system (ERP) is good, but over time the fit degrades, and the gap between how the business runs and how the software works widens. Consequently, the standard reports provided with the ERP software are less and less flexible enough to incorporate physical changes in the way the business operates.

FUTURE RESEARCH DIRECTIONS

The fear of the costs associated with excess inventory lead to material requirements planning (MRP) models for production planning. This was the driving force behind the development of material resource planning (MRP II) tools, which were the forerunners of today's ERP tools. This static view of business dynamics works less well in today's rapidly changing market driven world, where the cost of production (or over-production) is only one side of the equation. Today's managers have greater concerns on the demand side, trying to keep resources working efficiently all the while keeping track of changing customer requirements, an evolving value proposition, and changing production and delivery processes. Companies need to understand their cost base, and systems will provide the necessary rigour and discipline in maintaining visibility of costs, but people are required to focus on and understand these changes in customer behaviour. Once this is achieved, it remains the highly specialised task of understanding how these changes can be incorporated into the new business model and how this can be coded into the ERP application. At present, very few firms have staff capable of performing this task in real time. Many require outside help, which may be costly, or not available in the specific industry context.

ERP systems are essential in providing access to the raw transactional information which will relay back to managers the "current state of the road," but other demand based information has to be gathered in order to form a picture of the road ahead, and the ability to make the connection between the two sets of data remains at the heart of the unstructured work of managers that computer systems cannot replicate. For instance, many companies operate on a hybrid diet of live sales orders and sales forecast. However, it is not easy to feed this hybrid diet into an MRP module.

There are several interesting themes for further research raised in this study, and these articulate around the framework, tools, and resources that are put in place post-ERP to account for the gap between the operational information gathered at the transactional level and the planning information required at the managerial level.

As we have seen in this research, there are latency issues associated with the collection of data in highly integrated systems that limit their usefulness in a decision-making context. These are exactly analogous to the issue of integration with manufacturing execution systems (MES), whereby the ability to monitor and control processes is limited by the physical connectivity to process control equipment on the shop floor. It is difficult to capture accurately the parameters of certain physical processes, and in this sense control systems will always "lose sight of" inventory at some point in the process, only to pick it up again later, for example, at the finished goods level. The area of MES integration with ERP is itself a whole area of research coming very much to the fore as companies extend the notion of control down towards the physical interface to manufacturing equipment.

The authors believe that there is much value to be gained from the development of a framework to model the gap between the ERP system and the physical transactions that make up the administration of business processes, such that managers have a vocabulary to discuss, recognise, and measure this gap. It is often said that ERP systems provide a common grammar for the business, but this grammar needs to be extended to cover explicit, dynamic departures from template ERP processes. It also needs to incorporate the decisions that are typically made by managers, such that the underlying transactional information can be mapped at some point to the decisional layer most used by managers. For example, this framework could take into account multiple information layers as seen in Figure 1.

Arguably, an intelligent ERP application should be able to embody more than just the lower layer of this framework and could span as much as the bottom five layers.

At the present time, it has also been suggested that business intelligence tools give greater reporting flexibility to managers wanting to unlock the meaning behind their operational data. These tools rely on a layer of metalogic which encapsulates

Figure 1. Information layers

Decision layer

Reporting layer (pivots, broadcast rules, etc.)

Consolidation logic and reconciliation (metadata)

Interface layer (from planning systems)

Extraction layer (which tables, when, where, etc.)

Transaction layer (ERP)

the structure or lens through which managers need to see their business activity. The authors believe that a study of the requirements for these tools would reveal inconsistencies with the data model of the underlying ERP system. To what extent these inconsistencies can be modeled and therefore anticipated may provide a clue to the gap alluded to above. An obvious starting point here would be to research the integration of forward looking performance parameters (such as budgets, forecasts, and targets) into the ERP data model such that comparison to actual figures is automatic and real-time. The design implications of this type of automatic KPI generator (or dashboard) could best be explored through combined field research and laboratory prototype development.

Finally, the organisations studied show a common trait in the skill set required in the post-ERP organisation. It may be unclear as yet how best to develop such skills, however what is clear is the role they play in helping the organisation to get value from their investment in ERP software. These "functional analysts" are highly experienced business people who are inquisitive and understand the link between a business activity and the data recorded about it. They have an unshakeable belief in the value to be derived from the vast depositories of data available to them in the ERP system. We believe that in researching the skill sets as defined by these resources, companies will be able to set about more systematic methods for developing and maintaining the skills.

REFERENCES

Ackoff, R. L. (1967). Management misinformation systems. *Management Science*, *14*(4), 147-156.

Benjamin, R. I., & Blunt, J. (1992, Summer). Critical IT issues: The next ten years. *Sloan Management Review*, *33*(4), 7-19.

Davenport, T. H. (1998, July-August). Putting the enterprise into the enterprise system. *Harvard Business Review*, *76*(4), 121-131.

Dearden, J. (1972, January/February). MIS is a mirage. *Harvard Business Review*, *50*(1), 90-99.

Eisenhardt, K. M. (2002). Has strategy changed? *MIT Sloan Management Review, 43*(2), 88-91.

Galbraith, J. R. (1983). Organisation design: An information processing view. In Pennings, J. M. (Ed.), *Decision making: An organisational behavior approach* (pp. 131-139). New York: Markus Wiener Publishing Inc.

Gorry, G. A. (1971, Winter). The development of managerial models. *Sloan Management Review*, *12*(2), 1-16.

Gorry G. A., & Scott Morton, M. S. (1971, Fall). A framework for management information systems. *Sloan Management Review*, *13*(1), 55-68.

Holsapple, C. W., & Sena, M. P. (2001). Beyond transactions: The decision support benefits of ERP systems. *Journal of Decision Systems, 10*(1), 65-85.

Koch, C. (2001). ERP: Managing multiple processes in a state of flux. *Journal of Decision Systems*, *10*(1), 87-98.

Langley, A., Mintzberg, H., Pitcher, P., Posada, E., & Saint-Macary, J. (1995). Opening up decision making: The view from the black stool. *Organisation Science*, *6*(3), 260-279.

Leavitt, H.J. & Whisler, T.L. (1958). Management in the 1980's. *Harvard Business Review, 36*(6) 41-48.

Markus, M. L., Axline, S., Petrie, D., & Tanis, C. (2003). Learning from experiences with ERP: Problems encountered and successes achieved. In G. Shanks, P. B Seddon, & L. P. Willcocks (Eds.), *Second-wave enterprise resource planning systems*: *Implementing for effectiveness* (pp. 23-55). Cambridge, UK: Cambridge University Press.

Mason, R. O. (1969, October). Basic concepts for designing management information systems. *Accounting Information Systems* (Research Paper No. 8) (pp. 81-95). University of California, Graduate School of Management.

Mintzberg, H. (1989) *Mintzberg on management*. New York: The Free Press.

Pfeffer, J. (1992). *Managing with power, politics and influence in organisations*. Boston: Harvard Business School Press.

Shang, S., & Seddon, P. B. (2003). A comprehensive framework for assessing and managing the benefits of enterprise systems: The business manager's perspective. In G. Shanks, P. B. Seddon, & L. P. Willcocks (Eds.), *Second-wave enterprise resource planning systems*: *Implementing for Effectiveness* (pp. 23-55). Cambridge, UK: Cambridge University Press.

Staehr, L. Shanks, G., & Seddon, P. B. (2004). Understanding the business consequences of ERP systems. In F. Adam and D. Sammon (Eds.,) *The enterprise resource planning decade: Lessons learned and issues for the future*. Hershey, PA: Idea Group Publishing.

Stefanou, C. J. (2001). Organisational key success factors for implementing SCM/ERP systems to support decision making. *Journal of Decision Systems*, *10*(1), 49-64.

Zuboff, S. (1988). *In the age of the smart machine: The future of work and power.* New York: Basic Books.

ADDITIONAL READING

Boersma, K. & Kingsma, S. (2005). From means to ends: The transformation of ERP in a manufacturing company. *Journal of Strategic Information Systems*, *14*, 197-219.

Browne, J., Harhen J., & Shivnan, J., (1996). *Production management systems: An integrated perspective*. Guildford, UK: Addison-Wesley.

Lee, Z., & Lee, J. (2000). An ERP implementation case study from a knowledge transfer perspective. *Journal of Information Technology*, *15*, 281-288.

Mintzberg, H. (1981, January/February). Organization design: Fashion or fit? *Harvard Business Review*, *59*(1), 103-116.

Pinsonneault, A., & Kraemer, K. (1993, September). The impact of information technology on middle managers. *MIS Quarterly*, *17*(3), 271-292.

Chapter XV
Planning and Deployment of Dynamic Web Technologies for Supporting E-Business

John A. Hines
Arkansas State University, USA

ABSTRACT

More and more, internal applications are being moved from legacy systems into a more flexible Web-based environment. The issue concerning World Wide Web technologies is important to today's businesses. Decision making in this area is complex and needs to consider carefully the characteristics and needs of the entities employing these technologies. It has furthermore become clear that the Internet, in particular the World Wide Web, is playing an increasingly larger role in how people communicate. Through this research, technologies used to serve dynamic Web content are compared. This comparison includes performance as well as cost issues, the things that professionals in the business world face when deciding the best implementation of Web server technologies. Existing studies cover a limited scope of the overall picture, and research has thus been focused into very narrow aspects of the global entity. However, the continuing developments in Web technologies dictate the need for a broad scope approach to comparative studies in this field. Such a scope is pursued in this research.

INTRODUCTION

More and more, internal applications are being moved from legacy systems into a more flexible Web-based environment. The issue concerning World Wide Web technologies is important to today's businesses. Decision making in this area is complex and needs to consider carefully the characteristics and needs of the entities employing these technologies. It has furthermore become clear that the Internet, in particular the World Wide Web, is playing an increasingly larger role in how people communicate. Through this research, technologies used to serve dynamic Web content are compared. This comparison includes performance as well as cost issues, the things that professionals in the business world face when deciding the best implementation of Web server technologies. Existing studies cover a limited scope of the overall picture, and research has thus been focused into very narrow aspects of the global entity. However, the continuing developments in Web technologies dictate the need for a broad scope approach to comparative studies in this field. Such a scope is pursued in this research.

The planning and deployment of dynamic Web technologies is an in-depth and daunting task for many organizations. Purely scientific issues, such as performance, are important but a complete analysis must also include consideration of business perspectives like costs and benefits. Planning starts with a basic listing of what is to be accomplished through dynamic Web technology from a performance standpoint. Then the process steps through analysis of the technologies from a business perspective via a total cost of ownership (TCO) return on investment (ROI), and/or other financial analysis.

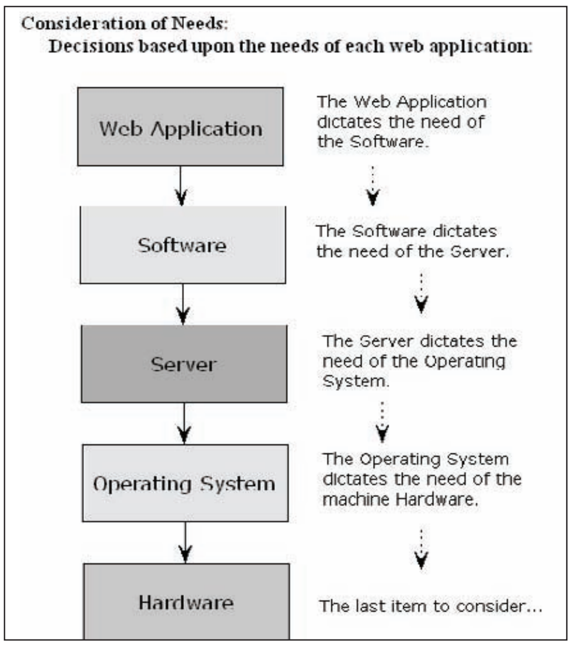
BACKGROUND

Choices to be Made

E-business platform planning begins with an analysis of the demands of the Web applications to be deployed. What is the Web application going to have to do? Does it need to store data? Is it an information service, or some other application? Serving as an example, a typical inventory management system will be used herein. This model is quite common among internal Web applications, and there is a multitude of ready-to-go applications available for purchase. While these may not exactly fit the needs of the organization, most can be tailored; so, for the purposes of this research, the inventory control system does represent a production-like application for demonstrating e-business platform planning and deployment. Such applications need features such as a database to store item inventories and product descriptions and a means to track session variables so the application will remember who is altering internal data and what information the employee has chosen to modify. Figure 1 summarizes the analysis, which starts from the top (designing the application) and works down through the different technologies that are needed.

Web server selection. Given an application, the first platform item to consider is the Web server software. Web server software comes in a variety of flavors and, in case of Windows operating systems, is included with the operating system. For

Figure 1. Assessing needs of Web applications using a top down approach (Hines, 2006, pp. 628)



a simple internal inventory control Web application this is normally sufficient, although features of other Web servers may render them more attractive. Some applications may need a third party commercial product; for example, Oracle Application Server may be required if the Web application is highly data intensive. Nevertheless, for the remainder of this study, we will utilize either what is included or what can be acquired through the open source initiative.

In conjunction with the selection of a Web server, the language and extra modules the Web application needs to run successfully must be determined. Languages such as PHP are supported by a third party language interpreter that is downloadable through the open source initiative. Other languages such as ASP require either a third party solution or support is through the operating system utilized. ASP is a Microsoft brand technology and is supported through the installation of Internet Information Services (IIS), a Web server software solution that is included with the Microsoft Windows operating system.

Also in conjunction with the planning of the Web server, database storage must be considered, since the inventory control Web applications nearly always must deal efficiently with large volumes of data. Database solutions can vary considerably in price. For example, the MySQL and PostgreSQL database solutions are licensed under the general public license (GPL), the same no-charge open source initiative licensing schema under which UNIX derivatives (e.g., Linux) are. In contrast, Microsoft offers its enterprise-level database management system, SQL Server, which can be very expensive and is licensed in a similar fashion to that used for Windows Server solutions. Similarly, Oracle offers what is by many considered to be the premier database solution, which is priced accordingly and was the driver for the development of the Oracle Application Server.

Operating System Choices. Once the Web server issues have been decided, the operating system software must be considered. Operating systems such as Linux (or other derivatives, including FreeBSD and OpenBSD, of the UNIX operating system) are available from the open source community and are free for downloading. In lieu of an up-front purchase, the purchase of a support contract for these open source distributions is normally arranged with the distributors. That is, for an organization that is operating a Web server for commercial use, a support contract is generally necessary in order to utilize the free distribution. At the other end of the continuum, proprietary operating systems such as Microsoft Windows Server 2003 Enterprise carry a heavy initial price tag, which includes support and maintenance. They are usually priced on a "per processor" basis.

Choosing the Hardware. The final aspect needing to be considered is the physical hardware (or, the machine) upon which the Web server and other software solutions are to run. There are multitudes of server hardware solutions available, such as the Dell PowerEdge, the HP Pavilion, and the IBM Blade servers. These servers can range in price from $1,000 to $35,000 depending on the internal hardware utilized, the included features, and vendor support. This is a very time sensitive issue, as hardware vendors are continually developing new and/or improved solutions, so basing this research on what is presently a robust and popular configuration seems appropriate. One of the most popular preconfigured servers is the Dell PowerEdge and through the remainder of this chapter it will be assumed this is the server architecture being utilized. This machine comes configured with 2 Dual Core Intel® Xeon® 5130, 4MB Cache, 2.00GHz, 1333MHz FSB processors, 2GB 533MHz (4x512MB), Single Ranked DIMM Random Access Memory chips, a 1x4 Backplane for 3.5-inch Hard Drives, a PERC 5/i, x4 Backplane, an Integrated Controller Card, an Integrated SAS/SATA RAID 1, PERC 5/i Integrated, two 73GB, SAS, 3.5-inch, 10K RPM Hard Drives, and a Riser with 3 PCI Slots at a cost of $4,184 at the time of this writing (Dell, 2006).

Financial Matters

A financial analysis is needed to provide a complete comparison of different operating systems, Web server software, and database software, as well as the staff requirements for administering these systems and the physical hardware. The TCO approach is initially a good way to understand future costs; however, the focus is on costs, without regard to benefits, whether tangible or intangible. Therefore, an ROI analysis is also needed to develop an understanding of the benefits with respect to costs. According to Wettemann (2003), the problem with total cost of ownership is that, used alone, TCO provides a very narrow view of just the costs associated with an application and completely ignores the benefits. The objective should not be to choose the cheapest application, but to choose the application that provides the greatest benefit or return for the company. Further,

It's [TCO is] frequently used to benchmark the costs of managing a vendor's application or piece of hardware against the costs for industry rivals' products, which isn't always an apples-to-apples comparison. And a product can have a low cost of ownership but not be the best managed. (Hoffman, 2002)

By adding ROI analysis to the scope of the overall analysis, TCO can be taken into account and adjusted by the added consideration of investment evaluation which will give a better overall picture of the costs and benefits. The ROI for an application is a percentage rate that measures the relationship between the amount the organization gets back from an investment and the amount invested (Whitten, Bentley, & Dittman, 2001, pp. 376). To accomplish this, the magnitude and timing of expected gains are compared with the initial investment cost (Schwalbe, 2002, pp. 92-93). The decision of which technology is to be used will be partially determined by higher return of investment. The final decision will depend on all factors equally.

COMPARING THE OPTIONS

The different available solutions and needs of the organization will dictate what server software, hardware, database solution, language, and the type of employees needed to develop and maintain the Web application.

Web Applications Development

While the internal inventory control application is being used to demonstrate the platform selection process for e-business planning and deployment, various features can and are typically incorporated. Such features include information portals, education portals, internal Web applications for other internal data processing, Web logs, discussion forum or bulletin boards, Web mail, and last but not least, multiplayer online games. Questions consequently arise about data storage and the sensitivity of those data. In the case of a pure inventory control application it would be required at the least to have items data concerning available for purchase made public. Beyond that, the information gathered from customers (e.g., their personal or financial information) will need to be secured so no outside source may access that information. Obviously, as features are added beyond the pure shopping cart application, the data storage and management questions become more complex.

Whatever the features incorporated into the Web application, consideration needs to be made for the functionality of the application. In addition to data management, for example, graphic formats and file sizes need consideration. If the graphics are too large in file size the Web application will be slower to move from one Web page to the next. Arguments can be made about this issue concerning the available bandwidth of the client but for development purposes it should always be assumed that the client will have slow connectivity and outdated hardware.

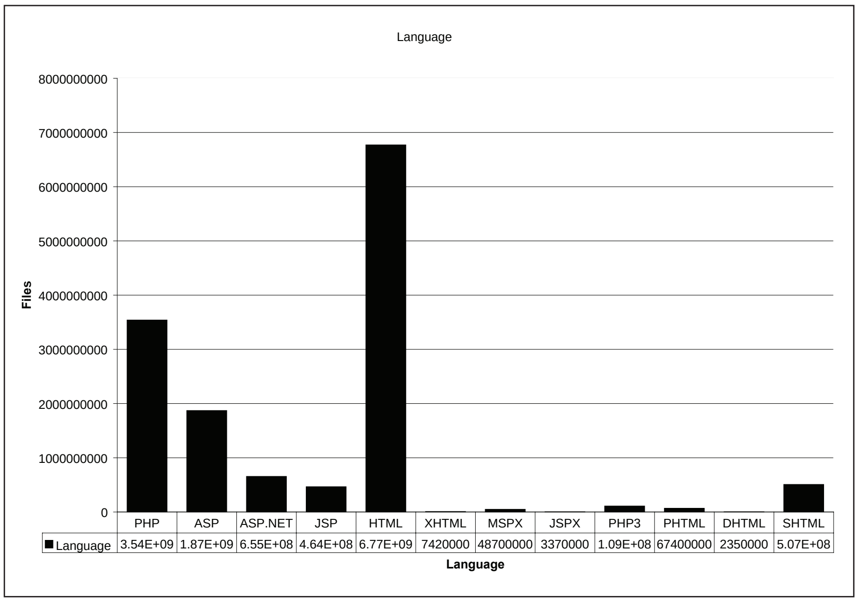
To act as middleware in a Web application there are many choices available. The clear forerunner is J2EE (a Java programming platform), which is commonly used in Web-based enterprise applications. Application frameworks – sets of modules or other such tools that make the creation of Web applications easier – are also available for use. Libraries such as Perl DBI or PEAR provide strong functionality and drastically reduce development time. Other frameworks, such as the Microsoft .NET framework, provide tools and functionality for applications that are not specific to the Web but may be used for Web application development. A large list of different libraries and frameworks for Web application development is available on the Wikipedia Web site (Wikipedia, 2006). Predeveloped packages present a variety of further options that might be considered. There are packages that are, in essence, plug and play. That is, they can be downloaded and installed with minimal to no Web language knowledge. Solutions like OpenPro (2006) and Compiere (2006) are complete solutions that employ administration interfaces to give developers full control over the content and aesthetics of the Web applications. Some of these packages are released to the general public under the GPL, the Limited GPL, or other such free licensing schema. In addition, there are other Web applications which are available for purchase (e.g., the Epicor Vantage/Vista or Intuitive ERP packages), and these can range in price from $2,000 to $500,000.

When Web applications are to be fully developed in-house, rather than acquired as packages, language support for the development of those applications is important. Nevertheless, this support can generally be considered de minimus due to the fact that most Web languages are supported within the Web server software. PHP, an exception, is open source and easily attainable, as well as compatible with most platforms. Further, under a Linux environment, PHP support is usually packaged along with the distribution used and is installable as the operating system is initially installed. Alternatively, PHP under the Windows environment is installable as an extra package available from the PHP Web site.

Note that PHP as a server-side scripting language is referenced throughout this research because of its popularity, as evidenced by a Google “allinurl” search. An “allinurl” search on the Google search engine using a file extension (e.g., “asp,” “php,” “htm,” etc.) as an argument will provide a rough guideline as to the market share for a given Web programming language. The term “allinurl:.php” produces 3.5 billion results. When compared to other file extensions in this manner (Figure 2), PHP is the front runner in server-side Web languages. In fact, the only other file extension to have produced as many or more hits is the “html” extension, which is the basic hypertext protocol and is not rendered on the Web server but, in fact, is rendered by the client’s Web browser; that is, HTML is not a server-side scripting language. Even the combination of the entire examined lot of Microsoft file extensions (.asp, .aspx, .mspx, .jsp, etc.) do not add up to the market share PHP appears to enjoy.

Web languages such as PHP or ASP typically will perform best on their native platforms; for example, PHP pages running on Apache in Linux environment will outperform ASP pages being processed on that platform (Hines, 2006). Nevertheless, PHP running in its native environment has also been shown to outperform other scripting languages running on their native platforms (Hines, 2006). While ASP and other languages offer powerful functionality, PHP has the most utility built in. For example, PHP boasts built in support for FTP transfer, e-mail, data compression, MD5, XML, and file upload capabilities, where other Web languages could and often do require third party solutions that could become quite expensive. ASP does include built in support for using Microsoft SQL server and Microsoft

Figure 2. Google "allinurl" searches produce rough estimates of language market share by file extension



Access files for data storage. PHP can do the same with the addition of built in support for MySQL, PostgreSQL, Oracle, and many other data solutions (Anstey, 2003). It is safe to assume that PHP's popularity can more than likely be attributed to its high level of built in features, functionality, and the fact that the PHP project is open source. In Figure 3, some of the more popular Web languages are outlined along with licensing information and/or cost.

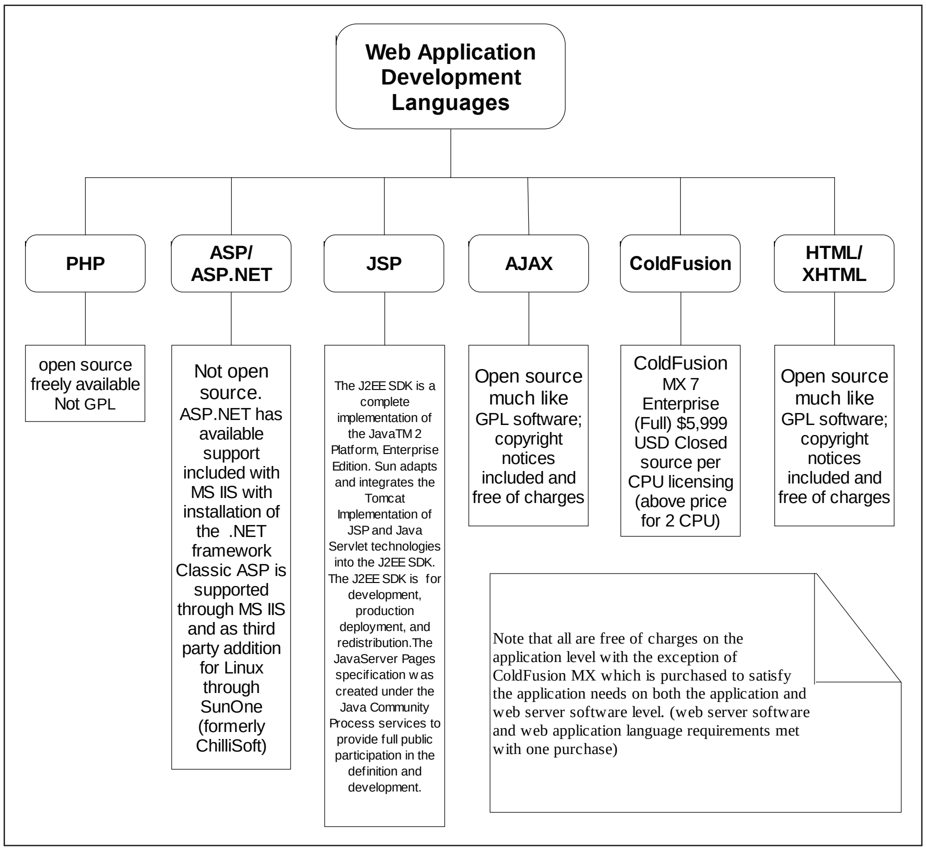
Hiring a Web developer to construct and maintain the application can be a major component of the total cost associated with Web applications. According to a recent article from CIO Update (Haber, 2005), a lead application programmer can command a salary of $72,000 to $98,250 depending on experience. Presented with that information a small business owner or a newly formed small business could easily panic. It may become necessary for a smaller organization to make compromises in this area, even so far as to have their existing IT staff learn how to construct Web applications. However, third party solutions for Web application development are also available. Depending on the requirements a contracted Web developer may charge in the neighborhood of $10,000 for a basic but complete inventory management package to serve a small to medium-sized business.

Because of its general popularity, as well as its generally strong comparative performance, PHP will herein be considered the language of choice. It will also be assumed that the small business owner will attempt construction of a Web application, a reasonable assumption given the plethora of technical, as well as nontechnical, resources and tutorials available on the World Wide Web.

Web Server Software Options

When Web server software is being considered, the issue is twofold: there must be consideration of both the database server and the Web server, which generally work in conjunction with each

Figure 3. An overview of Web development languages



other. Nevertheless, database servers and Web servers are not the same, and different issues affect the selection of each.

When data storage is examined, many ways to store data will be discovered. For example, data can be stored directly to files or can be inserted into a data solution through a TCP/IP connection from within the Web application. Furthermore, database solutions have the highest spread in price range of any requirement for the Web application. Open source database solutions such as MySQL, PostgreSQL, and DB4 are released under the GPL and thus have no up-front purchase involved. Other commercial products, such as Microsoft SQL Server, can be considerably more expensive; for example, Oracle's database solution, Oracle 10g, at the high end of the price range, carries a price tag of $96,000 for a one-CPU license of their Enterprise Edition. *PC Magazine* (Dyck, 2002), which performed a benchmark study on databases, states

We hesitated to award the blue ribbon to a product that costs twice as much as its nearest competitors. But in the end, Oracle9i prevailed thanks to its superior feature set, reliability, and performance. Oracle9i offers the most advanced management and performance tools, and it topped the other products on our throughput tests

(Oracle's current release is version 10g.). According to a comparison published on the Microsoft

Web site (Microsoft SQL Server, 2004), Microsoft SQL Server 2000 costs half as much and has the power to make it a contender in the market vs. Oracle 10g.

There have been vast improvements in database solutions, as well as options, since the publication of the *PC Magazine* article and the Microsoft comparison. For example, when that article was published MySQL version 4 was examined and is currently up to version 5.0. The current version has introduced a much more feature rich set of administration tools, a greater amount of functionality, and drivers that will let MySQL provide connectivity to a number of different dynamic Web scripting engines and libraries including the .NET framework. In an article published on the World Wide Web (Turvey, 2005), four different database solutions were compared: Oracle 10g, MySQL 4.1.14, Microsoft SQL Server Express, and DB2 Express. The author states:

Oracle 10g and DB2 are acknowledged as industry leaders in the database field and both are strong in terms of features and are surprisingly easy to use with clean GUIs. However, one of the other databases tested is also acknowledged by the industry and has proven itself in terms of reliability and performance in many very large installations around the globe. And what's more it is free!

The improvements since this study in regards to MySQL and its recent 5.0 release have given the competition a run for their money. With add-on software that is provided for free download on MySQL's Web site, administration and implementation of a MySQL database server can be performed by people who have relatively little experience with databases.

While database servers manage the extensive amounts of data and are crucial to the efficient operation of a Web site, they nevertheless represent a tier of support beyond the basic client-server architecture of the Web. Web server software on the other hand is the core of the delivery of a Web application, interacting directly with the Internet and the client browsers employed by end users. The Web server utilizes the scripting engines that allow the processing of dynamic Web languages

Figure 4. Most common Web server software options

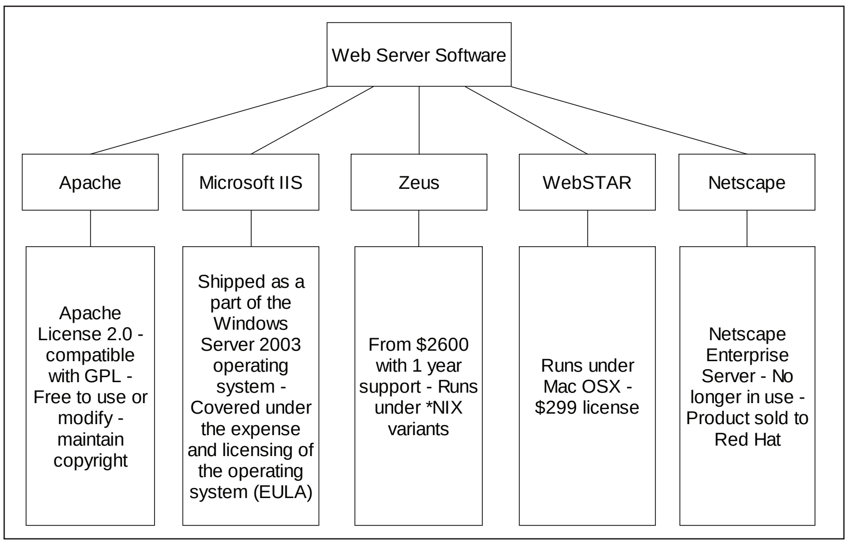
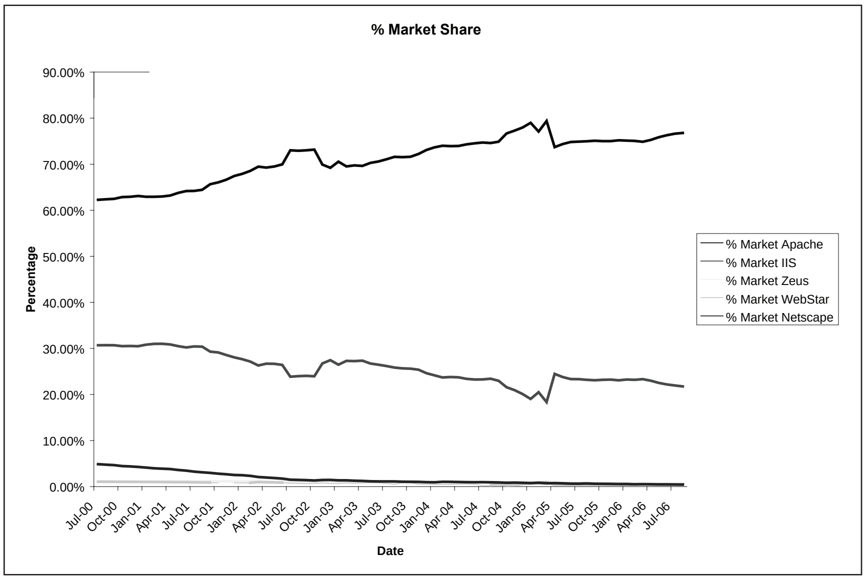


Figure 5. Percentage of market share for the top five Web server software packages



such as ASP, ASP.Net, PHP, Java (for servlets), JSP, and many others. As might be expected, performance of a Web server is proportional to the amount of work load placed upon it. In Figure 4 the top five Web server software packages are listed along with some general information about cost and/or licensing.

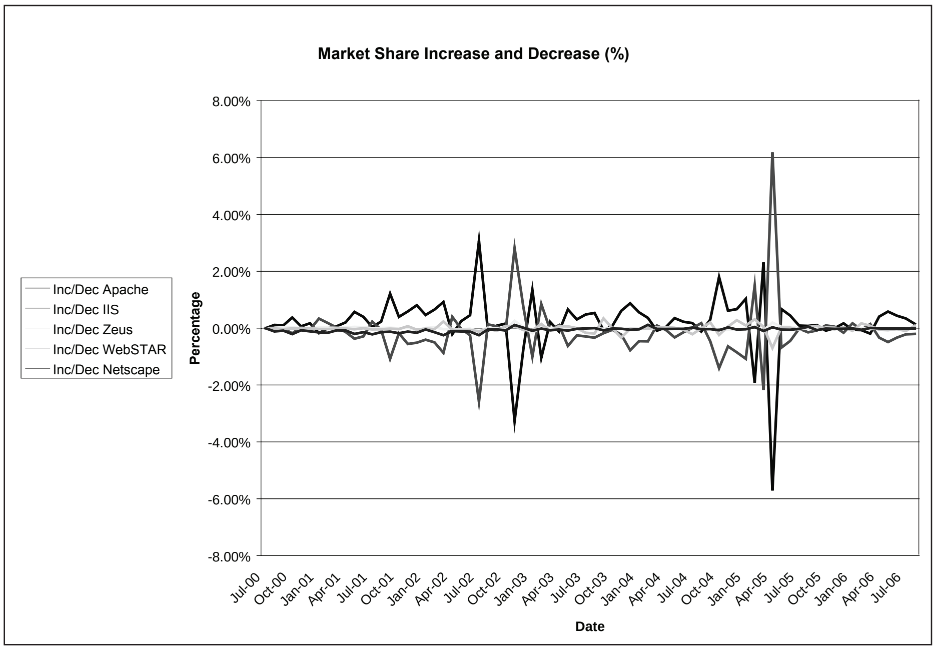
Different Web server packages will perform better in specialized areas. Nevertheless, most if not all will do their main job function, which is serving Web content, quite adequately. Through the use of the SecuritySpace Web site (SecuritySpace, 2006), Web server market share statistics were gathered and analyzed for the top five Web server software packages. Over twenty million Web sites were surveyed, and the results calculated from the collected data show that the Apache Web server enjoys a 77% market share of the top five Web server data (Figure 5).

It is notable that Web server software market share is, in essence, defined by the top two software packages, the Apache Web server and Microsoft Internet Information Services. Looking at Figure 6, one can further see this dominance in the Web server world by Apache and Microsoft IIS. When the Apache package gains market share Microsoft loses close to the same amount, and vice versa.

However, the SecuritySpace results do not match those that can be found on Port80Software's Web site (Port80Software, 2005) stating that Microsoft Internet Information Services hold a commanding 54% of the current market share followed by the Apache Group with 23% market share. One notable aspect of Port80's survey is the selected survey group of Fortune 1000 members. The Port80 group states

While the well-known Netcraft Web server surveys attempt to look at the whole of the Internet, Port80's surveys focus on a small but important pool: the 1000 largest corporations in the U.S. We firmly believe that this yields more reliable results and, more importantly, the Port80 Software survey looks to those sites that demand the most of their Web servers.

Figure 6. Increase and decrease in Web server market share shown as a percentage. Only the top five by popularity are included.



The number of Web servers, according to SecuritySpace, just considering domains with a .com extension reaches nearly the nine million mark. If each of the Fortune 1000 had 10 Web servers, this figure (10,000 instances) would not come close to explaining Netcraft's or SecuritySpace's survey results for just the .com extension. Very possibly, the fact that IIS is the choice for 75% of all intranet Web servers (Alwang, 1998) and that the Port80 survey questions did not distinguish between intranet and Internet applications is responsible for the disparity in market share estimates.

One aspect of choosing a Web server is Web site portability. While Microsoft's IIS is not portable to any environment other than Microsoft Windows, the Apache Web server is quite portable and will run in a variety of environments. This is important for keeping a business from becoming locked into any particular hardware/software platform for multiple decision cycles. Further, the choices concerning other aspects of the e-business platform, such as scripting language, are affected, as IIS functions at its best only in conjunction with Microsoft brand technologies and vice versa. For example, the fact that ASP-enabled (or any other) Web applications will suffer minor performance drawbacks in non-native environments (Hines, 2006) is worth noting. These non-native implementations will require third party software such as Sun Microsystems ASP ONE server (Sun Java System Active Server Pages 4.0).

Operating System Considerations

On the operating system front, there are many heated debates about which operating system performs better, has better security, and/or is cheaper, not only initially but also in the long run.

Whether for security reasons or other concerns, many people in the system administration world have developed a favorite; for example, some favor Microsoft Windows over Linux and others vice versa. Application developers choose a favorite for other reasons, but in essence everyone has a brand to which loyalties lie. There are many alternatives available to Microsoft Windows or the various distributions of Linux. For example, Sun Microsystems has a UNIX operating system available called Solaris, which is very stable and robust. Other UNIX distributions include Digital UNIX, FreeBSD, IBM AIX, NetBSD, and OpenBSD, all of which perform suitably to support the needs of a Web application, run robustly, and have good stability "right out of the box."

From the Microsoft (i.e., Windows) world, for the hosting of limited Web applications even the desktop operating system, Windows XP Professional, will do the job. However, the support and power of Windows Server 2003 in conjunction with the security and power of the included Internet Information Services 6.0 (as opposed to IIS 5.1, which is included with Windows XP Professional), makes the enterprise level option (Server 2003) a desired choice over any desktop solution. An OEM version of Microsoft Windows Server 2003 small business server premium (5 client license) is available for purchase for approximately $830.

For the initial server set up, clearly the open source alternatives provide the lowest cost choice as far as operating systems are concerned. Nevertheless, even with the Windows Server, the cost for acquiring an operating system is minimal or, in many cases, negligible, as an operating system is often installed when a server machine is purchased. The market dominance of Microsoft Windows means, however, it is a larger target for viruses, and new security vulnerabilities are discovered daily, so a qualified administrator must have experience dealing with security issues. It might be concluded that, while the same applies to an administrator for alternative operating sys-

Table 1. System administrator salaries (The Enterprise Systems Staff, 2006)

Year to Year Change						
2006	**2005**	**2004**	**2002**	**2001**	**1-year change**	**4-year change**
$61,600	$60,500	$56,000	$62,000	N/A	+1.8%	-0.7%

Average Annual Bonus						
2006	**2005**	**2004**	**2002**	**2001**	**1-year change**	**4-year change**
$3,000	$2,600	$1,800	$2,400	N/A	+15.4%	+21.0%

By Experience Level			
< 5 years	**< 10 years**	**10+ years**	**Differential**
$55,100	$61,300	$67,000	+21.6%

By Operating System					
AIX/UNIX	**Windows**	**Midrange**	**Mainframe**	**Non-Mainframe Linux**	**Windows Only (Non-mainframe)**
$63,400	$63,400	$64,500	$62,000	$59,900	$57,600

By Application Environment					
ERP	**CRM**	**B2B**	**B2C**	**Supply Chain**	**Data Warehouse**
$61,800	$61,800	$65,800	$65,600	$63,200	$62,700

tems, security for those systems tends to be less of an issue, as the alternatives are a smaller target. Such a conclusion is simply untrue. Security, no matter what the operating system, is a large issue in administrator consideration and should be a priority. Information retained about customers is vital and should be taken very seriously.

Over the life of the Web application, there are other aspects of operating systems costs to consider, in particular, operating and administration costs. To begin, employee and other related staff salaries must be considered, and these can vary according to the platform being supported. There is controversy over the skills it takes to be a Linux (or UNIX) administrator, as compared to a Windows administrator. Schenkenfelder argues that the typical Linux administrator can handle more than the typical Windows administrator. "What I've found is that a Linux administrator who knows what he's doing should be able to administer two to three times the amount of boxes a Windows administrator should be able to administer," he said (Gross, 2003). According to a salary survey performed by The Enterprise Systems Staff (2006), Linux system administrators do enjoy a somewhat higher salary than do Microsoft Windows system administrators.

Consideration of all costs, both long term and short term, is important but can seem to muddle the selection process for operating systems. Short term savings of free distributions may or may not outweigh the long term savings from the hiring of the less-skilled administrators supposedly needed for proprietary, more "user-friendly" operating systems.

SELECTING THE HARDWARE

Generally speaking, computing machines on the World Wide Web, or any other type of network where machines communicate with one another, are classified as either a server machines or client machines. When constructing or purchasing a machine intended to be used as a server machine, it must be expected that this machine will serve many client machines. In the case of the World Wide Web, many client machines, in this case operated by end-users who want or want to adjust information and/or to purchase materials or services, will connect to a server that is running a Web server. Consequently, the server machine can be considered the mechanical representative of the business or organization; this representative provides access to (and serves as a gatekeeper for) the information or products that the organization wants to make available to management. The server machine is thus the vital gateway into the virtual world of corporate data to be made available to and for management. This gateway must be clearly defined and must be easily accessible to those who need service.

Having a server machine that is quick to respond and deliver information is a key factor of productivity. In some cases, in a small business that does house its own Web content; there may be the temptation to have the same machine hardware act simultaneously as a user workstation, an e-mail server, and/or a database server. Furthermore, developers might also use that server as a development machine, and this means the Web server will take dramatic performance hits when the developer is working on compiling code, rendering high definition graphics, or other CPU-intensive and/or memory intensive applications. To house a Web server the hardware used must be able to serve many clients at the same time. For example, some Web sites serve hundreds of thousands of end users in a month's time. If the Web server is not powerful enough, it will limit network activity and may cause network failures, thus denying service to end users and placing future income and productivity for the firm at risk.

For simultaneous access to a Web application, CPU configuration and RAM will play key roles, as will available network bandwidth. The main objective of configuring a machine to serve Web applications should then be to reduce as much as possible or even eliminate the likelihood of bottlenecks that an end user would experience. Disk access is also of great importance, since so much content will need to be delivered so frequently. Therefore, setting up a redundant array of inexpensive disks (RAID) strip (SCSI) utilizing three hard disks for speed, and possibly a fourth hot swappable disk, for access will yield the highest benefit. Of course, the machine that does the most reading and writing of data, that is, the database server, will benefit most from a RAID strip. Hence, if the database for a Web site is housed on a machine different from the Web server, the machine with the database will benefit more from this particular setup. A multiprocessor machine will also boost performance by letting jobs process faster through the CPU. If a single machine houses both the database and the Web server, a multiprocessor machine can be beneficial even for single threaded services, such as the ASP interpreter, because of the performance load placed upon the hardware from the operating system, the Web server, and the database server.

Internal components present further decisions to be made, although one way to avoid some of this decision process is simply to purchase a preconfigured machine from a vendor such as Dell, Hewlett Packard, or Sun Microsystems. It is important, nevertheless, whether a firm is purchasing components for building a custom machine or purchasing a preconfigured server, to determine what the minimum acceptable configuration will be. This choice is dependent on the type of content to be processed by the Web application. For a purely informational Web application, that is, a Web site with static content only (atypical for an ecommerce Web site), not much is required as essentially no processing is involved at the server other than simply to deliver requested files. However, in the case of file transfer or streaming media, demands of the Web application and server hardware increase, and a fast connection to the Internet becomes necessary as clients are connected to the Web server machine for greater volumes of interaction and potentially much greater periods of time without excellent Internet connections. Further, a Web application that utilizes the common gateway interface (CGI) or otherwise serves active content (as with ASP or PHP Web pages) and involves database access will require faster CPU cycling.

A business that provides application interfacing, such as gaming or Web-based e-mail services, will require highly reliable network and fast computing performance. An essential item to include for any Web server regardless of its role is an uninterruptible power supply (UPS) to enable continued operations (or at least facilitate dependable backup and recovery) in the case of a power outage. In general, the items requiring a UPS are anything and everything involved in the connection between customer access points and the server machine itself. If one component in the chain between the network and server machine goes out of service the entire service is disabled.

If an organization has someone capable of constructing computing machinery, money can be saved by purchasing individual components and building a customized machine to act as a Web server. In this case warranty tracking for individual components becomes an important issue as components are generally purchased from multiple hardware vendors. In contrast, when a premanufactured server machine is purchased, service and warranty issues are placed in one location with a single level of responsibility involved. Usually, the vendors of premanufactured machines also provide extended service and repair contracts that can be arranged when equipment is acquired from those vendors. Care should nevertheless be taken when a warranty is included with the purchase of a premanufactured machine. Some warranties will not cover hardware used in conjunction with other hardware purchased from different vendors or if there is a balance due associated with the purchased machine. Warranties vary in contract length and pricing, but, if an item outlasts its warranty, the extended warranty should not be considered as having been a bad decision. The quality of hardware components may be undetectably poor, whether as a result of manufacturing defects, bad design, or other circumstances, and a warranty helps to balance the unknown level of risk involved. A small business can therefore benefit from the extra investment into an extended warranty to avoid being stuck with a "lemon."

Another Alternative to Consider: Outsourcing

When the costs involved with developing a solution to deliver Web content are examined many factors come into play, and outsourcing generally surfaces as an option. To simplify the decision process many organizations will outsource development work and Web application hosting to third parties and will thus bypass an examination of the complete picture (needs, performance, costs, etc.). While overall economic analysis is needed, some small businesses do not possess the knowledge or tools to perform such an analysis. Such operations, for example, the family or home business, use their resources in order simply to feed the family and pay the mortgage or rent. In other words, the extensive use of the World Wide Web has "leveled the playing field" for small businesses and has enabled them to compete with larger organizations, but "getting into the game" still requires assistance often not available to small businesses.

Acquiring the licensing and software, as well as the hardware, needed to develop an effective Web site can be quite expensive. Many small to medium-sized businesses will thus wisely choose to outsource its development and hosting to a third party solution provider. A third party would then be responsible for the maintenance and upgrades to server machines. It would also be responsible for features such as access to database solutions, third party development software, and promised uptime and network connectivity. Hosting companies provide numerous types of hosting solutions from shared Linux or UNIX hosting accounts with minimal features from $10 per month to dedicated, one customer server machines or Microsoft Windows Server (2000 or 2003) priced from $75 monthly to $200 monthly.

Hiring a third party solution will consequently take away part or all of the need for skilled employees and replace it with a monthly bill that may cover site development, server administration, database maintenance, and/or other aspects of supporting the e-business platform.

A Summary

Based upon the comparisons provided above, example costs involved in developing and deploying an e-business Web site are summarized below in Table 2. The cost analysis is somewhat simplistic but should provide a general idea of the costs involving deployment of Web applications. Furthermore, the table provides a structure to guide a firm in developing its own situation-specific comparisons and analyses. That is, a formal detailed cost analysis should be performed in practice, and the intention here is to guide the decision making, rather than to provide it.

Clearly, at first glance, one can see that the option to outsource, whether completely or partially, presents significant cost savings. In order to develop and house a Web application internally the resources (particularly monetary) must be present, and this commonly poses a major hurdle to small businesses in particular. It is nevertheless possible (and sometimes desirable) for a small business to host and develop a Web application internally, but the internal solution brings the sometimes hard to meet responsibility of machine and user security, as well as data integrity. Control of the entire platform would, however, be restricted to individuals working within the organization, and this may be a feature that outweighs the costs in-

Table 2. Rough guideline estimates for development and hosting costs

	Complete Internal	Partial Outsource	Complete Outsource
Machine/Hardware	$2,000 - $6,000	$10-200	$10-200
Operating System	$0-$2,600	$0	$0
Database Server	$0-$96,000	$0	$0
Web Server Software	$0-$5,000	$0	$0
Web Scripting Engine	$0-$300	$0	$0
Network Connectivity	$75-$1,000 Monthly	$75-$1,000 Monthly	$75-$1,000 Monthly
Server Machine Maintenance	$55,000-$67,000	$0	$0
E-mail Services	$0-$2,500	$0	$0
Development Software	$0-$10,000	$0-$10,000	$0-$10,000
Total Estimate:	**$57,075-$190400**	**$85-$11,200**	**$85-$11,200**

volved. That is, the intangible benefits of internal control may offset the tangible costs associated with the in-house option.

CONCLUDING REMARKS

For internal Web applications, it can be assumed that the Web application will be a replacement for a process that had to be performed manually. The elimination of personnel supported operations or proprietary equipment can be a cost saving tool that will pay for the development and implementation of a Web application in the short run. Close attention must be paid to how end users are affected by and use the Web application. The example of an inventory control system used within this research is a small yet primary function of internal Web applications. Other additional features needed within the Web application will serve as a multiplier to the requirements that must be fulfilled. Further expansion of Web applications from the explicitly internal application to a hybrid application for both production management and customer interfacing, such as order placement or online billing, may become necessary. In this case close attention is needed when addressing connection and user security. A business of any type has a responsibility to keep user information secure to not only customers but its employees as well.

Another aspect to be considered for future expansion is the physical expansion of the organization. Once information moves from within one location, more than likely behind a firewall, to a multilocation organization the information becomes easily exposed. Again, user security is at the forefront of the requirements to be fulfilled. Low confidence in application security will mean users will not trust the system and thus will not use it. Internally, the firewall will prevent a lack in trust in the application but when the scope becomes hybrid or external in any nature user security will become more of a major issue in regards to user information.

Issues pertaining to Web application development and deployment are important to any organization involved in e-business. While these might be considered strategically, addressing the issues as inexpensively as possible is what concerns most organizations, especially small to medium-sized businesses. The main problem for small businesses is the initial costs, so it is often this aspect is what is addressed by a small business. That is, small businesses tend to take on high cost issues in a nearsighted manner. The return on investment and total cost of ownership is often ignored much to the detriment of their future bottom line. Once the needs of the organization are determined, which can be done by applying any of the typical software engineering models such as the waterfall method or, more commonly, by use of a questionnaire, a proper design can be developed for a Web application. Once a design is outlined for the Web application the other requirements concerning hardware, software, operating system, and Web server technologies will become apparent. Once the needs are determined then a cost analysis can be performed. This process should be followed multiple times over the planning phase. In general, each of the specific issues concerning the planning and deployment of Web technologies are not independent, and should not be treated independently.

FUTURE RESEARCH DIRECTIONS

Constant changes occurring in information technology, and more specifically Web application development, provide an opportunity to review the current state of research in this field. The subject matter covered in this chapter concentrates on a broad scope look at the needs of an organization concerning Web technology. The hardware used in deploying this type of application, as well as the hardware used in production equipment, is the key determinate of efficient technology. The advancement of hardware, software, and development languages are not predictable and constant studies of these developments are a necessity in order to simply maintain knowledge in information technology. On the front of hardware advancements, Intel Corporation is leading the way with breakthroughs in central processor development. By the end of 2007, they will have a CPU with embedded 45 nanometer transistors that promises to push Moore's Law to the extreme while lowering power consumption. Computer numeric control (CNC) machinery depends on processor power in order to perform calculations for positioning. This processing power will be utilized by the software industry from application, to development environments, and the operating system. Future research in this field will have to follow the trends of technology's development.

The technological advances listed above will also provide opportunity to research and develop new methods for the production environment based upon newer hardware architecture and software development. Benchmarking existing systems as well as developing faster simulations of different production environments will become easier utilizing new technologies. Companies are constantly working to develop the very best products and will take advantage of technology to gain or remain one step ahead of the competition. For example, using faster hardware for sizing a pilot production line will help identify bottle necks that may go unnoticed on slower machines.

The research will introduce energy efficient methods for manufacture that will lower costs and increase profits. Energy efficiency is a prominent topic for the United States Department of Energy [DoE]. The DoE's Industrial Technologies Program supports research that addresses improvements in industrial energy efficiency. Investment in energy efficient innovative technology can benefit not only the financial bottom line but also the corporate image and community relations of an organization. Through technology, this can be achieved. The future for technology in a production environment must focus upon energy efficiency, and research supporting this is priority.

REFERENCES

Compiere. (2006). *Compiere: Smart open source ERP & CRM*. Retrieved December 22, 2006, from http://www.compiere.org/

Dell. (2006). *Dell online store*. Retrieved July 26, 2006, from http://configure.us.dell.com/dellstore/config.aspx?c=us&cs=04&kc=6W300&l=en&oc=pe2950-med&s=bsd&fb=1

Gross, G. (2003, January 3). Linux TCO edge: Lower labor costs. *ZDNet*. Retrieved July 22, 2006, from http://techupdate.zdnet.com/techupdate/stories/main/0,14179,2907876,00.html

Hines, J. (2006).Confronting the performance issues of dynamic web technologies. In *Proceedings of the 2006 Southwest Decision Sciences Institute* (pp. 627-636).

Hoffman, T. (2002, December 2). TCO: Flawed but useful. *Computerworld*. Retrieved July 22, 2006, from http://www.computerworld.com/managementtopics/roi/story/0,10801,76240,00.html

OpenPro. (2006) *OpenPro: ERP software solution*. Retrieved December 22, 2006, from http://www.openpro.com/proic.html

Schwalbe, K. (2002). *Information technology project management* (2nd ed.). Canada: Course Technology, Thomson Learning Inc.

ServerWatch. (2006). *ServerWatch's ServerCompare*. Retrieved July 26, 2006, from http://www.serverwatch.com/stypes/compare/index.php/compare2_18333,18097

The Enterprise Systems Staff. (2006, August 8). 2006 ESJ salary survey, part 1: Salaries grow for all IT staff positions. *Enterprise Systems*. Retrieved August 21, 2006, from http://esj.com/Enterprise/article.aspx?EditorialsID=2052

Turvey, S. (2005, December 23). *Road test: Four databases tested*. Retrieved October 1, 2006, from http://www.builderau.com.au/architect/database/soa/ Road_test_Four_databases_tested/0,339024547,339224962,00.htm

Wettemann, R. (2003, July 2). The strengths and weaknesses of TCO. *CIO Magazine*. Retrieved July 20, 2006, from http://www2.cio.com/analyst/report1538.html

Wikipedia. List of Web application frameworks. *Wikipedia*. Retrieved September 1, 2006, from http://en.wikipedia.org/wiki/List_of_Web_application_frameworks

Whitten, J., Bentley, L., & Dittman, K. (2001). *Systems analysis and design methods* (5th ed.). New York: McGraw-Hill Higher Education.

RECOMMENDED READINGS

Ardagna, D., Trubian, M., and Zhang, L. (2004). *SLA based profit optimization in multi-tier Web application systems*. Paper presented at the International Conference on Service Oriented Computing (pp. 173-182). New York.

Bautista, M., & Lee, B. (2006, November 5-8). A recharge model for end user computing support: yes! people pay us to fix their computers. In *Proceedings of the 34th Annual ACM SIGUCCS Conference on User Services, SIGUCCS '06*, Edmonton, Alberta, Canada, (pp. 16-21). New York: ACM Press. Retrieved August 20, 2007, from http://doi.acm.org/10.1145/1181216.1181220

Bogdanova, G., Todorov, T., Blagoev, D., & Todorova, M. (2003, June 19-20). Use of dynamic technologies for Web-enabled database management systems. In *Proceedings of the 4th international Conference Conference on Computer Systems and Technologies: E-Learning,* Rousse, Bulgaria. In B. Rachev & A. Smrikarov (Eds.), *CompSysTech '03* (pp. 202-207). New York: ACM Press . Retrieved August 20, 2007, from http://doi.acm.org/10.1145/973620.973654

Candan, K. S., Li, W., Luo, Q., Hsiung, W., & Agrawal, D. (2001, May 21-24). Enabling dynamic content caching for database-driven Web sites. In *Proceedings of the 2001 ACM SIGMOD* *international Conference on Management of Data*, Santa Barbara. In T. Sellis (Ed.), *SIGMOD '01* (pp. 532-543). New York: ACM Press. Retrieved August 20, 2007, from http://doi.acm.org/10.1145/375663.375736

Cegielski, C. G., Reithel, B. J., & Rebman, C. M. (2005, August). Emerging information technologies: Developing a timely IT strategy. *Commun. ACM*, *48*(8), 113-117. Retrieved August 20, 2007, from http://doi.acm.org/10.1145/1076211.1076214

Dijkstra, M. (2004). A model for the management of dynamic Web applications. *ACM International Conference Proceeding Series*, *60*. Retrieved March 1, 2007, from http://lawrence.library.astate.edu:2081/ft_gateway.cfm?id=1052271&type=pdf&coll=portal&dl=ACM&CFID=12922547&CFTOKEN=72642690

Dix, J. (2006). Managing the virtual world. *Network World*, *23*. Retrieved Decembger 11, 2006, from http://lawrence.library.astate.edu:2524/login.aspx?direct=true&db=iqh&AN=23305383&site=ehost-live

DoE. (2006). Tax incentives for energy-efficient buildings and products. *U.S. Department of Energy*. Retrieved February 28, 2007, from http://www.eere.energy.gov/buildings/info/tax_credit_2006.html

Donovan, J. J. (1976, December). Databsse system approach the management decision support. *ACM Transactions on Database Systems*, *1*(4), 344-369. Retrieved August 20, 2007, from http://doi.acm.org/10.1145/320493.320500

Gupta, P., & Moitra, D. (2004, February). Evolving a pervasive IT infrastructure: A technology integration approach. *Personal Ubiquitous Comput.*, *8*(1), 31-41. Retrieved August 20, 2007, from http://dx.doi.org/10.1007/s00779-003-0254-7

Kimbrel, T., Steinder, M., Sviridenko, M., & Tantawi A. (2005, May). Dynamic application placement under service and memory constraints. Paper presented at the International Workshop on Efficient and Experimental Algorithms, Santorini Island, Greece.

Kühn, W. (2006, December 3-6). Digital factory: Simulation enhancing the product and production engineering process. In *Proceedings of the 37th Conference on Winter Simulation*, Monterey, California. In L. F. Perrone, B. G. Lawson, J. Liu, & F. P. Wieland (Eds.), *Winter Simulation Conference* (pp. 1899-1906).

Mandal, P. (2002). Implementation of information technology in a job shop manufacturing company: A focus on ManuSoft. In M. Khosrow-Pour (Ed.), *Annals of cases on information technology* (pp. 103-118). Hershey, PA: IGI Global Inc.

Mandal, P. (2004). Global issues in e-business (editorial). *Journal of Global Information Technology Management Edited Special Issue*.

McGinnis, L. F., Huang, E., & Wu, K. (2006, December 3-6). Systems engineering and design of high-tech factories. In *Proceedings of the 37th Conference on Winter Simulation*, Monterey, California. In L. F. Perrone, B. G. Lawson, J. Liu, & F. P. Wieland (Eds.), *Winter Simulation Conference* (pp. 1880-1886).

Nassif, S. R. (2006, October 4-6). Model to hardware matching: for nano-meter scale technologies. In *Proceedings of the 2006 International Symposium on Low Power Electronics and Design, ISLPED '06,* Tegernsee, Bavaria, Germany, (pp. 203-206). New York: ACM Press. Retrieved August 20, 2007, from http://doi.acm.org/10.1145/1165573.1165621

Newbury, M (2007). Training for power quality. *Maintenance solutions.* Retrieved Feb. 26, 2007, from http://www.facilitiesnet.com/ms/article.asp?id=6140&keywords=infrared,%20training

Pacifici, G., Segmuller, W., Spreitzer, M., Steinder, M., Tantawi, A., & Youssef, A. (2005). *Managing* *the response time for multi-tiered Web applications* (Tech. Rep. RC 23651). IBM.

Porter, J. D. (1969, August). Application of cost/effectiveness analysis to EDP system selection. *Install Manag. Rev.*, *1*, 9-39. Retrieved August 20, 2007, from http://doi.acm.org/10.1145/1147418.1147422

Seydel, J. (2006).Data envelopment analysis for. *Industrial Management & Data*, *106*, 81-95.

Shankar, P. (2006). Bangalore: Dell's new development engine. *Siliconindia*, *9*. Retrieved Jan. 14, 2007, from http://lawrence.library.astate.edu:2524/login.aspx?direct=true&db=iqh&AN=23305383&site=ehost-live

Songini, M. (2007). Canadian beef processor touts RFID. *Computerworld*, *41*. Retrieved Feb. 2, 2007, from http://lawrence.library.astate.edu:2524/login.aspx?direct=true&db=iqh&AN=23738093&site=ehost-live

Surkan, Ml. (1998, June 6). Load balancing has its limits. *PC Week*, *15*(27), 98.

Topley, J. (2005). *The design of software: Dynamic Web technology adoption*. Retrieved Jan. 5, 2007, from http://discuss.techinterview.org/default.asp?design.4.56828.31

van der Mei, R. D., Ehrlich, W. K., Reeser, P. K., & Francisco, J. P. (2000, March). A decision support system for tuning Web servers in distributed object oriented network architectures. *SIGMETRICS Perform. Eval. Rev.*, *27*(4), 57-62. Retrieved August 20, 2007, from http://doi.acm.org/10.1145/346000.346020

Want, R. (2004, October). The magic of RFID. *Queue*, *2*(7), 40-48. Retrieved August 20, 2007, from http://doi.acm.org/10.1145/1035594.1035619

WebSphere eXtended Deployment (XD). Retrieved August 20, 2007, from http://www-306.ibm.com/software/Webservers/appserv/extend/

Zhao, W., & Schulzrinne, H. (2005). *DotSlash: Handling Web hotspots at dynamic content Web sites*. Paper presented at the IEEE Global Internet Symposium. Retrieved March 1, 2007, from http://www.cs.columbia.edu/~zwb/paper/gi05.pdf.

Chapter XVI
Web-Based Decision Support System:
Concept and Issues

Rajib Goswami
Tezpur University, India

Pankaj Barua
Assam Agricultural University, India

ABSTRACT

This chapter elaborates the basic concepts underlying the development of Web-based decision support systems (DSS) with a discussion on the key concepts and technical issues. The utility of Web-based decision support system in enhancing communication and decision-making capability in a distributed environment or a multiple stakeholder process (MSP) has been explained through examples with diverse application from the real world. Further, the chapter introduces a Web-based decision support system developed by the authors for water resources management on basin scale and also some evolving concepts like mobile agent technology to meet the challenges and problems associated with traditional Web-based DSS. The authors hope that better understanding of the key issues and concepts can bring together analysts, modelers, and the end users to build Web-based DSS that are understandable, accessible and acceptable to all, be it corporate or business houses, environmental agencies or government organizations.

INTRODUCTION

A decision support system (DSS) can be defined as a computer-based tool used to support complex decision making and problem solving. Although this definition applies very well to decision making in many purely technical areas, it falls short of reflecting one extremely important aspect of the decision-making process, that is, the role of human factor.

One of the biggest challenges for DSS in facilitating access to information by a broad spectrum of stakeholders is that available information must directly address their concerns and information needs. Therefore, it is important to know how the information is obtained from and presented to nonspecialists; what information is or should be presented, and how the access to the information is managed. Another challenge is associated with enabling nontechnical professionals (decision makers may not be technical people, they may be politicians or bureaucrats) to obtain answers to their questions, especially in cases where both questions and responses need not be expressed in technical terms. The information presented to nonspecialists cannot substitute or hide the facts. This information must contain the same value as far as real consequences of options, but the form of this information should allow for straight forward description of impacts, perils, and benefits in layman terms.

The only possible method to adequately respond to these challenges has been the balanced and targeted usage of DSS technologies combined with organizational adjustments to the decision-making process, for example, where nontechnical professionals and interest groups also have the right to participate in the evaluation of options and their impacts. Web-based DSS can be used effectively to overcome this problem. Web-based DSS can help retrieve, analyze, and display structured data from large multidimensional or relational database, provide access to multimedia documents and unstructured data, and facilitate communication and decision making in distributed teams or multiple stakeholder processes (MSP) (Power & Kaparthi, 2002).

Unlike traditional DSS implemented on a single computer or on a network, where a user (decision maker or stakeholder) has an account, the development and usage of Web-based DSS faces many conceptual and technical challenges. In the case of DSS implemented on a single machine or in a network, the user has DSS available either through the software installed on the operating system or through a user interface to a remote application server. In the latter case, the capabilities of the user interface also rely strongly on the operating system. The access to resources extends beyond physical resources of the computer, such as disk space, memory, and printers. The user working with DSS in an interactive mode may also access and manipulate models built into the DSS and their parameters. The user may “activate” or “deactivate” certain components of the system model, change preferences, select display, or print alternatives. Data used by a DSS can be accessed and modified to allow users to explore various situations and scenarios. Results obtained by the user can be stored for further use; working sessions can be suspended and then started again without loosing information and data created during commenced sessions. In the case of DSS implemented via the Internet, the situation is significantly different; the user is accessing the Internet through a Web browser, which does not offer the same level of capabilities as an operating system. In order to offer users of Internet based DSS the same control and operational capabilities as those available to a user in traditional IT environment, the owner of a particular Web server has to make additional technical and developmental efforts. Technical difficulties and costs associated with providing the Internet users with advanced control mechanisms over DSS cause implementation of a Web-based DSS to proceed at a slower pace.

From the above discussion it is clear that WWW technologies have created new opportunities for DSS research and also for developing new innovative DSS. The field of Web-based DSS is a new one and more research is needed to design methodologies for implementing DSS using Web technology, to investigate linking of models and database technologies in Web environment, and to define role and effects of user’s involvement in design and development of Web-based DSS.

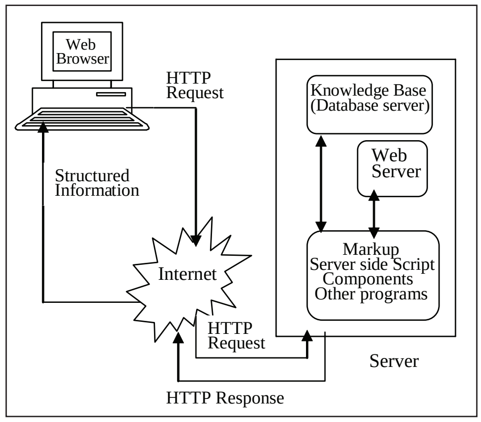
The present chapter contains discussion of the basic concepts underlying development of Web-based decision support systems. The chapter also reviews the key concept and technical issues. A prototype Web-based DSS developed by the author for analyzing consequences of various policies is also presented.

GENERAL ARCHITECTURE OF CONVENTIONAL WEB-BASED DSS

Web technologies can be used to implement any category or type of DSS. Web-based means the entire application is implemented using Web technologies; Web-enabled means key parts of an application like a database remain on a legacy system, but the application can be accessed from a Web-based component and displayed in a browser. Web-based DSS deliver decision support information or decision support tools to a manager or business analyst using a "thin-client" Web browser like Netscape Navigator or Internet Explorer that is accessing the global Internet or a corporate intranet. The computer server that is hosting the DSS application is linked to the user's computer by a network with the TCP/IP protocol. Web-based DSS can be communications-driven, data-driven, document-driven, knowledge-driven, model-driven, or a hybrid. (Nandalal & Simonovic, 2002).

A traditional decision support system does not provide flexible access to various DSS components in distributed enterprise-wide environment. As technology moved from local area network (LAN)-based or client-server based to Internet-based technologies; Web-based applications came up in a big way in decision-making system. Tools are developed using Web-accessible technology with knowledge-base built on relational database management system. Today most of the big organizations are creating data warehouses accessible by intranets and extranets based on

Figure 1. Traditional Web-based DSS



Internet technology. The main goal of a Web-based DSS is to retrieve a large volume of structured information using Web-based technologies. The common approach is to build such a DSS by using a commonly available relational database management system (RDBMS) for data storage and a common scripting language to forward the data in a suitable format to Web clients using the HTTP protocol.

By Web-based DSS generally we mean a system that provides structured decision support information to the decision makers using a "thin-client" (Web browser). In these systems a server hosts the DSS application and a Web browser may access this application. The database in this type of application is stored in a normal RDBMS. The updating of this database is again done through browser-based interface using hyper text markup language (HTML), extensible markup language (XML), scripting language, and so forth. Simply making a DSS accessible by using a Web browser does not mean a Web-based DSS. A true Web-based DSS is that which interacts with different widely distributed components of a decision-making system using Web technology and TCP/IP. The major task in developing such a system is to design the model of distributed decision-making components like (1) decision makers, (2) the interfaces, (3) the database components accessible through Internet, (4) the interactive processes for generating structured information, and (5) security.

BASIC TECHNOLOGIES USED IN WEB-BASED DSS

Web-based DSS are generally of multitier architecture. A user requests information through a Web browser (thin client) using hyper text transfer protocol (HTTP) to a Web server (Figure 1). The processing of the request is done at Web server using programs known as scripts. These scripts make connection to a database and process the request. The results are then sent back to the Web browser. These requests may be queries or to update a database. Sometimes results are again fed into another system for further processing to obtain structured information for decision making (e.g., distributed users may enter data to a database server using Web browser). Another group of distributed users over the Internet may process the database to generate some new kind of data. These data may be fed into systems like simulation software, system dynamics tools, or to artificial learning systems like artificial neural network (ANN), support vector machine (SVM), and so forth. The application programs are generally kept in a remote server and the user interface used is a Web-browser. The tools used for Web-based DSS are mainly HTML, extensible markup language (XML), common gateway interface (CGI) scripts, Web server application programming interface (API), Java applets and servlets, component-based tools like component object model (COM), Microsoft's distributed component object model (DCOM), Active-X components, and different scripting languages like Java scripts, Perl scripts, and so forth.

HTML and XML

The HTML is used to design the Web pages to be used as user interface with hypertext links. These interfaces may be parameter passing forms or a report. There are some technologies like the product WebDB of Oracle, where database query results/reports may be converted to HTML file by the system itself and the system may be used for generating HTML based data entry forms. Generally Java script, Perl script programs, and so forth are used for decision aid programs.

The extensible markup language (XML) is used for describing different data elements of a Web page. The XML codes are very effective in applications for databases. The XML tags converts Web pages into more structured documents. XML allows applications to process documents, data, and information more efficiently.

CGI Scripts

A CGI script is used to generate Web pages dynamically. Web browser invokes a program on the server that creates a new page. This new Web page may be based on either server data or it may process the results of a form submitted by a client through Web browser. The CGI programming is generally done in Perl, C, or AppleScript. CGI scripts or programs are initiated by the HTTP server every time a request for services is received. CGI programs are known as safe to run as the CGI program has very limited access to the server and a CGI program can be crashed without damaging the server in case of memory protected operating systems like UNIX. Thus CGI scripts are good tools for generating database reports as dynamic Web pages. The difference between a Web page written in static HTML and a Web page generated by a CGI program that takes no input is that the second one is done on the server side. There are other ways for Java programs to talk to servers like remote method invocation and servlets. Java is a simple, object-oriented, distributed, interpreted, robust, secure, architecture neutral, portable, high performance, multithreaded, and dynamic language. Therefore, Java applications can be used in diverse environments like Internet and so, Java language is a powerful tool for programmers to Internet-based DSS. JavaScript programming language is highly integrated with Web browser objects. JavaScripts are downloaded with HTML pages and Web browsers run these scripts when they are received.

Embedded Scripting

For developing an interactive Web-based system with dynamic output, embedded scripting languages are developed like the active server page (ASP) of Microsoft. With ASP, VBScript and JScript scripting languages are used. Similarly macromedia cold fusion is another embedded scripting tool that processes embedded cold fusion markup language (CFML).

Reusable Components

A reusable software component is another technology by which some controls and functionality can be added quickly to Web pages. "Activex controls" are some reusable software components of Microsoft like their earlier technologies of component object model (COM) and object linking and embedding (OLE). Using these components, one can embed interactive functionality to HTML pages. But these are mainly used with Microsoft's Web browser, Internet Explorer.

In Web-based DSS, keeping track of transactional information creates security and privacy issues. Another big problem is overload to the server. Of course, the load problem is directly related with the scalability of the computers, networking technologies, and software.

FEW EXAMPLES OF WEB-BASED DECISION SUPPORT SYSTEMS

A large number of organizations are presently working in Web-based distributed DSS. A simple Internet search related to the subject would result in an enormous number of "hits." This shows how widespread the notion of Web DSS has become. There are even dedicated Web sites for advertising and marketing Web-based DSS products (e.g., *www.DSSResources.com*).

Apart from the Web sites, the progress of Web-based DSS is well documented in literature dealing with the subject. The publications concerning the topic are reasonably huge and even a superficial review also exceeds the scope and space limitation of this chapter. Nevertheless a few Web-based DSS covering a diverse area of application is discussed in the following sections.

GDSI: A Web-based DSS for Effective Use of Clinical Practice Guidelines (Douglas, Rouse, Ko, & Niland, 2004)

Clinical practice guidelines (CPGs) are systematically developed healthcare recommendations. It is observed that though considerable effort and resources have been applied to CPG design, (development and deployment) the impact of CPGs on clinician behavior is not so consistent. Researchers from City of Hope National Medical Center's Division of Information Sciences developed a Web-based expert system to facilitate CPG utilization called graphical decision support interface (GDSI). The GDSI employs a relational database driven state machine architecture adapted from an instrument control system developed at National Medical Center's Division of Information Sciences. Here users enter information about an individual patient, and the system computes additional derived values and provides context-specific recommendations. It also provides reports on when and why clinicians intentionally deviate from guideline recommendations to assist with organizational benchmarking and CPG modification efforts.

The GDSI Web application was developed using Microsoft active server page (ASP) technology and deployed using Microsoft's Internet information server (IIS). The server (Web and database) hardware consists of two Dell 1650 servers with dual 1.4 GHz processors and 512 MB RAM. Here questions and each recommendation are modeled as a state in the GDSI state machine. A Microsoft SQL Server 2000 relational database is used to relate each state with its corresponding initialization, transition, status, and action rules. When a user responds, the questions and the order in which they were received were stored. If a user elects to revise an answer to an earlier question, the answers to all subsequent questions are deleted. This allows the GDSI to be used in a "what if" fashion to observe how the CPGs change in response to different clinical situations.

Long-Term Hydrological Impact Assessment (L-THIA) (Bernard, Choi, Harbor, & Pandey, 2003)

L-THIA was initially developed as a spreadsheet tool. Recently it has been integrated with a geographical information system (GIS). An interdisciplinary team at Purdue University has developed a Web-accessible version of L-THIA, overcoming several difficulties related to common data availability, ease-of-use, and widespread accessibility with interactive Web-based GIS. Developed by an interdisciplinary team, with input from users, the L-THIA Web DSS structure is comprised of a modeling system, a database system, and a graphical user interface, and includes special features for users with limited hydrology knowledge. In this DSS the primary decision-making context is land use change, which can include changes in agricultural uses in an area, or conversion to nonagricultural uses. The L-THIA model was rewritten in the "C" programming language and an executable L-THIA created to run within the Web-based L-THIA system with common gateway interface (CGI). In the L-THIA Web application, the user interacts with a Web interface written with HTML, Javascript, and Java to select the location of the site being analyzed and to provide information about the area of each land use and hydrologic soil group combination within the area of interest. A CGI script determines some values from the land use and hydrologic soil information provided by the user. Once the CGI scripts have generated the necessary information, L-THIA is run on the Web server using the rainfall and land use data and those CGI script determined values as input. The outputs are processed with CGI scripts, Javascript, and Java to provide Web-based tabular and graphical representations of the model output.

Stockpoint-Stockfinder (Stockfinder, 2006)

Stockfinder is a data-driven Web-based DSS available *at http://investor.stockpoint.com*. This system is related with another knowledge driven DSS called Investment Profile. This system helps investors to identify stocks based on criteria like price, type of industry, return, and so forth. Here a user answers a short questionnaire about personal financial goals, risk tolerance, and so forth. The DSS processes the answers and provides a list of possible investments that match the user's personal goal and budget constraints.

Bayer Corporation's Web-Based DSS Tool

Pharmaceutical Division of Bayer Corporation developed a Web-based DSS tool as a pioneering work in 1998. Through this system, managers at more than 500 cost centers create yearly budgets.

Other users for decision making access their planning information via Bayer's intranet. The system's backend database server was Oracle7.4 at one IBM RS6000 server machine.

EVOLVING CONCEPT: APPLICATION OF INTELLIGENT SOFTWARE AGENTS IN BUILDING WEB-BASED DSS

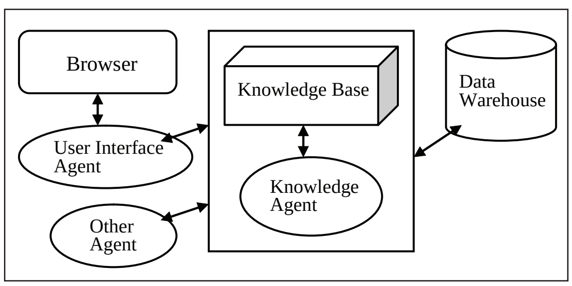
Web-based online expert systems are increasingly deployed for dynamic systems worldwide. Therefore, the delivery of services meet the changing needs of clients, becoming a major driver of development of many Web-based systems in business and government. As the Commission on Sustainable Development (CSD) report indicates, there have been many developments in this area since 1992: "The revolution in information technology has exploded into the 'new knowledge economy'," and "New information technologies are changing the ground rules for information flow in society." The importance of using new technologies to provide information access and enhance participation is emphasized in the CSD documents, and has led to new approaches to collaboration, and to increase participation in decision making. In these processes, interoperability can be either a key feature or a limiting factor (Miller, 2000). In the following sections some new approaches towards Web-based DSS is discussed.

A new approach towards Web-based DSS is by applying mobile agents and the distributed object technologies in constructing intelligent Web-based DSS. In these Web-based decision support systems, generally multiple mobile agents are used (Figure 2). These mobile agents are basically some independent computer programs that move around the network and act on behalf of the user or another entity (Pham & Ahmed, 1998). A mobile agent is a software entity that can travel from host to host to access resources stored at each visited host and interact with other agents. Mobile agents are not confined to the system where it begins execution. Thus the concept of mobile agents is useful for constructing distributed systems of heterogeneous environment. Distributed DSS requires frequent remote execution of decision models and needs access to the source data on remote servers. Thus distributed DSS can take advantage of this new concept. In most of the framework suggested in this approach for Web-based DSS, multiple agents are used. Some of the agents may be stationary and some are mobile agents.

Studies show that the emerging mobile agent technology can better meet the challenges and problems raised earlier compare to the traditional approaches (Glitho, Olougouna, & Pierre, 2004; Pham & Ahmed, 1998; Wang, Chen, Y, & Liu, 2004). The mobility gives mobile agents the advantage to reduce network load and overcome network latency by moving computations to the data sources. The ability to operate asynchronously and independently from the original process that created them makes mobile agents highly robust and fault-tolerant in building enterprise-wide network systems including DSS.

As the expectations from the design of a Web-based DSS are much greater than the design of a traditional DSS, a framework called flexible Web-based DSS generator (FWDSSG) utilizing software agents to enhance the functionality of existing DSS was proposed (Samaras, Pitoura, &

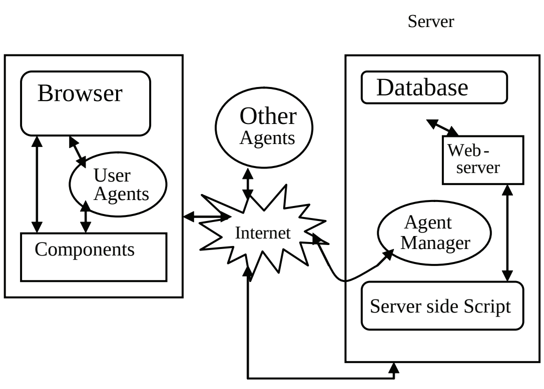
Figure 2. Agents in Web-based DSS



Evripidou, 1999). Multiple agents are employed in the framework to enable flexibility of a Web-based DSS Generator. These agents are:

1. User interface agents to perform systematic presentation for the system.
2. User agent (stationary and intelligent) to represent the user on the client computer to coordinate the agent manager on the server computer. The user requests the agent manager to execute the tasks and obtains the results back from the agent manager.
3. DSS component agent (stationary and interactive) to present the DSS component resources located in various Web sites.
4. DSS components mapping agents (mobile and Intelligent) to work with the DSS components agents to complete the assigned tasks. The agent manager located on the server computer creates various task agents to carry out the tasks after the agent manager has received the request from the user agent. The relationship between the agent manager and the mapping agents is that of master-slave. The agent manager creates and allocates the mapping agents to remote sites to carry out some specific tasks. The agent manager receives the feedback from the mapping agents and sends the results to the user agent.
5. DSS scenario agent (stationary and interactive) to represents the scenario used by the user.

Figure 3. Agent-based DSS architecture



Agents are mainly used to retrieve information through communication with other agents (Figure 3). For agent-to-agent communication a protocol called agent communication language (ACL) is used. The data can be passed back and forth in the form of XML.

Helen and Jia (2003) suggest an agent-enabled communication mechanisms in designing decentralized distributed decision support enterprise where each decision-support service center is self-contained and can function independently, and at the same time are all connected in an Internet-like enterprise-wide network environment for cooperative problem solving. They propose to implement a stationary master agent called *SrvAgent* to dynamically create and delegate decision models as tasks to its mobile subagents. This master agent is not movable, and always stays and is continuously running on the agent server to perform other tasks in parallel with the slave agent. However, the slave agents are mobile, and are created at runtime to really travel to the remote host.

In a framework proposed by Yang, Zhu, and Yan De (2004), the hierarchy of Web-based decision support system intelligent agents was divided into different layers called method layer, model layer, and application intelligent layer. Intelligent agents in this case were used on a Web-based DSS, based on data warehouses and data mining technology and concept. The method layer includes various quantitative analysis tools, including the linear regression analysis, time series analysis, neural network analysis, statistics analysis, and so forth. This layer includes some special economic analysis tool also. The data processing of data warehouses included three stages: data preparation and processing (DPP); data presentation (DP);

and data analysis (DA). The main tasks in the first two stages were to prepare data source to serve for the decision support system. The main task in the data analysis stage is to provide tools for decision system, in this way it can also be defined as analysis layer in data warehouse system. The decision support system analysis layer of this framework consists of online analytical processing (OLAP) query analysis tool, data mining tool, and DSS analysis forecast tool, each focused on a certain field and used by different user. The data warehouse system with these tools could process the huge precious information hidden in data warehouse. The model layer calls automatically and intelligently various module-constructing tools and try to integrate modules for the application layer to use. The application layer tries to gather various data in data warehouses and module in the model layer, and constitute the quantitative analysis report.

Another framework has been suggested based on agent and data warehouse, called Web-oriented warfare command DSS (Wang et al., 2005). In this framework, agents were classified as interface agent, assignment agent, communication agents, knowledge agents, and data warehouse agents. Interface agents were the interface between the front-end Web browsers and the knowledge base. Assignment agents talk to knowledge base and the model base and interact with communication agents and knowledge agents. It basically assigns missions for other agents. Communication agents were used for harmonizing user and agents. User actions and the solutions were kept in a knowledge base. As intelligent agents can learn by observation so it can maintain the knowledge base. The knowledge agents are mainly used for this purpose. Data warehouse agents acted as an agent of data warehouses that was capable of data mining.

ISSUES IN WEB-BASED DECISION SUPPORT SYSTEM (DSS)

Communication is a major bottleneck in implementation of DSS in the Internet environment. In most of the developing countries across the world, client computers are often connected to the Internet through low end communication and or telephone lines with relatively low transmission rates. Therefore, time needed to perform computation on server side can be relatively long. Significant investments by government and international agencies are required to provide this necessary infrastructure.

Another problem associated with Web-based DSS is integration of functional and user friendly model. Implement of DSS in the Internet medium imposes specific requirements and limitations for the developers. Many existing, proven, and user friendly models cannot be directly integrated for building Web-based DSS. Additional efforts and programming support are required to overcome this problem.

Lack of information is another important issue for development of Web-based DSS particularly in developing countries. Although significant progresses have been made in the recent years with respect to collection, storage, and usage of data including via remote sensing, availability of reliable, and consistent data are still difficult. In some cases data may be available but not shared. Authors of this chapter experienced the depth and seriousness of the problem while developing the prototype DSS for Brahmaputra Basin. There was not enough information and a trustworthy model for building the Web-based DSS. The river flows through four countries and some places within its basin have boundary conflict. Gathering information on the basin was very difficult and at times was nearly impossible. The problem was addressed by using a system dynamics model (black box kind of approach) that requires more of knowledge rather than data. In the followings sections the efforts by the authors in developing a Web-based DSS for water resources management of a large river system is presented.

A PROTOTYPE WEB-BASED DSS FOR A LARGE RIVER SYSTEM

The decision-making process concerning any large river system (those discharging 10,000 M^3/S or more) is a multistakeholder process (MSP). Growing demand for water, restricted availability, and resulting conflict over river water resources has generated considerable interest among the scientists, planners, and water professionals for its efficient management. Since rivers are the most potential source of fresh water, management planning for water resources is generally carried out in river basin scale. Complexity of the decision-making process concerning water resources management on basin scale calls for use of sophisticated yet easy to use decision support (DSS) tools for processing vast amount of diverse information. Since river basin decision-making process is a MSP concerning stakeholders spread over a number of states and countries, Web access to such a DSS would enhance its applicability through better exchange and sharing of information.

The authors were associated with designing a generic Web-based DSS framework on river basin scale named **BARISS**: **B**asin **A**nd **R**iver **I**nformation and **S**imulation **S**ystem. The main motivation behind development of **BARISS** was to promote access to information related to important decisions associated with utilization of water resources of any large river system. The water availability and consumption pattern in any large river system is dynamic over a certain time period; hence assessment of future water availability and consumption is required. Scenarios of future water availability, consumption and so forth are important constraints in decision making which needs to be visualized so as to evolve an effective strategy/planning of water resources consumption and distribution keeping availability in mind. This requires understanding the behavior pattern of future use, therefore system dynamics (SD) has been considered as an appropriate tool for building the model base of **BARISS.** To assess and simulate water availability on river basin scale, adequate information on water availability and consumption like flow rate of river, rainfall, ground water level, population, industry and so forth are needed. Such information generated in any river basin is usually from a diverse set of sources and incorporate diverse areas. In most countries of the world including India, such information is collected, viewed, stored, and (re)distributed by central and provincial governments, scientific organization, and their regional sub divisions. The information thus collected and stored by these organizations is hard to get. Because of the large volume, even the control over the information gets lost within the organization. At a time of increasing river management consciousness, it can not be sufficient enough just to generate information; it is also needed to properly store and analyze these and offer it for scientific analysis and policy planning. Relational database management system (RDBMS) has that capability and therefore the information system component of **BARISS** has been built using the RDBMS concept with Web-based user interface for easy sharing.

The **BARISS** framework has been tested using case study approach in a real world situation. The river system for the case study has been selected based on:

- Possibility of researcher's association and understanding of a particular river system.

- A river system for which similar effort on this line has not been done earlier
- A large and international river system.
- Concerns a controversial issue involving conflicting interest between states, countries, or stake holders.

The search led to the Brahmaputra River Basin being selected as a case study for the following reasons:

- Being from the same region, researcher's understanding of the river system is satisfactory
- The river is large and international in character (one of the largest river system of the world and flows through 4four countries)
- No earlier studies on similar line has been carried out for this river
- Government of India has an ambitious project for inter linking of all the rivers of India, which has been of concern to the neighboring country of Bangladesh, and also to the general public, and environmental scientists of the Brahmaputra Basin.

The test application of Web-based DSS framework **BARISS** on Brahmaputra has been named as **BRISS** (**B**rahmaputra **R**iver **I**nformation and **S**imulation **S**ystem).In the following sections BRISS has been discussed in brief.

BRISS

The BARISS framework has been tested on the Brahmaputra River and named as BRISS (Brahmaputra River Information and Simulation system). BRISS has been presented as a Web site. It consists of a river basin information system built on the RDBMS concept and river basin simulation model using the system dynamics theory. In the present set up under study the river basin Database has been built using "Oracle 8i" database. The Internet access to the database has been provided using "Oracle Web DB" (an ORACLE Product). SD software STELLA has been used for building the river basin SD model. Considering that in the study area a client's computer would be connected to the Internet through low end communication (telephone lines with poor bandwidth), therefore

Figure 4. Index page of BRISS

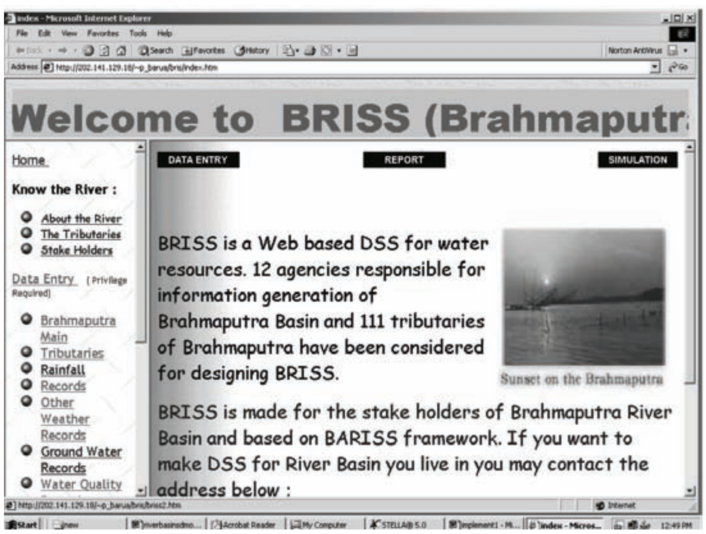
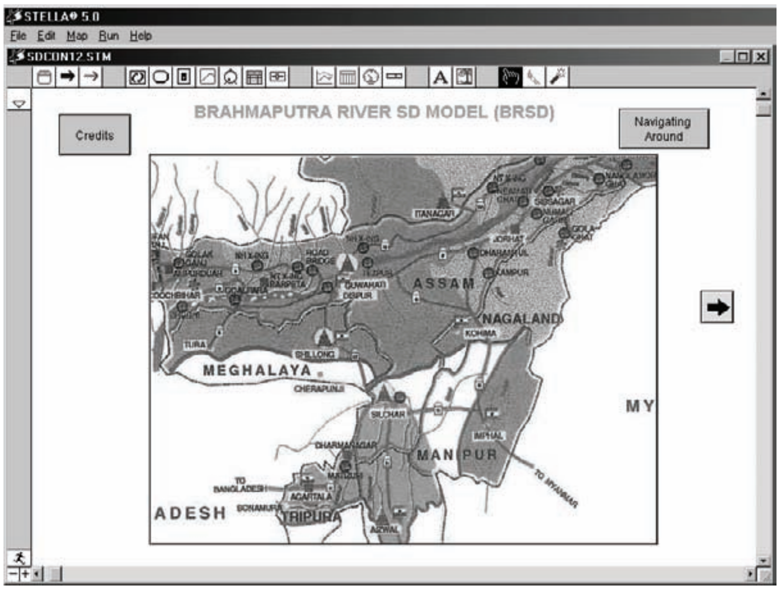


Figure 5. Welcome menu of BRSD



proposed implementation of BRISS has been through thick client and thin server concept.

When any client accesses BRISS through a Web browser, the screen as shown in Figure 4 will be displayed on the client's computer screen.

The Model Base of BRISS

The model base of BRISS comprise of a system dynamics model to assess the dynamics of water availability and consumption for the basin. System dynamics software STELLA has been used to build the Brahmaputra River system dynamics (BRSD) model. The simulation menu of BRISS leads to BRSD (Figure5). For simulation a user may query for data through BRISS and use the data to run the simulation.

While building a BRSD model, about 285 variables have been used to represent five sectors, namely, population, irrigation, industry, basin water, and sustainability. The part of the river basin passing through the state of Assam in India has been considered for the model. Thirty years of data (1971-2001) of the basin has been used for model validation. Both qualitative and quantitative validation has been done with good matching of data and trend. Sensitivity of the river basin water status to some selected parameters has been assessed through sensitivity analysis. The sensitivity analysis BRSD model has been used for scenario generation for the year 2025 and 2050. Scenarios generated have been compared and the availability status of water for the basin has been presented.

The application of **BARISS** revealed that apart from being a policy tool, it can also assist in participatory decision making, public opinion building, and conflict resolution in any large river system. In keeping pace with the developments in information technology associated with Internet changes are being incorporated in **BRISS**.

CONCLUSION

Use of Web-based DSS and Web-enabled DSS is increasingly being used by corporate and business houses, environmental agencies, and government organizations for decision making. These are primarily done in two ways: first, expansion of existing Web sites making room for free, unlimited and user friendly information, data, and literature along with downloadable models; second, creation of relatively modern Web-based DSS. The second approach is getting momentum in the last few years. Lots of new techniques are being used, some of which have been discussed earlier in this chapter. Development of new techniques is definitely a technical challenge. In spite of rapid progress made in the "soft" side of Web-based DSS development there are areas for improvement. One of the major difficulties regarding application of Web-based DSS is its inability to effectively communicate with a broad circle of users and stake holders. Significant improvements can be brought about when analysts and modelers work together with the end users to build Web-based DSS as understandable, accessible, and acceptable to all parties involved.

Another point of concern is that Web-based DSS as a tool for resolution of environmental dispute and conflicts concerning usage and sharing of natural resources is still rare. International organizations may use Web-based DSS as an independent, unbiased, and objective tool capable of addressing controversial issues (like global warming, water sharing, carbon load, etc.) arising between two or more countries in order to establish a neutral basis for communication and discussion to help resolve the conflict.

FUTURE RESEARCH DIRECTIONS

The future modeling tools and support systems for business houses, environmental agencies, and government organizations for decision making must include the use of intelligent technologies like machine intelligence, neural nets, genetic algorithms, and soft computing modeling like fuzzy logic, approximate reasoning, and probabilistic modeling. The next generation of decision support system for any environmental problem (like a river system management) should be adaptable with virtual teamwork in different places and in different time zones, decision support systems on mobile devices, and should be able to access multilayer networks (Internet(s), intranets).

Scientists around the world are working on ways to revolutionize the Internet by an interdisciplinary project collectively known as the Semantic Web. Researchers are working towards developing standards, protocols, and technologies that will advance the development of a more meaning-oriented Web. The aim of the Semantic Web efforts is to be able to find and access Web sites and Web resources not by keywords, but by descriptions of their contents and capabilities.

As the move to a smarter Web is taking far too long, some organizations are looking into other available technologies like agent technology to meet their needs. As a mobile agent can travel from one host to another caring its code, therefore, autonomous execution of programs on remote hosts is possible. Presently, a mobile agent-enabled approach in constructing large-scale distributed network services and its integration with the distributed objects model for decision support has been taken up. But future extensions in this line are to be done to improve flexibility in terms of network management and decision support. As the significance of the distributed DSS will grow in the future, research is needed in the design and empirical assessment of multiple stakeholder-based DSS. There is scope of development of algorithms for improving the tools for combinatorial pattern matching or similarity searching, handling of semistructured information, and so forth.

Scientists at several computer companies and institutes like Massachusetts Institute of Technology, the University of Chicago, Kyoto University, and Stanford University, are independently collaborating on the Semantic Web project, led by the World Wide Web Consortium (W3C). As the full open Semantic Web is projected as the future of Web technology, detailed analysis must be done to make the Semantic Web technology progressively profitable for development of Web-based decision support system.

REFERENCES

Bernard A. E., Choi, J., Harbor, J., & Pandey, S. (2003). Web-based DSS for hydrologic impact evaluation of small watershed land use changes. *Computers and Electronics in Agriculture, 39*, 241-249.

Douglas, C. S., Rouse, L., Ko, D., & Niland, J.C. (2004).GDSI: A Web-based decision support system to facilitate the efficient and effective use of clinical practice guidelines**.** In *Proceedings of the 37th Hawaii International Conference on System Sciences – 2004*. IEEE.

Glitho, R. H., Olougouna, E., & Pierre, S. (2002, January/February). Mobile agents and their use for information retrieval: A brief overview and an elaborate case study. *IEEE Network, 16*(1), 34-41.

Helen, S. D., & Jia, X. (2003). A framework for Web-based intelligent decision support enterprise. In *Proceedings of the 23rd International Conference on Distributed Computing Systems Workshops (ICDCSW'03)*. IEEE Computer Society.

Miller, P. (2000, June 21). Interoperability. What is it and why should I want it? *Ariadne, 24*. Retrieved September 10, 2006, from http://www.ariadne.ac.uk/issue24/interoperability/

Nandalal, K. D. W., & Simonovic, S. P. (Eds.). (2002, October). *State of the art report on system analysis methods for resolution of conflicts in water resources management* (Report Prepared for Division of Water Sciences). UNESCO.

Pham, V. A., & Ahmed, K. (1998, July). Mobile software agents: An overview. *IEEE Communication Magazine, 36*(7), 26-37.

Power, D. J., & Kaparthi, S. (2002, December). Building Web-based decision support systems. *Studies in Informatics and Control, 11*(4).

Samaras, G., Pitoura, E., & Evripidou, P. (1999, March). *Software models for wireless and mobile computing: Survey and case study* (Tech. Rep. TR-99-5). University of Cyprus.

Stock-finder. (2006). *Stockpoint-Stockfinder*. Retrieved September 10, 2006, from http://investor.stockpoint.com

Wang, X., Chen, Y., & Liu, Y. (2005). Web-oriented warfare command decision support system based on agent and data warehouse. In *Proceedings of the 2005 International Conference on Cyberworlds (CW'05)*. IEEE Computer Society.

Yang, W., Zhu, W., & Yan De, Y. L.(2004). The prototype research of a Web-based DSS intelligent agent over data warehouse. In *Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence* (WI'2004).

ADDITIONAL READING

AbouRizk, S. M., & Hajjar, D. (1998) A framework for applying simulation in construction. *Canadian Journal of Civil Engineering, 25*(3), 604-617.

Alter, S. (1981). *Transforming DSS jargon into principles for DSS success*. Paper presented at Transactions of the International Decision Support Conference.

Angehrn, A. A. (1993). Computers that criticize you: Stimulus-based decision support systems. *Interfaces, 23*, 3-16.

Angehrn, A. A., & Jelassi, M. T. (1994). DSS research and practice in perspective. *Decision Support Systems*, *12*, 267-275.

Balasubramaniam, R., & Kannan, M. (2001). Integrating group decision and negotiation support systems with work processes. In *Proceedings of the 34th Hawaii International Conference on System Sciences*, Hawaii.

Bordini, R. H. (2005) *Multi-agent programming: Languages, platforms, and applications*. Springer.

Du, H.S., Jia, X., & Chen, K. (2002). An integrated framework for heterogeneous decision support systems. In *Proceedings of the IASTED on Networks, Parallel and Distributed Processing and Applications*, *2002* (pp. 449-454).

Giorgini, P., & Henderson, B. (2005). *Agent-oriented methodologies.* Hershey, PA: IGI Global, Inc.

Jain, L. C., Baba, N., & Howlett, R. J. (2001) *Knowledge-based intelligent information engineering & allied technologies.* Paper presented at the KES'2001. IOS Press.

Jeffrey, X. Y. (2004) *Advanced Web technologies and applications.* Paper presented at the 6th Asia-Pacific Web Conference, APWeb. Springer.

Khosrowpour, M. (2000). *Challenges of information technology management in the 21st century.* Hershey, PA: IGI Global, Inc.

Kimball, R., & Merz, R. (2000). *The data Webhouse toolkit: Building the Web-enabled data warehouse.* John Wiley & Sons Inc.

Luck, M., Ashri, R., & D'Inverno, M. (2004). *Agent-based software development.* Norwood, MA: Artech House, Inc.

Müller, J. P. (1996). *The design of intelligent agents: A layered approach*. Springer.

Norman, T. J. J., Faratin, N. R. P., & Mamdani, E. H. (1996, August). Designing and implementing a multi-agent architecture for business process management. *Intelligent Agents III: Proceedings of the Third International Workshop on Agent Theories, Architectures, and Languages* (LNAI 1192, pp. 261-275). Berlin: Springer Verlag.

Padgham, L., & Winikoff, M. (2004). *Developing intelligent agent systems: A practical guide.* John Wiley & Sons, Inc.

Plekhanova, V. (2002). *Intelligent agent software engineering.* Hershey, PA: IGI Global, Inc.

Power, D. J. (2002) *Decision support systems concepts and resources for managers.* Westport, CT: QUORUM BOOKS.

Power, D. J. (2004). *Decision support systems: Frequently asked questions.* iUniverse Publishing.

Sauter, V. L. (1996) *Decision support systems: An applied managerial approach*. John Wiley & Sons, Inc.

Turk, A., Farnschläder, F., Silberhorn, H., Stoyan, H., & Kreutter, S. (2000, September). The Web consultant: A flexible framework for dynamic Web application. In *Proceedings of the Electronic Commerce and Web Technologies, First International Conference, ECWeb 2000,* London, UK, (pp. 13-24). Springer.

Woolbridge, M. (1996, August) Agents as a Rorschach test: A response to Franklin and Graesser. *Intelligent Agents Iii*, *Agent Theories, Architectures, and Languages* (LNCS 1193, pp. 47-48). Berlin: Springer Verlag.

Section III

Applications in Manufacturing and Production Management

Chapter XVII
Independent Component Analysis and its Applications to Manufacturing Problems

Xian-Chuan Yu
Bejing Normal University, China

Ting Zhang
Bejing Normal University, China

Li-Bao Zhang
Bejing Normal University, China

Hui He
Bejing Normal University, China

Wei Zou
Bejing Normal University, China

Meng Yang
Bejing Normal University, China

ABSTRACT

Independent component analysis (ICA) is a statistical method for transforming an observed multidimensional random vector into components that are as independent as possible. In this chapter, we introduce the background information, the theory of ICA, and present several common algorithms such as fast ICA, kernel ICA, and constrained ICA. It is first applied to mineral resources prediction and remote sensing imagery, while traditional methods cannot satisfy the complexity of the spatial data (prospecting geochemistry data, remote sensing data, etc.). In application cases, ICA is applied to analyze the spatial data in some districts of China. The result shows that some independent elements accord with the practical distribution better than conventional methods. Moreover, ICA can get rid of the various kinds of correlations in remote sensing imagery effectively and improve the classification accuracy. However, this method also has some limitations. At last, we list the future research directions of our work.

INTRODUCTION

Independent component analysis (ICA), which is developed from blind source separation (BSS), is the identification and separation of mixtures of sources with little prior information. It is also used in the feature extraction.

In the beginning, the basic ICA data model is linear and without noise. The model is often estimated by choosing an appropriate objective function then maximizing or minimizing it by the optimization algorithm. The objective function decides the statistic characters such as asymptotic variance, robustness, and so on. The optimization algorithm decides the algorithm's characteristics, such as the convergence speed, the numeric stability, and so forth. In the actual research, objective functions, those most extensively used, are likelihood and network entropy, negentropy, mutual information, Kullback-Leibler divergence, and high order cumulation.

In data mining and pattern recognition, independent component analysis is very useful and gradually becomes a hot problem in signal processing. Now ICA is widely applied in biomedicine signal processing, sound signal separation, communication, error diagnose, feature extraction, financial time sequence analysis, data mining, astronomy, and so forth. For example, ICA could provide solutions to denoise electroencephalogram (EEG), magneto encephalography (MEG), or functional magnetic resonance imaging (FMRI) signals for medical applications (Makeig, Bell, Jung, & Sejnowski, 1996; Mckeown, Jung, Makeig, et al., 1998; Vigário, Jousmaki, Hamalainen, Hari, & Oja, 1998), it could separate mixed sound signals (Li & Sejnowski, 1994), and reveal the factors to reduce the investment risk from the economic time sequence data (Mlroiu, Kiviluoto, & Oja, 2000). ICA could also be applications for code division multiple access (CDMA) communications (Ristaniemi, Raju, & Karhunen, 2002). At the same time, it has been used to extract features from natural images (Olshausen & Field, 1996; van Hateren & van der Schaaf, 1998). Recently, ICA has been explored for the feature extraction from multicolored and solid 3D images (Hoyer & Hyvrinen, 2000). The mixture model has been developed to extend to solving general unsupervised classification and data mining problems beside the simple ICA, which has been used as a tool for unsupervised analysis of hyper spectral images.

In manufacturing fields, ICA is also very important. For example, in voice control computer, ICA can help recognize the correct speech commands in noise environment and other signals if the locations of the sources and the receive devices are unknown.

In this chapter, we will introduce the main theory and algorithms of ICA, then present its two novel applications. One is mineral resources prediction of geology and the other is remote sensing imagery information extraction of geography.

BACKGROUND

For blind source separation, several methods have been proposed, such as principle component analysis (PCA), factor analysis (FA), projection pursuit, redundancy reduction, and blind deconvolution. PCA and FA belong to the second order separation method; however, the actual signification is usually explicit. PCA assumes the original data satisfying with the Gauss distribution emphasis to reduce the dimensions, so the compute precision is not very high. And the eigenvectors after PCA are orthonormal which do not accord with a great deal of natural information in the real world. The high order separation methods conclude projection pursuit, redundancy reduction and blind deconvolution. Whereas they also request that the data obey Gauss distribution.

ICA inherits the advantages of the theories above and modifies some disadvantages. For example, ICA assumes that the original data are non-Gauss distribution, which is also a redundancy reduction method. Meanwhile, the eigenvectors that got by ICA are not orthonormal but independent with each other. In fact, ICA can be regarded as factor analysis to non-Gauss original data so dimension reduction is just its minor target. Therefore, PCA can become the preprocessing step in ICA.

In 1986, a recursive neural network model and a learning algorithm based on Hebb learning rule was proposed by Herault and Jutten. It could separate several unknown independent sources by a feedback method. Later, Comon (1994) elaborated the concept of independent component analysis and proposed the cost function related to the approximate minimization of mutual information between the sensors. Many complicated algorithms based on cumulant were designed, which make the three order nonlinear characteristic widely spread.

Infomax of ICA was proposed by Tony Bell (Bell & Sejnowski, 1995). Mari realized that Infomax ICA can be developed by using natural grads. Cardoso (1997) developed Infomax ICA. Many practical problems were solved after that and the algorithm was greatly improved in computing scale. Working on the previous Infomax ICA just adapted to sub-gauss sources, Te-Won Lee and Mark Girolami developed the extent version of Infomax ICA and the algorithm can separate super- and sub-Gaussian (Lee, Girolami, & Sejnowski, 1999) sources. Their adaptive methods are more reasonable from a neural processing perspective than the cumulant-based cost functions proposed by Common.

Other algorithms for performing ICA have been proposed. Maximum likelihood estimation (MLE) approaches to ICA were first proposed by Gaeta and Lacoume and elaborated by Pearlmutter and Parru. Nonlinear PCA algorithms, which have been developed by Karhunen and Joutsensalo (1995), can also be viewed from the Infomax principle since they approximately minimize the mutual information of the network outputs. Francis R. Bach and Michael I. Jordan (2002) proposed kernal independent component analysis (KICA) at the international ICA meeting held in 2002. A new objective function based on canonical correlations in a reproducing kernel Hilbert space and two new algorithms based on the objective function are presented.

BASIC THEORY OF ICA

Independent component analysis was motivated from a blind source separation problem called *cocktail-party* problem. Imagine that two people are speaking in a room simultaneously. You put two microphones in different location to record the mixture of the two sound signals. Then you can get two time signals, which can be denoted by $x_1(t)$ and $x_2(t)$, where t is the time index. The speech signals of the two speakers in time t can be denoted by $s_1(t)$ and $s_2(t)$. If we do not consider the situation of noise, we can express the mixed signals as follows:

$$\begin{aligned} x_1(t) &= a_{11}s_1 + a_{12}s_2 \\ x_2(t) &= a_{21}s_1 + a_{22}s_2 \end{aligned} \tag{1}$$

where $a_{11}, a_{12}, a_{21}, a_{22}$ are coefficients which depend on the locations of microphones. So if we only know the mixed signals $x_1(t)$ and $x_2(t)$, while the original signals $s_1(t)$, $s_2(t)$ and the parameters $a_{11}, a_{12}, a_{21}, a_{22}$ are unknown, how to separate the original signal and the parameters is the problem that we research. For the random time index t, assuming that original signals are statistical independent, ICA can deal with this problem.

The estimation of the independent component analysis data model is usually performed by formulating an objective function and then minimizing or maximizing it (Hyvärinen, 2001). Therefore, the properties of the ICA method depend on the objective function and optimization algorithm. Generally, the statistical properties (e.g., consistency, asymptotic, variance, and robustness) depend on the choice of the objective function and the algorithmic properties (e.g., convergence speed, memory requirements, and numerical stability) depend on the optimization algorithm. These two classes of properties are independent in the sense that different optimization methods can be used to optimize a single objective function, and a single optimization method may be used to optimize different objective functions.

ICA has resemblance with PCA, such as they are both used in feature extraction and dimension reduction. However, PCA works for the second order moment of the data to produce uncorrelated components while ICA strives to generate components as independent as possible. Therefore, PCA can be regarded as the preprocessing program of ICA.

Assumptions of ICA Model

Assume that there is an M-dimensional zero mean vector s(t)=[s1 (t),… ,sM (t)]T, whose components are mutually independent. The vector s(t) corresponds to M independent scalar valued source signal si(t). We can write the multivariate p.d.f. of the vector as the product of marginal independent distributions.

$$p(s) = \prod_{i}^{M} p_i(s_i) \tag{2}$$

A data vector x(t)=[x1(t),…,x2(t)]T is observe at each point *t*, such that:

$$x(t) = As(t) \tag{3}$$

where A is an N*M scalar matrix. The goal of ICA is to find a linear transformation W of the dependent sensor signals *x* that makes the outputs as independent as possible:

$$u(t) = Wx(t) = WAs(t) \tag{4}$$

where *u* is an estimate of the sources.

For the linear mixing and immixing model, we adopt the following assumptions:

1. The number of sensors is greater than or equal to the number of sources: N≥M.
2. The source s (t) is instant mutually independent every time.
3. At most one source is normally distributed.
4. The time difference of sensor noise or only low additive noise signals are neglected in the model.

For the above four assumptions, Assumption 1 is needed to make a full rand matrix. Assumption 2 is the basis of ICA. For Assumption 3, the immixing of two Gaussian sources is ill posed when the sources are white random processes. Assumption 4 is necessary to satisfy the Informax condition. It is difficult to estimate the non-noise ICA model, so in order to get more satisfying results, we mainly use non-noise at present.

Because S and A are unknown, any scalar quantity change of s_i can be counteracted by the scalar quantity change of the matrix A.

$$x = \sum_i \left(\frac{1}{\alpha_i} a_i\right)(s_i \alpha_i) \tag{5}$$

In a general way, we can assume that the independent component has a fixed variance 1 because they are random variables.

$$E\{s_i^2\} = 1 \tag{6}$$

Adding minus to independent component will not influent the data model. Moreover, we cannot confirm the order of the independent component since S and A are unknown.

Selection of Objective Function

Likelihood and Network Entropy

It is possible to estimate the ICA model by a maximum likelihood method. Denoted by $W=(w_1,...,w_m)^T$ the matrix A^{-1}, the log-likelihood takes the form:

$$\boldsymbol{L}=\sum_{t=1}^{T}\sum_{i=1}^{m}\log f_i(w_i^T x(t)+T\ln|\det W|) \qquad (7)$$

where f_i is the density function of the s_i.

Another related objective function was derived from a neural network viewpoint. Assume that x is the input to the neural network whose outputs are of the form $g_i(w_i^T x)$, where the gi is some nonlinear scalar function, and the wi is the weight vector of the neurons. Because entropy can be the measure of non-Gaussian, ICA is to maximize the entropy of the outputs:

$$L=H(g_1(w_1^T x),\cdots,g_n(w_n^T x)) \qquad (8)$$

When the nonlinearities gi used in the neural network are chosen as the cumulative distribution function corresponding to densities fi, ie.,gi'(.)=fi(.), the principle of network entropy maximization, or "Infomax," is equivalent to maximum likelihood estimation. The advantage of the maximum likelihood approach is under some regularity conditions, it is asymptotically efficient. However, there are also some drawbacks. First, this approach requires the knowledge of the probability densities of the independent components. A second drawback is that the maximum likelihood solution may be very sensitive to outliers, while robustness against outliers is an important property of any estimator.

Nengentropy and Approximations of Negentropy

Negentropy is defined as follows:

$$J(y)=H(y_{gauss})-H(y) \qquad (9)$$

where y_{gauss} is a Gaussian random vector of the same covariance matrix as *y*. Negentropy, or negative normalized entropy, is always nonnegative, and is zero if and only if *y* has a Gaussian distribution. The advantage of using negentropy, or, equivalently, differential entropy, as a measure of non-Gaussianity is that it is well justified by statistical theory. Actually, in some sense, negentropy is the optimal estimator of non-Gaussianity, as far as statistical properties are concerned. The problem in using negentropy is, however, that it is exclusively difficult to compute. Estimating negentropy using the definition would require an estimate of the pdf. Therefore, this objective function remains mainly a theoretical one. In practice, some approximations have to be used. The classical method of approximating negentropy is using higher-order moments, for example as follows:

$$J(y)\approx\frac{1}{12}E\left\{y^3\right\}^2+\frac{1}{48}kurt(y)^2 \qquad (10)$$

However, the validity of such approximations may be rather limited. These approximations were based on the maximum-entropy principle. In general we obtain the following approximation:

$$J(y)\approx\sum_{i=1}^{p}k_i[E\left\{G_i(y)\right\}-E\left\{G_i(\mathrm{v})^2\right\} \qquad (11)$$

where k_i is some positive constant, and v is a Gaussian variable of zero mean and unit variance (i.e., standarlized). But the point here is that by choosing *G* wisely, one obtains approximations of negentropy that are much better. Particularly, choosing *G* that does not grow too fast, one obtains more robust estimators. The following choices of G have proved very useful:

$$G_1(u) = \frac{1}{a_1} \log \cosh a_1 u, \tag{12}$$

where 1≤α1≤2 is some suitable constant.

$$G_2(u) = -\exp[(-u^2 / 2)] \tag{13}$$

Mutual Information and Kullback-Leibler Divergence

Inspired by information theory, the most satisfying object function is mutual information. Mutual information can be interpreted by using the interpretation of entropies code length. We define the mutual information I between m random variables, y_i, i=1…..m, as follows:

$$I(y_1, y_2, \cdots, y_m) = \sum_{i=1}^{m} H(y_i) - H(y) \tag{14}$$

Mutual information is a natural measure of the dependence between random variables. It is always non-negative and zero if and only if the variables are independent statistically. A natural method of ICA approximation is to make the mutual information of s_i to the least. An important property of mutual information is that we have for an invertible linear transformation $y = Wx$:

$$I(y_1, y_2, \cdots, y_n) = \sum_i H(y_i) - H(x) - \log|\det W| \tag{15}$$

In fact, mutual information is equivalent to the well-known Kullback-Leibler divergence between the joint density of $y_1, y_2, ..., y_m$ and the product of its marginal densities (Amari, Cichocki, & Yang, 1996; Pham, 2002).

$$D(p_1, p_2) = \int p_1(y) \log \frac{p_1(y)}{p_2(y)} dy \tag{16}$$

The connection to the Kullback-Leibler divergence also shows the close connection between minimizing mutual information and maximizing likelihood. Actually, the likelihood can be presented as a Kullback-Leibler distance between the observed density and the factorized density assumed in the model. Both of those methods minimize the Kullback-Leibler distance between the observed density and a factorized density.

The problem with mutual information is that it is difficult to estimate. To use the definition of entropy, we need an estimate of the density. This problem has restricted the use of mutual information in ICA estimation severely.

Nonlinear Cross-Correlations

Several researchers have used the principle of cancelling nonlinear cross-correlations to obtain the independent components. Generally speaking, such nonlinear cross-correlation is of the form $E\{g_1(y_i)g_2(y_j)\}$ where g_1 and g_2 are some suitably chosen odd nonlinear ties. Under the assumption that the y_i and y_j have symmetric densities, if y_i and y_j are independent, these cross-correlations are zero. Often the objective function here is formulated only implicitly, and an exact objective function may not even exist yet. The nonlinear ties must be chosen according to the pdf's of the independent components.

Objective Function of KICA

KICA is mainly involved with objective function which is a canonical correlations based on reproducing kernel Hilbert space. Kernel methods can be used to define an objective function that can be used to estimate the parametric part of the ICA model, despite of the absence of a specific distribution on the source nodes. Building on recent developments in kernel methods, we show that this objective-function can be computed efficiently. We define the F-correlation based on reproducing kernel Hilbert space as follow:

$$\rho_F = \max_{f_1, f_2 \in F} corr(f_1(x_1), f_2(x_2)) = \max_{f_1, f_2 \in F} \frac{\operatorname{cov}(f_1(x_1), f_2(x_2))}{(\operatorname{var} f_1(x_1))^{1/2} (\operatorname{var} f_2(x_2))^{1/2}} \quad (17)$$

Where f_1, f_2 range over F, and clearly, if the variables X_1 and X_2 are independent, then the F-correlation is equal to zero. We can also define objective-function as follows:

$$I_{\rho F} = -\frac{1}{2}\log(1 - \rho_F) \quad (18)$$

$I_{\rho F}$ is always nonnegative and equal to zero if and only if the variables $x_1 and x_2$ are independent. Based on ρ_F, we can modify above objective-function and get a new ρ_F as follows:

$$I_{\rho F} = -\frac{1}{2}\Pi_i (1 - \rho_i^2) \quad (19)$$

Where $\Pi_i (1 - \rho_i^2)$ is the generalized variance. Minimizing these objective-functions will lead to two flexible and robust kernel ICA algorithms. These two algorithms are kernel canonical correlation analysis and kernel generalized variance algorithm.

ALGORITHMS FOR ICA

Processing of the Data

Before doing fast ICA, we must preprocess the observed data. The normal goal of preprocessing is to satisfy the assumption of the data model which is that the mean of the data are zero and makes the algorithm convergent faster.

Most ICA algorithms require a preliminary centering or whitening of the data. Centering is subtracting its mean vector so as to make x a zero-mean variable, and divide the difference between the maximum and the minimum of the vector.

$$X' = \frac{X - MinX}{MaxX - MinX} \quad (20)$$

Whitening means that linear transforms the observed data x to be a new vector v and the components of v are uncorrelated and their variance is 1(Hyvarinen & Oja, 2000). The method of whitening is as follows:

$$R = E\{(X - u)(X - u)^T\}$$
$$V = D^{-\frac{1}{2}} U^T X \quad (21)$$

Where D is the eigenvalue of the covariance matrix, and u is the eigenvector. The signal after whitening is zero means and has covariance matrix $E\{VV^T\} = I$.

However, before doing ICA, we need to use PCA to do the data pretreatment whose robustness is not consistent, especially for the outlier. An effective method is to compute the logarithm of them to reduce the influence of outlier to the whole data, and enhance the robustness of ICA.

Algorithms of ICA

Jutten-Herault Algorithm

Neural networks inspired the algorithm. Their algorithm was based on cancelling the nonlinear cross-correlations. The nondiagonal terms of the matrix W are updated according to:

$$\Delta W_{ij} \propto g_1(y_i) g_2(y_j), \quad i \neq j \quad (22)$$

where g_1 and g_2 are some odd nonlinear functions, and $y = (I + W)^{-1} x$. The diagonal term W_{ij} is set to zero. The y_i then gives, after convergence, estimates of the independent components. Unfortunately, the algorithm converges only under rather severe restrictions, so it is used widely in practice.

Algorithms for Maximum Likelihood or Infomax Estimation

Girolami (1997) shows that maximization of network entropy is equivalent to the maximum likelihood approach. Usually these algorithms are based on gradient ascent of the objective function. The natural gradient method simplifies the gradient method considerably, and makes it better conditioned. Amari et al. (1996) propose the following algorithm:

$$\Delta W \propto \frac{\partial H(y)}{\partial W} W^T W = [I - \phi(u)u^T]W \tag{23}$$

Where $\phi(u)$ is the gradient vector of the log likelihood called the score function. The general learning algorithm can be derived from several theoretical viewpoints such as MLE, Infomax, and negentropy maximization. Girolami (1997) employs a parametric density model for sub- and super-Gaussian sources resulting in a simple form:

$$\begin{aligned} \Delta W &\propto [I - K\tanh(u)u^T \\ &-uu^T]W \quad \begin{cases} K = 1 \text{ super} - Gaussian \\ K = -1 sub - Gaussian \end{cases} \end{aligned} \tag{24}$$

Dinh proposed an algorithm with low consumption by minimizing mutual information, and introduced a new cost-function. Then the relative gradient can be estimated accurately (Pham, 2002).

Xia Wu (Wu, 2004) put forward an improved Infomax ICA algorithm to avoid the disadvantages of Informax ICA. He gives a transfer function which can be modulated to add a parameter *a* to logistic function as following:

$$g_i = \frac{1}{1+e^{-au}} \qquad u = wx + w_0 \tag{25}$$

Thus, we can choose a density of a different input based on the probability, so that the function g_i can match the distributing of input better. There are two rules to choose *a*. One is based on correlative coefficient while the other is on entropy distance.

Nonlinear PCA Algorithm

Nonlinear extensions of the well-known neural PCA algorithms were developed by Oja, Ogawa, and Wangviwattana (1991) .The following nonlinear version of a hierarchical PCA learning rule was introduced:

$$\Delta W \propto g(y_i)x - g(y_i)\sum_{j=1}^{i} g(y_j)w_j \tag{26}$$

where *g* is a suitable nonlinear scalar function. The symmetric version of the learning rules can be extended for the nonlinear case in the same manner. In general, the introduction of nonlinearity means that the learning rule uses higher-order information in the learning. It was proved by Karhunen and Joutsensalo (1995) and Oaj (1997) that for well-chosen nonlinear functions, the learning rule does perform ICA, if the data are sphered. Algorithms for exact maximizing the nonlinear PCA criteria were introduced by Oaj (1999).

An interesting simplification of the nonlinear PCA algorithms is the big gradient algorithm. The feedback term in the learning is here replaced by a such simpler one, giving:

$$\begin{aligned} W(t+1) &= W(t) + \mu(t)g(W(t)x(t))x(t)^T \\ &+\alpha(I - W(t)W(t)^T)W(t) \end{aligned} \tag{27}$$

where μ(t) is the learning rate sequence, α is a constant on the range [0.5 ,1], the function *g* is applied separately on every component of the sector y=Wx, and the data are assumed to be sphered.

The Fast ICA Algorithm

A fixed-point algorithm, named fast ICA, was introduced using kurtosis, and the fast ICA algorithm was generalized for general objective functions.

The fast ICA algorithm (Hyvärinen & Oja, 1997) we often used is to maximize $\mathbf{w}^{\mathbf{T}}\mathbf{Z}$ and it combines fixed-point iteration method. First we note that the maximum of $J_G(W)$ is always obtained at certain optima of $E\{G(\mathbf{w}^{\mathbf{T}}\mathbf{Z})\}$. So the question is transformed to seek the conditional extremum of $E\{G(\mathbf{w}^{\mathbf{T}}\mathbf{Z})\}$ in the case of $E\{(W^TZ)^2\}=1$. We can use the method of Lagrange multipliers to get the extremer. If the Lagrange multiplier is assumed to be β, we will get the following formula:

$$\begin{aligned} H(W) &= E\{G(w^T z) - \beta(\|w\|^2 - 1) \\ &= E\{G(w^T z)\} - \beta(w^T w - 1) \end{aligned} \tag{28}$$

Because the derivative of $H(W)$ is zero, we can get the following equation:

$$E\{\mathbf{Z}g(\mathbf{w}^{\mathrm{T}}\mathbf{Z})\} + \beta\,\mathbf{w} = 0 \tag{29}$$

Now our task is to solve the equation using Newton iteration method (Li, Wang, & Yi, 2001). The left side of the equation is assumed to be F, and the gradient of the F is:

$$\frac{\partial F}{\partial \mathbf{w}} = E\{\mathbf{Z}\mathbf{Z}^{\mathrm{T}}\,\mathrm{g}'(\mathbf{w}^{\mathrm{T}}\mathbf{Z})\} + \beta\,\mathbf{I} \tag{30}$$

To simplify the transpose of the matrix, the first term of the above formula is approximated to the following formula because the data have been normalized.

$$\begin{aligned} &E\{\mathbf{Z}\mathbf{Z}^{\mathrm{T}} g'(\mathbf{w}^{\mathrm{T}}\mathbf{Z})\} \\ &\approx E\{\mathbf{Z}\mathbf{Z}^{\mathrm{T}}\}E\{g'(\mathbf{w}^{\mathrm{T}}\mathbf{Z})\} = E\{g'(\mathbf{w}^{\mathrm{T}}\mathbf{Z})\}\mathbf{I} \end{aligned}$$

Then we can get the following formula based on Newton iteration method.

$$\mathbf{w} \leftarrow \mathbf{w} - \left.\begin{matrix}[E\{\mathbf{Z}\,\mathrm{g}(\mathbf{w}^{\mathrm{T}}\mathbf{Z})\} \\ +\beta\,\mathbf{w}]\end{matrix}\right/ \begin{matrix}[E\{\mathrm{g}'(\mathbf{w}^{\mathrm{T}}\mathbf{Z})\} \\ +\beta\,]\end{matrix} \tag{31}$$

The above formula can be further simplified as below:

$$\mathbf{w} \leftarrow E\{\mathbf{Z}\,\mathrm{g}(\mathbf{w}^{\mathrm{T}}\mathbf{Z}) - E\{\mathrm{g}'(\mathbf{w}^{\mathrm{T}}\mathbf{Z})\}\mathbf{w}\} \tag{32}$$

g' can be chosen from g_1', g_2', g_3'

$$\begin{aligned} &\mathrm{g}_1'(y) = a_1(1-\tanh^2(a_1 y)) \\ &\mathrm{g}_2'(y) = (1-y^2)\exp(-y^2/2) \\ &g_3'(y) = 3y^2 \end{aligned} \tag{33}$$

The above conclusion is for one independent component, and if we want to estimate more than one independent component, we must use the symmetric approach of fast ICA and then we can get the independent components at the same time. The symmetric approach can reduce the time complexity. The steps of the algorithm are shown in the following:

1. Change the mean value of the data into 0 and the variance into 1
2. Choose the number of the independent components
3. Initiate the value of W_i,i=1,…,m
4. $\mathbf{w} \leftarrow E\{\mathbf{z}\,\mathrm{g}(\mathbf{w}^{\mathbf{T}}\mathbf{z}) - E\{\mathrm{g}'(\mathbf{w}^{\mathbf{T}}\mathbf{z})\}\mathbf{w}\}$, g' can be chosen from g_1', g_2', g_3'
5. Suppose:

   $\mathbf{W} = (\mathbf{w}_1, ..., \mathbf{w}_m)^{\mathrm{T}}$, $\mathbf{W} \leftarrow (\mathbf{W}\mathbf{W}^{\mathrm{T}})^{-1/2}\,\mathbf{W}$ as
6. Repeat Steps (4) and (5) until convergence

To estimate several independent components, we need to run fast ICA algorithm using several units (Delfosse & Loubaton, 1995) with weight vectors $w_1,\ldots, w_n$. To prevent different vectors from converging to the same maxima we need to decorrelate the outputs $w_1^T x,\ldots, w_n^T x$ after every iteration.

A simple way of decorrelation is to estimate the independent components one by one. When p independent components are estimated, that is, $w_1,\ldots, w_p$, we run the one-unit fixed-point algorithm for w_{p+1}, and subtract from w_{p+1} the "projection" $w^T_{p+1} Cw_j\ w_j$, j=1,…,p of the previously estimated p vectors, and then renormalize w_{p+1}:

1. $$w_{p+1} = w_{p+1} - \sum_{j=1}^{p} w_{p=1}^T C w_j w_j \qquad (34)$$

2. $$w_{p+1} = \frac{w_{p+1}}{\sqrt{w_{p+1}^T C w_{p+1}}} \qquad (35)$$

The covariance matrix $C = E\{XX^T\}$ is equal to I, if the data are sphere.

Units using this fast ICA algorithm can then be combined, just as in the case of neural learning rules, into systems that estimate several independent components. Such systems may either estimate the independent component one-by-one using hierarchical decorrelation, or estimate all the independent components in parallel, with symmetric decorrelation.

The fast ICA algorithm is neural in that it is parallel and distributed, but it is not adaptive. Instead of using every data point immediately for learning, fast ICA uses sample averages computed over larger samples of the data. The convergence speed of the fixed-point algorithms is clearly superior to those of the more neural algorithms. Speed-up factors in the range from 10 to 100 are usually observed.

Tensor-Based Algorithms

A large amount of research has been done on algorithms utilizing the fourth-order cumulant tensor for estimation of ICA. Because the coefficients of the fourth-order tensor must be stored in memory, the algorithm requires O (m^4) units of memory. The algorithms also tend to be quite complicated to program, which require sophisticated matrix manipulations.

Kernel ICA Algorithm

The algorithm simulates a variety of source distributions, and it is proved by experiments that kernel ICA is superior to other ICA algorithms. Bach and Jordan propose (2002) two new ICA algorithms based on reproducing kernal hilbert space (RKHS), kernel canonical correlation analysis (KCCA), and kernel generalized variance (KGV) algorithm (Bach et al., 2001). We have introduced the object function of kernel ICA. The algorithm is illustrated in Box 1.

Box 1. Kernal ICA algorithm

Given input data: mixed vectors x:K×n dimensions,

Output data: mixture W

1. Decide the algorithm parameter such as the objective function is KCCA or KGV, Gaussian kernel width σ, standardization parameter κ, the dimension of independent component which is request N, the maximum loop mumber maxit, the precision of objective function tolJ, the precision of distance Amari tolW by users' input.

continued on following page

Box 1. continued

2. Pretreat the mixed vector x
   a) Centering: count the mean of each element of x, then deduce the mean from each element of x and get the data xc
   b) Reduce the dimensions and whiten the data: reduce the data from K- dimension to N- dimension then get the data white-sig.
3. Generate a random matrix Wrand, count its objective function C(Wrand), the produce will be given below.
4. Count the matrix Wopt that let the objective function C (Wrand) minimum by the method Stiefel manifold.
   a) Let tmin=1, iter=0 and the optimization parameter
   b) Count the value of the objective function Jold and the transform gradient gradJ, that is, the search direction dirSearch
      i. count the current value of the objective function, that is, Jold;
      ii. set the angle change radio dr=0.001, then for each element of W0, circumgyrate it by dr degrees, that is:

$$W_0([i\ \ j],:) = \begin{bmatrix} \cos(dr) & \sin(dr) \\ \sin(-dr) & \cos(dr) \end{bmatrix} \times W_0([i\ \ j],:) \tag{36}$$

      count the current value of the objective function J and set WTgradF(i,j) = WTgradF(j,i) = (J-J0)/dr;
      iii. Change the gradient gradJ = $W_0 \times W_T$gradF
   c) Search the minimum t by the golden section along geodesic then get tmin, and return the value of the current objective function Jmin at the same time.
      i. Search in two directions through the geodesic in the range [0 tmin], and decide the range [ax , cx] of new tmin' and the value of its corresponding objective function fax , fcx;
      ii. Minimize the objective function using the golden section in the range [ax , cx], then count the value of the current minimum objective function Jmin and its corresponding tmin;
   d) Count the new matrix Wnew using W , tmin , the search direction dirSearch and Stiefel manifold method.

$$W_{new} = W \times \exp(t_{\min} \times \frac{W' \times dirSearch - dirSearch' \times W}{2}) \tag{37}$$

   e) Count the distance Amari between the current matrix and the new matrix Wnew [the formula is shown as the formula (37)], if the value of the current objective function Jmin is much less than Jold then evaluate Wnew the same as the matrix W;
   f) If errW is not less than the optimistic parameter tolW, the change vale errJ of the contrast is less than the precision tolJ, and the loop number is not less than the stated maximum loop number maxit then return to a), iter=iter+1, repeat the steps above; or record the current matrix W and the value of the current minimum objective function Jmin and quit the loop;
   g) Output the matrix Wopt;

The algorithm of objective function C (W)

A. The objective function of KCCA
   1. count the Gram matrices K_1, K_2, …, K_m after centering the estimated source vectors
      i. for each xi of vector x, count
         (Li)a,b = K(xia,,xib),

         where $K(x, y) = \exp(-\frac{(x-y)^2}{2\sigma^2})$

continued on following page

Box 1. continued

ii. $p = I - \frac{1}{N} I$

iii. $K_i = PL_i P$

2. Count the minimum eigenvalue of the matrix K_k

   i. Use Cholesky decomposition

   $K_i \approx G_i G_i^T = U_i \Lambda_i^T$

   for each Gram matrix Ki

   ii. normalize|→λ(λ + Nκ/2) to get the matrix Ri for the diagonal matrix Λi

   iii. Construct the matrix **R**k,

   where $(R_k)_{ij} = R_i U_i^T U_j R_j (i \neq j), (R_k)_{ii} = I$

   iv. count the minimum eigenvalue of **R**k, that is, the minimum eigenvalueof Kk

3. count the number of the objective function by the formula

$$C(W) = -\frac{1}{2} \log \lambda_F (K_1, K_2, \cdots, K_m)$$

B. The objective function of KGV

1. count the Gram matrices K_1, K_2, …, K_m after centering the estimated source vectors

   i. For each xi of vector x, count

   $(L_i)_{a,b} = K(\chi_i^a, \chi_i^b)$

   where $K(x, y) = \exp(-\frac{(x-y)^2}{2\sigma^2})$

   ii. $P = I - \frac{1}{N} I$

   iii. $K_i = PL_i P$

2. Count the minimum eigenvalue of the matrix K_k

   i. use Cholesky decomposition

   $K_i \approx G_i G_i T = U_i \Lambda_i U_i^T$ for each Gram matrix Ki

   ii. Normalize|→λ(λ + Nκ/2) to get the matrix Ri for the diagonal matrix Λi

   iii. Construct the matrix **R**k,

   where $(R_k)_{ij} = R_i U_i^T U_j R_j (i \neq j), (R_k)_{ii} = I$

   iv. Count det(K_k)=det($\mathbf{R}_k$)

3. Count the value of the objective function by the formula

$$C(W) = -1/2 \log(\det(K_k) / \det(D_k))$$

Constrained ICA Algorithm

When dealing with practical signals, the sequence of independent components is quite important to explain the results. Therefore, eliminating indeterminacy in the permutation and dilation is useful to produce a unique ICA solution with systematically ordered signals and normalized demixing matrix. Constrained ICA uses Lagrange multiplier methods to obtain unique ICs (Lu & Rajapakse, 2000).

The independent components are ordered in a descent manner according to a certain statistical measure defined as index $L(u)$. The constrained optimization problem to constrained independent component analysis (CICA) is then defined as follows:

Minimize mutual information $M(W)$
subject to:

$$g(W) \leq 0, g(W) = [g_1(W) \cdots g_{M-1}(W)]^T \quad (38)$$

where $g(W)$ is a set of $(M-1)$ inequality constraints, $g_i(w) = L(u_{i+1}) - L(u_i)$ defines the descent order, and $L(u_i)$ is the index of some statistical measures of output components u_i, such as variance, normalized kurtosis.

Using Lagrange multiplier methods, the augmented Lagrangian function is defined as:

$$L(W,u) = M(W) + \frac{1}{2r}\sum_{i=1}^{M-1}\{[\max\{0, \overline{g_i}(W)\}]^2 - u_i^2\} \quad (39)$$

where $u = [u_1 \cdots u_m]^T$ is a set of Lagrange multiplier, r is the scalar penalty parameter, $\overline{g_i}(X) = u_i + rg_i(X)$.

The iterative procedure to determine the demixing matrix W is given as follows:

$$W \leftarrow W + \alpha(L + \Psi(u)u^T)W$$
$$w_i \leftarrow \frac{w_i}{\|w_i\|}, i = 1, 2, \ldots, N \quad (40)$$

where:

$$\Psi(u) = \begin{pmatrix} g_1(u_1) - u_1 L'(u_1) \\ g_2(u_2) + (u_1 - u_2)L'(u_2) \\ \vdots \\ g_{M-1}(u_{M-1}) + (u_{M-2} - u_{M-1})L'(u_{M-1}) \\ g_M(u_M) + u_{M-1}L'(u_M) \end{pmatrix}$$

APPLICATIONS

Mineral Resources Prediction

Gold Ore Bodies Localization Prognosis of Peripheral

Nowadays, more and more researches are concerned to how to find gold ore resources effectively in lower cost and summarize some novel methods to extract mineral resources. The following are the common steps: first, establish the looking for mine emblem of typical deposit, use every kind of the looking for mineral resources information which is related to emblem and has been fined; second, practice the target localization prognosis with the looking for mine near by the fined deposit, measure up to purpose of the looking for mine.

In Wulaga gold deposit, which is located in the east of Heilongjiang province in China, we adopt the method above, especially gamma-ray spectrometry, to explore the gold deposit. Because the gold deposits are controlled by lithology, structure, and alteration, and the alteration can result in the radioactivity difference between ore bearing structure zone and wall rock, as a result, the prevenient work ascertained five target locations and the actual project verified the nicety of prognosis. The five locations are displayed as follows. In the Figure 1, the square areas in green are the prognosis locations.

The data which we used for the experiment are a .dat file. The file includes 6,779 samples; that means sampling in 6,779 different coordinates and every coordinate has the content of 11 elements (Au, As, Hg, Sb, Bi, Cu, Pb, Zn, Ag, Mo, and W). However, the element we pay attention to is Au.

In the experiment, the data were processed by fast ICA first. However, the algorithm can be affected a lot by outliers in the data. Outliers present special relation with the majority of the sample. Although, outliers may be correct, they may also cause error mostly. To avoid the situation,

Figure 1. The five prognosis locations

we modify the outliers with the rule (Delfosse et al., 1995):

$$x = \begin{cases} E + 4\sigma & if\ x > E + 4\sigma \\ E - 4\sigma & if\ x < E - 4\sigma \end{cases} \qquad (41)$$

where E is the mean of the data and σ is the variation of the data. After the preprocessing in the original data, we do fast ICA again. In the last, the two results are compared also with the five prognosis locations.

Because the sequence of ICA outcome cannot be confirmed, the corresponding elements can be seen through comparing with the two plots above.

After processing with ICA, we find some of the results fit the prognosis locations best. The result in fast ICA with the same original data and its corresponding results after robust processed are shown as following. In the figures, the big red or blue dots indicate that there are mineral resources in this location.

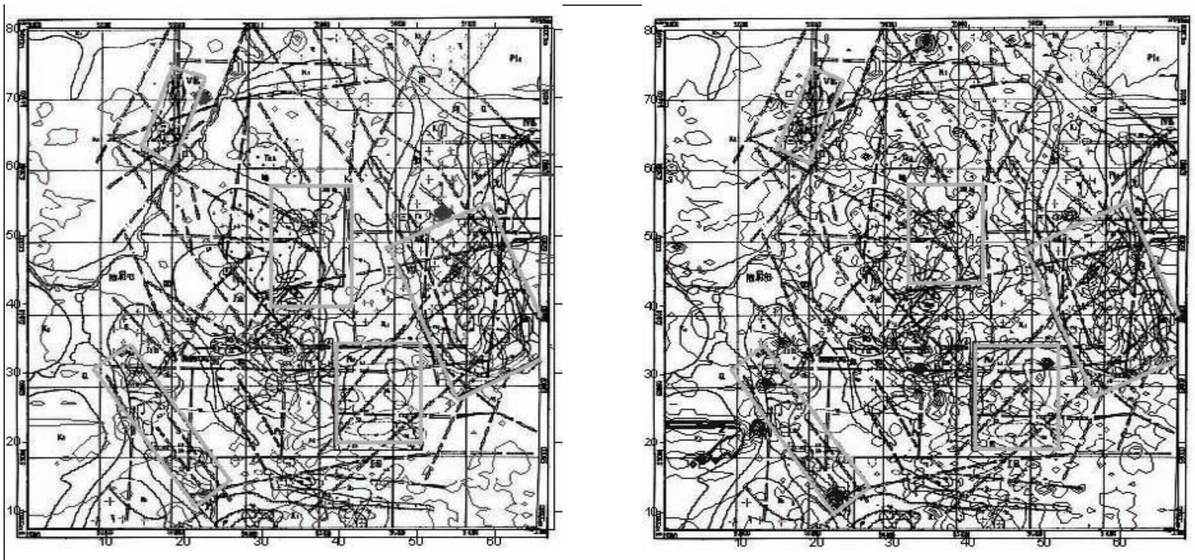
As a result, we can see that ICA can extract the main characters in a great deal of data, which means it can denote the location of gold deposit, and after robust process, the result can reflect the actual gold deposit distribution better.

Polymetallic Mines Localization Prognosis

Mineral resources are one of the crucial factors in continuous development. First, predicting mineral resources is to detect and discover the geochemical exploration exception from disordered original data; then the next step is predicting on the base of the exceptional information by mathematical models and information science. In the recent years, the optical means to predict mineral resources are using manifold information such as geology, mineral products, remote sensing, physical geography, geochemistry, and so on. Moreover, exacting the valid information of geostatistical information is the pivotal and the most difficult step in mineral resources prediction. Essentially, the physical geography exception, geochemistry exception, and remote sensing exception are all the reflection of geological exception in different aspects. Multivariate statistical analysis, geostatistics, artificial neural network, and many methods which are applied to mineral resources prediction enrich the theory and practice.

The main process methods have some liitations and cannot satisfy the complexity of the geodata. For example, principal component analysis emphasizes reducing the dimensions and supposes that the sources must be Gauss distribution. The eigenvector is orthogonal, which the majority of natural information in the realistic world is, and does not meet the request. Due to the limitations of the theory and the complexity of the geodata, results of some conventional methods are not ideal. Therefore, a new way to predict mineral resources is proposed which is independent component analysis (ICA). ICA inherits the advantages of the former theory and overcomes the deficiency of these methods. The result is more close to actual data.

Figure 2. The third element of the ICA result



(a) The result without robustness preprocessed

(b) The result after robustness preprocessing

Figure 3. The eighth elements of the ICA result

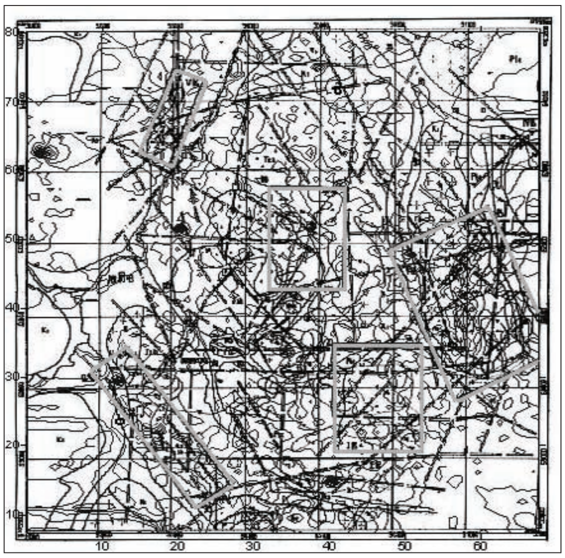
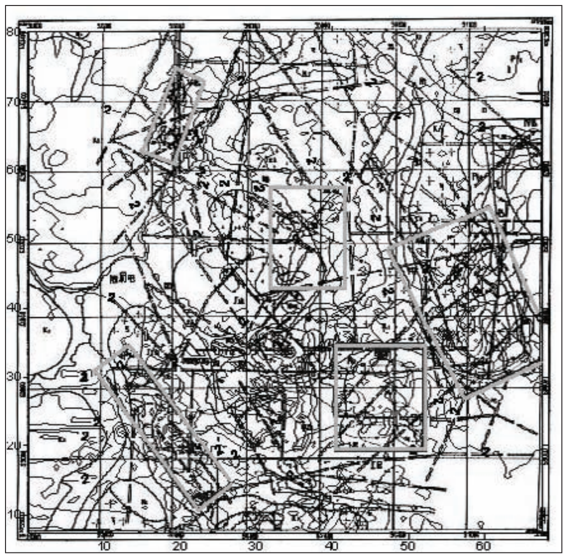


Figure 4. The 10^{th} element of the ICA result



We apply fast ICA to analyze one geochemistry datum, which was collected in a gold deposit area of Inner Mongolia province in China, has 10,855 samples, and the content of 18 elements (Ag, As, Au, Bi, Cd, Co, Cr, Cu, Fe2O3, Hg, Mn, Mo, Ni, Pb, Sb, Sn, W, and Zn).

Actually, our attention is on five elements, which are Ag, Au, Cu, Pb, Zn, and their polymetallic mines. Our task is to get relatively independent components and their linear combination by fast ICA then, predict the areas where the minerals may exist. Our source data are an excel file, and each row of this data includes sample number, line number, point number, x-coordinate, y-coordinate, and the content of each element. The original plot of the polymetallic mines is shown in Figure 5. The elementary mineral distribution in this area has been confirmed, so the Figure 6 is a background.

To display the result more clearly, we add the background to the contour in Figure 7. In Figure 7, the areas in the red circles mean that the original data accord with the actual data, while the ones in the blue circles mean that the original data do not accord with the actual data.

After we preprocessed the original data (Einax & Soldt, 1998), we do fast ICA to the preprocessed data. In our experiment, the non-linear function we chosen is a Gaussian function of. $G(y) = -\exp(-y^2 / 2)$. The iteration number

Figure 5. The original sample distribution of the polymetallic mines

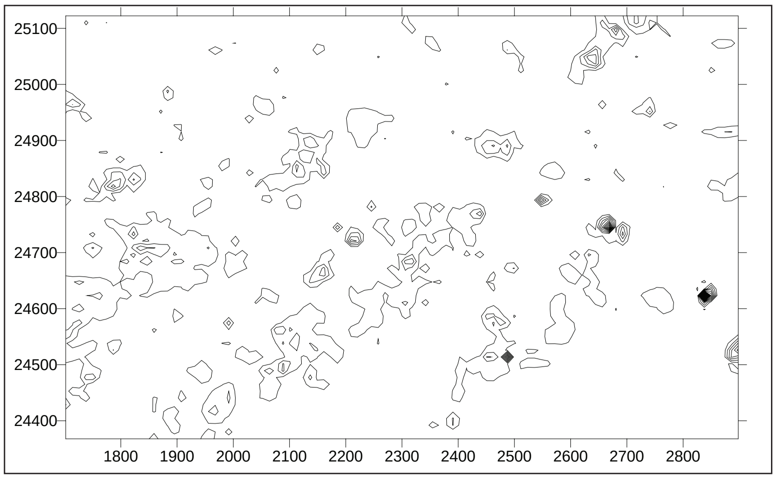


Figure 6. The already confirmed distribution of the polymetallic mines

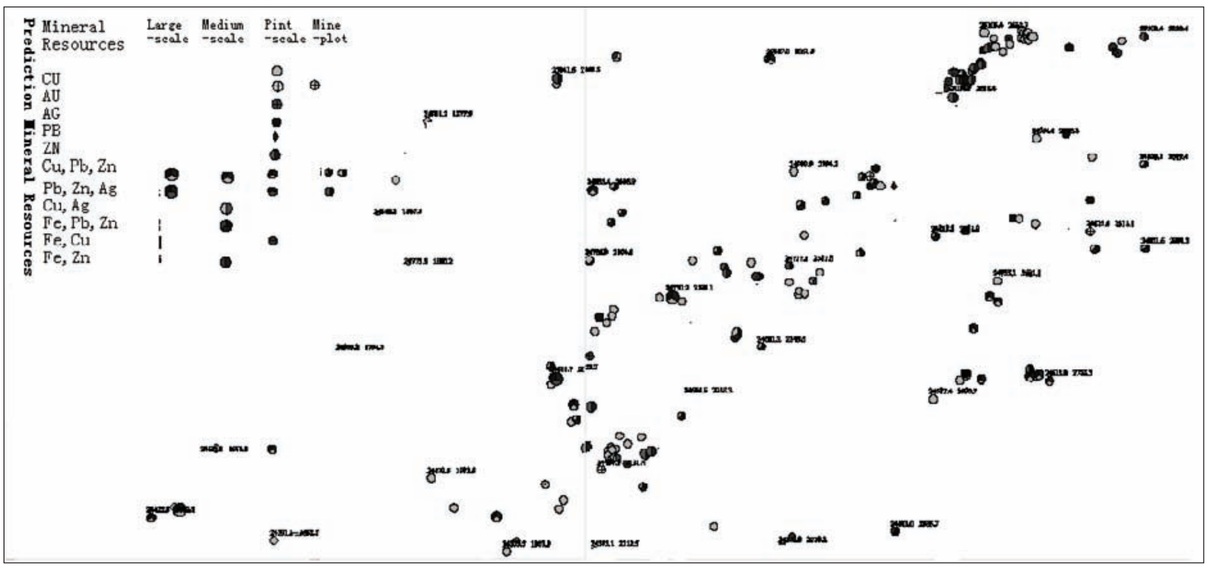


Figure 7. The original sample distribution of the polymetallic mines in the background

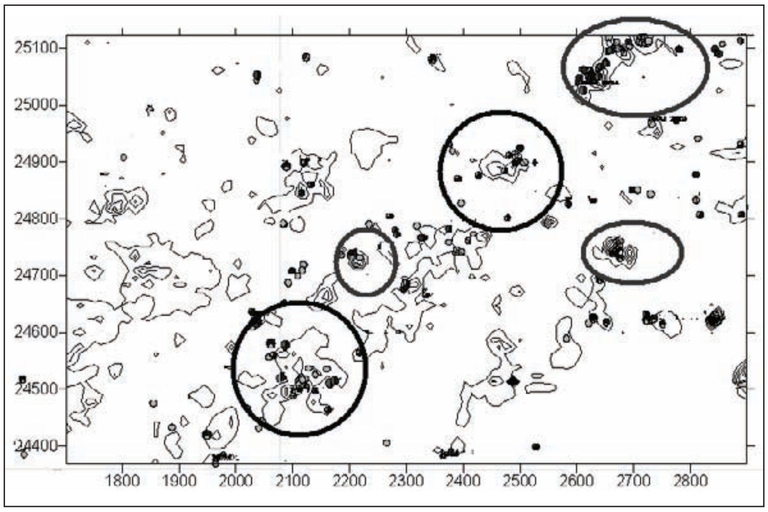
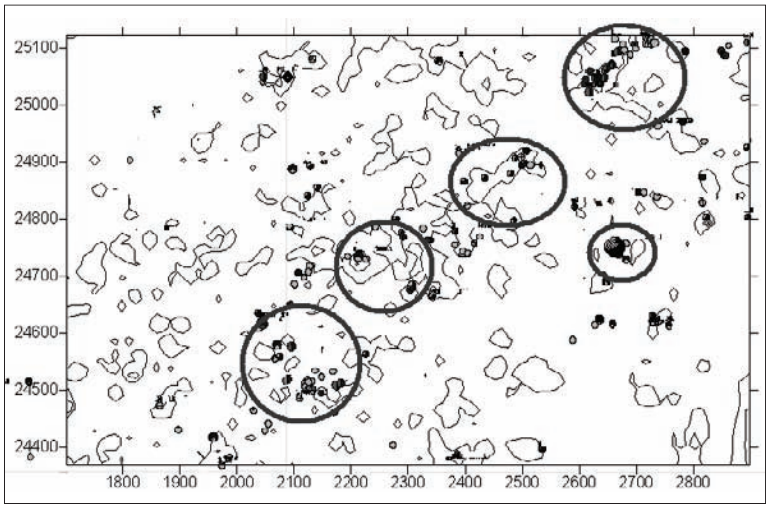


Figure 8. The independent component the polymetallic mines in the background

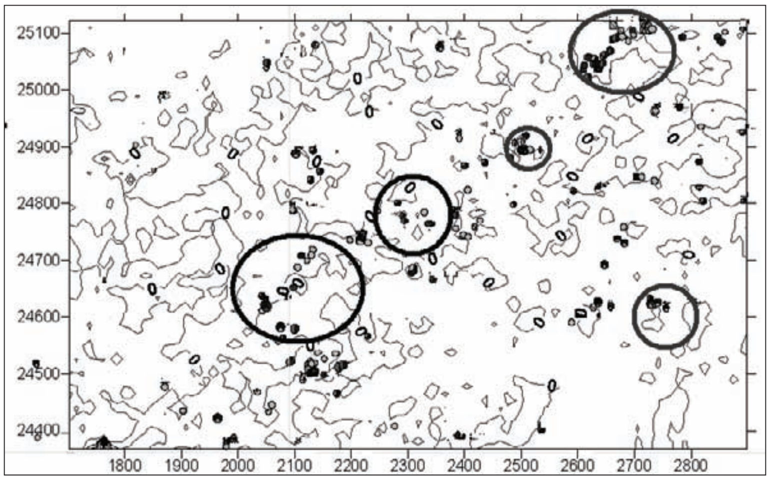


is 1,000. The result of fast ICA with the same background is plotted in Figure 8.

Contrasted to Figure 7 and Figure 8, we found that the result after eliminating outliers is closer to the practical circulations in gold deposit area of Inner Mongolia province. Meanwhile, the result is also compared to PCA. In Figure 9, it is obvious that ICA result satisfies with the background better than the PCA result.

Referred to other mineral datum, our conclusion is that the figures above can represent the distribution of a single element or the linear combination of some elements, such as element As, element Sb, element Au, and element Bi are corresponding to the forth, fifth, seventh, and tenth figures. It is instructive for the mineral resources prediction and exploitation.

Figure 9. PCA result in the background



Extracting the Information of Remote Sensing Imagery

Increasing with multispectrum and hyper-spectrum remote sensing imagery (RSI) data, scientists need effective methods to find out and analyze these data. At present, one of the most effective ways to collect information of an image is image classification (Szu & Zhang, 2002). The highly efficient and accurate automatic classification algorithm of remote sensing imagery is the key for dynamic monitoring, appraising, and predicting changes to the environment. The relatively ripe classification methods are minimum distance classification, K-means method, maximum likelihood classification, neural network method, and principal component analysis (PCA) at present (Zeng & Wang, 2004). PCA approach is to change the information of the multidimensional spectrum into several principal components which include more than 99% information of the image and has improved the classification efficiency. PCA as an orthogonal transform can reduce or dispel the influence of correlation between multibands on classification, thus improving the classification accuracy (Shah, Watanachaturaporn, Arora, & Varshney, 2003). But on the other hand, its scope of application is confined because it requires the dimension to be relatively low, its calculation accuracy is not high, and the feature vectors it received are orthogonal. However, a large number of natural information in the real world (sound, image, etc.) can not meet this demand (YU & Cheng, 2004). So in order to further improve classification accuracy; a better method is required to improve the quality and separability of the RSI to be classified.

To prevent the disadvantage effect of outliers, the first step is alternate computing to remove the outliers, then whitening and centering them.

As the referenced ICA model signal vectors in the experiment, at first, in each layer of the input image matrix, we connect lines end-to-end to form a row vector then each parataxis row vector is the input of ICA model.

We do fast ICA with varied data sources and parameter combinations, moreover, conclude the best parameter combination. Then, we display the independent components and write them in .tif files.

Intendment components can be regarded as input, and we classify the image by maximum likelihood classification (MLC) in the software IMAGE ERDAS. To remove the slender parts in the result, post process else is taken. At last, we choose Kappa parameter in confusion matrix to check the precision of classification.

The original experiment data contain a TM image coming from Shunyi District of Beijing, from April 1999 (Figure 10), a ETM+ image from Zhaoqing district, from December 2002 (Figure 11), and a radar image in Zhaoqing district in the summer of 2003 (Figure 12). The former two images have band 1, 2, 3, 4, 5, 6 while the radar image has four single bands. We chose different band combinations based on the distinct characters of each image to get the best result and reduce calculation. Varied assorted parameters are used in the same image which ensures the comparableness of the result. The independent components are shown in Figures 13, 14, and 15. The experiment data are from small to big (from 128×128 to the whole image, several data groups in all) to check the robustness of ICA. In order to solve the information loss, we try to do ICA with blocks and the result is shown in Figure 13(3). The three

Figure 10. Original image (256×256)

Figure 11. Original image (512×512)

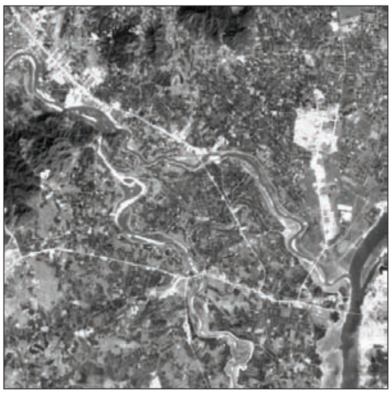
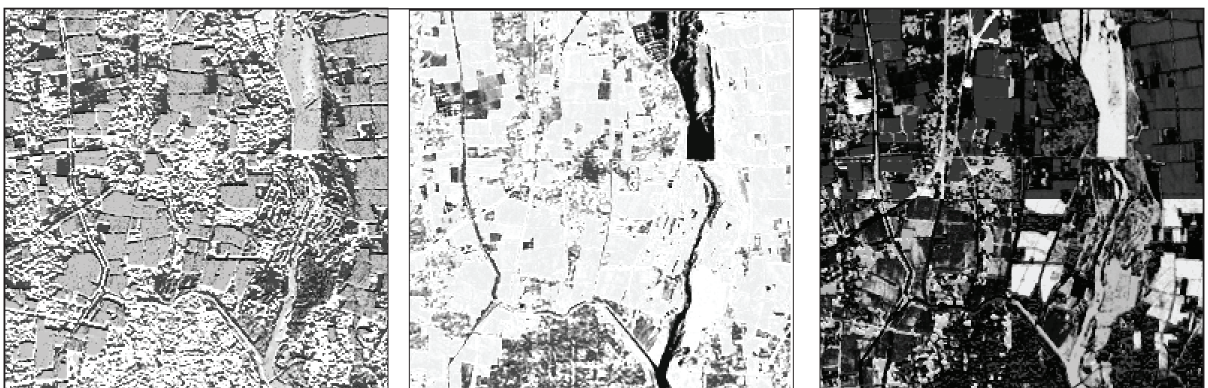


Figure 12. Original image (300×309)

Figure 13. Independent components images (Shunyi District)

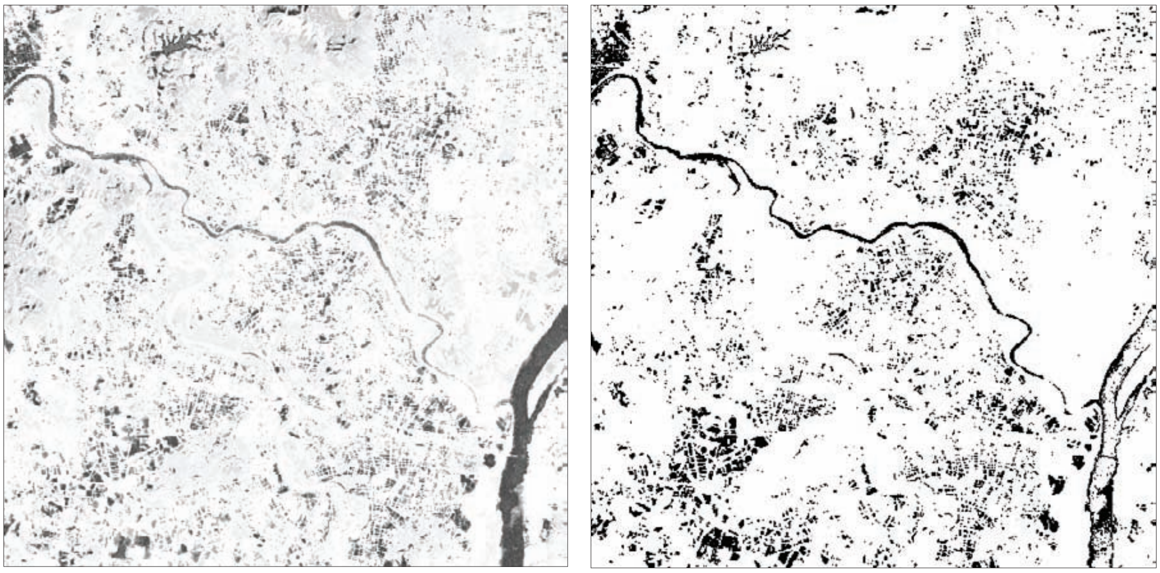


(1) The nonlinear function is pow

(2) The nonlinear function is tanh

(3) Deal with four blocks then put them together

Figure 14. Independent components images (Zhaoqing District)



(1) The bands are 5, 4, 3 (2) The bands are 6, 4, 3

Figure 15. Independent components images (Zhaoqing District SAR)

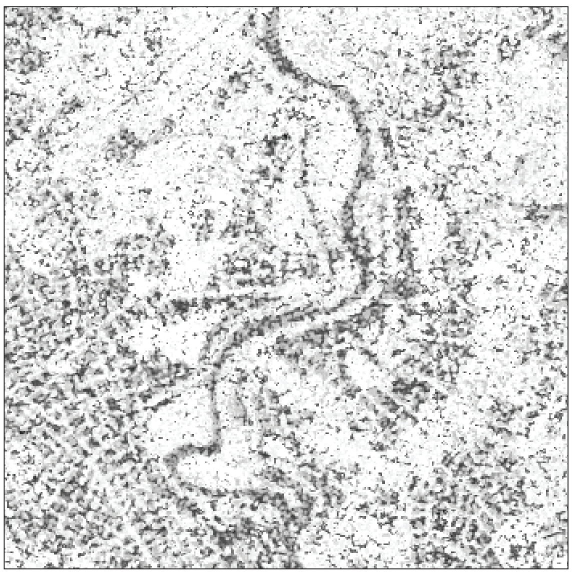


Table 1. Correlation matrix of original image

Band	5	4	3
5	1.000000	0.710994	0.721963
4	0.710994	1.000000	0.584670
3	0.721963	0.584670	1.000000

Table 2. Correlation matrix of omage processed by PCA

PC	1	2	3
1	1.000000	0.521387	0.579347
2	0.521387	1.000000	0.308342
3	0.579347	0.308342	1.000000

Table 3. Correlation matrix of image processed by ICA

IC	1	2	3
1	1.000000	-0.111410	-0.057916
2	-0.111410	1.000000	0.058096
3	-0.057916	0.058096	1.000000

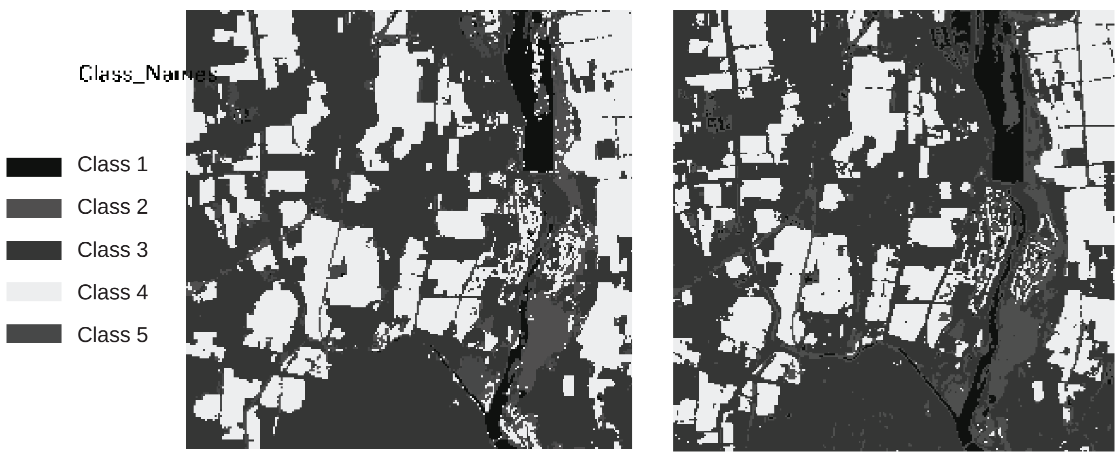
independent components after ICA, are taken as R, G, and B components of color image to form the independent component image (ICI).

From Figure 13 to 15, we can see that fastICA can get the independent components, however, it lost some texture information. For the image in Shunyi District, because its vegetation information is abundance, assembling band 5, 4,,3 and choose pow as the nonlinear function can get the best result (Figure 13[1]). As to the remote sensing imagery in Zhaoqing, because different vegetation types (woodland, vegetable plot, paddy field, and so on) are needed to be divided, the independent component by band 6, 4, 3 is the best (Figure 14).

From the preprocessed images above, we can see that images have better arrangement and separability than the original images. And from Table 1 through 3, we can find that fast ICA has better decorrelation result than PCA, which will make for classification. These ICI are the input images to be classified by MLC (Belouchrani & Cardoso, 1995). Figure 16 is the result of the classification (limited to the pages, only the best effect is shown).

From the results above, we can see that after fast ICA, the classification result is more clear and the assemble effect is better. However, some features are lost especially the nonlinear feature and texture information.

Figure 16. The classification result image (256×256)



(a) The classification result of original image

(b) The classification result of independent components

Table 4. The precision statistics (%)

classification precision / algorithm	maximal precision	minimal precision	average precision
fast ICA	76.52	43.21	59.86

To compare the classification results impersonally, we apply the Kappa parameter in confusion matrix to verify the classification precision, then get the precision appraise statistical results as shown in Table 4.

Each band image of remote sensing imagery (RSI) can be regarded as a vector in multidimensional image space. It is too complicated to use multidimensional image for classification directly. ICA can reduce dimension remarkably, and make a relatively low subspace vector standing for each band image. But different from PCA, ICA does not need the vector of subspace to be orthogonal each other, but points to the direction of the same density and finds the independent vector in the end (Lee, 1998).

Another experiment is to compare the performance of different ICA algorithms and PCA as preprocess for remote sensing imagery (RSI) classification. The original RSI is a TM imagery with band 1, 2, 3, 4, 5, 7, coming from Shunyi District of Beijing in 1999.

In this experiment, the input matrix X corresponding to the m band original RSI, which will be described in the fourth part, is linearly combined pixel by pixel. It is equivalent to stacking all the rows of one band of the RSI together from beginning to end and forming a vector of the length (1×1600•1197), and then X(*m*×1600•1197) is composed of such vector one by one. Then X is centered and whitened, yielding a matrix of whitened data Y. This is the input to ICA models in the experiment.

Each of the ICA algorithms outputs an unmixed matrix W which can be applied to the matrix Y to recover estimates of the independent components by the transformation S=WY.

PCA, fast ICA, kernel ICA-KGV, and efficient independent component analysis (EFFICA) methods are applied to process the original RSI (1600×1197) respectively at first. In the experiment, because there is an abundance of vegetation information in the original imagery, only band 3, 4, and 5 are chosen to simplify computing. Processing the same image with different methods has guaranteed the comparability of the experimental results. The original imagery is shown in Figure 17 and the principle components image (PCI) yielded by PCA and the independent component images (ICI) produced by ICA methods used in this experiment are shown from Figure 8 to Figure 21. (Only part of the images (256×256) are presented because of limited space in the article). Regard the three pieces of independent components retained from ICA processing as R, G, and B composition in colored image so as to form the independent component image (ICI), such as the PCI.

The covariance of the image matrix is calculated both before and after preprocessing. Results show that PCA method gets rid of the correlations among bands as well as ICA. However, the third and fourth order cumulants are also calculated and the results confirm that ICA method can get high-ordered statistical independent characteristic, which is the advantage over PCA and yields better class separability. We can also see from Figure 19 to Figure 21 that all of the ICIs seem easier for visual interpretation techniques than the original imagery and the PCI, however, fast ICA loses some texture information, while others keep very abundant texture information and linear characters and EFFICA performs the best.

In order to compare the convergence of these ICA algorithms, the number of iterations of the three ICA algorithms at the same error rate has also been enumerated in Figure 22. The results show that EFFICA converges the fastest and because kernel ICA algorithm looks for one contrast function among the whole function space, its quantity of calculation is great (Bach et al., 2001); therefore it has a high demand for memory and it cannot process the data size which is bigger than 512×512 directly, instead, a method to separate the image matrix (1600×1197) into several subimage matrixes to solve the “out of memory” problem is needed. While to EFFICA, it generally converged after no more than 5*R(R is the number of compo-

Figure 17. Original RSI

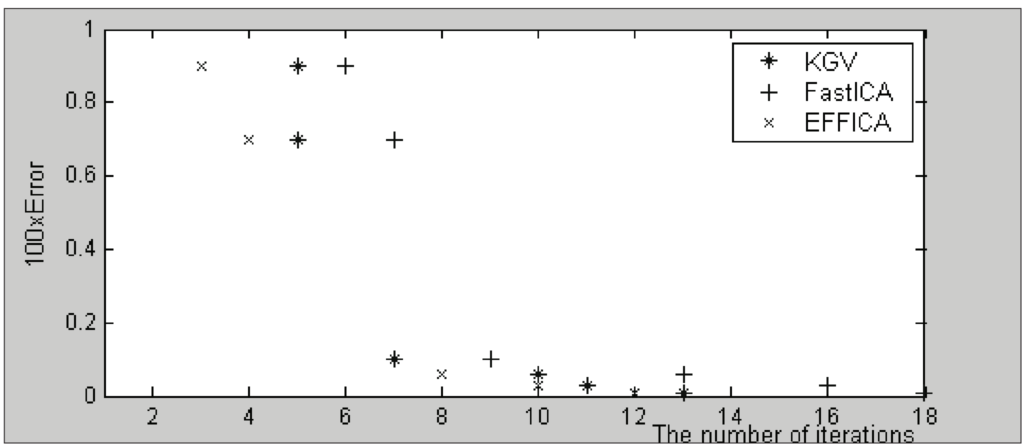
Figure 18. PCI

Figure 19. ICI of fast ICA

Figure 20. ICI of KGV

Figure 21. ICI of EFFICA

Figure 22. A comparison of convergence tempo between three ICA algorithms when they come at the same error threshold (Y-coordinate is the true error×100)



nents needed to estimate) iterations and because it is a one-step maximum likelihood estimate (MLE) which reaches asymptotic Fisher efficiency, the quantity of its computation is quite small so that its demand of memory comes down greatly, which settled the "out of memory" problem.

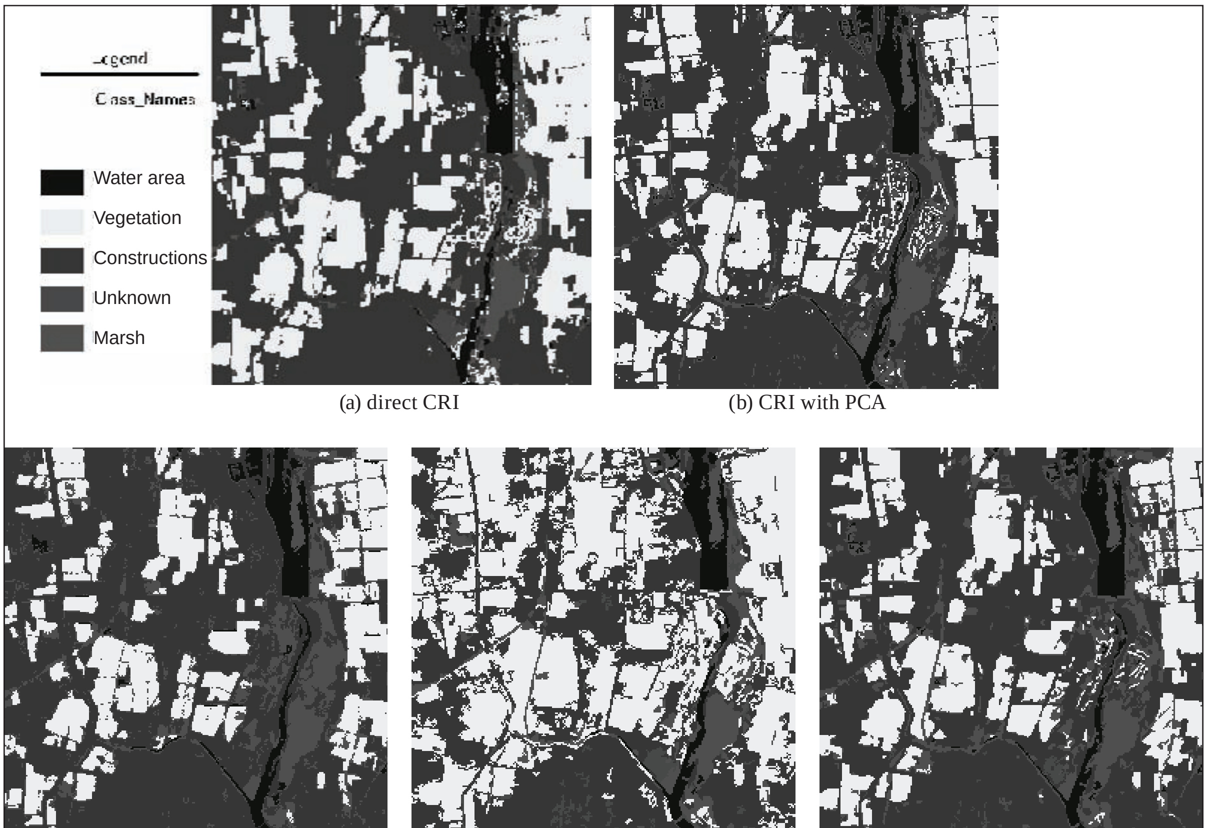
The PCI and all of the ICIs are regarded as the input of the classification model. A MLC method uniting ISODATA is used for the classification. Classification resulting images (CRI) are shown in Figure 23.

From CRIs in Figure 23, we can find that the CRIs with preprocessed data have more manifest outlines than that with raw imagery data. And compared with PCA, the performance of ICA algorithms is better. Among all the ICA algorithms, KGV seems to outperform fast ICA. KGV has not only high accuracy but also keeps abundant texture information, while fast ICA loses some objects with small pixel size. And the EFFICA improves on KGV.

Except the experiment mentioned above, an iteration method is used to get rid of the outliers in the original data foremost and more confident and robust results are derived.

The results show that the samples trained by the independent components perform better than those trained by the principal components and the

Figure 23. Classification resulting images (256×256)



(a) direct CRI (b) CRI with PCA

(c) CRI with fast ICA (d) CRI with KGV (e) CRI with EFFICA

Accuracy = 73.33% Accuracy = 72.35% Accuracy = 73.25% Accuracy = 80.33% Accuracy = 86.00%

raw data, and among the ICA methods EFFICA performs the best. More confident results are derived from gathering all the results which are obtained from different methods.

Independent component analysis decomposes the observed signal into some independent components according to statistical independent principle and by optimizing algorithms, and these independent components are a kind of estimation of the source signal. There is always certain dependence between the bands of the multiband RSI, which is unfavorable to the image classification. ICA utilizes high-ordered statistical information so that it can obtain result components as independent as possible, which enable ICA to get rid of the various kinds of correlations in RSI effectively and improve the classification accuracy. But unlike ordinary signals, the RSI often has a large number of data points and often with nonlinear noises and some outliers. From the result we can see that applying fast ICA to remote sensing image, there will be a lost of some texture information. As a result, we will continue to do the research for a better solution to avoid the problem.

CONCLUSION

ICA is a very flexible and widely-applicable tool which searches the linear transformation of the observed data into statistically maximally independent components. It is also interesting to note that the methods to compute ICA (e.g., maximum negentropy, minimum mutual Information, and maximum likelihood) are equivalent to each other (at least in the statistical sense). There is also resemblance between the forms of the gradient descent (Newton Raphson) algorithm and the fast ICA algorithm. Comparing to the other blind source separation methods, ICA can fit the actual situation better.

In many manufacturing problems, data mining and pattern recognition problems are the vital and difficult parts, while ICA can deal with blind source separation problems; it is used in many prospects, including audio (signal) processing, image processing, telecommunication, finance, education, and other manufacturing problems.

In this chapter, the main theory and some common algorithms of Independent component analysis are introduced, and we list two latest applications in ICA. ICA models have been developed and applied to solve many practical problems. However, some problems of ICA that still exist to be solve include overcomplete and nonstationary problems. Finding the most suitable model to each practical area is very important.

FUTURE RESEARCH DIRECTIONS

The ICA data model we choose in the chapter above is simply called ICA because it is a noise-free model. However, the actual data are always noisy observations. Simply ICA model just considers the ideal situation. Moreover, ICA has indeterminacy on the phase of the independent components, while in mineral resources prediction, the negative results are meaningless. Many novel approaches for systematically enhancing the ICA's usability and performance have been presented in recent years, such as constrained ICA, noisy ICA, overcomplete ICA, and so forth. As for the non-negative independent components, a new statistic analysis method has become a hot research direction world wide which is called non-negative matrix factorization (NMF). NMF can ensure the matrices after factorization are still non-negative. However each algorithm we mentioned above can just solve one aspect of the problems. One of our next research directions is to find a synthetical algorithm that can eliminate the indeterminacy of permutation, dilation, and the resulting error due to the presence of noise. Concretely, we must find a new contrast function and iterative rule.

Considering the practical application field we have been concerned, the geodata includes geochemistry data, physical geography data, geological information, mineral information, and remote sensing information. Mineral resources prediction needs analyze and evaluate the data above synthetically. All experiments we introduce in the chapter just use single kind of the data. Therefore, the other research direction is how to utilize multi-ype information of a certain area. Although many researches have been done in spatial ICA, in fact, the real spatial relativity has not been considered. Especially, for the points of spatial coordinate, a lot of work just studies the variability of single point but not unite other points, which have the relations with it in space. The next step of our work is to use the results based on remote sensing data in the areas which are mentioned in the chapter to verify or assist the mineral resources distribution then get the better prediction result. We also consider introducing the concept variance function of kriging so that the new method will accord with the spatial characteristic of the geodata.

ACKNOWLEDGMENT

Our researches have been supported by the National Natural Science Foundation of China (Grant No. 40372129, 40202030), the Natural Science Found of Beijing (Grant No. 4062020),a nd Program for New Century Excellent Talents in University. Also, thanks to Jiayi Li from Beijing Normal University, and Purnendu Mandal and Dipak Laha who are the editors of this book.

REFERENCES

Amari, S., Cichocki, A., & Yang, H. (1996). A new learning algorithm for blind signal separation. *Advances in Neural Information Processing System 8 (Proc. NIPS'95)* (pp. 757-763). Cambridge, MA: MIT Press.

Bach, F. R., Jordan, M. I., & Shawe-Taylor, J. (2002). Kernel independent component analysis. *Journal of Machine Learning Research, 3*(7), 1-48.

Bell, A. J., & Sejnowski, T. J. (1995). An information maximisation approach to blind separation and blind deconvolution. *Neural Computation, 7*(6), 1129-1159.

Belouchrani, A., & Cardoso, J. F. (1995). Maximum likelihood source separation by the expectation-maximization technique: Deterministic and stochastic implementation. In *Proceedings of the NOLTA* (pp. 49-53).

Cardoso, J. (1997). Infomax and maximum likelihood for source separation. *IEEE Letters on Signal Processing, 4*(4), 112-114.

Comon, P. (1994). Independent component analysis: A new concept? *Signal Processing, 36*(3), 287 -314.

Delfosse, N., & Loubaton, P. (1995). Adaptive blind separation of independent sources: A deflation approach. *Signal Processing, 45*(1), 59-83.

Einax, J. W., & Soldt, U. (1998). Multivariate geostatistical analysis of soil contaminations. *Fresenius Journal of Analytical Chemistry, 361*, 10-14.

Gaeta, M., Lacoume, J. L., & Lagunas, M.A. (1990). Sources separation without a prior knowledge [A]. The maximum likelihood solution. In *Proceedings of the EUSIPCO[C],* Barcelona, Spain (pp. 621-624).

Girolami, M. (1997). *Self-organizing artificial neural networks for signal separation*. Unpublished doctoral thesis, Paisley University, Scotland, Department of Computing and Information System.

Girolami, M. (1998). An alternative perspective on adaptive independent component analysis algorithms. *Neural Computation, 10*(8), 2103-2114.

Herault, J. & Jutten G. (1991). Blind seperation of sources, part I: An adaptive algorithm based on neuromimetic architecture. *Signal Processing, 24*, 1-10.

Hoyer, P. O., & Hyvrinen, A. (2000). Independent component analysis applied to feature extraction from color and stereo images. *Network: Computation in Neural Systems, 11*(3), 191-210.

Hyvärinen, A. (2001). Survey on independent component analysis. *IEEE Signal Processing Letters*, *6*(6), 145-147.

Hyvärinen, A., & Oja, E. (1997). A fast fixed point algorithm for independent component analysis. *Neural Computation*, *9*(7), 1483-1492.

Hyvarinen, A., & Oja, E. (2000). Independent component analysis: Algorithms and applications. *Neural Networks*, *13*(4-5), 411-430.

Karhunen, J., & Joutsensalo, J. (1995). Generalizations of principal component analysis, optimization problems, and neural networks. *Neural Networks*, *8*(4), 549-562.

Lee, T. (1998). *Independent component analysis: Theory and application*. Boston: Kluwer Academic Publishers.

Lee, T. W., Girolami, M., & Sejnowski, T .J. (1999). Independent component analysis using an extended infomax algorithm for mixed sub-Gaussian and super-Gaussian sources. *Neural Computation*, *11*(2), 409-433.

Li, S., & Sejnowski, T. J. (1994). Adaptive separation of mixed broadband sound sources with delays by a beam-forming Herault-Jutten network. *IEEE Journal of Oceanic Engineering*, *20*(1), 73-79.

Li, Q., Wang, N., & Yi, D. (Eds.). (2001). *Numerical value analysis* (4th ed.). The Springer Publishing Company.

Lu, W., & Rajapakse, J. C. (2000). Constrained independent component analysis. *In Advances in Neural Information Processing Systems (NIPS 2000), 13,* 570-576. Cambridge, MA: MIT Press.

Makeig, S., Bell, A. J., Jung., T. P., & Sejnowski, T. J. (1996). Independent component analysis of electroencephalographic data. *Advances in Neural Information Processing Systems*, *8*, 145-151.

Mckeown, M. J., Jung, T., & Makeig, S., Brown, G., Kindermann, S. S., Lee, T., et al. (1998). Spatially independent activity patterns in functional MRI data during the Stroop color-naming task. *Proceedings of the National Academy of Sciences*, *95*(3), 803-810.

Mlroiu, S., Kiviluoto, K., & Oja, E. (2000). ICA preprocessing for time series prediction. In *Proceedings of the 2nd Int. Workshop on Indep. Comp. Anal. and Blind Source Separation*, Helsinki, Finland (pp. 453-457).

Oja, E. (1997). The nonlinear PCA learning rule in independent component analysis. *Neurocomputing*, *17*(1), 25-46.

Oja, E. (1999). Nonlinear PCA criterion and maximum likelihood in independent component analysis. In *Proceedings of the* International *Workshop on Independent Component Analysis and Signal Separation (ICA'99)*, Aussois, France (pp. 143-148).

Oja, E., Ogawa, H., & Wangviwattana, J. (1991). Learning nonlinear constrained Hebbian networks. In T. Kohonen, et al. (Eds.), *Artificial neural networks*, *Proceedings of the. ICANN'91* (pp. 385-390). Espoo, Finland, North-Holland, Amsterdam.

Olshausen, B. A., & Field, D. J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, *381*, 607-609.

Pearlmutter, B.A. & Parra, L.C. (1997). Maximum likelihood blind source separation: A context sensitive generalization of ICA. *ICONIP'96* (pp. 151-157).

Pham, D. T. (2002a). Fast algorithms for mutual information based independent component analysis. *IEEE Transactions on Signal Processing, 52*(10), 2690-2700.

Pham, D. T. (2002b). Mutual information approach to blind separation of stationary sources. *IEEE Transactions on Information Theory, 48*(7).

Ristaniemi, T., Raju, K., & Karhunen, J. (2002). Jammer mitigation in DS-CDMA array systems using independent component analysis. In *Proceedings of the 2002 IEEE International Conference on Communications,* New York.

Shah, C. A., Watanachaturaporn, P., Arora, M. K., & Varshney, P. K. (2003). *Some recent results on hyperspectral image classification*. Paper presented at the IEEE Workshop on Advances in Techniques for Analysis of Remotely Sensed Data, NASA Goddard Spaceflight Center, Greenbelt, Maryland.

Szu, H., & Zhang, L. (2002). *Pattern process of the intellectual vision: The unsupervising learning and other methods applied to image of multicenter.* The Publishing House of the Education of Science and Technology of Shanghai.

van Hateren, J. H., & van der Schaaf, A. (1998). Independent component filters of natural images compared with simple cells in primary visual cortex. *Proceedings of the Royal Society, Series B, 265* (pp. 359-366).

Vigário R., Jousmaki, V., Hamalainen, M., Hari, R., & Oja, E. (1998). Independent component analysis for identification of artifacts in magnetoencephalographic recordings. In *Advances in neural information processing 10 (NIPS'97)* (pp. 229-235). Cambridge, MA: MIT Press.

Wu, X. (2004). *The discussion of the application of independent component analysis in fMRI data*. Unpublished master's dissertation, Beijing Normal University.

Yu, X., & Cheng, X. (2004). Research of independent component analysis (EI). *2004 IEEE International Conference on Systems, Man & Cybernetics, 1*(7), 4804-4809.

Zeng, S., & Wang, X. (2004). Technology of remote sensing image classification based on independent component analysis. *Remote Sensing Journal, 8*(2).

ADDITIONAL READING

Bach, F. R., & Jordan, M. I. (2002). Kernel independent component analysis. *Journal of Machine Learning Research, 3*(7), 1-48.

Bingham, E. (2003, December). *Advances in independent component analysis with application to data mining.* Helsinki University of Technology.

Blanco, Y., & Zazo, S. (2003). New Gaussianity measures based on order statistics: Application to ICA. *Neurocomputing (S0925-2312), 51,* 303-320.

Boscolo, R., Pan, H., & Roychowdhury, V. P. (2004). Independent component analysis based on nonparametric density estimation. *IEEE Transactions on Neural Networks (S1045-9227), 15*(1), 55-65.

Douglas, S. C., Cichocki, A., & Amari, S. (1998). A bias removal technique for blind source separation with noisy measurements. *Electronic Letters, 34,* 1379-1380.

Einax, J. W., & Soldt, U. (1998). Multivariate geostatistical analysis of soil contaminations. *Fresenius' Journal of Analytical Chemistry, 361,* 10-14. Springer-Verlag.

Haykin, S. (1998). *Neural networks: A comprehensive foundation* (2nd ed.). Prentice Hall.

HE, H., & Yu, X. (2005). A comparison between FastICA and KERNELICA in remote sensing image classification. *2005 IEEE International Geoscience and Remote Sensing Proceedings, Seoul,Korea, 1*(8), 1658-1661.

Hyvärinen, A. (1999). Fast independent component analysis with noisy data using gaussian moments. In *Proceedings of International Symposium on Circuits and Systems*, Orlando, Florida (pp. V57-V61).

Hyvärinen, A. (2001a). Blind source separation by nonstationarity of variance: A cumulant-based approach. *IEEE Transactions on Neural Networks, 12*(6), 1471-1474.

Hyvärinen, A. (2001b). Survey on independent component analysis. *IEEE Signal Processing Letters, 6*(6), 145-147.

Hyvärinen, A., & Oja, E. (1997). A fast fixed point algorithm for independent component analysis. *Neural Computation, 9*(7), 1483-1492.

Hyvärinen, A., & Oja, E. (2000). Independent component analysis: Algorithms and applications. *Neural Networks, 13*(4-5), 411-430.

Jutten, C., Babaie-Zadeh, M., & Hosseini, S. (2004). Three easy ways for separating mixtures? *Signal Processing, 84*(2), 217-229.

Lee, D., & Seung, H. S. (2001). Algorithms for non-negative matrix factorization. *Advances in Neural Information Processing Systems, 13*, 556-562.

Lewicki, M. S., & Sejnowski, T. J. (1998). Learning overcomplete representations. *Neural Computation*, *12*(2), 337-365.

Lu, W., & Rajapakse, J. C., & Leen, T. K. (Eds.). (2000). Constrained independent component analysis. *Advances in Neural Information Processing Systems 13(NIPS 2000)* (pp. 570-576).

Mckeown, M. J., Jung, T. P., & Makeig, S., et al. (1998). Spatially independent activity patterns in functional MRI data during the Stroop color-naming task. In *Proceedings of Natl.Acad.Sci. USA, 95* (pp. 803-810).

Theis, F. J. (2004). Uniqueness of complex and multimensional independent component analysis. *Signal Processing (S0165-1684), 84*(5), 951-956.

Wang, X., Tian, J., Li, X., Dai, J., & Ai, L. (2004). Detecting brain activations by constrained non-negative matrix factorization from task-related BOLD fMRI. *Proceedings of the SPIE - The International Society for Optical Engineering, 5369*(1), 675-682.

Yu, X., Yu, C., Jing, R. H., & Yang, M. (2004). The theory of disjunctive kriging and its application in grade estimate. *International Geoscience and Remote Sensing Symposium Proceedings* (pp. 295-299). Anchorage, Alaska.

Yu, X., & Cheng, X. (2004). Research of independent component analysis. *2004 IEEE International Conference on Systems, Man & Cybernetics, 1*(7), 4804-4809.

Zhang L. Q., Cichocki, A., & Amari, S. J. (1999). Natural gradient algorithm for blind separation of overdetermined mixture with additive noise. *IEEE Signal Processing Letters, 6*(11), 293-295.

Zheng, C., Huang, D., Sun, Z., Lyu, M. R., & Lok, T. M. (2006). Nonnegative independent component analysis based on minimizing mutual information technique. *Neurocomputing, 69*, 878-883.

Chapter XVIII
Swarm Intelligence in Production Management and Engineering

Swagatam Das
Jadavpur University, India

Amit Konar
Jadavpur University, India

ABSTRACT

This chapter explores the scope of biologically inspired swarm intelligence (SI) into production management with special emphasis in two specific problems of vehicle routing and motion planning of mobile robots. Computer simulations undertaken for this study have also been included to demonstrate the elegance in the application of the proposed theory in the said real-world problems. Possible directions of future research and industrial applications have also been appended at the end of the chapter.

INTRODUCTION

Biologically inspired computing is currently given importance for its immense parallelism and simplicity in computation. In recent times, quite a large number of biologically motivated algorithms have been invented, and are being used for handling many complex problems of the real world. For instance, *neural computing* (Haykins, 1999) attempts to mimic the biological nervous systems of the living creatures to ensure a significant amount of parallel and distributed processing in computation. *Genetic algorithms* (Golberg, 1989; Holland, 1975) imitate the Darwinian evolutionary process through cross-over and mutation of biological chromosomes. They have successfully been used in intelligent search, optimization, and machine learning applications. *Artificial life*

(Langton, 1995) attempts to model the group behavior of a biological species to interpret complex phenomena of many primitive macrostructures. One typical example of artificial life algorithm includes *artificial immune systems* (Jerne, 1983), which have proved themselves successful in many engineering applications, including behavioral robotics and antivirus vaccines.

In this chapter, we would like to examine the collective behavior of a group of biological creatures such as ants, bees, termites, and wasps. This family of algorithms is called *swarm intelligence (SI)* (Engelbrecht, 2005). The focus of our study is centered on two distinct types of SI algorithms and their applications in appropriate production management and engineering problems.

The behavior of a single ant, bee, termite, and wasp often is too simple, but their collective and social behavior is of paramount significance. A look at National Geographic TV Channel also reveals that advanced mammals including lions enjoy social lives, perhaps for their self-existence at old age and in particular when they are wounded. The collective and social behavior of living creatures motivated researchers to undertake the study of swarm intelligence. Historically, the phrase SI was coined by Beny and Wang in 1989 in the context of cellular robotics. A group of researchers in different parts of the world started working almost at the same time to study the versatile behavior of different living creatures. Inspired by the collective behavior of ants, Marco Dorigo (1992) came up with an interesting solution to a class of optimization problems with his so-called *ant systems*. Dorigo's classical ant systems have undergone an evolution in its fifteen years lifespan, which finally resulted in the *ant colony optimization* (ACO) *algorithm* (Dorigo, Di Caro, & Gambardella, 1999).

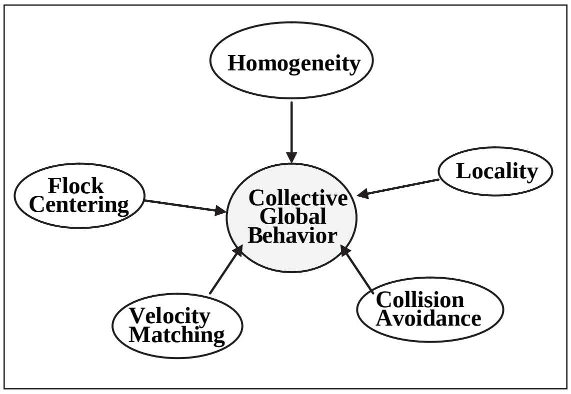
Sociological behavior of birds and fish schools also motivated researchers to study their collective characteristics, such as movements in groups, hunting patterns, and selection of breeding places. Kennedy and Eberhart (1995) enunciates an interesting dynamics of *artificial swarm* to determine the optima in a search landscape. Their algorithm is popularly known as *particle swarm optimization* (PSO). PSO has already entered varieties of engineering and scientific design, optimization, classification, control, machine learning, and other applications. In this chapter, we examine the scope of ACO and PSO in production management problems.

General Characteristics of Swarm Intelligence Algorithms

An agent is an entity capable of performing/executing certain operations. SI systems are typically made up of a population of simple agents interacting locally with one another and with their environment. Although there is normally no centralized control structure dictating how individual agents should behave, local interactions between such agents often lead to the emergence of global behavior. Many biological creatures such as fish schools and bird flocks clearly display structural order, with the behavior of the organisms so integrated that even though they may change shape and direction, they appear to move as a single coherent entity (Couzin, Krause, James, Ruxton, & Franks, 2002). The main properties of the collective behavior can be pointed out as

- **Homogeneity:** Every bird in flock has the same behavioral model. The flock moves without a leader, even though temporary leaders seem to appear.
- **Locality:** Its nearest flock-mates only influence the motion of each bird. Vision is considered to be the most important sense for flock organization.
- **Collision Avoidance:** Avoid colliding with nearby flock mates.
- **Velocity Matching:** Attempt to match velocity with nearby flock mates.
- **Flock Centering:** Attempt to stay close to nearby flock mates

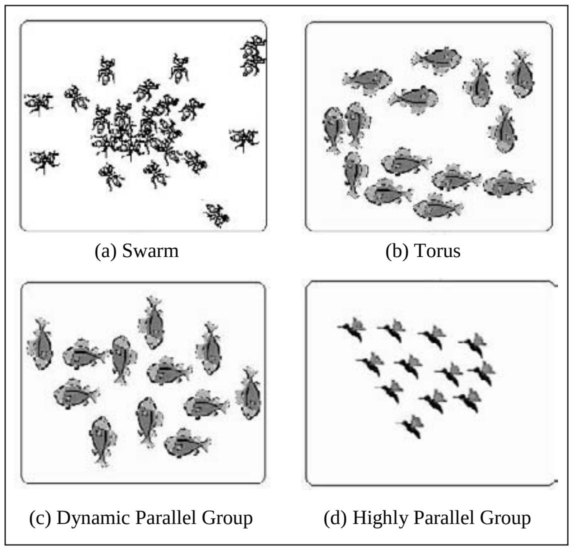
Figure 1. Main traits of collective behavior



Individuals attempt to maintain a minimum distance between themselves and others at all times. This rule is given the highest priority and corresponds to a frequently observed behavior of animals in nature (Krause & Ruxton, 2002). If individuals are not performing an avoidance maneuver they tend to be attracted towards other individuals (to avoid being isolated) and to align themselves with neighbors (Partridge, 1982; Partridge & Pitcher, 1980). Couzin et al. (2002) identifies four collective dynamical behaviors as illustrated in Figure 2:

- **Swarm:** An aggregate with cohesion, but a low level of polarization (parallel alignment) among members.
- **Torus:** Individuals perpetually rotate around an empty core (milling). The direction of rotation is random.
- **Dynamic parallel group:** The individuals are polarized and move as a coherent group, but individuals can move throughout the group and density and group form can fluctuate.
- **Highly parallel group:** Much more static in terms of exchange of spatial positions within the group than the dynamic parallel group and the variation in density and form is minimal.

Figure 2. Different models of collective behavior



As mentioned by Fayyad, Piatestku-Shapio, Smyth and Uthurusamy (1996), at a high-level, a swarm can be viewed as a group of agents cooperating to achieve some purposeful behavior and achieve some goal (see Figure 2). This collective intelligence seems to emerge from what are often large groups (Grosan, Abraham, & Monica, 2006).

According to Milonas (1994), five basic principles define the SI paradigm. First is the proximity principle where the swarm should be able to carry out simple space and time computations. Second is the quality principle in which the swarm should be able to respond to quality factors in the environment. Third is the principle of diverse response where the swarm should not commit its activities along excessively narrow channels. Fourth is the principle of stability where the swarm should not change its mode of behavior every time the environment changes. Fifth is the principle of adaptability in which the swarm must be able to change behavior mote when it is worth the computational price. Note that principles four and five are the opposite sides of the same coin.

As it appears, "self-organization" is one of the fundamental features of any SI system. However, it is not a simple term to define. In general, it refers to the various mechanisms by which pattern, structure, and order emerge spontaneously in complex systems. Examples of such structures and patterns include the stripes of zebras, the pattern of sand ripples in a dune, the coordinated movements of flocks of birds or schools of fish, the intricate earthen nests of termites, the patterns on seashells, the whorls of our fingerprints, the colorful patterns of fish, and even the spatial pattern of stars in a spiral galaxy. Bonabeau, Dorigo, and Theraulaz (1998) try to define self-organization in the following language:

Self-organization is a set of dynamical mechanisms whereby structures appear at the global level of a system from interactions of its lower-level components.

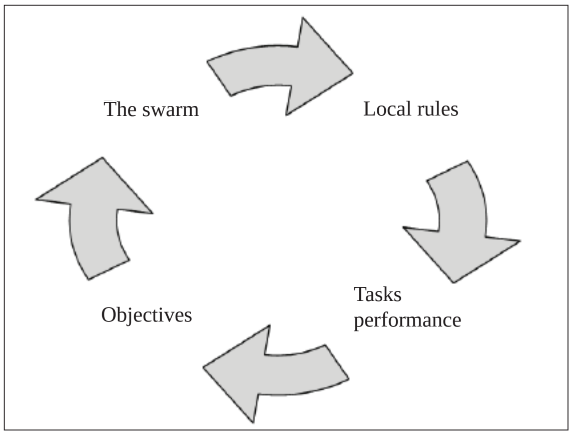
Serra and Zanarini (1990) describe the concept of self-organization generally as "highly organized behavior even in the absence of a pre-ordained design." They go on to further describe examples such as the resonance phenomenon in lasers, and in cellular automata where "unexpected and complex behaviours can be considered as self-organized." Self-organization was originally introduced in the context of physics and chemistry to describe how microscopic processes give rise to macroscopic structures in out-of-equilibrium systems. Recent research, however, suggests that it provides a concise description of a wide rage of collective phenomena in animals, especially in social insects. This description does not rely on individual complexity to account for complex spatial-temporal features, which emerge at the colony level, but rather assumes that interactions among simple individuals can produce highly structured collective behaviors. There are four main features that govern the self-organization in insect colonies:

- Positive feedback (amplification)
- Negative feedback (for counter-balance and stabilization)
- Amplification of fluctuations (randomness, errors, and random walks)
- Multiple interactions

At a high-level, a swarm can be viewed as a group of agents cooperating to achieve some purposeful behavior and achieve some goal (see Figure 3). This collective intelligence seems to emerge from what are often large groups of relatively simple agents. The agents use simple local rules to govern their actions and via the interactions of the entire group, the swarm achieves its objectives. A type of self-organization emerges from the collection of actions of the group.

An autonomous agent is a subsystem that interacts with its environment, which probably consists of other agents, but acts relatively independently from all other agents (Fayyad, Piatestku-Shapio, Smyth, & Uthurusamy, 1996). The autonomous agent does not follow commands from a leader, or some global plan (Flake, 1990). For example, for a bird to participate in a flock, it only adjusts its movements to coordinate with the movements of its flock mates, typically its neighbors that are

Figure 3. The scheme of a simple swarm



close to it in the flock. A bird in a flock simply tries to stay close to its neighbors, but avoids collisions with them. Each bird does not take commands from any leader bird since there is no lead bird. Any bird can in the front, center, and back of the swarm. Swarm behavior helps birds take advantage of several things including protection from predators (especially for birds in the middle of the flock), and searching for food (essentially each bird is exploiting the eyes of every other bird).

Below we discuss in details two algorithms from SI domain, which have gained huge popularity in a relatively short span of time all over the world. One of these algorithms, known as ant colony optimization (ACO), mimics the behavior of a group of real ants in multiagent cooperative search problems. The latter one is referred to as particle swarm optimization (PSO), which draws inspiration from the behavior of particles, the boids method of Craig Reynolds, and sociocognition (Kennedy, Eberhart, & Shi, 2001).

Ant Colony Systems: An Overview

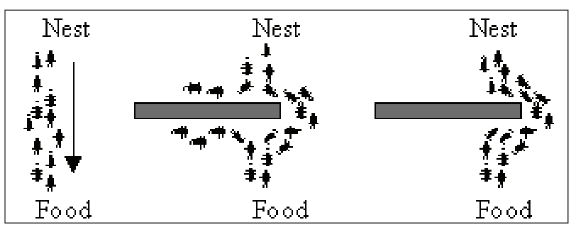
Insects like ants, bees, wasps, and termites are quite social. They live in colonies and follow their own routine of tasks independent of each other. However, when acting as a community, these insects even with very limited individual capability can jointly (cooperatively) perform many complex tasks necessary for their survival (Bonabeau et al., 1998). Problems like finding and storing foods and selecting and picking up materials for future usage require detailed planning, and are solved by insect colonies without any kind of supervisor or controller.

It is a natural observation that a group of "almost blind" ants can figure out the shortest route between a cube of sugar and their nest without any visual information. They are capable of adapting to the changes in the environment as well (Dorigo & Gambardella, 1996). It is interesting to note that ants while crawling deposit trails of a chemical substance known as pheromone to help other members of their team to follow its trace. The resulting collective behavior can be described as a loop of positive feedback where the probability of an ant choosing a path increases as the count of ants that already passed by that path increases (Dorigo, Di Caro, & Gambardella, 1999; Dorigo & Gambardella, 1996).

The basic idea of a real ant system is illustrated in Figure 4. In the left picture, the ants move in a straight line to the food. The middle picture illustrates the situation soon after an obstacle is inserted between the nest and the food. To avoid the obstacle, initially each ant chooses to turn left or right at random. Let us assume that ants move at the same speed depositing pheromone in the trail uniformly. However, the ants that, by chance, choose to turn left will reach the food sooner, whereas the ants that go around the obstacle turning right will follow a longer path, and so will take longer time to circumvent the obstacle. As a result, pheromone accumulates faster in the shorter path around the obstacle. Since ants prefer to follow trails with larger amounts of pheromone, eventually all the ants converge to the shorter path around the obstacle, as shown in the right picture.

An artificial ant colony system (ACS) is an agent-based system, which simulates the natural behavior of ants and develops mechanisms of cooperation and learning. ACS was proposed by Dorigo and Gambardella (1997) as a new heuristic to solve combinatorial optimization

Figure 4. Illustrating the behavior of real ant movements



problems. This new heuristic, called ant colony optimization (ACO) has been found to be both robust and versatile in handling a wide range of combinatorial optimization problems.

The ACO Algorithm

The main idea of ACO is to model a problem as the search for a minimum cost path in a graph. Artificial ants as if walking on this graph, look for cheaper paths. Each ant has a rather simple behavior capable of finding relatively costlier paths. Cheaper paths are found as the emergent result of the global cooperation among ants in the colony. The behavior of artificial ants is inspired from real ants; they lay pheromone trails (obviously in a mathematical form) on the graph edges and choose their path with respect to probabilities that depend on pheromone trails. These pheromone trails progressively decrease by evaporation. In addition, artificial ants have some extra features not seen in their counterpart in real ants. In particular, they live in a discrete world (a graph) and their moves consist of transitions from nodes to nodes.

Pheromone placed on the edges acts like a *distributed long term memory* (Dorigo & Gambardella, 1997). The memory, instead of being stored locally within individual ants, remains distributed on the edges of the graph. This indirectly provides a means of communication among the ants called *stigmergy*. In most cases, pheromone trails are updated only after having constructed a complete path and not during the walk, and the amount of pheromone deposited is usually a function of the quality of the path. Finally, the probability for an artificial ant to choose an edge not only depends on pheromones deposited on that edge in the past, but also on some problem dependent local heuristic functions.

Solving the Classical TSP Problem by ACO

Below we illustrate the use of ACO in finding the optimal tour in the classical traveling salesman problem (TSP). Given a set of n cities and a set of distances between them, the problem is to determine a minimum traversal of the cities and return to the home-station at the end. It is indeed important to note that the traversal should in no way include a city more than once. Let r (C_x, C_y) be a measure of cost for traversal from city C_x to C_y. Naturally, the total cost of traversing n cities indexed by $i_1, i_2, i_3, \ldots, i_n$ in order is given by the following expression:

$$\mathrm{COST}(i_1, i_2, ..., i_n) = \sum_{j=1}^{n-1} r(Ci_j, Ci_{j+1}) + r(Ci_n, Ci_1) \tag{1}$$

The ACO algorithm is employed to find an optimal order of traversal of the cities. Let τ be a mathematical entity modeling the pheromone and η_{ij} = 1/r (i , j) is a local heuristic. Also let $\text{allowed}_k(t)$ be the set of cities that are yet to be visited by ant k located in city i. Then according to the classical ant system (Dorigo and Gamberdella, 1999) the probability that ant k in city i visits city j is given by:

$$p_{ij}^k(t) = \frac{\left[\tau_{ij}(t)\right]^{\alpha} . \left[\eta_{ij}\right]^{\beta}}{\sum_{h \in allowed_k(t)} \left[\tau_{ih}(t)\right]^{\alpha} . \left[\eta_{ih}\right]^{\beta}}, \tag{2}$$

if j ∈ $\text{allowed}_k(t)$ = 0, otherwise.

In equation (2), shorter edges with greater amount of pheromone are favored by multiplying the pheromone on edge (i , j) by the corresponding heuristic value η(i, j). Parameters α (> 0) and β(>

0) determine the relative importance of pheromone vs. cost. Now in ant systems, pheromone trails are updated as follows. Let D_k be the length of the tour performed by ant k, $\Delta\tau_k$ (i , j)= $1/D_k$ if (i, j) $\in$ tour done by ant k and = 0 otherwise, and finally let $\rho \in [0,1]$ be a pheromone decay parameter which takes care of the occasional evaporation of the pheromone from the visited edges. Then once all ants have built their tours, pheromone is updated on all the ages as:

$$\tau\ (i\,,\,j) = (1-\rho).\ \tau(i,\ j) + \sum_{k=1}^{m} {}_{k}\,\Delta\tau\ (i,j) \tag{3}$$

From equation (3), we can guess that pheromone updating attempts to accumulate greater amount of pheromone to shorter tours (which corresponds to high value of the second term in equation (3) so as to compensate for any loss of pheromone due to the first term). This conceptually resembles a reinforcement-learning scheme, where better solutions receive a higher reinforcement.

The ACS (Dorigo & Gambardella, 1997) differs from the classical ant system in the sense that here the pheromone trails are updated in two ways. First, when ants construct a tour they locally change the amount of pheromone on the visited edges by a local updating rule. Now if we let γ to be a decay parameter and $\Delta\tau(i, j) = \tau_0$ such that τ_0 is the initial pheromone level, then the local rule may be stated as:

$$\tau\ (i, j) = (1-\gamma).\ \tau\ (i, j) + \gamma\ .\ \Delta\tau\ (i, j) \tag{4}$$

Second, after all the ants have built their individual tours, a global updating rule is applied to modify the pheromone level on the edges that belong to the best ant tour found so far. If κ is the usual pheromone evaporation constant, D_{gb} is the length of the globally best tour from the beginning of the trial, and $\Delta\tau^{/}$ (i , j) = 1/ D_{gb} only when the edge (i, j) belongs to global-best-tour and zero otherwise, then we may express the global rule as:

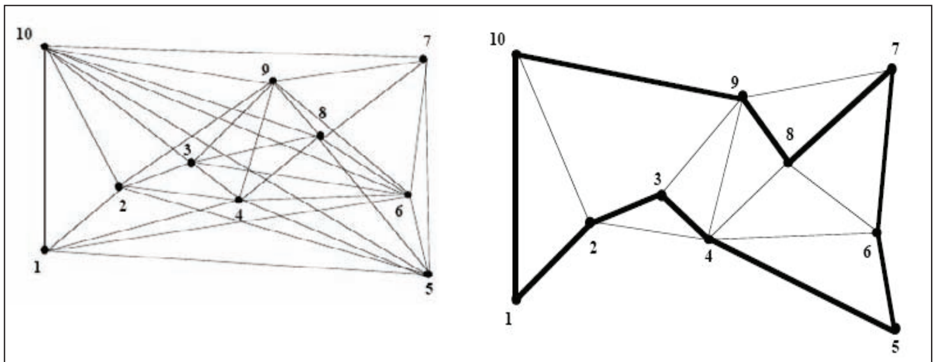
$$\tau\ (i, j) = (1-\kappa).\ \tau\ (i\ ,j) + \kappa.\ \Delta\tau^{/}\ (i, j) \tag{5}$$

The main steps of ACS algorithm are presented below:

Procedure ACS
Begin
Initialize pheromone trails;
Repeat
Begin /*at this stage each loop is called an iteration */
 Each ant is positioned on a starting node;
 Repeat
 Begin /* at this level each loop is called a step */
Each ant applies a *state transition rule like Rule (2)* to
 incrementally build a solution and a *local pheromone-updating rule like Rule (4);*
 Until all ants have built a complete solution;
 A *global pheromone-updating rule like Rule (5)* is
 applied.
Until Terminating_condition is reached;
End

In the above algorithm, the first loop (iteration) starts with m ants being placed in n cities chosen according to some initialization rule (e.g., randomly). In the embedded loop (step) each ant builds a tour (i.e., an acceptable solution to the TSP) by repeatedly applying a stochastic state transition rule. While building its tour, the ant can modify the pheromone level on the visited edges by applying the local updating rule given by equation (4). Once all the ants have terminated their tour, the pheromone trails are modified again by the global updating rule given in equation (5). In Figure 5, we show the computer simulation of the ACO working on a 10 city TSP problem.

Figure 5. Solving the TSP problem with ACO algorithm(Dorigo et al., 1996)



(a) Trail distribution at the beginning (b) Trail distribution after 100 iterations

The Particle Swarm Optimisation (PSO)

The concept of particle swarms, although initially introduced for simulating human social behaviors, has become very popular these days as an efficient search and optimization technique. The particle swarm optimization (PSO), as it is called now, does not require any gradient information of the function to be optimized, but uses only primitive mathematical operators and is conceptually very simple. Since its advent in 1995, PSO has attracted the attention of a lot of researchers all over the world resulting into a huge number of variants of the basic algorithm as well as many parameter automation strategies.

PSO (Kennedy & Eberhart, 2001) is in principle a multiagent parallel search technique. Particles are conceptual entities which fly through the multidimensional search space. At any particular instant each particle has a position and a velocity. The position vector of a particle with respect to the origin of the search space represents a trial solution of the search problem. At the beginning a population of particles is initialized with random positions marked by vectors $\vec{X}_i$ and random velocities $\vec{V}_i$. The population of such particles is called a "swarm" S. A neighborhood relation N is defined in the swarm. N determines for any two particles P_i and P_j whether they are neighbors or not. Thus for any particle P, a neighborhood can be assigned as $N(P)$, containing all the neighbors of that particle. Different neighborhood topologies and their effect on the swarm performance have been discussed by Kennedy (1999). The PSO used in this work, implicitly uses a so-called fully connected neighborhood topology (or *gbest*). Every particle is a neighbor of every other particle.

Each particle P has two state variables:

1. Its current position $\vec{x}(t)$.
2. Its current velocity $\vec{v}(t)$.

And also a small memory comprising of:

1. Its previous best position $\vec{p}(t)$, that is, personal best experience in terms of the objective function value $f(\vec{p}(t))$.
2. The best $\vec{p}(t)$ of all $P \in N(P)$:$\vec{g}(t)$, that is, the best position found so far in the neighborhood of the particle.

The PSO scheme has the following algorithmic parameters:

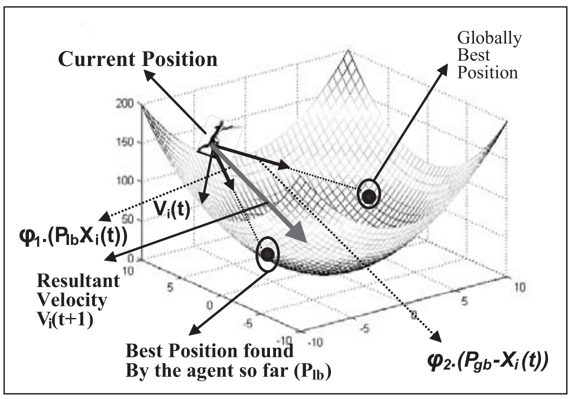
1. V_{max} or maximum velocity which restricts $\vec{V}_i(t)$ within the interval $[-V_{max}, V_{max}]$.
2. An inertial weight factor ω.
3. Two uniformly distributed random numbers φ_1 and φ_2 which respectively determine the influence of $\vec{p}(t)$ and $\vec{g}(t)$ on the velocity update formula.
4. Two constant multiplier terms C_1 and C_2 known as "self confidence" and "swarm confidence" respectively.

Initially the settings for $\vec{p}(t)$ and $\vec{g}(t)$ are $\vec{p}(0) = \vec{g}(0) = \vec{x}(0)$ for all particles. Once the particles are initialized, the iterative optimization process begins where the positions and velocities of all the particles are altered by the following recursive equations. The equations are presented for the d-th dimension of the position and velocity of the i-th particle.

$$\left.\begin{aligned} V_{id}(t+1) &= \omega V_{id}(t) + C_1\varphi_1 . (P_d(t) \\ &\quad - X_{id}(t)) + C_2\varphi_2 . (g_d(t) - X_{id}(t)) \\ X_{id}(t+1) &= X_{id}(t) + V_{id}(t+1) \end{aligned}\right\} \qquad (6)$$

The first term in the velocity updating formula represents the inertial velocity of the particle. The second term involving $\vec{P}(t)$ represents the personal experience of each particle and is referred to as "cognitive part." The last term of the same relation is interpreted as the "social term" which represents how an individual particle is influenced by the other members of its society. Typically, this process is iterated for a certain number of time steps, or until some acceptable solution has been found by the algorithm or until an upper limit of CPU usage has been reached. Once the iterations are terminated, most of the particles are expected to converge to a small radius surrounding the global optima of the search space. The ideal distribution of the particles after the algorithm is stopped is shown in Figure 6.

Figure 6. Illustrating the velocity updating scheme of basic PSO



A pseudo code for the PSO algorithm may be put forward as:

The PSO Algorithm

Input: Randomly initialized position and velocity of the particles: $\vec{X}_i(0)$ and $\vec{V}_i(0)$
Output: Position of the approximate global optima $\vec{X}^*$

```
Begin
 While terminating condition is not reached do
 Begin
  for i = 1 to number of particles
  Evaluate the fitness: = f(X_i);
  Update P_lbi and P_gb;
  Adapt velocity of the particle using equation (6);
  Update the position of the particle;
  increase i;
end while
end
```

The PSO algorithm can be seen as a set of vectors whose trajectories oscillate around a region defined by each individual previous best position and the best position of some other individuals (Kennedy & Eberhart, 2001). There are different neighborhood topologies used to identify which particles from the swarm can influence the individuals. The most common ones are known as the *gbest* and *lbest*. In the *gbest* swarm, the trajectory

Figure 7. Graphical representation of (a) gbest swarm (b) lbest swarm

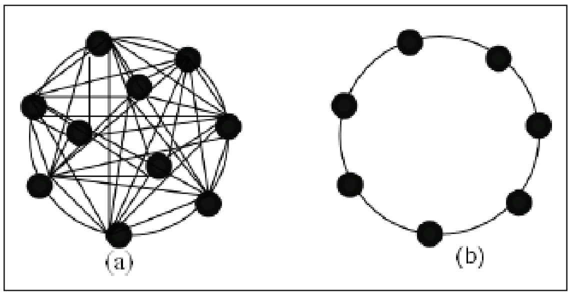
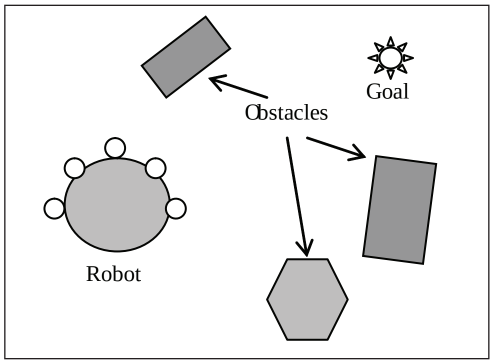


Figure 8. An instance of the motion planning problem



of each individual (particle) is influenced by the best individual found in the entire swarm. It is assumed that *gbest* swarms converge fast, as all the particles are attracted simultaneously to the best part of the search space. However, if the global optimum is not close to the best particle, it may be impossible for the swarm to explore other areas and, consequently, the swarm can be trapped in local optima (Kennedy & Mendes, 2002) In the *lbest* swarm, each individual is influenced by a smaller number of its neighbors (which are seen as adjacent members of the swarm array). Typically, *lbest* neighborhoods comprise of two neighbors with one on the right side and one on the left side (a ring lattice). This type of swarm will converge slower but can locate the global optimum with a greater chance. An *lbest* swarm is able to flow around local optima, with subswarms being able to explore different optima (Kennedy and Mendes, 2002). A graphical representation of a *gbest* swarm and an *lbest* swarm respectively is depicted in Figure 7 according to Kennedy and Mendes (2002).

Watts (Watts & Strogatz, 1998) introduced the small-world network model which allows interpolating between regular low-dimensional lattices and random networks, by introducing a certain amount of random long-range connections into an initially regular network (Dall'Asta, Baronchelli, Barrat, & Loreto, 2006). Starting from here, several models have been developed; icing model (Barrat & Weight, 2000), spreading of epidemics (Moore & Newman, 2000), and evolution of random walks (Jasch & Blumen, 2001) are some of them.

PSO FOR ROBOT MOTION PLANNING

A robot is a controlled manipulator capable of performing complex task in the real world. Consider the motion-planning problem of a wheeled mobile robot with ultrasonic transducers to determine the range of the obstacles in different directions around the robot. Figure 8 illustrates an instance of the motion planning problem, where the big circle denotes the robot and the circles around the big circle denote the ultrasonic transducers. The problem here is to determine a path for the robot to reach the goal position from a given starting position without touching any obstacle. Typically there are two types of planning called the local and the global planning. In case of local planning, the robot determines its next position and moves there. This is one cycle of planning. To reach the goal position, the robot needs to execute several cycles of planning. We in this chapter present a scheme for local motion planning of a robot by using PSO algorithm.

In order to efficiently use PSO in this application, we need to transform the motion planning problem into a search problem. Since we need the robot to traverse minimum distance in its journey from the starting point to the goal point, we can use the Euclidean distance metric between the current position of the robot and the goal position as a fitness function for the particles in the PSO algorithm. Let (x_c, y_c) be the current position and (x_g, y_g) be the goal position of the robot. Thus, the Euclidean distance describing the fitness function may be given by:

$$f = \sqrt{(x_c - x_g)^2 + (y_c - y_g)^2} \qquad (7)$$

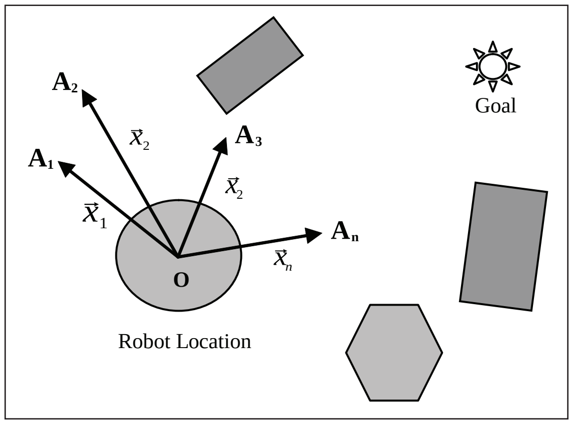
Readers, however, may wonder that on many occasions, this distance cannot truly measure the distance to be left for the robot in subsequent motion planning. This is because of the phenomenon that the positions (x_c, y_c) be such that the straight line joining (x_c, y_c) and (x_g, y_g) may have an overlap with one or more obstacles. To make sure that the particles are not befooled to traverse the obstacles, we can add a penalty factor to the previously constructed fitness function. Suppose the total length of the straight line overlapped with the obstacles is d and the distance between (x_c, y_c) and (x_g, y_g) is D. then we can construct a penalty factor

$$(1 - \frac{d}{D})^{-1}$$

and use it as a factor to the fitness function previously constructed. How does this factor influence the particles to avoid the obstacles? This is because of the phenomenon that the fitness of the particles is discounted by the factor

$$(1 - \frac{d}{D})^{-1}$$

Figure 9. Orientation of the particles at the initial phase



once the factor starts moving over an obstacle. Naturally, the velocities of those particles may be reduced in the PSO algorithm. It is apparent from our discussion that the particles in the present context are simply two-dimensional points (x, y). The initialization phase of the algorithm can be represented here by a collection of particles directed in different orientations at the current position of the robot. Figure 9 describes this phenomenon.

Let the current position of the robot be at O and the particles are represented by vectors **OA**$_1$, **OA**$_2$…with vector tips denoted by, $\vec{x}_2$,…, $\vec{x}_n$ with corresponding vectors $\vec{v}_i$ for the i-th particle.

Consider a population of 20 particles in the present context, that is, n = 20 here. Now in the first iteration of the algorithm, we evaluate the fitness of each i-th particle ($\vec{x}_i$) by,

$$f_i' = [\sqrt{(x_c - x_g)^2 + (y_c - y_g)^2}].(1 - \frac{d}{D})^{-1} \qquad (8)$$

We then add up the velocity and position of the particles by the basic two PSO equations as in equation (6). Results of the computer simulation for this problem has been shown in Figure 10 for a more difficult robot world map. The optimal path as found out by PSO algorithm has been marked in red color.

The question that remains is how to effectively generate coordinated motions of a group of robots when the information about the environment, such as the states of other robots, is not available or too costly to obtain. This issue is particularly relevant if we are to develop group robots that can work collectively and adaptively on a common task, even though each robot can only sense and hence react to its environment locally. In what follows we demonstrate how PSO can be applied to modify local motion strategies of individual mobile robots without the aid of any centralized modeling or control except for a high level criterion for measuring the quality of collective task performance.

In the present context, we assume that the robots have limited sensing, communication, and onboard computation capabilities. More specifically:

1. **Sensing:** The group robots are capable of distinguishing which is the box to be pushed and which is the group mate (e.g., through color detection). Due to the limitation in sensing range the robot will not be able to detect the exact shape of the box.
2. **Communication:** Although there is no scope of cross communication between the group robots, each robot can establish a communication channel with a remote evolutionary agent for receiving the next movement command.
3. **Onboard Computation:** All coordinated movement evaluation and selection are handled by a remote agent. The onboard computation of a robot is responsible only for its motor-level control, local sensing, and communication.

Figure 11 demonstrates a contact pushing force on a cylindrical box that leads to its motion. Each robot i exerts a vector force F_i making some angle β_i with the straight line connecting the CG of the box and the goal right at their point of contact. If the line of action of the box does not pass through the CG of the box, then there will be a torque acting on the box due to each of the robots given by:

Figure 10. Simulation result for the PSO based motion planning algorithm

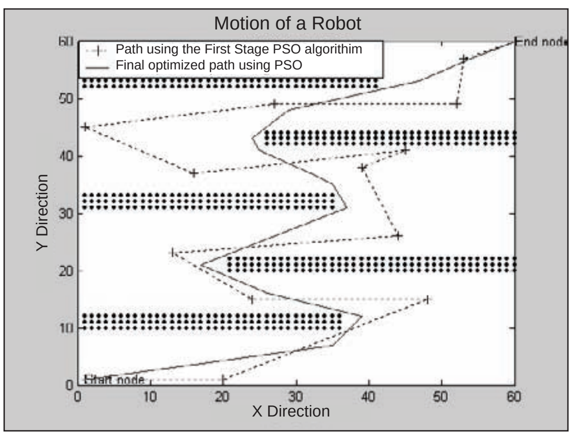
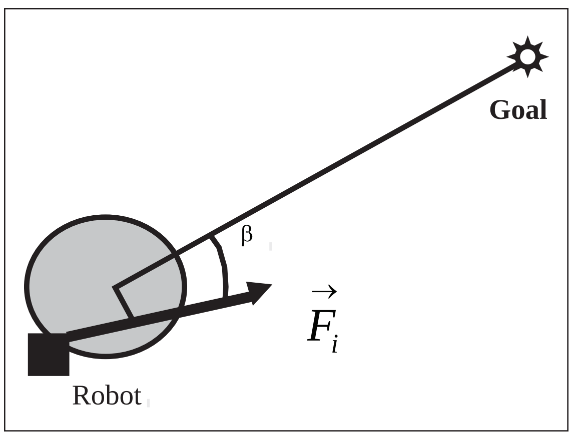


Figure 11. A single robot applying force and torque on the box through point contact



$$J_i = \pm |\vec{F}_i| . l_i \tag{9}$$

with l_i being the distance from the center of mass to the direction of the pushing force.

In this case we measure the performance of a group behavior using the fitness function to be maximized:

$$S_f = s_1 . s_2 . s_3 \tag{10}$$

$$s_1 = \sum_{1_i=1}^{3} F_i \cos \beta_i \tag{11}$$

$$s_2 = \left\| \sum_{i=1}^{3} J_i \right\|^{-1} \tag{12}$$

$$s_3 = 1 + \cos\Theta \tag{13}$$

where Θ is the direction of the net contact force by all the three robots. In the above definition of the fitness function S_1 implies that the robots must have to maximize the projection of the pushing forces along the shortest possible path between the box and the goal. S_2 implies that the rotation of the box should be discouraged during collective pushing, and S_3 measures how much the block moves correctly along the desired goal direction.

Now, while applying the PSO algorithm, each particle is defined for a 10 dimensional hyper-space where the position vector of each particle is shown in Box 1.

For All experiments we set C_1=C_2= 1.494 and ω=0.729 in equation (6) for PSO.

Figure 12 below illustrates through three snapshots the movement steps of three robots pushing a cylindrical box towards a desired goal

Box 1. The position vector of each particle

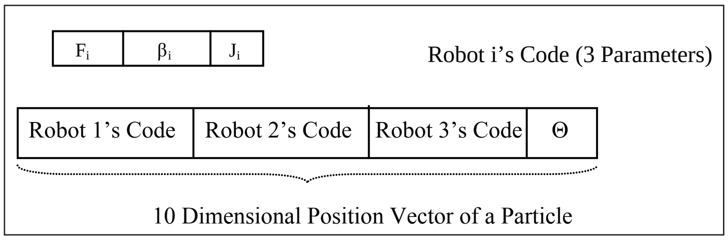
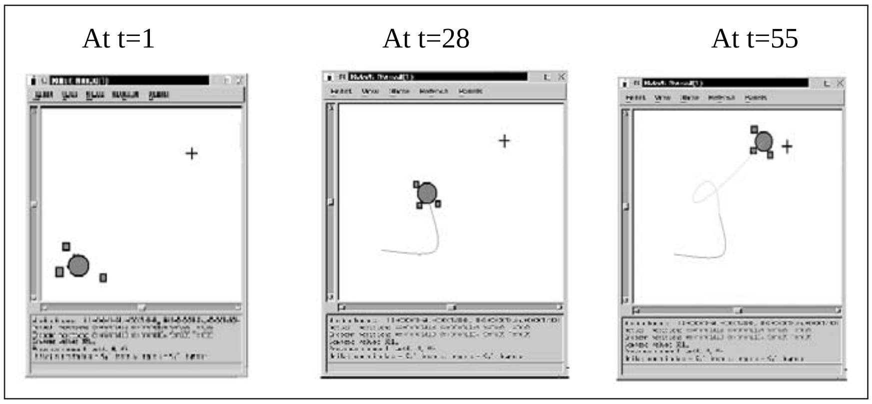


Figure 12. Three snapshots of collective box pushing by three mobile robots



location (marked by + symbol in the simulation window). Snapshots at more regular intervals cannot be given due to the space consideration. At the beginning of box-pushing group robots move in a rather chaotic manner as the positions of the particles in a population are randomly initialized.

In all experiments performed under the scope of this study, we note that PSO is a very fast and robust tool for selection of local movement strategies of the group robot during the collective box pushing. In both the problems, collective movement of all the three robots stabilize quickly after about 20 generations.

APPLICATION OF PSO IN PID TUNING

In industrial process control proportional-integral-derivative (PID) type controllers are widely being used. Let u(t) be the controller output and e(t) be the error signal, evaluated by taking the difference of the feedback process output from the reference input.

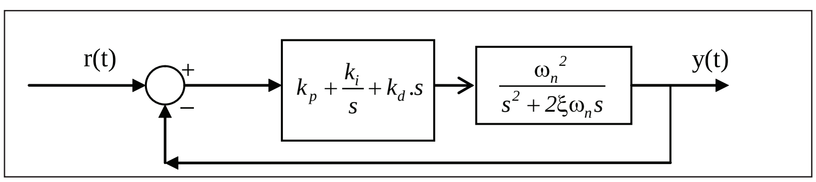
$$u(t) = K_p e(t) + K_i \int e(t)dt + K_D de/dt \qquad (14)$$

where K_p, K_D, and K_i denote the proportional gain, derivative coefficient, and integral coefficient respectively. To keep the controller response linear with the signed error, the proportional gain term is used in the above expression. To speed up the controller response, the integral action is needed, and the integral constant K_i is selected judiciously, so as to limit the excessive overshoot of the process response. The derivative action is an anticipatory control over the integral action, and is needed to reduce the percentage overshoot in the process response and thereby reducing the settling time (Kuo, 1976) for the process. The derivative action is also needed to enhance the system inertia when load fluctuation takes place. In industrial plants, usually the controller starts with a proportional action, followed by an integral action and derivative action in order. The time delays between each two control actions depend on the inertia of the process response.

For most commercial processes, Ziglar Nichol's chart is commonly used to select the coefficients K_p, K_D, and K_I uniquely. However, it has been noted that a fixed choice of K_p, K_D, and K_I degrades the controller performance especially when load fluctuation takes place. Researchers in control engineering thus prefer time varying coefficients of $K_p(t)$ $K_D(t)$, and $K_I(t)$. Usually, the current and the last few sample process responses/errors are used to determine the adaptation of the above time-varying coefficients.

In recent times, researchers are keen to employ evolutionary algorithms to autonomously adapt the above coefficients. For instance, McDonell in a recent review presents different schemes for tuning parameters of a control system using genetic algorithms (GA). In this section, we provide an alternative for adaptation of K_p, K_D, and K_I by invoking the PSO algorithm.

Figure 13. Block diagram of a second order closed loop system with a PID controller.



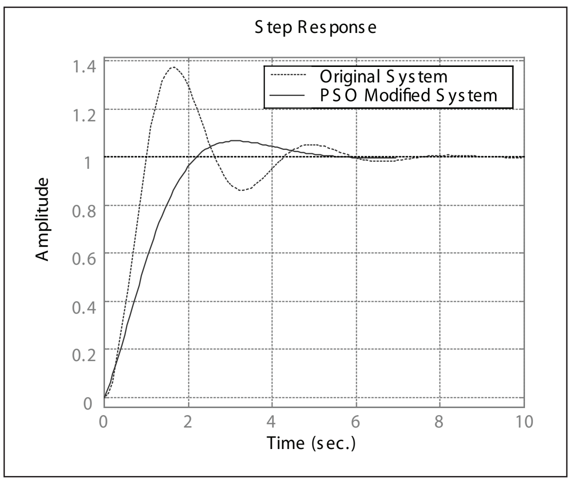
In this section, we briefly illustrate how the PSO algorithm can be used for tuning a PID controller. First we have considered the simple closed loop system given the open loop transfer function G(s) and the feed back gain of H(s) =1. From that we calculate the rise time. Now from the user we get the required value of rise time t_R. Transfer function for the PID controller is given by:

$$U(s) = K_p + K_i/s + K_d s \tag{15}$$

Figure 13 provides a scheme for controlling a second order system using a PID controller.

The design methodology adapted here is that we only specify the required rise time (T_R) of the second order system. The PSO optimization program then finds out one of the possible values of the coefficients K_p, K_i, and K_d so that the system has a desired rise time T_R. In this problem we are using PSO to tune the parameter K_p, K_d, and K_i of the PID controller such that the objective function $F = (t_R - t_r)^2$ is minimized where t_R=Desired rise time and t_r=Rise time are calculated by the PSO algorithm at k-th iteration.

Figure 13. Step response of a second order system with a PSO based PID controller



As an example, consider the open loop transfer function of a plant given by:

$$Y(s) = \frac{4}{s^2 + 1.2s}$$

Then for a unity feedback system, the closed loop transfer function will be given by:

$$C(s) = \frac{4}{s^2 + 1.2s + 4}$$

Now the original rise time of the system is t = 0.990 s. Suppose we would like to design a PID controller such that the desired rise time becomes t_r = 2.2 s. We run a small PSO algorithm comprising of 20 particles with the usual settings of parameters C_1=C_2=2.0 and ω = 0.749. The swarm tries to minimize the objective function $F = (t_R - t_r)^2$. The algorithm is run for 300 iterations. As a mean of 20 such independent runs of the algorithm, we get the transfer function of the required PID controller as $0.1168s^2 + 2.515s + 2.156$. Figure 14 provides the step responses due to the original system without any controller and the system controlled with a PID controller designed by PSO algorithm.

ANT COLONY SYSTEMS IN VEHICLE ROUTING AND ADVANCED LOGISTIC PROBLEMS

We can represent the well known vehicle routing problem (VRP) as a directed graph G = (V, A, d) where $V = \{v_0, v_1, v_2, ..., v_n)$ is a set of vertices and $A = \{(v_i, v_j) : i \neq j\}$ is a set of arcs. The vertex v_0 denotes the depot, the other vertices of V represent cities or customers, and the non-negative weights d_{ij}, which are associated with each arc (v_i, v_j) that represent the distance (or the traveling time or the traveling cost) between v_i and v_j. For each customer vi a non-negative demand q_i and non-negative service time δ_i is given. The aim is to find the minimum cost of the vehicle routes where:

- Every customer is visited exactly once by exactly one vehicle.
- All vehicle routes begin and end at the depot.
- For every vehicle route the total demand does not exceed the vehicle capacity Q.
- For every vehicle route the total route length (including service times) does not exceed a given bound L.

The VRP is a well known combinatorial optimization problem, and therefore no polynomial time solutions for such problem exist. In this section, we attempt to model the VRP as an ant system. The roots constructed by artificial ants will be are analogous to the selection of successive cities for the VRP. For the selection of a nonvisited city, we consider two aspects: (1) the degree or level of goodness about the choice of that city as the next destination; and (2) how promising is the choice of that city. The former aspect is measured in terms of the store pheromone trails τ_{ij} associated with an arc between nodes v_i and v_j. The later aspect is measured by a parameter called *visibility* denoted by η_{ij}. In VRP, visibility is usually defined as the reciprocal of the traversed distance. Symbolically, we define

$\eta_{ij} = 1/d_{ij}$ where d_{ij} denotes the distance between the cities v_i and v_j.

Given the set of cities $\Omega = \{v_j \in V : V_j$ is feasible to be visited$\} \bigcup v_0$ where v_j is selected to be visited after city v_i and v_0 is the starting city. We now provide a random proportional rule that includes the above two aspects for the selection of the next city in a probabilistic sense. The probability of selecting city j from city i is given by:

$$p_{ij}(t) = \frac{\left[\tau_{ij}(t)\right]^{\alpha} . \left[\eta_{ij}\right]^{\beta}}{\sum_{h \in \Omega} \left[\tau_{ih}\right]^{\alpha} \left[\eta_{ih}\right]^{\beta}}, \text{ if } v_j \in \Omega \qquad (16)$$

= 0, otherwise.

This formula includes parameters α and β that determine the relative influence of the trails and visibility in the selection of the next city j t city i.

To find a feasible solution we now increase the pheromone trail for the k-th ant by

$$\Delta\tau_{ij} = 1/L_k$$

where L_k denotes the objective value for the given problem. To determine the so far best solution, that is, to attain optimal objective value L*, we set

$$\Delta\tau_{ij} = 1/L^*$$

The ant that increases the pheromone trail following the last formula is called an elitist ant. The trail intensities thus are updated for m number of ants as follows:

$$\tau_{ij}^{new} = \rho\, \tau_{ij}^{old} + \sum_{k=1}^{m} \Delta\tau_{ij}^{k} + \sigma\, \Delta\tau_{ij}^{*}$$

where we presume σ number of elitist ants and ρ is a rate constant that controls the rate of increase in pheromone trails between nodes i and j. Given the number of customers, we select as many ants as the number of customers before the iteration of the VRP starts. The basic two steps of construction of vehicle routes and trail updates are then used recursively until no further change in trail intensity (τ_{ij}) is noted.

A traditional business model is articulated in three stages: production, distribution, and sales. Each one of these activities is usually managed by a different company, or by a different branch of the same company. Research has been trying to integrate these activities since the 60s when multiechelon inventory systems were first investigated (Clark & Scarf, 1960); but, in the late 70s, the discipline which is now widely known as supply chain management was not delivering what was expected, since the integration of data and management procedures was too hard to achieve, given the lack of real integration between the enterprise resource planning (ERP) and the enterprise data processing (EDP) systems (Sodhi, 2001). Only in the early 1990s did ERP vendors start to deploy products able to exploit the pervasive expansion of EDP systems at all levels of the supply chain. The moment was ripe for a new breed of companies, such as SAP, i2, Manugistics, and others to put data to work and start to implement and commercialize advanced logistics systems (ALS), whose aim is to optimize the supply chain seen as a unique process from the start to the end.

The first ALSs were the preserve of big companies, who could afford the investment in research and development required to study their case and to customize the application to interact with the existing EDP systems. Moreover, the available optimization algorithms required massive computational resources, especially for combinatorial problems such as vehicle routing. While ALSs were first deployed, researchers in the field of operational research were first investigating new “metaheuristics,” or heuristic methods that can be applied to a wide class of problems, such as ant colony optimization (ACO) (Bonabeau et al., 2000; Dorigo et al., 1996). Algorithms based on ACO are multiagent systems that exploit artificial stigmergy for the solution of combinatorial optimization problems; they draw their inspiration from the behavior of real ants, which always find the shortest path between their nest and a food source, thanks to local message exchange via the deposition of pheromone trails. The remarkable advantage of ACO based algorithm over traditional optimization algorithms is the ability to produce a good suboptimal solution in a very short time. Moreover, for some problem instances, ACO algorithms enhanced with local optimization capabilities have been proven to be the best overall (Gambardella & Dorigo, 2000).

Logistic Chain and the Work Flow Loop in a Company

Sales and distribution processes require the ability to *forecast* customer demand and to *optimally plan* the fine distribution of the products to the consumers. These two strategic activities, forecast and optimization, must be tightly interconnected in order to improve the performance of the system as a whole (Gambardella et al., 2001). In the work flow process of a company, the *sales department* generates new orders by contacting the customers (old and new ones) to check whether they need a new delivery.

New orders are then processed by the *planning department*, which, according to the quantity requested, the location of the customers, and the time windows for the delivery, decides how many vehicles to employ and computes the best routes for the delivery in order to minimize the total travel time and space. This task may be assisted by ACO algorithms solving the static vehicle routing problem (SVRP).

The vehicle tours are then assigned to the fleet, which is monitored by the *fleet operational control station*, which monitors the evolution of deliveries in real time. Here the use of ACO algorithms for solving the dynamic vehicle routing problem (DVRP) is needed. Finally, after vehicles have returned to the depot, delivery data are off-loaded and transferred back to the company database.

CONCLUSION

The chapter proposed modern approaches for handling some well-known problems in production management and engineering using swarm intelligence (SI). Many of the classical problems in production management are NP-complete, and thus no well known polynomial time solutions are amenable to deterministically handle these problems. The problems include robot motion planning, VRP, and PID tuning. Stochastic algorithms capable of performing real time optimizations have been employed to solve these problems. The stochastic selection process in the proposed algorithms has been motivated by the group behavior of social insects, the algorithms following the inventors fall in the class of swarm intelligence.

Significant progress in swarm intelligence algorithms have been observed during the last five years. Parameter selection in these algorithms is an important consideration and needs focused research efforts in future. Many interesting recommendations on parameter selection have been addressed elsewhere (Engelbrecht, 2005; Kennedy & Eberhart, 2001). On the other hand, the agents (like particles in PSO or ants in ACO) could be made more adaptive. Instead of tuning every parameter of a coordination mechanism, each agent should learn the best parameter setting for the application at hand. Future research should specify a generic extension of agents that operate in the production management system to support learning of agent parameters, and reinforcement learning may be an attractive candidate.

FUTURE RESEARCH DIRECTION

Results gained from the research of complex dynamical systems should be applied to support the designer in the tuning and evaluation process. The stability of emergent system-level features in the face of external changes and disturbances is an important issue in the evaluation of an application system. Self-organization often includes multistability and the behavior of the interacting individuals may be nonlinear. Hence stability analysis seems to be an important future research area for these multiagent swarms.

In recent times, synergism of swarm intelligence with other soft computing algorithms has opened up new avenues for the next generation computing systems. Engineering search and optimization problems including production management will find new dimensions in the light of hybridization of swarm intelligence with other algorithms.

REFERENCES

Barrat, A., & Weight, M. (2000). On the properties of small-world network models. *The European Physical Journal*, *13*, 547-560.

Beni, G., & Wang, U. (1989). *Swarm intelligence in cellular robotic systems.* Paper presented at the NATO Advanced Workshop on Robots and Biological Systems, Il Ciocco, Tuscany, Italy.

Bonabeau, E., Dorigo, M., & Theraulaz, G. (1998). *Swarm intelligence*. Oxford University Press.

Couzin, I. D., Krause, J., James, R., Ruxton, G. D., & Franks, N. R. (2002). Collective memory and spatial sorting in animal groups. *Journal of Theoretical Biology, 218*, 1-11.

Dall'Asta, L., Baronchelli, A., Barrat, A., & Loreto, V. (2006). Agreement dynamics on small-world networks. *Europhysics Letters.*

Dorigo, M. (1992). *Optimization, learning, and natural algorithms.* Unpublished doctoral dissertation (in Italian), Politecnico di Milano, Dipartimento di Elettronica, Milano, Italy.

Dorigo, M., Di Caro, G., & Gambardella, L. M. (1999a). Ant algorithms for discrete optimization. *Artificial Life, 5*(2), 137-172.

Dorigo, M., Di Caro, G., & Gambardella, L. (1999b). Ant colony optimization: A new meta-heuristic. In P. J. Angeline, Z. Michalewicz, M. Schoenauer, X. Yao, & A. Zalzala, (Eds.), *Proceedings of the congress on evolutionary computation* (Vol. 2, pp. 1470-1477). IEEE Press.

Dorigo, M., & Gambardella, L. M. (1996). A study of some properties of ant Q. In *Proceedings of* *the PPSN IV – 4th International Conference on Parallel Problem Solving From Nature*, Berlin, Germany (pp. 656-665). Springer-Verlag.

Dorigo, M., & Gambardella, L. M. (1997, April). Ant colony system: A cooperative learning approach to the traveling salesman problem. *IEEE Transactions on Evolutionary Computation*, *1*, 53-66.

Engelbrecht, A. P. (2005). *Fundamentals of computational swarm intelligence*. Wiley.

Fayyad, U., Piatestku-Shapio, G., Smyth, P., & Uthurusamy, R. *Advances in knowledge*.

Flake, G. (1999). *The computational beauty of nature*. Cambridge, MA: MIT Press.

Goldberg, D. E. (1989). *Genetic algorithms in search, optimization and machine learning.* Reading, MA: Addison-Wesley.

Grosan, C., Abraham, A., & Monica, C. (2006). Swarm intelligence in data mining. In A. Abraham, C. Grosan, & V. Ramos (Eds.), *Swarm intelligence in data mining* (pp. 1-16). Springer.

Haykin, S. (1999). *Neural networks: A comprehensive foundation*. Prentice-Hall.

Holland, J. H. (1975). *Adaptation in natural and artificial systems*. University of Michigan Press.

Jasch, F., & Blumen, A. (2001). Trapping of random walks on small-world networks. *Physical Review E, 64,* 066104.

Jerne, N. K. (1983). The generative grammar of the immune system. *EMBO Journal, 4*(4).

Kennedy, J. (1999). Small worlds and mega-minds: Effects of neighborhood topology on particle swarm performance. In *Proceedings of the 1999 Congress of Evolutionary Computation* (Vol. 3, pp. 1931-1938). IEEE Press.

Kennedy, J., & Eberhart, R. (1995). Particle swarm optimization. In *Proceedings of IEEE International conference on Neural Networks* (pp. 1942-1948).

Kennedy, J., & Eberhart, R. C. (2001). *Swarm intelligence*. Morgan Kaufmann Academic Press.

Kennedy, J., Eberhart, R., & Shi, Y. (2001). *Swarm intelligence*. Morgan Kaufmann Academic Press.

Kennedy, J., & Mendes, R. (2002).Population structure and particle swarm performance. In *Proceedings of the IEEE Congress on Evolutionary Computation (CEC)* (pp. 1671-1676).

Krause, J., & Ruxton, G. D. (2002). *Living in groups.* Oxford University Press.

Kuo, B. C. (1976). *Automatic control systems.* Englewood Cliffs, NJ: Prentice Hall.

Langton, C. G. (Ed.). (1995). *Artificial life: An overview*. Cambridge, MA: MIT Press.

McDonnell, J. R. (1997). Control. In T. Back, D. B. Fogel,, & Z. Michalewicz, (Eds.), *Handbook of evolutionary computation*. New York: IOP Oxford Publishing Ltd.

Milonas, M. M. (1994). Swarms, phase transitions, and collective intelligence. In C.G. Langton (Ed.), *Artificial life III.* Reading, MA: Addison Wesley.

Moore, C., & Newman, M. E. J. (2000). Epidemics and percolation in small-world networks. *Physics Review E, 61*, 5678-5682.

Partridge, B. L. (1982). The structure and function of fish schools. *Science American, 245*, 90-99.

Partridge, B. L., & Pitcher, T. J. (1980). The sensory basis of fish schools: Relative role of lateral line and vision. *Journal of Comparative Physiology, 135*, 315-325.

Serra, R., & Zanarini, G. (1990). *Complex systems and cognitive processes*. New York: Springer-Verlag.

Watts, D. J., & Strogatz, S. H. (1998). Collective dynamics of small-world networks. *Nature, 393*, 440-442.

ADDITIONAL READINGS

1. For an important application of ant systems in load balancing in telecommunication networks, the readers may go through an interesting paper by Schooerwoerd, Holland, Bruten, and Rothkrantz (1997).
2. Di Caro and Dorigo (1998) propose the 'AntNet' algorithm for computer networks. In their work, intelligent packets – *ants* – are introduced in the networks, which interact to keep the contents of the routing tables up-to-date.
3. An interesting application of ACO to job shop scheduling problems can be traced in the works of Ventresca and Ombuki (2004). They proposed a pheromone alteration strategy which improves the basic ant system by utilizing the behaviour of artificial ant. Experiments using well-known job shop problems show that this approach improves on the solution quality obtained by the basic ant system and is competitive with another recently proposed extension of the ant system, the MAX-MIN algorithm. Pasia, Hartl, and Doerner (2006) have recently reported an important application of pareto-ant systems in bi-objective flow shop scheduling.
4. For interesting applications of Particle Swarm Optimization in multiagent robotics the readers may go through the works by Pugh, Zhang, and Martinoli (2005) and Qin, Sun, Li, and Cen (2004).
5. It is very interesting to note that emulating the way ants find the shortest path to a new food supply has led researchers at Hewlett-Packard to develop software programs that can find the most efficient way to route phone traffic over a telecommunications network. Southwest Airlines has used a similar model to efficiently route cargo. To allocate labor, honeybees appear to follow one simple but powerful rule--they seem to specialize in a particular activity unless they perceive an important need to perform another function. Using that model, researchers at Northwestern University have devised a system for painting trucks that can automatically adapt to changing conditions. In an e-document released by Eric Bonabeau and Christopher Meyer (2001), the authors have explored the connections between the paradigm of swarm intelligence and the future enterprises. In the future, the authors speculate, a company might structure its entire business using the principles of swarm intelligence.

Bonabeau, E., & Meyer, C. (2001, May). Swarm intelligence: A whole new way to think about business. *Harvard Business Review*, 107-114.

Di Caro, G., & Dorigo, M. (1996a). AntNet: Distributed stigmergetic control for communications networks. *Journal of Artificial Intelligence Research*, *9*, 317-365.

Di Caro, G., & Dorigo, M. (1996b). *Discovery and data mining.* AAAI/MIT Press.

Pasia, J. M., Hartl, R. F., & Doerner, K.F. (2006, September 4-7). *Solving a bi-objective flowshop scheduling problem by pareto-ant colony optimization*. Paper presented at the ANTS 2006: Fifth International Workshop on Ant Colony Optimization and Swarm Intelligence, Bussels.

Pugh, J., Zhang, Y., & Martinoli, A. (2005, June). *Particle swarm optimization for unsupervised robotic learning.* Paper presented at the Swarm Intelligence Symposium (pp. 92-99), Pasadena, California.

Qin, Y., Sun, D., Li, N., & Cen, Y. (2004, August). Path planning for mobile robot using the particle swarm optimization with mutation operator. In *Proceedings of the 3rd International Conference on Machine Learning & Cybernetics*, Shanghai, China (pp. 2473-2478).

Schoonderwoerd, R., Holland, O., Bruten, J., & Rothkrantz, L. (1997). Ant-based load balancing in telecommunication networks. *Adaptive Behavior, 5*(2).

Ventrescam, M., & Ombuki, B. (2004). Ant colony optimization for job shop scheduling problem. In *Proceeding (451) Artificial Intelligence and Soft Computing.*

Chapter XIX
Artificial Neural Network and Metaheuristic Strategies:
Emerging Tools for Metal Cutting Process Optimization

Indrajit Mukherjee
Bengal Engineering and Science University, India

Pradip Kumar Ray
Indian Institute of Technology Kharagpur, India

ABSTRACT

Application of optimization tools and techniques is necessary and an essential requirement for any metal cutting-based manufacturing unit to respond effectively to severe competitiveness and increasing demand of quality product in the global market. However, both problem types and techniques employed are diverse. Often the context of the problem involves building nonlinear inferential response surface model(s) of the process(s), and then determine levels of inputs and in-process parameters that result in best (or significantly improved than existing) measures of process quality improvement and effectiveness. Selecting the appropriate levels or settings of inputs and in-process variables is a typical example of desired process effectiveness. However, determination of optimal process conditions, using appropriate solution methodology through cost-effective inferential nonlinear response surface model(s) is a challenging and continual research endeavour for researchers and practitioners. In this context, artificial neural network (ANN) and metaheuristic strategies, such as genetic algorithm (GA), simulated annealing (SA), and tabu search (TS), either in its original form or its variant, has been shown to yield promising outcomes for solving nonlinear response surface optimization problems in metal cutting process(s). The goal of this chapter is to assess the status and scope of artificial neural network-based inferential model, GA, SA, and TS-based metaheuristic search stategies in metal cutting processes. Subsequently, a solution methodology for nonlinear response surface optimization in metal cutting processes is proposed for the

benefits of selection of an appropriate technique. Specific application in a multiple response grinding process optimization problem using ANN, real-valued genetic algorithm, simulated annealing, and a modified tabu search is also provided for a clearer understanding of the settings, where the proposed methodology is being used.

INTRODUCTION

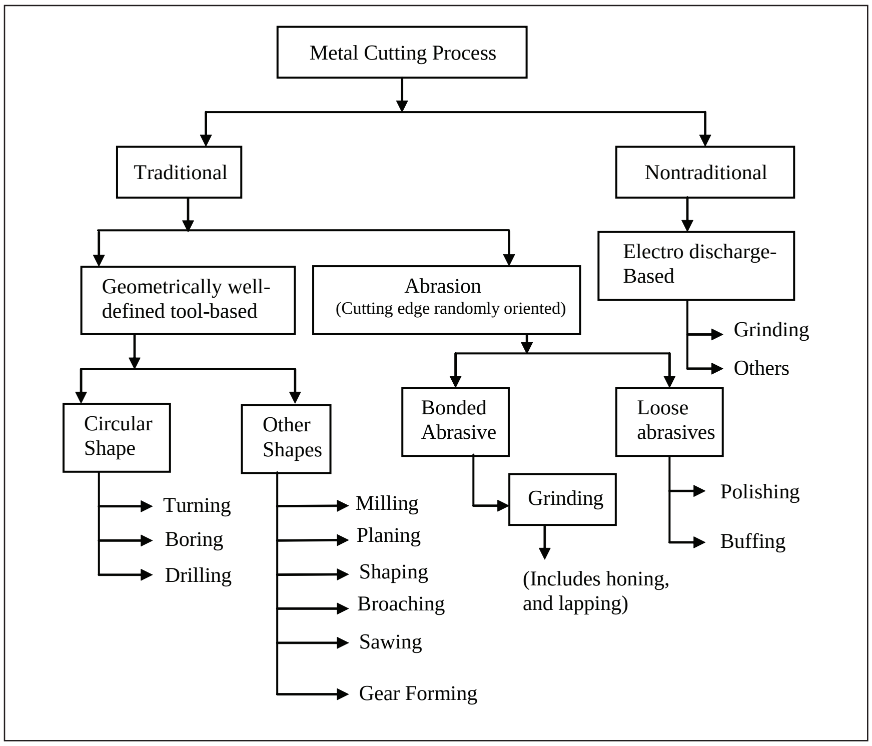
Modelling an abrasive metal cutting process and determining its optimal process conditions with respect to a particular component or part with a desired quality to be manufactured are the two main objectives in the context of process parameter optimization. These two objectives are to be achieved in two successive stages. In the first stage, modelling of the cutting process is to be carried out. In the second stage, an appropriate optimization methodology needs to be developed and used. However, modelling and optimization is considered to be a formidable and challenging task in almost all situations, particularly in a complex abrasive metal cutting process (Chen & Kumara, 1998; Krajnik, Kopac, & Sluga, 2005; Petri, Billo, & Bidanda, 1998; Sathyanarayanan, Lin, & Chen, 1992; Shaji & Radhakrisnan, 2003; Shin & Vishnupad, 1996).

In this context, in order to produce a wide variety of components or parts having different shapes, sizes, and surface texture, a large number of metal cutting processes have been developed. These processes are classified into traditional and nontraditional categories (El-Hofy, 2005) on the basis of the types of machining and machine tools or mechanical tool(s) or abrasives used. The metal cutting processes are grouped into several categories as shown in Figure 1. In a typical metal cutting process, materials are removed by plastic deformation of the work materials by harder tool materials, such as *SiC* and diamond. However, producing shapes of the tools with a hard material may be uneconomical because of tool wear. Grains or grits or abrasives are the only viable alternatives in such a situation (Farago, 1976; Ghosh & Mallik, 1999). As the parts and components are hardened to extend their service life, abrasive cutting is generally the only way that can readily cut hardened materials to achieve required dimensional accuracy and surface texture (Waters, 1996).

The use of abrasives, which are bonded in wheels, for machining a high quality part or component at low cost represents one of the important and emerging areas of metal cutting (Farago, 1976; Shin & Vishnupad, 1996). Grinding, a bonded-abrasive based metal cutting process, is the only possible practical and economic means of shaping parts into finished products with required surface finish, acceptable surface integrity, and high geometric accuracy (Feng, Wang, & Yu, 2002; Maksoud & Atia, 2004). In case of grinding, metal is removed in the form of small chips produced by plowing, rather than conventional cutting mechanism, using geometrically undefined abrasive cutting edges (El-Hofy, 2005; Farago, 1976; Waters, 1996). The term "grinding" has been defined as removing either metallic or other materials by the use of a solid grinding wheel, and includes processes such as honing and lapping (King & Hahn, 1986). In a typical grinding process, a hard bonded abrasive surface is pressed against a workpiece, resulting in the removal of material from both the work piece and the abrasive. Material removal rate in a typical grinding process is usually low, with chip thickness varying between 0.002 to 0.05mm (Stephenson & Agapiou, 1997). Whereas, for turning, drilling, boring, and milling, chip thickness is generally greater than 0.05mm (King & Hahn, 1986; Stephenson & Agapiou, 1997).

Inferential modeling and optimization have been conceived as important means to control and improve the complex grinding process, which attempts to mimic the jobs earlier performed by human workers based on their knowledge and experience (Gupta, Shishodia, & Sekhon, 2001;

Figure 1. Classification of metal cutting processes



Shin & Vishnupad, 1996). Application of such tools and techniques may lead to appropriate selection of process conditions, ensure stable processes, and provide required dimensional precision and surface texture. However, considering the inherent complexity of a grinding process, modeling and optimization remains one of the most critical and difficult tasks for researchers and practitioners (Luong & Spedding, 1995; Shin & Vishnupad, 1996). As has been mentioned in the literature, modelling and optimization of grinding process is a challenging problem mainly because of the following reasons:

1. Unlike in many conventional machining processes, where cutting is performed by a geometrically well-defined cutting edge(s), grinding is performed by a number of sharp edge abrasive particles, which are randomly distributed in the cutting area, and also randomly oriented in a grinding stone (Ghosh & Mallik, 1999). It is in this context, the process of material removal is random in nature, and it is too difficult to maintain close surface finish or control dimensional accuracy of hard particles, which affect the cutting process (Chen & Kumara, 1998).
2. There is a need for comprehensive inferential grinding model relating various input variables, in-process parameters, and multiple responses to all process conditions (Lee & Shin, 2000).
3. Research works on modelling and optimization of grinding process is fragmented (Chen & Kumara, 1998). Attempts to utilize all available and proposed techniques by process engineer(s), for a specific grinding problem, may be cumbersome

and practically infeasible. A generic and easy-to-implement framework for process engineers, which compare and contrast the limitations, potentials, and scope of application of various tools and techniques in metal cutting process is lacking. Lack of in-depth knowledge may lead to improper selection of modelling and optimization tools and techniques.

4. Experimentation at different control conditions of grinding process may lead to optimal process conditions or settings. However, in many kinds of production systems (Coit, Jackson, & Smith, 1998; Dabade, 1997; Petri et al., 1998), experiments based on the concept of statistical design of experiment and its variants, such as factorial and fractional factorial design (Montgomery, 2001), and response surface design, Taguchi methods, and evolutionary operations (Box & Draper, 1987b) may not be feasible. Sequential experimentation in such situations can be expensive, time consuming, and may eventually lead to delay or interruption of process. Furthermore, experimentation conditions may be biased toward certain procedures and operators, and as such, may not include the effects of actual production environment (Coit et al., 1998).
5. A typical grinding process may have multiple response characteristics. The exact number of response characteristics to be optimized simultaneously may vary depending on specific application, and other customer requirements (internal or external). Modelling and optimization may become more difficult if dependency (correlation) between responses is significant, as has been the case in many real life problems (Carlyle, Montgomery, & Runger, 2000; Chiao & Hamada, 2001; Kim & Lin, 2000).

In view of the abovementioned critical features, and complexity of the issues involved in grinding operation, the scope of this chapter is specific to the field of grinding process parameter optimization. Grinding process optimization may be considered as a two-step problem. At the first step, interrelationship between identified inputs, in-process parameters, and outputs needs to be determined using appropriate modelling technique(s). At the second step, using the inferential model function(s), and considering other process and variable constraints, the overall optimization problem is first formulated. Subsequently, an appropriate optimal solution technique (algorithm or strategy) attempts to determine the best (or significantly better than existing) process input conditions, which maximizes (or minimizes) the overall objective function(s) of the formulated problem.

Inferential Modelling of Grinding Process

The first necessary step for ensuring improvement in the performance of a grinding process is to understand the principles governing the metal cutting process(s). Model (explicit inferential type) based on their principals may be of two types: mechanistic and empirical (Box & Draper, 1987a). The functional relationship between the outputs, inputs, and in-process parameters as determined analytically (or theoretically) for a cutting process is referred to as "mechanistic model." Ideally, the mechanistic models as proposed by researchers (Ghosh & Mallick , 1999; Maksoud & Atria, 2004; Moulik, Yang, & Chandrasekhar, 2001), are derived from the basic physical principles and based on understanding about the kinds of relationships and their characteristic features by the analysts. However, as the physical interrelationships for complex grinding operation cannot be easily and accurately defined (Maksoud & Atria, 2004), mechanistic models are unavoidably limited by their underlying assumptions. Furthermore, the mechanistic models may provide only a partial description of the grinding process, and there is a lack of adequate and acceptable analytical model for varied metal cutting situations (Luong & Spedding, 1995; Maksoud & Atria, 2004). Application of empirical modelling technique(s) may be a viable alternative under such situations.

As has been found in the literature (Mukherjee & Ray, 2006b), there have been numerous applications of empirical modelling techniques based on statistical regression (Montgomery & Peck, 1992), artificial neural network (Fu, 2003), and fuzzy set theory (Zadeh, 1965). Although these types of modelling techniques may be working satisfactorily in varied problem situations, they are not devoid of certain assumptions and shortcomings, which bound the use of a technique in specific domain of a problem (Mukherjee & Ray, 2006b).

Optimal Solution Techniques for Grinding Process

With time, complexity in grinding process mechanism has increased and thereby problems related to determination of optimal or near-optimal cutting condition(s) are faced with discrete and continuous parameter spaces with multimodal, differentiable, as well as nondifferentiable objective function(s) or response(s) (Deb, 2002; Stephenson & Agapiou, 1997). Search for optimal or near-optimal solution(s) by suitable optimization methodology is a critical and difficult task (Chen & Tsai, 1996; Hui, Leung, & Linn, 2001). In this context, a large number of tools and techniques have been developed to solve parameter optimization problems, which may be further classified as conventional and nonconventional approaches (Mukherjee & Ray, 2006b).

Whereas, conventional techniques attempt to provide an exact optimal solution to the problem, a nonconventional technique determines near-optimal (approximate) solution(s). Conventional optimization techniques include statistical design of experiment, with Taguchi methods and response surface methodology (RSM), iterative mathematical search techniques, such as linear, nonlinear, and dynamic programming (LP, NLP, and DP, respectively). Nonconventional metaheuristic techniques include genetic algorithm (GA), tabu search (TS), and simulated annealing (SA), which are sufficiently general and extensively used by researchers in recent times. Although, conventional and nonconventional optimization techniques may work satisfactorily in varied problem situations, they are based on a set of assumptions and shortcomings, which may restrict their use in a specific domain of a problem, and universal acceptance of such techniques ensuring global optimal solution in different kinds optimization problems can be debated (Mukherjee & Ray, 2006b; Voβ, 2001).

In view of the several shortcomings of the existing approaches and methodologies used for abrasive metal cutting process modelling and optimization as mentioned, there is a need for developing a better approach and methodology for its modelling and optimization. Considering the abovementioned criticalities, the important research issues are:

1. In many industrial situations, designed experimentation may be costly and time consuming. Experimentation in a production line may result in many types of interruptions in the production process, and nonavailability of production lines to process required parts and components (Coit et al., 1998). Moreover, the optimal conditions as determined through off-line experimentation might substantially differ from the actual condition. These differences result in working of the process at a condition that may be significantly deviated from optimal or near-optimal cutting conditions. Under such circumstances, an appropriate solution methodology for the optimization problem needs to be developed.
2. The quality of a typical finished product is generally a desirable combination or com-

posite of a family of properties, which can often be interacting or correlated, and nearly always measured in a variety of units (Harrington, 1965). The desired combination of response properties can be said as a measure of the degree of customer satisfaction (Kim & Lin, 2000). Conventional approach using response surface design methodology is limited to consideration of a maximum of two responses at a time (Carlyle et al., 2000), and a Taguchi method for simultaneous optimization of multiple responses is more of judgmental and qualitative in decision making in such situations (Phadke, 1989). As has been mentioned in the literature, simultaneous optimization of more than two responses, balancing one against the other, is a major issue of concern for researchers and practitioners.

It is in this context that the objective of this chapter is to develop a comprehensive solution methodology for optimization problem in correlated and nonlinear multiresponse abrasive metal cutting process using artificial neural network and metaheuristic strategies, and show the usefulness of the proposed methodology in a representative abrasive metal cutting process under different conditions and constraints.

BACKGROUND

With time, as complexity in dynamics of cutting processes increased substantially, researchers and practitioners have focused on various mathematical modelling and optimization techniques to determine optimal or near-optimal cutting condition(s) with respect to various objective criteria and constraints (Tan & Creese, 1995). However, there exists no universal input-output and in-process parameter relationship inferential model, which is applicable to all kinds of metal cutting processes, and can predict cutting behaviour over a wide range of cutting conditions (Hassan & Suliman, 1990; Luong & Spedding, 1995). In addition, optimization techniques also have certain assumptions and limitations for implementation in real-life cutting process problems (Carlyle et al., 2000; Dabade & Ray, 1996; Mukherjee & Ray, 2006b; Osborne & Armacost, 1996; Youssef, Sait, & Adiche, 2001). In this section, an attempt has been made to review critically the existing and frequently used artificial neural network-based inferential models and metaheuristic search-based optimization techniques, specific to metal cutting processes.

Artificial Neural Network Models

Modelling techniques of input-output and in-process parameter relationship using ANN offering a distribution-free alternative have attracted attention of practitioners and researchers alike in manufacturing when faced with difficulties in building empirical nonlinear inferential models in metal cutting process control. These techniques may offer a cost effective alternative in the field of machine tool design and manufacturing approaches, receiving wide attention in recent years.

ANN may handle complex input-output and in-process parameter relationship of machining control problems. The learning ability of nonlinear relationship in a cutting operation without going deep into the mathematical complexity, or prior assumptions on the functional form of the relationship between input(s), in-process parameter(s), and output(s) (e.g., linear, quadratic, higher order polynomial, and exponential) makes ANN an attractive alternative choice for many researchers to model cutting processes (Petri et al., 1998; Zhang & Huang, 1995). Being a multivariable, dynamic, nonlinear estimator, it solves problems by self-learning and self-organization (Fu, 2003). Coit et al. (1998) consider practical aspects of building and validating ANN models.

Several applications of ANN-based input-output relationship modelling for metal cutting processes are reported in the literature. Back propagation neural network, proposed by Rumelhart, Hilton, and Williams (1986), have been successfully applied by Sathyanarayanan et al. (1992), Jain, Jain, and Kalra (1999), and Feng et al. (2002) for modelling a typical creep feed super alloy-grinding, prediction of material removal rate and surface finish parameter of a typical abrasive flow machining, and a honing operation of engine cylinder liners, respectively. There are certain assumptions, constraints, and limitations inherent in these approaches, which may be worth mentioning. ANN techniques are attempted only when regression techniques fail to provide an adequate model. Some of the drawbacks of ANN techniques are: (1) model parameters may be uninterpretable for nonlinear relationship; (2) it is dependent on voluminous data sets, as sparse data relative to a number of input and output variables may result in over-fitting or terminate training before network error reaches optimal or near-optimal point (Coit et Al., 1998; Howard & Beale, 1998); and (3) identification of influential observations, outliers, and significance of various predictors may not be possible by this technique. There is always an uncertainty in finite convergence of the intrinsic algorithms used in ANN-based modelling technique, and generally, convergence criteria are set based on prior experiences gained from earlier applications. No universal rules exist regarding choice of a particular ANN architecture for any typical metal cutting process problem.

Determination of Optimal or Near-Optimal Cutting Condition(s)

With time, complexity in metal cutting process dynamics has increased and as a consequence, problems related to determination of optimal or near-optimal cutting condition(s) are faced with discrete and continuous parameter spaces with multimodal, differentiable, as well as nondifferentiable objective function or response(s). A large number of techniques have been developed by researchers to solve these types of parameter optimization problems, and may be classified as conventional and nonconventional optimization techniques (Mukherjee & Ray, 2006b).

Metaheuristic Strategies

Nonconventional solution strategies, namely heuristics and metaheuristic, based on extrinsic model or resultant objective (fitness) function as developed is only an approximation of exact optimal, and attempt to provide best or significantly improved that existing cutting conditions. Heuristics, generally providing simple means of indicating which among several alternative solutions seems to be the most effective one in order to achieve some goal, consist of a rule or a set of rules seeking acceptable solution(s) at a reasonable computational cost (Voβ, 2001). Heuristic-based search techniques may be very useful for cases where conventional optimization techniques are not suitable, such as problems with high-dimensional search space with many local optima. Researchers and practitioners prefer alternative cost effective near-optimal (or approximate) solution(s) than exact optimal, as it may be extremely difficult to find exact optimal point in higher dimension and multimodal search space.

Although heuristic search may offer near-optimal solution(s), they are mainly problem-specific (de Werra & Hertz, 1983). Researchers suggest several alternatives to problem specific heuristics, also called generalized iterative master strategy or "metaheuristic" (Glover, 1986; Glover & Laguna, 2002), which guide and modify other heuristics to produce solutions that are normally generated in a quest for local optimality. As has been reported in the literature, three types of metaheuristic-based search algorithms, namely genetic algorithm (GA), simulated annealing (SA), and tabu search (TS), are applied in the domain of cutting process parameter optimization. These techniques are derivative-free, and are not based on functional form of relationship existing between response(s) and decision variables for its search direction. Each of these techniques is explained in the following sections along with their application potential and limitations.

Genetic Algorithm (GA)

The working of a GA (Deb, 2002; Goldberg, 2002), generally preferred for large and complex cutting process parameter optimization problems, is based on three basic operators, namely reproduction, crossover, and mutation, in order to offer a population of solutions. The algorithm creates new population from an initial random population (obtained from different feasible combination of process decision variables) by reproduction, crossover, and mutation in an iterative process. A GA is very appealing for single and multiobjective optimization problems (Deb, 2002), and some of its advantages are as follows: (1) it is not based on gradient-based information, and does not require the continuity or convexity of the design space; (2) it can explore large search space and its search direction or transition rule is probabilistic, and not deterministic in nature, therefore, the chance of avoiding local optimality is high; (3) it works with a population of solution points rather than a single solution point as in conventional techniques, and provides multiple near-optimal solutions; and (4) it has the ability to solve convex and multimodal function, multiple objectives, and nonlinear response function problems, and it may be applied to both discrete and continuous type of objective function(s).

Several applications of GA-based techniques in metal cutting process parameter optimization problems have been reported in the literature. Onwubolu and Kumalo (2001) propose a local search GA-based technique in multipass turning operation with mathematical formulation in line with work by Chen and Tsai (1996) with simulated annealing-based techniques. Krimpenis and Vosniakos (2002) use a GA-based optimization tool for sculptured surface computer numeric controlled (CNC) milling operation to achieve optimal machining time and maximum material removal. Chowdhury, Pratihar, and Pal (2002) apply a GA-based optimization technique for near optimal cutting conditions selection in a single-pass turning operation, and claim that GA outperforms goal programming techniques in terms of unit production time at all the solution points. Schrader (2003) illustrates the usability of a GA-based technique for simultaneous process parameter optimization in multipass turning operations.

Simulated Annealing (SA)

A SA technique (Aarts & Korst,1989; Kirkpatrick, Gelett, & Vecchi, 1983), based on the concept of modelling and simulation of a thermodynamic system, may be used to solve many combinatorial cutting process optimization problems. This technique starts with the selection of an initial random process decision vector, and moves to new neighbourhood decision vector that improves objective function value. A SA technique may accept inferior decision vector based on certain probabilistic measure to avoid local optimal in a multimodal response function. The probability that there is a move to an inferior decision vector (or the decision vector which provides degraded objective function value) decreases as the value of a “temperature parameter” defined in the algorithm, decreases, which is analogous with slow cooling in an annealing process to attain perfect crystalline state.

A number of different versions and applications of SA-based techniques in metal cutting process problems is reported in the literature. Chen and Tsai (1996) combine SA and Hooks-Jeeves pattern search techniques (Hooks & Jeeves, 1961) for optimizing cutting conditions in complex machining (multipass turning operation) to minimize unit operation cost. Chen and Su (1998) determine near optimal machining conditions for a continuous profile turning operation in CNC by using a SA algorithm, and claim that the algorithm delivers high-quality heuristic solution with reasonable computational requirements. For optimization of CNC turning process, Juan, Yu, and Lee (2003) apply SA-based techniques to attain optimal cutting conditions of high speed milling operation.

Tabu Search (TS)

A local search algorithm-based technique, called “tabu search” (*TS*), developed by Glover (Glover, 1989, 1990a, 1990b; Glover & Laguna, 2002) derives its attractiveness due to its greater flexibility and ease of implementation in combinatorial optimization problems. TS strategy starts with an initial feasible solution point (obtained from random feasible combination of process decision variables), and moves stepwise towards an improved solution point. A sample of decision space vectors in the neighbourhood of the current decision vector is generated, and the best vector within the sample is determined based on a heuristic approach. A move is made from current decision vector to a best decision vector not in tabu list, which provides improved objective function value in a single step by simple modifications of current decision vector. A tabu list contains a certain number of last decision vectors visited. The best decision vector replaces the oldest vector in the tabu list, and the survival vectors in the list are given a tabu active status, which reduces risk of cycling of same decision vector (i.e., modification in current decision vector, which would bring back previously visited vector). In subsequent iteration, uses of tabu active vectors are forbidden (so called ‘tabu moves’) for creating a sample of decision vectors in the neighbourhood of current decision vector space. Tabu active status of a decision vector is overridden only based on certain aspiration level criteria, such as acceptance of the modification on current vector that improves objective function value (de Werra & Hertz, 1983).

Although Kolahan and Liang (1996), while exploring the potentials of TS-based techniques for simultaneous decision making, attempt to minimize drilling cost by setting a number of machining parameters, such as machine cutting speed, tool travel, tool switch, and tool selection for a drilling operation in a plastic injection mould, there is hardly any report indicating an application of this technique for metal cutting process parameter optimization.

Although, metaheuristic may be considered to be a good solution approximation (near-optimal point) to solve complex combinatorial optimization problems within a reasonable amount of computational time as compared to search for exact optimal point, there are certain constraints and assumptions inherent in this techniques.

A few shortcomings of these techniques may be worth mentioning: (1) convergence of the metaheuristics are not always assured; (2) no universal rule exists for appropriate choice of algorithm parameters, such as population size, number of generations to be evaluated, crossover probability, mutation probability, and string length of GA; (3) convergence of SA may be strongly affected by the parameters of cooling schedule, and no universally acceptable levels of control parameters in cooling schedule exist for different types of cutting process parameter optimization problem; (4) selection of aspiration level criterion, tabu moves, neighbourhood generation, intensification, and diversification scheme plays a key roles in randomization of search to unexplored feasible state space, and effectiveness of TS strategy for continuous variable multidimensional state space; and (5) the repeatability of the near-optimal solution obtained by metaheuristics with same initial cutting conditions is not guaranteed.

ISSUES, CONTROVERSIES, PROBLEM AND SOLUTION METHODOLOGY

The working and performance (in terms of response quality) of a typical grinding process depends on the kind (capability) of grinding machine in use, the quality levels of the input materials and variables, and the settings of the machine parameters. In most real-life industrial environments, quality of a product is considered to be a composite of a family of properties, which can often be interacting or correlated with one another, and nearly always measured in a variety of units (Carlyle et al., 2000; Chiao & Hamada, 2001; Del Castillo, Montgomery, & McCarville, 1996; Harrington, 1965). Quality of a typical grinded product is seldom defined as a single response characteristic, and a composite of quality attributes for a grinded product needs to be expressed in quantitative terms. The desired combination of response properties or trade-off between responses is so-called "composite desirability" or "degree of customer satisfaction" (Kim & Lin, 2000). The trade-off represents some explicit compromise among conflicting conditions. Optimal combination rarely results in individual optimal of each response (Vining, 1998).

In this context, cost models for quality improvement are getting fewer acceptances from the view point of industrial practitioners, as it is difficult, if not impossible, to determine various cost elements present in actual manufacturing environment (Teeravaraprug & Cho, 2002). Therefore, the objective of study in many industrial-grinding situations is response(s) optimization, and it may be assumed that the cost incurred for experimentations and data collection in the short-term may be significantly less than long-term benefits achieved by response surface optimization.

In case of multiple responses, the problem is often called "multiple responses optimization" problem or "simultaneous optimization" problem (Carlyle et al., 2000). In such types of problems, one of the difficulties lies in the different properties of multiple responses having different types of optimality, such as nominal-the-best (NTB), dmaller-the-better (STB), and larger-the-better (LTB), and may be measured in varied units with substantial different magnitudes. Moreover, the multiple responses might not increase or decrease simultaneously, and a complex trade-off or compromise becomes essential.

Harrington (1965) proposes an analytical technique for optimizing multiple response, using an exponential function for transforming measure response into desirability scale "d" ($0 \leq d \leq 1$), and a geometric mean of "d" values determines overall desirability or composite desirability. For dual response system, Mayer and Carter (1973) propose a contour overlap plot, which maximizes a "primal response" function subject to condition that a constrained secondary response function takes on some specified or desired value. However, the idea of dual response approach for simultaneous modelling of mean and variance may not be appropriate for multiple quality characteristics problems (Xu, Lin, Tang, & Xie, 2004). In addition, choice of primal and secondary responses to be objective function is difficult when the number of responses is large. Derringer and Suich (1980) extend Harrington's technique by introducing more generalized transformation of responses into desirability measures. However, this approach does not consider the correlation of responses. Ignoring correlation may alter the form of desirability function, and jeopardize the determination of optimal process conditions. Khuri and Conlon (1981) present an approach in which generalized distance measure is employed to indicate the weighted distance of each response from individual optimal point. The variance-covariance of responses is used to determine the weights. Then the solution is found that minimizes the generalized distance. Kim and Lin (2000) propose a "*maximin*" approach using "minimum" operator for aggregating multiple responses, which can handle potential dependency between responses, and existence of dependency does not affect the proposed approach. The approach is easy to implement with little mathematical or statistical knowledge, having a basic physical interpretation. The "*maximin*" approach does not require any assumptions regarding form and degree of estimated response surface inferential models.

Considering the problem complexity as discussed, controversies and issues of various multiresponse optimization approaches, the methods adopted for converting multiple responses to single aggregate objective function, so-called "degree of customer satisfaction," is based on two approaches, namely, generalized transformation function as proposed by Derringer and Suich (1980), and "*maximin*" approach using a "minimum" operator as proposed by Kim and Lin (2000). In the following section, an overall nonlinear multiresponse grinding process optimization problem, having continuous multivariate state space is first mathematically formulated, and thereby overall near-optimal process conditions are determine that meet customer requirements as specified (primary objective). It is also to be ensured that the quality of the "degree of customer satisfaction" (Kim & Lin, 2000) of the process reaches an acceptable level (secondary constrained objective). In order to provide an optimal solution methodology, an integrated approach comprising artificial neural network-based nonlinear inferential model, desirability functions, "*maximin*" composite desirability approach, and metaheuristic search technique is proposed.

Problem Formulation

For the formulation of the problem, the following notations are used:

- s : particular grinding stage under consideration
- $p_{(s)}$: number of response variables considered at the s-th stage
- $X_{I(s-1)}$: input vector state space considered at the s-th stage
- $X_{P(s)}$: in-process parameter vector state space considered at the s-th stage
- $x_{(s)}$: set of input conditional state space at the s-th stage, consisting of $X_{I(s-1)}$ and $X_{P(s)}$
- $X_{s(i)}$: i-th component of multidimensional state space vector ($X_{(s)}$) of the input conditional state space for i = (1,…, m), where m is the total number of inputs and in-process variables at s-th stage.
- h_s : inferential abstract functional relationship between input, in-process parameters, and responses of the s-th stage
- g_c : common functional relationships to convert individual predicted response to scale free desirability measure, at the s-th stage
- $d_{j(s)}$: individual desirability measure of j-th response characteristic for the s-th stage of operation, based on g_c function for j = (1,…, $p_{(s)}$)
- m_c : common functional relationship to convert multiple $d_{j(s)}$ values of the s-th stage to a single composite desirability measure, so-called $\lambda_{(s)}$ for j = (1,…, $p_{(s)}$)
- $\lambda_{(s)}$: composite desirability measure for the s-th stage, based on $d_{j(s)}$ values and m_c function for j = (1,…, $p_{(s)}$)
- $\hat{Y}_{js}$: predicted j-th response at the s-th stage based on empirical model function h_s for j = (1,…, $p_{(s)}$)
- Y_{js} : observed j-th response at the s-th stage or $X_{I(i)}$ for j = (1,…, $p_{(s)}$)
- $X_{(s)}$: set of input state space at the s-th stage, $\{X_{I(s-1)}, X_{P(s)}\}$

The assumptions underlying for formulation of the problem are as follows:

1. The grinding process is assumed to be stable and under statistical control.

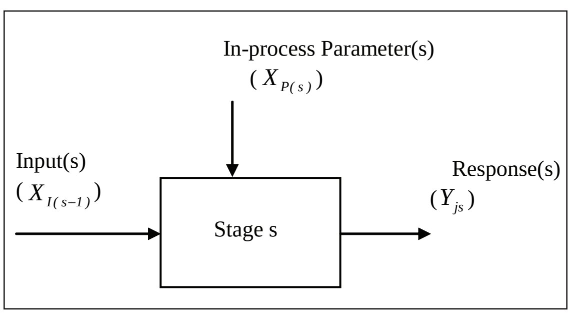
2. Both the input variables and the response characteristics as identified for the grinding process are assumed to be continuous random variables.
3. The random fluctuations in the behaviour of the grinding process are due to the presence of a number of nuisance factors only, which are uncontrollable.
4. The responses for a process may be correlated with one another.
5. The grinding process is assumed to be an isolated system and an independent one.
6. The grinding process outputs are independent of any influence of input conditions of previous stages

The set of assumptions as mentioned above are valid for all types of inferential models that have been developed and used. However, depending on a particular techniques used for modelling, the technique specific assumptions are also valid.

Inputs and In-Process Variables for Inferential Modelling

To develop an inferential response surface model, it is first necessary to identify various critical inputs and in-process parameters. The effect of change of input conditions to responses, their interdependency are typically well understood by skilled workers and technicians, knowledgeable and experienced in the process. What is least understood is the mathematical relationship or how these factors interact to determine desired multiple response quality characteristics. The development of an inferential model(s) leads to understanding of these interactive effects and attempts to capture the behavior of these input conditions to accurately predict response behavior. The input space vector of a typical grinding process consists of input variables, and in-process parameters. The input variables refer to diameter, ovality, and taper of cylinder liner bore, material properties (hardness and composition). Responses variables refer to finished diameter, ovality, and taper of cylinder liner bore, surface texture, and honing angle. In-process variables may be cutting pressure, cutting time, cutting oil temperature, and dog length (for honing process). A typical grinding process with its various inputs, in-process parameters, and responses at s-th stages are schematically presented in Figure 2.

Figure 2. Schematic diagram of the s-th stage grinding process with its inputs, in-process parameters and outputs



The number of inputs and in-process variables usually determine the dimension of the search space. The inferential model(s) provide the necessary functional input(s), to determine optimal or near-optimal process conditions. However, there may be numerous practical process and variable constraints for the underlying problem, which need to be considered.

Desirability Functions

The concept of desirability function, which maps the individual influence of each response into a single scalar response, have been proposed to be applied for modelling multiresponse cases (Derringer, 1994; Derringer & Suich, 1980; Harrington, 1965; Kim & Lin, 2000). The desirability function approach may be applied for getting a combined representation of the influence of several individual responses (either desirable or undesirable). The value of the desirability function ranges between 0 and 1. If the predicted response is same as customer specified desired response, the desirability value is assumed to be 1, and if the value of the response lies beyond its customer specified desired interval, the value of the desirability function is assumed to be 0. Using the concept of desirability, each predicted j-th response ($\hat{Y}_{js}$) at s-th stage is transformed to a scale free desirability value $d_{j(s)}$ $(0 \le d_{j(s)} \le 1)$. The desirability function is a scale-independent index, which enables quality characteristics to be compared to various units, and an effective means to simultaneously optimize multiple responses (Tong et al., 2001). An overall composite desirability measure is obtained by aggregating individual desirability values ($d_{j(s)}$). Derringer and Suich (1980) provide a systematic transformation scheme from $\hat{Y}_{js}(X_{(s)})$ to $d_{j(s)}$ based on function $g_c(\hat{Y}_{js}(X_{(s)}))$, and for nominal-the-best (*NTB*) type response it is expressed as:

$$d_{j(s)} = g_c(\hat{Y}_{js}(X_{(s)})) = \begin{cases} 0 & \text{if } \hat{Y}_{js}(X_{(s)}) < Y_{j(\min)} \\ & \text{or } (\hat{Y}_{js}\ X_{(s)}) > Y_{js(\max)} \\ \left[\dfrac{\hat{Y}_{js}(X_{(s)}) - Y_{js(\min)}}{T_{js} - Y_{js(\min)}}\right]^u & \text{if} \\ & (Y_{js(\min)} \le \hat{Y}_{js}\ X_{(s)}) \le T_{js} \\ \left[\dfrac{\hat{Y}_{js}(X_{(s)}) - Y_{js(\max)}}{T_{js} - Y_{js(\max)}}\right]^t & \text{if} \\ & T_{js} < \hat{Y}_{js}(X_{(s)}) \le Y_{js(\max)} \end{cases} \qquad (1)$$

where, $Y_{js(\min)}$ and $Y_{js(\max)}$ are the minimum and maximum bounds of j-th response at s-th stage. u and t are the exponential parameters that determine the shape of desirability function, and T_{js} is the desired target value of the j-th response at s-th stage. Using a composite desirability approach, a multiple response function may be transformed into a single objective or fitness function.

Composite Desirability or Dimensionality Reduction Strategy

This method converts a multiple response problem to a single scale-free aggregated objective function, which has been often defined as a composite desirability function (Jeong & Kim, 2003; Kim & Lin, 2000) sometimes referred to as "degree of customer satisfaction." This approach aims to identify the setting of the input conditions, which maximizes the composite desirability function. If all the quality characteristics reach their ideal values, the composite desirability is 1, and if any one of them does not reach its ideal value, the composite desirability is below 1. In this context, if any one of the responses does not meet customer requirements, the individual desirability is assumed to be 0, and overall composite desirability is also assumed to be *0*. The minimum of the desirability values of these responses is considered to be the value of the degree of customer satisfaction (Kim & Lin, 2000). The concept of "minimum operator" for aggregating the individual desirability ($d_{j(s)}$) is expressed as:

$$\lambda_{(s)} = \text{minimum}\{d_{1(s)}, d_{2(s)}, ..., d_{p(s)}\} \quad \text{for} \quad 0 \le \lambda_{(s)} \le 1 \qquad (2)$$

Some of the basic advantages (Kim and Lin, 2000) of selecting "minimum operator" composite desirability, over other existing dimension reduction approaches (Derringer & Suich, 1980; Harrington, 1965; Khuri & Conlon, 1980) are as follows:

1. Robust to the potential dependency (linear or nonlinear) between responses or existence of dependency between responses does not affect the proposed approach.

2. Provides a better balance between all the responses compared to other existing approaches.
3. The objective function expressed in term of $\lambda_{(s)}$ allows a good physical interpretation, and allows meaningful comparison between points in input conditional state space (Kim and Lin, 2000).
4. Easy to understand and implement with little mathematical knowledge.
5. Does not require any assumptions regarding the functional form or degrees of the estimated response models.

Process Constraint(s) or Requirement(s)

As optimal process conditions for a grinding process are to be determined taking into account several types of constraints related to the process and its variables, it is to be ensured that the value of the suggested composite desirability function reaches an acceptable level under process constraint or requirement. The set of process constraints at the *s*-th stage, denoted as, $\lambda_{(s)}$, is mathematically expressed as:

$$\lambda_s \geq \lambda_{sD} \tag{3}$$

where, λ_{sD} is the problem-specific minimum value of the acceptable degree of customer satisfaction measure of the process. The process constraint are referred to as the set of secondary objectives.

Formulation of Objective Criteria

Based on the inferential model function, predicted j-th response at s-th stage, $\hat{Y}_{js}$ is expressed as:

$$\hat{Y}_{js} = h_s[X_{(s)}] \quad \text{for j} = (1,\ldots, p_{(s)}) \tag{4}$$

where, $p_{(s)}$ is the number of responses at s-th stage, $X_{(s)}$ represents input conditions at s-th stage , which is a set of $X_{I(s-1)}$, and $X_{P(s)}$.

Individual predicted response are converted to scale-free desirability value by using desirability function (g_c) and expressed as:

$$d_{j(s)} = g_c(\hat{Y}_{js}) \text{ for j} = 1,\ldots, p_{(i)} \tag{5}$$

All the desirability values are converted into a single scale-free composite desirability value using an m_c function and expressed as:

$$\lambda_{(s)} = m_c(d_{j(s)}) \text{ for j} = (1,\ldots, p_{(s)}) \tag{6}$$

Therefore, an s-th stage multiple responses grinding process optimization problem may be mathematically stated as:

$$\underset{X}{Maximize} \ \ \lambda_{(s)} \quad \text{(primary objective)} \tag{7}$$

subject to

$$\lambda_s \geq \lambda_{sD} \quad \text{(secondary objective)} \tag{8}$$

$$\mathrm{X}_{I(s-1)\min} \leq \mathrm{X}_{I(s-1)} \leq \mathrm{X}_{I(s-1)\max} \quad \text{(input variable bounds)} \tag{9}$$

$$\mathrm{X}_{P(s)\min} \leq \mathrm{X}_{P(s)} \leq \mathrm{X}_{P(s)\max} \quad \text{(in-process parameter bounds)} \tag{10}$$

where, equation (8) represents specific process constraint or secondary objective. Equations (9) and (10) are variable state space bound. $\mathrm{X}_{I(s-1)\min}$ and $\mathrm{X}_{I(s-1)\max}$ are the minimum and maximum bounds of input variables, respectively. $\mathrm{X}_{P(s)\min}$ and $\mathrm{X}_{P(s)\max}$ are the minimum and maximum bounds of in-process variables, respectively.

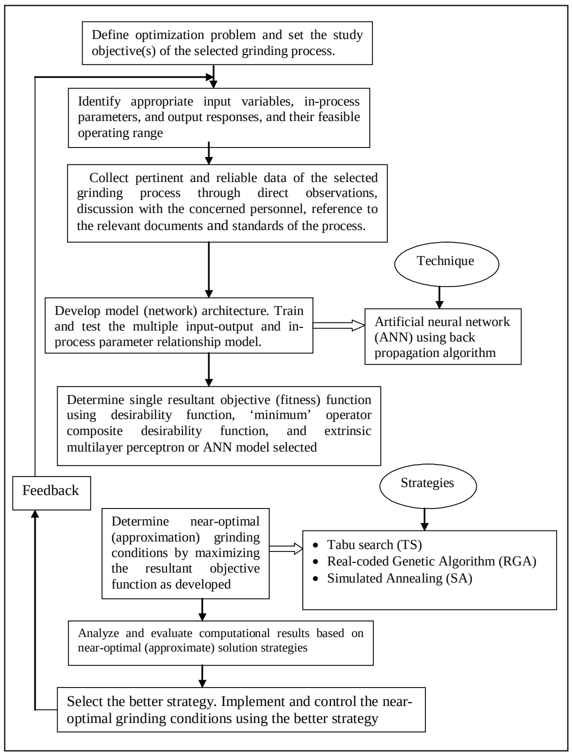
Solution Methodology

In order to provide an optimal solution for the multiple response optimization problems, as formulated in Equations (7-10), an integrated approach comprising multivariate inferential process modelling, desirability functions, and metaheuristic search techniques, is required. In this particular chapter, linear desirability function and equal weight of responses are used only to illustrate the potential and effectiveness of the integrated approach. However, nonlinear desirability functions or weighted responses may be used, based on expert knowledge and experience on the shape of desirability functions and also relative importance or criticality of responses for the particular problem under consideration. It is to be mentioned that the form or degree of desirability function does not affect the integrity of approach or violate any other conditions for implementation of the proposed solution methodology. Moreover, the solution methodology is flexible to incorporate varied forms or degree of multivariate response at s-th stage of grinding.

Among the metaheuristic, RGA, SA, and TS, including their variants, are recommended in order to obtain near-optimal (approximate) solution(s) for such types of problems. The optimal solution methodology, as proposed, is illustrated with the help of a flow diagram in Figure 3.

The techniques selected are discussed in brief in the following sections.

Figure 3. Schematic diagram of the solution methodology



Error Back Propagation-Based ANN (BPNN)

1. Multilayered perceptron (MLP) networks trained using error back propagation (BP) algorithm (Rumelhart et al., 1986), is a key development and popular choice for researchers in various process modelling applications (Coit et al., 1998; Feng et al., 2002; Jain et al., 1999; Sathyanarayanan et al., 1992; Wasserman, 1993; Zhang & Huang, 1995). The learning ability of nonlinear relationship, having unknown interactions, in a cutting process, without prior assumptions on the functional form of the relationship between input(s), and in-process parameter(s) and output(s) (e.g., linear, quadratic, higher order polynomial, and exponential), makes BPNN an attractive choice for nonlinear multivariate modelling problems.

Some of the attractive features of BPNN are as follows:

2. Ability to handle imprecision and fuzzy information and capability of generalization.
3. Learning and adoptability allows the process to update or modify its internal structure according to changing environment.

4. noise-insensitivity provides accurate prediction in the presence of uncertain data and measuremental error (Basheer & Hajmeer, 2000),
5. simple and uses local computations only for processing information in the neural network.

In case of a fully interconnected artificial neural network with simple neuron processing units, each neuron has an adjustable weight factor (w_{ij}) associated with it. The net input (v_j) or net effect for the j-th node is expressed as:

$$v_j = \sum_{j-1} w_{ij} v_i \quad (11)$$

where, i is the node in the previous layer, w_{ij} is the connecting weights of j, and i-th nodes in two different layers. The corresponding activation, $v_j^{'}$, of the neuron is determined using a transfer function(θ) (Basheer & Hajmeer, 2000; Rumelhart et al., 1986; Wasserman, 1993) that converts the total signal into a real number from a bounded interval, and expressed as:

$$v_j^{'} = \theta(v_j) = \theta(\sum_{j-1} w_{ij} v_i) \quad (12)$$

As this technique distributes backward the error starting from output layer down through hidden layer, it is so-called back propagation of error with modified delta rule (Basheer & Hajmeer, 2000; Rumelhart et al., 1986). An error measure for final check on training and testing performance is normalized root-mean-square of error (RMSE) (Tsai and Wang, 2001), and expressed as:

$$RMSE = \frac{1}{Length[o_j(t) - v'_j(t)]} \sqrt{\sum_p \frac{1}{2}[o_j(t) - v'_j(t)]^2} \quad (13)$$

where, $o_j(t)$ is the actual or observed output and

$$v'_j = \theta(v_j) = \theta(\sum_{j-1} w_{ij} v_i) = \tanh\left(\frac{1-e^{-v_j}}{1+e^{-v_j}}\right) \quad (14)$$

After learning is completed, the final values of weights are fixed. These final weight values are used for testing and other network functional "recalling" sessions. In this study, a BPNN with a modified performance function (MSEREG) (Howard & Beale, 1998; Mukherjee & Ray, 2006a), constant momentum, variable learning rate (Howard & Beale, 1998), hyperbolic tangent squashing function, and batch gradient decent algorithm (Howard & Beale, 1998; Mukherjee & Ray, in press, 2006a; Wasserman, 1993) is preferred, which attempts to improve generalization, stabilize learning, and improve the performance of steepest decent algorithm.

Real-Coded Genetic Algorithm (RGA)

In this particular study, real-coded variables, blending method crossover, and uniform mutation rate are selected. The basic statements of programming RGA in matlab (version 6.5) environment are based on Haupt and Haupt (1998). The matlab code is modified according to the problem, which was originally developed by Houck, Joines, and Kay (1998). The statements or pseudo-code of RGA to maximize $\lambda_{(n)}$ is provided in Figure 4.

Simulated Annealing (SA)

The basic statements (pseudo-codes) of SA for maximizing $\lambda_{(n)}$, which is implemented in matlab (version *6.5*) programming environment is base on the work by Eglese (1990), and provided in Figure 5.

The cooling schedule selected for the study is a modified version of power function as proposed by Kirkpatrick et al. (1983), and expressed as:

Figure 4. Pseudo-code of RGA for maximization of composite desirability function

Begin
Set initial generation number *g=0*;
Choose (Construct) a set (population size is 200) of initial feasible random variables of $x_{s(i)}$ [say *pop* (X_n)];
Evaluate *pop* (X_n); %Assess the fitness function/objective function $\lambda_{(s)}$ of every individuals in pop (X_n)
Record the best $X_n = X_n*$, and best $\lambda_{(s)} = \lambda*_{(s)}$; % $\lambda_{(n)}$ (X_n*) = $\lambda*_{(s)}$
Select and record best half of pop (X_n), e.g. *pop'* (X_n); % best half selected based on $\lambda_{(s)}$ (X_n) values
Repeat
g=g+1;
Select and record best half of *pop'* (X_n) as *pop''* (X_n);
Select mating pairs and reproduce from *pop''* (X_n); %reproduces same number of new offspring by predefined artificial blending method crossover. Record all new offspring in a set *pop'''* (X_n);
Select from *pop''* (X_n) and *pop'''* (X_n) to form new *pop'* (X_n); % size of *pop'* (X_n) remains same
Set new *pop'* (X_n) ← {*pop''* (X_n), *pop'* (X_n)};
Mutate and evaluate new *pop'* (X_n); % Assess the fitness function/objective function, *f* (X_n) of individuals in pop' (X_n)

IF best $\lambda_{(s)}$ of new pop' (X_n)> $\lambda*_{(s)}$ and $\lambda_{(s)}$ ($X_{(s)}$') $\geq$ *0.5* **THEN** % $X_{(s)}$' is a input conditional state space point for the *s*-th stage which provide best $\lambda_{(s)}$
Replace and record new best $X_n = X_n*$, and best $\lambda_{(s)} = \lambda*_{(s)}$; % $\lambda_{(s)}$ (X_n*) = $\lambda*_{(s)}$
END;
Until *g*=200;
Return the best X_n* visited [$x_{s(i)}$'s], and corresponding elite $\lambda*_{(s)} = \lambda_{(s)}$ (X_n*) value

$$T = T_0 * (T_n / T_0)^{\wedge(t/\alpha)} \qquad (15)$$

where, T_0, T_n, t, and α are initial temperature, final temperature, temperature change counter, and scale factor, respectively. A SA technique may also accept inferior state space to generate new neighborhood state space based on an acceptance probabilistic measure to avoid trap to local optimal. The acceptance probability selected for this study is based on the following expression.

$$\text{Probability } (\textit{Acceptance of a bad move}) = \exp^{(-\Delta E*\beta/T)} \qquad (16)$$

where, β is a positive number to scale parameter, and ΔE is the change in $\lambda_{(s)}$value.

Tabu Search (TS)

The most critical and difficult task to implement a TS strategy for continuous variable optimization problem is to define an appropriate tabu restriction zone. For discrete variable, identifying similar moves is far easier than for continuous variable problems. In case of a two-dimensional problem (single predicted response having single predictor), the tabu zone may be easily identified as a predefined circular area around current trial point (X). For three-dimensional problems (single response and two predictor), it may be a predefined spherical area. The problem complexity further increases as the nonlinearity and dimensionality of state space increases. As per literatures, researchers (Glover, 1994; Siarry & Berthiau, 1997)

Figure 5. Pseudo-code of SA for maximization of composite desirability function

```
Begin
Construct an initial random feasible solution ( X_n (0)); % X_n (0) is the initial feasible random
                                                            variables, x_s(i) values in the state
                                                            space bounds
Set best objective/fitness function λ*_(s) = λ_(s) ( X_n (0));
Set best solution X_n *= X_n (0);
Set current solution X_n = X_n (0);
Set the initial pseudo temperature T_0 as 50; % T_0>0;
Set a final pseudo temperature T_N as 45; %T_0 > T_N>0;
Set initial temperature change counter t=0;
Set constant temperature increment δ_t as 0.1 % increment of t
Repeat
Set repetition counter i=0;
Repeat
Let X_n ' be a near-neighbourhood feasible point of X_n % Determination of near-
                                                            neighbourhood point of X_n is
                                                            discussed in chapter-III
Let ΔE=λ_(s) ( X_n *) - λ_(s) ( X_n ');
IF ΔE≤0 and λ_(s) ( X_n ') ≥ 0.5 THEN set X_n = X_n '% Uphill movement
Set λ*_(s) = λ_(s) ( X_n ');
Set X_n *= X_n ';
IF ΔE>0 THEN set X_n = X_n ' if p_T=exp^(-ΔE* β /T) > p=random number (1, 0) % p_T is
                                                            acceptance probability of bad move or downhill
                                                            movement
i=i+1;
Until i=50 % 50 is the maximum number of iteration to be performed at any particular
                                                            temperature

t=t+0.1;
Set T=Cooling schedule function f (T_0, T_N, t); % f is the cooling schedule function
Until 2000 iteration is reached
Return the best X_n * visited and corresponding elite λ*_(s) value
```

prefer to identify tabu restriction based on specific attribute(s) of components of state space vector in such situations. However, working with components of state space vector and their attributes, rather than complete vector in higher dimensional search space may be computationally expensive, and rather complicated. In addition, the proposed works on continuous variables using TS strategy are found suitable up to two-to-three dimensional state space problem (Siarry & Berthiau, 1997). However, there is lack of successful attempts for higher dimensional state space problem. In this context, multidimensional complete vector distance concept to identify tabu move (Mukherjee & Ray, 2006b) may be a viable, easy-to-implement, and reasonable alternative for both isolated and interdependent stage multidimensional state space problems. Moreover, an idea of complete multivariate vector distance in restrictive or tabu zone (hyperplane or hypersurface) is relatively easier to physically interpret than working with attributes of component of vector concept to identify tabu zone.

It is in this context, a modified TS (MTS) strategy as proposed by Mukherjee and Ray (in press) for continuous variable grinding problem, having multidimensional state space is used. An absolute Mahalanobis vector distance (Maesschalck, Jouan-Rimbaud, Massart, 2000; Mahalanobis, 1936; Rego & Alidaee, 2005) approach is used to identify tabu moves, rather than components of vectors (Glover, 1994). Absolute Mahalanobis distance between any two instances ($X_{(a)}$, $X_{(b)}$) in the input conditional state space may be expressed as:

$$MD_{a,b} = \sqrt{(X_{(a)} - X_{(b)}) S_X^{-1} (X_{(a)} - X_{(b)})^T} \quad (17)$$

where, S_X is the sample variance-covariance matrix of multiple multivariate q dimensional $X_{(1)}$ observations. However, the primary idea of MTS strategy is based on scatter search (Glover et al., 2000) intensification scheme as proposed by Glover (1994). The basic statement of MTS strategy (as shown in Figure 6) is proposed by Mukherjee and Ray (in press), and is based on the work by Glover (1994), Gendreau (2003), and Adra (2003). In this study, a simplest idea used as aspiration criterion is to allow move, even if it is within tabu zone, if it results in a solution with improved objective value, which is better than current best-known or elite solution point (Gendreau, 2003). A weighted combination of reference points previously created is used to generate new trial points for intensification strategy. A typical diversification technique is adopted by generating random trial points (unmapped or not based on scatter search approach) in absence of improving moves for predefined number of iteration (Lokketangen & Glover, 1998), hoping

Figure 6. MTS strategy for maximization of degree of customer satisfaction for grinding optimization problem

Begin
Choose (Construct) initial random feasible trial point(s) (X_0); %Population of random X_n values may be selected and X_0 is the potential best trial point ($x_{s(i)}$] in the population based on $\lambda_{(s)}$ value
Set current solution (X_n) = X_0;
Set best objective function value $\lambda^*_{(s)} = \lambda_{(s)}$ (X_0);
Set best $X_n^* = X_0$;
Create an empty Tabu list (T) =Ø; % Tabu list consist of recently visited L (size of T) input multi-dimensional state space

While *2000* iteration not met **Do** % Termination criteria may be time based or number of main loop iteration based

Generate $X_{(n)}$' feasible neighbourhood point(s) of $X_{(n)}$ and not in T; % Set tabu criteria using a pre-defined absolute critical Mahalanobis vector distance concept and thereby determine tabu restriction zone or tabu move in input multi-dimensional state space. The detail of calculation is provided in Chapter-IV

IF $\lambda_{(s)}$ ($X_{(n)}$')> $\lambda_{(s)}$* and $\lambda_{(s)}$ ($X_{(n)}$') $\geq$ *0.5*
Set $\lambda^*_{(s)} = \lambda_{(s)}$ (X_n') and X_n*= X_n';
Set X_n=X_n';
End
End While
Return the best X_n* or $x^*_{s(i)}$ and corresponding $\lambda^*_{(s)}$

for an improvement in objective value. A move to a new trial point in the state space is defined as tabu (or forbidden) if the absolute Mahalanobis distance (MD) between any new trial points ($X'_{(1)}$) and any one of the recent moves stored in the three dimensional tabu list (T) array is less than the problem-specified critical MD.

The problem-specific critical absolute MD selected is calculated based on the following expression:

$$MD\{Critical\} = (1/c)*[MD_{\max} - MD_{\min}] \quad (18)$$

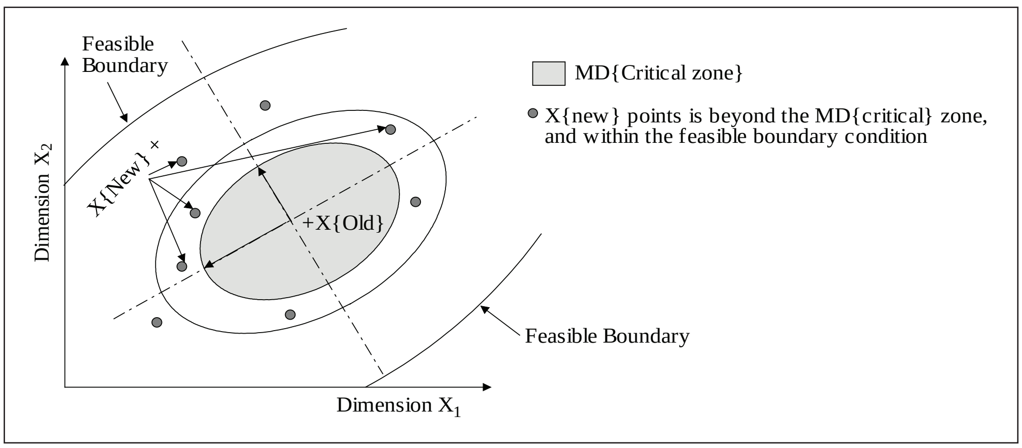
where, $MD_{(max)}$ and $MD_{(min)}$ are the maximum and minimum absolute MD from the CP of the cloud of primary state space of all the multivariate observations of $X_{(1)}$ in primary state space, respectively. c is a positive denominator selected, which divide the range of observed absolute MD from center point (CP) of state space, into c equal number of intervals. The elliptical zone that can be constructed based on calculated MD $\{critical\}$ of Equation (26) is the forbidden or tabu restriction zone/area selected. A typical tabu restriction elliptical zone in a two-dimensional state space is illustrated in Figure 7.

CASE EXAMPLE

This section describes in detail the use and effectiveness of the recommended solution methodology to optimize a grinding (honing) process in an automobile manufacturing unit. A two-pass (rough and finish) hydraulic oil pressure-based vertical honing (abrasive machining) operation for the liner bores of a typical 6-cylinder diesel engine is considered for the optimization study.

Finished cylinder block liner is one of the most critical engine components characterized by its internal surface quality having a bearing on the level of smoke emission, lubrication oil consumption, and longevity of diesel engine (Feng et al., 2002). Internal surface quality characteristics, such as surface finish, honing angle (cross-hatch angle), and dimensional accuracy also affect the lubrication oil retention property at high engine combustion temperature, and its resistance to wear. Presently, honing is the only machining process available that simultaneously provides required surface finish, cross hatch angle, and dimensional accuracy for engine cylinder block liner (Feng et al., 2002). In case of honing, geometrically undefined and random directional mul-

Figure 7. Tabu restriction



tiedge abrasive stone carries out cutting operation requiring stone expansion, rotation, and stroking that provide required surface finish, cross-hatch lay direction, and dimensional accuracy.

In this particular case example, the engine manufacturer is facing acute problem of simultaneously maintaining and controlling desired surface finish, cross-hatch angle, and dimensional accuracy of the liner, resulting in harmful side effects such as high smoke emission (measured in Hatridge smoke unit or HSU) and lubrication oil consumption (measured in gms per hour at the rated crankshaft rpm). The best-possible combination of input conditions (in terms of three specified inputs variables and seven process parameters), ensuring minimum variability in the outputs or responses (measured with five response characteristics in this case), remains unknown.

In view of these specific problems, which are of critical and chronic nature, the optimization problem is based on consideration of the following important aspects:

1. Process parameter settings for honing operation need to be optimally determined so that operational performance of the cylinder block is at an acceptable level
2. The grinding operation must ensure acceptable degree of customer satisfaction at the finish stage, which is basically dependent on optimal parameter settings
3. A number of harmful side effects, such as smoke emission and lubrication oil consumption, need to be reduced
4. Lubricant oil retention at high engine combustion temperature needs to be improved
5. Optimal process conditions as determined and implemented should ensure stability in the honing operation, mainly affected by chance or random causes.

Once the input variables, in-process parameters, and responses of the grinding (honing) process are identified at *i*-th stage, pertinent and reliable production data (Coit et al., 1998) are collected through direct observation, discussion with the concerned personnel, reference to the relevant documents, and standard operating practices of the manufacturing unit. Data are considered to be representative of the entire feasible operating conditions under which the inferential models will be used. The data were collected during the period of 2004-2005 in a finish honing processes (Gehring-make machine) required for diesel engine cylinder blocks, which are used for commercial heavy utility vehicles, in one of the largest automotive manufacturing plant in private sectors in eastern India. Real time data are collected at different time points, based on appropiate sampling plan (Montgomery, 1991).Periodic calibration and maintenance of the gauges and measuring instrument during data collection ensure accuracy, precision, and minimum measurement error.

The objective of the study is to determine the optimal levels of seven in-process parameters ($X_{P(s)}$) at the two-pass finish honing stage (e.g., dog-length, $x_{s(1)}$; vertical stroke speed, $x_{s(2)}$; rough-pass hydraulic pressure, $x_{s(3)}$; finish-pass hydraulic pressure, $x_{s(4)}$; rough-pass honing time, $x_{s(5)}$; finish-pass honing time, $x_{s(6)}$; and cutting oil temperature $x_{s(7)}$) and three input variables ($X_{I(s-1)}$) (e.g., input liner bore diameter, $x_{s(8)}$; input liner bore ovality, $x_{s(9)}$; and input liner bore taper, $x_{s(10)}$), such that the degree of customer satisfaction in relation to five specific response characteristics (e.g., surface finish (Y_{1s}); honing angle {or cross-hatch angle} (Y_{2s}),finished liner bore diameter (Y_{3s}); finished liner bore ovality (Y_{4s}); and finished liner bore diameter taper (Y_{5s})) is maximized. Diameter, surface finish, and honing angle are calculated based on the average of six readings taken at different bore location. Whereas, ovality and taper are measured based on maximum reading. All the inputs, in process parameters, and response variable data are preprocessed to standardized values or scale free variables (Rencher, 1995; Timm, 2002) before final model development.

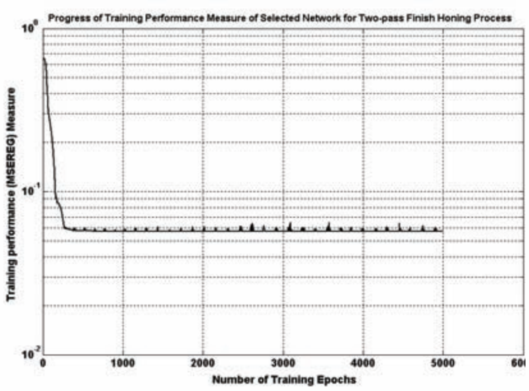
The final model as development and acceptable solution(s) as determined by MTS, RGA, and SA is discussed in the following section.

BPNN Algorithm-Based Grinding Process Model

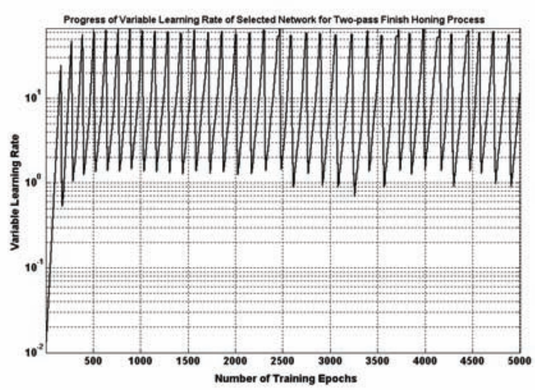
BPNN architecture selected for two-pass finish honing process ha 10 input layer (inputs and in-process parameters), 5 output layer (responses), and 30 neurons in single hidden layer. Hyperbolic tangent sigmoidal nonlinear transfer functions are adopted for representing the activation to other neurons (Hagan, Demuth, & Beale, 2002). The total data are divided into training and testing (untrained) half in an 80%-20% ratio (Coit et al., 1998). The training data set is selected in such as way so that it consist of extreme points in multi-dimensional state space, and thereby eliminate or minimize extrapolation of model while predicting responses of distinct test sample observations. In other words, training set is selected in such a way so that the test data should lie within the multidimensional state space boundary of training data (Basheer & Hajmeer, 2000). The network is trained in the batch training mode, using momentum constant and variable (adaptive) learning rate (Hagan et al., 200) for stable training. To improve the generalization of the final network architecture or prevent over-fitting of data, MSEREG is selected as discussed earlier. For a MSEREG function, γ_k is selected as 0.5 (Howard & Beale, 1998). The performance goal of the networks during trial runs is set to 0.001 or if the maximum number of epoch reaches 5000 (Tsai & Wang, 2001). The change in MSEREG and adaptive learning rate with number of epoch during training is provided in Figure 8. Figure 9 ([a] and [b]) shows the plot of actual and predicted responses for the test sample set of input conditions and trained BPNN architecture.

The RMSE value of training and testing data set based on BPNN model are 0.0962 and 0.2171, respectively.

Figure 8. Change in (a) MSEREG and (b) variable learning rate for two-pass finish honing network during training

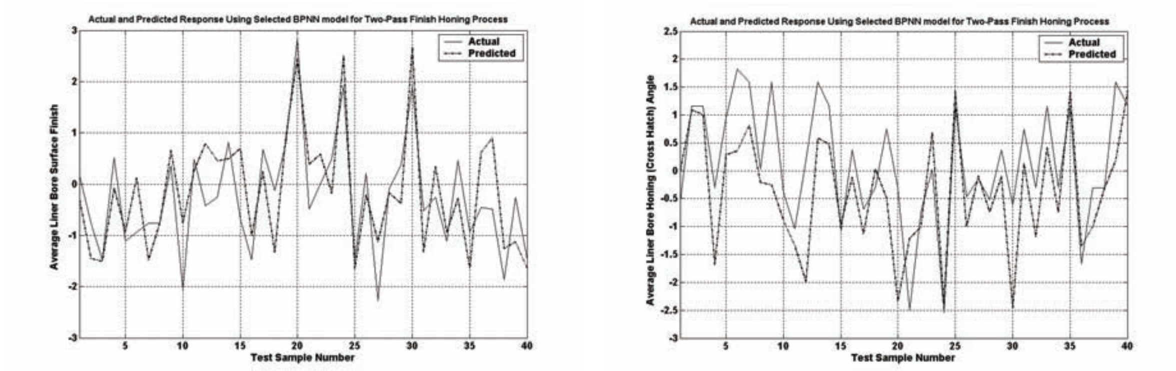


(a) Change in MSEREG during training



(b) Change in variable learning rate during training

Figure 9. Actual and predicted responses for the test sample set of input conditions using trained BPNN architecture



(a) Actual and predicted average liner bore surface finish

(b) Actual and predicted average liner bore honing angle

Metaheuristic Approach

The inferential process models as developed in the forth section provide the necessary functional inputs to determine near-optimal cutting conditions by any particular metaheuristic search technique. The bounds on $X_{I(s-1)}$, and $X_{P(s)}$ determine the search space for a grinding optimization problem. The desired values of response quality characteristics are also identified at this stage of analysis. The necessary steps involved to determine near-optimal process conditions are as follows:

- **Step 1:** Formulate the single-stage optimization problem.
- **Step 2:** Define the intrinsic parameters for selected metaheuristic strategy.
- **Step 3:** Specify the input conditional state space bounds and desired response values
- **Step 4:** Determine the elite near-optimal solution based on proposed solution methodology for single-stage system, using h_s inferential model function, desirability functions, and metaheuristic strategy.
- **Step 5:** Repeat **Step 3** to **Step 4** for defined number of independent computational runs.

For the implementation of RGA, the search conditions, such as population size of 200, blending method crossover, and mutation at a rate of 4% (or $\mu = 0.04$) are selected. The parametric values (search conditions) of T_0, T_N, δ_t, α, and β selected for SA are 50, 45, 0.1, 200, and 800, respectively, for both single and two-stage system. The ratio of T_n/T_0 is selected at 0.9 (Kirkpatrick et al., 1983). The intrinsic parametric values for MTS, selected for analysis, such as the length of tabu list (L), and c are 15 and 80, respectively. The termination criteria as determined by initial trial runs for MTS and SA are set at 2000 maximum number of main loop iteration, and for RGA it is 200 maximum number of generation. In this study if degree of customer satisfaction is less than 0.5, the solution is not accepted or considered in the list of elite solution in any particular run. A sample size of 30 consecutive independent computational run results, for any particular metaheuristic strategy, is selected to determine the best attained and average values of $\lambda^*_{(s)}$, and average computational times. All the computational runs are executed in a single specific personal computer (PC) having basic configuration as Intel Pentium (*IV*) processor, 1.8 GHz CPU, and compatible to 256 MB RAM. The summary statistics of independent and consecutive 30 computational runs for MTS, RGA, and SA are provided in Table 1.

The summary statistics show the suitability and better consistency of MTS (in terms of sample mean and standard deviation of $\lambda_{(i)}$ values), over RGA and SA.

Figure 10 [(a), (b), and (c)] illustrates the progress of metaheuristic search for the maximum attained (highest) $\lambda_{(s)}$ value using MTS, RGA, and SA, respectively, using ***BPNN***-based process functional approximation and considering two-pass finish honing as an isolated system optimization problem. Table 2 provides implementation results based on 30 existing and predicted $\lambda_{(s)}$ values (degree of customer satisfaction) based on MTS.

The results show a considerable improvement in degree of customer satisfaction measure, ensuring an acceptable $\lambda_{(s)}$ value (or $\lambda_{(s)} \geq 0.5$). BPNN-based model sensitivity on the MTS solution is also performed to understand the possible impact of variation (Mukherjee & Ray, in press).

FUTURE RESEARCH DIRECTIONS

Someof the important areas where opportunity exists for carrying out further research work are as follows:

1. In order to involve expert knowledge, fuzzy set-based modelling technique may be employed.

Table 1. Summary statistics based on 30 computational runs

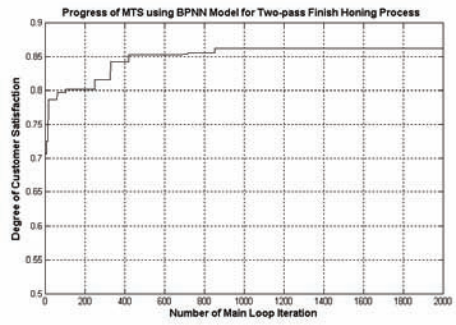
Strategy	**Highest $\lambda_{(s)}$ value attained**	**Minimum $\lambda_{(s)}$ value attained**	**Mean $\lambda_{(s)}$ attained**	**Standard deviation (std.dev.)**	**Average CPU time (in seconds)**
MTS	0.86109	0.85012	0.855806	0.003165	634.8017
RGA	0.88055	0.71881	0.836517	0.038992	93.645
SA	0.82697	0.80172	0.803824	0.007289	19345.08

Table 2. Best existing value and expected implementation results (based on MTS) at the finish honing process

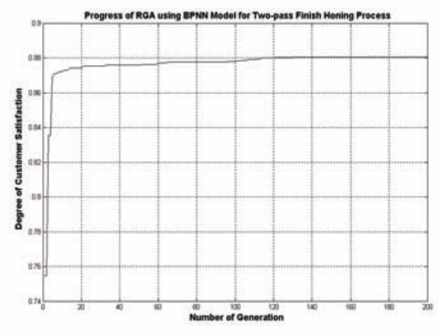
Settings	Implementation Results			
	Maximum $\lambda_{(s)}$	Minimum $\lambda_{(s)}$	Average $\lambda_{(s)}$	SD[a] ($\lambda_{(s)}$)
Existing	0.4991	0.0505	0.1876	0.1162
Proposed	0.8611[b]	0.8501[b]	0.8558[b]	0.0032[b]

[a] Sample standard deviation; [b] To nearest decimal place

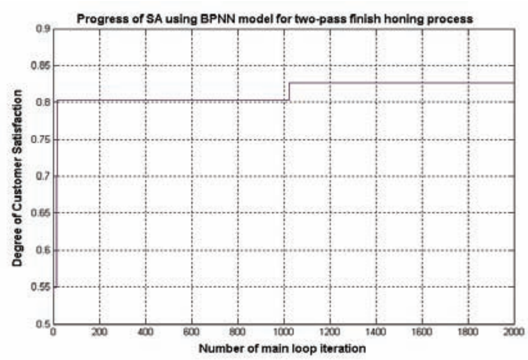
Figure 10. Progress of (a) MTS, (b) RGA, (c) SA-based search for the highest degree of customer satisfaction value using BPNN model



(a) MTS-based search



(b) RGA-based search



(c) SA-based search

2. The scope exist for developing metaheuristic techniques based on ant colony system, particle swarm optimization, and hybrid metaheuristic-based strategy, which may be employed in order to verify their applicability for abrasive metal cutting processes.
3. The proposed inferential models can be further perfected with the addition of designed and experimental data, providing additional information on extreme ranges of process conditions.

It is expected that the future trend of research will be directed considering these opportunities in abrasive metal cutting processes.

CONCLUSION

A systematic approach of artificial neural network-based modelling and determination of near-optimal cutting conditions by metaheuristic strategy has shown an interesting potential in both product and process quality improvement of abrasive metal cutting operation. The solution methodology for process parameter optimization in abrasive metal cutting operation attempts to provide a single, unified, and systematic approach to determine near-optimal cutting conditions in various kinds of abrasive metal cutting problems. It incorporates the use of one or more of the existing metaheuristic techniques, making the methodology a unified and effective means. Moreover, it attempts to provide the user with flexibility to adopt suitable techniques based on their inherent potential as discussed and on-hand problem complexity highlighting the importance of problem criticality, data collection, and analysis.

ACKNOWLEDGMENT

The authors gratefully acknowledge helpful and valuable technical suggestions and references suggestions provided by Professor Fred Glover (Leeds School of Business, University of Colorado, USA) regarding tabu search and neighborhood selection strategy, in early phases of this work. The authors are also grateful to the anonymous referees for their valuable comments and suggestions. The authors would also like to acknowledge the support and assistance provided by Human Resource, Quality Assurance (Engine Division), and Production (Engine Division) Department of M/s Tata Motors Limited, Jamshedpur, eastern India plant for their cooperation and support during data collection, and providing other process technical details.

REFERENCES

Aarts, E., & Korst, J. (1989). *Simulated annealing and Boltzmann machines*. New York: John Wiley and Sons.

Adra, S. (2003). Optimisation techniques for gas turbine engine control systems. *MSC dissertation report*. UK: University of Sheffield, Department of Computer Science.

Basheer, I. A., & Hajmeer, M. (2000). Artificial neural networks: Fundamentals, computing, design, and application. *Journal of Microbiological Methods*, *43*, 3-31.

Box, G. E. P., & Draper, N. R. (1987a). *Empirical model-building and response surface*. New York: John Wiley and Sons.

Box, G. E. P., & Draper, N. R. (1987b). *Evolutionary operation: A statistical method for process improvement*. New York: John Wiley & Sons.

Carlyle, W. M., Montgomery, D. C., & Runger, G. (2000). Optimization problem and method in quality control and improvement. *Journal of Quality Technology*, *32*(1), 1-17.

Chen, Y. T., & Kumara, S. R. T. (1998). Fuzzy logic and neural network for design of process parameters: A grinding process application. *International Journal of Production Research*, *36*(2), 395-415.

Chen, M .C. & Su, C. T. (1998). Optimization of machining conditions for turning cylinder Ssocks into continuous finish profiles. *International Journal of Production Research*, *36*(8), 2115-2130.

Chen, M. C., & Tsai, D. M. (1996). A simulated annealing approach for optimization of multipass turning operations. *International Journal of Production Research*, *34*(10), 2803-2825.

Chiao, C., & Hamada, M. (2001). Analyzing experiments with correlated multiple responses. *Journal of Quality Technology, 33*(4), 451-465.

Chowdhury, K., Pratihar, D. K., & Pal, D. K. (2002). Multi-objective optimization in turning: Using GA Algorithm. *Journal of Institute of Engineers (Ind)*, *82*, 37-44.

Coit, D. W., Jackson, B. T., & Smith, A. E. (1998). Static neural network process models: Considerations and case studies. *International Journal of Production Research*, *36*(11), 2953-2967.

Dabade, B. M. (1997). *Studies on some aspects of quality engineering in production plants*. Unpublished doctoral thesis. Indian Institute of Technology, Department of Industrial Engineering and Management, Kharagpur.

Dabade, B. M., & Ray, P. K. (1996). Quality engineering for continuous performance improvement in products and processes: A review and reflections. *Quality and Reliability Engineering International*, *12*, 173-189.

De Maesschalck, R., Jouan-Rimbaud, D. & Massart, D. L. (2000). Tutorial: The Mahalanobis distance. *Chemometrics and Intelligent Laboratory Systems*, *50*, 1-18.

Deb, K. (2002). *Multi-objective optimization using evolutionary algorithm* (1st ed.). John Wiley & Sons Ltd.

Del Castillo, E., Montgomery, D. C., & McCarville, D. R. (1996). Modified desirability functions for multiple response optimization. *Journal of Quality Technology*, *28*(3), 337-345.

Derringer, G. C. (1994). A balancing act: Optimizing a product's properties. *Quality Progress*, *27*, 51- 57.

Derringer, G., & Suich, R. (1980). Simultaneous optimization of several response variables. *Journal of Quality Technology*, *12*, 214-219.

de Werra, D., & Hertz, A.(1983). Tabu search techniques. *OR Spectrum*, *11*, 131-141.

Eglese, R. W. (1990). Simulated annealing: A tool for operational research. *European Journal of Operational Research*, *46*, 271-281.

El-Hofy, H. (2005). *Advanced machining processes: Nontraditional and hybrid machining processes*. New York: MacGraw-Hill.

Farago, F. T. (1976). *Abrasive methods engineering*. New York: Industrial Press Inc.

Feng, C., Wang, X., & Yu, Z. (2002). Neural networks modelling of honing surface roughness parameter defined by ISO 13565. *SIAM Journal of Manufacturing Systems*, *21*(8), 1-35.

Fu, L. (2003). *Neural network in computer intelligence*. India: Tata Mc Graw Hill Edition.

Gendreau, M. (2003) *Handbook of metaheuristics* (F. Glover & G. A. Kochenberger, Eds.) (pp. 40-54). New York: Kluwer Academic Publisher.

Ghosh, A., & Mallik, A.K. (1999). *Manufacturing science*. New Delhi: East West Press Private Limited.

Glover, F. (1986). Future paths for integer programming and links to artificial intelligence. *Computer & Industrial Engineering*, *13*, 533-549.

Glover, F. (1989). Tabu search: Part I. *ORSA Journal on Computing*, *1*, 190-206.

Glover, F. (1990a). Tabu search: A tutorial. *Interface*, *20*, 74-94.

Glover, F. (1990b). Tabu search: Part II. *ORSA Journal on Computing*, *2*(1), 4-32.

Glover, F. (1994). Tabu search for nonlinear and parameter optimization (with links to genetic algorithm). *Discrete Applied Mathematics*. *49*, 213-255.

Glover, F., & Laguna, M. (2002). *Tabu search* (6th ed.). London: Kluwer Academic Publisher.

Glover, F., Laguna, M., & Marti, R. (2000). Fundamentals of scatter search and path relinking. *Control and Cybernatics*, *29*(3), 653-684.

Goldberg, D. E. (2002). *Genetic algorithms in search, optimization, and machine learning* (5th ed.). Pearson Education Asia.

Gupta, R., Shishodia, K. S., Sekhon, G. S. (2001). Optimization of grinding parameters using enumeration method. *Journal of Material Processing Technology, 112,* 63-67.

Hagan, M. T., Demuth, H. B., & Beale, M. (2002). *Neural network design*. India: Thomson Learning.

Harrington, E. C., Jr. (1965). The desirability function. *Industrial Quality Control*, *21*, 494-498.

Hassan, G. A., & Suliman, S. M .A. (1990). Experimental modeling and optimization of turning medium carbon steel. *International Journal of Production Research*, *28*(6), 1057-1065.

Haupt, R. L., & Haupt, S. E. (1998). *Practical genetic algorithms*. New York: John Wiley & Sons, Inc.

Hooks, R., & Jeeves, T. A. (1961). Direct search solution of numerical and statistical problems. *Journal of the Association of Computing Machinery, 8*(2), 212-229.

Houck, C. R., Joines, J. A., & Kay, M. G. (1998). Binary and real-valued simulation evolution for Matlab GAOT V2. A genetic algorithm for function optimization: A Matlab implementation. *ACM transactions on mathematical software*. Free Software Foundation, Inc.

Howard, D., & Beale, M. (1998). *Neural network toolbox: For use with Matlab; user's guide* (Version 3). The Mathworks Inc.

Hui, Y. V., Leung, L. C., & Linn, R. (2001). Optimal machining conditions with cost of quality and tool maintenance for turning. *International Journal of Production Research, 39*(4), 647-665.

Jain, R. K., Jain, V. K., & Kalra, P. K.(1999). Modelling of abrasive flow machining process: A neural network approach. *Wear, 231*, 242-248.

Jeong, I., & Kim, K. (2003). Interactive desirability function approach to multi-response surface optimization. *International Journal of Reliability, Quality & Safety Engineering*, *10*(2), 205-217.

Juan, H., Yu, S. F., & Lee, B. Y. (2003). The optimal cutting parameter selection of production cost in HSM for SKD61 tool steels. *International Journal of Machine Tools & Manufacturing*, *43*, 679-686.

Khuri, A. I., & Conlon, M. (1981). Simultaneous optimization of multiple responses represented by polynomial regression functions. *Technometrics*, *23*, 363-375

Kim, K. J., & Lin, D. K. J. (2000). Simultaneous optimization of mechanical properties of steel by maximizing exponential desirability functions. *Applied Statistics*, *49*(3), 311-325.

King, R. I., & Hahn, R. S. (1986). *Handbook of modern grinding technology*. New York: Chapman & Hall.

Kirkpatrick, K., Gelett, C. D., Jr., & Vecchi, M. P. (1983). Optimization by simulated annealing. *Sciences*, *220*, 671-680.

Kolahan, F., & Liang, M. (1996). A tabu search approach to optimization of drilling operations. *Computer & Industrial Engineering*, *31*(1/2), 371-374.

Krajnik, P., Kopac, J., & Sluga, A. (2005). Design of grinding factors based on response surface methodology. *Journal of Materials Processing Technology*, *162-163*, 629-636.

Krimpenis, A. & Vosniakos, G.C. (2002). Optimization of multiple tool NC rough machining of a hemisphere as a genetic algorithm paradigm application. *International Journal of Advance Manufacturing Technology, 20,* 727-734.

Lee, C. W, & Shin, Y. C. (2000). Evolutionary modelling & optimization of grinding processes. *International Journal of Production Research*, *38*(12), 2787-2831.

Lokketangen, A., & Glover, F. (1998). Solving zero-one mixed integer programming problems using tabu search. *European Journal of Operations Research, 106,* 624-658.

Luong, L. H. S., & Spedding, T. A. (1995). A neural-network system for predicting machining behaviour. *Journal of Materials Processing Technology, 52,* 585-591.

Mahalanobis, P. (1936). On the generalized distance in statistics. In *Proceedings of the Indian National Institute Science,* Calcutta, India (Vol. 2, pp. 49-55).

Maksoud, T. M. A, & Atia, M. R. (2004). Review of intelligent grinding and dressing operations. *Machine Science & Technology, 8*(2), 263-276

Mayers, R. H., & Carter, Jr. (1973). Response surface technique for dual response system. *Technometrics, 15*(2), 301-317.

Montgomery, D.C. (1991). *Introduction to statistical quality control* (2nd ed.). New York: John Wiley & Sons.

Montgomery, D. C. (2001). *Design and analysis of experiment* (5th ed.). New York: Wiley & Sons, Inc.

Montgomery, D. C., & Peck, E. A. (1992). *Introduction to linear regression analysis* (2nd ed.). New York: John Wiley & Sons, Inc.

Moulik, P. N., Yang, H. T. Y., & Chandrasekhar, S. (2001). Simulation of thermal stresses due to grinding. *International Journal of Mechanical Science, 43,* 831-851.

Mukherjee, I., & Ray, P. K. (in press). A modified tabu search strategy for multiple-response grinding process optimization. *International Journal of Intelligent Systems Technologies & Applications (IJISTA) (Special Issue on Advanced Evolutionary Computational Techniques for Design, Manufacturing, Logistics and Supply Chain Problems).*

Mukherjee, I., & Ray, P. K. (2006a, May). Metaheuristic strategies for non-linear multi-response grinding process optimization. *The ICFAI Journal of Operations Management, 5*(2), 35-56.

Mukherjee, I., & Ray, P. K (2006b). A review of optimization techniques in metal cutting processes. *International Journal of Computers & Industrial Engineering, 50,* 15-34.

Onwubolu, G. C., & Kumalo, T. (2001). Optimization of multi-pass turning operation with genetic algorithm. *International Journal of Production Research, 39*(16), 3727-3745.

Osborne, D. M., & Armacost, R. L. (1996). Review of techniques for optimizing multiple quality characteristics in product development. *Computer & Industrial Engineering, 31*(1/2), 107-110.

Petri, K. L., Billo, R. E., & Bidanda, B. (1998). A neural network process model for abrasive flow machining operations. *Journal of Manufacturing Systems, 17*(1), 52-64.

Phadke, M. S. (1989). *Quality engineering using robust design*. AT&T Bells Laboratory, New Jersey: Prentice Hall.

Rego, C., & Alidaee, B. (2005). *Metaheuristic optimization via memory and evolution: Tabu search and scatter search*. Kluwer Academic Publisher.

Rencher, A. C. (1995). *Method of multivariate analysis.* John Wiley & Sons, Inc.

Rumelhart, D. E., Hilton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature, 323*(9), 533-536.

Sathyanarayanan, G., Lin, I. J., & Chen, M. (1992). Neural network modelling and multiobjective optimization of creep feed grinding of superalloys. *International Journal of Production Research, 30*(10), 2421-2438.

Schrader, R. S. (2003). *Optimization of cost and cycle time for multi-pass turning operation using GA* (Rep. No. 2003010, pp. 205-2065). University of Illinois.

Shaji, S., & Radhakrisnan, V. (2003). Analysis of process parameters in surface grinding with graphite as lubricant based on Taguchi method. *Journal of Material Processing Technology*, *141*(1), 1-9.

Shin, Y. C., & Vishnupad, P. (1996). Neuro-fuzzy control of complex manufacturing processes. *International Journal of Production Research*, *34*(12), 3291-3309.

Siarry, P., & Berthiau, G. (1997). Fitting of tabu search to optimize functions of continuous variables. *International Journal for Numerical Methods in Engineering*, *40*, 2449-2457.

Stephenson, D. A., & Aqapiou, J .S. (1997). *Metal cutting theory and practices: Manufacturing engineering and material processing.* New York: Marcel Dekker.

Tan, F. P., & Creese, R. C. (1995). A generalized multi-pass machining model for machining parameter selection in turning. *International Journal of Production Research*, *33*(5), 1467-1487.

Teeravaraprug, J., & Cho, B. R. (2002). Designing the optimal process target levels for multiple quality characteristics. *International Journal of Production Research*, *40*(1), 37-54.

Timm, N. H. (2002). Applied multivariate analysis. Springer text in statistics. *Principal component, canonical correlation, and exploratory factor analysis* (pp. 447-448). New York: Springer-Verlag.

Tong, L., Wang, C., Houng, J., & Chen, J. (2001). Optimizing dynamic multiresponse problems using the dual-response-surface method. *Quality Engineering*, *14*(1), 115-125.

Tsai, K., & Wang, P. (2001). Predictions of surface finish in electrical discharge machine based upon neural network models. *International Journal of Machine Tools & Manufacture*, *41*, 1385-1403.

Vining, G. G. (1998). A compromise approach to multiresponse optimization. *Journal of Quality Technology*, *30*(4), 309-313.

Voβ, S. (2001). Meta-heuristics: The state of art. In A. Nareyak (Ed.), *Lecture notes in computer science* (Vol. 2148, pp. 1-23). Springer Verlag.

Wasserman, P. D. (1993). *Advanced methods in neural computing.* New York: Van Nostrand Reinhold.

Waters, T. F. (1996). *Fundamentals of manufacturing for engineers.* UCL Press/Taylor & Francis Group.

Xu, K., Lin, D. K. J., Tang, L., & Xie, M. (2004). Multiresponse system optimization using goal attainment approach. *IIE Transaction*, *36*, 433-445.

Youssef, H., Sait, S. M., & Adiche, H. (2001). Evolutionary algorithms, simulated annealing and tabu search: A comparative study. *Engineering Applications of Artificial Intelligence*, *14*, 167-181.

Zadeh, L. A. (1965). Fuzzy sets. *Information and Control*, *8*, 338-353.

Zang, H. C., & Huang, S. H. (1995). Application of neural network in manufacturing: A state of art survey. *International Journal of Production Research*, *33*(3), 705-728.

ADDITIONAL READING

Bohachevsky, I. O., Johnson, M. E., & Stein, M. L. (1986). Generalized simulated annealing for function optimization. *Technometrics*, *28*(3), 209-217

Box, G. E .P. (1985). Discussion of offline quality control, parameter design, and Taguchi's method (with discussion) by R N Kaker. *Journal of Quality Technology*, *17*, 189-190.

Chams, M., Hertz, A., & de Werra, D. (1987). Some experiments with simulated annealing for coloring graphs. *European Journal of Operations Research*, *32*, 260-266.

Connolly, D. T. (1989). An improved annealing scheme for the QAP. *European Journal of Operations Research*, *46*, 93-100.

Cus, F., & Balic, J. (2003). Optimization of cutting process by GA approach. *Robotics and Computer Integrated Manufacturing*, *19*, 113-121.

Cvijovic, D., & Klinowki, J. (1995).Taboo search: An approach to multi-minima problem. *Science*, *267*, 664-666.

Fleischer, M. (1995). Simulated annealing: Past, present, and future. In *Proceedings of the Winter Simulation Conference* (pp. 155-161).

Gen, M., & Cheng, R. (2000). *Genetic algorithms and engineering optimization*. New York: John Wiley and Sons, Inc.

Glover, F. (1977). Heuristics for integer programming using surrogate constraints. *Decision* Sciences, *8*(1), 156-166.

Glover, F., & Greenberg, H. J. (1989). New approach for heuristic search: A bilateral linkage with artificial intelligence. *European Journal of Operations Research*, *39*, 119-130.

Glover, F., Kelly, J. P., & Laguna, M. (1995). Genetic algorithm and tabu search: Hybrids for optimization. *Computers and Operations Research*, *22*(1), 111-134.

Govindhasamy, J. J., MacLoone, S. F., Irwin, G. W., French, J. J., & Doyle, R. P. (2005). Neural modelling, control and optimization of an industrial grinding process. *Control Engineering Practice*, *13*, 1243-1258.

Henderson , D., Jacobson, S. H., & Johnson, A. W. (2002). The theory and practice of simulated annealing. In F. Glover & G.A. Kochenberger (Eds.), *Handbook of metaheuristics* (pp. 287-319). New York: Kluwer Academic Publishers.

Hillier, F. S., & Liebermann, G. J. (1999). *Operations research* (2nd ed.). Delhi, India: CBS publishers and Distributors.

Kim, I. S., Lee, S. H., & Yarlagadda, P. K. D. V. (2003). Comparison of multiple regression and back propagation neural network approaches in modelling top bead height of multipass gas metal arc welds. *Science and Technology of Welding and Joining*, *8*(5), 347-352.

Klir, G. J., & Yuan, B. (2002). *Fuzzy sets and fuzzy logic, theory and applications* (6th ed.). Prentice Hall of India Private Limited.

Koulamas, C., Antony, Sr., & Jaen, R. (1994). A survey of simulated annealing applications to operations research problems. *Omega*, *22*(1), 41-56.

Malkin, S. (1984). Grinding of metals: Theory and application. *Journal of Applied Metalwork*, *3*, 95-109.

Markos, S., Viharos, Z. J., & Monostori, L. (1998). *Quality-oriented, comprehensive modelling of machining processes.* Paper presented at the 6th ISMQC IMEKO Symposium on Metrology for Quality Control in Production (pp. 67-74).

Mitra, A. (2001). *Fundamentals of quality control and improvement* (2nd ed.). India: Pearson Education Asia, Inc.

Mukherjee I., & Ray, P.K. (in press). Optimal process design of two-stage multiple response grinding processes using desirability functions and metaheuristic techniques. *International Journal of Applied Soft Computing.*

Nair, V. N. (Ed). (1992). Taguchi's parameter design: A panel discursion. *Technometrics*, *34*, 127-161.

Reeves, C. R., & Rowe, J. E. (2002). *Genetic algorithms: Principles and perspective, A guide to GA theory*. New York: Kluwer Academic Publishers.

Ross, P. J. (1989). *Taguchi techniques for quality engineering.* McGraw-Hill Book Company.

Salhi, S. (2002). Defining tabu list size and aspiration criterion within tabu search method. *Computer and Operations Research*, *29*, 67-86.

Taguchi, G. (1996). The role of D.O.E. for robust engineering: A commentary. *International Journal of Quality and Reliability Engineering*, *12*, 73-74.

Tao, L. (2004). *Research incubator: Combinatorial optimization* (Tech. Rep.). Pace University.

Widrow, B., & Lehr, M. A. (1990). 30 years of adaptive neural networks: Perceptron, madaline, and back propagation. *Proceedings of IEEE*, *78*(9), 1415-1442.

Zadeh, L. A. (1973). Outline of a new approach to the analysis of complex systems and decision processes. *IEEE Transactions. Systems, Man, and Cybernatics*, *SMC*, *3*, 28-44.

Chapter XX
Intelligent Laser Scanning of 3D Surfaces Using Optical Camera Data

A. Denby
The Nottingham Trent University, UK

J. F. Poliakoff
The Nottingham Trent University, UK

C. Langensiepen
The Nottingham Trent University, UK

N. Sherkat
The Nottingham Trent University, UK

ABSTRACT

In CAD/CAM, reverse engineering involves obtaining a CAD model from an object that already exists. An exact replica can then be produced, or modifications can be made before manufacture. Single-perspective triangulation sensors provide an inexpensive method for data acquisition. However, such sensors are subject to localised distortions caused by secondary reflections or occlusion of the returning beam, depending on the orientation of the sensor relative to the object. This chapter describes an investigation into integrating optical camera data to improve the scanning process and reduce such effects, and intelligent algorithms, based on image analysis, which identify the problem regions, so that the sensor path and orientation can be planned before the scan, thereby reducing distortions.

INTRODUCTION

Ideally an object is designed on a CAD system to provide the data needed to control the CAM equipment to manufacture the object. However, there is often a need to copy objects for which no prior CAD data are available, for example, when making replacement parts. Machining such objects by hand is possible but expensive, as is redesigning the objects on a CAD system. Therefore there is a real need for an inexpensive method for generating the required data from the object which maintains an acceptable degree of accuracy.

One approach is to use a laser sensor to measure the surface. Unfortunately, laser scanning is subject to localised distortions, which are often caused by occlusion or secondary reflections of the beam, depending on the orientation of the laser head relative to the object. Without prior knowledge of the object, a 'blind' scan must be implemented. We have investigated the integration of an optical camera into the system to provide such knowledge. Image analysis allows the path and orientation of the laser sensor to be planned before the scan, thereby reducing the distortions. Scanning time can also be shortened by reducing scan resolution in 'low interest' regions.

It has been found that simple edge detection algorithms such as Canny can determine a single best orientation, but a combination of algorithms is needed to eliminate noise and create continuous edge segments, which can then be used to develop scan regions of appropriate orientation. We have developed new vectorisation algorithms to identify edge segments. Calibration of the camera image relative to the scanner is important to avoid errors. Discrepancies between scan data from different orientations can be prevented by careful calibration of the scanner rotation system.

BACKGROUND

In traditional 'forward' engineering, concepts and models are transformed into real parts. Reverse engineering starts with real parts or prototypes and transforms them into engineering models. Typically, the process begins by measuring an existing object to provide a model, in order to exploit the advantages of CAD/CAM technologies. Such techniques are used in a wide variety of applications, including medicine and animation as well as more traditional production. A typical application is the re-engineering of an existing structure for input into a CAD or other 3D modeling program, where analysis and modifications are required to make a new, improved product. The data acquisition phase is a crucial step in this procedure and data acquisition methods can be either tactile or nontactile.

Laser triangulation is a popular nontactile data acquisition method in which a laser beam is projected onto the surface of interest and the reflected spot is detected by one or more photosensitive devices. The position of a surface point is then calculated using triangulation. Laser triangulation can acquire data at very fast rates; however the technique is subject to errors, as shown in Figure 1. Before describing the types of errors that occur, we explain how the triangulation process works.

The laser scanner consists of a unit with an emitter and detector which moves over an object and outputs readings corresponding to the distance of the object from the scanner. The emitter projects a laser beam onto the object in a (normally) vertical direction, as shown in Figure 2. For ease of explanation we have based our descriptions on the assumption that the beam is vertical but the principle can equally well be used for other configurations. The sensor detects light returning from the spot where the beam hits the object and measures the direction at which it returns, here indicated by angle φ. The distance of the spot from the emitter can then be calculated by $h = w \cot \varphi$, where w is the distance between emitter and sensor. As the scanner moves over the object, the positions of points on the surface are measured and collected to form the point cloud.

Our group has worked for many years at the interface of engineering and computer science. We have made contributions in the fields of control of CNC machines (Chow, Poliakoff, & Thomas, 2002; Poliakoff, Chow, Orton, Howson, & Al-Dabass, 2005) and measurement of surfaces (Denby, Langensiepen, Poliakoff & Sherkat, 2005; Sacchi, Poliakoff, Thomas & Häfele, 2004; Wong, Poliakoff, & Thomas, 2001). The work described in this chapter arose from an investigation into the errors that occur in laser triangulation. Wong (2002) found that the majority of these errors fall into three broad categories: (systemic) noise, transitional errors caused by changes in reflectivity across the object, and errors due to the geometry of the object. Wong investigated how to reduce the geometric errors in the third category, which are caused when the object itself interferes with the measurement of the height of the primary spot. He found that sometimes there are secondary reflections of the light from the primary spot onto other parts of the object within the sensor's field of view; if these are also detected by the sensor, there will be an error in the reading obtained. In other cases the primary spot could be occluded

Figure 1. A single scan of a small bottle top with orientation parallel to the x axis (i.e., left to right). Note: The vertical scale is exaggerated to help show the errors, which can be seen where the edge is roughly perpendicular to the sensor orientation: upward and downward 'spikes' on the left; smaller 'bow wave' errors on the right, which extend further from the object. To view a colour version of this figure, please see http://www.igi-global.com/downloads/pdf/laha/20color.pdf

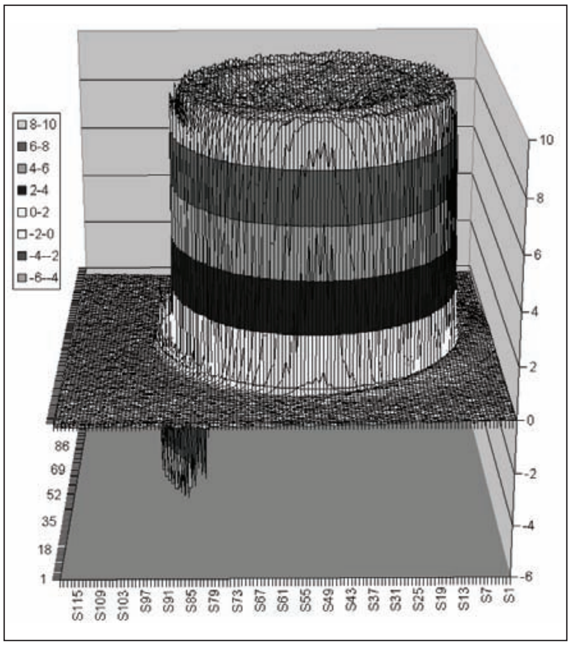


Figure 2. Diagram to illustrate how the heights of different parts of the object are measured using triangulation. The triangulation angle, φ , varies as the height changes; the height is calculated as h = w cot φ, where w is the distance between emitter and detector. To view a colour version of this figure, please see http://www.igi-global.com/downloads/pdf/laha/20color.pdf

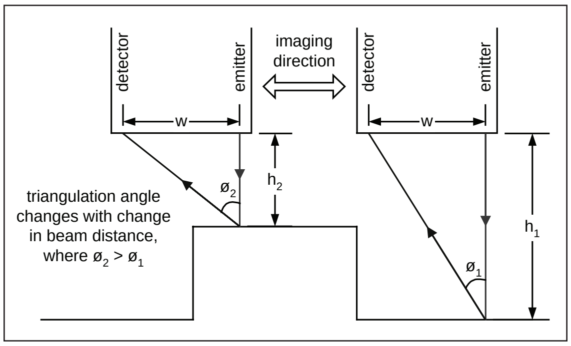
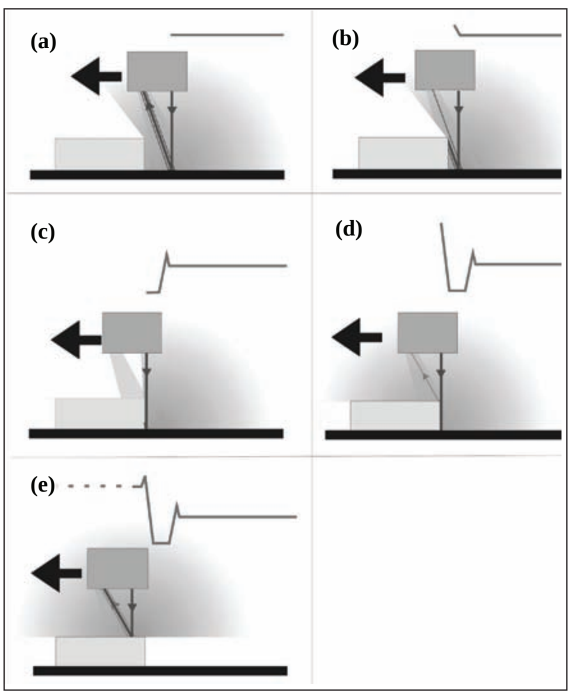


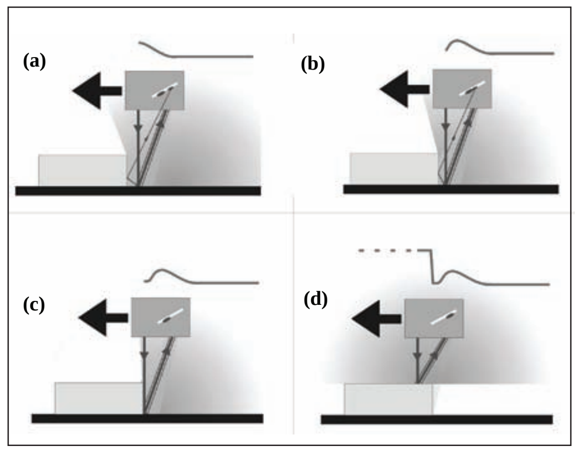
Figure 3. Illustration of how the 'spikes' distortion occurs for a simple object. The scanner is shown moving from right to left and building up the output profile in green, although the errors obtained do not depend on the direction of motion. Note: At (a) and (e) there are no distortions. Occlusion begins at (b) with part of the beam occluded and a small rise in the output. At (c) the beam is completely occluded and the output is zero, giving a trough, whereas at (d) part of the spot is on the upper surface and another small rise is seen in the output. If the surface has high reflectivity, light from secondary reflections during complete occlusion (c) may produce the large rise or 'spike' effect. To view a colour version of this figure, please see http://www.igi-global.com/downloads/pdf/laha/20color.pdf



from the sensor by another part of the object, and then the error is likely to be even larger.

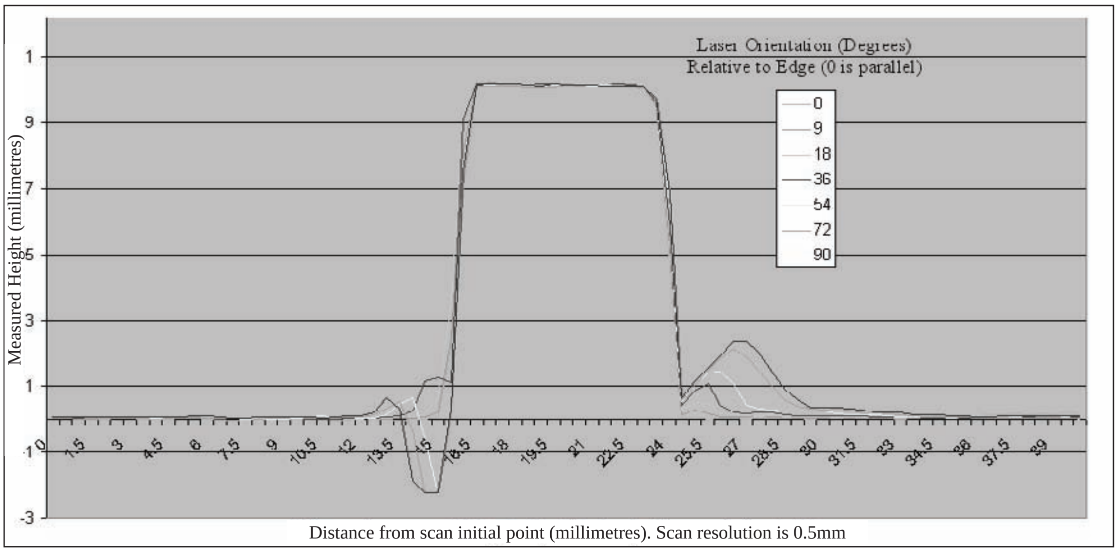
Figures 3 and 4 summarise Wong's findings. When the reflected signal is occluded from the detector by part of the object, for example, at

Figure 4. Illustration of how the 'bow wave' distortion occurs for a simple object. The scanner moves from right to left, as in Figure 3. The bow wave grows as it approaches the object as secondary reflections cause the output to rise until a maximum bow wave is reached at (a). Note: At (b) it is still closer to the object, so that the secondary reflections have a smaller effect. Just before the spot reaches the edge (c) the output dips further. When the spot is on the top of the object at (d) there are no distortions. To view a colour version of this figure, please see http://www.igi-global.com/downloads/pdf/laha/20color.pdf



(b), (c), and (d) in Figure 3, false readings with substantial errors are obtained (which we refer to as 'spikes'). In other cases, for example, at (a) and (b) in Figure 4, secondary reflections also cause false readings, for which the errors are not so large but more extensive (which we refer to as 'bow wave'), as also shown in Figure 1. These 'geometric' errors close to the edges of the object are worst when the laser scan head is oriented approximately *perpendicular* to the edge (to the left and right in Figure 1). Whereas, if the laser is oriented *parallel* to the edge, the errors are minimised. Figure 5 shows how the errors increase as the relative orientation changes from parallel towards perpendicular.

Figure 5. A plot of the scan output for different orientations of the scanner, showing how the distortion are least at 0º, when the scanner is oriented parallel to the edge, and largest at 90º, when it is perpendicular. The scale in the vertical axis is exaggerated. To view a colour version of this figure, please see http://www.igi-global.com/downloads/pdf/laha/20color.pdf



Wong addressed these 'geometric' errors by exploiting the fact that knowledge of the orientation of an edge will allow the system to identify the laser unit orientation likely to produce least distortion near that edge. Initially a number of complete scans of the object are made with different orientations. Intelligent software, based on comparisons between the scans, is used to identify problem regions and associate an edge with each such region. Then, the appropriate scan is selected with orientation chosen to minimise the error on that region. Wong's method has resulted in much improved scan quality. Unfortunately, the major drawback of this approach is a greatly increased overall scan time, because at least eight complete scans (at 45º intervals) are needed. Typically a single scan of 0.05mm pitch requires about 1 hour to execute for an object of 30 sq. cm, so eight complete scans require about 8 hours. Our latest investigation has aimed for similar improvement but without the need for repeated complete scans.

OPTICAL CAMERA DATA INTEGRATION

Most of the data collected using Wong's approach are eventually discarded, having only been used in order to identify error regions and select the appropriate scan orientation for each such region. Our new approach is to recognise the problem regions *prior* to the laser scan process, in order to determine an optimal scanning strategy for the object. By integrating an optical camera with the laser triangulation system we can obtain knowledge of the object's geometry prior to scanning. Then we can identify regions where optimal choice of sensor orientation is crucial, rather than relying on postprocessing of multiple

Figure 6. Overview of method 1

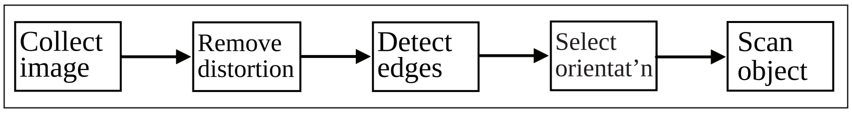
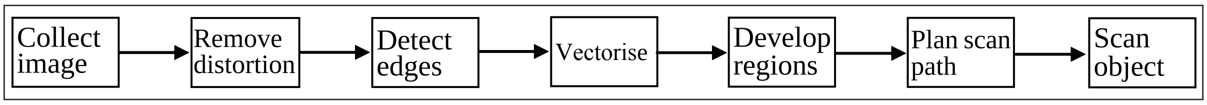


Figure 7. Overview of the methods 2 and 3



scans to do this. The path planning then involves selecting the best sensor orientation for different regions of the object, thereby minimising such 'geometric' errors. However, unlike Wong's method, much redundancy in scanning the same region many times is avoided, thus reducing the total scan time.

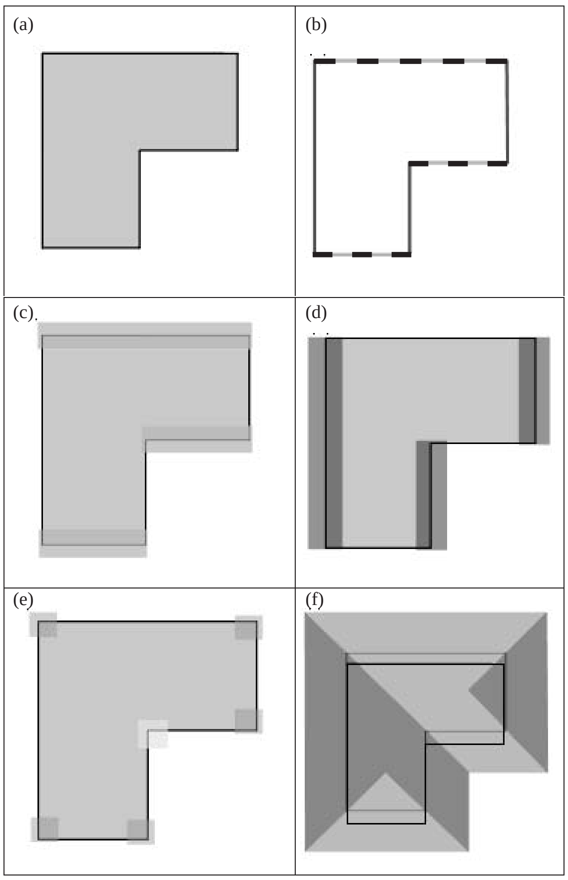
We use information captured from a simple (2D) CCD camera as an a priori guide to orientation and path-planning for the scanning laser. The camera image is analysed using edge detection techniques and then vectorised to provide an edge map. A scanning path plan that minimises errors due to object geometry is generated based on the edge map and corresponding region-segmented images.

The simplest strategy, Method 1, involves performing a single scan with a fixed orientation, as illustrated in Figure 6. The (small) improvement here is that the orientation for the scan is chosen to minimise the number of error points in the point cloud. This method is not much slower than a standard single scan, because the only additional time is for the edge detection software to run. However, it has the disadvantage that generally there will still be many distortions in the resulting point cloud, so we include it merely for comparison.

A better strategy, Method 2 (Figure 7), is to generate a number of different partial scan paths, each covering a number of regions in the image that share a common 'optimal' scan orientation. Figure 8(a) through (e) illustrates the idea behind this method for a simple L-shaped object. A number of partial scans are then executed, with minimal overlap between them, with the scan head being reoriented between the scans. Remaining parts can be scanned with any orientation. This method is slower than a single scan but faster than Wong's method. Where the regions overlap some procedure is needed to resolve any conflict between the data from the different regions, for example by simple averaging or by some sort of blending. If the regions can be chosen without any overlaps, then this method can be replaced by the approach of Method 3.

As illustrated in Figures 7 and 8(f), Method 3 uses a single scan sweep over the object with continuous real-time automatic optimisation of the laser scan head orientation. This method requires careful monitoring to ensure continued accuracy throughout the scan and requires equipment that is capable of controlling the rotation of the scan head as needed in real time. Again this method is faster than Wong's method but slower than a single scan. It may be faster than Method 2 but that could depend on the complexity of the object's geometry. The time taken may have to be slower than a single scan with fixed orientation, in order to accommodate the real-time rotation of the scan head.

Figure 8. Illustration of the idea behind methods 2 and 3 for a simple L-shaped object (a). Note: The detected vectors are shown in (b), with dashed lines representing horizontal (0º) and solid vertical (90º). Here Method 2 uses scan region Algorithm A and the scan regions for 0º and 90º are shown in (c) (shaded) and (d) (darker shaded), respectively. The places where the scan regions overlap are shown in (e), where the shaded outer corners are expected to give good results, because of their location. In the case of an inner corner, it may be difficult to resolve the output value. Method 3 uses scan region Algorithm B and there are no overlaps (f); the darker region is to be scanned at 0º and the lighter scanned at 90º. (The grey region can be scanned with any orientation.) To view a colour version of this figure, please see http://www.igi-global.com/downloads/pdf/laha/20color.pdf



OUR INVESTIGATION

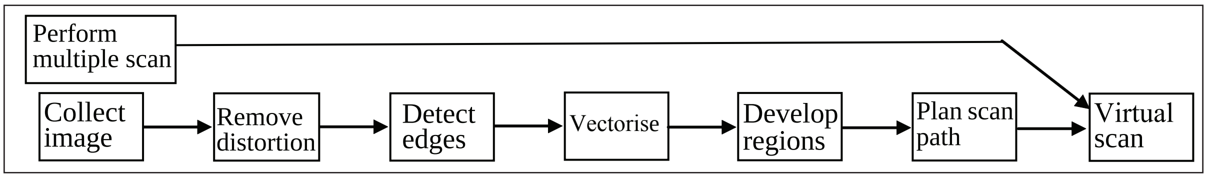
In order to investigate the three methods described above, we chose a slightly different approach. Because the scanning process is very time-consuming, we wanted to avoid performing many experiments with different path plans. Therefore we came up with the strategy of virtual scans based on a number of complete scans covering all possible scan orientations within a certain approximation. Then a path plan could be 'tested' using a virtual scan, by selecting the required values from the scan data for the relevant orientation. Further work would be needed for Method 3 but many questions can be answered by using such a virtual scan.

So, for each object a set of 10 complete scans was obtained at 18° intervals, thus covering a range of 180°. This then provided a scan of any orientation to within 9° of the required orientation, because two orientations 180° apart can be covered by the same scan. Again for further refinements of the proposed methods some parts of the other 10 scans might be needed.

In this way we could also perform a virtual partial scan based on the scan data already collected. Each virtual partial scan is performed by choosing the appropriate data from the pre-existing scan. Although the initial collection of complete scan data for all the orientations was time-consuming, it avoided much repeated performance of partial scans during the investigation. This approach does not allow us to choose every possible orientation but we have found that it is sufficient to demonstrate the principle. Figure 9 is a modification of Figure 7 showing how the virtual scan is incorporated.

The next section describes the image processing and edge detection algorithms needed to produce a map of all the edge pixels from the image. From these the optimum orientation can be found for Method 1. The vectorisation and region development algorithms are presented in the fol-

Figure 9. Overview of the scan process used for our investigation



lowing section, which can then be used to produce the path plans needed for Methods 2 and 3.

IMAGE PROCESSING FOR EDGE DETECTION

Since colour images provide more information than grey value images, more detailed edge information might be expected from color edge detection, however Novak and Shafer (1987) found that 90% of edges are approximately the same in grey value and in color images, although it is possible that the remaining 10% may make a significant difference to the overall edge continuity. Although colour image output was available from the digital cameras, it was decided initially to implement several common edge detection methods using greyscale images to test the hypotheses presented in this project because of the complexity of colour edge detection methods. For this project it was often necessary to spray-coat the objects being scanned and imaged in order to reduce the chance of specular reflections occurring in the scan data. The sprayed surface means that images are generally monochromatic so there is no advantage (at this stage) in using colour images.

The image processing begins with compensation for distortions in the image caused by deficiencies in the camera. Before edge detection is performed it may be useful to smooth the image. For both smoothing and edge detection the method of 'convolution' is often used. At each position a mask is used to calculate a weighted sum of nearby pixels to replace the original pixel value. For smoothing, the mean filter takes a simple average, while the gaussian mask is symmetrical with the highest weight in the centre.

Edge detection involves the identification of pixels on the image where there are discontinuities or abrupt changes in grey level (intensity) or colour, or the gradient of this intensity or colour. Edges usually correspond to significant variation in reflectance, illumination, orientation, and depth of surfaces and are typically associated with photometric, geometric, and physical characteristics of objects within the image (Ziou, 1998). A wide range of methods have been used for edge detection, some of which are described here. First and second derivative operators have been used to identify edge pixels, which then require further processing to thin the detected edges. Other approaches include region growing, which can identify regions which are not edges, and segmentation using texture, as explained below.

Examples of first derivative operators are the Roberts Cross operator (Ziou, 1998), the Sobel operator (Lyers, 1988), the Prewitt gradient method (Prewitt, 1970), and the Frei-Chen method (Ziou, 1998). These all involve a different square masks for a 'convolution' with the image to produce a new output image (Low, 1991). The Canny edge detector (Canny, 1986) employs a gaussian smoothing function and simple first derivative masks for detection. The technique was extended by Deriche (1987) to employ both a gradient magnitude

Table 1. Comparison of time to execute first derivative operators (640×480 greyscale image)

Operator	Mask Size	Number of Masks	Mean time to execute (ms) (averaged over 5 samples)
Robert's Cross	2x2	2	245.3
Prewitt Gradient	3x3	2	325.5
Sobel	3x3	2	320.6
Frei-Chen	3x3	9	850.5
Canny (convolution only)	1x3	2	233.5

and a gradient orientation map to determine the direction for the edge-tracking and to assign an orientation to each edge pixel.

In conjunction with a thresholding method, such as the edge-following technique used by the Canny operator, most of these edge detectors provide a useful edge response for vectorisation. However the Frei-Chen masks do not provide a means of recovering orientation information. A selection of established edge detection algorithms have been implemented using a common code platform in order to compare them in a controlled manner and provide unbiased results. The Roberts Cross, Prewitt gradient, Sobel, and Frei-Chen methods were implemented and compared with the Canny edge detector without the gaussian smoothing and the results are shown in Table 1.

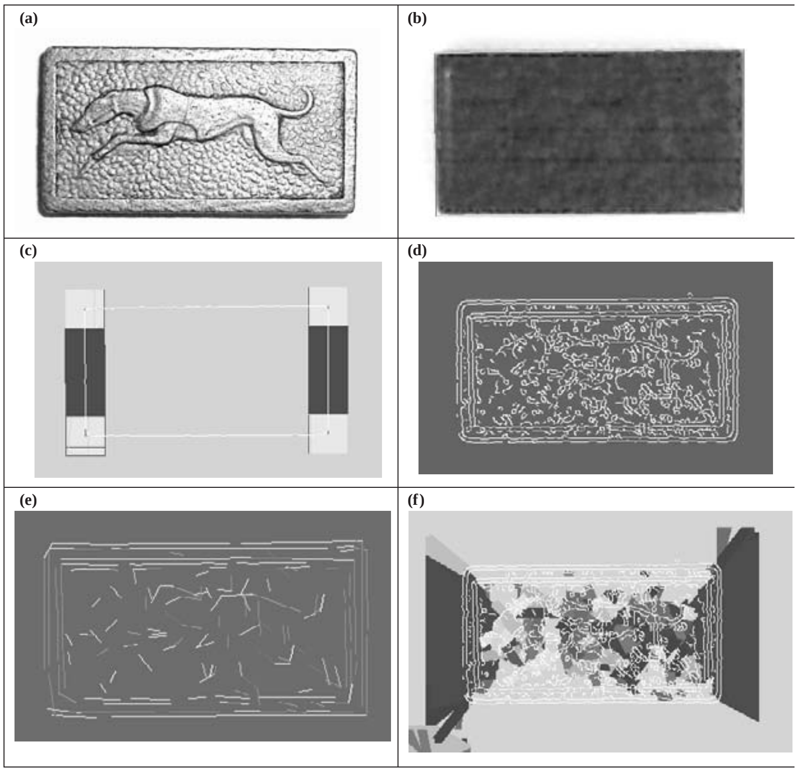
The Sobel operator also generally produces considerably higher output values for similar edges, compared with the Roberts Cross. The Roberts and Sobel gradient masks are more sensitive to diagonal edges. The Prewitt gradient mask is more sensitive to horizontal and vertical edges. The Frei-Chen edge detector has equal sensitivity for diagonal, vertical, and horizontal edges. However, the time taken is more than three times longer than for the Canny edge detector, which we therefore used for our investigation.

A number of edge detection algorithms have been evaluated for their suitability for the identification of edges in physical objects for path planning in preparation for laser scanning. Our findings support the view of Heath, Sarkar, Sanocki, and Bowyer (1998) that the Canny operator provides the best general edge detector where the parameters can be tuned for each image.

The Canny algorithm uses gaussian smoothing followed by computation of the gradient magnitude and direction, giving a pair of values for each pixel. The edges are then localised by using a process of 'nonmaximal suppression,' which provides an improvement to the basic skeletonisation technique. The idea is to determine the local maximum of the gradient magnitude and then track along the top of the gradient 'ridge' in both directions along the orientation of the expected edge (i.e., perpendicular to the gradient direction). A pixel is considered to be the local maximum in this context if the magnitude of the gradient for that pixel is greater than that of the two neighbouring pixels in the direction of the gradient. All edge pixels that are not maximal (i.e., on the top of the edge) are set to zero, giving a line of single pixel width in the output.

Thresholding techniques are reviewed by Sezgin and Sankur (2004) and we use thresholding as a method to localise the edges. The tracking process is controlled by two thresholds: *Tmax* and *Tmin* where *Tmax* > *Tmin*. Edge tracking can only begin at a point on a locally maximal edge pixel with a gradient modulus higher than *Tmax*. Tracking then continues in both directions out from that point until the height of the ridge falls below *Tmin*. This edge hysteresis helps to ensure

Figure 10. Image processing results for a black domino, initially unsprayed. Note: (a) photograph of the domino; (b) vectors; (c) scan regions using Algorithm A with detected edges superimposed in white. After the domino has been sprayed, more detail is detected: (d) edges; (e) vectors; (f) scan regions using Algorithm B (where the regions at bottom left were caused by an alignment spot). The different colours indicate different orientations, for example, turquoise for 0º. To view a colour version of this figure, please see http://www.igi-global.com/downloads/pdf/laha/20color.pdf



that noisy edges are not broken up into multiple edge fragments. Usually, the upper tracking threshold can be set quite high and the lower threshold quite low for good results. Setting the lower threshold too high will cause noisy edges to break up. Setting the upper threshold too low increases the number of spurious and undesirable edge fragments appearing in the output.

Texture information is an important consideration for edge detection but we have not used it here. Pure texture segmentation gives only a coarse segmentation, so it can only be used as an auxiliary tool to check segmentation and texture parameters for the segmented regions. Haralick, Shanmugam, and Dinstein (1973) provide definitions of texture and derive a number of texture parameters including contrast, correlation, direction, entropy, homogeneity, and uniformity. One approach commonly used to handle texture is to smooth the image using a gaussian or other blurring filter. However, this can cause problems, because, as the strength of the blurring filter increases, it becomes more difficult to detect the position of the edges accurately (i.e., edge

Figure 11. The combined scan results for the domino using the plan in Figure 10(c), showing downward spike errors (indicated by the arrows) in the top surface of the domino. The scale in the vertical axis is exaggerated to help show errors. To view a colour version of this figure, please see http://www.igi-global.com/downloads/pdf/laha/20color.pdf

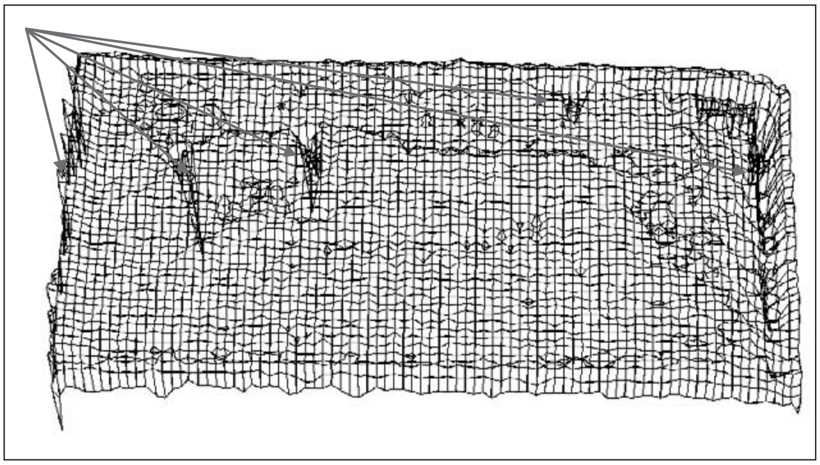
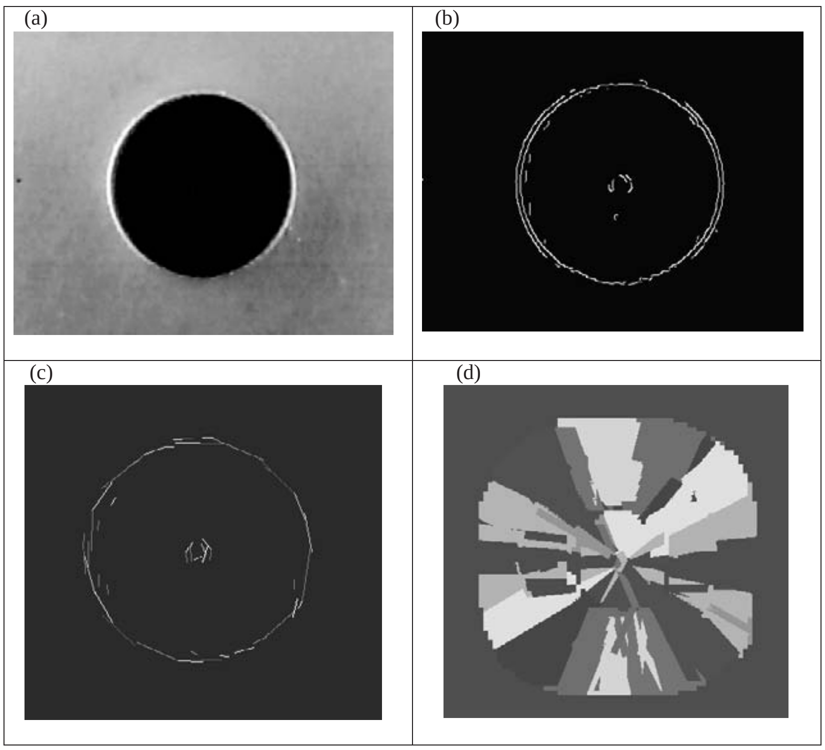


Figure 12. Image processing results for the small bottle top from Figure 1, which was also sprayed: (a) photograph of the object; (b) edge detection; (c) vectorisation and (d) scan regions obtained using Algorithm B. Again, the different colours indicate different orientations, for example, turquoise for 0º. To view a colour version of this figure, please see http://www.igi-global.com/downloads/pdf/laha/20color.pdf



localisation suffers) and also fine detail which we may want to keep becomes lost along with the texture 'noise.'

Corner finding has also been investigated in the context of edge detection. Kitchen and Rosenfeld (1982) use a local quadratic fit to find corners. Wang and Binford (1994) model the effects of shading on the direction of the image gradient, creating a detector insensitive to shading. In practice most corner detectors are usually not very robust and often require expert supervision to prevent the effect of individual errors from dominating the recognition task. Smith and Brady (1997) propose the smallest univalue segment assimilating nucleus (SUSAN) corner detector, which performs well in inherently noisy 'real world' images because it does not rely on image derivatives. It may be that corner finding could be used to identify potential problems where two or more edges meet but we have not used it so far.

Split-and-merge edge detection is a two stage process that combines the advantages of both region growing and region splitting methods. First the image is split recursively until each region in the image meets the specified criteria of homogeneity then adjacent regions are merged together if they satisfy the same criteria. Because both split and merge processing options are available, the starting segmentation does not have to satisfy any of the homogeneity conditions. In more complex images even more complicated criteria may not be enough to give acceptable results. The split-and-merge techniques introduced by Chen and Pavlidis (1981) and later developed by Spann and Wilson (1985) use a linked pyramid and statistical decision criteria to combine global and local region information.

The white lines in Figures 10(c), 10(d), and 12(b) show the results of edge detection for the domino shown in Figure 10(a), unsprayed and sprayed, respectively, and the bottle top shown in Figure 12(a), sprayed. The next stage is to vectorise the identified edge pixels so that they can be used as the basis for developing scan regions, as described in the next section.

VECTORISATION AND SCAN REGION DEVELOPMENT

Many vectorisation methods have been proposed in the scenario of converting line drawing images (e.g., engineering drawings) directly into CAD models, which is in many ways analogous to the process that is being investigated in this project. However none of the methods proposed work perfectly (Tombre, 1998). The output of vectorisation should represent the shape of original image as faithfully as possible but, the vectors do not always correspond with 'ground-truth' graphic objects (Tombre, 2000). Some postprocessing is usually necessary to rebuild graphic objects from vectors with geometric constraints. Many of these methods are based on skeletonisation or other thinning methods that are not required here (since the edges are already thinned to a single pixel width by the non-maximal suppression process). One method of storing the component pixels in vectorisation is chain encoding, such as Freeman chain codes (Freeman, 1970) in which the direction of neighbouring edge pixels are stored sequentially according to their relative positions. Line and arc-fitting algorithms are often employed to convert the original image into a low-level vector format represented by short line segments and short arcs. Line-fitting methods are popular but since they depend on approximating a sequence of adjacent line segments, a ground-truth line cannot be recovered correctly if some parts of it are missing or have serious distortions (Hori, 1993).

The Hough transform is one method that has often been used to detect known geometrical shapes such as lines and circles (or other known shapes) in images (Ballard, 1981). The main advantage is that it is tolerant of gaps in feature boundaries and is relatively unaffected by image noise. However it is computationally- and memory-intensive (Risse, 1989). A number of authors have proposed improvements to the standard Hough transform to improve line detection and localisation. Ji and Xie (2003) review the approaches taken by a number of authors contributing to the use of Hough transforms, as well as propose a method by which edge localisation may be improved. Conceptually, the Hough transform considers lines at all possible positions and orientations and counts the number of pixels that fall on each line. A 'transform space' is created where points in the 'Hough space' map to lines in the image space. This method finds many lines in the image, but we have found that it has several unwanted effects. First, quantisation of the pixels in the image space and of the accumulator cells in Hough space can lead to a cluster of points in Hough space. When these points are mapped back into the image space they form a group of lines of slightly different orientations intersecting at a common point, giving a 'bow tie' effect. Secondly, lines that pass through many pixels are favoured by the accumulator because there is more pixel evidence for them, which means that shorter lines will be removed by thresholding. Such short lines are equally valid as edges which cause the distortions with which we are concerned. We investigated a 'windowed' Hough approach to avoid losing shorter lines but found it even more computationally intensive than the standard Hough.

For this application we have developed a vectorisation algorithm which can overcome the problems encountered with the other methods. The idea is to select each edge pixel in the image and attempt to 'grow' a straight line starting at that pixel. All eight neighbouring pixels are examined in turn and when an edge pixel is found the 'growth' can begin. The line (defined by a start and end pixel) is grown, together with an accompanying list of pixels which have contributed to the development of the line. Growing continues until a stopping condition is reached. The first stage is to seek to extend the line by looking for an edge pixel, first at the pixel closest to the extension of the line and adjacent to the end pixel of the line, and then at the pixels on either side (relative to the line). If one or more new edge pixels are found, the one closest to the line is added to the list of pixels and a new line is fitted by least squares fitting (see below for more details). Then the signed distances of all the pixels in the list from the new line are calculated. The growth is stopped if one (or more) of the following cases holds:

1. No extension is possible (no edge pixels found).
2. The new edge pixel is already in the list.
3. The new line has a different start pixel from the original line.
4. The change in direction of the new line compared to the previous one exceeds a given tolerance: the tangent of the tolerance angle is inversely proportional to the length of the line (so the tolerance angle is reduced as the line grows).
5. The distance of one or more of the list of pixels from the new line exceeds a given tolerance.
6. The sum of the signed distances exceeds a given tolerance.
7. The growth becomes one-sided (i.e., there is a sequence of pixels in the list with the following property: the first is at a distance greater in magnitude than the mean distance, all are on one side of the fitted line and the number at greater distance than the previous one exceeds a given tolerance).
8. The line overlaps other lines already discovered, either completely or very closely, so it can be discarded, because nothing new will be found.

The mean distance is calculated as the sum of the signed distances of all the pixels from the fitted line. After the least squares fitting has been done, the start and end pixels of the new line are chosen as follows. If the magnitude of the gradient of the fitted line is less than 1, the *y* value of the start pixel (respectively end pixel) for the new line is changed, if necessary, so that it lies on the fitted line, and otherwise the *x* value is changed.

We have found that the above algorithm sometimes produces lines which deviate considerably from the previous trend by the addition of the last few pixels, with the effect of moving the line direction too far in small increments. This situation is not prevented by conditions 5, 6, and 7, unless the tolerances are made smaller to prevent this. However, with the smaller tolerances many lines are terminated prematurely. Therefore when the growth is stopped, an attempt is made to 'rewind' the line until the magnitude of the distance for the last pixel is less than the magnitude of the mean distance for the rewound line. This is done, rather than merely testing for the above condition at each stage of growth, because sometimes the line direction can 'return' closer to the trend after a small deviation.

In order to keep the vectors fitted closely to the edges discovered in the image the initial vectorisation tolerance parameters must be quite strict. Simple (artificial) images are reduced to a minimum sufficient set of vectors quite efficiently using this method, however in 'real' images there is often some redundancy in vectors, given that the edges are often curved and several straight lines may be fit around that curve. This often leads to the generation of many short line segments, which can sometimes be merged. There are also situations where discovered vectors can be merged, because the initial tolerances caused the break-up of long vectors into shorter segments. The vectorisation algorithm may also result in the generation of parallel overlapping segments from a single line. In such cases it is possible to merge these segments into a single vector using a 'combine' algorithm. Figures 10(b), 10(e), and 12(c) show the results of vectorisation for the edges shown in Figures 10(c), 10(d), and 12(b) as white lines. The circular shape of the bottle top consists of short lines at different orientations.

The development of scan regions based on the vectors can be done in various ways. We have implemented two different such algorithms, A and B. The first of which is only appropriate for Method 2, because it has overlapping regions. The second has no overlapping regions, so it can be used for either Method 2 or Method 3.

Algorithm A

For every vector a scan region is developed around the vector in order to cover all potential error points associated with the corresponding edge in the object. If the object is scanned with the sensor oriented parallel to the vector, and therefore to the edge, the error should be minimised. For parts where there are no scan regions the orientation of the scanner should not affect the result, so it can be chosen to be the default value, or some other if it is more convenient. Figure 8(a) through (e) illustrates this algorithm for the case of a simple L-shaped object. The width of the scan region needs to be large enough to cover all potential bow waves, because their extent is always greater than that for the spike distortions. Thus the width will need to be larger as the height of the object increases. The problem with this approach is that there are often overlaps between two or more scan regions and then the conflict between the different orientations needs to be resolved. We have taken a simple average of all the scan regions involved and, unfortunately, this can lead to errors. A better approach would be to use another algorithm for these problem areas, such as Algorithm B or even Wong's method. Both these algorithms take longer but the additional time would not be very large in many cases, provided that the scan region widths are made as small as possible.

Algorithm B

This algorithm uses the 'nearest vector' approach, whereby it is assumed that the vector (and therefore the edge) most likely to affect a point is the one nearest to that point. Thus there are no overlaps but the time taken to develop the regions is much greater. Figure 8(f) illustrates this algorithm for the simple L-shaped object from Figure 8(a) with vectors shown in Figure 8(b). In order to reduce the time taken, we have developed the idea of 'regions of influence,' so that vectors which are not within a region of influence associated with a point can be assumed to have no influence on the errors at that point. The way we have done this means that some scan regions encroach into the default regions but that will not have a detrimental effect. The jagged edges of these encroachments can be seen in Figure 12(d).

CALIBRATION

In order to calibrate the image space against the scan space, calibration markers (small circular disks painted with a black and white quarter pattern) were placed at known coordinates on the laser scan bed around the object to be scanned, using the laser itself as a guide to positioning the markers. The markers were placed around the object, then the camera was positioned over the object and the height adjusted until all the calibration markers were visible in the camera's field of view. Once the coordinates have been found for these known points in the image space, the parameters for an affine transformation can be calculated, which can the be used to transform between the two spaces (Foley, Van Dam, Feiner, Hughes, & Phillips, 1994).

The camera needs to be correctly aligned so that it is facing the scan bed perpendicularly in order to minimise parallax errors in position of edges. This can be validated by moving the camera up and down and checking that a mark at the centre of the image remains at the centre, and correcting if necessary. For objects which are more than 3-4cm across, a correction will also be needed to allow for the parallax effect. Again this can be done by measuring the change in position of each edge as the camera moves up and down. This also opens up the possibility of automatic self-calibration of the system.

It is important that the laser scanner and sensor are set up such that the beam is aligned as closely as possible with the z axis of the CNC machine. If it is far from vertical, then some of the errors we are trying to avoid will be exacerbated and will be harder to avoid. The direction of the laser beam should be close to vertical but a small discrepancy can be compensated if both the discrepancy in angle and its orientation are known. All three of the corrected coordinates of the measured point will generally be slightly different from the values read in.

It is also important that the scanner rotation mechanism is aligned as closely as possible with the z axis of the machine. Again a small discrepancy can be compensated, provided that its parameters are known. If the emitter is not on the axis of rotation, then the spot will move in a circle as the scanner is rotated. For Method 2 this can be compensated once the radius is known. However, for Method 3 it would be difficult to accommodate more than a very small radius, because the measured point would be too far away from the intended position.

RESULTS

Figure 1 shows the results obtained for a single scan of the small bottle top with the sensor oriented parallel to the x axis. No distortions are seen when the direction of the edge is roughly parallel to the x axis. But there are distortions when the edge is close to the direction of the y axis with 'spikes' at the top on the left and 'bow wave' near the base on the right.

Figure 13. Combined scan results for the sprayed domino from Figure 10 using (a) Algorithm A and (b) Algorithm B. Again the scale of the vertical axis is exaggerated. To view a colour version of this figure, please see http://www.igi-global.com/downloads/pdf/laha/20color.pdf

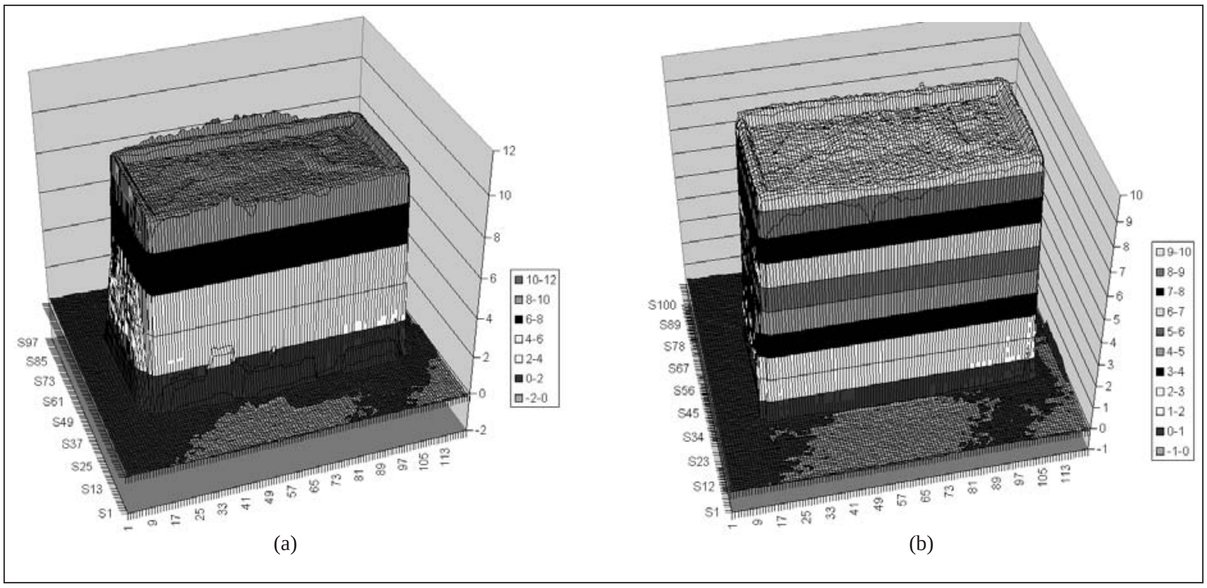


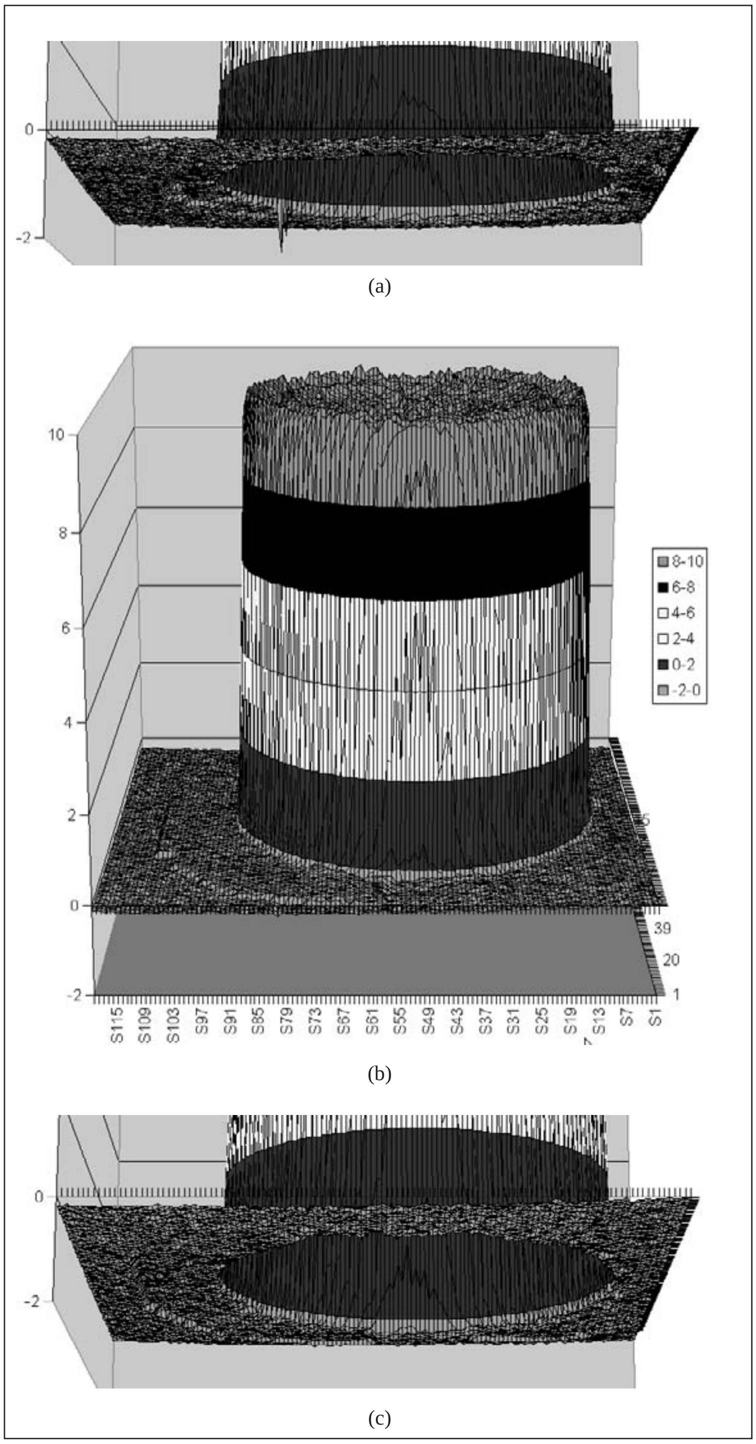
Figure 10 shows the results obtained for a domino with slightly rounded corners. In Figure 10(b) and (c) the edge detection, vectorisation, and regions obtained by Algorithm A look promising. Unfortunately the scan of the top of the domino (Figure 11) shows several downward spikes. This has occurred because of the low reflectivity of the black domino. Therefore we sprayed the domino with white powder and repeated the experiments as shown in Figure 10(d) through (f). The regions obtained for this by Algorithm A are not shown, because the overlaps make it confusing to interpret, but those for Algorithm B show how the scan regions relate to the vectors. The combined scans using both Algorithms A and B are shown in Figure 13(a) and (b). Algorithm B appears to give better results than Algorithm A in the overlapping parts where the averaging is used in Algorithm A.

Figure 12 shows the results for the small bottle top which has been sprayed. Again only the regions obtained for this with Algorithm B are shown, because those for A are confusing. Figure 1 shows the results for a single orientation scan with distortions present and Figure 14 shows the combined scan results. In Figure 14(a) the detail shows that with Algorithm A some downward spikes occur. However with Algorithm B, as seen in Figure 14(b) and (c), there are no significant distortions.

DISCUSSION

From the results shown it appears that Algorithm B is better than Algorithm A, regardless of whether Method 2 or 3 is used. However, we have found that Algorithm B takes considerably longer for the software to run. Therefore a compromise method to find the regions may be the best approach. We propose to use Algorithm A initially but then use the output from it to identify all overlapping regions. For these regions Algorithm B can be used to

Figure 14. Scan results for the sprayed bottle top from Figures 1 and 12. Note: (a) using Algorithm A, showing a detail viewed from below with a downward spike still present; (b) and (c) using Algorithm B, with views from above and a detail from below, showing no downward spikes. (The scale of the vertical axis is exaggerated.) To view a colour version of this figure, please see http://www.igi-global.com/downloads/pdf/laha/20color.pdf



resolve the conflict between several orientations, thus reducing the extra time required.

Another approach to reducing the time taken is to use Method 1 to identify the 'best' orientation and start performing a complete scan with this orientation. While this scan is taking place, the rest of the software can be run in parallel to plan the paths for the other orientations for Method 2.

There are a number of situations where it is not possible to resolve the problem of errors, because scans of all orientations will be distorted. Such situations occur when there is a small hole or an internal corner, as shown in orange in Figure 8(e) and no orientation can be relied upon to give undistorted results. For a narrow hole it may be impossible to detect light returning from the inside with any scanner orientation, which is a problem inherent in the nature of the sensor. Then all that can be done is to highlight such a problem region to the user. It may be possible to use extrapolation to try to predict the distortion near an internal corner, based on the way distortions occur further away from the corner. However, this relies on some assumption of homogeneity of the object. If such results are likely to be unreliable, those regions should also be highlighted to the user.

Other approaches could be used to reduce the total scan time. For example, it may be possible to estimate the heights of edges prior to scanning by using two images taken with the camera at different heights. The change in position will depend on the height of the edge and allow scan region width to be chosen appropriately. In addition the direction of movement will indicate which side is the lower, and therefore where the distortions are expected at some orientations. The scan region width for the other (higher) side can then be made much smaller.

Image analysis prior to scanning may also allow us to vary the scan resolution based on the complexity of the image. In areas of 'high complexity' we can increase the sample rate to achieve a better resolution where a higher level of detail is required.

CONCLUSION

At the start of this chapter, we specified that the replicating an existing solid object via CAD/CAM requires a cost effective method to generate data for input into the CAM system from that solid object. We have shown here that the integration of an optical camera into a single-perspective laser scanner system provides a major step towards such a method. The availability of reasonably inexpensive CCD cameras of reasonable pixel resolution makes this a cost effective solution to the distortion problems faced by such a scanner. Intelligent algorithms based on the fusion of the two forms of sensor output can give high quality results at a reduced cost when compared to more expensive laser scanners. Image analysis techniques allow the problem regions of the object to be identified, so that the path and orientation of the laser sensor can be planned before the scan, thereby reducing distortions.

Care needs to be taken with both the setting up and calibration of the camera relative to the scanner and of the scanner rotation system, in order to ensure that other errors are not introduced. However, small misalignments can be detected and compensated in the software system. Work is in progress in our laboratory towards integrating such a system into operational equipment.

ACKNOWLEDGMENT

The support of the EPSRC and Axiomatic Technology Ltd. in funding and facilities for this project is gratefully acknowledged.

FUTURE RESEARCH DIRECTIONS

This chapter has outlined work which represents the first step in implementing the use of our inexpensive and cost effective approach to scanning 3D objects, so that they can be replicated by conventional CAD/CAM techniques. There remains a significant body of research and development which will be needed before our approach can be implemented as a routine technique in mechanical engineering workshops and similar industrial environments. The first stage involves testing our approach on a wide range of objects, including those which contain holes and concave features. There is no reason to suppose that this method will not work with such features, except where occlusion of the beam by another part of the object is impossible to avoid from any orientation. However, it is important to validate the technique with real examples. The second stage involves designing a robust scanner for carrying out these measurements, as opposed to the modified laboratory instrument that was used in our studies described above. Given the increasing move towards CAD/CAM technology and the decline in the number of engineers trained in traditional machining techniques, we believe that our approach will become increasingly viable in years to come.

REFERENCES

Ballard, D.H. (1981). Generalizing the hough transform to detect arbitrary shapes. *Pattern Recognition, 13*(2) 111-122.

Canny, J.F.A. (1986). Computational approach to edge detection. *IEEE Transactions on Pattern Analysis & Machine Intelligence, 8*(6), 679-698.

Chen, P.C., & Pavlidis, T. (1981). Image segmentation as an estimation problem. In A. Rosenfeld (Ed.), *Image modelling* (pp. 9-28). Academic Press.

Chow, Y.K., Poliakoff, J.F., & Thomas, P. D. (2002). Interpolation and acceleration algorithms for Stepper Motors: A parametric approach. In *Proceedings of the 8th IEEE International Conference on Methods, Models and Robotics*. ISBN 83-88764-51-9.

Denby, A., Langensiepen, C., Poliakoff, J., & Sherkat, N. (2005). Use of edge models to guide optimal 3D scanning,. In *Proceedings of the 8th International Conference on Computer Modelling and Simulation*. ISBN: 1-84233-111-6.

Deriche, R. (1987). Using Canny's criteria to derive a recursively implemented optimal edge detector. *The International Journal of Computer Vision*, *1*(2), 167-187.

Foley, J.D., Van Dam, A., Feiner, S.K., Hughes, J.F., & Phillips, R.L. (1994). *Introduction to computer graphics* (pp.168-177). MA: Addison-Wesley.

Freeman, H. (1970). Boundary encoding and processing. *Picture processing and psychopictories.* New York: Academic Press.

Haralick, R., Shanmugam, K., & Dinstein, I. (1973). Texture features for image classification. *IEEE Transactions on Systems, Man, and Cybernetics, SMC, 3*(6), 610-621.

Heath, M., Sarkar, S., Sanocki, T., & Bowyer, K. (1998). Comparison of edge detectors: A methodology and initial study. *Computer Vision and Image Understanding, 69*(1), 38-54.

Hori, O., & Tanigawa, S. (1993). Raster-to vector conversion by line fitting based on contours and skeletons. In *Proceedings of ICDAR'93,* Tsukuba, Japan, (pp. 353-358).

Ji, Q., & Xie, Y. (2003). Randomised Hough transform with error propagation for line and circle detection. *Pattern Analysis Applications, 6*, 55-64.

Kitchen, L., & Rosenfeld, A. (1982). Gray-level corner detection. *Pattern Recognition Letters, 1,* 95-102.

Low, A. (1991). *Introductory computer vision and image processing* (p. 69). McGraw-Hill.

Lyvers, E.P., & Mitchell, O.R. (1988). Precision edge contrast and orientation estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 10*(6), 927-937.

Novak, C.L., & Shafer, S.A. (1987). Color edge detection. In *Proceedings of the DARPA Image Understanding Workshop,* Los Angeles, California (Vol. 1, pp. 35-37).

Poliakoff, J.F., Chow, Y.K., Orton, P.A., Howson, M., & Al-Dabass, D. (2005). Evaluation by simulation of interpolation and acceleration algorithms for Stepper Motors. *International Journal of Simulation Systems, Science & Technology, 6*(7-8), 18-29. ISSN 1473-804x [online].

Poliakoff, J.F., & Thomas, P.D. (1998). Error correction in scanned engineering drawings using 3-D knowledge-based reconstruction. In R. Kasturi & A.K. Chhabra (Eds.), *Graphics recognition* (LNCS, Vol. 1389, pp. 280-290). Springer-Verlag.

Prewitt, J.M.S. (1970). Object enhancement and extraction. In B. S. Lipkin & A. Rosenfeld (Eds.), *Picture processing and psychopictories.* New York: Academic Press.

Risse, T. (1989). Hough transform for line recognition: Complexity of evidence accumulation and cluster detection. *CVGIP, 46*(3) 327-345.

Sacchi, R., Poliakoff, J.F., Thomas, P.D., & Häfele, K.-H. (2004). Segmentation of scanned surfaces: Improved extraction of planes. In M. Sarfraz (Ed.), *Geometric modeling: Techniques, applications, systems and tools* (pp. 251-262). Wiley.

Sezgin, M., & Sankur, B. (2004). Survey over image thresholding and quantitative performance evaluation. *Journal of Electronic Imaging, 13*(1), 146-165.

Smith, S.M., & Brady, J.M. (1997). SUSAN: A new approach to low level image processing. *International Journal of Computer Vision, 23*(1), 45-78.

Spann, M., & Wilson, R.G. (1985). A quad-tree approach to image segmentation which combines statistical and spatial information. *Pattern Recognition, 18*(3/4), 257-269.

Tombre, K. (1998). Analysis of engineering drawing: State of the art and challenges. In K. Tombre & A.K. Chhabra (Eds.), *Graphics recognition: Algorithms and systems* (pp. 257-264). Berlin: Springer-Verlag.

Tombre, K., & Tabbone, S. (2000). Vectorization in graphics recognition: To thin or not to thin. In *Proceedings of the ICPR'00* (Vol., pp. 291-96).

Wang, S.J., & Binford, T.O. (1994). *Generic, model-based estimation and detection of discontinuities in image surfaces* (Vol. 2, pp. 113-116). ARPA IUW.

Wong, K.H. (2002) *Compensation for distortion in the imaging process for 3D surfaces.* Unpublished doctoral thesis, Nottingham Trent University, Nottingham, UK.

Wong, K.H., Poliakoff, J.F., & Thomas, P.D. (2001). Compensation techniques for distortions from a single-perspective optical triangulation sensor. In *Proceedings of the Fifth Conference on Optical 3-D Measurement Techniques.*

Ziou, D., & Tabbone, S. (1998). Edge detection techniques: An overview. *Pattern Recognition and Image Analysis, 8*(4), 537-559.

ADDITIONAL READING

Beraldin, J.A., Blais, F., Cournoyer, L., Rioux, M., El-Hakim, S.H., Rodella, R., et al. (1999). 3D digital imaging and modeling on remote sites. In *Proceedings of the Second International Conference on 3D Digital Imaging and Modelling,* Ottawa, Canada.

Blais, F. (2004). Review of 20 years of range sensor development. *Journal of Electronic Imaging, 13*(1), 231-240.

Chai, I., & Dori, D. (1992). Orthogonal zig-zag: An efficient method for extracting lines from engineering drawings. In C. Arcelli, L.P. Cordella, & G. Saruiti di Baja (Eds.), *Visual form* (pp. 127-136). New York: Plenum Press.

Chen, F., Brown, G.M., & Song, M. (2000). Overview of three-dimensional shape measurement using optical methods. *Optical Engineering, 39*, 10-22.

Curless, B. (1997). *New methods for surface reconstruction from range images*. Unpublished doctoral thesis, Stanford University.

De Bakker, M., Verbeek, P.W., & Steenvoorden, G. K. (2000). Smart PSD array for sheet-of-light range imaging. *Proceedings of SPIE, 3965*, 21-32.

Dorsch, R., Herrmann, J., & Häusler, G. (1994). Laser triangulation: Fundamental uncertainty of distance measurement. *Applied Optics, 33*(7), 1306-1314.

Faugeras, F. (1993). *Three-dimensional computer vision*. MIT Press.

Faugeras, O.D., Luong, Q.-T., & Maybank, S.J. (1992). Camera self-calibration: Theory and experiments. In *Proceedings of the European Conference on Computer Vision* (pp. 321-234).

Häusler, G. (1999). Three-dimensional sensors: Potentials and limitations. *Handbook of Computer Vision and Applications, 1*(19), 485-506.

Ingle, K.A. (1994). *Reverse engineering*. McGraw-Hill Professional Publishing.

Jansa, J., Huang, Y. R., & Trinder, J. C. (1993). Problems of precise target location and camera orientation in digital close-range photogrammetry. *Proceedings of SPIE, Videometrics II, 2067*, 151-161.

Kanaugo, T., Jaisimha, M., Palmer J., & Haralick, R. (1995). A methodology for quantitative performance evaluation of edge detection algorithms. *IEEE Transactions on Image Processing, 4*(12), 1667-1673.

Nalwa, V.S. (1993). *A guided tour of computer vision*. Addison-Wesley.

Reichmann, W. (1995). Fast object recording by means of structured light and photogrammetric techniques. *Proceedings of IAPRS, 30*, 195-200.

Ruzon, M., & Tomasi, C. (1999). Color edge detection with the compass operator. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Ft. Collins, Colorado (Vol. 2, pp. 160-166).

Shu, C., & Xi, F. (1999). Model-based scanning path generation for inspection. In *Proceedings of the Second International Conference on 3-D Digital Imaging and Modeling (3DIM '99)*, Ottawa, Canada (pp. 118-124).

Thomas, P.D., Poliakoff, J.F., Razzaq, S.M., & Whitrow, R.J. (1996). A combined high and low level approach to interpreting scanned engineering drawings. In R. Kasturi & K. Tombre (Eds.), *Graphics recognition: Methods and applications* (LNCS, Vol. 1072). Springer-Verlag.

Vàrady, T., Martin, R.R., & Cox, J. (1997). Reverse engineering of geometric models: An introduction. *Computer Aided Design, 29*(4), 255-268.

Xu, Y., Tang, J., & Zhong, W. (2000). New method of processing the signals of a position sensitive detector. *Proceedings of SPIE, Advanced Photonic Sensors: Technology and Applications, 4220*, 260-263.

Zheng, Z., Wang, H., & Teoh, E. (1999). Analysis of gray level corner detection. *Pattern Recognition Letters, 20*, 149-162.

Chapter XXI
Using Data Mining for Forecasting Data Management Needs

Qingyu Zhang
Arkansas State University, USA

Richard S. Segall
Arkansas State University, USA

ABSTRACT

This chapter illustrates the use of data mining as a computational intelligence methodology for forecasting data management needs. Specifically, this chapter discusses the use of data mining with multidimensional databases for determining data management needs for the selected biotechnology data of forest cover data (63,377 rows and 54 attributes) and human lung cancer data set (12,600 rows of transcript sequences and 156 columns of gene types). The data mining is performed using four selected software of SAS® Enterprise Miner™, Megaputer PolyAnalyst® 5.0, NeuralWare Predict®, and BioDiscovery GeneSight®. The analysis and results will be used to enhance the intelligence capabilities of biotechnology research by improving data visualization and forecasting for organizations. The tools and techniques discussed here can be representative of those applicable in a typical manufacturing and production environment. Screen shots of each of the four selected software are presented, as are conclusions and future directions.

INTRODUCTION

Mining biological, medical, or behavioral data is an emerging area for research on bioinformatics (Cohen & Hersh, 2005; Wang & Yang, 2005). This chapter illustrates the use of data mining as a computational intelligence methodology for forecasting data management needs. Specifically, this chapter discusses the use of data mining with multidimensional databases for determining data management needs for the selected biotechnology data of forest cover data and human lung cancer data sets. The analysis and results will be used to enhance the intelligence capabilities of biotechnology research by improving data visualization and forecasting for organizations. The tools and techniques discussed here can be representative of those applicable in a typical manufacturing and production environment. The chapter also helps organizations to choose proper data mining software for their forecasting data management needs.

The data mining is performed using four selected software of SAS® Enterprise Miner™, Megaputer PolyAnalyst® 5.0, NeuralWare Predict®, and BioDiscovery GeneSight®. One of the databases is that of forest cover type's data that is a very large database composed of 63,377 rows and 54 attributes. The other database is that composed of human lung carcinomas cancer data and is a smaller database with data elements at the human gene level that comprise a microarray database consisting of 12,600 rows of transcript sequences and 156 columns of gene types. Background on related literature and software are also presented. Screen shots of each of the four selected software are presented, as are conclusions and future directions.

BACKGROUND

The study of forecasting started in the 1960s with two categories of linear (e.g., regression) and nonlinear forecasting techniques (e.g., artificial neural network and self-organizing map). Most data mining techniques combine both linear and nonlinear models (He & Xu, 2005) and use different data analysis tools to discover relationships and knowledge in data that may be used to make valid classification and predictions (Chen, Diao, Dulong, et al., 2005; Nielson, 2005).

Data mining has been used in many fields for effective discovery and prediction of new knowledge. Neaga and Harding (2005) present an enterprise integration and management framework based on data mining. Alverez-Macias and Mata-Vazquez (2004) use data mining for the management of software development process. Wu, Chen, and Chian (2006) implement data mining techniques in a product quality control system. Padmanabhan, Zheng, and Kimbrough (2006) use data mining techniques for analyzing user data tracked online for e-business. Rubinov, Soukhorokova, and Ugon (2006) discuss the clusters and classification in data analysis. To improve the prediction accuracy, Li and Ye (2006) present a supervised clustering and classification algorithm for mining data with both numeric and nominal variables in medical diagnosis.

As discussed by Segall (2006) in a chapter in *Encyclopedia of Data Warehousing and Mining*:

Microarray informatics is a rapidly expanding discipline in which large amounts of multi-dimensional data are compressed into small storage units. Data mining of microarrays can be performed using techniques such as drill-down analysis rather than classical data analysis on a record-by-record basis. Both data and metadata can be captured in microarray experiments.

An important issue of this chapter is what benefits your organization can drive from a properly implemented storage management policy and specifically for databases of varying dimensionalities such as at the microarray level.

Segall (2006) further discusses the following background on microarray databases by Schena (2003) and National Center for Biotechnology Information (NCBI):

A Microarray has been defined by Schena (2003) as 'an ordered array of microscopic elements in a planar substrate that allows the specific binding of genes or gene products.' Schena (2003) claims microarray databases as "a widely recognized *next revolution in molecular biology that enables scientists to analyze genes, proteins, and other biological molecules on a genomic scale.*

According to an article (2004) on the *National Center for Biotechnology Information* (NCBI) Web site, "because microarrays can be used to examine the expression of hundreds or thousands of genes at once, it promises to revolutionize the way scientists examine gene expression," and "this technology is still considered to be in its infancy."

Probabilistic models for large-scale DNA-sequencing such as the Human Genome Project as well as for proteins and nucleic acids and other biological sequence analysis are discussed by Durbin, Eddy, Krogh, and Mitchson (2000). Students at Ray Soin College of Business (2005) at Wright State University discussed emerging business models in data management and data mining including those for digital fixed-content storage, and implantable radio-frequency identification (RFID) about the size of a grain of rice for healthcare.

Edelstein (2006) discusses building profitable customer relationships with data mining, indicating that

the sheer volume of customer information and increasingly complex interactions with customers have propelled data mining to the forefront of making your customer relationships profitable…because of a greatly improved ability to respond to each individual contact in the best way, and reduced costs due to properly allocating your resources.

Both SAS (2004) and Delmater and Hancock (2001) explain how data mining unites customer relationship management (CRM) and business intelligence and paves the way to e-commerce. Williams (2006) discusses performance capacity management for the world's online marketplace with application to the development of eBay's data warehouse. Report mining was used as an easier way to access corporate information and effective cash management by white papers posted on the Datawatch (2005a, 2005b) Web pages. The Geneva Research Collaboration (2002) discusses the use of data mining for high performance computing and complex system modeling for very large data sets.

United States General Accounting Office (GAO, 2004) wrote a report on the wide range of uses of data mining for federal efforts that include managing human resources and analyzing intelligence and detecting terrorist activities for each of the federal agencies. According to this GAO (2004) report, the U.S. Air Force Oracle Human Resources software uses data mining to provide information on promotions and pay grades.

MAIN THRUST OF THE CHAPTER: ISSUES, CONTROVERSIES, PROBLEMS

The issues of this chapter are the applications of data mining for forecasting data management needs. Four software have been utilized for illustrating this with their selection being determined by their grants of software and technical support to perform this research. This university is currently a partner with the SAS® Academic Alliance and utilizes SAS® Enterprise Miner in teaching of its data mining courses.

Other issues in this research are the selection of the data and specific data tools for each of the software to be used in this research are as described below.

The controversies are whether or not data mining can be used to accurately forecast data management needs; and if so, the problems are to determine which of these data mining software and tools are better than the others for the selected databases of varying dimensionalities including that at the microarray level.

Issues: Data Used in Research

Each of the four selected software is applied to two databases of differing dimensionalities. One is a very large database of forest cover type that is available on the same Web site of the Machine Learning Repository at the University of California at Irvine by Newman, Hettich, Blake, and Merz (1998) for which results are shown in Segall and Zhang (2006a, 2006b) for different datasets of numerical abalone fish data and discrete nominal-valued mushroom data. The forest cover type dataset is owned by the Remote Sensing and Geographical Information Systems (GIS) Program of the Department of Forest Sciences at Colorado State University in Fort Collins, Colorado.

The forest cover type's database consists of 63,377 records each with 54 attributes that can be used as inputs to predictive models to support decision-making processes of natural resource managers. The 54 columns of data are composed of 10 quantitative variables, 4 binary variables for wilderness areas, and 40 binary variables of soil types. The forest cover type's classes include Spruce-Fir, Lodgepole Pine, Ponderosa Pine, Cottonwood/Willow, Aspen, Douglas-Fir, Krummholz, and others.

The second database used in this chapter is that of the Broad Institute (2006). The Broad Institute is a research collaboration of Massachusetts Institute of Technology (MIT), Harvard, and its affiliated hospitals of Beth Israel Deaconess Medical Center, Brigham and Women's Hospital, Children's Hospital Boston, Dana-Farber Cancer Institute, and Massachusetts General Hospital. According to its Web page, the mission of the Broad Institute is "to bring the power of genomics to medicine."

The data set selected from the Broad Institute is one of those posted as available with unrestricted access as one of the Web links posted on the Web page of the Broad Institute Cancer Program Data Sets (2006) and is that which is related to the "Classification of Human Lung Carcinomas by mRNA Expression Profiling" research project of the Broad Institute. According to the Web page of the Broad Institute Cancer Program Publications (2006), lung carcinoma is the leading cause of cancer death in the United States and worldwide. The selected data base for this research was used by Bhattacharjee et al. (2001) using oligonucleotide microarrays to analyze "mRNA expression levels corresponding to 12,600 transcript sequences in 186 lung tumor samples, including 139 adenocarcinomas resected from the lung."

Issues: Data Mining Tools Used

The proposed research is to entail data mining of the two datasets described above using the four selected software indicated above as SAS® Enterprise Miner™, Megaputer PolyAnalyst® 5.0, NeuralWare Predict®, and BioDiscovery GeneSight®. The data mining algorithms to be performed include those for neural networks, genetic algorithms, clustering, and decision trees. Much of this data mining has already been performed by the authors for these two data sets and for the selected software.

The computer output included in the manuscript for the data mining includes plots as generated by each of these software for the two data sets of differing dimensionalities. These plots include histograms of the data, hierarchical clustering, box plots, time series plots, principal component analysis (PCA) plots, k means clustering, self-organizing-maps (SOM) clustering, and two-dimensional SOM plots. Each of these would be compared for the corresponding plots for each of the two data bases of differing dimensionalities (e.g., forest cover type data set of 63,377 records with 54 attributes vs. the human lung cancer microarray database of 12, 600 rows and 156 columns of gene types).

Problems and Results: Data Mining for Forecasting Data Management Needs

Result 1a: Data Mining Using BioDiscovery GeneSight® for Forest Cover Type Data

Figure 1 shows the window of GeneSight® with the forest cover type dataset with respective variables of elevation, aspect, slope, horizontal distance to hydrology, vertical distance to hydrology, horizontal distance to roadways, hillshade 9am, hillshade noon, hillshade 3pm, horizontal distance to fire points, wilderness area, soil types (40 types), and cover types (7 types). The boxplots of Figure 2 for these seven variables have greatest variability evident for the variable of horizontal distance to roadways and the variable of cover types. The hierarchical clustering of global variations using the Euclidean distance metric is shown in Figure 3. Figure 4 shows the time series plot of the forest cover type data.

Figure 1. Window of GeneSight® with the forest cover type dataset

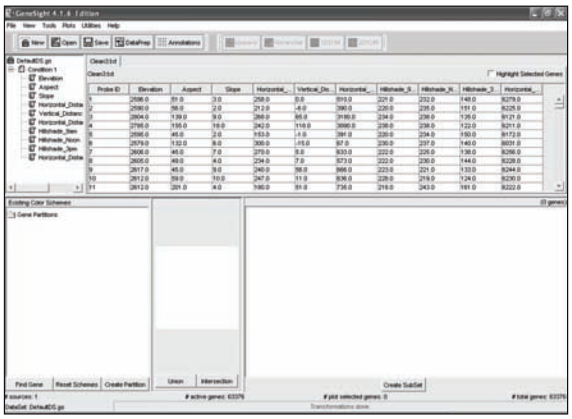


Figure 2. Boxplots of seven cover types using GeneSight®

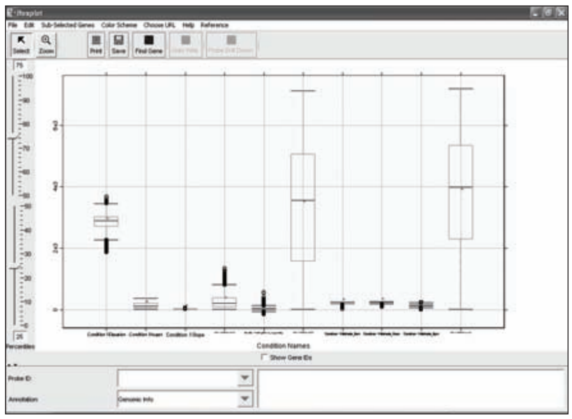


Figure 3. Hierarchical clustering of global variations with the Euclidean distance metric using GeneSight®

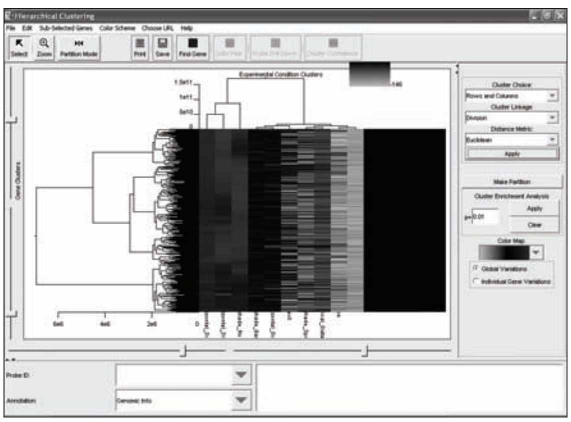
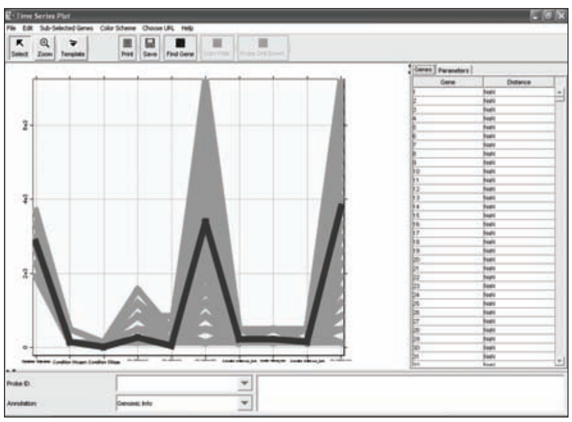


Figure 4. Time series plot of the forest cover type data using GeneSight®



The scatter plot of actual data given by Figure 5 illustrates a greater density on the right side of this figure. Figure 6 shows the two-dimensional self-organizing map (SOM) for eleven variables for all of the data using the squared Euclidean metric. These variabilities are respectively differing from those based on the average and standard deviation option of GeneSight® for the eleven variables of the forest cover type using the squared Euclidean metric, and also those using the Chebychev distance metric.

Result 1b: Data Mining Using BioDiscovery GeneSight® for Human Lung Microarray Data

The window of GeneSight® with the human lung carcinomas cancer data for each variable of the 156 gene types is illustrated in Figure 7. Figure 8 shows boxplots for the first thirteen (13) gene types that appear not to be significantly different from one another. Figure 9 shows the individual gene variations of hierarchical clustering for the

Figure 5. Scatter data plot of actual forest cover type data using GeneSight®

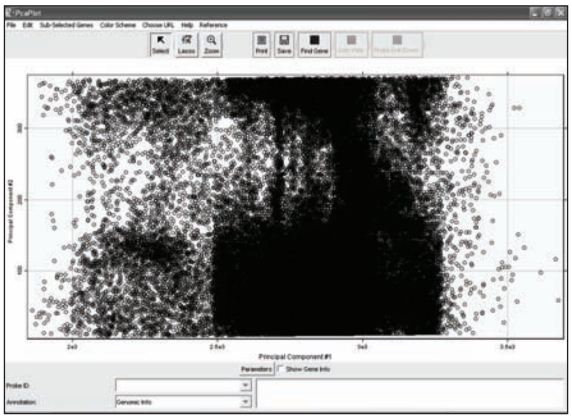


Figure 6. Self-organizing map (SOM) with the squared Euclidean metric using GeneSight®

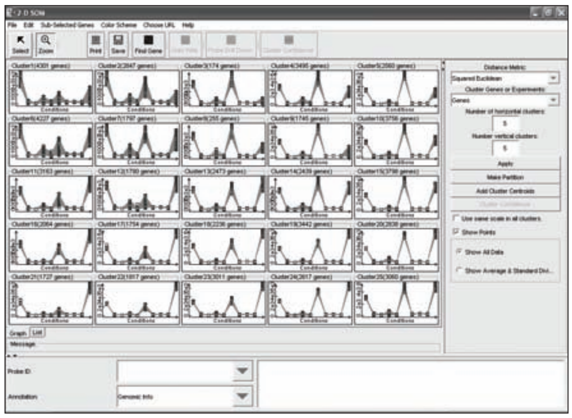


Figure 7. Window of GeneSight® with the human lung carcinomas cancer data

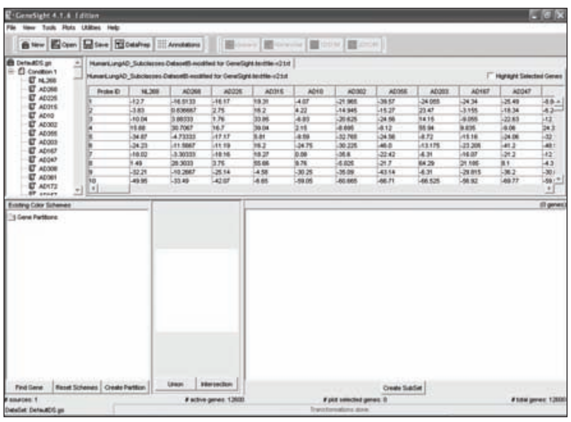
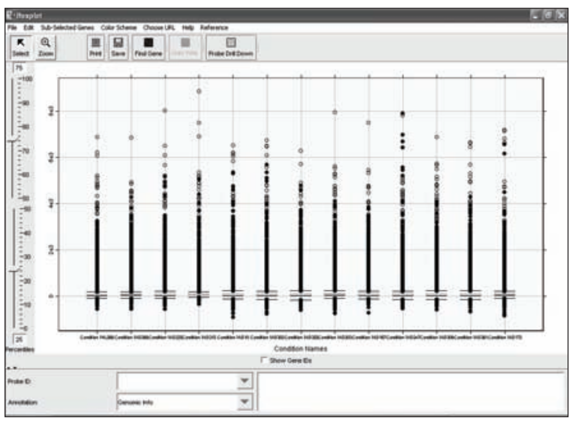


Figure 8. Boxplots for the first thirteen (13) gene types using GeneSight®



human lung cancer mircoarray data using the Euclidean metric. This figure shows more distinct gene clusters as evident by the greater variation in colorings of the clusters than that for the global variation using the same distance metric. The Chebychev metric was also run and has more distinct global variations of k-means clustering than those for the Euclidean metric.

Global variations of the self-organizing map (SOM) clustering using the Euclidean distance metric for the first thirteen (13) gene types are illustrated by Figure 10. These same plots using the Chebychev metric for the human lung caner microarray databases have also been generated but are not shown due to space limitations. The time series plots for the human lung carcinomas cancer data are given by Figure 11. Figure 12 shows the principle-component analysis (PCA) distribution plots for the first thirteen gene types and illustrates the scatter plots of the PCA as very distinctly clustered to the left of this figure.

Figure 9. Individual gene variations of hierarchical clustering for the human lung cancer mircoarray data with the Euclidean metric using GeneSight®

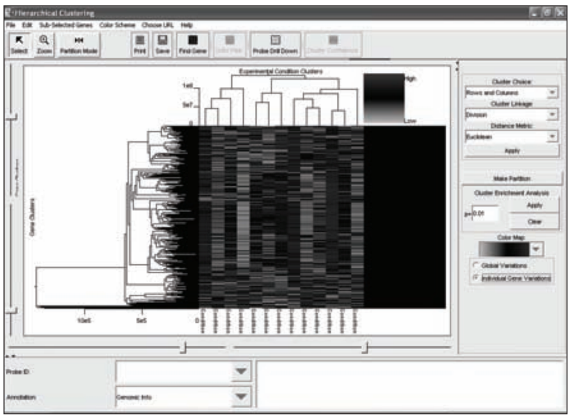


Figure 10. Global variations of the self-organizing map (SOM) clustering with the Euclidean distance metric using GeneSight®

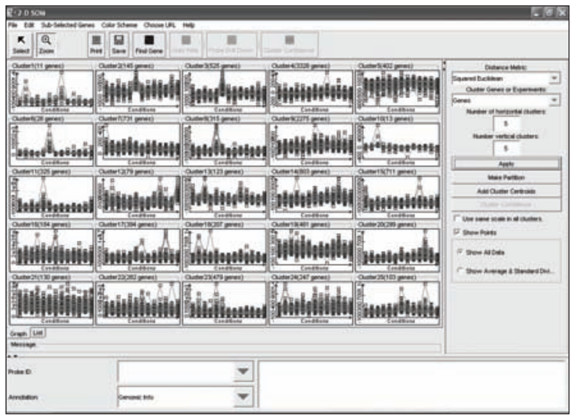


Figure 11. Time series plots for the human lung carcinomas cancer data using GeneSight®

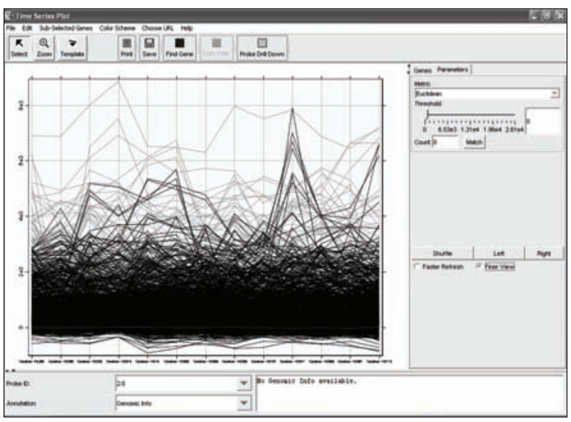
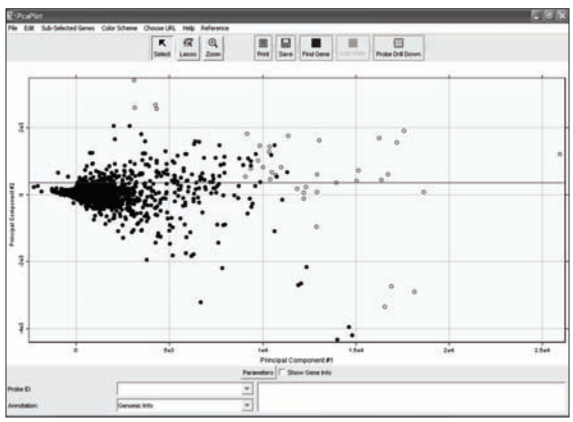


Figure 12. Scatter plots of the PCA for human lung data using GeneSight®



Result 2a: Data Mining Using Megaputer PolyAnalyst® 5.0 for Forest Cover Type Data

A histogram of the forest cover type data as generated by Megaputer PolyAnalyst® 5.0 indicates the presence of substantially more of the Lodgepole Pine cover type. The decision tree report indicates a classification probability of 80.19% with a total classification error of 19.81%. Per PolyAnalyst® output, the decision tree has a tree depth of 100 with 210 leaves, a depth of constructed tree of 16, and a classification efficiency of 47.52%. Figure 13 illustrates the complexities and compositions of the cluster analysis for the various cover types.

Result 2b: Data Mining Using Megaputer PolyAnalyst® 5.0 for Human Lung Microarray Data

Megaputer PolyAnalyst® 5.0 was applied for the first five gene types of the Human Lung microarray data yielding graphs, rules mining models, reports, and dimension matrices. A histogram of the frequency counts for the NL268 gene type has been compiled to illustrate an almost equal frequency count for each of the five bins.

Figure 14 illustrates the link chart for the five selected gene types for the five bins of gene AD268 as very symmetrical. The bin selection rule for gene type NL268 is shown by Figure 15. The decision tree yielded a classification probability of 69.49% and classification efficiency of 61.74%. The decision tree per PolyAnalyst® output yielded a total classification error of 30.51% for a tree depth of 1,000 with 38 leaves, depth of constructed tree of 8, and a classification efficiency of 61.74%.

Result 3a: Data Mining Using SAS® Enterprise MinerTM for Forest Cover Type Data

The workspace of SAS Enterprise Miner™ that was used in the data mining of the forest cover dataset is shown in Figure 16. Figure 17 shows the results of using data mining tool of decision trees with SAS Enterprise Miner™. Note that Figure 25 combines four views of results of the data mining using decision trees in a single window. As can be seen in Figure 25, decision tree analysis was completed within 13 leaves of the decision tree using 40% of the sample for the training, 30% validation, and 30% testing.

Figure 13. Cluster analysis of the cover types using PolyAnalyst® 5.0

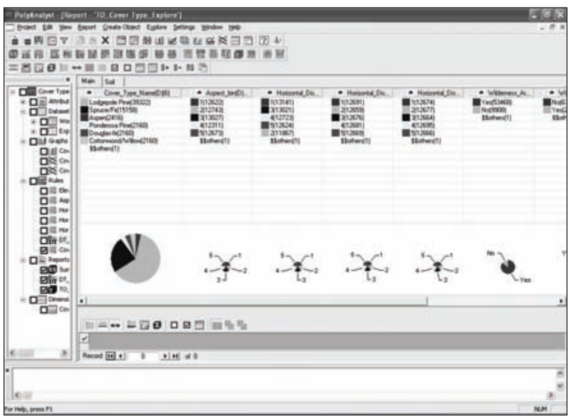


Figure 14. Link chart for NL268 and AD268 gene types of human lung dataset using PolyAnalyst® 5

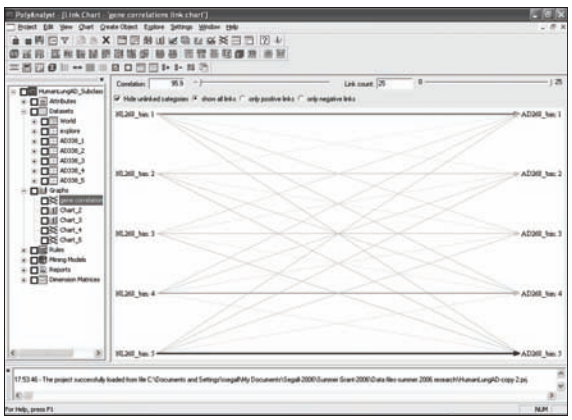


Figure 15. Decision Tree Report for NL268 gene type data bin of human lung Project showing open data tag for AD115<2.38 node using Poly-Analyst® 5

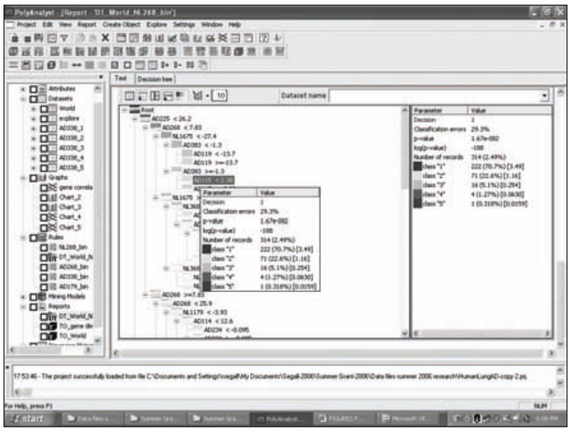


Figure 16. Workspace of SAS Enterprise Miner™ for the forest cover dataset

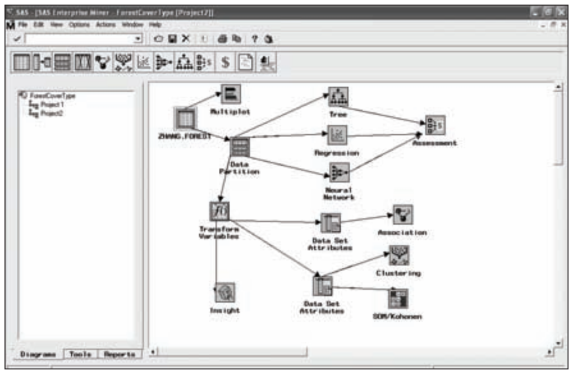


Figure 17. Decision trees output of the forest cover data using SAS Enterprise Miner™

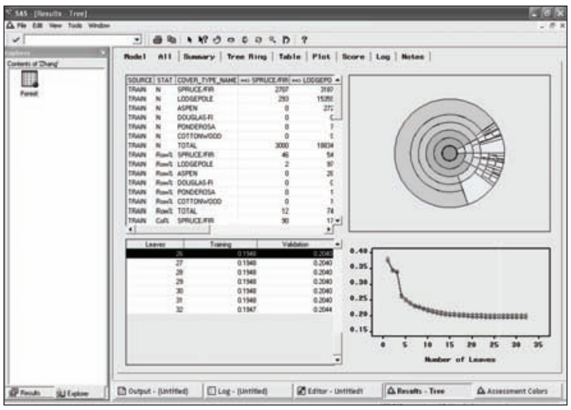
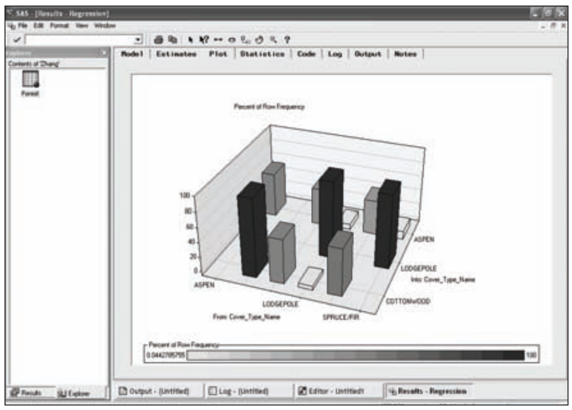


Figure 18. 3-D plot for regression modeling of the forest cover data using SAS Enterprise Miner™



A 3-D plot for regression modeling of the forest cover data for training is shown in Figure 18, which has 15 model degrees of freedom and an R-squared value of 92.8%. The SAS process monitor for the forest cover data has been used to indicate an almost parallel objective process with that of the validation objective process values.

Figure 19 shows a 2x3 self-organized map (SOM) that provides results in the form of an interactive map that illustrates the characteristics of the clusters and importance of each variable. Figure 19 shows the normalized means for the clusters of the variables and cluster proximities respectively. The same clustering criteria have been used for each of the six clusters with the top three greatest frequencies of 8,544 for Cluster 6, 7,487 for Cluster 2, and 5,170 for Cluster 3. The cover types of Spruce-Fir and Lodgepole are associated with a support of 34.92%, a confidence of 73.68%, and a lift of 1.40.

Figure 19. Self-Organized Map (SOM) normalized means of clusters for the forest cover data using SAS Enterprise Miner™

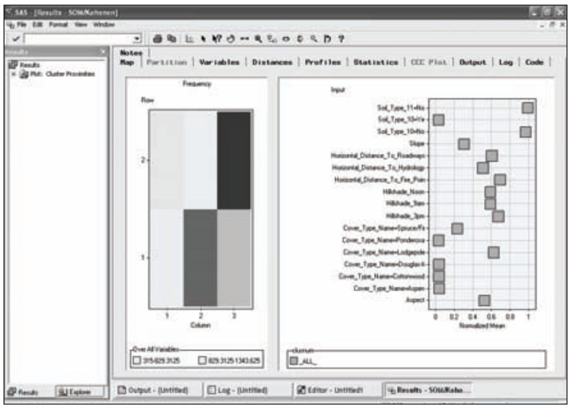
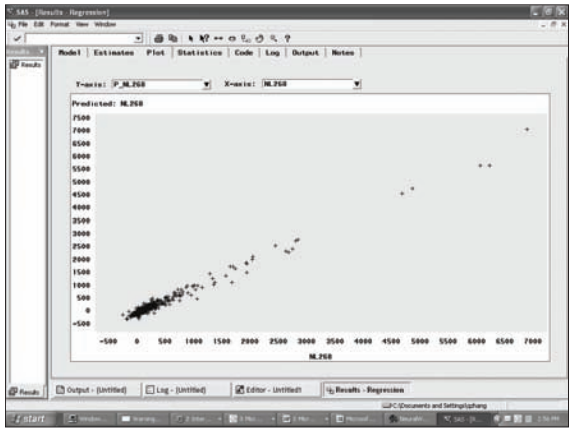


Figure 20. Regression scatter plot of predicted versus actual NL268 gene type for human lung data using SAS Enterprise Miner™



Result 3b: Data Mining Using SAS® Enterprise MinerTM for Human Lung Microarray Data

The results of using data mining tool of decision trees with SAS Enterprise Miner™ for training and validation were completed within 13 leaves of the decision tree using a sample of 5,040 for the training and sample of 3,780 for the training.

Figure 20 shows a regression scatter plot of predicted vs. actual gene type of NL268 for the human lung data. As visualized in Figure 20, most of the data values are less than 1,500 and a linear fit is evident. A summary of the fit statistics for regression modeling of the human lung data for training had 106 model degrees of freedom, and 4,934 degrees of freedom for error. Analysis of variance was performed by SAS Enterprise Miner™ for the human lung data. The data mining has provided an extremely good modeling to the human lung data as indicated by the r-squared value of 97.81% and adjusted r-squared value of 98.13%.

A summary table of the fit statistics with target variable of NL268 effect for human lung data using neural networks for data mining was obtained with SAS Enterprise Miner™. Figure 21 provides a view of the SAS process monitor the human lung data that indicates an almost parallel objective process with that of the validation objective process values. Figure 21 also shows the variations in average error for the neural network training and validation for the human lung data. As evidenced by Figure 21, the average error decreases dramatically within the first 20 iterations and then tails off to uniform rate.

The results of using cluster analysis for the data mining of the human lung data are shown in Figure 22. As visualized by Figure 22, the normalized means for the gene types listed are almost uniform with the exception of that for AD122 and AD043 gene types. It was found that several of the clusters have the same cluster proximities which are to be expected from the almost uniform normalized means, as shown in Figure 22 and Figure 23, of the frequency distributions of the clusters.

Self-organized maps (SOM) were also obtained that illustrate the characteristics of the clusters and importance of each variable. These

SOM showed that the normalized means for the cluster proximities of the gene types variables are now scattered and not as uniform as that in Figure 16 using standard cluster analysis techniques. The numerical importance of the gene types with the most important gene types ranged from about .53 for gene type AD327 to .09 for gene type AD114. The same clustering criteria has been also used for each of the nine (9) clusters with the greatest frequency of 3,364 for Cluster 4, which is significantly different from those of the other eight clusters.

Figure 21. Neural network training and validation variations in average error for human lung data using SAS Enterprise Miner™

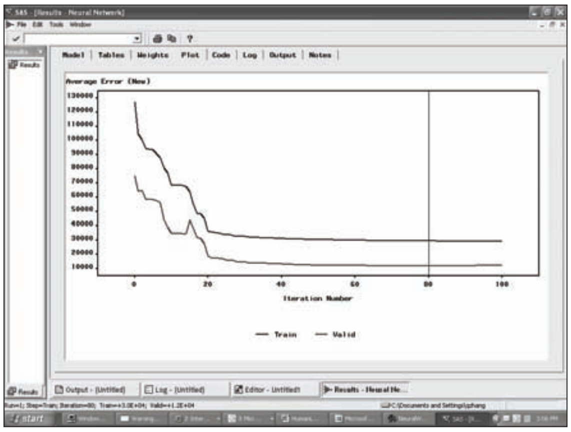


Figure 22. Cluster analysis for transformed training human lung data using SAS Enterprise Miner™

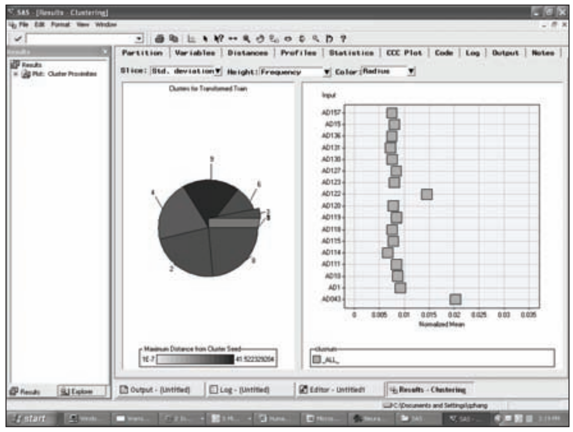


Figure 23. Frequency distributions for clusters for human lung data using SAS Enterprise Miner™

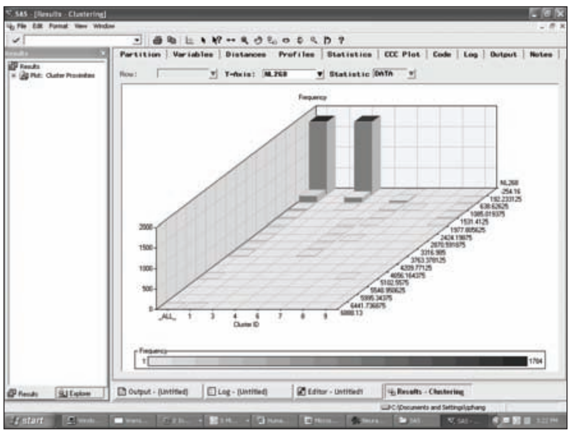
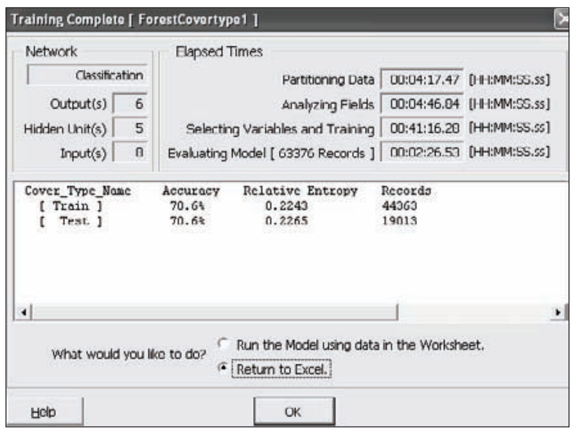


Figure 24. Prediction using NeuralWare Predict® with the forest cover type name as dependent variable



Result 4a: Data Mining Using NeuralWare Predict® for Forest Cover Type Data

The prediction using NeuralWare Predict® with the forest cover type name as dependent variable is shown in Figure 24. The result indicates an overall accuracy of 70.6%. Figure 25 show specifically the prediction accuracy in each category of Spruce-Fir, Lodgepole Pine, Ponderosa Pine, Cottonwood/Willow, Aspen, Douglas-Fir, and Krummholz.

Result 4b: Data Mining Using NeuralWare Predict® for Human Lung Microarray Data

The prediction using NeuralWare Predict® with NL 268 as dependent variable is illustrated in Figure 26. The result indicates an overall accuracy of 85%. Figure 27 shows the importance of predictive variables. The results indicate that the two most important variables in predicting NL 279 are NL 279 and AD 311.

Figure 25. Classification and prediction accuracy for the forest cover type data using NeuralWare Predict®

Accuracy	Rel. Entropy	Cover_Type_ Name	aspen	spruce/ fir	douglas- fir	lodgepole pine	ponderosa pine	cottonwood/ willow	Total
0.878725	0.127288	aspen	2123	52	67	115	59	0	2416
0.765932	0.118256	spruce/fir	230	11610	14	3302	2	0	15158
0.72963	0.243464	douglas-fir	57	0	1576	0	429	98	2160
0.799603	0.216509	lodgepole pine	2310	5426	86	31442	58	0	39322
0.665278	0.211929	ponderosa pine	51	0	517	0	1437	155	2160
0.922222	0.111137	cottonwood/willow	0	0	72	0	96	1992	2160
0.791782	0.186779	Total	4771	17088	2332	34859	2081	2245	63376

Figure 26. Prediction using NeuralWare Predict® for human lung data with NL 268 as dependent variable

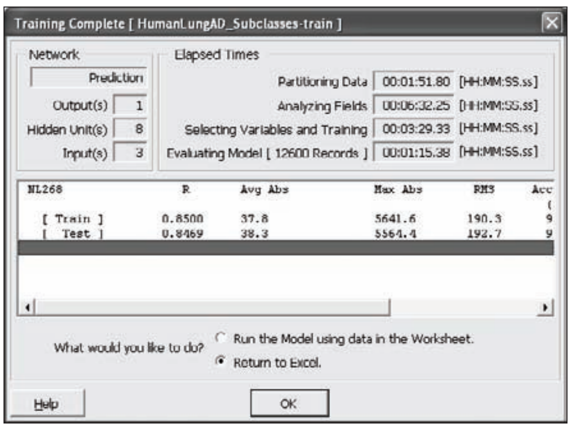
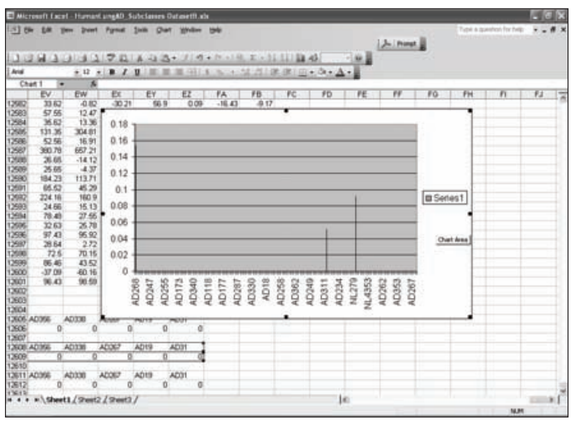


Figure 27. Importance of predictive variables for human lung data using NeuralWare Predict®



CONCLUSION

This chapter has provided results of using four well-recognized software for data mining of forest cover dataset (63,377 rows and 54 attributes) and human lung cancer datasets from approximately 12,600 patients for 186 gene types.

NeuralWare Predict® and BioDiscovery GeneSight® have less data mining functions than SAS and PolyAnalyst® do. NeuralWare® Predict is used with Microsoft Excel so that it is easy to use. It is also cheaper. It produces comparable results with other software in terms of prediction using neural networks. BioDiscovery GeneSight® is primarily for clustering analysis in bioinformatics, and is able to provide a variety of data mining visualization charts and colors.

In contrast, PolyAnalyst® and SAS Enterprise Miner have more data mining functions. SAS Enterprise Miner is powerful and effective. For example, it provides extremely good r-squared values in the analysis of variance generated from the data mining using regression analysis. PolyAnalyst® is able to provide data mining visualization using pie charts and color-coded frequencies for the report as shown in Figure 13. It provides visual link charts that the other software are unable to provide.

In addition, the data mining of the microarray datasets of human lung cancer was insightful of new information and especially in the determination of the significant gene types. This later finding is important in determining the association of specific gene types with human lung cancer using data mining techniques for microarray databases as a relatively new area of research in bioinformatics.

FUTURE RESEARCH DIRECTIONS

Future directions of the research include using other data mining tools for other databases of differing dimensionalities to contrast the findings of the research presented in this chapter. SAS Enterprise Miner version 4 was used for the analysis of the data obtained. Since the time of the writing of this manuscript, the College of Business at this university has acquired and installed SAS Enterprise version 5.2, which is substantially different than the previous version.

There are several categories in which data mining algorithms can be grouped and these include classification and prediction, regression, clustering, association rules, temporal data mining, and text mining. The future directions of this research would investigate algorithms and heuristics for data mining in each of these categories as well as possibly other categories that would evolve with this research. The heuristics would include those of genetic algorithms, simulated annealing, tabu search, and artificial neural networks (ANN). These research directions have been proposed for funding by Zhang and Segall (2006).

The future directions of this research also include applications of some of the techniques discussed within the additional readings section of this chapter. These additional techniques could include hybrid data mining techniques, periodicity detection in time series databases, analysis using self-organized maps (SOM), and regression forest techniques.

The sources of additional databases to continue additional research include those available on the Web from University of California at Irvine Machine Learning Lab, and the Broad Institute. As stated in the background section of this chapter the Broad Institute is an organization created in 2003 in collaboration with MIT, Harvard, and affiliated hospitals, and the Whitehead Institute. The Web site of the Broad Institute has other datasets in the Cancer Program Data Sets (2006) that are expected to be updated and added to for which additional data mining can be performed.

The data resources available on Web site of the Broad Institute (2007) includes those posted with Web links for the labeled categories of genome sequences, maps, and annotations, genetic variations, cancer datasets, RNAi, and chemical biology. The genome sequence includes those data sets for human and other mammals, nonmammalian vertebrates and invertebrates, plants, fungi, and bacteria. The genetic variation datasets posted on the Broad Institute (2007) Web pages are single nucleotide polymorphism (SNP) libraries of human, mouse, chimpanzee, and dog, as well as human and mouse haplotype maps. The data set labeled as RNAi on the Broad Institute (2007) Web site are complete sets of small inhibitory RNAs against human and mouse genes generated by the RNAi Consortium. All of the data resources available on the Broad Institute (2007) Web site exhibit the strong potential of continuing this research in a multitude of future directions.

ACKNOWLEDGMENT

The authors first and forth most wish to acknowledge the support provided by a 2006 Summer Faculty Research Grant as awarded to them by the College of Business without whose program and support the proposal for this research would have not have been written and the research moved forward to obtain the software and results as described here within this chapter.

The authors also want to acknowledge each of the four software manufactures for their support of this research. SAS Inc. provided a partnership with the College of Business providing SAS Enterprise Miner™ and training. Megaputer Intelligence Inc., NeuralWare, and BioDiscovery Inc. each provided generous technical support for their respective software of PolyAnalyst® 5.0, NeuralWare Predict®, and GeneSight®.

REFERENCES

Alvarez-Macias, J., Mata-Vazquez, J., & Riquelme-Santos, J. (2004). Data mining for the management of software development process. *International Journal of Software Engineering, 14*(6), 665-695.

Bhattacharjee, A., et al. (2001). Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. *Proceedings of National Academy of Sciences, 98*(24), 13790-13795. Broad Institute Cancer Program Publication. Retrieved August 15, 2007, from http://www.pnas.org/cgi/content/full/191502998v1

Broad Institute. (2006). Retrieved August 15, 2007, from http://www.broad.mit.edu/index.html

Broad Institute. (2007). *Data resources.* Retrieved August 15, 2007, from http://www.broad.mit.edu/tools/data.html

Broad Institute Cancer Program. (2006). Retrieved August 15, 2007, from http://www.broad.mit.edu/cancer

Broad Institute Cancer Program Data Sets. (2006). Retrieved August 15, 2007, from http://www.broad.mit.edu/cancer/datasets.cgi

Broad Institute Cancer Program Publications. (2006). Retrieved August 15, 2007, from http://www.broad.mit.edu/cgi-bin/cancer/publications/pub paper.cgi!mode=view&paper id=62

Chen, Y., Diao, Q., Dulong, C., et al. (2005). Performance scalability of data mining workloads in bioinformatics. *Intel Technology Journal, 9*(2), 131-142.

Cohen, A., & Hersh, W. (2005). A survey of current work in biomedical text mining. *Briefings in Bioinformatics, 6*(1), 57-71.

DataWatch. (2005a). *Effective cash management: Powered by report mining.* Retreived August 15, 2007, from http://www.datawatch.com/dataconversionsoftware/product-resources/white-papers.asp

DataWatch. (2005b). *Report mining: An easier way to access corporate information*. Retrieved August 15, 2007, from http://www.datawatch.com/dataconversionsoftware/product-resources/white-papers.asp

Delmater, R., & Hancock, M. (2001). *Data mining explained: A manager's guide to customer-centric business intelligence*. Digital Press.

Durbin, R., Eddy, S. R., Krogh, A., & Mitchson, G. (2000). *Biological sequence analysis: Probabilistic models of proteins and nucleic acids*. Cambridge, UK: Cambridge University Press.

Edelstein, H. (2006). *Building profitable customer relationships with data mining*. Retrieved August 15, 2007, from www.twocrows.com/crm-dm.pdf and www.cio.com/km/data/links.htm

Geneva Research Collaboration. (2002). *Research directions*. Retrieved August 15, 2007, from http://grc.web.cern.ch/grc/directions.html

He, C., & Xu, X. (2005). Combination of forecasts using self-organizing algorithms. *Journal of Forecast, 24*, 269-278.

Li, X., & Ye, N. (2006). A supervised clustering and classification algorithm for mining data with mixed variables. *IEEE Transactions on Systems, Man, and Cybernetics, 36*(2), 396-406.

Neaga, E., & Harding, J. (2005). An enterprise modeling and integration framework based on knowledge discovery and data mining. *International Journal of Production Research*, *43*(6), 1089-1108.

Newman, D. J., Hettich, S., Blake, C. L., & Merz, C. J. (1998). UCI Repository of machine learning databases. *Irvine, CA: University of California, Department of Information and Computer Science*. Retrieved August 15, 2007, from http://www.ics.uci.edu/~mlearn/MLRepository.html

Nielsen, B. (2005). Strategic knowledge management research: Tracing the co-evolution of strategic management and knowledge management perspectives. *CR*, *15*(1), 1-13.

Padmanabhan, B., Zheng, Z., & Kimbrough, S. (2006). An empirical analysis of the value of complete information for eCRM models. *MIS Quarterly, 30*(2), 247-267.

Raj Soin College of Business. (2005). Emerging business models in data management and data mining. *Wright State University*. Retrieved August 15, 2007, from http://www.geocities.com/innovating_competitively/data/?200620

Rubinov, A., Soukhorokova, N., & Ugon. J. (2005). Classes and clusters in data analysis. *European Journal of Operational Research, 173*, 849-865.

SAS. (2004). *Data mining Q & A with Dr. Kudyba*. Retrieved August 15, 2007, from http://www.dmreview.com/article_sub.cfm?articleId=1002434

Schena, M. (2003*)*. *Microarray analysis*. New York: John Wiley & Sons, Inc.

Segall, R. S. (2006). Microarray databases for biotechnology. In J. Wang (Ed.), *Encyclopedia of data warehousing and mining* (pp. 734-739). Hershey, PA: IGI Global, Inc.

Segall, R. S., & Zhang, Q. (2004). *Applications of modern heuristics and data mining techniques in knowledge discovery* (Funded proposal submitted to Arkansas State University College of Business Summer Research Grant Committee).

Segall, R. S., & Zhang, Q. (2005). *Continuation of research on applications of modern heuristics and data mining techniques in knowledge discovery* (Funded proposal submitted to Arkansas State University College of Business Summer Research Grant Committee).

Segall, R. S., & Zhang, Q. (2006a, March 2-4). Applications of neural network and genetic algorithm data mining techniques in bioinformatics knowledge discovery: A preliminary study. *Proceedings of the Thirty-seventh Annual Conference* *of the Southwest Decision Sciences Institute*, *37*(1). Oklahoma City

Segall, R. S., & Zhang, Q. (2006b). Data visualization and data mining of continuous numerical and discrete nominal-valued microarray databases for biotechnology. *Kybernetes: International Journal of Systems and Cybernetics*, *35*(9/10), 1538-1566.

United States General Accounting Office (GAO). (2004, May). *Data mining: Federal efforts cover a wide range of uses* (Rep. No. GAO-04-548). Report to the Ranking Minority Member, Subcommittee on Financial Management, the Budget, and International Security, Committee on Government Affairs, U.S. Senate . Retrieved August 15, 2007, from http://www.gao.gov/new.items/d04548.pdf

Wang, W., & Yang, J., (2005). Guest editors' introduction: Special issue on mining biological data. IEEE *Transactions on Knowledge and Data Engineering*, *17*(8), 1019-1020.

Williams, S. (2006, June 5-6). *Performance and capacity management for the world's online marketplace*. Paper presented at the SAS Conference, Cary, North Carolina.

Wu, R., Chen, R., & Chian, S. (2006). Design of a product quality control system based on the use of data mining techniques. *IIE Transactions*, *38*, 39-51.

Zhang, Q., & Segall, R. S. (2006), *Further Continuation of Research on Applications of Data Mining Techniques in Knowledge Discovery: An In-depth Investigation on Algorithms and Heuristics* (Proposal submitted to Arkansas State University College of Business Summer Research Grant Committee).

ADDITIONAL READINGS

Aref, W. G., Elfeky, M. G., & Elmagarmid, A. K. (2004). Incremental, online, and merge mining of partial periodic patterns in time-series databases. *IEEE Transactions on Knowledge & Data Engineering*, *16*(3), 332-342.

BaFail, A. O. (2004). Applying data mining techniques to forecast number of airline passengers in Saudi Arabia (domestic and international travels). *Journal of Air Transportation*, *9*(1), 100-115.

Chen, S. Y., & Liu, X. (2004). The contribution of data mining to information science. *Journal of Information Science*, *30*(6), 550-558.

Chun, S., & Kim, S. H. (2004a). Automated generation of new knowledge to support managerial decision-making: case study in forecasting a stock market. *Expert Systems*, *21*(4), 192-207.

Chun, S., & Kim, S. H. (2004b). Data mining for financial prediction and trading: Application to single and multiple markets. *Expert Systems with Applications*, *26*(2), 131-139.

Chun, S.-H., & Park, Y.-J. (2006). A new hybrid data mining technique using a regression case based reasoning: Application to financial forecasting. *Expert Systems with Applications*, *31*(2), 329-336.

Damle, C., & Yalcin, A. (2007). Flood prediction using time series data mining. *Journal of Hydrology*, *333*(2-4), 305-316.

Elfeky, M. G., Aref, W. G., & Elmagarmid, A. K. (2005). Periodicity detection in time series databases. *IEEE Transactions on Knowledge & Data Engineering*, *17*(7), 875-887.

Enke, D., & Thawornwong, S. (2005). The use of data mining and neural networks for forecasting stock market returns. *Expert Systems with Applications*, *29*(4), 927-940.

Garcia, M. P., & Kirschen, D. S. (2006). Forecasting system imbalance volumes in competitive electricity markets. *IEEE Transactions on Power Systems*, *21*(1), 240-248.

Ghiassi, M., Zimbra, D. K., & Saidane, H. (2006). Medium term system load forecasting with a dynamic artificial neural network model. *Electric Power Systems Research*, *76*(5), 302-316.

Gutiérrez, J. M., Cano, R., Cofiño, A. S., & Sordo, S. (2005). Analysis and downscaling multi-model seasonal forecasts in Peru using self-organizing maps. *Tellus: Series A*, *57*(3), 435-447.

Iliadis, L. S., & Spartalis, S. I. (2005). Fundamental fuzzy relation concepts of a D.S.S. for the estimation of natural disasters' risk (The case of a trapezoidal membership function). *Mathematical & Computer Modelling*, *42*(7/8), 747-758.

Kim, K. (2006). Artificial neural networks with evolutionary instance selection for financial forecasting. *Expert Systems with Applications*, *30*(3), 519-526.

Larivière, B., & Van den Poel, D. (2005). Predicting customer retention and profitability by using random forests and regression forests techniques. *Expert Systems with Applications*, *29*(2), 472-484.

Leigh, W., Hightower, R., & Modani, N. (2005). Forecasting the New York stock exchange composite index with past price and interest rate on condition of volume spike. *Expert Systems with Applications*, *28*(1), 1-8.

Menzies, T., Greenwald, J., & Frank, A. (2007). Data mining static code attributes to learn defect predictors. *IEEE Transactions on Software Engineering*, *33*(1), 2-13.

Petrovic, D., Xie, Y., & Burnham, K. (2006). Fuzzy decision support system for demand forecasting with a learning mechanism. *Fuzzy Sets & Systems*, *157*(12), 1713-1725.

Safer, A. M. (2003). A comparison of two data mining techniques to predict abnormal stock market returns. *Intelligent Data Analysis*, *7*(1), 3-13.

Schlink, U., Herbarth, O., Richter, M., Dorling, S., Nunnari, G., Cawley, G., et al. (2006). Statistical models to assess the health effects and to forecast ground-level ozone. *Environmental Modelling & Software*, *21*(4), 547-558.

Suh, S. C., Saffer, S. I., Li, D., & Gao, J. (2004). A new insight into prediction modeling systems. *Journal of Integrated Design & Process Science*, *8*(2), 85-104.

Sung, S. Y., Zhao L., Tan, C. L., & Ng, P. A. (2003). Forecasting association rules using existing data sets. *IEEE Transactions on Knowledge & Data Engineering*, *15*(6), 1448-1459.

Wang, J., Ma, Z., & Li, L. (2005). Detection, mining and forecasting of impact load in power load forecasting. *Applied Mathematics & Computation*, *168*(1), 29-39.

Wang, T.-C., & Chang, T.-H. (2007). Forecasting the probability of successful knowledge management by consistent fuzzy preference relations. *Expert Systems with Applications*, *32*(3), 801-813.

Yang, Y., Lin, H., Guo, Z., & Jiang, J. (2007). A data mining approach for heavy rainfall forecasting based on satellite image sequence analysis. *Computers & Geosciences*, *33*(1), 20-30.

Yoon-Joo, P., Byung-Chun, K., & Se-Hak, C. (2006). New knowledge extraction technique using probability for case-based reasoning: Application to medical diagnosis. *Expert Systems*, *23*(1), 2-20.

Yubin, Y., Hui, L., & Jixi, J. (2006). Cloud analysis by modeling the integration of heterogeneous satellite data and imaging. *IEEE Transactions on Systems, Man & Cybernetics: Part A*, *36*(1), 162-172.

Zhi-Hua, Z., &Yang, Y. (2005). Ensembling local learners through multimodal perturbation. *IEEE Transactions on Systems, Man & Cybernetics: Part B*, *35*(4), 725-735.

Chapter XXII
Supply Network Planning Models Using Enterprise Resource Planning Systems

Sundar Srinivasan
University of Missouri-Rolla, USA

Scott E. Grasman
University of Missouri-Rolla, USA

ABSTRACT

The advent of the Web as a major means of conducting business transactions and business-to-business communications, coupled with evolving Web-based supply chain management (SCM) technology, has resulted in a transition period from "linear" supply chain models to "networked" supply chain models. Various software industry studies indicate that over the next five to seven years, interenterprise business relationships, information structures, and processes will evolve dramatically. Enterprises will blend internal production and supply chain processes with those of their external trading partners. Currently, organizations are finding creative ways to mitigate supply chain costs while maintaining operational efficiency. New approaches, technologies, and methodologies are aiding with these cost-cutting measures to drastically reduce supply chain costs and increase customer satisfaction. This chapter discusses the background of supply chain planning and execution systems, their role in an organization, and how they are aiding in collaboration. The chapter concludes with a case study on how a supply chain management system could help an organization be more effective.

INTRODUCTION TO ERP SYSTEMS

Enterprise resource planning (ERP) systems aim to integrate all business functions and data of an organization into a single integrated system. The main component of an ERP system is the use of a common database. A typical ERP system landscape consists of a variety of hardware and software to help integrate the business functions and data. The goal of an ERP system is to provide a unified scheme to perform and record all the business activities of an organization and ensure organization, classification, and structure of the business processes and data.

An ERP system can be viewed as a group of processes, applications, and technology and consists of the following:

- Databases
- Applications to support business processes
- Network and systems infrastructure
- Middleware (group of software that aid integration of the various components)

This chapter discusses the evolution of ERP systems, provides brief information on the various ERP vendors and details the role and the impact of the integrated business software in manufacturing intelligence.

Evolution of ERP Systems

The concept of ERP has been around since the 1960s, and has its beginning in materials requirements planning (MRP). It was meant to provide an integrated approach to reduce inventory and process times and better manage procurement and production. In the 1970s, ERP systems evolved into manufacturing resource planning (MRP II) to involve financial and human resource planning in a limited capability. MRP and MRP II had their own limitations in terms of handling multiple locations, product aggregations, capacity constraints, and so forth. These limitations resulted in the development of ERP systems.

ERP systems, in the simplest sense, can be considered as a single, integrated database that gathers, stores, and helps analyze the data of an organization. Until the early 1990s ERP products were running on mainframes; however, with the advent of the client-server architecture in the mid-1990s, a majority of the ERP systems run on client-server architectures. The emergence of ERP systems went hand in hand with the idea of concentrating on single enterprises. The primary goal of these systems was to integrate the business processes of a single company. The business process integration capabilities were also very limited even if multiple companies of a conglomerate used the same information system. Among the biggest hurdles for this integration was the cost of technology. However, with significant development in technology over the last 5-7 years, the idea of cross-enterprise integration has become achievable and affordable. This led to the next generation of integrated business software products, commonly referred to as ERP II.

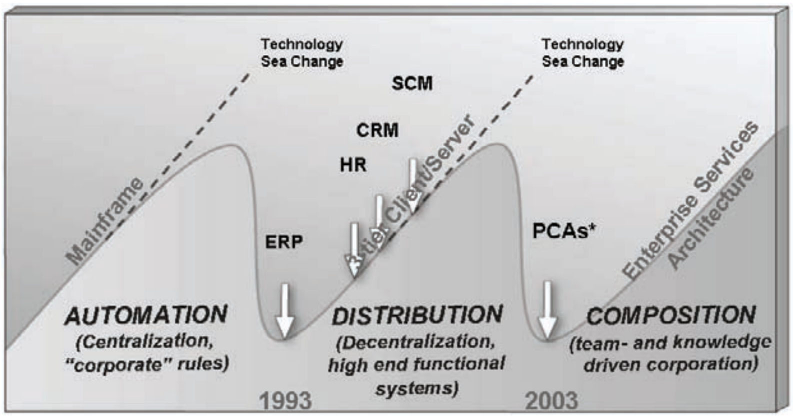
ERP II is the latest evolution that adapts ERP to the e-commerce environment through changes in functionality, technology, and architecture. The most evident change from ERP to ERP II is a change in focus of a business process from enterprise-centric to a collaborative environment. ERP II extends the scope of the business processes from an individual organization to all the stake holders in the supply chain. According to the Gartner Group, ERP II is “a business strategy and set of industry-domain-specific applications that build customer and shareholder community’s value network system by enabling and optimizing enterprise and inter-enterprise collaborative operational and financial processes” (Gartner Group, May 2001).

Table 1 represents the timeframe, industry needs, and the progress of technology of these systems.

Table 1. History of ERP systems

Year	Industry Need	Technology Progress
1970 - 1990	Real-time Automated systems	Automation systems, Transactional systems (OLTP)
Early 1990s	Scalable Integrated business processes	Analytical systems (OLAP), ERP
Late 1990s	Heterogeneous business processes	Integration, Web-services, “e-commerce”
2000s	Packaged composite business processes	ERP II

Figure 1. Paradigm shift in technology (Source: Gartner Group, 2003)



The 1970s through the 1990s saw a significant need for real-time automated systems to enable faster transaction entry. In the early 1990s this expanded by requiring analytical systems to process the transaction data in order to better understand the business. During the late 1990s, the industry requirement further expanded to requiring integration of Web-enabled heterogeneous systems. Organizations used proprietary interfaces to communicate between systems. This approach treats the Web application tier as just another silo, rather than as the integration hub through which all transactions flow, resulting in the need for the development of a technology that could reuse the applications developed. This technology is referred to as packaged composite application software, and the concept is very similar to objects in software development, where applications are built by reusing logic from two or more existing applications to form a new application without having to start from scratch. A composite application consists of functionality drawn from several different sources integrated by a technology platform. The components may be individual Web services, selected functions from within other applications, or entire systems whose outputs have been packaged as Web services.

Figure 1 represents the paradigm shift in technology and the products developed in the respective time frames.

Until the early 1990s, the ERP solutions were hosted on mainframes. With mainframe software architectures all intelligence is within the central host computer, and users interact with the host through a terminal that captures keystrokes and sends that information to the host. A limitation of mainframe software architectures is that they do not easily support graphical user interfaces or access to multiple databases from geographically dispersed sites. As a result of the limitations of file sharing architectures, the client/server architecture emerged in the mid-1990s. This approach introduced a database server to replace the file server. Using a relational database management system, user queries were answered directly. The client/server architecture reduced network traffic by providing a query response rather than total file transfer, improving multiuser updating through a GUI front end to a shared database. This technology evolution resulted in the IT landscape of most organizations being heterogeneous with a variety of enterprise suites, best-of-breed systems, and legacy systems. Also, they have a mixture of hardware platforms, operating systems, and databases.

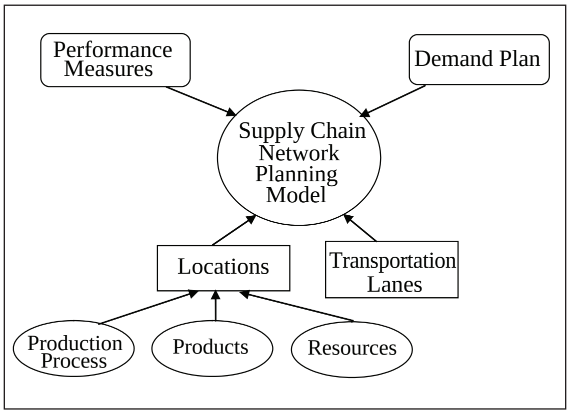
Further complications occur when organizations go through mergers, acquisitions, and divestitures. The traditional applications are designed for efficiency, not reuse, since there is no clear distinction between user interface, logic, and data, and they usually have a high cost of modification. In 2003, the concept of service-oriented architecture (SOA) was developed, which is the underlying structure supporting communications between services. An SOA solution consists of a composite set of business services that realize an end-to-end business process. SOA architecture is an application architecture that is designed to map directly to business requirements and is key to achieving the agility required. SOA, together with the emergence of service-oriented business applications (SOBAs) and service-oriented development of applications (SODA) are among the most significant shifts in IT.

Software is often completely inflexible once deployed, thus changing the nature of software can improve agility. For this to work, packaged business applications need to be broken down into smaller pieces that are easier to change. SOA makes it possible to look at software in pieces. As technology and standards evolve SOA will allow enterprises to mix applications services with others, regardless of the original supplier and the hardware and software platforms they use. We can think of it as separating business functions into subroutines/methods with the subroutines existing in any machine. SOA for an enterprise is usually referred to as the enterprise service architecture (ESA), and is the technology format currently being adopted by all the major ERP vendors in their product development. The evolution of SOA architectures and systems affects all levels, from business-IT alignment to hardware and devices. It will help software developers develop business applications faster and will help users to realize the full potential of integration.

It is important to understand that ERP systems were originally developed as transaction-based systems. With increased competition and shorter product life cycle, there was a significant demand for robust planning systems. It has also become very important for an organization to have near real-time collaboration and integration with all the stakeholders in the supply chain. Organizations have better understood the need for a higher level of information exchange with the vendors to ensure optimized cost and quality (Turban, McLean, & Wetherbe, 2004). Similarly, the impact of increased levels of collaboration with the distributors and the customers has also been recognized (Hitt, Wu, & Zhou, 2002).

The demand for faster and increased levels of information exchange saw the development of functionality-specific planning systems (e.g., supply chain management systems, product life cycle management systems, customer relationship management systems, and supplier relationship management systems, etc.). In addition, the roles

Figure 2. SAP Corporation Product Suite



of planning and execution in a business process are starting to converge. The future integrated business software products will be adaptive in nature, and groups of systems will continuously interact, exchange information, and iteratively plan and execute. For example, Figure 2 represents the product suite from SAP Corporation.

ERP Systems

mySAP ERP serves as the primary transaction processing system, while the rest of the systems are primarily designed for analytical purposes. Netweaver is the technology platform that integrates all these products. mySAP ERP is SAP's latest version of the core ERP software. It is primarily used as an online transaction processing system. Some of the capabilities of mySAP ERP include:

- Streamlining operations and optimizing the use of corporate resources and assets,
- Accelerating time to market and time to value,
- Delivering higher levels of service and more individualized products and services, and
- Enhancing customer satisfaction.

mySAP ERP includes financials (financial and management accounting and financial supply chain management), human capital management (talent management, core HR processes, and workforce deployment), corporate services (managing real estate; enterprise assets; project portfolios; corporate travel, environment, health, and safety [EH&S], quality, and global trade services), and operations (end-to-end procurement and logistics business processes, including discrete and process manufacturing).

ERP systems are the core of an organization's ERP landscape. They have become the primary transaction processing system. The historical transactions are stored in a data warehouse and are diced and sliced by the analytical systems – supply chain, customer relation, supplier relation, and product lifecycle – to help make better business decisions.

Customer Relationship Management Systems

Customer relationship management (CRM) is an information industry term for methodologies, software, and usually Internet capabilities that help an enterprise manage customer relationships in an organized way. For example, an enterprise might build a database about its customers that described relationships in sufficient detail so that management, salespeople, service providers, and perhaps the customer could directly access information, match customer needs with product plans and offerings, remind customers of service requirements, know what other products a customer had purchased, and so forth.

CRM consists of helping an enterprise enable its marketing departments to identify and target their best customers, manage marketing campaigns with clear goals and objectives, and generate quality leads for the sales team. CRM also involves assisting the organization to improve telesales, account, and sales management by optimizing information shared by multiple employees, and streamlining existing processes (e.g., taking orders using mobile devices). Enabling CRM allows the formation of individualized relationships with customers, with the aim of improving customer satisfaction and maximizing profits, and identifying the most valuable customers and providing them the highest level of service. CRM also provides employees with the information and processes necessary to know their customers, understand their needs, and effectively build relationships between the company, its customer base, and distribution partners.

According to the Gartner Group, there are three aspects of CRM, each of which can each be implemented in isolation:

1. **Operational CRM:** Automation or support of customer processes that include a company's sales or service representatives
2. **Collaborative CRM:** Direct communication with customers that does not include a company's sales or service representatives ("self service")
3. **Analytical CRM:** Analysis of customer data for a broad range of purposes

For example, mySAP customer relationship management provides solutions so that organizations can act immediately to improve sales, service, and marketing effectiveness. mySAP CRM provides the functionality to:

- Analyze, plan, develop, and execute all marketing activities.
- Help acquire, grow, and retain profitable.
- Drive service revenue and profitability with support for service sales and marketing.
- Enable e-commerce to increase sales and reduce transaction costs.

CRM systems have gained importance in the last three to four years. For CRM to work, companies must bring together a number of disparate processes, systems, and types of data, regardless of where they reside, to deliver an integrated, unified view of the customer that drives a consistent approach to interactions that is proactive as well as reactive. Adapting an integration technology, like ESA, will help organizations realize system.

Supply Chain Management Systems

Supply chain management (SCM) is the oversight of materials, information, and finances as they move in a process from supplier to manufacturer to wholesaler to retailer to consumer. Supply chain management involves coordinating and integrating these flows both within and among companies. The ultimate goal of any effective supply chain management system is to reduce inventory (with the assumption that products are available when needed). As a solution for successful supply chain management, sophisticated Web-enabled software systems are available.

Supply chain management flows can be divided into three main flows:

- Product flow.
- Information flow.
- Financial flow.

The product flow includes the movement of goods from a supplier to a customer, as well as any customer returns or service needs. The information flow involves transmitting orders and updating the status of delivery, as well as other inventory related data. The financial flow consists of credit terms, payment schedules, and consignment and title ownership arrangements.

There are two main types of SCM software: planning applications and execution applications. Planning applications use advanced algorithms to determine the best way to fill an order. Execution applications track the physical status of goods, the management of materials, and financial information involving all parties. Some SCM applications are based on open data models that support the sharing of data both inside and outside the enterprise (this is called the extended enterprise, and includes key suppliers, manufacturers, and end customers of a specific company). This shared data may reside in diverse database systems, or data warehouses, at several different sites and companies.

By sharing this data "upstream" (with a company's suppliers) and "downstream" (with a company's clients), SCM applications have the potential to improve the time-to-market of products, reduce costs, and allow all parties in the supply chain to better manage current resources and plan for future needs. For example, mySAP supply chain management enables adaptive supply chain networks by providing planning and execution capabilities to manage enterprise operations. mySAP SCM supports supply chain functionality to:

- Model existing supply chain and ensure a profitable match of supply and demand.
- Enable supply chain planning and with distribution, transportation, and logistics integrated into real-time planning processes.
- Provide network-wide visibility across your extended supply chain to perform strategic as well as day-to-day planning.

Supply chain management systems will yield improvements in the areas of cost, time, and quality. A very effectively implemented supply chain management system will help an organization with making processes more transparent and improve flexibility. SCM systems will not only concentrate on the order fulfillment cycle, but also incorporate product design, process recovery, and customer relationship.

Supplier Relationship Management Systems

Supplier relationship management is a comprehensive approach to managing an enterprise's interactions with the organizations that supply the goods and services it uses. The goal of supplier relationship management (SRM) is to streamline and make the processes between an enterprise and its suppliers more effective, just as customer relationship management (CRM) is intended to streamline and make more effective the processes between an enterprise and its customers.

SRM includes both business practices and software and is part of the information flow component of supply chain management (SCM). SRM practices create a common frame of reference to enable effective communication between an enterprise and suppliers who may use quite different business practices and terminology. As a result, SRM increases the efficiency of processes associated with acquiring goods and services, managing inventory, and processing materials.

For example, mySAP supplier relationship management helps simplify and automate procurement, and integrate strategic practices for supplier qualification, negotiation, and contract management more tightly and cost-effectively with other enterprise functions and their suppliers' processes. mySAP SRM supports:

- Supplier qualification, more efficient supplier negotiation, and better contract management.
- Requisitions, orders, goods receipt, and invoice settlement.
- Linking suppliers to purchasing processes and collaborate more effectively.

If correctly put into practice, SRM systems in a firm can enhance supplier selection, improve business interactions, and accelerate purchasing cycle time. SRM systems can definitely be an asset to companies, especially those who are trying to cut down costs.

Product Lifecycle Management Systems

Product lifecycle management (PLM) is a strategic business approach that applies a consistent set of business solutions that support the collaborative creation, management, dissemination, and use of product definition information. PLM supports the extended enterprise (customers, design and supply partners, etc.). PLM spans from concept to end of life of a product or plant and integrates people, processes, business systems, and information.

It is important to note that PLM is not a definition of a piece, or pieces, of technology. It is a definition of a business approach to solving the problem of managing the complete set of product definition information by creating that information, managing it through its life, and disseminating and using it throughout the lifecycle of the product. Three core or fundamental concepts of PLM include universal, secure, managed access, and use of product definition information. PLM helps maintain the integrity of that product definition and related information throughout the life of the product or plant. PLM also helps manage and maintain business processes used to create, manage, disseminate, share, and use the information.

PLM includes management of all product-related information from requirements, through design, manufacturing, and deployment. This information ranges from marketing requirements, product specifications, and test instructions and data to the as-maintained configuration data from the field. The PLM solution links information from many different authoring tools and other systems to the evolving product configuration. At the same time, the lifecycle began to include production-focused attributes and information. Manufacturing and operational process plans are also now viewed as an inherent part of PLM. Processes, and the workflow engines that control them, ensure complete digital feedback to both users and other business systems throughout each lifecycle stage.

The mySAP product lifecycle management application provides an integrated, single source of all product-related information needed for collaborating with business partners and supporting processes including product innovation, design and engineering, quality and maintenance management, and control of environmental issues. mySAP PLM had the functionality to:

- Manage specifications, bills of materials, routing and resource data, project structures, and related technical documentation throughout the product life cycle.
- Helps plan, manage, and control the complete product development process.
- Support collaborative engineering and project management.
- Provide integrated quality management for all industries.
- Coordinate enterprise asset management.
- Monitor, environment, health, and safety.

Enterprises are adopting PLM solutions to meet a variety of challenges. They are increasingly discovering that PLM helps them deal with daunting growth, global operations, and highly competitive market demands. PLM solutions connect people to work collaboratively, and centralize and improve the management of the product data. It streamlines the process steps to create, manufacture, and support products throughout their lifecycle from concept to retirement.

ERP VENDORS

ERP vendors are those organizations that develop, sell, and support ERP systems. The biggest ERP vendors include SAP AG, Oracle Corporation, and Microsoft Corporation. AMR Research expects the enterprise applications market to grow from

$47.8 billion in 2004 to $64.8 billion by 2009 (AMR Research, 2005). Table 2 lists the various enterprise applications, their current market size, and market forecasts until 2009.

Manugistics (recently acquired by JDA) and i2 were among the pioneers in developing supply chain management software. However, the traditional ERP software manufacturers like SAP and Oracle have also been developing supply chain management solutions.

Organizations implementing the entire suite of integrated software solutions have the option of selecting a single integrated system approach or the best-of-breed approach. The single integrated system provides efficient and reliable interfaces between the analytical and transactional systems. Examples of vendors providing a single integrated system include SAP and Oracle. The best-of-breed approach is adopted by organizations that use different "brands" of products for transactional and analytical systems. The best-of-breed approach may provide organizations with better functionality, but lacks the tight integration between the transactional and analytical systems. Historically, the best-of-breed approach has not provided good overall solutions (*Optimize*, September 2006).

The developments in technology have revolutionized manufacturing processes. The rest of this chapter discusses the impact of the technology and the latest software, and how they have enabled firms to be more profitable.

MANUFACTURING PROCESSES

Traditionally, manufacturing processes included planning and execution. However, with the understanding of the importance of the supply chain and the advances in technology, collaboration has also become an important process in manufacturing.

Until now, most ERP systems performed both strategic and operational level activities, with primary goal being operational activities. The systems performed online transaction processing and online analytical processing. As the planning criteria became complex, the traditional ERP systems were unable to provide accurate plan-

Table 2. Enterprise applications market size and forecasts, 2004-2009

Application	2004	2005	2006	2007	2008	2009	Five-Year CAGR
Supply Chain and PLM/CAD	**15131**	**15954**	**17301**	**18768**	**20242**	**21837**	**8%**
Supply Chain Management	5473	5523	5827	6147	6485	6842	5%
PLM	3960	4297	4957	5705	6493	7378	13%
CAD	5698	6133	6517	6916	7264	7618	5%
Enterprise Management	**21759**	**22042**	**23297**	**24404**	**25505**	**27226**	**5%**
Core ERP*	14899	14568	15140	15570	15964	16922	3%
Procurement	1991	2186	2394	2609	2818	3043	9%
Human Capital Management/HR	4869	5288	5764	6225	6723	7261	7%
Customer Management	**10902**	**11454**	**12260**	**13236**	**14427**	**15726**	**6%**
Total Market	**47792**	**49450**	**52858**	**56408**	**60174**	**64789**	**4.7%**

** Core ERP = Traditional ERP modules only: manufacturing, EAM, financial and accounting, B2B exchange platforms, integration, knowledge management, portals, and analytics/BI.* *Source: AMR Research, 2005* ning results. The primary reason is that the data structure of a transaction processing system is very inefficient for analytics purposes. However, with increasing demand for better planning capabilities, ERP vendors have developed products specific for these requirements. These products are referred to as supply chain management systems. The ERP systems have to be tightly integrated with the supply chain management systems to take advantage of its planning capabilities.

Planning

Manufacturing planning requires determining the location, type of product, and the magnitude of the customer demand. The planning results will differ depending on the time frame under analysis. For example, a strategic plan (long-term) may involve determining the product group requirement for a particular region. Tactical (medium-term) and operational (short-term) planning may involve more details about the product, a specific location in the region, and so forth. The first step in manufacturing planning involves calculating the independent requirements based on the forecast values and requirements from the sales information system and costing/profitability analysis. The next step involves the planning of only those items that are critical to the overall process. The final step involves creating the materials requirement planning. The output of MRP is a planned production order, purchase requisition, or a planned purchase order.

Execution

Manufacturing execution involves procurement and supply of goods and services among all the stakeholders in the supply chain. The first step in manufacturing execution is the release of the production order. The next step in this process involves the issues of the materials from the storage location to the work center. After the manufacturing process is complete, goods are received into stock.

Collaboration

Manufacturing companies have increased their productivity and efficiency over time by implementing new strategies like total quality management, lean manufacturing, and Six Sigma. Although these have significantly improved the efficiency, there is an increasing recognition that companies are competing as supply chains, not individual entities. To maintain a competitive advantage, manufacturers must make a significant change in strategy to effectively synchronize activities among functionally and geographically dispersed organizations. All the stakeholders of the supply chain need to collaborate. A collaborative manufacturing strategy would help a company maximize the effectiveness of its value chain and, hence, be more profitable. Collaborative manufacturing strategies will play a crucial role in helping world-class companies increase business value in the emerging global economy. To successfully meet marketplace requirements, manufacturers must create business processes that leverage shared information. ERP solutions provide the perfect platform to aid this collaboration.

A well integrated ERP solution provides all the stakeholders in the supply chain with relevant, real-time information and analysis to be efficient not only as an individual organization, but also as an entire supply chain.

ERP SYSTEMS AND MANUFACTURING PROCESSES

Supply Chain Management Systems

Planning is one of the most important tasks in a supply chain. The software systems that help perform planning are called as supply chain management systems. Advanced planning systems is the generic name for this breed of software. Advanced planning systems are by no means a replacement to the ERP systems. The aim of an advanced planning system is to address the deficiencies of an ERP system for planning. The two main characteristics of an advanced planning system are:

- **Integral planning:** Plan for the entire supply chain
- **True optimization:** Defining objectives and constraints for every part of the supply chain and solving it

However, it must be acknowledged that it is impossible to plan for the entire supply chain and at the same time perform optimization for every piece of the puzzle. Hence, a new architecture called the hierarchical planning system was developed. Hierarchical planning systems decompose the planning tasks into partial plans based on aggregation and disaggregating capabilities of time, products, resources, and so forth. Hierarchical planning systems provide the feasibility required for addressing integral planning and true optimization at the same time. Advanced planning systems are based on the hierarchical planning architecture, and address the planning requirements for all the four stages of the supply chain (i.e., procurement, production, distribution and sales). They also address planning for all time horizons (i.e., long-term, medium-term and short-term).

Product Lifecycle Management Systems

Product lifecycle management (PLM) systems describes a framework of technology and services that permit manufacturing companies and their partners and customers to collaboratively conceptualize, design, build, and manage products throughout their entire lifecycle. PLM systems enable organizations to create digital product information, and facilitate collaboration during the product development phase. PLM systems also control and automate critical processes such as release to manufacturing, change, and configuration management, throughout the product's lifecycle.

PLM systems have emerged as the primary means by which manufacturing companies can achieve significant improvements in their product development process. PLM systems are unique from other enterprise applications because they manage digital product information and optimize the digital product value chain. PLM systems also act as a document management system throughout the life cycle of a product. PLM systems have to integrate with other enterprise systems including traditional enterprise resource planning (transactional) and supply chain management systems (analytical).

CASE STUDY

Supply chain models give an overview of all elements in the supply chain and their relationship to each other, and are used to describe the strategic view of a supply chain from a planning perspective. There are a lot of models that break down the total supply chain into simpler subsystems and optimize each of them individually; however, optimized subsystems do not necessarily mean the total supply chain is optimized.

Supply network planning models integrates purchasing, production, distribution (of demands), and transportation so that comprehensive mid-term to long-term tactical planning and sourcing decisions can be simulated and performed on the basis of a single, global consistent model.

The case below discusses the methodology to define a supply chain model and suggests techniques for implementation of a supply chain network planning model in an ERP environment.

The model, if implemented, would serve as a very valuable tool to enhance learning experience.

Task 1: Defining the Supply Network Planning Model Agents

Supply network planning includes all the processes from the demand plan to the delivery of goods/services to the customer. Definition of the supply network planning process consists of three main subtasks and is represented in Figure 3.

Task 1a: Defining the Supply Chain Model

The supply chain model is a combination of the following agents: production process, locations, products, resources, and transportation lanes. Production process defines the detailed information required for manufacturing a product, and contains the recipe and the routing for the goods/services to be manufactured/rendered. A location is a logical or physical place at which products or resources are managed on a quantity basis. Location includes production plant, distribution centers, customers, and vendors. Resources enable the definition of capacities of equipment, machines, personnel, means of transport, and warehouses. Transportation lanes represent a direct route between two locations that can be used to source and transport products.

A new agent-based supply chain model, as shown in Figure 4, with production process, locations, products, and resources will be created in this research effort.

The location is assigned to the model and the production process, products, and resources are assigned to the location. The supply chain model will be used to synchronize activities and plan the flow of material along the supply chain, thus, creating feasible plans for purchasing, manufacturing, inventory, and transportation, as well as service enterprises.

Task 1b: Analysis of Supply Network Model

For analysis of the supply network model, optimization-based, heuristics-based, and supply/demand propagation-based planning techniques

Figure 3. Supply network planning process (Source: SAP AG, 2004.)

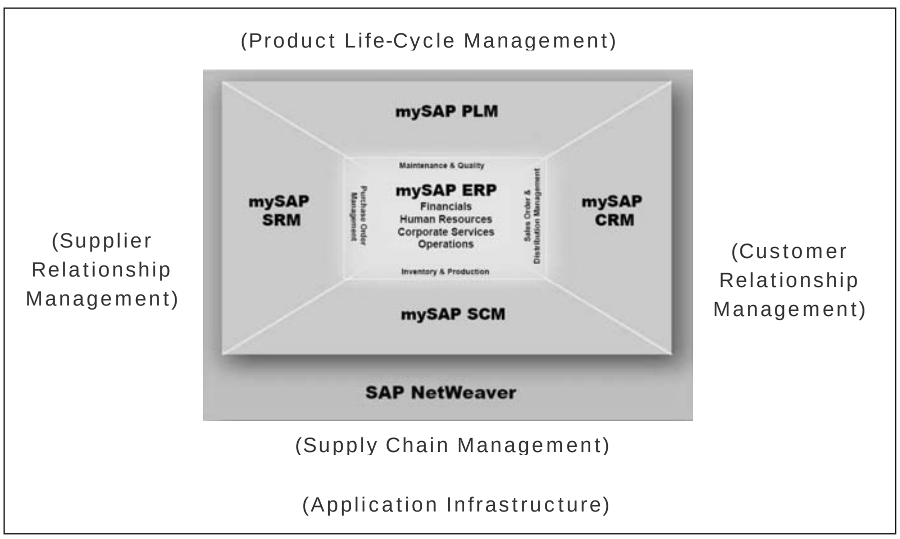
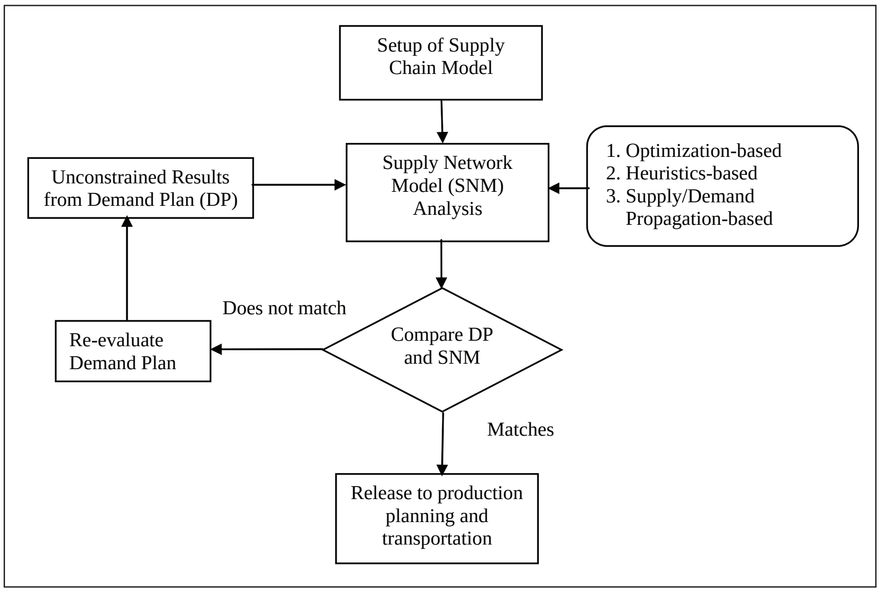


Figure 4. Agent-based supply chain network planning model



could be investigated to determine their feasibility and appropriateness. Optimization-based planning techniques are typically based on quantitative models aimed at minimizing costs (or maximizing profits) subject to constraints. Heuristics-based planning techniques work to create feasible (often nonoptimal) plans, and supply/demand-based planning techniques are based on statistical analysis of time series data. Quantitative and qualitative forecasting techniques can be used to develop a demand plan for comparison prior to release of the plan. The results from the demand plan (either real or simulated) are released to the supply network model and they form the basis for the analysis. The results from the demand plan do not include any constraints (e.g., production and distribution).

Task 1c: Validation and Implementation of Supply Network Planning Process

Validation of the model will be based on comparing the results from the constrained supply network model and the unconstrained demand plan. Since the initial model is formulated using an unconstrained demand plan, results may not be feasible when constraints are added, and the model must be modified. For example, available capacity may be less than the planned demand, hence, pricing parameters or resource constraints must be reevaluated.

The agents of the supply chain model represent the enterprise-wide integration, and the analytical agent-based model will be used to determine rules for agent behavior. Once the initial model and rules are developed, they would be evaluated and improved.

Task 2: Model Evaluation and Improvement

Measuring the effectiveness of the supply chain is very important. Successful evaluation and improvement of the agent-based supply chain integration model depends on concentrating on specific key business processes and developing an appropriate set of key performance indicators (KPIs) applicable at the enterprise level to measure the effectiveness of the supply chain. KPIs are quantifiable measurements that reflect the critical success factors of an organization, and depend on the product/service offered by an organization.

Task 2a: Definition of Performance Measures

Five main performance measures could be used for the evaluation of the supply chain models. The performance measures listed below are based on the guidelines from the SCOR model:

- **Supply Chain Delivery Reliability:** The performance of the supply chain in delivering the correct product, to the correct place, at the correct time, in the correct condition and packaging, in the correct quantity, with the correct documentation, to the correct customer.
- **Supply Chain Responsiveness:** The speed at which at which a supply chain provides products to the customer.
- **Supply Chain Flexibility:** The agility of a supply chain in responding to marketplace changes to gain or maintain competitive advantage.
- **Supply Chain Costs:** The costs associated with operating the supply chain.
- **Supply Chain Asset Management Efficiency:** The effectiveness of an organization in managing assets to support demand satisfaction. This includes the management of all assets, fixed and working capital.

Task 2b: Evaluation and Improvement of Performance Measures

The performance measures should be then evaluated for appropriateness for an enterprise-wide model in an ERP environment. Additional performance measures could be developed based on results of the definition and analysis/evaluation of the model. Comparison of the performance of the developed supply chain model with historic data and theoretical results provides an opportunity for benchmarking the performance of the model, as well as continuous improvement. The performance measures defined by SCOR will be studied for their validity and application on an enterprise-wide basis and additional measures will be developed, if necessary.

Task 3: Implementation of Supply Chain Network Planning Model in ERP Environment

Enterprise resource planning (ERP) is a software-driven business management system that integrates all facets of the business, including planning, manufacturing, sales, and marketing. The integrative capability of the ERP software makes it attractive for implementation of the agent-based supply chain integration model presented in this business case.

SAP advanced planner and optimizer (APO) is a component of the mySAP supply chain management solution that is used for planning and optimizing supply chain processes at a strategic, tactical, and operational planning level. APO is used for creating the model agents defined earlier, and assigning agents such as locations, products, resources, and production process models. After the initial assignments are made, agents for transportation lanes are added to link supply to demand locations, allocate products to the transportation lanes, and maintain quota arrangements. The developed model will be capable of tracking and evaluating supply chain agents including products, production process models and material handling, production, storage, and transportation. APO acts as the interface that acts as a top enterprise planning layer covering other planning areas such as manufacturing, demand, distribution, and transportation.

APO supply network planning (SNP) integrates purchasing, manufacturing, distribution, and transportation so that comprehensive tactical planning and sourcing decisions can be simulated and implemented on the basis of a single, globally consistent model. Supply network planning uses advanced optimization techniques based on constraints and penalties to plan product flow along the supply chain. The result is improved purchasing, production, and distribution decisions, reduced order fulfillment times and inventory levels, and improved customer service.

Starting from a demand plan, supply network planning determines a permissible short- to medium-term plan for fulfilling the estimated sales volumes. This plan covers both the quantities that must be transported between two locations (e.g., distribution center to customer or production plant to distribution center), and the quantities to be produced and procured. When making a recommendation, supply network planning compares all logistical activities to the available capacity.

The deployment function determines how and when inventory should be deployed to distribution centers, customers, and vendor-managed inventory accounts. It produces optimized distribution plans based on constraints (i.e., transportation capacities) and business rules (i.e., minimum cost approach, or replenishment strategies). The transport load builder (TLB) function maximizes transport capacities by optimizing load building.

For the implemented scenario, interactive demand planning will be used to create a demand plan, which is released to supply network planning to determine production planning. The process is simulated in APO utilizing planning techniques (cost-optimization, heuristics, and supply/demand propagation). Implementation of the supply chain model in an ERP environment will provide a test bed for validation, benchmarking studies, as well as further research on development and evaluation of analytical supply chain models. As an additional bonus, the implementation provides an infrastructure for creating case studies and exercises utilizing real-world data.

FUTURE RESEARCH DIRECTIONS

Further research in the field of supply chain management should consider analysis to determine the level of implementation of collaborative planning (such as CPFR) and collaborative replenishment (such as CRP) systems in different industries and the effects of this knowledge sharing on the performance of the supply chain. Collaborative planning includes the forecasting aspect of the demand management process, while collaborative replenishment covers the synchronization part of this supply chain process. Research and development of multidecision models which consider the objectives of the different companies will play an important role in these decision and planning systems.

Further research in the field of customer relationship management should try to analyze how different business units can use the same customer data: What specific type of integration alternatives are available for organizations working with rapidly changing CRM and SCM support technologies? Future studies should also develop decision and operations research tools to analyze the large amount of data gathered through the Internet. Further progress in research should be made in order to analyze intraorganizational and interorganizational effects simultaneously.

Further research in the field of order fulfillment needs to address the better use of information and creation of knowledge by using actual and new analytical and decision tools. The access to more data and information will put more emphasis in global optimization along the entire supply chain, instead of the usual models that focus on local optimization. More work considering the global supply chain will appear in this area. Also, more models using multicriteria decision making that reflect the integration and collaboration aspects of the e-fulfillment process will be the subject of future work.

Researchers should develop more decision models that take into account the global aspects of the supply chain to help to improve the manufacturing planning. Significant research is also required to help develop real-time tools, modeling, and decision systems that use real-time data available through the Internet.

CONCLUSION

With the continuous improvement in technology with software and computational capabilities organizations are able to better plan and execute their vision. The role of ERP software in this success cannot be undermined. However, there have also been instances where ERP implementations have resulted in large financial losses, even bankruptcy. It is important for organizations to recognize their core mission and capabilities and understand how ERP software can help achieve better efficiencies at their capabilities. If implemented the right way, ERP software unequivocally can help in the success of an organization.

REFERENCES

AMR Research. (2005). *Market analytics report: Enterprise resource planning, 2004-2009.* New York.

Baldwin, H. (2006). Are you the villain In the ERP Versus best-of-breed battle? *Optimize, 59.*

Chopra, S., & Meindl, P. (2003). *Supply chain management: Strategy, planning and operation*. New York: Prentice-Hall.

Gartner Research. (2001). *Decision framework research note DF-12-752*. New York.

Hitt, L.M., Wu, D.J., & Zhou, X. (2002). Investment in enterprise resource planning: Business impact and productivity measures. *Journal of Management Information Systems, 19*(1), 71-98.

Persson, F., & Olhager, J. (2004). Performance simulation of supply chain designs. *International Journal of Production Economics*, *77*, 213-245.

Quinn, F.J. (1999). The payoff potential in supply chain management. *Supply Chain Management Review*, *1.*

Turban, E., McLean, E., & Wetherbe, J. (2004). *Information technology for management* (4th ed.). New York: John Wiley & Sons.

ADDITIONAL READING

Barnes, J.C. (2001). *A guide to business continuity planning.* New York: John Wiley & Sons.

Chopra, S., & Sodhi, M. (2004). Managing risk to avoid supply-chain breakdown. *Sloan Management Review*, *46*(1), 53-61.

Fine, C., & Hax, A. (1985). Manufacturing strategy: A methodology and illustration. *Interfaces, 15*(6), 28-46.

Fisher, M. (2002). What is the right supply chain for your product? *Harvard Business Review, 75*(2), 105-116.

George, D., & Schoemaker, M. (2004). Driving through the fog managing at the edge. *Long Range Planning*, *37*(2), 127-142.

Hau, L. (2004). The triple-a supply chain. *Harvard Business Review*, *82*(10), 102-112.

Hayes, R., & Wheelwright, S. (1979). The dynamics of process-product life cycles. *Harvard Business Review*, *57*(2), 127-136.

Jeffrey, L., & Choi, T. (2004). Building deep supplier relationships. *Harvard Business Review*, *82*(12), 104-113.

Kopczak, L., & Johnson, E. (2003). The supply chain management effect. *Sloan Management Review*, *44*(3), 27-34

Lee, H.L., & Wolfe, M. (2003). Supply chain security without tears. *Supply Chain Management Review*, *7*(1), 12-20.

Pyke, D.F. (2000). Matching B2B e-commerce to supply chain strategy. *Supply Chain Management Review, Global Supplement*, 16-19.

Silver, E.A., Pyke, D.F., & Peterson, R. (1998). *Inventory management and production planning and scheduling.* New York: John Wiley.

Wise, R., & Morrison, D. (2000). Beyond the exchange: The future of B2B. *Harvard Business Review*, *78*(6), 86-96.

Chapter XXIII
Modeling and Analysis for Production Performance:
Analysis of U.S. Manufacturing Companies

Purnendu Mandal
Lamar University, USA

Enrique (Henry) Venta
Lamar University, USA

ABSTRACT

Modeling is a great approach to analyze long-term consequences of policy options in manufacturing. In this chapter two modeling approaches are discussed for understanding the intertwined relationships among factors which influence the performance and competitiveness of manufacturing: the system dynamics approach and the quantitative survey approach. The system dynamics approach is used to develop a conceptual model of the strategic issues that influence the performance and competitiveness of manufacturing, and the results of a quantitative survey are used to understand the actual extent of the influences of various factors in the current situation.

INTRODUCTION

The world of manufacturing has changed dramatically over the last decade. Increasing competitive pressures are being experienced in both domestic and international manufacturing markets; rapid advancement of information technology has opened the door to global markets. Indeed, one of the most important changes affecting manufacturing is the globalisation of business, which has affected the nature and rules of competition for nations as well as for firms. The competitiveness of a country is being linked to the performance of its firms in global competition (Porter, 1991).

Trading nations in the world economy have become increasingly interdependent. Consequently, integration with world industry and the use of international rather than domestic benchmarks is a fundamental determinant for company success.

Manufacturing companies wanting to compete in global markets must acquire the ability to become globally efficient and nationally responsive. Technological innovation and leadership is a major requirement at the company as well as at the national level. During the last decade, however, the manufacturing industry in many industrialized countries has experienced a sharp decline in productivity and manufacturing competitiveness. This has been exacerbated by a high degree of under utilized capacity, low productivity, high costs, and technological and social obsolescence. The decline in productivity suggests that the economic and work practices of the past may no longer be adequate to sustain living standards or provide employment for a new generation of workers (Love & Gunasekaran, 1997; Petzall, Selvarajah, & Willis, 1996).

The literature suggests that there are a variety of ways by which a manufacturing system can become competitive and several different kinds of solutions to the problem of achieving competitiveness have been proposed. Samson (1991) suggests that the formulation and implementation of a manufacturing strategy that focuses on a strategic manufacturing mission (SMM) and the definition of the relationship between manufacturing and other functions should be undertaken to improve competitiveness. But before the SMM can be developed, it is necessary that an understanding of the factors and relationships that influence the performance and competitiveness of manufacturing be obtained.

This chapter discusses two modeling approaches: the system dynamics approach and the quantitative survey approach. These two approaches are very useful for understanding the intertwined relationships among factors which influence the performance and competitiveness of manufacturing. The system dynamics model is very powerful in conceptualizing the strategic issues, whereas the quantitative survey is useful in understanding the actual extent of the influences of various factors in the current situation.

SYSTEM DYNAMICS MODELING

Essentially, system dynamics is the result of the cross-fertilization between the elements of traditional management, feedback control theory, and computer simulation. Feedback control provides a structure for building a model and a way of selecting the most appropriate information for decision making. Indeed, system dynamics has been principally developed as a methodology for improving the effectiveness of the decision-making process (Graham, Morecroft, Senge, & Sterman, 1992; Morecroft, 1988; Senge, 1990). During the 1950s system dynamics was known as *industrial dynamics*. Forrester (1958) applies this technique to production-inventory problems.

Understanding Causal Relations

To assist the decision maker a technique known as causal loop diagramming is used in system dynamics to show the relationships between various factors. Senge (1990) has undertaken interesting work in the area of causal relations. Senge (1990) has used this concept to show why certain process or patterns develop over time. He theorizes that there are patterns of causal behavior (or archetypes) that can explain why events happen in certain ways. For example, one archetype defined by Senge (1990) is the *vicious circle* as illustrated in Figure 1. This is interpreted as “A implies an increase in B which implies an increase in A which implies an increase in B……” and so on. For instance, population growth rate will increase population, and population will increase growth rate which will increase population further, resulting in a vicious circle of ever increasing population.

Many real life systems exhibit short term behavioral pattern that are similar to patterns generated by vicious circles. Growth in population, product demand, spread of epidemics, and so forth are some of the examples. Ultimately, however nothing grows forever. In the long run growth slows down and the system exhibits a stagnant or declining behavior. The effects of "limiting factors" become dominant at this stage. Senge (1990) explains the behavior of such systems using the "limits to growth" archetype structure. Figure 2 presents a simple "limits to growth" archetype for growth behavior in manufacturing output.

The left hand side part of the diagram (represented by the positive loop +A, which can be interpreted as a vicious circle) is responsible for an ever increasing growth in manufacturing output. (A positive feedback loop reinforces behaviour in the same direction). Investment leads to more manufacturing capacity and therefore more manufacturing output. More manufacturing output provides more investment in manufacturing. However, as it is clear, manufacturing output does not grow indefinitely and at some point of time the growth slows down. The right hand side of the diagram (represented by the negative loop -B) explains why the growth in manufacturing output slows down. (A negative feedback loop reinforces behaviour in the opposite direction). As manufacturing output increases, the economy reaches a higher level of manufacturing activities and the cost of inputs to manufacturing increases. As low cost inputs are used up first, the cost of inputs (such as raw materials, skill, technology, etc.) increases at the latter stages of industrialisation. Higher costs of inputs imply lower manufacturing output. Manufacturing suffers from high cost of production. The increasing manufacturing costs, however, can be contained by productivity improvement policies, but then there is a limit to which gains in productivity can be achieved. The combined behaviour of positive loop A and negative loop B over time is that the manufacturing output will continue to grow in the early stages of industrialisation, but the growth will come to a stand still or even decline at the latter stages of development.

Figure 1. The vicious circle

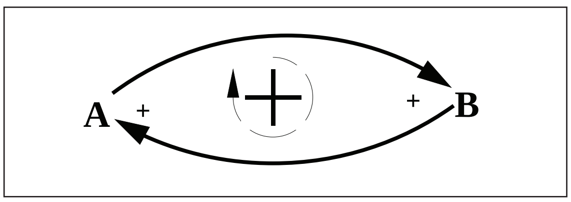
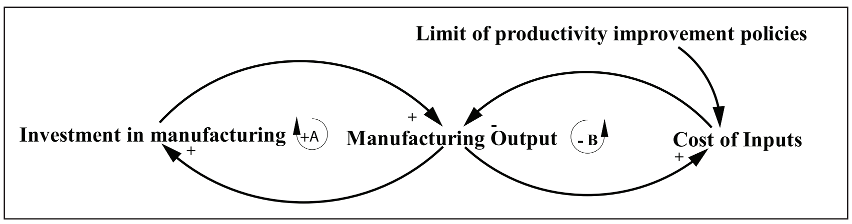


Figure 2. Limits to growth archetype for manufacturing output



Fundamentally, the manufacturing system can be characterised as being comprised of inputs, conversions, and outputs. Manufacturing focuses on converting inputs to desired outputs that satisfy certain requirements. To maintain a desired output, it is necessary to understand how the level of inputs and the changes in the conversion process are affected over time. Essentially, managers should be concerned with how inputs, such as investment, level of capacity, manpower, and workforce skill influence the conversion process by determining output measures, such as material utilisation, waste reduction, and quality. Furthermore, management should investigate and determine how process outputs influence inputs.

Modeling Characteristics

System dynamics has its own paradigm and has established itself as a powerful methodology (Mohapatra & Mandal, 1989). The modeling process is iterative, though the stages to be followed may appear to be sequential. Implicitly, Mohapatra and Mandal (1989) suggest that system dynamics can fulfill certain modeling requirements, especially in the context of the manufacturing system. These include:

1. **A holistic view of the problem:** The modeler can integrate a number of subsystems to give an overall picture of the manufacturing system.
2. **The policy as the central focus:** The modeling effort goes towards experimentation of policy issues.
3. **Long time frames:** Since system dynamics is a methodology for analyzing strategic issues, a long time horizon is desirable.
4. **Construction of causal relationships among variables:** System dynamics treats intrasystem interactions and the interactions between the system and its physical environment as cause and effect relationships. Thus, a system dynamics model is basically a dynamic picture of the perceived cause and effect relationships among the real system elements.
5. **Feedback loops as the basic building blocks of a model:** As was explained before, a feedback loop is created whenever an input to a system is affected by its output. For example, a company while selling goods from built up inventories usually takes into account the inventory level when planning future production. If the inventory is too high production will usually be slashed and vice-versa. This creates a loop in which inventory affects production which affects inventory, which, in turn, influence production (Figure 3). The phrases inside the brackets pertain to the inventory situation.

Figure 3. Feedback loop structure for action

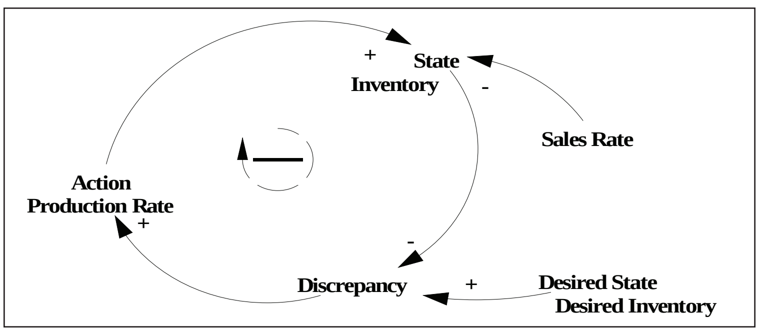


Figure 3 illustrates the general structure of a feedback loop and contains variables such as state, desired state, and action.

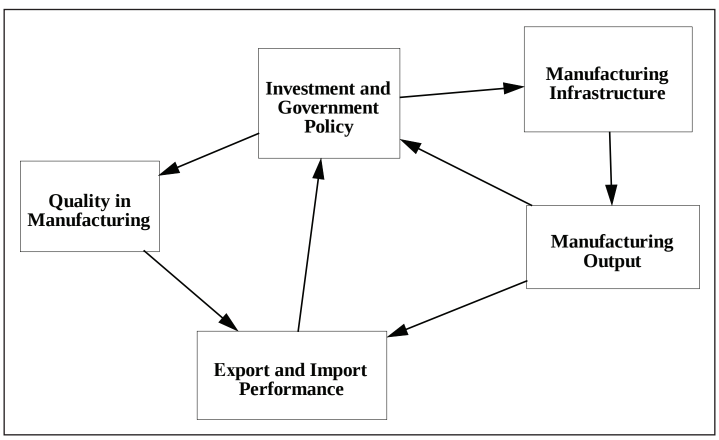
6. **Policy decisions are embedded in feedback loops:** The decision about the production rate (as shown in Figure 3) is based on the discrepancy, and the discrepancy is dependent on the actual inventory. In other words, the decision is dependent on the other variables that are connected by the feedback loops.
7. **Endogenous explanation of system behavior:** Modeling causal structures in feedback loops helps in providing endogenous explanations of real system behavior.
8. **Structure rather than the parameter values for a system:** A feedback structure can generate similar behavior for a wide range of values for each of the parameters. Therefore, model structures are given priority over model parameters.
9. **Validation through multistage procedures:** Building confidence in a system dynamics model and its usefulness are considered more important than the absolute model validity. Model validity in system dynamics is established through qualitative judgment and quantitative analysis.
10. **Model understanding as the basis for designing new policy:** System dynamics requires a complete understanding of the causal structures and the mechanisms that generate the behavior. This understanding paves the way for designing new policy structures.

SD Overview Of Manufacturing System

The importance of manufacturing in building a nation can not be over emphasized. For strategic reasons the manufacturing sector of a nation should attain a desired level of growth. If manufacturing attains international standards it can earn foreign exchange and improve dramatically the national economy and the domestic standard of living. However, creating a strong manufacturing base in order to meet international standards is a gradual process and requires consistent policies and effort on the part of government and industry.

A sectoral overview diagram of the major subsystems (or elements) of the manufacturing system at the national level is shown in Figure 4. The most important subsystems perceived to

Figure 4. Sectoral overview diagram of manufacturing system



be associated with a model of the manufacturing system are investment, manufacturing infrastructure, manufacturing output, export/import sector, and quality performance. The interrelationships and the direction of major influences among the subsystems are shown by the connecting lines and arrows.

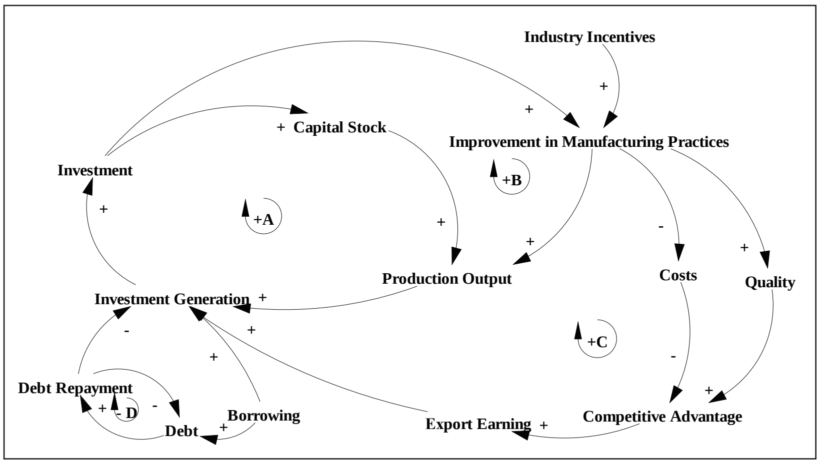
Implicitly, these influences will determine the importance of manufacturing in the national economy. If the national manufacturing system can meet the demands and expectations of the domestic market, then the contribution of manufacturing to gross domestic product (GDP) may increase and there will be greater pressure for improving the quality of goods and services. This in turn may call for more investment in manufacturing activities and further improvement in management practices. Both investment and better management practices may contribute to increasing output in quantitative terms and also the quality of products produced. Indeed quality is a major factor that will determine manufacturing capability in foreign and domestic markets (Yetton, 1990).

If the quality of products and services improves, the image of manufacturing in international and global markets may also improve, which in turn may increase export earnings. With increased export earnings the government will be in a position to invest more in manufacturing infrastructure as its capacity to earn more external earnings will strengthen its importance in the economy. Furthermore as the quality and quantity of goods and services improves demand will increase and this will be met by increased production. As a result, the contribution of manufacturing to GDP increases and generates more internal resources for investment.

Qualitative Analysis Of Policy Options

Figure 5 shows the overall conceptual causal loop diagram of a national manufacturing system. The causal loop diagram highlights a number of feedback loops that contribute to the development of the overall manufacturing system. The identification and qualitative analysis of the feedback loops can be used to determine likely development scenarios in the future. An attempt is made herein after to explain the impacts of policy options on the overall growth of manufacturing.

Figure 5. The overall causal loop diagram of manufacturing system



Option 1: Growth through Investment

The positive feedback loop "A" in Figure 5 indicates that the growth in output can be maintained by building capital stock in the manufacturing sector. Investment increases the level of capital stock, leading to increases in output. In the process, as output increases investment availability improves and ultimately reinforces the build-up of capital. The feedback loop "A" in Figure 5 shows behavior similar to the behavioral pattern generated by a "vicious circle" (Senge, 1990).

In reality however, there are many factors which act against or modify unrestricted growth. The major factors are the depreciation of capital, technological obsolescence, and types of manufacturing infrastructure. There could be other factors, but only these three factors are considered here for further explanation. Figure 6 is a refined and elaborate causal diagram of the aggregate feedback loop "A" of Figure 5. This diagram models two types of manufacturing infrastructure (high capital intensive manufacturing [HCIM] and low capital intensive manufacturing [LCIM]) and their respective outputs. Capital stocks in HCIM and LCIM increase as investment increases and decrease with capital depreciation. Capital depreciation is assumed here to depend on the average life of infrastructure. Capital depreciation is also assumed to depend on technological obsolescence (after all, what is the value of a piece of capital equipment if that technology is no longer in use!). A fraction of the total manufacturing output is assumed to have been saved and invested in HCIM and LCIM based on an investment policy. The causal relations shown in Figure 6 have been converted into computer equations using the Vensim package.

As an explanation of the effect of technological obsolescence on capital stock in HCIM and LCIM and manufacturing output, Figure 7 displays representative computer output from the model simulation. Time dependent behavior of capital stock in HCIM, capital stock of LCIM, and manufacturing output can be observed from the graphical output of the model. The simulation assumes a 10% per annum depreciation in the value of capital stock due to technological

Figure 6. A detailed model of investment, capital stock and manufacturing output

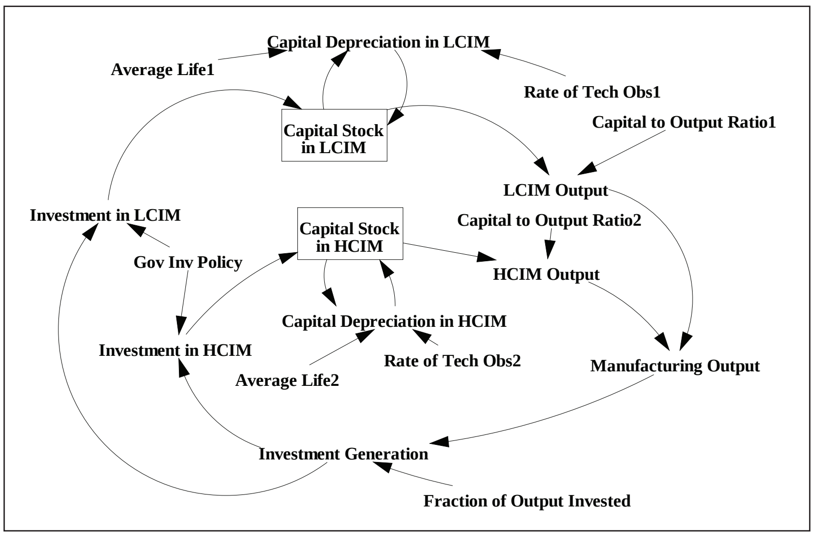
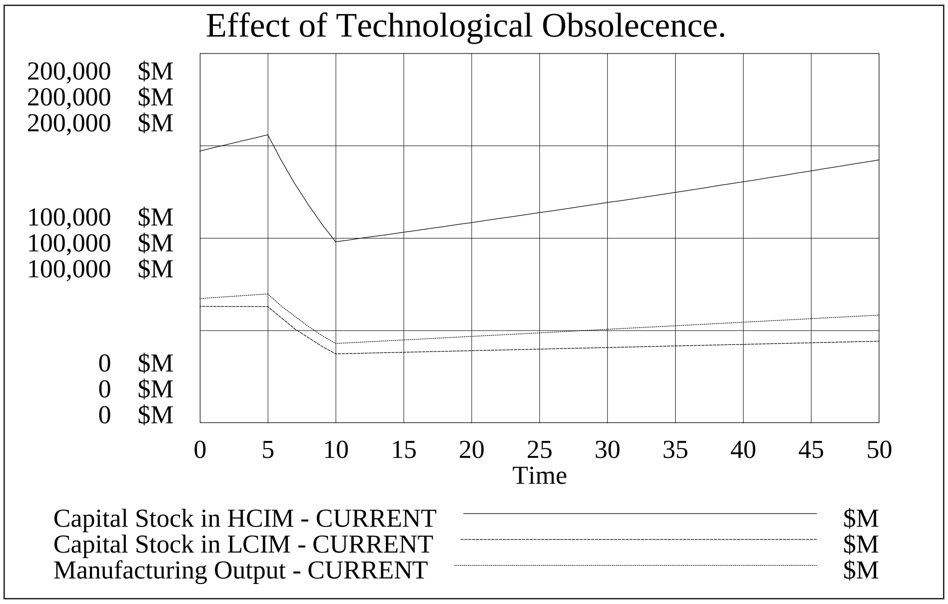


Figure 7. Effect of technological obsolescence to capital stock and manufacturing output



obsolescence for five years, from the fifth to tenth year of simulation. The effect is dramatic. Even at the end simulation period, the level of capital stocks remains lower than the initial level.

The investment rate needs to be more than the depreciation rate if growth is to be maintained. Furthermore, investment policy can also alter the growth scenario. The amount of investment in high and low capital intensive manufacturing can sometimes determine the creation of capacity and alter the output rate in the long-term. The simulation output presented in Figure 7 is for a 70/30 investment policy (70% investment in high capital intensive manufacturing and 30% in low capital intensive manufacturing). The testing of other investment policies could provide additional insights. The development of heavy industries such as metal manufacturing, iron, and steel, created by investments in high capital intensive capacities, may eventually provide manufacturing with a strong base. Research suggests that investments in high capital intensive areas such as advanced manufacturing technology are taking place (Sohal, Samson, & Weill, 1991).

Option 2: Growth through Improvement in Manufacturing Practices

Improvements in manufacturing management practices can influence the growth of the manufacturing sector. This is shown by the positive feedback loop "B" in Figure 5. Investments for improvement in management practices would in the long run keep manufacturing costs down and improve product quality.

Option 3: Growth through Competitiveness

National policies for manufacturing should focus on striving for higher quality, low cost, and short-lead times so that a competitive advantage can be attained in national, international, and global markets. Industry should achieve both economies of scale and economies of scope by adopting appropriate technologies. Improvements through quality can have a major impact on competitiveness. Research undertaken by Sohal, Ramsay, and Samson (1992) suggests that companies with a quality improvement program achieved a reduction in scrap, rework, stocks, machine down time, lead times, and warranty payments. The research of Sohal et al. (1992) indicates that companies are achieving both tangible and intangible results. These benefits, however, are not acquired overnight. If Australian industries become internationally competitive and increase exports, this will earn foreign exchange that will be invested to create additional capacity. Feedback loop "C" in Figure 5 indicates how the mechanism of export earning can strengthen growth.

Lessons Learned From System Dynamics Modeling

The management of technology in manufacturing is a complex issue and as such much depends on our understanding of the intricate relationships that may exist between types of manufacturing capacities, technological obsolescence, investment policies, cost reduction, and quality improvement strategies. The major influences in manufacturing from a system's perspective have been described. Major variables and their causal relationships have been established. By linking the various factors for a common theme the individual causal loop diagrams have been developed. Finally, with the assistance of an overall causal loop diagram the effects of various policy options are described.

System dynamics has been used to gain an understanding of the interdependencies of the subsystems of the manufacturing system and for determining their influences in the long-term dynamic behavior of the industry. There is an enormous research potential for modeling and analyzing policy issues in manufacturing through system dynamics.

The research presented here has been based on qualitative analysis, an important component of system dynamics modeling. Qualitative analysis is a prerequisite for understanding a system. It provides an avenue and direction for further investigation. It is envisaged that the ideas presented in this work will stimulate further discussion and the use of systems dynamics as a tool to assist managers in their future decision making and policy formulation in manufacturing.

QUANTITATIVE SURVEY APPROACH

There are needs for research on specific issues which require focused study on data sets for specific periods. Statistical analysis is often used in those situations. Data sets are normally generated through structured surveys or secondary data sources, and the focus is primarily on numerical values and attributes of the data sets. This chapter presents a modeling approach to U.S. manufacturing industry which required deployment of a nationwide survey in generating the data sets.

During the last decade, U.S. manufacturing firms have gone through major organizational restructuring, cultural changes, and other changes in manufacturing practices. There are doubts about the inherent strengths of U.S. manufacturing; various agencies, decision makers, and researchers are not sure whether the manufacturing industry is in a position to meet tough global competition. This work, based on a national survey of manufacturing companies, presents an overview of the present state of U.S. manufacturing practices and firms' performance. This survey makes a significant contribution to an ongoing effort of country specific studies spreading over four continents.

Much of the stagnation of U.S. manufacturing since the turn of the new millennium can be explained in terms of the worldwide economic glut and increasing competition from Third World economies. The current state of manufacturing has been a major concern to manufacturers and their associations. In a recent study, the Council of Manufacturing Associations (CMA) contends that "while manufacturing has been the engine for healthy economic growth and good jobs, intense global competition and the rising cost of doing business in the U.S. threaten manufacturing's capacity to maintain the nation's economic strength and standard of living" (CMA, 2003).

In order to sustain strong economic growth, U.S. manufacturers must improve business practices to achieve higher productivity, profitability, innovation, and other essential aspects of business leading firms to become globally competitive. Studies have shown (Fawcett & Myers, 2001) that organizational performance is directly influenced by organizational strategy and the structure of the organization. Both strategic focus and structural adjustments are changing rapidly due to advancements in information technologies, global and domestic competition, and organizational culture. Businesses are adjusting processes by incorporating advanced technologies, implementing total quality management (TQM), forming partnerships, and other mechanisms.

Business practices and performance are two different things; the hope is that good practices will lead to superior performance. Researchers have devoted considerable effort in classifying and categorizing various facets of manufacturing practices and hypothesizing about their impact on organizational performance. Ungan (2005) classifies the best practice context into three elements: best practice factors, organizational factors, and environmental factors. Through multiple regression analysis, with a survey data of 93 respondents, he established a positive significant association between management support and organizational resource availability, external pressures, perceived operational benefits, and compatibility.

Nahm, Vonderembse and Koufteros (2004) propose that organizational cultural factors affect manufacturing practices and performance. Based on Schein's (1993) work on three levels of cultural phenomena, they developed a framework to relate culture and manufacturing practices to performance. Their findings indicate that higher levels of customer orientation in the organization lead to higher levels of advanced manufacturing practices, which lead to high performance. Fawcett and Myers (2001) also proposed a conceptual framework for advanced manufacturing practices which tie the product and employee development practices to the manufacturing process practices of just-in-time production and manufacturing automation. This study looked into four popular manufacturing practices/strategies (integrated product development, employee development, just-in-time manufacturing, and manufacturing automation) and examined their interrelationships and impact on organizational performance. The study established significant and positive impacts of all four practices on organizational competitiveness.

Berndt and Morrison (1992) report on a study to determine the extent to which investments in "high-tech" office and information technology capital have resulted in reduced costs and increased productivity growth. They examine the relationships between investments in high-tech office and information technology capital and alternative industry performance measures such as labor and multifactor productivity, gross returns to capital, real expost internal rates of return, and markups over variable costs using data from U.S. Bureau of Economic Analysis and the Census and Annual Survey of Manufactures. They found only very limited evidence of a positive correlation between industry profitability and the share of information technology capital in the total physical capital stock.

Black and Lynch (2004) argue that changes in workplace organization, including re-engineering, teams, incentive pay, and employee voice, have been a significant component of the turnaround in productivity growth in the U.S. during the 1990s. Their work finds that workplace innovation appears to explain a large part of the movement in multifactor productivity in the U.S. over the period of 1993 through 1996.

Heshmati (2003) contributes a survey of recent contributions to, and developments of, the relationship between outsourcing, efficiency, and productivity growth in manufacturing and services. Issues of measurement of productivity growth, parametric, and nonparametric approaches to productivity measurement, econometric approaches to efficiency analysis, and the relationship between outsourcing and productivity growth in manufacturing and services are discussed.

Tippins and Sohi (2003) consider the payoff of investing heavily in information technology (IT) and propose that organization learning plays a significant role in determining the outcomes of IT. The authors develop the concept of IT competency, and using structural equations modeling with data collected from managers in 271 manufacturing firms, show that organizational learning plays a significant role in mediating the effects of IT competency on firm performance.

Black and Lynch (2001) use data from a nationally representative sample of union businesses to examine the impact of workplace practices, information technology, and human capital investments on productivity. They found that how a particular work practice is actually implemented within the establishment, rather than which work practice is adopted, is what is associated with higher productivity.

Ketokivi and Schroder (2004) offer a sound theoretical foundation for the proposition that manufacturing practices have competitive value. They build a theoretical argument of manufacturing practices, strategic contingency, and performance and test it in a sample of 164 manufacturing plants using a series of regression analyses. Results show that both the best practice and strategic contingency arguments have merit in explaining operational performance.

Laugen, Acur, Boer, and Frick (2005) assume that the best performing companies must be the ones deploying the best practices. In order to find out what are those practices, the highest performing companies in the 2002 International Manufacturing Strategy Survey database are identified, and the role 14 practices play in these companies was investigated. They conclude that process focus, pull production, equipment productivity, and environmental compatibility appear to qualify as best practices. Quality management and information and communication technology (ICT) may have been best practice previously, but have lost that status. E-business, new product development (NPD), supplier strategy, and outsourcing are relatively new, and cannot yet be qualified as, but may develop into, best practices.

The main purpose of the survey approach here is to present an overview of the business practices, environment, culture, and strategies practiced by manufacturing companies in the U.S. A thorough literature review was conducted to gain insights into a general framework through which one can link organizational strategy, culture, practices, and business environment to organizational performance. Figure 8 shows the linkages and elements in each category. The research and the consequent research questionnaire development are based on this general framework.

The following research questions are considered for investigation in this study:

- Is there a significant relationship between process management practices and product quality?
- Is there a significant relationship between organizational culture and financial performance?

- Is there a significant relationship between technology management practices and financial performance?
- Is there a significant relationship between company strategy and innovation performance?

Research Method

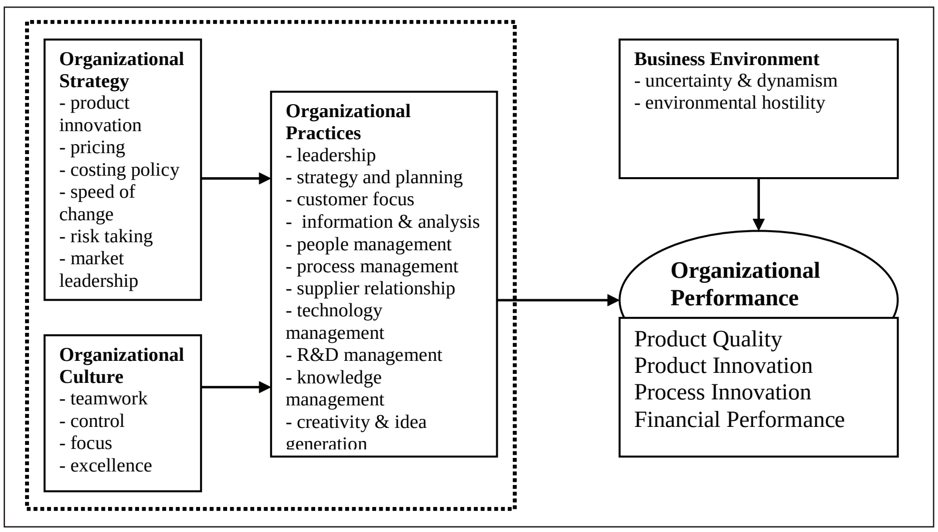
Empirical data was collected through a questionnaire survey of U.S. manufacturing companies. A six-page six-part questionnaire was developed which included questions on organization profile, organizational practices, organizational performance, business environment, organizational strategy, and organizational culture. The organizational practices part was designed to capture detailed input in areas of leadership, strategy and planning process, customer focus, information and analysis, people management, process management, supplier relationship, technology management, R&D management, knowledge management, and creativity, and idea generation. The organizational performance part asked for detailed input in the areas of product quality, product innovation, process innovation, and financial performance.

Two rounds of mail surveys were conducted. In the first mailing, roughly 1,500 letters were posted requesting CEOs or Presidents to respond to the survey. In the second mailing, 2,200 companies (including many of those approached in the first mailing) were approached. Altogether, 108 responses were received. The data was entered into a statistical package for social science (SPSS) file and analyzed.

Results Analysis

Table 1 presents some vital information about the companies who took part in the survey. Out of 108 companies in the sample, the majority (about 60%) of the companies employed between 101 and 500 workers. The average annual revenue was $314 million. About 80% of the companies were both ISO 9000 certified and had established TQM programs.

Figure 8. A framework of linking organizational strategy, culture, practices and environment to organizational performance



Process Management Practices vs. Product Quality

Question: Is there a relationship between process management and organizational performance related to product quality?

The perception of the product quality of the manufacturing companies responding is above average in their industry. The quality attributes (performance, conformance, reliability and, durability) ranked greater than 4 in the scale 1 (worst in industry) to 5 (best in industry).

Table 3 depicts the relationships between process management practices and the dimensions of product quality. Across the board, the strongest relationships were observed for "fool-

Table 1. Company profile

Number of Employees	Frequency	Percent	Average Annual Revenue	ISO 9000 Certification	TQM Program
< 100	4	3.7	$314 million	Certified = 89 companies Not Certified = 16 companies	Has TQM program = 86 No TQM program = 19
101 - 500	62	57.4			
501 - 1000	34	31.5			
> 1000	8	7.4			
Total	108	100.0			

Table 2. Product quality related performances

Organizational Performance: Product Quality (relative to major competitors in industry)	(1= worst in industry; 5= best in industry) Mean Score
Performance of products	4.43
Conformance to specifications	4.31
Reliability of products	4.41
Durability of products	4.42

Table 3. Correlation coefficients between quality practices and quality attributes

	Performance of products	Conformance to specifications of products	Reliability of products	Durability of products
The concept of "internal customer" is well understood		0.192 *		
Design processes in our organization is "fool-proof" (preventive-oriented)	0.271 **	0.302 **	0.173 *	0.177 *
Clear, standardized and documented instructions which are well understood by employees	0.242 **	0.274 **	0.177 *	0.183 *
Making extensive use of statistical techniques to improve production processes and reduce variation				

* weak relationship, correlation is significant at the 0.05 level (2-tailed)
** strong relationship, correlation is significant at the 0.01 level (2-tailed)

proof" design process, and clear and standardized instructions. It is interesting to note that extensive use of statistical techniques for process improvement and reduction of variation was not significantly related to any of the dimensions of product quality.

Organizational Culture vs. Financial Performance

Question: Is there any relationship between organizational culture and financial performance?

The perception of the financial performances of the manufacturing companies responding is above average in their industries, but not as high as the perception of product quality.

The table below shows the relationship between organizational culture attributes and dimensions of financial performance.

An organizational culture that believes in expansion, growth, and development was most strongly linked to financial performance. Also, showing strong relationship with some dimensions of financial performance were cultural attributes of teamwork and cohesion, stability, focus on task accomplishment and goals, focus on goal clarity, focus on efficiency, and belief in quality excellence. Centralization (or lack there of) and formalization and structure do not appear to be related to the respondents' perception of financial performance.

Table 4. Level of financial performance

Organizational Performance: **Financial Performance** (relative to major competitors in industry)	(1= worst in industry; 5= best in industry) Mean score
Sales Growth	3.66
Market Share	3.83
Profitability	3.81

Table 5. Correlation coefficients between cultural factors and financial performance

	Sales growth relative to major competitors	Market share relative to major competitors	Profitability relative to major competitors
Value of human relations, teamwork, and cohesion	0.251 **	0.179 *	0.248 **
Belief in expansion, growth, and development	0.388 **	0.263 **	0.304 **
Do not believe in control, centralization	0.199 *		
Do not believe in reutilization, formalization, and structure			
Promote stability, continuity, and order	0.169 *	0.264 **	
Focus on task, accomplishment, and goal achievement	0.218 *	0.198 *	0.250 **
Focus on direction, objective setting, and goal clarity	0.220 **	0.213 *	0.231 **
Focus on efficiency, productivity, and profitability	0.178 *	0.238 **	
Believe in excellence in outcome, and quality	0.210 *	0.215 *	0.228 **

* weak relationship, correlation is significant at the 0.05 level (2-tailed)
** strong relationship, correlation is significant at the 0.01 level (2-tailed)

Technology Management Practices vs. Financial Performance

Question: Is there any relationship between technology management practices and financial performances?

Table 6 shows the relationships between technology management practices and the dimensions of financial performance.

Overall, relationships between technology and financial performance were generally strong. It appears that respondents consider technological leadership an important way to maintain and improve financial performance.

Organizational Strategy vs. Product and Process Innovation

Question: Is there any relationship between organizational strategy and organizational performance in product innovation?

Table 6. Correlation coefficients between technology practices and financial performance

	Sales growth relative to major competitors	**Market share relative to major competitors**	**Profitability relative to major competitors**
Company always attempts to stay on the leading edge of new technology	0.229 **	0.184 *	0.274 **
Make an effort to anticipate full potential of new practices and technologies	0.273 **	0.249 **	0.286 **
Pursue long-range programs in order to acquire technological capabilities in advance of needs	0.254 **	0.217 **	0.277 **
Constantly thinking of next generation of technology	0.257 **	0.178 *	0.267 **

* weak relationship, correlation is significant at the 0.05 level (2-tailed)
** strong relationship, correlation is significant at the 0.01 level (2-tailed)

Table 7. Level of product innovation and process innovation

Organizational Performance: Product Innovation (relative to major competitors in industry)	**(1= worst in industry; 5= best in industry)**
Level of newness (novelty) in new products	3.63
Use of latest innovation in product development	3.68
Speed of new product introduction	3.34
Number of new products introduced to the market	3.50
Number of new products first-to-market	3.35
Organizational Performance: Process Innovation (relative to major competitors in industry)	**(1= worst in industry; 5= best in industry)**
Technological competitiveness	3.83
Speed of adoption of latest innovation	3.39
Newness of technology in processes	3.53
Rate of change in processes, techniques and technology	3.56

Table 8. Correlation coefficients between organizational strategies and product innovation considerations

	Product innovation of the company relative to major competitors in industry -				
	Level of newness (novelty) of company's new product	Use of latest technological innovations in new product development	Speed of new product introduction	Number of new products introduced	Number of new products first-to-market
Development and introduction of major and frequent product innovations is primary strategy	0.489 **	0.404 **	0.352 **	0.445 **	0.394 **
Attempts to be ahead of competitions in product novelty or speed of innovation	0.560 **	0.472 **	0.437 **	0.604 **	0.592 **
Company is growth-, innovation-, and development-oriented rather than favoring tried and true market	0.524 **	0.412 **	0.367 **	0.494 **	0.444**
Pursue a tough "undo the competitors" philosophy rather than trying to cooperate and coexist with competitors	0.309 **	0.221 **	0.234 **	0.312 **	0.170 *
Company has a strong inclination or tendency for high risk projects with chances of very high returns rather than low-risk projects with normal and certain rates of return	0.303 **	0.272 **	0.170 *	0.334 **	0.332 **
Price cutting and minimization of expenditures is very important strategy	- 0.198 *		- 0.169 *		- 0.183 *
Cost centers and fixing standard costs by analyzing variances for cost control is used frequently throughout the firm instead of only rarely or for a small part of operations					
Prefer to explore and make decisions on the basis of gradual and incremental change					

* weak relationship, correlation is significant at the 0.05 level (2-tailed)
** strong relationship, correlation is significant at the 0.01 level (2-tailed)

The perception of product innovation performance and process innovation performance of the manufacturing companies responding is above average in their industries.

Lessons Learned from the Survey

The main result of this study is not surprising: the characteristics, practices and culture of manufacturing organizations are strongly related to their financial performance, the quality of their products, and the degree of product and process innovation that they exhibit. There are also some relationships, or lack of relationships, that are surprising.

In the area of product quality much of what was found was expected. "Fool-proof" processes and its close relative, clear, standardized, and documented instructions are important in achieving high product quality. The concept of internal customer was not rated as an important determinant of product quality. The lack of relationship between the use of statistical techniques and any dimension of product quality was surprising. We

Table 9. Correlation coefficients between organizational strategies and process innovation considerations

	Process innovation of the company relative to major competitors in industry -			
	Technological Competitiveness	Speed in adoption of latest technologies in processes	Currency of technology used in processes	Rate of change in processes, techniques and technology
Development and introduction of major and frequent product innovations is primary strategy	0.393 **	0.447 **	0.431 **	0.397 **
Attempts to be ahead of competitions in product novelty or speed of innovation	0.476 **	0.461 **	0.422 **	0.487 **
Company is growth-, innovation-, and development-oriented rather than favoring tried and true market	0.444 **	0.484 **	0.449 **	0.474 **
Pursue a tough "undo the competitors" philosophy rather than trying to cooperate and coexist with competitors	0.234 **		0.173 *	0.219 **
Company has a strong inclination or tendency for high risk projects with chances of very high returns rather than low-risk projects with normal and certain rates of return	0.269 **	0.312 **	0.251 **	0.298 **
Price cutting and minimization of expenditures is very important strategy	- 0.261 **			- 0.225 **
Cost centers and fixing standard costs by analyzing variances for cost control is used frequently throughout the firm instead of only rarely or for a small part of operations				
Prefer to explore and make decisions on the basis of gradual and incremental change				

* weak relationship, correlation is significant at the 0.05 level (2-tailed)
** strong relationship, correlation is significant at the 0.01 level (2-tailed)

hypothesize that the use of statistical process control tools is now so ingrained into manufacturing processes that the survey respondents did not consider the use of those "routine" tools to be exceptional.

The elements of organizational culture are in many cases strongly related to financial performance as measured by sales growth, market share, and profitability. Teamwork, a philosophy of growth and expansion, and a strong focus on goals are all strongly related to financial performance. On the other hand, centralization or decentralization and a focus on efficiency were not as strongly related to financial performance. A belief in reutilization, formalization, and structure is unrelated to financial performance indicating a relaxation of traditional hierarchical structures in manufacturing organizations. U.S. manufacturers also indicate that technology management practices are directly related to financial performance.

The most interesting results show up in the relationships between organizational strategy and innovation. As expected, innovation is inversely related to price and cost cutting. The inverse relationships are particularly strong in the areas of technological competitiveness and rate of change in processes. Conversely, organizations that attempt to be leaders in innovation achieve strong relationships with speed of new product and process introduction and with the number of new products introduced.

A notable lack of relationship is between firms that use gradual or incremental change and product or process innovation. It has been widely understood that U.S. manufacturers tend to "play for the home run" rather that use the incremental approach of many Asian companies. This study reinforces the notion.

It is interesting to note that U.S. manufacturers rated their product quality much higher in their industries than they did their financial performance or their degree of innovation (product or process). These ratings raise some interesting questions that remain unanswered by the present study:

- Is there a time lag between achieving high product quality and seeing the results in the "bottom line"?
- Is product quality really a differentiating factor for the competitiveness of U.S. manufacturing?
- Is the U.S. still the world leader in innovation?

DISCUSSION

As demonstrated, the purposes of the two modeling approaches (system dynamics and quantitative survey) are 180 degrees opposite. One focuses on long term issues, while the other is more concerned with short term issues. One (system dynamics) produces dynamics behavioral trends to add to our understanding of issues; the other provides statistical data to justify answers to specific questions. Both the system dynamics model and survey served different purposes in our understanding of the U.S. manufacturing industry.

The purpose of a model is to help in explaining, understanding, or improving a system. A model can be used as an aid to thought, communication, or experimentation. Whenever a model is constructed it is done with a purpose or objective. The modeler must be specific about the objective and construct the model accordingly. The different probable objectives of constructing a model are:

- To use the model as an aid to understanding the system. The behavior of a complex system is decided by the interactions of the variables of the system. It may not be possible to visualize the interactions of the parts directly in the actual system. In such cases it becomes easier to study the system by analyzing the interactions in a model. Besides, a model can be used to isolate the factors that affect system performance to a high degree. It helps in recognizing the elements of the system that are necessary to achieve the desired output.
- To use the model as a guide to judgment and intuitive decisions. With an available system structure a model can generate the most likely behavior for a typical decision which is arrived at intuitively. In other words, the model can be used to answer if-then type of questions.
- To use the model to help establish desired policies. In an actual system it is not possible to test different policies before finally adopting one of them. Once a policy is adopted, its outcome will be observed in due course of time, and, the policy implementation cannot be undone. Moreover, it is not possible to undo the outcome, and another policy cannot be tried for the same time period. However, the model can be used for testing several policies regarding input, organizational structures, and so forth over the same time period and the best policy can be determined.

Thus, the main purpose of developing a model is to see that it helps in generating the behavioral characteristics of the system it represents. The model should be able to represent the nature of the actual system.

FUTURE RESEARCH DIRECTIONS

In model building exercises, it is very important to decide which factors to include in the model. This is done keeping in mind the purpose for which the model is constructed. One should realize that a system under investigation is a subsystem of a suprasystem. This subsystem is to be isolated from the rest of the suprasystem. Thus, it would be logical to think of an imaginary line that separates what is considered to be inside the system and what is considered to be outside. This imaginary line constitutes the system boundary. Such a boundary is drawn out of necessity. Only those elements or activities which are inside the system boundary would be considered while constructing a model.

This is perhaps the most important and, at the same time, the most difficult task in the entire exercise of model building. Exclusion of a part of a system from the model will result in the exclusion of the modes of behavior which are dependent on that part of the system. On the other hand inclusion of more and more parts of the system would result in increased complexity of the model. If it is intended to include the whole chain of relationships of any variable with others, it would go on stretching to more and more areas, and it may eventually lose the very basic purpose for which the model construction exercise was taken up. There must be a limit to the areas to be considered within a model.

The limit exercised in restricting the areas under the purview of a model automatically sets the boundary of the system. The model should include inside the boundary only those elements or activities that are dynamically significant for the purpose of the model.

In designing a dynamic model of a system, the factors that must be included arise directly from the questions that are to be answered. It should be kept in mind that there is no unique model for any system. For the same system, there may well be different models for different classes of questions. A particular model for a system may have to be altered and extended as new questions are explored. Skill and experience at model building help in deciding the factors to be included in the model. An experienced model builder will always ask the important questions pertinent to the system behavior (those having important answers). Questions that are too general or restrictive should be avoided.

A good understanding of model building helps in judging the validity of including a factor in a model. Here skill and practice in handling dynamic models provide aids in developing new models. Whether a factor is significant or insignificant for the model could be judged from past experience of model building.

Objectives of the model constitute a good starting point for structuring the model. The objectives are in relation to some output of the system. The model builder is to explore the group of factors related to this output. These factors may be technical, managerial, legal, economic, psychological, financial, social, and historical factors. Initially, only a few factors may be recognized and the skeletal model may be developed. Later, concentrating on each factor individually, other areas related to that particular factor can be explored. However, such a method may lead to inclusion of insignificant factors too. Good judgment on the part of the model-builder and thorough understanding of the system will help prevent such inclusions. Clearly, there are lots of research potentials in this area.

REFERENCES

Berndt, E. R., & Morrison, C. J. (1992). *High-tech capital formation and economic performance on U.S. manufacturing industries: An exploratory analysis* (Working Paper 3419-92). Sloan School of Management, Massachusetts Institute of Technology.

Black, S. E., & Lynch, L. M. (2001). How to compete: The impact of workplace practices and information technology on productivity. *The Review of Economics and Statistics, 83*(1), 1-28.

Black, S. E., & Lynch, L. M. (2004). What's driving the new economy? The benefits of workplace innovation. *The Economic Journal, 114*(493).

CMA. (2003, June 10). *The case for a strong manufacturing base* (CMA Press Release). Council of Manufacturing Association.

Fawcett, S. E., & Myers, M. B. (2001) Product and employee development in advanced manufacturing: Implementation and impact. *International Journal of Production Research*, *39*(1), 65-79.

Forrester, J. W. (1958). Industrial dynamics: A major breakthrough for decision-makers. *Harvard Business Review, 36*, 37-66.

Graham, A. K., Morecroft, J. D. W., Senge, P. M., & Sterman, J. D. (1992). Model-supported case studies for management education. *European Journal of Operational Research*, *59*(1), 151-166.

Heshmati, A. (2003). Productivity growth, efficiency and outsourcing in manufacturing and service industries. *Journal of Economic Surveys, 17*(1), 79.

Ketokivi, M., & Schroeder, R. (2004). Manufacturing practices, strategic fit and performance: A routine-based view. *International Journal of Operations & Production Management*, *24*(2), 171-191.

Laugen, B. T., Acur, N., Boer, H., & Frick, J. (2005). Best manufacturing practices: What do the best-performing companies do? *International Journal of Operations and Production Management*, *25*(2), 131-150.

Love, P. E. D., & Gunasekaran, A. (1997). Process reengineering: A review of enablers. *International Journal of Production Economics*, *50*, 183-197.

Mohapatra, P. K. J., & Mandal, P. (1989). The system dynamics paradigm. *Decision*, *16*(4), 251-266.

Morecroft, J. D. W. (1988). System dynamics and microworlds for policy makers. *European Journal of Operational Research*, *35*(3), 301-320.

Nahm, A.Y., Vonderembse, M. A., & Koufteros, X. A. (2004). The impact of organizational culture on time-based manufacturing and performance. *Decision Sciences*, *35*(4), 579-607.

Petzall, S. B., Selvarajah, C. T., & Willis, Q. F. (1996). *Management: A behavioural approach*. Longman.

Porter, M. (1991). *The competitive advantage of nations.* New York: Macillan Free Press.

Samson, D. (1991). *Manufacturing and operations strategy.* Australia: Prentice Hall.

Schein, E.H. (1993). How can organizations learn faster? The challenge of entering the green room. *Sloan Management Review, 34*(2), 85-92.

Senge, P. M. (1990). *The fifth discipline*. New York: Double Day.

Sohal, A. S., Ramsay, L., & Samson, D. (1992). Quality management practices in Australian industry. *Total Quality Management*, *3*(3), 283-299.

Sohal, A. S., Samson, D., & Weill, P. (1991). Manufacturing and technology strategy: A survey of planning for AMT. *Computer-Integrated Manufacturing Systems*, *4*(2), 71-79.

Tippins, M. J., & Sohi, R. S. (2003). IT competency and firm performance: Is organizational learning a missing link? *Strategic Management Journal, 24*(8), 745-761.

Ungan, M. (2005). Management support for the adoption of manufacturing best practices: Key factors. *International Journal of Production Research, 43*(18), 3803-3820.

Yetton, P. (1990). Going international. In internationalising Australia's economy: Vision and means in the garnet and Australian manufacturing council reports. In I. Marsh (Ed.), *Australian Journal of Management (Special Ed), 16*(3), 195-203.

ADDITIONAL READING

Checkland, P. (1998). *Systems thinking, systems practice*. New York: John Wiley & Sons.

Gunasekaran, A., Patel, C., & Tirtiroglu, E. (2001). Performance measures and metrics in a supply chain environment. *International Journal of Operations and Production Management, 21*(1/2), 71-87.

Maani, K. E., & Cavana, R. Y. (2000). *Systems thinking and modeling: Understanding change and complexity*. New Zealand: Prentice Hall.

Mandal, P., Howell, A., & Sohal, A. S. (1998). A systemic approach to quality improvement: The interactions between the technical, human and quality systems. *Total Quality Management Journal*, *9*(1).

Mohapatra, P. K. J., Mandal, P., & Bora, M. C. (1994). *Introduction to system dynamics modelling.* Hydrebad: Universities Press.

Rungtusanatham, M., Forza, C., Koka, B. R., Salvador, F., & Nie, W. (2005). TQM across multiple countries: Convergence hypothesis versus national specificity arguments. *Journal of Operations Management, 23,* 43-63.

Senge, P. (1992). *The fifth discipline: The art and practice of the learning organisation*. Random House.

Stair, R., & Reynolds, G. W. (2004). *Principles of information systems* (6th ed.). Cambridge: Course Technology.

Sterman, J. D. (2000). *Business dynamics: Systems thinking and modeling for a complex world*. McGraw-Hill.

About the Contributors

Dipak Laha is a reader in the Department of Mechanical Engineering at Jadavpur University, Calcutta, India. He received a BE in mechanical engineering from Bengal Engineering College, Shibpur, an MTech in industrial management from Indian Institute of Technology, Kharagpur and a PhD in Industrial Engineering from Jadavpur University. His teaching and research are in the areas of operations research, production planning and control, computer integrated manufacturing, evolutionary optimization, heuristics, operations management, and artificial intelligence.

Purnendu Mandal is professor and chair at the Information Systems and Analysis Department, Lamar University, Beaumont, Texas. His teaching and research interests are in the areas of database management systems, e-commerce, strategic management information systems, management information systems, and system dynamics. His research papers have appeared in *European Journal of Operational Research*, *International Journal of Production Economics*, *Management Decision*, *International Journal of Operations & Production Management*, *International Journal of Quality & Reliability Management*, *Logistics Information Management*, *Intelligent Automation and Soft Computing: An International Journal*, *International Journal of Technology Management*, *ASCE Journal of Management in Engineering*, *Decision*, *Applied Mathematical Modeling*, and so forth.

*** * ***

M. Reza Abdi is a lecturer in operations and information management at Bradford University School of Management. He has several years of teaching and industrial experience in various universities and manufacturing companies. His specialist areas are operations management, manufacturing system design, decision support systems, analytical hierarchical process (AHP), analytical network process (ANP), information management, computer modeling and simulation, and fuzzy logic.

Frédéric Adam is a senior lecturer in business information systems at University College Cork and visiting research fellow from the London School of Economics (LSE). He holds PhDs from the National University of Ireland and Université Paris VI (France). His research has been published in international journals including *Journal of Strategic Information Systems*, *Decision Support Systems*, and *Journal of Information Technology*. He is the editor-in-chief of the *Journal of Decision Systems* and vice-chair of the Working Group 8.3 on DSSs of the International Federation for Information Processing (IFIP).

Yoseph Tafesse Azene received his advanced diploma in mechanical technology from Bahir-Dar Poly Technique Institute, Ethiopia. Yoseph joined University of London, East in 1999 and earned a degree in manufacturing system engineering; obtained MSc in IT for product realization from Cranfield University. He worked as a team member on the development of a simulation training tool for engineers using the ERP- SAP R/3 system in Airbus wing manufacturing plant-UK. He is currently pursuing a PhD degree at Cranfield University, School of Applied Science. The research, titled "Rolling System Design Optimisation using a Soft Computing Technique," is sponsored by CORUS research and development, UK and EPSRC. Yoseph is a member of the Institution of Engineering Technology (IET).

Pankaj Barua received BTech and MTech degrees in agricultural engineering from Jawaharlal Nehru Agricultural University, Jabalpur, India and Punjab Agricultural University, Ludhiana, India. He obtained PhD degree from Tezpur University, Assam, India in management sciences. A part of his research for PhD involved development of Web-based DSS for water resources management. During his professional career he worked in the Department of Soil Conservation, Government of Assam, India and Assam Agricultural University, Jorhat, India at various capacities. At present he is working (since 2005) as Scientist, Precision Farming, Assam Agricultural University, Jorhat, Assam, India.

Fergal Carton is a college lecturer in business information systems at University College Cork in Ireland. With 15 years experience as an information systems consultant, starting with the Boston Consulting Group in London, he has worked on projects for over 30 major business customers involving the definition of IT strategy, enterprise software selection, and systems implementation. Most recently he spent four years implementing a global Oracle system in EMC^2's international subsidiaries. With a primary degree in computer science (University College Dublin), and an MSc in management from the European School of Management Studies (ESCP-EAP) in Paris, Fergal's research focus is on the value of enterprise-wide systems, and particularly their impact on organisation, effectiveness, and culture.

Narendra S. Chaudhari has obtained a PhD, MTech (computer science), and BTech in electrical engineering (with distinction) degrees from I.I.T. Bombay in 1988, 1983, and 1981, respectively. For his doctoral work, he worked on graph isomorphism problem where, besides other heuristic approaches, he developed an approach for strongly distinguishing regular graphs based on interior point approach (applied on polytopes defined on eigenspaces of graphs). He has been fellow, Institute of Electronics and Telecom Engineers (IETE), India since 2000. He has a varied experience in industry and R&D organizations, and he had been with Devi Ahilya University, Indore, India from 1988 to 2001. Since 2001, he has been with Division of Information Systems, School of Computer Engineering (SCE), Nanyang Technological University (NTU), Singapore, and currently he is an associate professor. He has more than 145 publications and four books to his credit. He has supervised more than fifteen PhDs in computer science and engineering. His current research interests are in the areas of game AI, specialized neural networks (e.g., binary neural networks), fuzzy inference, optimization algorithms, computational learning, and theoretical computer science. He teaches courses on discrete structures, algorithms, database systems, formal languages, automata and computability, programming languages, and compiler construction.

Swagatam Das is currently a lecturer in the Department of Electronics and Telecommunication Engineering, Jadavpur University, Kolkata. Das received his ME TelE and BE TelE degrees from the same university in 2005 and 2003, respectively, and is also pursuing his research towards a PhD degree. Das has contributed significantly in the broad spectrum of swarm intelligence and its applications in diverse fields of engineering and computing. His works have been reported/are due to appear in IEEE Transactions, Elsevier, World Scientific, and other top publishing houses.

Alistair Denby graduated from Nottingham Trent University in 1999 and worked as a software designer at Nortel Networks for three years before returning to university to pursue the post-graduate research on which this chapter is based. He completed his PhD entitled "Intelligent Laser Scanning for Computer Aided Manufacture" in 2006.

Prasanta Kumar Dey is a senior lecturer in Aston Business School, Birmingham, UK. He specializes in operations and information management. Prior to joining Aston, he worked in Indian Oil Corporation limited (a Fortune 500 company), India as a project executive for 13 years and worked in various oil refining and pipelines transportation projects. Subsequently, he directed and lectured in a graduate project management program in the University of the West Indies, Barbados for five years. Dr. Dey holds a bachelor in mechanical engineering from Jadavpur University in Calcutta, a master's degree in industrial engineering and management from Asian Institute of Technology, Bangkok, Thailand, and PhD in production engineering from Jadavpur University, Calcutta. He has published more than 70 articles in the international refereed journals like *International Journal of Project Management*, *International Journal of Production Economic*, *International Journal of Operations and Production Management*, *IEEE Transactions on Engineering Management*, *International Journal of Technology Management*, *TQM Magazine*, *Impact Assessment and Project Appraisal*, *Environmental Impact Review*, *Managing Services Quality*, and presented more than 50 articles in the international refereed conferences across the world. He is currently the joint chief editor of *International Journal of Energy Sector Management*, an Emerald journal. He has undertaken several research projects in the area of operations management in India, Thailand, Vietnam, UK, and the Caribbean countries and successfully completed those. His current research interest includes sustainability issues of industry, energy management, supply chain management, total quality management, risk management, and project management in both manufacturing and service industries.

Petr Fiala is a professor and the deputy head of the Department of Econometrics, University of Economics in Prague, the Czech Republic. He is a graduate of the University of Economics in Prague, Charles University in Prague, and the Rochester Institute of Technology in Rochester, NY. He holds a PhD from the University in Prague. His current research interests include modeling of production systems, supply chain management, project management, and multicriteria and group decision making. He is a member of professional societies for Operations Research and Multiple Criteria Decision Making and is the vice president of the Czech Society of Operations Research and the Czech Republic representative in EURO (The Association of European Operational Research Societies). He is the author of books, lecture notes, and papers in quantitative models and methods in management and economics.

Rajib Goswami obtained his master's degree in computer applications from Assam Engineering College, Gauhati University, India in 1993. At present he is working as systems analyst at Department of Computer Science and Engineering, Tezpur University, Tezpur, Assam, India. Before joining Tezpur University, he was associated with different manufacturing industries in ERP implementation projects.

He has developed several financial and logistic management systems and Web-based decision support systems. His research interests include soft-computing techniques in management, artificial learning algorithms, and decision support systems.

Gokop Goteng had his first degree and master's degree in computer science at Abubakar Tafawa Balewa University Bauchi, Nigeria in 1997 and 2003, respectively. He worked with an IT/Accounting firm from 1998 to 2003. He also had another master's degree in IT for product realisation at Cranfield University, UK in 2004 under the PTDF scholarship of Nigerian Government. Gokop worked briefly as a Programmer in Enterprise Integration Department in Cranfield University during which he developed part of the V-CES online software and E-RM software in 2005. Currently, he is a doctoral student at Cranfield University. He is a member of the British Computer Society (MBCS) and a student member of IEEE Computer Society.

Angappa Gunasekaran is a professor of operations management in the Charlton College of Business at the University of Massachusetts (North Dartmouth, USA). Previously, he has held academic positions in Canada, India, Finland, Australia and Great Britain. He has BE and ME from the University of Madras and a PhD from the Indian Institute of Technology. He teaches and conducts research in operations management and information systems. He serves on the editorial boards of 20 journals and edits a journal. Dr. Gunasekaran has published about 200 articles in journals, 60 articles in conference proceedings, and 2 edited books. In addition, he has organized several conferences in the emerging areas of operations management and information systems. He has extensive editorial experience that includes the guest editor of many high profile journals and the editor of several journals in operations management and information systems. He has received outstanding paper and excellence in teaching awards. His current areas of research include supply chain management, enterprise resource planning, e-commerce, and benchmarking. Also, Dr. Gunasekaran is the Chairperson of the Department of Decision & Information Sciences and the Director of Business Innovation Research Center (BIRC).

Scott E. Grasman (http://www.umr.edu/~grasmans/) is an associate professor of engineering management and systems engineering at the University of Missouri, Rolla. He received his BSE, MSE, and PhD degrees in industrial and operations engineering from the University of Michigan. Prior to joining UMR he was an Adjunct Assistant Professor at the University of Michigan. His research interests and areas of expertise include operations research, management science, supply chain, and logistics systems.

Hui He graduated from Inner Mongolia University of Science & Technology with a bachelor's degree in 2003. She got her master's degree from Beijing Normal University in computer science in 2006. During this time, she attended the National Natural Science Foundation project, and published five papers in International and national conferences. Now, she is the lecturer of Information Technology College in Beijing Normal University Zhuhai Campus. Her current research interests include remote sensing imagery processing and GIS.

John Hines is an e-commerce veteran, having operated an online custom dye house and manufacturing firm, Weefcraft Tie-dyed Apparel (TieDye.org), for over 10 years. A large international clientele attests to the success of the firm. Over the same period of time, Hines has also provided technical consulting services for a number of other small online businesses and he has been involved with extensive development of a successful nonprofit multiuser online gaming service. Currently, Hines, a student at Arkansas State University, is working toward a masters degree in e-commerce. Hines received the SWDSI Outstanding Student Paper award for his work, “Confronting the Performance Issues Related to Dynamic Web Technology,” which appeared in the 2006 SWDSI Proceedings. In addition, he has also presented research papers at numerous other regional and national conferences.

Dimitrios Karyampas was born in Athens in 1984 and finished undergraduate studies in economics at the Athens University of Economics and Business (AUEB). He previously worked in the Greek Ministry of Economy and Finance, in the Eleventh Public Economic Office. He is currently finishing an MSc in finance and econometrics at the University of York, UK. He is working on option pricing theory and risk management. His research interest is on pricing methods and forecasts of VaR using stochastic volatility models and numerical (computational) methods with applications in Finance.

Amit Konar is currently a professor in the Department of Electronics and Telecommunication Engineering, Jadavpur University, Kolkata. Konar received his PhD, MPhil, and ME TelE degrees from Jadavpur University in 1994, 1988, and 1985 respectively. He has published research articles on almost in all areas of artificial intelligence and soft computing. He is also an author/co-author of six books on different aspects of machine intelligence. His first book entitled *Artificial Intelligence and Soft Computing* (1999), published by CRC Press in Florida has remained a popular text in many U.S. and European universities. Konar was invited as a visiting professor in University of Missouri at St. Louis, in 2006. He is the recipient of the prestigious 1997-2000 career award for young researchers, offered by All India Council for Technical Education (India). He is a member of the editorial boards of reputed journals like *Neurocomputing, Elsevier and International Journal of Hybrid Intelligent Systems* (*IJHIS*), and IOS Press, The Netherlands.

Caroline Langensiepen is currently a senior lecturer in computing at Nottingham Trent University. Her previous history was as a physicist and then for many years as a system/software designer on military and other high reliability real-time systems. Her research continues this interest in real-time aspects, and includes graphics, image processing, and sensor data acquisition systems.

Mahendran Maliapen completed his PhD in decision science with system dynamics and data mining applications in health informatics and clinical data management in 2003. He is a certified information systems professional in Canada, UK, and Australia. He was nominated as IBM certified consultant in 2001. He has experience in the design and development of information systems and in particular in the discipline of decision support and data management for the pharmaceutical, life sciences and healthcare, telecommunications, insurance, and banking and finance industries. He has experience in developing predictive analytics and segmentation models and the completed segment migration analysis using Markov models. Mahendran has a distinguished career as a lead consultant and technology architect working with clients’ teams of sizes of between six to ten persons with high-level departmental responsibilities.

Tshilidzi Marwala is a chair and full professor of systems and control engineering at the University of Witwatersrand in South Africa. He holds a BS in mechanical engineering (Magna Cum Laude) from Case Western Reserve University, a Master of Engineering from the University of Pretoria, a PhD in engineering from University of Cambridge, and post-doctoral studies at Imperial College (London). In years 2006-2007 he was a visiting fellow at Harvard University. His research interests include theory and application of computational intelligence. He has published over 130 papers in journals, proceedings, and book chapters and has supervised 30 master and PhD theses.

Jose D. Montero (JoseDMontero@yahoo.com) is a doctoral student at Utah State University focusing his research on data mining, data warehousing, and data integration. He holds a bachelor's degree in mathematics and computer science from Goshen College, and a Master of Science in applied mathematics/ computer science from Western Michigan University. He has 30 years of industrial experience in computer science, information technology, and business information systems in areas such as insurance, banking, retail, and manufacturing. In the last nine years he has worked as manager and senior software developer at ATK Launch Systems Group. ATK is the world leader in the design, development, and production of rocket motors for space, strategic missile defense, and tactical applications.

Reinaldo Moraga is an assistant professor in the Department of Industrial and Systems Engineering at the Northern Illinois University. He received two BSc degrees in wood engineering in 1988 and industrial engineering in 1993 from Universidad del Bio Bio, in Chile, and has degrees of MS and PhD in industrial engineering from the University of Central Florida. His research interests are simulation modeling, operations research, and supply chain management.

Indrajit Mukherjee received his MTech (Quality, Reliability and Operations Research) form Indian Statistical Institute, Kolkata, West Bengal, India and BE (mechanical) from Jalpaiguri Government Engineering College, Jalpaiguri, West Bengal, India. He worked as a senior engineer in the department of central quality in Tata Engineering Limited, India, during 2001-2002. He submitted PhD thesis from the Department of Industrial Engineering and Management, Indian Institute of Technology, Kharagpur, India in August 2006, and he is presently working as a lecturer in the Department of Mechanical Engineering, Birla Institute of Technology and Science, Pilani, Rajasthan, India. He has published several papers in international and national journals of repute, and also a number of papers in international and national conferences mainly in the areas of inferential modeling and optimization of abrasive metal cutting processes. His areas of interest are quality engineering and applied operations research. He is a student member of several professional bodies such as ASME, and IIE, USA.

S. Mukhopadhyay is professor in the Department of Computer Science, Burdwan University, Burdwan, India. His research interest has been in the area of artificial intelligence, optimization techniques and cryptography. He has teaching experience of more than twenty-five years at the post graduate level and has published several research papers in reputed national and international journals.

Janet Poliakoff is a senior lecturer in computing at Nottingham Trent University. After a degree in mathematics at the University of Cambridge, she worked for the Open University as a tutor in mathematics before starting research at Nottingham Trent University, where she received a PhD in computing in 1993 on "The Digital Representation of Two-Dimensional Cutter Paths." Her research interests include digital representation of curves and surfaces, fairing of B-spline curves, measurement of surfaces using laser scanners, smooth motion generation for stepper motors, and monitoring of rotary motion.

Anoop Prakash is a research scholar at Research Promotion Cell (RPC), National Institute of Foundry & Forge Technology, India. His research interest includes machine loading, corporate memory, supply chain logistics, evolutionary algorithms, information retrieval, process planning and scheduling, and so forth. His research articles have been published in the reputed conferences like IMLF 2006, Taiwan, and so forth.

Pradip Kumar Ray is presently a professor and head in the Department of Industrial Engineering and Management, Indian Institute of Technology (IIT), Kharagpur, India, and received his PhD (in 1991) and MTech (in 1981) degrees from IIT, Kharagpur, and Bachelor of Mechanical Engineering (in 1979) degree from Calcutta University, India. Ray has over twenty-six years of diversified experience; eight years as senior industrial engineer at General Electric Company of India in Calcutta, two years as associate professor at Eastern Mediterranean University in Cyprus, and sixteen years of teaching and research experience at IIT. He has published over 70 papers in international and national journals and conferences in the areas of productivity measurement and evaluation, quality engineering, TQM, operations research, ergonomics, and other related topics. His areas of interest and research include productivity modeling, quality engineering, ergonomics, and JIT-based operations management. Ray is a certified lead assessor for ISO9000 registration, and actively involved in a number of industrial consulting in his interest areas. He is a member of several professional bodies such as INFORMS, IIMM, and a Fellow of World Academy Productivity Science.

Luis Rabelo is an associate professor in the Department of Industrial Engineering and Management Systems at the University of Central Florida and a NASA Fellow. He holds MS degrees in electrical engineering and engineering management, and a PhD in engineering management. In addition, he has a joint MS in engineering and business management from MIT. Rabelo has worked as a program manager for several Fortune 500 companies. He has three international patents and over 140 publications.

Rajkumar Roy has a background in manufacturing engineering and artificial intelligence. He started his professional career in manufacturing industry back in 1987, and worked in the area of knowledge engineering, decision support, and shop floor implementation of expert systems. His research projects have a strong focus on industrial applications. Professor Roy is currently leading the research in the decision engineering area at Cranfield. The research theme includes engineering cost estimating, design optimisation, and micro-knowledge management. He is a chartered engineer, and a member of international organisations such as IEEE, IED, and ACostE and has co-authored four books in his area of research. Roy is also involved in developing further research in national and European level.

Supriyo Roy is a senior faculty member and HOD in the Department of MBA, Management Institute of Durgapur, Durgapur, India. His research area includes optimization techniques in production management via soft computing/computational intelligence. He has published several research papers in reputed national and international journals like *Information System and Operational Research* (*INFOR*), Canada; *AMSE*, France; *JIOS*, Delhi; *KINDLER*, Kolkata; *SAJOSPS*, Kerala, and so forth.

Alfonso Sarmiento is a PhD candidate at the Industrial Engineering and Management Systems Department at the University of Central Florida. He has a MS degree in industrial engineering at the University of Florida, and a BS in industrial engineering at the University of Lima, Peru. Currently he is working as a graduate research associate at the Engineering Technology Department at the University of Central Florida. His research interests include supply chain management, operations management, and decision support systems.

Richard S. Segall is an associate professor of computer & information technology in the College of Business, Arkansas State University, Jonesboro, AR. He holds BS and MS in mathematics, and MS in operations research and statistics, all from Rensselaer Polytechnic Institute, and a PhD in operations research from University of Massachusetts at Amherst. He has served on the faculty of several universities including University of Louisville, University of New Hampshire and Texas Tech University. His publications have appeared in *Applied Mathematical Modelling, Kybernetes: International Journal of Systems and Cybernetics, Journal of the Operational Research Society*, and *Encyclopedia of Data Warehousing and Mining.* His research interests include data mining, database management, mathematical modeling, and applied statistics.

P.P. Sengupta earned his masters in economics, PhD from Jadavpur University, Kolkata, India and is a professor at National Institute of Technology, Durgapur, India. He has more than twenty years of teaching experience in graduate and post graduate level. His research interest has been in the area of international and development economics, operational research, and production management. Presently, he has been engaged in post-doctoral work in the area of international and development economics at National Institute of Technology, Durgapur.

Ravi Shankar is group chair of sectoral management and associate professor of operations and information technology management at the Department of Management Studies at Indian Institute of Technology Delhi, India. He has nearly 23 years of teaching and research experience. His areas of interest are supply chain management, knowledge management, flexible manufacturing systems, quantitative modeling, and so forth. He has over 180 publications in journals and conference proceedings. His publications have appeared in various journals including the *European Journal of Operational Research, IEEE-Systems Man and Cybernetics, International Journal of Production Research, Computers and Industrial Engineering, International Journal of Production Economics, Computers and Operations Research, International Journal of Supply Chain Management*, and so forth. He is the executive editor of *Journal of Advances in Management Research.*

Nasser Sherkat is professor of intelligent recognition at Nottingham Trent University. After a degree in mechanical engineering at the University of Nottingham, he joined Nottingham Trent University and received his PhD in 1989. He now leads the intelligent recognition research group and his research interests include multimodal human-computer interaction, pattern recognition, and classification algorithms.

Nagesh Shukla is a research scholar associated with Research Promotion Cell (RPC), National Institute of Foundry & Forge Technology, India. His research interests are related to developing evolutionary algorithms, learning of Bayesian networks and its application as diagnostic tool, data mining, statistical analysis and modeling of manufacturing systems, reliability optimization, vehicle routing problem, supply chain management, computer networks, and other soft computing paradigms. His research articles have been published in the reputed conferences and journals like *ASME Design Conference*, *IMechE,* *International Journal of Production Research*, and so forth. He is also acting as a reviewer of a few journals including *International Journal of Production Research* and so forth.

Sundara Raghavan Srinivasan received his masters in chemical engineering from the University of Alabama in Huntsville. He received his PhD with the Department of Engineering Management at the University of Missouri-Rolla in 2004. He was a post-doc with the School of Management and Information Systems University of Missouri-Rolla. Srinivasan is currently serving as director for the Center for ERP at the University of Missouri-Rolla. His of research include ERP, supply chain management systems, new product development, and product cannibalization.

Petros Theodorou holds a PhD in strategy of information systems, a PostDoct in finance and strategy and a HBSc in economics. Theodorou is currently an adjunct professor of corporate finance and managerial accounting in Athens University of Economics and Business and a senior researcher in the Department of Strategy and Planning in the Public Power Corporation (DEH/PPC S.A). His previous working experience was in Computer Logic, Astron/PEP, and so forth and on the board of directors in various firms. Previous teaching and research experience include business strategy and corporate finance and quantitative methods at the Technological Educational Institution of Thessaloniki, Piraeus and a permanent position of assistant professor at the Technological Educational Institution of Kavala in the Department of Management. Furthermore, he worked as a researcher at the Aristotle University in the Department of Economics and at the University of Macedonia in the Department of Applied Informatics. Theodorou is a member of NYA, Economic Chamber of Greece, Management of Technology Organization, Who's Who Marquis and so forth. He has published as the author and co-author on information technology strategy and financial performance in various sources such as Idea Group Publishing, Elsevier Science, The University Publishers of Crete, International *Journal of Management Science, Journal of Cases on Information Technology,* and in many conferences proceedings.

Ashutosh Tiwari graduated in mechanical engineering from the Indian Institute of Technology (IIT) Kanpur. He started his research career at Cranfield University in 1998 through the MSc in Engineering and Management of Manufacturing Systems. From November 1999 to January 2002 he worked as a Research Officer in Cranfield University on an EPSRC project entitled "Flexible Optimisation within CAD/CAM Environment (FLEXO)," which was jointly sponsored by Nissan Technical Centre—Europe (NTC-E) and SDRC (UK) Ltd. His PhD is on "Evolutionary Computing Techniques for Handling Variable Interaction in Engineering Design Optimisation." He is actively involved in research, virtual organisation teaching, I-DEAS teaching in postgraduate modules and short courses, CAD/CAM software evaluation and support, preparation of training and support material, and professional activities.

Manoj Kumar Tiwari is a professor of industrial engineering and operations research at National Institute of Foundry Forge Technology, Ranchi. He is currently associated with Research Promotion Cell (RPC) NIFFT and also handling the TEQIP project, sponsored by NPIU, Noida, India. Dr. Tiwari has contributed more than 150 articles in various leading International journals of repute and International conferences that include *IEEE—Systems Man and Cybernetics, European Journal of Operation Research* (*EJOR*), *International Journal of Production Research (IJPR), Journal of Engineering Manufacture, Part-B*, and so forth. He is serving as an editorial board member of various international journals like

IJPR, *Robotics and CIM*, *IJAMT*, and so forth. He has been pursuing collaborative research with several faculty members working in various universities of high repute worldwide.

Enrique (Henry) R. Venta was named dean at Lamar University in 2002. Prior to that time he had been at Loyola University Chicago since 1979. His teaching and research activities are in the areas of strategic issues in operations, work systems design, operations planning, total quality management, managing the service sector, and clinical decision making. His academic accomplishments include PhD and MS degrees in operations research and economics from Northwestern University and a BS degree in industrial engineering from the University of Puerto Rico. Venta has published or presented more than 40 articles, chapters, and monographs. He has been an internal consultant for Rock Island Railroad, a research associate at Argonne National Laboratory, and an associate of a firm of management consultants. Venta has also served as independent consultant for various government agencies, and national and international corporations. Until his move to Texas, he served on Chicago Mayor Daley's Council of Technology Advisors. Venta serves on the board of directors of the South East Texas Economic Development District and is a member of the Economic Development Council of the Beaumont Chamber of Commerce. He serves on the board of several civic and service organizations.

Christina Busisiwe Vilakazi received her BS and MS in electrical engineering with cum laude at the University of the Witwatersrand. She is currently an engineer at Eskom the sole electricity generator in South Africa. Her interests are in the theory and application of computational intelligence. She has published seven papers in journals, proceedings, and book chapters.

Meng Yang graduated from College of Information Science and Technology in Beijing Normal University in 2000 with a bachelor's degree in computer science. Then he completed a Master of Technology in computer science from Beijing Normal University in 2007. His interested research area is computer graphics.

Xianchuan Yu graduated from Jilin University with his PhD degree in mathematical geology (computing and its application of geology). From 1995 to 1998, he worked in the post-doctoral research center of computer software and theory in Jilin University. His current research interests include geosciences information data mining, mineral sources prediction, image processing, and pattern recognition. Then, he worked for College of Information Science and Technology of Beijing Normal University (BNU). Now, he is the academic leader of intelligent information processing at BNU. He is a member of IEEE CS (Computer Society) and SMC (Systems, Man, and Cybernetics), an executive member of Chinese Mathmatical Geoscience Society, and an executive member of Chinese Computer Graphics and GIS Society. He is also the reviewer of *Journal of Software*, *Natural Science Progress*, and so forth.

Xue-Ming Yuan is a research scientist of Singapore Institute of Manufacturing Technology (SIMTech) and adjunct associate professor of Department of Industrial and Systems Engineering, National University of Singapore (NUS). He has held various academic positions in China, France, Hong Kong, and Singapore. He has published over 50 papers in the leading journals which include *IEEE Transactions*, *IIE Transactions*, *OR*, *EJOR*, *JAP*, *Stochastic Models*, and so forth. He has been leading many research and industry projects in the areas of demand forecasting, inventory network optimization, global supply chain optimization and management, stochastic models and algorithms, and e-commerce fulfilment. His academic and industrial contributions and achievements have been included in Marquis Who's Who in the World, and Marquis Who's Who in Science and Engineering.

Libao Zhang graduated from Jilin industrial university in 1999 with a technology bachelor's degree. In 2002, he achieved the technology master's degree from Jilin University in signal and the information process specialty. In 2005, he graduated from the Jilin University communication and information system specialty with the PhD degree. Now he is the lecturer of College of Information Science and Technology of Beijing Normal University. His current research interests include image compression and code, image information retrieval, and human face recognition.

Qingyu Zhang is an associate professor of decision sciences & MIS at Arkansas State University. He has published in *Journal of Operations Management, International Journal of Operations and Production Management, International Journal of Production Research, International Journal of Production Economics, International Journal of Logistics Management, Kybernetes: International Journal of Systems and Cybernetics*, *Industrial Management & Data Systems*, and elsewhere. He has edited a book, *E-Supply Chain Technologies and Management*, published by IGI Global. His research interests are data mining, Web-based information systems, e-commerce/e-business, supply chain management, flexibility, and product development.

Ting Zhang graduated from College of Information Science and Technology in Beijing Normal University in 2000 with a bachelor degree in computer science. Then she completed a Master of Technology in computer science from Beijing Normal University in 2007. Her current research interests include data mining and digital image processing. She has published four papers in international and national conferences.

Wei Zou graduated from College of Information Science and Technology in Beijing Normal University in 2000 with a bachelor's degree in computer science. Then he completed a Master of Technology in computer science from Beijing Normal University in 2007. His interested research area is computer graphics.

Index

D

E

F

G

H

I

J

K

L

M

N

O

P

Q

R

S