

UNDERSTANDING  
COMPLEX SYSTEMS

Springer:  
COMPLEXITY

Manish Patel  
Sylvia Nagl

# The Role of Model Integration in Complex Systems Modelling

An Example from  
Cancer Biology



Springer

# Springer Complexity

---

Springer Complexity is an interdisciplinary program publishing the best research and academic-level teaching on both fundamental and applied aspects of complex systems - cutting across all traditional disciplines of the natural and life sciences, engineering, economics, medicine, neuroscience, social and computer science.

Complex Systems are systems that comprise many interacting parts with the ability to generate a new quality of macroscopic collective behavior the manifestations of which are the spontaneous formation of distinctive temporal, spatial or functional structures. Models of such systems can be successfully mapped onto quite diverse "real-life" situations like the climate, the coherent emission of light from lasers, chemical reaction-diffusion systems, biological cellular networks, the dynamics of stock markets and of the internet, earthquake statistics and prediction, freeway traffic, the human brain, or the formation of opinions in social systems, to name just some of the popular applications.

Although their scope and methodologies overlap somewhat, one can distinguish the following main concepts and tools: self-organization, nonlinear dynamics, synergetics, turbulence, dynamical systems, catastrophes, instabilities, stochastic processes, chaos, graphs and networks, cellular automata, adaptive systems, genetic algorithms and computational intelligence.

The two major book publication platforms of the Springer Complexity program are the monograph series "Understanding Complex Systems" focusing on the various applications of complexity, and the "Springer Series in Synergetics", which is devoted to the quantitative theoretical and methodological foundations. In addition to the books in these two core series, the program also incorporates individual titles ranging from textbooks to major reference works.

## Editorial and Programme Advisory Board

Dan Braha

New England Complex Systems, Institute and University of Massachusetts, Dartmouth

Péter Érdi

Center for Complex Systems Studies, Kalamazoo College, USA and Hungarian Academy of Sciences, Budapest, Hungary

Karl Friston

Institute of Cognitive Neuroscience, University College London, London, UK

Hermann Haken

Center of Synergetics, University of Stuttgart, Stuttgart, Germany

Janusz Kacprzyk

System Research, Polish Academy of Sciences, Warsaw, Poland

Scott Kelso

Center for Complex Systems and Brain Sciences, Florida Atlantic University, Boca Raton, USA

Jürgen Kurths

Potsdam Institute for Climate Impact Research (PIK), Potsdam, Germany

Linda Reichl

Center for Complex Quantum Systems, University of Texas, Austin, USA

Peter Schuster

Theoretical Chemistry and Structural Biology, University of Vienna, Vienna, Austria

Frank Schweitzer

System Design, ETH Zürich, Zürich, Switzerland

Didier Sornette

Entrepreneurial Risk, ETH Zürich, Zürich, Switzerland

# Understanding Complex Systems

---

**Founding Editor: J.A. Scott Kelso**

Future scientific and technological developments in many fields will necessarily depend upon coming to grips with complex systems. Such systems are complex in both their composition - typically many different kinds of components interacting simultaneously and nonlinearly with each other and their environments on multiple levels - and in the rich diversity of behavior of which they are capable.

The Springer Series in Understanding Complex Systems series (UCS) promotes new strategies and paradigms for understanding and realizing applications of complex systems research in a wide variety of fields and endeavors. UCS is explicitly transdisciplinary. It has three main goals: First, to elaborate the concepts, methods and tools of complex systems at all levels of description and in all scientific fields, especially newly emerging areas within the life, social, behavioral, economic, neuroand cognitive sciences (and derivatives thereof); second, to encourage novel applications of these ideas in various fields of engineering and computation such as robotics, nano-technology and informatics; third, to provide a single forum within which commonalities and differences in the workings of complex systems may be discerned, hence leading to deeper insight and understanding.

UCS will publish monographs, lecture notes and selected edited contributions aimed at communicating new findings to a large multidisciplinary audience.

Manish Patel and Sylvia Nagl

---

# The Role of Model Integration in Complex Systems Modelling

An Example from Cancer Biology

 Springer

## Authors

Manish Patel  
University College London  
Gower Street  
London, WC1E 6BT  
UK  
E-mail: Manni.patel@gmail.com

Sylvia Nagl  
University College London  
Gower Street  
London, WC1E 6BT  
UK  
E-mail: sylvia.nagl@gmail.com

ISBN 978-3-642-15602-1

e-ISBN 978-3-642-15603-8

DOI 10.1007/978-3-642-15603-8

Understanding Complex Systems

ISSN 1860-0832

Library of Congress Control Number: 2010933872

© 2010 Springer-Verlag Berlin Heidelberg

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilm or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer. Violations are liable to prosecution under the German Copyright Law.

The use of general descriptive names, registered names, trademarks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

*Typeset & Cover Design:* Scientific Publishing Services Pvt. Ltd., Chennai, India.

Printed on acid-free paper

9 8 7 6 5 4 3 2 1

springer.com

# Contents

<b>1</b>	<b>Introduction</b> .....	1
	1.1 Complex Biological Systems, Modelling and Model Integration .....	1
<b>2</b>	<b>Nature to Numbers: Complex Systems Modelling of Cancer</b> .....	5
	2.1 Mathematical Modelling in Tumour Systems Biology .....	5
	2.1.1 Growth Models .....	7
	2.1.2 Angiogenesis Models .....	16
	2.1.3 Treatment Response Models .....	21
	2.1.4 Modelling Methodologies .....	23
	2.1.5 Dynamic Pathway Models .....	27
	2.1.6 Other Models .....	31
	2.2 Summary .....	32
	2.3 Conclusion .....	32
<b>3</b>	<b>Coping with Complexity: Modelling of Complex Systems</b> .....	33
	3.1 Preamble .....	33
	3.2 Dynamic Complex Systems .....	35
	3.2.1 Properties of Complex Systems .....	37
	3.3 A Systems-Theoretic Approach in Biology: System Biology .....	43
	3.4 Coping with Complexity: Modelling of Complex Systems ...	46
	3.4.1 Individual-Based Modelling Methods .....	48
	3.4.2 Simulation Software .....	51
	3.5 Summary .....	54
	3.6 Conclusion .....	55

<b>4</b>	<b>Complexity and Model Integration: Formalisations</b>	57
4.1	Preamble	57
4.2	The Nature of Models	58
4.2.1	Abstract Views of Models	59
4.3	Model Integration: Previous Work	71
4.3.1	Model Integration in Management and Environmental Sciences	71
4.3.2	Model Integration in Systems Biology	74
4.4	Summary	76
4.5	Conclusion	76
<b>5</b>	<b>Novel Strategies for Integrating Models into Systems-Level Simulations</b>	77
5.1	Motivation	77
5.2	Novel Model Integration Formalisations	79
5.2.1	Linear Integration Strategy	79
5.2.2	Agent-Based Integration	81
5.2.3	A Knowledge-Driven Approach (KDA) for Addressing Model Focus and Scope	86
5.2.4	Knowledge-Driven ABI: A Novel Formal Protocol for Model Integration	90
5.2.5	Expected Results and Validation	94
5.3	Summary	94
5.4	Conclusion	95
<b>6</b>	<b>Experiments in Model Integration</b>	97
6.1	Model Integration Experiments	97
6.2	Knowledge Driven Approach: BIT-Building	98
6.2.1	Chen (2004)	99
6.2.2	De Pillis (2005)	99
6.2.3	Mallet (2006)	100
6.2.4	Markus (1999)	101
6.2.5	Zhang (2007)	102
6.3	ABI: Model Decompositions	104
6.3.1	Decomposition of the Diffusion Model	104
6.3.2	Decomposition of the Gompertz Model	107
6.4	KDA+ABI: Formal Model Integrations	111
6.4.1	Gompertz Model Integration with ODEs and Discrete Models	112
6.5	Summary	125
6.6	Conclusion	125

- 7 Discussion** ..... 127
  - 7.1 KDA/ABI Performance ..... 127
    - 7.1.1 The Knowledge-Driven Approach..... 127
    - 7.1.2 Agent-Based Integration ..... 135
  - 7.2 Conclusion ..... 152
  
- References** ..... 153
  
- Index** ..... 165



# List of Figures

2.1	Current state of the tumour modelling literature. . . . .	6
2.2	Gompertz curve . . . . .	9
2.3	MTS and Tumour cords . . . . .	10
2.4	Gleevec pathway . . . . .	24
3.1	Cellular automata . . . . .	49
4.1	Model granularity . . . . .	60
4.2	Model anatomy . . . . .	65
4.3	model abstractions . . . . .	65
4.4	Structured modelling . . . . .	72
5.1	Model relation to the real system . . . . .	89
5.2	Knowledge-driven approach . . . . .	94
6.1	Chen (2004) BIT . . . . .	100
6.2	De Pillis (2005) BIT . . . . .	101
6.3	Mallet (2006) BIT . . . . .	102
6.4	Markus (1999) BIT . . . . .	102
6.5	Zhang (2007) BIT . . . . .	103
6.6	Diffusion model . . . . .	105
6.7	PDE solution to the diffusion model . . . . .	108
6.8	GA-driven diffusion model solution . . . . .	108
6.9	Gompertz model BIT . . . . .	110
6.10	GA-driven solution to Gompertz model . . . . .	111
6.11	Ribba (2006) model . . . . .	112
6.12	Ribba (2006) BIT . . . . .	113
6.13	Tumour growth according to the Ribba (2006) . . . . .	116
6.14	Tumour growth according to the Ribba (2006), adding vessels . . . . .	116
6.15	Gompertz/Ribba BIT integration . . . . .	117

6.16	Gompertz/Ribba model integration . . . . .	119
6.17	Anderson (1998) BIT . . . . .	120
6.18	Integrated BIT for Anderson, Ribba and Gompertz models . . . . .	122
6.19	Integration of Ribba and Anderson models . . . . .	123
6.20	Angionenesis after model integration of Ribba/Anderson models . . . . .	123
6.21	Angionenesis after model integration of Ribba/Anderson models . . . . .	124
6.22	Fluctuations in hypoxia, i.e. total number of cells in a lattice site whose hypoxia state is set to true, whilst vessels grow. . . . .	125
7.1	Varying parameters in the integrated model . . . . .	142
7.2	Desired integration between SOAs and model integration software . . . . .	146
7.3	Linear integration in ABI . . . . .	147

# Chapter 1

## Introduction

**Abstract.** Model integration is central to cancer modelling and systems biology in general, but progress has as yet been stunted by a lack of cohesive effort and a common ground on which such an effort could occur. With rapid development in computational techniques, and significant advances within the systems biology community the real power of model integration is now coming to light. Here we describe how such a subject matter can be approached given the tremendous complexity of biological systems.

### 1.1 Complex Biological Systems, Modelling and Model Integration

Biology is shifting into an era of research where the use of mathematical and computational methods is becoming an integral part of the investigative cycle. Whilst traditionally engineering and computer science-oriented fields have guided biology in this respect, keeping themselves separate as disciplines, systems biology - a growing field in which biology, computer science and engineering meet - is quickly becoming a focal point for complexity research and modelling. This has been driven by technological innovation, e.g. microarrays and high-performance computing, enormous endeavours in scientific discovery, such as the genome projects, and the sheer necessity of disease prevention and cure. The fact is that there are no known systems more complex than the human body and other biological systems and therefore the opportunity for research is unique.

Currently at the core of the systems biology effort is the determination of behaviour at all levels of the system, from molecular to whole-organism. The benefits of such a wide understanding, once achieved, will be immense - not only will it yield greater understanding for the purposes of medicine, there is additionally a

reciprocal stream by which disciplines engage ideas from biology, e.g. bio-inspired computing<sup>1</sup>.

Modelling plays a fundamental role in this challenge. It is invariably a multi-faceted process that starts with conceptualising the target system - often simplifying it and dissecting it into manageable modules - upon which a set of mathematical and computational formalisms is based. It can be said of these formalisms:

- Their structure and convention can be extremely diverse.
  - For example, take as a random sample differential equations and Boolean networks. Not only are these fundamentally different in design, the process of solving and executing them is completely different.
- Their components can become extremely complex.
  - As will be seen, some modelling efforts encapsulate such depth of detail as modelling formalisms for complex systems are often difficult to solve without simplification.
- The model and its solver are sometimes difficult to separate.
  - An example of a solver technology is a differential equation solver. Some models have actually been published as complete packages wherein the model and the solver are tightly coupled.
- The semantics of such models are often difficult to conceptualise into a precise computable form.
  - Assumptions, context, data sources, caveats and simplifications are all usually written in free text since there is no standardised way to describe such metadata.
- Biological systems, and hence models, operate on a multitude of scales.

What is the nature of this diversity? How does this diversity relate to the real system? What is the commonality between these diverse formalisms? Can there be any *unity* in this diversity? These are crucial questions that have to be asked if modellers are to make their simulations more representative of the real system. The reason is that each effort focuses on different aspects of the system and they ultimately need to be integrated. Moreover the composition of elements within the same system can be such that they warrant different modelling approaches. Unifying principles must therefore be found in order to harness the full potential of individual modelling efforts, paving the way to a more complete understanding.

The most common name given to this unification is *model integration*. Implementation has taken different forms within the few disciplines in which it has been tackled, however the underlying principle and aim has remained constant: model

---

<sup>1</sup> This has been seen before in computer science, for example, where understanding of evolutionary processes has led to a great many number of applications in the form of evolutionary algorithms.

integration in its most common conceptualisation is a solution by which disparate models are executed in such a way that the resulting model and output(s) are concordant with the original models and represent a model of a larger system of which the individual models are constituent parts.

The difference in implementations is completely dependant on the nature of modelling methodologies within those fields of research. For example, as will be discussed in Chapter 4, the Management and Operations Research sciences have developed a kind of integration strategy that is applicable to the sequential nature of the models found there.

There does not seem to be any universal method by which one can integrate models. The necessity for such a technology is widely recognised, and is equally evident in the field of cancer systems biology. However, its development is somewhat stunted when compared to other areas of research. Primarily the reason for this is the progressive complexity of models and simulations used and hence their integration is not forthcoming.

The recent rapid development of mathematical models describing complex biological processes in cancer at all scales, encouraged by the pressing requirement for improved therapy, makes tumour systems a compelling target for model integration research.

For these reasons this book has aimed to meet this challenge head-on and apply it to cancer systems biology. Firstly an in-depth literature review is given in Chapter 2. This review discusses the current state of the tumour modelling literature, where research is heading and the benefits and shortcomings of current methods. Not only is this a review of representative models and techniques, it serves as the basis on which the aforementioned diversity of models is analysed in more detail. This analysis provides the information needed to investigate the precise challenges that model integration presents.

The types of approaches in any modelling-oriented field are parallel to the nature of the real system itself. This is especially true in tumour systems biology where the mathematical and computational techniques applied to tumour biology are found to be correspondingly intricate to be able to explain its highly non-linear behaviour. Therefore the science of complexity and systems in the context of cancer biology is described in relevant detail in Chapter 3.

Based on the findings of the literature review and insights into complex systems, a two-phase approach is developed, explained in Chapters 4 and 5, and is implemented as a parallelisable application as a proof of concept. The method borrows ideas from complex systems as agent-oriented systems and object-oriented concepts from software engineering to tackle model complexity and semantics. The theory is tested by integrating models from the literature and an analysis of the results is made in Chapter 6. Finally, the theoretical basis of the method is extrapolated to include additional technologies in Chapter 7 to show that not only is it a multi-disciplinary approach, but also to illustrate that it is a well-rounded approach that shows promise for cancer modelling as well as other modelling disciplines.

## Chapter 2

# Nature to Numbers: Complex Systems Modelling of Cancer

**Abstract.** This chapter represents a survey of the literature pertaining to mathematical modelling of tumour biology and techniques used to achieve tumour simulations. The purpose of this review is to (a) portray the vast array of techniques currently in use in the field, including mathematical and computational formalisms; (b) develop an appreciation of the enormous complexity of the problem domain, i.e. the tumour system, and hence portray the parallel complexity of the techniques needed to describe the system; (c) develop a detailed survey of model attributes so that the challenges that model integration poses, as well as possible solutions, can be defined.

### 2.1 Mathematical Modelling in Tumour Systems Biology

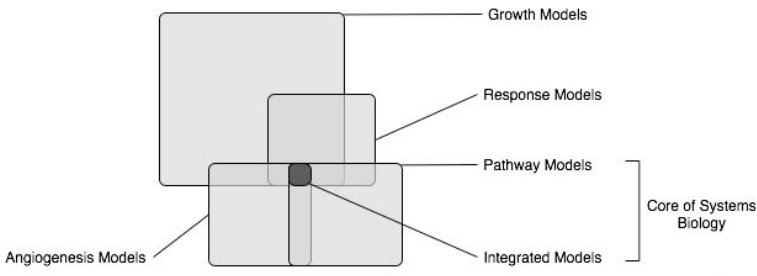
**Definition 1 (Model).** A model can be defined as an abstraction of a real system [131]. The abstractions of interest here are those that are expressed in some algorithmic form, be it mathematical, computational or logical. Simulation and model are often used interchangeably in the literature, however the term simulation can be defined as a model that is executed via a solver over a parametric variable, e.g. time. The terms model and solver are defined formally in more detail in Chapter 4.

Mathematical techniques in cancer biology have been applied for many years and their use has dramatically increased with the arrival of fast and accessible computation. The literature shows numerous examples of the use of mathematical modelling techniques, many of which have been previously applied to fields such as physics and environmental/meteorological sciences. It is evident that the application of such techniques is now ubiquitous in cancer research reflecting a general trend in almost all areas of biology and biomedicine. These analytical methods have been used to assist understanding of the emergence of complex tumour behaviour, which will be discussed in this chapter.

The types of available models focused on cancer biology range from general models, e.g. tumour growth, angiogenesis and signalling transduction pathway models, to specific models, e.g. behaviour in response to specific stimuli. Models of

biological processes are employed for diverse research purposes, notably, inquiries into fundamental tumour dynamics and drug target prediction. Algorithms used by these models range from differential equations, use of fractal theory [15], stochastic approaches [201] and more recently artificial intelligence techniques such as clustering and classification, neural networks, application of fuzzy logic [34], inductive/stochastic logic programming [163], Bayesian networks [98] and many more. Epidemiological models are also abundant in the literature, which are informative for the clinician in terms of modelling prognoses and diagnoses given various biological data<sup>1</sup>.

The types of models presented in this report have been broadly categorised into the following classifications: Growth Models, Angiogenesis Models, Treatment Response Models and Dynamic Pathways Models. Of course not all of the models presented here can be strictly classified into these types - in reality all are hybrids to some degree but most comfortably fall into one of the categorisations, as portrayed by Figure 2.1. As this figure suggests, there are virtually no models that incorporate all of these aspects to describe tumour behaviour. The following sections will highlight the fact that inclusion of diverse behaviours yield richer and more precise models. Indeed this is a key point in this project - models have traditionally dissected behaviour in such a way that it is often not obvious how to integrate them back together again. The methods described in this project endeavour to remedy that situation.



**Fig. 2.1** Current state of the tumour modelling literature. The majority of models can be classed as growth models; this aspect of cancer has attracted the most attention from modellers since it is one of the main clinically important behavioural features. More mechanistic models include response to treatment, angiogenesis and pathway dynamics. Relatively few models include a mixture of all these aspects. As will be discussed in later chapters, the core of systems biology has always been pathway/mechanistic-centric. Importantly, the issue of model integration is central to creating more representative models.

<sup>1</sup> This model class includes statistical models, which does not directly tackle systems behaviour and hence are not included in this review.

### 2.1.1 Growth Models

Growth models are the most abundant types of model for solid tumours in that all other modelling exercises must take into account the intrinsic growth functions that exist within the system.

In many ways growth models would seem to be the obvious target of tumour simulations as, ultimately, it is the aggressive growth behaviour of the system that leads to symptomatic disease. Indeed it is also this type of model, and findings related to these models, that seems to attract most attention at the media level. For example in 2008 ABC News published findings on biomarkers to help predict tumour size<sup>2</sup>.

Chignola (2000) [44] point out two important features that characterise the growth of tumours. Firstly, tumours are inherently made up of a heterogeneous population and therefore the growth dynamics can be quite complex. When modelling overall growth one is actually trying to model the growth of several phenotypically different populations. Secondly, growth is limited and cannot continue indefinitely. The constraint is placed by limitations in resources and space. All models in the literature seem to recognise these characteristics to some degree though it is not always explicitly stated.

The majority of the tumour modelling effort seems to be in the analysis of mechanisms that *control* growth since this is one of the most important factors that affect tumour development and response to treatment [33]. As a reflection of this one can see a rich variety of growth models in the literature dating back to the early 20th century. However, as yet, there is no universally accepted formalism that describes either vascular or avascular tumour growth [76].

#### 2.1.1.1 Generalised Formalisms

Growth models can be further subdivided by the the actual behaviour of growth that they model. Tumours are complex systems (explained in later chapters) and hence there are many facets on which to focus. For example, Cristini and coworkers at the University of Houston in 2009 explain "growth" in terms of invasiveness and metastases<sup>3</sup>. However, for the most part the growth modelling methodologies are concerned purely with a single body of parameterised growth.

The most simplistic formalisms include, at least in part, exponential models within defined bounds based on given restraints and difference equations [131]. Difference equations have the form:

$$x_t = R \cdot x_{t-1} \quad (2.1)$$

where  $x_t$  is the next discrete state of the system;  $x_{t-1}$  is the current state of the system (previous time index);  $R$  is a constant that denotes the rate of increase or decrease;

<sup>2</sup> <http://abcnews.go.com/Health/Healthday/story?id=5614404&page=1>

<sup>3</sup> [http://www.nanotech-now.com/news.cgi?story\\_id=33696](http://www.nanotech-now.com/news.cgi?story_id=33696)



As Mehrara (2010) and his colleagues of the University of Gothenburg explain, such simplistic formalisations rarely yield representative models<sup>4</sup>. More evolved models tend to have complicated functions in place of the parameter  $R$  rather than a constant. Collectively this can be termed as the state transition function.

A closed form of Equation 2.2 can be expressed:

$$x_t = x_0 \cdot R^t \quad (2.2)$$

which can be approximated to the following exponential when  $t$  is large:

$$x_t = x_0 \cdot e^{Rt} \quad (2.3)$$

where  $x_0$  is equivalent to the state at the first state/time index (e.g. tumour radius or mass).

One of the first generalised functions used to describe changes in population number, i.e. cell number in tumours, is the Gompertz function, formalised by Benjamin Gompertz in 1825. It was introduced specifically for tumour growth by Winsor in 1932 [193]. Other well-known functions include the logistic growth function and the von Bertalanffy growth function (these are very similar to the Gompertz function, and have been used to describe tumour growth previously, but will not be discussed any further here since the explanation of the Gompertz model here suitably covers the extent to which these kinds of model can describe tumour dynamics; the reader is directed to Marusic et al., 1994 [124], for a comparative review).

The Gompertz function reads as follows, although many forms of the equation exist:

$$Y = A + Ce^{-B(t-M)} \quad (2.4)$$

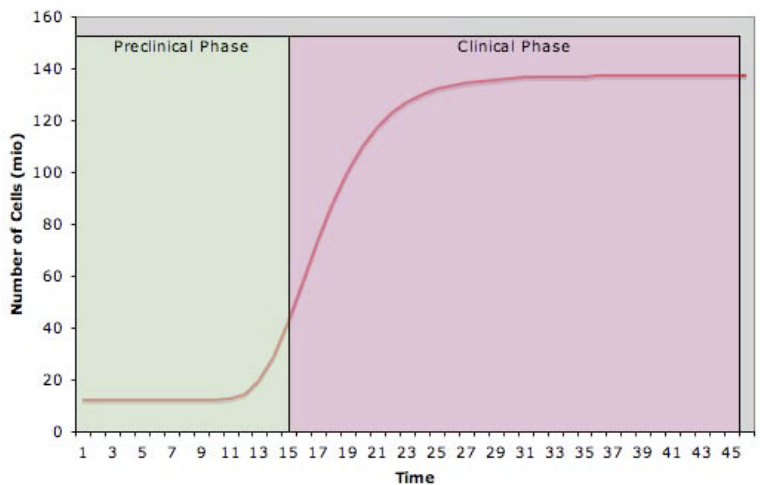
where  $Y$  represents growth, e.g. tumour volume or radius;  $A$  is the initial value, i.e. the lower asymptote;  $C$  is the maximum asymptote;  $B$  is the growth rate;  $t$  is time;  $M$  is the time of maximum growth rate;  $A$ ,  $C$ ,  $B$  and  $M$  are all constants.

The Gompertz curve is of course a gross simplification of population increase but is used ubiquitously across many fields, from tumour biology to population studies, ecology and psychology.

The literature suggests that this function can be used successfully to describe tumour growth [11, 159, 108]. The Gompertz function yields a curve that exhibits an initial stage of exponential growth, which eventually levels off into a plateau (Fig. 2.2). This type of curve suggests an intrinsic increase in growth retardation imposed on a growing number of individuals in the population until equilibrium such that growth and retardation, due to death or stasis, is reached.

However, the complexities of the tumour system is such that a model needs to be able to capture the rich and highly connected functional modules that the system exhibits, e.g. particular processes orchestrated by cellular pathways, to truly have the capacity to qualitatively and quantitatively mimic the system or command any clinical meaning. Hill was one of the first to realise this and investigated diffusion of molecules through tissue and its responsibility for tumour behaviour, hence creating

<sup>4</sup> <http://www.bjhcim.co.uk/news/2010/n1003020.htm>



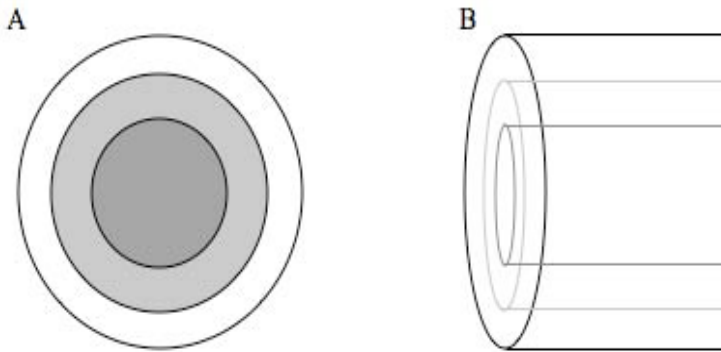
**Fig. 2.2** A generalised Gompertz curve showing the population increase in relation to time. The graph also shows an estimate of the size at which the tumour is detectable, i.e. a tumour is said to be in the clinical phase when it has actually been detected and is being treated.

one of the first algorithms for growth that integrated tumour biology into mathematics. The following decades saw the emergence of growth models that focused on the actual growth dynamics rather than deeper, analytical algorithms [11].

However not many of these algorithms at the time could actually be validated against the real biological system due to the lack of an experimental tumour system. This changed with the introduction of *in vitro* multicellular tumour spheroids.

### 2.1.1.2 Multicellular Tumour Spheroids and Tumour Cords

*In vitro* multicellular tumour spheroids (MTS) were introduced by a number of groups, the most cited being [176], as model experimental systems and a real alternative to the somewhat limited monolayer techniques for the exploration of tumour behaviour [23]. They have been successfully applied to many areas of the field including therapy resistance, drug penetration, invasion and tumour cell metabolism [108,59]. The familiar morphology of a necrotic centre surrounded by a layer of hypoxic tissue, which in turn is surrounded by a layer of normoxic tissue is observed in these *in vitro* model systems (Figure 2.3). Since there is no vasculature and therefore no additional nutrient supply (apart from that which is available from the surface) the tumour only reaches some small maximum radius, which is termed diffusion-limited growth [33]. In the wake of the accelerated use of this experimental system a substantial number of mathematical models for tumour growth were developed, which accurately describe avascular growth retardation, especially that of spheroid growth.



**Fig. 2.3** A. Idealised representation of a MTS; normoxic cells at the periphery, hypoxic layer and necrotic core (inner). Maximum radius 1-3mm [120]. B. Idealised representation of tumour cord: vascular centre (outer) surrounded by normoxic layer, hypoxic layer and necrotic layer (inner). Maximum radius, including vessel, is approximately 60-140 micrometres [158].

However, there are some critical issues that must be addressed before accepting that MTS effectively reproduce *in vivo* processes. The first objection that can be raised is the fact that many primary tumours are rarely discovered before neovascularisation has occurred and therefore the clinical significance of a spheroid mathematical model is automatically made immaterial [76]. Once a tumour begins to exhibit neovascularisation the entire morphology and behaviour of the tumour system changes. The supplementation of nutrients from blood vessels enables cells that were otherwise quiescent to re-establish their mitotic paths [120] unless other local inhibitory factors exist, e.g. cellular overcrowding [44].

It is also apparent that all too often the models are again simplified even more by assuming the boundaries between the morphological layers are crisp, discrete and symmetrical [161]. The genetic hetero/homogeneity of the MTS has also been brought into question [57], i.e. the symmetry of MTS would seem to suggest genetic homogeneity in the cell population whereas it has been shown that this is rarely the case in tumours. However most experts agree that the microenvironment itself, given the context, accurately mimics the *in vivo* system although some biological complexity is lost, e.g. MTS generally exhibit stable chromosome number [59,108].

These limitations considered, one must wonder how much analytical insight, pertinent to clinical phase tumours, models of avascular growth can actually offer. Even so the vast majority of growth models in the literature are spheroid models [11]. At the very least, the formulation and analysis of these models, and subsequent refinement, can form a basis for discovery of the factors involved in early tumour growth. It can therefore be seen as a valuable starting point in itself, even in the light of the known shortcomings of this approach.

As can be seen from Figure 2.3, the tumour cord is the morphological opposite of the MTS, having the nutrient supply in the middle of the tumour and the layers of cell state types (normoxic, hypoxic, necrotic) surrounding it. This configuration can

be thought of as a micro-system of an *in vivo* vascularised tumour. The vascularised tumour in turn can be thought of as a collection of cords. Vascular-driven growth is discussed shortly.

### 2.1.1.3 Avascular Growth Modelling Formalisms

Assumptions for preliminary modelling of avascular tumour growth almost invariably include a genetically homogenous cell population for each layer, spatial homogeneity and a whole-structure radial symmetry [33,73]. Many of the models in the literature stop short of vascularisation but there are some groups who have extended their avascular models into the vascular stage.

It has already been stated that the Gompertz equation forms the basis of many models, and indeed also forms the validation of many models, which will be shown later. The Gompertz function itself, however, is of course purely phenomenological and does not offer any insight to the mechanisms that actually contribute to growth. Chignola [44] take the formalism one step further by formulating a stochastic Gompertz-like growth equation that takes into account the variability in growth dynamics of a heterogeneous population. Even so the model does not give real insight into molecular or cellular mechanisms nor does it shed any light on probably morphology.

**Definition 2 (Phenomenological and Mechanistic Models).** Phenomenological models provide relatively simple constructs that can mimic a system by means of considering (usually macroscopic) observables only. Conversely mechanistic models in this book are defined as those that take into account finer-grained behaviours and components of the system.

Ordinary differential equations (ODE) and partial differential equations (PDE) are by far the most popular mathematical techniques to describe avascular, and indeed vascular, tumour growth, as well as systems biology<sup>5</sup> in general [11,33,73]. The most basic equations define tumour growth in terms of available nutrients and/or oxygen supply. Therefore Friedman (2004) states concentration gradients to be described as a PDE due to Ficks 2nd Law:

$$\varepsilon_0 \frac{\partial c}{\partial t} = \nabla^2 c - \lambda c, \lambda > 0 \quad (2.5)$$

$$\varepsilon_0 = \frac{T_{diffusion}}{T_{growth}} \quad (2.6)$$

where:  $\varepsilon_0$  is a small positive coefficient (e.g. 1 minute/1 day);  $T_{diffusion}$  is the diffusion time-scale;  $T_{Growth}$  is the tumour growth time-scale;  $c$  is concentration;  $\lambda$  is

---

<sup>5</sup> The field of systems biology is discussed in more detail in the next chapter. For now, systems biology can be defined as a field of study that encompasses both system complexity and biology.

a molecular, e.g. nutrient, degradation/uptake coefficient (or rate);  $\nabla$  is the gradient operator (Laplacian) given by the equation below:

$$c \cdot \left( \frac{\partial}{\partial x} + \frac{\partial}{\partial y} + \frac{\partial}{\partial z} \right) \quad (2.7)$$

where  $x, y, z$  are Cartesian directions.

Dasu (2003) [53] make extensive use of diffusion equations to model oxygen gradients, incorporating static vascular structure (fixed nutrient source), and subsequent effects on tumour growth.

Byrne (1999) [33] describes basic avascular (symmetrical) tumour growth in the form of an ODE:

$$\frac{1}{3} \frac{d}{dt} (R^3) = R^2 \frac{dR}{dt} = \int_0^R S(c)H(r - R_n)r^2 dr - \int_0^R N(c)H(R_n - r)r^2 dr \quad (2.8)$$

where  $\frac{1}{3} \frac{d}{dt} (R^3)$  is rate of change of tumour volume over time;  $\int_0^R S(c)H(r - R_n)r^2 dr$  is net rate of cell proliferation ( $S(c)$ );  $\int_0^R N(c)H(R_n - r)r^2 dr$  is rate of necrotic cell death ( $N(c)$ );  $R_N$  is the necrotic radius;  $R/r$  is overall radius and hypoxic radius respectively;  $H$  is the Heaviside step function.

Byrne goes on to describe how one can extend Equation 2.8 into further models that account for asymmetry by vascularisation and multiple cell populations. A more heterogeneous approach can be taken by reapplying the above equations with different coefficient terms to reflect different growth characteristics of multiple cells types. Friedman (2004) [73] also shows how these equations can be extended to multiple cell populations but simultaneously recognises that mixed population models, where different cell-type populations are continuously present everywhere in the tumour, are more realistic than the normal segregated models where cell-type populations are assumed to have discrete boundaries. Sherrat (2001) [161] tackle the problem of discrete interfaces of cell populations by incorporating ODE models that describe densities of cell populations with cellular movement along pressure gradients.

The differential equation approaches described above suffer from some fundamental limitations. It is generally agreed that differential equations by themselves do not capture enough resolution in terms of spatiotemporal behaviour of the system. At the very least, to get more information out of the models a great number of equations must be generated and solved [174], which might prove impossible in some applications, as some physical data needed for parameterisation might not be available. Modelling 3D heterogeneity becomes a computationally challenging task. Another problem, which is apparent from the equations stated above, is that the use of differential equations automatically assumes that the current state of the system is a consequence of the previous global state of the system. In reality, a single cell will behave according to its immediate locality, not according to the state of the tumour as a whole, and local environments are differing. Indeed, as Gatenby (2003) [76] point out, too often we are content with work that is entirely phenomenological - curve-fitting data - without developing mechanistic models that provide real insights

into the critical parameters that control system dynamics. However there are indications of a paradigm shift in the literature with regard to mathematical modelling of both avascular and vascular tumours.

The literature appears to be drifting away from the traditional compartmental, ODE and PDE approaches to mathematically simpler methods such as cellular automata (CA) and agent-based modelling (ABM). This could be due to the increase in access and popularity of powerful computational techniques and software as well as the surge of interest in biological systems from the computer science community.

#### 2.1.1.4 Non-traditional Avascular Growth Models

The use of CA<sup>6</sup> models has become more prominent in recent years. One of the earliest CA models for tumour growth was introduced by Duchting (1981) [68]. Qi (1993) [149] took its use further, formulating a stochastic two-dimensional CA that took into account proliferation, nutrient supply, mechanical pressures (and thus motion of cells within the tumour) and immune surveillance rules. The model quite plainly shows how simple rule definitions in CA can yield complex behaviour and, in this case, qualitatively accurate simulation results. The discrete grid cells react according to a neighbourhood of four other discrete cells (von Neumann neighbourhood) and can be occupied simultaneously by cancer/normal cells and macrophage cells. The discrete cells of the CA therefore represent populations of biological cells, or densities. Macrophages can engulf cancer cells (if they are coexisting), which is defined by a fixed probability, and/or cancer cells can divide. Cancer cells are also allowed to invade neighbouring discrete cells occupied by normal tissue. Both proliferation and invasion have a space restriction and global growth, i.e. overall radius, is defined as a function of the internal pressure of the tumour (which itself is a function of cellular densities and proliferation). These simple rules are formalised into the CA and it was found that the resulting growth followed a Gompertz-like growth law. Scalerandi and co-workers describe a discretised two-dimensional approach to modelling cord growth dynamics considering competition for nutrients and conservation of energy [157,158]. They highlight the use of locality-interaction modelling where cells react to their local environment and consequently demonstrate good agreement with experimental data. This is in sharp contrast with the phenomenological models mentioned earlier where the focus is on the global state of the system. The local interaction approach has the added benefit of incorporating heterogeneity in a way that is clearly not possible with phenomenological or differential equation based techniques.

Stamatikos [168,169], provide a prime example of how one can model a tumour without using the complex mathematics of differential equations but at the same time capture rich spatiotemporal qualities of avascular tumour growth. The structure of the simulated tumour bears remarkable resemblance to the spheroid structure. Simple rules and phase transitions of cellular cytokinetics were used to simulate

---

<sup>6</sup> For an online demonstration of the use of CA in growth models, see <http://demonstrations.wolfram.com/TumorGrowthModel/>

avascular tumour growth. A discrete space/time formalisation was employed with a rule base very much akin to the CA formalism. The virtual cells are made to run simplified cell cycle phases, which is reminiscent of the multiphase approach adopted by Please (1998) [147]. More recently the same group developed integration between an improved version of the model, including the oxygen enhancement ratio (OER) and linear quadratic (LQ) models (Equations 2.10, 2.11, 2.12) in the rules for cells, and an algorithm that relates neovascularisation information from brain scan images of gliomas [9]. Dionysiou (2006) [60] further incorporate p53 status to the radiosensitivity model thereby yielding a simulation that can predict the effects of different radiotherapy regimens. These simulations have led the group to wider and more representative glioma via grid technology [109].

It is important to realise the immediate difference between this approach and the differential equation approach as described earlier - in the CA-like formalism the local interactions form the basis of the state of the whole system, which is in fact reflective of the real system. Even with gross simplifications in the form of local interaction rules the simulation achieves satisfactory agreement to in vitro small cell lung cancer MTS with relatively richer spatiotemporal dynamics compared with the O/PDEs presented earlier.

Voitkova (1998) [189] describes a stochastic non-Markovian simulation of immune response to small avascular tumour tissue having a rule-base that includes immunity-tumour interactions, cellular random walks, growth and necrosis within a CA framework. All rules are probabilistic, an approach that is known to reduce computational load compared to a purely fine-grained deterministic method and also underlines the fact that the nature of many mechanisms on the smaller scale are not known. Results of this approach suggest that the immune-tumour interaction may have an oscillatory effect on tumour growth, reflective of the fluctuations in probabilities that immune cells will find a cancer cell target and kill it. Interestingly the tumour geometry is observed to be fractal in nature. Voitkova also boldly claims that this simple yet powerful approach exhibits tumour behaviour simulation better than any O/PDE based model.

More recently, Dormann (2002) [65] use a two-dimensional hybrid lattice-gas CA approach that incorporates cell migration as well as the usual nutrient flow, proliferation and cell necrosis. The lattice-gas flavours of CA are frequently used to model fluid dynamics and specifically contain rules that idealise conservation of momentum and flow of particles. In this case the particles are biological cells whose positions are in flux due to pressure and/or chemotactic gradients. Here the chemotaxis is mediated by a hypothetical signal emanating from near-necrotic cells, which attracts other cells towards them - a rule that is meant to account for the experimental observation of normoxic cells moving towards the necrotic core. Dormann aptly highlights the fact that the purpose of such spatiotemporal modelling is that it offers insights into the mechanisms responsible for collective organisational behaviour at the microscopic level, rather than having a purely phenomenological model that offers relatively fewer insights. The model then quite elegantly shows that the familiar ring structure (normoxic to hypoxic to necrotic) can be recreated relatively easily

and therefore proves the point that local interaction-based modelling provides a behaviourally rich and realistic simulation.

Patel (2001) [140], using a similar CA hybrid approach, focus on a different target in the microenvironment for tumour growth - that of acidity levels and its effects on invasion. The hybrid model incorporates the local interaction-centric advantages of a CA whilst at the same time utilising the power of differential equations solved on the discrete space/time grid to describe the diffusion of nutrients and H<sup>+</sup> ions in the interstitial space. The difference to other work is the fact that it is one of the very few that actually incorporate the presence of vasculature from which nutrients can reach tumour tissue, albeit randomly distributed in the lattice. Therefore, although compared to and described as early tumour MTS-like growth, strictly speaking this model actually describes vascular growth. The results show the power of the approach - the model showed that an increase in H<sup>+</sup> ions promotes tumour growth and invasion; leading to the prediction that the introduction of vascular density above a certain threshold therefore may actually decrease the growth rate since the blood flow reduces the concentration of acidity in the interstitial space.

The T-7 group of Los Alamos National Laboratory<sup>7</sup> (LANL), in particular Jiang, Pjesivac, Freyer and other co-workers, have also taken up these newer mathematical techniques to model MTS dynamics. Like Patel (2001) they have used a hybridised approach combining the powers of both CA and PDEs. Their model incorporates the effects of mitosis, mutation, necrosis, nutrient uptake and metabolic waste whilst being able to follow the fate of individual cells, which is not possible with differential equations. Chemical reaction-diffusion dynamics is governed primarily by three PDEs. With this approach the group preliminarily report a strong correlation with experimental data with respect to growth dynamics.

### 2.1.1.5 Other Growth Related Models

Other models for growth mostly consist of terms for metastasis, which is inherently tied in with angiogenesis (next section) and cellular adhesion. Araujo (2004) [11] review recent efforts in great detail, which will not be repeated here. A clear point of interest is that there is a surge in use of computational techniques such as CA as well as ABM, e.g. the works of dos Reis (2001) [150] who employ an agent-based approach to simulate the effects of cellular adhesion on tumour morphology. In summary, local interaction-based models conclusively exhibit more faithful behaviour when compared to the real system compared to traditional techniques.

However, a major issue in using agent-based simulations to study complex systems such as tumours is the fact that simulations are opaque. Methods for identifying the interactive mechanisms between agents giving rise to global system behaviours have been lacking. Recently, Chen and coworkers at University College London have developed a novel formal modelling language for ABM, complex event types (CETs), and novel computational methods which allow statistical dependencies between behaviours at different levels to be established [40, 41, 42]. This is providing

---

<sup>7</sup> <http://math.lanl.gov/>



a computational method for distinguishing between alternative interactive mechanisms underlying a higher level behaviour. This method can, for example, be applied to simulations of tumour initiation and growth using agent-based models of abnormal cell populations. Competition and selection events can be identified in simulation at both the individual and clonal level, allowing us to computationally test hypotheses about the interactive mechanisms underlying a neoplastic clones success.

### 2.1.2 *Angiogenesis Models*

Vascularisation of tumours is the most important event preceding malignancy, providing the pipeline for metastases, and is arguably one of the most important factors that determine the overall behaviour of a clinically aggressive tumour [120]. The fact that tumours can induce nearby capillaries to undergo neovascularisation has provided the impetus for research into molecular targets for therapy [84]. The modelling of vascularisation is one of the most complex problems in tumour modelling because of the multiple scales at which all critical events occur. Ranging from molecular interactions to gradients and diffusion through media of differing viscosity, branching and meeting of capillaries and overall tumour dynamics, as well as the complexities of blood flow itself, angiogenesis models must incorporate an exceptional level of spatiotemporal dynamics. Angiogenesis is distinct from the context of the tumour and can be classed as a complex system in its own right [113].

#### 2.1.2.1 **Basic Physiology of Angiogenesis**

In angiogenesis blood vessels are formed from the scaffold vessel structure resultant from a previous vasculogenesis process. The basic steps of tumour-induced angiogenesis are as follows [120]:

1. Hypoxic cells, under the strain of competition for nutrients, produce and emit growth factors generally known as Tumour Angiogenic Factors (TAFs), e.g. VEGF. TAFs can either be angiogenic (e.g. VEGF) or anti-angiogenic (e.g.  $\alpha$ -interferon, endostatin) and sometimes both depending on the local state.
2. The endothelial cells (EC) of existing capillaries, which have specific receptors for TAFs, receive the signal that diffuses through the extracellular matrix (ECM). This sets off a chain of intracellular events that eventually leads to cell proliferation. The EC then exhibits an activated mesenchymal phenotype (rather than its normal quiescent phenotype).
3. The ECs begin to excrete matrix metalloproteases (MMP), which break down the basement membrane, made of pericytes, and the fibres in the surrounding ECM.
4. As the ECM is broken down it is the leading EC, chemotactically migrating up the signal gradient, that guides a following posterior mass of proliferating ECs. The migration is also influenced by gradients of cellular adhesiveness mediated mostly by fibronectins and integrins. It should be noted that proliferation of ECs

only begins to take place 36-48 hours after the initial EC response - before this point ECs are actually recruited from the parent vessel [143].

5. Capillary tubes form once ECs begin to re-associate and align. Intracellular vacuoles extend to the poles, forming a hollow tube.
6. The immature vessels are said to become mature when pericytes envelope the newly formed structure. Maturation is a critical step in vascularisation [10].

Once the above stages of vascularisation have occurred anastomosis<sup>8</sup>, the process by which independent branches join, takes place before complete maturation [120]. Whilst it is easy to think of this process as diffusion-driven and that anastomosis occurs as a result of following simple diffusion gradients, angiogenesis is still an extremely difficult subsystem to model - the exact processes and mechanisms are still not fully understood. This is illustrated, for example, by the fact that whilst independent branches can loop they can also unexpectedly split and the reasons for this are not known at present [7].

Angiogenesis-related molecular species form the basis of most angiogenesis modelling implementations. In tumour-induced angiogenesis the process becomes somewhat anomalous because of the intensity and volatility of extracellular signals resulting from extensive mutations, aneuploidy and hypoxic stress of tumour cells [53].

One of the most well-known TAFs is the VEGF isoform family of extracellular proteins. There exists a corresponding family of receptors (most commonly mentioned is the VEGFR-1 receptor) that are expressed on the membranes of endothelial cells. Once VEGF has bound to its receptor the cascade is triggered, resulting in the preliminary stages of angiogenesis. This not only prepares the cells for angiogenesis but also increases the permeability of the vessel, which is in agreement with experimental data suggesting that tumour-associated vessels are extremely leaky [139, 106]. Angiopoietins are another class of extracellular signals involved in angiogenesis. It is worth mentioning that there is a severe under-representation in the literature with respect to models that incorporate the action of these proteins. Angiopoietins bind to the Tie family of receptors, which can be found on the EC membrane surface, and mediate sprout formation and branching [139]. Other major diffusible molecular factors include TGF- $\beta$ , FGF, EGF, the ephrins, PDGF, angiogenin, angiostatin, angiotropin, the interleukins, TIMP and interferon families and TNF- $\alpha$ .

The ECM is composed mainly of interstitial tissue, collagen, fibronectin, vitronectin, fibrin, von Willebrand factor and other structurally important molecules and plays a critical role in the leading ECs migration [7]. Fibronectin remains static in the ECM, although gradients of fibronectin exist, and plays an important role in cellular adhesion by binding to the integrin family of cell surface receptors. Cadherins are particularly important cell surface molecules, mediating cell-cell interactions. It is therefore clear that though an important role is played by chemotaxis (i.e. VEGF, etc.) there is also the haptotactic response to take under consideration, which many models ignore. The frictional forces that the ECM imposes, due to dense

<sup>8</sup> Synonymous with anastomosis in the literature.

content (i.e. collagen, fibronectin etc.), have a massive impact on the path the ECs take when moving up attractive concentration gradients [120].

Finally, the mechanical forces imposed by blood flow itself, which can have a significant influence on vessel formation and maturation, are another important factor in neovascularisation [139]. Stress in the vessels can induce transcription of PDGF and TGF- $\beta$ , which affect matrix consistency and vessel formation respectively [151]. Tumour blood flow models are relatively scarce in the literature and most of the research is conducted in light of existing models of hydrodynamics, a well-developed domain in the field of engineering. When applied to blood flow this is termed haemodynamics. However, the majority of the literature in this case is still applied to cardiac physiology-related haemodynamics rather than tumour-related haemodynamics. This type of modelling utilises complex PDEs and Lattice-Boltzmann flavours of CA.

The brief summary of key molecular, cellular and tissue-level events in angiogenesis presented here highlights the complexity of the process. Furthermore, although many of the molecular species involved have been identified it is generally accepted that still very little is known about the biochemical interactions that are taking place (both intra and intercellular) to satisfactorily explain the rich behaviour of angiogenic tumour systems [120]. This restricts the modelling of angiogenesis to being merely qualitative rather than quantitative.

### 2.1.2.2 Modelling Methodologies

The vast majority of evolved mathematical models that incorporate angiogenesis terms do so by characterising diffusion gradients of TAFs and subsequent responses of nearby capillaries. Spatiotemporal resolution is of course very important in these models and this is exactly what is now emerging from the modelling community, although earlier models have still lagged with the traditional differential equation approaches. The angiogenesis modelling literature can be loosely categorised into the stages on which they focus, though the stages are of course overlapping. Many modelling efforts have gone into focussing on the location and time at which ECs begin to respond to TAF signals. Other efforts have focussed on EC migration, the dynamic structure of vasculature, blood flow and turbulence.

Mantzaris (2004) [120] further define three basic categories of models for angiogenesis that can be found in the current literature:

1. *Continuum models*: Cellular populations and molecular species are treated as continuous variables. Cells are usually described in terms of densities. This approach invariably involves the use of an O/PDE approach.
2. *Mechanochemical models*: This type of model takes into account the gradients of chemical species in the ECM (e.g. fibronectin) and the migration of cells through the ECM.
3. *Discrete models*: These have appeared more recently, and like their growth model counterparts, include CA-related approaches. Typically this approach can be described as a discrete realisation of a continuum model, as will be portrayed by

the works of Anderson (1998) [7]. In ABM approaches individual cell fates can be followed.

To summarise, the generalised equations given by Mantzaris (2004) [120], which will not be repeated here, are mostly O/PDE models that describe (i) the density of cells in one to three dimensions and (ii) the concentration gradient from source to capillary of TAF(s) in one to three physical dimensions with appropriate boundary conditions. Diffusion of particles can be modelled as random walks, multiset rewrites [82] or simple Ficks-like PDEs.

More recently, as with growth modelling, hybrid CA and hybrids thereof have become increasingly popular. This should come as no surprise as the CA approach has proved to be extremely apt at capturing not only rich spatiotemporal behaviour, which is imperative when considering vessel structure, but also heterogeneity. The work of Patel described earlier illustrates the power of CA in vascularised tumour growth. Note here, however, that angiogenesis models do not pertain to mere growth models that incorporate vascular presence, as in Patel and co-workers model. Rather, angiogenesis models include vascular dynamics - i.e. sprout formation, extension and maturation.

Anderson (1998) [7] demonstrate quite elegantly the strengths and weaknesses of both continuous and discrete modelling methodologies, though this was not the specific goal, by first creating a set of differential equations to describe chemotaxis as well as haptotaxis and EC density dynamics. The model is then solved using an extension of the conventional finite difference method in a discrete 2-dimensional CA setting to gain better spatiotemporal insight. The authors highlight the fact that very few models successfully capture rich enough behaviour to be able to predict complex observable phenomena such as the brush-border effect (the increased level of neo-vessel branching when it approaches the TAF source). Moreover the model also incorporates random walks of leading ECs in order to mimic sprout extension (branching, anastomosis and cell proliferation). Once the continuum model was defined the relative effects of haptotaxis and chemotaxis were simulated in two dimensions. It was shown that concentration gradient geometry and magnitude is important, stressing the fact that assuming symmetrical dispersion of molecular signals in a traditional source-sink methodology does not sufficiently model reality. Additionally the simulations show the crucial role haptotaxis, i.e. the interplay of fibronectin, laminin, and other macromolecules, plays in angiogenesis, which is quite a unique development in the angiogenesis modelling literature. The model can be readily converted into a 3D setting, as demonstrated by Stephanou (2005) [170]. The 3D model is shown to be quite different both in terms of actual vasculature dynamics and drug delivery, highlighting the fact that any serious attempt at a tumour-level simulation must be performed in a 3D setting. The group reports that the most striking result obtained is that, when particular vascular structures are formed, only less than 3% of the drug actually reaches the tumour - the rest bypasses it. If correct, this of course bears important implications for therapy.

The discrete model that Anderson and Chaplain [7] then formulate incorporates biased random walk EC behaviour under chemotactic and haptotactic conditions. Since there is a lack of data as to what the anastomotic causal mechanisms actually

are, the authors incorporate an age (of sprout) and space constraint rule into the CA to model branching and looping. The individual cells in the CA represent approximately  $10\mu\text{m}$ , or 1-2 ECs, collated into a 200 by 200 grid. The resulting model is shown to enable both qualitative and quantitative comparisons with *in vivo* experiments. As with the continuum model, the effects of haptotaxis and chemotaxis were simulated and again showed that haptotaxis is likely to play a crucial role in neovascularisation - the simulations suggested that without it no anastomosis could be established and only minimal branching could occur. Anderson and co-workers recognise the importance of granular (i.e. mechanistic) approaches to tumour modelling over and above traditional methods as superior models of the system, illustrated in a more recent paper [8].

The CA approach is found to be far simpler to manipulate, especially in the absence of precise parameterisation data compared to the PDE approach. Of course, individual-based modelling techniques also have the distinct advantage of enabling the modeller to track the behaviour of cells individually, which is of enormous value in the context of angiogenesis.

The random-walk approach was also adopted by Levine (2001) [113] who also modelled tumour-induced angiogenesis in terms of haptotaxis and chemotaxis, although no formal comparison with *in vivo* angiogenesis is apparent. They also incorporate biochemical enzyme kinetics (receptor-ligand binding, etc.), making a rich model that can be interrogated right down to the molecular dynamics scale - given enough computational power this type of cross-scale modelling is possible without losing the rich fine-grained behaviours of the system, which may well be the case for differential equation DE approaches.

The work described above is a good example of how the tumour system has been dissected mathematically into subsystems, e.g. sprout formation and elongation, and studied in a closed fashion. In other words, by investigating the effects of chemo/haptotaxis on sprout fate for example, the focus is purely on angiogenesis, and other parts of the system are ignored, such as growth dynamics of the tumour itself. One must remember that the process of angiogenesis is repeated again and again and can take days to complete. During this time hypoxic regions could be growing or shrinking and therefore TAF signalling would be in flux. However, since the overall growth of the tumour is assumed to be static the modelling of angiogenesis itself has been restricted to prediction of location of sprout formation and subsequent vessel formation with steady TAF diffusion gradients. Moreover, important additional factors, such as the maturation stage of the new vessels would be omitted in closed sub-system modelling. In immature vessels blood flow is rather poor [139]. If perfusion is extremely poor or the levels of VEGF and other TAFs falls in the vicinity of immature vessels (IV) then the vessel can regress by, for example, mediation of Ang-2 [88].

Arakelyan (2002) [10] recognise the fact that IV and mature vessel (MV) structure stabilisations are critical in modelling angiogenesis since vasculature has been observed to be a dynamic rather than static structure. This is in sharp contrast to previous models where, once a leading EC has set a path for a new vessel, the vessel remains static and is assumed to be instantaneously mature and perfused. This

assumption is made in both continuum models (in the form of densities of vessels) and discrete models. Arakelyan and co-workers therefore set about producing a discrete model to describe not only the angiogenesis processes of vessel formation and maturation but also the simultaneous fluctuation of tumour growth and TAF signalling (no account is taken for the influence of ECM macromolecules). However it must be noted that even though the model incorporates vessel destabilisation it makes the simplification that IVs and MVs actually have the same tissue perfusion efficiency, which is known not to be true, and therefore enumerate both types of vasculature generically as effective vessel density. The algorithm employed by Arakelyan includes a fusion of abstract parameterisations and an extensive Boolean network. A Boolean network is simply a graphical model incorporating yes/no decision-making arcs and process nodes and is therefore essentially a Heaviside-step threshold-based system. The algorithm is described as multi-scale since it covers three organisational levels - molecular, cellular and organ-level. The model was tested by performing *in silico* simulation of functional knockout experiments, which would be very difficult to perform experimentally. For example, it was shown that in the absence of vessel destabilisation and maturation tumour and vessel volume follow exponential growth laws. Subsequent incorporation of these behaviours results in fluctuations in IV and MV volumes, and was shown to be in marked agreement with experimental data (although the agreement is qualitative rather than quantitative).

In summary angiogenesis models illustrate the power of individual-based models over traditional techniques. The geometry of vessel growth and its importance to growth is very well suited to individual-based models. Furthermore angiogenesis illustrates the complexity of the tumour modelling problem, consisting of a large number of stimulants and inhibitors with complex interactions over large time and space scales.

### ***2.1.3 Treatment Response Models***

Response behaviour is the ultimate goal of tumour modelling since it is in the successful simulation and prediction of critical parameters of this behaviour that will finally lead to clinical improvement. Response models, as a matter of necessity, must include growth aspects of tumours and, preferably, vascularisation aspects. Angiogenesis-response model hybrids have become more prominent in the literature only recently with the arrival of anti-angiogenic drugs. For example, Ledzewicz (2010) and co-workers demonstrate the use of optimisation and Gompertzian growth methods, including pharmakenergetics, to determine optimal dosage for anti-angiogenic therapy [111]. The introduction of pathway modelling (i.e. the pharmakenergetics) in this manner is slowly becoming more prevalent only very recently with respect to treatment response - however one can expect that the inclusion of pathway dynamics in response models will become more popular with the continuing explosion of knowledge of cellular signalling and the concept of targeted therapy.

Two kinds of tumour response models can be observed in the literature:

1. Mechanistic-based response models:
  - a. Chemotherapy (and subsequent immune-) response to tumours and use of therapies to boost immune response. This includes radio-immunotherapy.
  - b. Tumour response to radiation therapy.
  - c. Dynamic pathway response to the two therapies listed above.
2. Therapy optimisation: These models focus on exactly how and when therapy should be administered rather than focusing specifically on tumour dynamics.

The same formalisms that were discussed for tumour growth apply here - the tumour volume, density, number of cells and/or tumour radius are the primary outputs of response models although some models also go as far as including metastases. Also, the same temporal pattern of use of formalisms appears in the literature, i.e. first use of DE approaches for avascular tumours and later use of more descriptive computational approaches, such as CA and ABM.

### 2.1.3.1 Generalised Formalisms

A basic differential equation approach can be expressed as follows<sup>9</sup>:

$$\frac{dn}{dt} = pn_t - k_t n_t \quad (2.9)$$

where  $n_t$  is the number of cells at time  $t$ ;  $p$  is the proliferation rate of the tumour;  $k_t$  is the therapy effectiveness at time  $t$  and so  $k_t n_t$  denotes kill rate.

The variable  $p$  can be generally determined for tumour cell types and  $k_t$  is drug-dependant. This model is phenomenological and includes no spatial dynamics. However, it forms the basis of most response models that use the DE approach.

Equation 2.10 can be used to estimate another term that is extensively used in modelling - the so-called Tumour Control Probability<sup>10</sup> (TCP), which is used mostly in radiotherapy models:

$$TCP = e^{-\sum_{i=1}^Q N_i S_i D} \quad (2.10)$$

where  $TCP$  is the probability that no tumour cells survive (tumour cure probability);  $i$  and  $Q$  denote the total number of cells in the considered 3D space  $i$ ;  $N_i$  is the initial number of cells;  $S_i(D)$  is the surviving fraction of cells after dose  $D$  in region  $i$ .

The LQ model is another general formalism for radiotherapy and describes the survival probability of a cell. It reads as follows:

$$S(d) = e^{-(\alpha d + \beta d^2)} \quad (2.11)$$

<sup>9</sup> <http://www.ap.univie.ac.at/users/ves/cp0102/dx/node126.html>

<sup>10</sup> For an in-depth discussion on TCP, its applications and extensions the reader is directed to Zaider (2000) [204].

where:  $S(d)$  is the survival probability of the cell given dose  $d$ ;  $\alpha$  and  $\beta$  are parameters specific to the cell/tumour and reflect intrinsic radiobiological properties pertaining to sub/lethal DNA damage due to radioactivity;  $d$  is the dose (usually in Gy, where 1 Gy = 1 Jkg<sup>-1</sup>).

The above equation can additionally be extended to account for the sensitivity of hypoxic tissue by taking into account the OER:

$$S(d) = e^{\left[ -\frac{\alpha d}{OER} + \frac{\beta d^2}{OER^2} \right]} \quad (2.12)$$

### 2.1.4 Modelling Methodologies

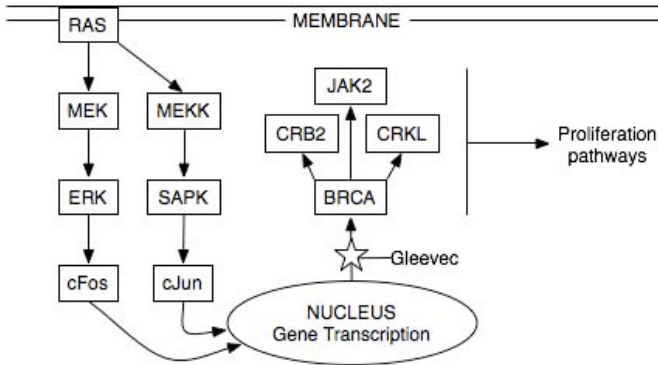
Immune response and therapy models are abundant in the literature (see [11]). The works of de Pillis and Radunskaya (2001, 2003) demonstrate the state-of-the-art in this kind of modelling using a combination of traditional approaches with optimal control theory to develop an avascular model (homogenous, MTS-like) that can qualitatively predict *in vivo* tumour dynamics. The model recognises three types of interaction that have previously been modelled separately: (i) interaction between tumour cells and immune system cells, (ii) interaction between normal cells and tumour cells and, (iii) interaction between tumour cells and chemotherapeutic agents. The model therefore effectively idealises competition between immune system cells and tumour cells (and their response to therapy) and the corresponding immune response to the presence of a tumour. Differential equations are formulated that determine the number of normal, tumour and immune cells and the cell kill relation for a therapeutic agent. The resulting model is reported to capture two interesting behavioural characteristics that are observed *in vivo*. Firstly, so-called Jeffs Phenomenon is observed - an asynchronous oscillation of tumour growth with respect to chemotherapy dose time. Secondly, and perhaps clinically more interesting, is the model's ability to predict tumour dormancy - a phenomenon that occurs when the tumour shrinks to an undetectable size only to re-emerge to grow exponentially again. Both of these behaviours can be explained by the interactions of the tumour and immune cells. This example alone goes to show the rich and qualitative (and eventually, in the future, quantitative) dynamics that mathematical modelling can actually generate which serves to better current understanding. The goal is of course to reduce tumour cell counts whilst keeping normal cell counts within safe bounds by finding appropriate times to administer certain amounts of dose in a given time interval. Optimisation and control theory is applied to achieve this and is extensively tested in more recent papers [55, 56]. A drug protocol is thence determined where the tumour cells are eradicated whilst keeping the total normal cell count around the tumour site above 75%, though the oscillations in tumour size observed in the simulation would admittedly not be clinically desirable.

Arciero (2004) [12] demonstrates another interesting theoretical approach. This example is particularly appealing because it models an immunotherapeutic strategy that has not yet been tested extensively *in vivo* - so called small interfering RNA (siRNA) immunotherapy - and so this scenario depicts a unique opportunity for



mathematical modelling to prove itself as a viable technology with which one can satisfactorily predict tumour dynamics. The siRNA molecules are only around 22 nucleotides long and interfere with certain transcripts such as that of TGF- $\beta$ , thereby blocking gene function. TGF- $\beta$  is involved in masking the tumour from immunosurveillance, it is also strongly angiogenic and relatively well understood, making it a good target. The proposed model uses a system of only five ODEs that describe immune cell numbers, tumour cell numbers (incorporating the phenomenological logistic function), IL-2 dynamics and the effects of TGF- $\beta$  on tumour growth and immunosurveillance and subsequent effects of siRNA that blocks TGF- $\beta$  and degradation of the siRNA. Administration of treatment is found to have oscillatory effects on tumour growth dynamics, just as de Pillis and Radunskaya described earlier.

This touches upon another active area of research in cancer systems biology - that of the effects of drugs on signalling pathways. Surprisingly the number of models that actually incorporate pathway dynamics directly with drug action is quite limited (i.e. both aspects together - the literature in terms of the individual aspects are actually substantial). These types of models will be discussed in more detail later, however to make a case in point the work of Charusanti (2004) [38] is briefly explained. The model proposed in this case demonstrates the power of simulations and makes use of traditional ODE approaches to predict the effects of STI-571 (Gleevec) on the Crkl pathway (Figure 2.4) in Chronic Myeloid Leukaemia (CML).



**Fig. 2.4** Gleevec action on the Bcr-abl oncogene. CML is probably the only type of cancer that is associated almost exclusively with a single type of mutation, i.e. 9-22 chromosome translocation (the Philadelphia Chromosome). Deregulated phosphorylation, due to this mutation and caused by the over-active Bcr-Abl gene, causes certain signalling pathways to be constitutively switched on (e.g. proliferation pathways, not shown) and others to be switched off (e.g. apoptosis, bottom right of figure).

The ODEs consider the concentration differential of certain populations of molecular species over time and the Gleevec pharmacokinetics (compartmental model of intraperitoneal cavity, blood and tumour). Even though no specific attention is paid to the spatial arrangement of tumour cells and normal cells, the model is considered

satisfactory since cancers originating from the haematopoietic pool in the marrow, like CML, are only quasi-solid. ODE approaches in this particular situation therefore are appropriate since no spatial dynamics need to be modelled (unless drug delivery itself is the focus)<sup>[11]</sup>. The resulting simulation suggests an important previously unrecognised dynamic of Gleevec treatment, which was later experimentally shown to be a real behaviour - during blast crisis the clearance of Gleevec from cells is very rapid and therefore the effectiveness of the drug is reduced. Here is an excellent example of *in silico* prediction and experimentation - the simulation immediately suggests that mechanisms of clearance must be addressed to enable Gleevec to function with greater efficiency.

Modelling has also established its place in therapy response with respect to angiogenesis<sup>[12]</sup>, and it has already been explained why targeting of angiogenic processes has become an important topic in oncology. Though the focus in Arakelyan's work is not entirely on the effects of drugs on tumour dynamics, by incorporating ODEs for drug action (VEGF and Ang-1 production inhibitors) into the model potentially important qualitative behaviour can be observed - their simulations suggest that, for an aggressive tumour that has relatively low sensitivity to anti-angiogenic drugs, monotherapy is not sufficient to eradicate the tumour. The model showed that only combinatorial therapy successfully killed all tumour cells. Stoll (2003) <sup>[172]</sup> take a slightly different approach by considering the contribution of endothelial progenitor cells (EPC) from the bone marrow to angiogenesis as well as the contribution of ECs in the vicinity of the tumour. Stoll and co-workers establish a set of ODEs that describe vascular densities and flow of EPCs into the region and re-parameterise the equations to account for drug action. Drug actions considered include the effects of therapy that target EPC migration, growth factor and angiogenic factor signalling, chemotherapy by anti-angiogenic scheduling and combined therapies. The model is found to be in agreement with clinical data.

Whilst the works briefly described above take a semi-mechanistic approach to modelling, there are still those models that are almost totally phenomenological such as those presented by Ubezio (2004) <sup>[181]</sup>. In this model cell cycle features are incorporated into a simple compartmental approach and curve fitting to cytometry data is used to parameterise the ODEs. Drug effects are then simply modelled by stochastically altering the behaviour of the compartments, e.g. by giving a cell a probability of being blocked in certain stages of the cell cycle.

In the immune response modelling area CA approaches are less prominent than traditional approaches - this could be because response models often focus on the drug delivery aspect, dosimetry, direct effects on growth and related optimisations as opposed to individual-based understanding, e.g. cell to cell interactions. Here again however it can be argued that delivery and diffusion of drugs through the (dynamic)

---

<sup>11</sup> Spatiotemporal modelling of these types of tumour will be very difficult compared to solid tumour modelling and hence the next step in modelling in the years to come will involve the consideration of quasi-solid tumour dynamics.

<sup>12</sup> A whole host of response-to-angiogenesis targeted therapy models exist but will not be summarised here. The reader is referred to the works of <sup>[11]</sup> for a good review and, for specific models, <sup>[125, 146, 159]</sup>.

vasculature and heterogeneous flow through the ECM all have observable impacts on tumour dynamics and cannot be modelled by traditional methods. Therefore one can expect that CA or ABM models to take up a more prominent role in this area.

McDougall (2002) [125] describe an extension of the model by Anderson and Chaplain by incorporating a perfusion function into the angiogenesis model thereby formulating a 2D simulation for drug delivery in a vascularised tumour. Interestingly the flow model was actually adapted from CA-like petroleum flow models, which simulated the perfusion of petroleum through pore networks in solid rock. Though this formalisation only models perfusion around the tumour rather than through it, some very interesting results were procured pertaining to delivery strategy. Having simulated two chemo-agent introduction schema (one-off bolus injection and continuous infusion) it was predicted that large one-off doses may sometimes lead to dilution in the over-perfused regions and therefore actually never reach the tumour, which obviously would have serious implications for therapy. One must remember, however, that this model does not take into account maturation of vessels, i.e. vessels are assumed to be immediately mature. This is of course not concordant with the real system.

Many other chemotherapy response models exist that focus on different clinical aspects. Pinho (2002) [145], for example, formulate a continuous model where the interactions between normal and cancer cells are considered in the context of primary and secondary tumour sites, i.e. metastases, and interactions with a chemotherapeutic agent. There are many different clinical strategies for optimal therapy and therefore the optimisation problem has become very complex consisting of many sub-problems. For example, there are those models that focus specifically on pulsed therapy as well as drug resistance, which have become increasingly important issues. Lakmeche (2001) [110] develop such a model, focussing on pulsed therapy and chemotherapy-induced drug resistance by utilising complex sets of both ODEs and discrete algebraic equations. Similarly Jackson (2000) [97] describe a PDE model that accounts for vasculature as well as effects of drug resistance.

Optimisation of treatment has been discussed with respect to chemotherapy above. However the vast majority of literature that deals with this branch of modelling is devoted to radiotherapy. Often the modelling of radiotherapy seems to employ a different line of enquiry with respect to other types of models - rather than focusing on tumour dynamics the vast majority of models are phenomenological, which derive quick and simple solutions to treatment problems such as fractionation, hence the TCP and LQ models, and extensions thereof. Barazzuel (2010) demonstrate this with the use of an extended LQ model for the treatment of glioblastomas to predict the effectiveness of conjugal radiotherapy with drug temozolomide (TMZ) [21].

Only relatively recently have the concepts of tumour dynamics and response been fused. Traditionally the optimisation approach always takes into account four main tenets, the so-called Four Rs described as follows [171]:

1. *Repair effects*: a lethal dose of radiation causes a double-strand break in the DNA of the recipient cell, causing the cell to die. Single strand breaks (sub-lethal events) can also occur but these are repairable by the recipient.

2. *Reoxygenation*: hypoxic cells become reoxygenated either by cell migration or vascularisation.
3. *Redistribution of cell cycle*: quiescent (hypoxic) cells are relatively harder to kill. Once these cells re-enter the cell cycle they are more susceptible to therapy.
4. *Repopulation*: due to the first three Rs.

Stamatakos' research describes the utilisation of a Monte Carlo approach to modelling in 3D which is very similar to that described earlier (except that the cell cycle model is more detailed). The model is CA-like in nature and simulates avascular growth; excellent agreement with experimental data is reported [203]. Since space is discretised the LQ model is applied to individual voxels, which themselves represent cells - in fact the same methodology is applied by Stewart and Traub, explained above. The radiobiological effect is modelled as a function of the cell cycle stage in which those cells are situated. Although this paper does not allude to any formal optimisation approach to find the best course of treatment, apart from the declaration of four fractionation regimes, the simulation predicts that all fractionation tests fail to eradicate all tumour cells. The overall simulated tumour dynamics exhibit good agreement with experimental results. A simple iterative process to this simulation can yield a good *in silico* optimisation strategy.

A similar approach is also adopted by Borkenstein (2004) [28]. The model takes into account cell cycling and response to radiation, as explained already for the models above, but also is one of the few models that incorporates the effects of angiogenesis and subsequent growth. This makes it one of the most powerful radiobiological response simulations available currently. Since the response formulae are applied in a discrete fashion a heterogeneous population of cells with differing sensitivities and responses to radiation according to cell state and immediate environment can be considered. Simulations revealed interesting characteristics of growth after fractional treatment, highlighting the importance of timing and repopulation of quiescent cells. Interestingly the results were quite similar to that produced by Stamatakos' group - an accelerated fractionation dose was predicted to give best chances for tumour control. Another intriguing result that the Stamatakos model could not have simulated, however, was the effect of angiogenesis to response - since vascularisation decreased hypoxia the tumour became more sensitive to treatment but only if proliferation of normoxic cells was fast. If proliferation was slow then the therapeutic benefit of having reduced hypoxia was eliminated and angiogenesis led to decreased response.

### 2.1.5 *Dynamic Pathway Models*

The field of intra/intercellular molecular pathway research and associated mathematical and computational methods embodies the core of the systems biology field. This is because ultimately the behaviour of pathways that will define the behaviour of the cell, which in turn contributes to the behaviour of the entire system [45, 208]. Pathways are currently an extremely active area of research and constitute an enormous body of literature.

Generically, three types of intracellular pathways can be defined from the literature that is qualitatively different in nature:

1. *Metabolic networks*, which have been dealt with for many years and involve enzyme/substrate biochemical reactions.
2. *Gene networks*, which are networks that are controlled both on the proteomic level (e.g. protein-protein repressors) and on the DNA/RNA levels (e.g. DNA-protein repressors and self-looping RNA), and are sometimes synonymous with the third kind of pathway in the literature.
3. *Signal transduction pathways*, which are dynamic in structure (especially in systems where mutation is a key factor) and in molecular flux and are involved in signalling of extra/intracellular stimuli, ultimately resulting in gene transcription (i.e. gene networks).

Intercellular signalling is generally considered comparatively less complex and has already been discussed to some degree in terms of diffusion in the angiogenesis modelling section and so will not be discussed any further here. The methodologies discussed below will focus on signal transduction as this has been predominant in the cancer-related modelling literature.

### 2.1.5.1 Modelling Methodologies

With the arrival of high-throughput technologies, such as microarrays, two types of distinct algorithms can be observed in the literature [104]: (i) data mining of high-throughput datasets to predict and define causal relationships between genes and proteins to build up conceptualisations of pathways; (ii) simulations of established conceptualisations. A vast amount of literature is devoted to the former. For example, Wang (2010) develop a complex non-linear mathematical model to infer genetic regulatory networks via transcription factors such as p53 [190]. These types of models, however, will not be discussed here in any detail in light of the main goals of this book<sup>13</sup>. Stochastic models that are derivatives or relatives of Bayes theorem dominate this type of modelling [16, 178, 174].

Computational methods have been extensively used in pathways research, both in terms of visualisation and simulation, which can be seen by the appearance of computational methodologies and techniques such as Systems Biology Mark-up Language (SBML)<sup>14</sup>. The number of databases that currently exist, e.g. Biocarta, KEGG<sup>15</sup>, DIP<sup>16</sup> and many others, reflects the amount of research that has gone into representations and visualisations<sup>17</sup>. Dynamic pathway models can be very difficult

<sup>13</sup> Usually these algorithms themselves are not mathematical models of the system. However, often the outputs of these algorithms, e.g. possible protein networks, do represent models of the system in a more direct way.

<sup>14</sup> <http://sbml.org/index.psp>

<sup>15</sup> <http://www.genome.jp/kegg/>

<sup>16</sup> <http://dip.doe-mbi.ucla.edu/>

<sup>17</sup> Visualisation methods are not discussed here - the reader is directed to Kitano (2003) for a review.

to compile and verify, which is possibly due to the lack of data to which models can be fitted [45]. As will become apparent in the next few paragraphs the vast majority of dynamic pathway models are based on ODEs, which of course need to be parameterised. Whilst microarray data, though extremely noisy, provide at least some data there exists no such comparative technology as yet for protein measurements, though some attempts have been made at high-throughput proteomics [133].

Cho and Wolkenhauer (2003) [45] highlight the importance of spatiotemporal modelling of the individual cell. Not only are molecular interactions highly non-linear, they are highly organised both in space and time, which means that serious models must also take into account the spatial organisation of organelles within the cell as well as the molecules themselves. Zhu (2003) [208] have responded to this problem with a call for new methods that can tackle the transient molecular interaction capacity and flexibility of gene regulatory networks and signalling pathways in general. ODE approaches are said to be an inappropriate strategy since the network itself is evolving in situ. They make a special distinction for metabolic networks and claim that these are relatively static (assuming no mutation). Crucially, Zhu and co-workers put forward a software engineering project that involves the integration of different types of model, i.e. continuous versus discrete, stochastic versus deterministic, qualitative versus quantitative, to build a hybrid platform<sup>18</sup> on which one can simulate the spatiotemporal evolution of dynamic networks (each integration is programmer-defined). The so-called Grid Cellware project presents a set of optimisation algorithms and solvers and exploits a parallel computing infrastructure to solve user-defined models (usually DE-based).

The Michaelis-Menten (MM) equations are a fundamental formalism for enzyme/substrate kinetics and essentially reduce to a simple ODE (not covered in detail here because the role of these models is not as prominent as others in tumour modelling).

The MM equation, which gives rate of change of product in time, can be expressed as:

$$\frac{d[P]}{dt} = V_{max} \frac{[S]}{K_m + [S]} \quad (2.13)$$

where  $[P]$  and  $[S]$  are the concentrations of product and substrate respectively;  $V_{max}$  is the optimum enzymatic kinetics on  $S$ ;  $K_m$  is the so-called MM constant and represents  $S$  at  $\frac{1}{2}V_{max}$ .

An interaction network can then be described as a contiguous set of MM equations.

Signalling networks are modelled slightly differently but almost all still use the differential equation approach to describe pathway dynamics. Tyson (2001) [180], for example, apply different types of (mostly) differentials to certain aspects of cell physiology - genetic regulatory circuits described as ODEs or Boolean networks; spatially-oriented signalling by PDEs and cellular automata; functional or integro-differentials for time delays and spatial averaging; stochastic models when there are small numbers of molecules. Cho and Wolkenhauer illustrate the use of differentials

<sup>18</sup> Cellware, <http://www.bii.a-star.edu.sg/research/sbg/cellware/index.asp>

with respect to a simplified version of the NF- $\kappa$ B signal transduction pathway. This particular pathway plays a central role in control of proliferation and apoptosis and is therefore sometimes found to be constitutively switched on in many tumour types. The general rate equation used is as follows:

$$\frac{d\mathbf{m}(t)}{dt} = \mathbf{h}[\mathbf{m}(t), \mathbf{k}(t), \xi(t)] \quad (2.14)$$

where  $m_i(t)$  denotes the concentration of the  $i^{\text{th}}$  molecule in the network;  $k_i(t)$  denotes the  $j^{\text{th}}$  rate parameter;  $\xi(t)$  denotes an uncertainty (noise) function at time  $t$ .

Most differential based models are derivations of this basic ODE. Accordingly Schoeberl (2002) [160] formulated an EGF signal transduction pathway model, which is still one of the most well defined dynamic pathways for tumour biology. Initial values are of course needed before the network can be simulated. This sort of model has huge potential - one can predict the behaviour of the pathway given certain conditions and therefore begin to develop drug-targeting strategies. Schoeberl and co-workers perform a number of simulations and describe molecular dynamics that would lead to deregulated cell proliferation, e.g. a simple increase in the number of EGF receptors. Similarly, Wiley (2002) [192] also describe mitogenic activity by modelling the increase of internalisation of the EGF receptor. However, to the best of the authors knowledge, no formal attempt has been made to extend these particular models to describe the dynamics of the cell or host system (e.g. tumour), which would require model integration. For example a growth model can be developed by linking these pathway dynamics to an individual-based model of growth, i.e. each cell implements the pathway model. However such a linking has important implications for scalability. This angle on model integration is discussed in Chapter 4.

Another notable effort, and extension of the EGF pathway described above, is reported by Miller (2004) [126]. It has been argued that ionising radiation can inadvertently cause the over-expression of both MAPK and TGF- $\alpha$  signalling pathways and therefore contribute to sustained proliferation. The authors therefore modelled the autocrine signalling induced by radiation-exposed tumour cells. MATLAB<sup>19</sup> is used to simulate the network of 148 chemical reactions and 104 ODEs. The resulting model agreed with experimental data, at least for short-term radiation response. Yet again however, the effects on (dynamics of) the whole tumour system are not considered in any serious way - the pathway model is considered in isolation though it does seem very well suited to be a component of a therapy model, such as that described by Stamatakis group described in the previous section. Such an integration would yield a simulation that is far more representative of a clinical case.

Stochastic methodologies are also quite popular in this area. Wolkenhauer (2004) [199] reviews the latest methods including the popular chemical master equation, Gillespie algorithm and  $\tau$ -Leap method (all of which are used extensively in the literature). Meng et al. (2004) strongly advocate the use of stochastic methods by

---

<sup>19</sup> www.mathworks.com

arguing the stochastic nature of natural systems such as signal transduction and therefore the inaptness of deterministic models. A distinction in the types of stochasticity is established by virtue of origin - extrinsic (randomness from outside the system) and intrinsic (randomness generated from within the system). Meng and co-workers then set out formalisations for these types of stochasticity. These stochastic models however are computer-intensive and so Lok (2004) [116] complements the stochastic modelling appraisals by reviewing parallel computing and shortcuts for the chemical master equation-based stochastic algorithms.

It can be argued that the models explained thus far are fundamentally flawed in that the pathways themselves exhibit static topologies. To tackle this problem Shmulevich (2002) [162] develop a Boolean/Bayesian hybrid model that fuses the conditional stochastic concepts of Bayesian networks with the simplicity of Boolean networks in a so-called Probabilistic Boolean Network (PBN) formalisation. The mathematics is quite complex and not shown here. However, simply put, a previously fixed topology of genetic regulatory networks are given a choice of reaction functions, yielding a more dynamic network. To the best of the authors knowledge such an implementation for signalling networks is not yet present in the literature though the method does seem to be flexible enough to do so. However, the method as yet remains untested against biological data.

Various other methodologies have also been suggested in recent years, many coming from concepts in computer science. Particularly, ABM has become more popular [70,39]. This permits spatio-temporal modelling and single molecular tracking akin to CA. Petri Nets are also becoming a popular method [206], as well as knowledge-based reasoning approaches [20].

Only very recently has there been any convincing attempt at directly linking pathway dynamics to tumour dynamics. Perhaps the first concerted effort in this respect has been presented by Christopher (2004) [46] who employ a forward-modelling paradigm to mine through data to first build up a pathway for proliferation (Description Cell Language for representation and ODEs for dynamic description) and then apply it in the context of a cancer cell. The proliferative potential of the cell can then be applied to a wider context of a heterogeneous population of cells (i.e. a tumour) for a tumour growth simulation and/or cell level drug response simulation. The group use the Virtual Cell platform for simulation, which will be discussed later. Jiang (2005) [99] alternatively use a Boolean network to model the behaviour of the cell cycle in response to oxic conditions and hence the effects on tumour growth. With the continuing rise in access to high power computing resources such multiscale models are expected to become more prominent.

### 2.1.6 Other Models

The review given above is by no means exhaustive with many more techniques only just emerging, e.g. the use of fractal mathematics in vascular networks [77] and drug delivery [61]. Model focus is also diversifying with increasing attention being given to point mutations [134] and chromosomal aberrations [75]. Finally, there is a plethora of modelling papers that focus on classification or clustering of some kind



of data and then deal with tumour prognosis or diagnosis in some way. None of these models deal with the dynamics of the tumour itself and can be categorised as phenomenological. In fact a huge number of models have been developed in response to the availability of large datasets like those offered by microarray experiments and other molecular-based techniques. For example, Catto (2003) [34] use a neuro-fuzzy approach in which linguistic (i.e. qualitative data) could be used to classify patients into prognosis groups. The group report an eye-opening 88-95% efficiency.

Recently David Gavaghan's research group at Oxford University<sup>20</sup> are tackling the modelling challenge from a number of angles, including differential equation approaches inspired by the Physiome Project (discussed in later chapters), agent-based methods as well as more phenomenological methods, e.g. the use of the finite-element method and nonlinear elasticity to build a three dimensional, anatomically accurate models of breast tumours and surrounding tissue [142].

## 2.2 Summary

A whole range of models and techniques that appear in the tumour modelling literature have been discussed. The inadequacies of traditional techniques, i.e. DE models, over computational methods such as ABM and CA have been highlighted. The tumour modelling literature can be dissected by the target system the models attempt to emulate, i.e. growth, response, angiogenesis and pathway models - though it must be pointed out that there is significant overlap across these sets.

## 2.3 Conclusion

Many different aspects of tumour dynamics, with a diverse collection of methodologies, have been modelled but there is no unifying principle between these models. There is currently little or no scope to reuse these models without extensive redesign and development, especially when the complexity of semantics can be so varied from model to model. To do so would require model integration technologies. The survey made here will help to highlight the points at which such integrations can be made, especially from the viewpoint of model behaviour, which will be further explained in subsequent chapters.

---

<sup>20</sup> <http://web.comlab.ox.ac.uk/>

## Chapter 3

# Coping with Complexity: Modelling of Complex Systems

**Abstract.** This chapter will detail the most salient points regarding complex systems, their behaviour and properties, with a focus on the tumour systems, and define related concepts. The concept of local interactivity in multi-agent systems will be introduced as a direct explanation of behaviour of complex systems. Systems biology will be described from a systems-theoretic point of view and a review of the state of the art in terms of modelling technologies will be discussed.

### 3.1 Preamble

Systems biology aims at describing biological processes as interactions of molecular components structured by regulatory wirings. [5]

Systems biology is an academic field that seeks to integrate high-throughput biological studies to understand how biological systems function<sup>1</sup>.

A systemic approach to biology ought to put the emphasis on the entire system; insofar as it is concerned with components at all, it is to explain their roles in meeting the needs of the system as a whole [50].

Systems biology addresses the missing links between molecules and physiology [32].

(There seems to be a consensus that) systems biology will progressively complement the conventional mode of study by facilitating the understanding of biological networks and mechanisms in terms of their dynamic system behaviour on different levels of organization [67].

Model integration strategies are the focus of this book. To formalise these strategies a detailed understanding of the nature of the very system that is being modelled is required. The field of systems biology, one of the fastest growing fields, strives to do precisely this - though opinion over its direction is quite mixed, as the quotes above show. In fact systems biology is central to the topic of model integration (in

---

<sup>1</sup> [www.wikipedia.org](http://www.wikipedia.org)

the application of tumour modelling given here) since it provides the basis on which an integration strategy must be placed, i.e. what are the (sub)systems involved; how do they interact; what are their components; what is their behaviour? A systemic view, according to Cornish-Bowden (2006) [50] above, should strive to answer these questions. Is this the direction in which the field is actually going? These questions will be explored concisely in this chapter.

Accordingly, tumours will be examined as complex, adaptive and dynamic systems. Based on this, the nature of modelling methodologies will be formalised in the next chapter, which will make the challenge of model integration clear and will make it possible to formalise solutions.

**Definition 3 (State).** A state is a quantitative or qualitative variable of the system that defines some condition of being. A state can also be defined as a specification of a system or process at a particular instant [198]. For the purposes of explanation in this book, the term is also applicable to abstract concepts that do not relate directly to any physical components.

**Definition 4 (Behaviour).** Behaviour is a defined pattern of state changes wherein the said state(s) may refer to both environmental and system states. A pattern of state changes constitutes a regular configuration, which can be identified (observed) specifically, of state changes over time (or other parameter), possibly in relation to or in conjunction with other states, within some predefined limits. Behaviours will often be referred to as non-linear or complex, which will be discussed further in following sections.

Diverse aspects of biological systems, from the underlying reasons of their complexity to overall behaviour, have been subject to debate and investigation in recent years. This is embodied in the large complex systems research literature and the impact of this research field and system modelling is enormous - some benefits have already been discussed in the previous chapter. However, the formalisation of concepts pertaining to the characteristics of complex systems is still an active field of research.

Relating certain key characteristics of complex systems to how they are typically represented in models *in silico* has direct implications on the engineering of model integration strategies<sup>2</sup>:

- *Complexity in behaviour:* States of the system as well as individual models that mimic the system can be highly nonlinear (non/linearity will be defined more precisely shortly). The integration of models must preserve the behavioural complexity of the original individual entities.
- *Complexity of component parts:* As will be explained in the following sections, complex systems are composed of components that may well exhibit a degree of complexity themselves. It may be necessary to make this layered complexity explicit in a model integration platform.

---

<sup>2</sup> A full definition of model integration from a formal perspective is given in Chapter 4

- *Compartmentalisation*: Modelling often involves compartmentalisation of the system, which can be based on both real compartments and artificial ones, to simplify the modelling process. A model integration strategy must respect any compartmentalisation that exists within the original models since these constitute implicit assumptions about what the models represent.
- *Environmental context*: Complex systems often interact with the environment, which can affect overall behaviour (though the environment itself is seldom modelled as seen in the previous chapter).

The following sections review these fundamental characteristics within the framework of complex systems theory. They will be discussed with respect to their applications to biology and, in particular, tumour modelling and, eventually, model integration (next chapter).

## 3.2 Dynamic Complex Systems

Most biological systems, including macro-systems such as those found in ecology, are said to be dynamic complex systems [82]. Although there are a small number of shared characteristics amongst different complex systems (CS) a succinct definition of a complex system is difficult to articulate since there is still no general consensus. An initial definition of a CS can be given as:

- A system that exhibits behaviour that is often non-linear due to a large number of interacting components. The interactions between the components, e.g. cells, may well be non-linear too.
- A system whose behaviour cannot be predicted merely by investigating the behaviour of the individual components in isolation.

**Definition 5 (Component, of a complex system).** A component is any part of a complex system that can be reasonably compartmentalised, either by physical or abstract partitions, on the basis that it forms a functional (but not necessarily function-critical) part of the whole.

**Definition 6 (Linearity and Non-linearity).** Linearity, mathematically, is defined as a behaviour that satisfies the properties:

$$\text{additivity} : f(x) + f(y) = f(x + y) \quad (3.1)$$

$$\text{homogeneity} : a \cdot f(x) = f(ax) \quad (3.2)$$

Neither of these properties is present in pure non-linearity: consider the combination of behaviours in complex systems as mathematical functional outputs, as defined above, which produce other observable behaviours where additivity and homogeneity do not hold (e.g. consider the signalling pathways of all the different molecular species within a cell; even when taken collectively this cannot explain the overall behaviour of cell). To define non/linearity more clearly, a linear behaviour, in terms

of system states, can be predicted with knowledge of any previous state and then aggregating a fixed (or fixed functional) amount over a parametric index (usually time) to the desired index, whereas this is not always possible with non-linear behaviour. In the case of complex systems non-linearity is a key property [130,18,194]. In this book, unless otherwise stated, the term complex behaviour is used interchangeably with non-linear behaviour.

*Open* complex systems invariably interact with the environment(s), which provides a context to their behaviour through feedback mechanisms. In many cases the environment is also considered complex, e.g. the tumour in the wider context of the surrounding tissue and body. Conversely, *closed* complex systems have no such interaction and are therefore said to be independently complex, though the existence of such systems is arguable. Whole organisms are often presumed closed despite the fact that, of course, there is ample interaction with the environment. In reality closed systems do not really exist but are in fact used in modelling scenarios to simplify the complexity problem.

**Definition 7 (Interaction).** An interaction in the context of complex systems can be defined as any communicative action between two or more entities where the entities can be either physical objects, such as molecules or cells, or abstract concepts, such as behaviours. The medium of communication is an abstract concept when the entities involved are not directly interacting physically, e.g. a causal interaction within a chain of events. The three main types of interaction described in this chapter are:

1. *Physical* interactions, e.g. enzyme-substrate reaction.
2. *Informational* interactions where two entities exchange information without physically interacting. This is an artefact of simulating physical systems where the actual physical interactions are not specifically simulated but "metamodelled" such that state changes in one entity affect state changes of another purely by mutual observation. Informational exchange will be used to simplify the way one can describe complex physical interactions where minute interactions need not be modelled specifically.
3. *Causal* interactions are similar to informational interactions in that state changes of entities, which can be linked into networks (where vertices represent entities and edges represent directed causal interactions), affect state changes in other entities. The difference is that a causal interaction is not simultaneously bidirectional, i.e. it is not an exchange via mutual observation, though feedback is possible.

**Definition 8 (Environment).** If an artificial division can be drawn such that it encapsulates the subset of all behaviours and entities of the system of interest; the remaining subset of behaviours and entities in the universe of discourse can be considered environmental.

**Definition 9 (Universe of Discourse).** The entire set of considered entities, variables, interactions, etc. in any given modelling situation.

### 3.2.1 *Properties of Complex Systems*

A complex system in its entirety represents behavioural dynamics that are the result of interacting components, which are sometimes referred to as agents, on many spatiotemporal levels.

**Definition 10 (Agent).** Agent has a varied meaning in the computer science literature depending on the context [191]. In this book, unless otherwise stated, an agent can be considered an entity/component of a complex system that is assumed to have some degree of autonomy of functionality within the system (as well as across systems via the environment) [39]. Typically an agent is considered to interact locally within the system, though in an abstracted model this need not be the case.

Besides agents, complex systems are frequently composed of, or interact closely with, other subsystems that are often complex and dynamic in their own right [32]. A tumour is a complex system demonstrating highly non-linear behaviour [57], which can be said to be a function of the interaction between many subsystems under varying environmental conditions, e.g. cells, microvasculature angiogenesis and ECM. Complex systems science must therefore not only encompass the scale between the individual component (agent, subsystem) and the whole (system) but also must capture an understanding of the dynamics of interactions within the system and environmental feedback [30].

A distinction that must be made before further exploring properties of CS is that of a *complex* system from a mere *complicated* one. Often one thinks of a complex system as something that has many interacting components. Indeed this could also be a qualification for complicated systems but it does not go far enough to define complexity. Similarly, behaviour may well be observed to be non-linear in complicated systems too albeit to a smaller degree. One of the most prominent and observable attributes of a complex system that sets it apart from other systems is that of emergent phenomena [17]. An emergent behaviour or property by definition is a phenomenon that cannot be predicted merely by examining the parts of the system by themselves - it is in evidence where the whole really is greater than the sum of its parts [100] and is therefore parallel to the concept of non-linearity, i.e. non-additivity and non-homogeneity. For example, consciousness itself is considered to be an emergent phenomenon arising from non-linear neuronal activity [129]. Emergence is a difficult property to gauge quantitatively. It is also very difficult to generalise the most favourable conditions for such a phenomenon to occur, however it is generally agreed that a complex system whose components are many must have the right balance of interaction between components such that interactions are not

---

<sup>3</sup> However this does not need to be the case - complex systems can also be made up of many simple components. Conversely, a simple system can be made up of components that are themselves complex, e.g. the earth contains many complex systems but its behaviour as a planet in the solar system is comparatively simple.

<sup>4</sup> Emergence is an oft-abused pseudonym for complex behaviour that is not easily explainable. Emergence shall be treated strictly as per the definitions given in the main text to avoid any argument about its actual relevance outside of the context given here.

so sparse as to render the system completely disorganised and not so much that the system becomes too orderly and rigid [47,100].

Prior to further discussion one must note that a great deal of debate surrounding emergence in the literature is not only due to differing philosophical viewpoints but also the sheer variety of definitions of so-called emergent phenomena. Emergentists Charlie D. Broad, John Stuart Mill and Samuel Alexander, as well as others over the last century, had set down much of the philosophical foundations of this subject. Based on these works one can tentatively define emergence in a general form.

**Definition 11 (Emergence).** A property is emergent if it cannot be deduced from the behaviour of the systems components, even with the most complete knowledge of those components, when those components are studied in isolation or in other wholes [27]. In other words, components and their interactions must be considered *together*, which make up the whole (system). Later explanations and theories of emergence have included the notion of flows of causality in the system, as will be explained shortly.

Note that the definition for emergence suffices for the work presented here only and is not representative of all views. The present and most popular views are derived largely from that of Alexander from which the concepts of weak and strong emergence have come to light (discussed shortly).

**Definition 12 (Causality).** Causality can be defined as a concept by which physical or abstract events occur or properties and/or states are changed as a result of other physical/abstract events, properties and/or states (i.e. causes) [167]. In other words there is an actual dependence of effects, or probabilities of effects, on causes. Formally, A is the cause of B if and only if bringing about A would be an effective way of bringing about B [132]. This is termed agency-oriented view of causality. Note that this is one of many views, however the agency-oriented view is most relevant to the work here and suffices so far as (i) model integration and its implications are concerned and (ii) its medical applications are concerned (i.e. medical reasoning). Extending this to the context of complex systems: Any event, property or state that is deemed a causal source (i.e. a point of origin, which may well be the endpoint from a larger perspective) can be linked to a number of downstream (cause-to-effect direction) events, properties and states. Note that this need not be linear, i.e. the direction can be cyclic. This will be explained in the main text shortly.

Two types of emergence can be defined at this point, namely weak and strong [137]. Weak emergence describes the behaviour arising from micro to macrostates resultant of the interactions on the microscopic level. Note that a complete knowledge (if such a thing is possible to achieve) of component properties is not equivalent to a complete knowledge of simultaneous and collective component interactions. In strong emergence, however, the behaviour of the whole can influence the behaviour of the component parts in a form of feedback, i.e. there is a top-down causal flow.

Critically, strong emergence differs from the weak form in that the causal flow is genuinely separate from the influence of the component parts and is therefore

said to be irreducible with respect to those components [24, 35]. In essence this irreducibility is similar to Broads notion of non-deducibility - the behaviour of the whole cannot be explained by the behaviour of the parts. In the arguments presented below, as well as the rest of this book, it will be maintained that though unexpected behaviours can arise, and can subsequently be labelled emergent (depending on its definition), a depth of knowledge of component properties as well as their interactions taken together, acting simultaneously, in the context of whole can lead to the development of a framework in which such behaviour, inclusive of the top-down variety described above, can be predicted.

Complex systems and emergent phenomena are said to arise from both bottom-up and top-down collective behaviours [100]. The top-down perspective, or *systems* view, is a globally directed phenomenon in contrast with the purely bottom-up situation, the so-called *reductionist* view [18]. Though behaviours arise from the short-range interactions of the component parts, e.g. short-range communication of cells in the tumour system, which embodies the concept of weak emergence, the concept of strong emergence dictates that some behaviour affects the component parts from a global perspective. The components behaviour in turn globally affects the system by feedback. As an example of an apparent strongly emergent behaviour take tumour growth (as well as normal tissue growth in this case) that can be constrained by the global pressure exerted by its own outer cells, e.g. the viable layer of a MTS. This pressure constraint (which is a global state, i.e. top-down, induced by the original microscopic system state itself) will manifest itself on the local scale such that each tumour cell is compressed into a smaller space, which in turn forces cells to slip into the  $G_0$  (quiescent) phase of the cell cycle - applied globally, this results in a retardation of tumour growth. Top-down control in this case has revealed itself ultimately as a combination of local interactions (the crowding affect of cells) and cell properties (induction of  $G_0$ ) - importantly the combination is a simultaneous behaviour across the entire system in conjunction with other system behaviours. It is from here that one can trace the communication of behaviour through the scale and structural hierarchy of the system, from tumour (system) level down to cell level and eventually the molecular dynamics that initiate rest phase, and the chain of causality continues back up the hierarchy to affect the system globally again. Note that the property of space constraint-based retardation cannot be predicted from analysis of the state-space dynamics of the cell or molecular agents alone nor would it be deducible if it were placed in a different whole. Additionally it cannot be predicted solely from interactions of agents. It is only when the cell interactions and properties are considered within the context of other cells, i.e. local interaction, in a systemic view, i.e. in the wider context of the tumour, does this property reveal itself.

Thus one must consider the nature of communication of such behaviour, i.e. how top-down behaviour is defined on the smaller scales of the system. From the above example it can be argued that top-down control, if it can indeed be defined as such, manifests itself as short-range interactions between and within components and its effects therefore become cyclically bottom-up - both system and reductionist



views in this case are simultaneously correct<sup>5</sup> because causality can flow in both directions. Non-simultaneous bidirectional causality need not be paradoxical when viewed in the language of the microcosm - top-down behaviour, the prime symptom of strong emergence, is merely the tail end of a causal chain, which is microscopically weak (in the sense of weak emergence) in origin, and temporarily continues via the global system and terminates back at the local level cyclically. This reduces strong emergence to the weak form.

When viewed in such a way it can be seen that the original notion of bottom-up system complexity (and ensuing causality) is not violated by the relatively transient causality transposed by global-local communication. The implication here is that for closed systems, where complex behaviour is independent of environmental states, any apparent top-down causality of the system is indeed bottom-up in origin. In the case of biological systems, where the systems are very much open to environmental conditions, only a subset of all behaviour is attributable to the small-scale and overall global dynamics is due to both bottom-up causality and environmental interaction-induced top-down causality.

**Definition 13 (Environment).** Corollary to Definition 8. The environment itself can be considered a system (be it complex or simple) in its own right if the function of the system of interest is not entirely dependant on the functionality of the environment. Of course this distinction of dependency is a matter of degree; a precise partition between system and environment, in the case of modelling, is defined by the perspective of the modeller. Naturally, the environment represents a system(s) that is larger or equal (in physical space) to the system of interest. The environment therefore typically interacts with the system of interest in a top-down fashion (whereas the reciprocation is of course bottom-up).

Take, for example, de Pillis and Radunskayas explanation of Jeffs phenomenon in the previous chapter wherein oscillatory growth dynamics is observed as a result of immune cell interactions *in vivo*. Such oscillations can affect the distribution of nutrients, cell cycle dynamics, mutation and adaptation rates of tumour cells, oxic-level population dynamics, etc., and can therefore even influence vascularisation and further disease progression. One can assert that this oscillatory behaviour, and its downstream effects on growth and tumour dynamics, possesses genuine causal power that cannot be deduced from examination of the components, which fulfils the basic requirements for strong emergence [24]. Such an assertion is, however, superficial at best - the effects of the immune system is *independent* of the tumour when considered as an environmental interaction. Critically, the interplay of these environmental systems and their integration with the tumour give rise to complex behaviours that can be considered weak emergence when their interaction is viewed to form an integrated whole. In other words, when one considers all of these interactions within the context of the larger system, in this case the human body, then the

---

<sup>5</sup> Although there is no reason to suggest that reductionism and systemism are mutually exclusive. Indeed, from the reducible systems point of view, systemism actually subsumes reductionism.

origin of all behaviour can be traced back to fundamental physical laws that govern the components and simultaneously the rules that govern the *interaction between* components - again, *interaction* and properties together are the key to complex behaviour. In this case, the components of the immune system, e.g. macrophages, and their interactions with the tumour cells give rise to the aforementioned oscillatory behaviour on the macroscopic scale - the immune and tumour systems as an integrated whole have given rise to the behaviour and there is no genuine top-down causality.

**Definition 14 (System Independence).** Systems A and B can be considered independent iff:

1. There exists a subset of distinct functions  $f_a$  and  $f_b$ , independent of the superset of functions that exist as a result of A/B interactions, where  $f_a \cup f_b = \emptyset$ , that are carried out by A and B respectively; Or:
2. A and B are totally functionally independent of each other - either A or B, or both, can exist and function in the absence of the other.

Note that system independence is a matter of degree - condition 1 is a relaxed form of condition 2 to account for the case of naturally interacting systems. Also note that condition 1 reinforces the definition of the environment with respect to the system of interest. Condition 2 does not enforce any stipulation on behavioural dynamics when A and B are together or apart. This is because the behaviour of each system may well be very different when interacting with each other, which could be the case in the naturally occurring case, e.g. tumour and immune systems, though the systems can theoretically be thought of as independent. In such cases one can consider one system as environmental (in the case of the above example, the immune system). Modularisation of systems by this criterion of independence, e.g. modularisation of all human systems, implies that complex systems can be viewed as networks of interconnected modules of functional and/or component interactions. A modular approach to complex systems is in fact fundamental to understanding complex behaviour and modelling of such systems, which will be discussed in more detail in the next chapter. System independence also raises the issue of system and component boundaries. In conditions 1 and 2 the boundaries are abstract (functional) but can also include physical aspects.

Systems in general, both natural and artificial, exhibit dynamic behaviour. To delineate a distinction from so-called chaotic systems where state dynamics are also in non-linear flux, albeit in an extremely disorganised fashion, behaviour in complex systems is more organised, structured and hierarchical as described above<sup>6</sup>. To define this distinction further, complex system behaviours are relatively insensitive to initial conditions and eventually stabilise, whereas chaotic system behaviours

---

<sup>6</sup> Self-organisation (described later) is another behaviour that sets CS apart from chaotic systems - organisational ability implies a reduction in entropy of the whole system. Indeed living systems have been shown to be able to dynamically change their Gibbs free energy potential (defined as the energy available to carry out any process, e.g. synthesis, taxis, etc.) to push the system into decreased entropic states [27].

vary hugely even with small initial state changes [17,194]. This convergence of behaviour from differing initial conditions, and indeed the resilience of CS to external stimuli, can be viewed as a type of adaptation, e.g. the tumour system has the adaptive ability to dampen sensitivities to adverse conditions [13]. Tumour systems have therefore been described as complex adaptive systems (CAS) with respect to self-sustenance and therapy resistance [144] though the property of adaptation is far more fundamental than this - supporting cellular systems in a dynamic environment itself is an extremely complex adaptive behaviour. The induction of angiogenesis by over-expression of TAFs and avoidance of apoptosis, for example, can be viewed as adaptive behaviours in a highly dynamic environment. Most often the mechanism of adaptation in the case of tumours is an accelerated mutation rate [134] where a selection pressure, such as application of therapy, together with a genetically diverse cell population can give rise to therapy-resistant cells. The tumour system then replenishes with the resistant cell variety and disease persists by this Darwinian process, where one can identify individual cells as mutual competitors. There are a number of sensitivities for which the tumour system cannot adapt too such as susceptibility to rapid mutation leading to unviable cells and, to take an extreme example, death of the host. Subsystem controls and microscopic dependencies are responsible for this robust but fragile behaviour [52], which has been described for natural and artificial systems alike.

Another property of complex systems is self-organisation [47], e.g. the organisation of the cytoskeleton and spatiotemporal organisation of molecules participating in signal transduction cascades within a cell. Self-organisation on the macroscopic level is defined to be the result of structural organisation on the meso/microscopic level, which in turn is due to local interaction between agents [51]. The properties of self-organisation and adaptation due to the interactions of agents are perhaps more evident in biological systems than any other set of complex systems. Therefore the recent pattern, as described in the previous chapter, of some modellers moving away from traditional mathematical techniques in favour of interaction/rule-based techniques should not be so surprising since a concept such as self-organisation is very difficult to describe mathematically without trivialising mechanistic behaviours<sup>7</sup>. In the context of tumour systems self-organisation becomes an interesting problem since it is the very loss of normal organisational control that causes over-proliferation - the normally robust tissue organisational behaviour, tightly controlled at both the cellular and molecular levels, reveals its fragility to induced and/or spontaneous mutations that escape apoptosis and early immune detection.

In summary complex systems have been discussed in the context of their properties, the most notable being that of emergence, nonlinear behaviour and self-organisation. Critically it has been argued that it is the interactions between agents within the system and across the environment, as well as their properties, when considered simultaneously in the context of whole (i.e. all of these interactions and properties *together*) that are responsible for overall observable behaviour. Moreover an explanation of systems on multiple physical and time scales is entirely conducive

---

<sup>7</sup> Moreover note that biological systems can exhibit a dynamic structure of organisation, not just a rigid one.

to an interactions-centric understanding of complexity that goes hand-in-hand with properties-centric understanding on the microscale. Importantly such an understanding holds the key to model integration strategies, which will be explained in the next chapter.

### 3.3 A Systems-Theoretic Approach in Biology: System Biology

The volume of research pertaining to simulations of biological phenomena, including cancer, has accelerated dramatically in recent years with high performance computing, e.g. GRID technology, and the availability of enormous amounts of molecular data via high-throughput technologies such as microarrays [31] has generated a surge of interest in understanding biological systems quantitatively on the minute scale. The field of systems biology embodies a sizeable portion of this research. As the name suggests, the field in principle spans entire cell-, tissue-, organ-, organism- and community-level understanding via modelling, simulations and visualisations though, as will be explained below, the main focus has centred on intracellular pathways. Systems biology is therefore a fusion of systems theory and mathematics with computer science applied to biology and readily lends itself to the inclusion of engineering principles. In addition to the systems perspective the eventual use of systems biology is closely linked, but not equivalent, to bioinformatics. Systems biology is not as new a field as one might be led to think - questions as to the relevance and application of systems theory to biology were seriously being considered as early as the 1950s [49], though the uptake by the larger community was still rather reductionist in spirit, e.g. elucidation of vast metabolic pathways of mammalian and plant cells, and one could argue that it is as old as systems theory itself (in fact Ludwig von Bertalanffy, the founder of systems science in the early 20<sup>th</sup> century, was himself a biologist by profession). Indeed systems thinking has often gone hand-in-hand with biology though applications have not directly been biological, e.g. John von Neumanns cellular automata theory in the 1940s was described in the context of reproductive artificial life (hence Conways Game of Life in 1970) and predator-prey equations developed by Lotka and Volterra in the mid 1920s which still play an important role in describing both ecology and cellular competition in many current modelling schemes.

Currently systems biology is shifting deeper into systems thinking such that it is paving the way to ever more complex applications of modelling and simulations to solve biological problems, and conversely the understanding of complex biological systems is providing a drive in computer science, mathematics and engineering to improve on current technology and methods. These two channels embody subfields that can be described as computational biology and biological computing respectively [82].

Wolkenhauer (2003) [200] describes the key challenges for systems biology as follows:

- Methodologies for parameter estimation.
  - As can be seen from the heavy reliance on differential equation and rule-based approaches in the tumour modelling literature, parameterisation is an important issue for accurate simulations. This can only be obtained from wet-lab experimentation and data mining. As will be explained in following chapters, the model integration process can also drive the search for parameter values.
- Modular representations and simulation of large-scale system dynamics.
- Scaling models across time-scales and description levels (genes, transcripts, proteins, etc.)
  - Model integration is recognised as a key technology in systems biology [69]. Here it is defined further as a technology that addresses the scale and model semantics problem too. Simulator software integration is one of many integration problems tackled by Systems Biology Workbench<sup>8</sup> (SBW) [156] and SCIPATH [141], as well as other middleware.
- Visualisation and fusion of information, integration of models and simulators.
  - Visualisation is a critically important aspect of systems science [87, 66]. Visualisation is pivotal in turning complex data and computation into human understanding.

Fundamentally, amongst the majority of currently active systems biologists, the bulleted points above distil to a single purpose: to understand the mechanisms of regulation, and therefore the dynamics, of biological systems [199] via enabling technologies and visualisations. Ultimately this involves modelling of the interplay between environment and microscopic local interactions of cells, molecules and other entities of interest, and the system of interest on the mesoscopic scale. Perhaps due to the huge efforts put into genome projects and greater understanding of molecular structure and events via improved technologies, the original systemic focus has reduced for many researchers to regulation of the genome. It is thus little wonder that the central topic that has dominated systems biology for the last eight years is molecular networks (metabolic and signal transduction), even though the field portrays itself as a richly varied mixture of perspectives as shown by the quotes at the beginning of this chapter. Mathematical modelling and simulations of pathways have received a great deal of attention from a computer science perspective [156] whereas the topology of networks [2], modes and (de)centralisation of control [19] and robustness [105] have been tackled from an engineering perspective. With a plethora of pathway modelling devices and databases, this trend of research is set to continue.

Given the current pathways-oriented focus in systems biology it would seem that the presumed goal is a fuller understanding of the cell, rather than a fuller understanding of tissue, organ or whole-organism systems (though it may well be the case that a hierarchical understanding, starting from the cell level, is sought). The academic question in this regard is whether pathway dynamics really *should* embody

---

<sup>8</sup> www.sbw-sbml.org

the bulk of systems biology research. Is it not the case that aiming to understand the component agents (pathways) to explain the behaviour of the cell system is reductionism by definition? Given the previous discussion on complex systems the answer is affirmative if pathways are considered isolated components of the whole, i.e. taken out from the context of the system. Unless this context is considered, the determination of pathway dynamics, whilst extremely powerful in its own right, is essentially an attempt to describe biological systems *serially*<sup>9</sup> and purely bottom-up and is therefore contrary to the view of systems thinking (i.e. consideration of properties and interactions of components simultaneously in the system). Indeed the vast majority of pathway analyses and models are not placed back into systemic context, i.e. system dynamics is not considered. It is not implied here that such modelling is invalid - indeed pathways are components of the whole (cell) and serial additions make knowledge of that subsystem of the cell more complete. However, in the opinion of the author, whether this will complete knowledge of cellular behaviour as a whole is an open question. The cell is considered a complex system its own right; its behavioural dependencies in environmental conditions raises the question of whether an isolationist view of pathway dynamics will ever really reveal cell behaviour to the point of tissue-, organ- or organism-level understanding (as may well be required for cancer systems, for example). Encouragingly, developments such as the Physiome Project (discussed shortly) seem to be guiding this research into a more systemic view.

This pathways-oriented view has been provided by the strong influence of engineering principles within the field - Hiroaki Kitano (System Biology Institute, Japan), John Doyle (Caltech, USA), Uri Alon (Weizmann Institute of Science, Israel) and Douglas Kell (University of Manchester, UK) who have all come from engineering-oriented fields, and groups such as ERATO (Japan) that produced Systems Biology Mark-up Language (SBML) with Caltech computer scientists, have made extremely influential contributions to the field via literature and software that have spurred on the continued research in pathways. Of course from the engineering perspective, where similar pathway principles have been drawn from artificial systems, the understanding of pathways is systems-oriented - networks of interacting individuals form the basis for overall behaviour as explained earlier but when applied to biological systems this has been capped to the cell level.

Cell-level research is of course extremely useful in its own right, as recent developments have shown. Genome sequencing, two-hybrid and microarray experimentation, all play an important role in driving pathways research to progress the understanding of cellular circuitry [104]. However, assuming the experimental data and thence the models and simulations are correct, to what extent will the understanding of these complex dynamics contribute to understanding the behaviour of the cell as a community member of an *in vivo* tissue? This is as yet an open question since the coupling of pathway dynamics with cellular behaviour is not always clear - pathway dynamics itself (i.e. flux of molecular constituents) is the primary

---

<sup>9</sup> Pathways in the real system are incredibly complex and so the intuitive way to build up a systemic picture is to break them down into smaller pathways that are examined in isolation and later joined serially by common nodes.

output but the collective profile of that output is often difficult to relate to cellular behaviour. The danger is that the systems biology viewpoint has focused on the cell system as the proverbial whole and has treated either molecules as agents, wherein molecules are linked to other molecules to form more complex pathways and networks, or pathways as agents, wherein pathways are linked to other pathways in an additive fashion. The behavioural dynamics of pathways, and ultimately cells, have therefore been reduced with the exclusion of other cell behaviours. The previous chapter presented projects modelling pathways in isolation from the rest of the potential, but unknown, network linkages (and linkages to other stimuli such as molecular-level changes in the membrane due to physical stress and strain), i.e. there is a real danger of developing pseudo-networks due to incomplete knowledge. When one thinks about as yet undetected interactions, which are omitted by necessity from the models, serious questions are raised as to how well these models actually represent reality. In summary, though systems biology promises a solution to understanding the complexity of life it is still a field with many serious challenges.

There is, however, a very real sense of systemic-oriented change within the field. In fact many systems biologists have turned to the wider context of the whole with software infrastructures, such as E-Cell and Virtual Cell<sup>10</sup>, and global ventures such as the PhysioMe Project, of which there have been offshoot projects, notably the Cardiome, Epitheliome and Beacon Projects (discussed later). One can see that the focus is changing from a cellular perspective to a systems-wide perspective with recognition of the fact that the cell, a complex system unto itself, is only a small component of much larger systems and that to gain any real insight, especially for clinical use, the mesoscopic scale (cellular focus) must be expanded towards tissue and organ levels. Tissues and organs are of course themselves components of the whole organism. Obviously the importance of complex agent and component functions cannot be overstated - the molecular level dynamics in the cell, for example, is vital in understanding overall behaviour and so even if the shift in focus is towards a larger (physical and/or time) scale the smaller scale dynamics must not be ignored. This is a very important challenge in modelling biological systems - building cooperative models at different scales that constitute the whole, wherein the inherent layered hierarchy of complexity of the overall system is addressed, is highly non-trivial. The role of model integration is to play a critical part in the solution to this problem, as will be explained in the next chapter.

### 3.4 Coping with Complexity: Modelling of Complex Systems

It is clear from the previous sections that an interaction-centric approach to complex systems is key to understanding the diversity of behaviour and this is especially so in the biological case. This raises the next question of how can any system behaviour be quantified formally? Systems can be viewed in terms of states and state

---

<sup>10</sup> Note that though the original focus of these projects were centred on the cell level they are moving to the tissue level with the availability of more computational power.

dependencies, which are the prerequisite to understanding behavioural dynamics as seen in the previous chapter. To formalise further:

- States, or *observables*, are described in terms of *state variables* (e.g. nutrient concentrations, diffusion, proliferation rates, etc.).
- States can change over time according to a *transition function*. The transition function itself is sometimes subject to change, and therefore one can obtain several orders of transition function for a particular system<sup>11</sup>. In practice this is sometimes difficult to conceptualise but some formalisations do exist, such as finite state machines.
- The set of all possible states is termed the *phase* or *state space*, which is not always possible to define explicitly.
- The transitions of states over many time indexes can be recorded as a *trajectory*.

There are many techniques by which one can begin to model complex systems given basic mathematical and computational tools. Some of these are discussed below and form the basis, either directly or indirectly, of many of the modelling efforts summarised in the previous chapter.

When viewed as observables states are seldom elemental. This means that other states of the system, which may not be directly measurable, collectively contribute to the value of the state variable. In that case it is conceivable that there exists a number of sub-state combinations that yield identical phenomenological state (observable) values. A complex system can then be modelled abstractly as a hierarchical tree with state and sub-state relationships (becoming more atomic as one reaches the termini of the tree) including combinatorial identities. Wolkenhauer (2001) [198] and Bar Yam (2004) [18] both ascribe to this phenomenological view of the system wherein only the observables are taken into account. The rationale is that the more atomic (given some modeller-defined cut-off) state dependencies and dynamics need not be modelled explicitly, thereby saving on computational cost, since collective observables can be used reasonably to explain behaviour. A simple formalisation is as follows:

$$\psi_t : \{s_{t,1}, s_{t,2} \dots s_{t,x}\} \quad (3.3)$$

yields the definition of a system given  $y$  observables of interest where:  $\psi_t$  is the overall system state at time  $t$ , consisting of a set of observables;  $s_{t,x}$  is the  $x^{\text{th}}$  state (observable) variable at time  $t$ .

Every state undergoes change by some state transition function:

$$s_{t,i} \xrightarrow{f_s} s_{t+1,i} \quad (3.4)$$

The transition function,  $f_s$ , can be more complex, involving multiple time steps, states and environmental inputs. Additionally, if more detail is required, the model can be extended by considering collectives of states that contribute to more phenomenological states. When considering state variables as discrete one can view the

---

<sup>11</sup> Though systems are rarely modelled with something more complex than a 3rd order function.

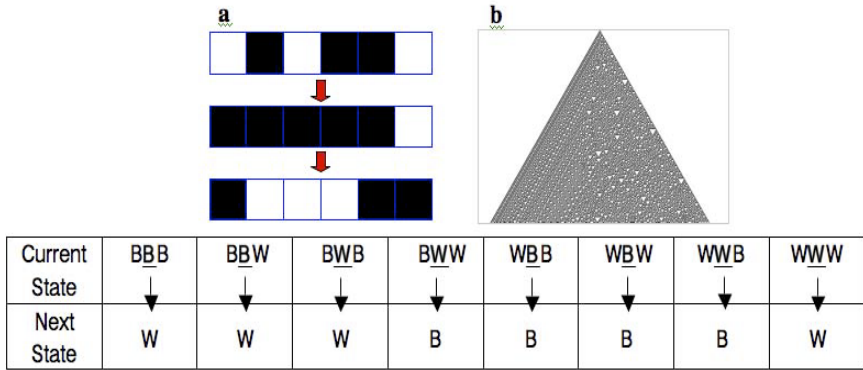


transition of states via the state function as a model of a finite state machine (by cloning the transition function for every state-to-state transition) - in fact a state machine may well be the starting point in a simple modelling exercise. Superficially the transition functions relation to the state(s), along with  $\psi_t$  as a whole, constitutes the definition of system behaviour. Whilst this will suffice as a rudimentary model, one can see immediately that it is no more in substance than the phenomenological models described in the previous chapter. The mechanistic basis of behaviour is hidden within the transition function. Whilst there are very obvious advantages to this methodology, as will be seen in the explanation of metamodelling approaches in the next chapter, there are also distinct disadvantages. For example, in the case of tumour models, research questions will often revolve around the mechanistic basis of behaviour since this knowledge can greatly benefit how therapy can be developed and improved: what are the molecular explanations of cellular behaviour (e.g. pathway dynamics that contribute apoptotic decision-making)? What are factors that affect the spatiotemporal development of neovasculature? Why does growth retardation increase even if there are enough nutrients as the tumour grows? These are just a few questions that can be raised. In short, if more mechanistic explanations of behaviour are needed for detailed analysis then this phenomenological approach is insufficient.

### ***3.4.1 Individual-Based Modelling Methods***

As can be seen from the literature review, individual-based models such as CA and ABM have gained more popularity as viable methods to cope with the complexity of biology. The relative simplicity and flexibility of such methods coupled with readily available software, efficiency, amenability to parallelisation and superiority over traditional techniques as explained earlier makes individual-based modelling an attractive alternative to cope with biological complexity [4]. Individual-based techniques shift the focus of modelling from a phenomenological to a local interaction-centric perspective [25]. In the case of bio-simulations with ABMs, for example, the bulk of the modelling effort is focussed on the agent, e.g. cell, and its interactions and behaviour rather than global dynamics of the tissue or organ. The volume of related modelling literature is enormous and only a small fraction with particular relevance to the work presented here is explained in any detail.

CA were first developed by one of the greatest mathematicians of the 20<sup>th</sup> century, Professor Jon von Neumann and his contemporary Stanislas Ulam [196]. Since then CA have been used and developed in many applications that include quantum mechanics, diffusion, hydrodynamics and engineering, general systems theory, cryptography, game theory, economics and biology [17]. The concept of the cellular automaton is extremely simple and is illustrated in Figure 3.1 in its simplest form - a 1-dimensional CA.



**Fig. 3.1** *a.* Discrete blocks (cell, or finite state automaton) make up a single line (i.e. 1 dimension). This initial state is called the configuration. The table above is called the rule-base, which explicates the rules that each cell must follow at each discrete time step. So if a cell is Black, and the two cells to its left and right are black, then its next state is White (the first rule in the table). This particular rule-base is defined as Rule 30 by Wolfram, which comes from turning the logical rules into binary (not discussed any further here). The rules operate on a neighbourhood of radius 1, i.e. they only observe the state of cells 1-cell away, however many different types of neighbourhood have been used (the most common being the von Neumann and Moore neighbourhoods, which are diamond- and dice-shaped respectively). The singular sequence is usually considered toroidal (i.e. left side is joined to the right side to make a ring). A donut, or toroid, shape is made if the CA is 2D). *b.* Rule 30 exhibits a complex pattern. Notice that one can observe tiny triangles emerging from the simple rules. These can be interpreted as fractals.

The deterministic transition rules determine the next state of a cell<sup>12</sup>, which is totally dependant on the present state of the cell itself (with some variations that include a concept of limited cell memory) and the present states of neighbouring cells, i.e. cells only behave according to their local neighbourhood. The direct relevance to biological systems here is obvious wherein biological components such as living cells behave according to their own states as well as the environmental state(s). The one major difference here is that the real system is continuous in nature whereas a CA is discrete both in space and time. It is easy to see how this can be transported into higher dimensions (of course only up to three for real-world simulations). Stochastic approaches have been incorporated into CA, e.g. Qi (1993) [149] use stochastic rules for cellular interactions in a growth model. According to the transition rules and configuration the resulting behaviour is observed to be very complex, repetitive or pseudo-random [129], however this kind of distinction of behaviour has only ever been examined rigorously with 1D binary state CA. Stephan Wolfram, who greatly popularised the utilisation of CA in the 1980s and 1990s, characterised the complexity of behaviour of 1D binary state CA but failed

<sup>12</sup> Note use of terminology: the term cell is used for the individual in CA as well as the biological type.

to describe their dynamics with traditional mathematical methods, leading him to propose that complex systems dynamics therefore cannot be described purely by traditional mathematical methods [194].

Many different variations of the basic CA described in Figure 3.1 exist. Rules can be complicated but are generally kept quite simple in most applications. Two of the most popular CA that are used routinely in systems science and hydrodynamics are the so-called Lattice Gas (LG) and Lattice Boltzmann (LB) flavours, which incorporate important physical laws into the rules, e.g. Newtonian mechanics<sup>13</sup>.

CA cells can be considered as a type of static (in space) agent and so CA can be viewed as a special case of agent-based systems (ABS). Table 3.1 illustrates the fundamental similarities and differences between CA and ABS.

**Definition 15 (Computational Agent).** (Corollary to Definition 10) A computational agent can be defined as an autonomous, or at the very least semi-autonomous, entity within a software system and is designed to perform some function based on predefined logical rules [83]. Agents are often encoded with some degree of intelligence to achieve local or global goals but communication with other agents is key to agent functionality (communication with the parental system is sometimes also just as important). States (usually discrete) can be embedded into agents, which can then have downstream effects on behaviour via a memory or learning. They can therefore also be viewed as state machines. Agents have been very successful in many application areas from optimisation and mathematical solving tasks to simulation avatars and industry-standard AI deployment [183]. Agents are logically separated from the environment in which they operate. One can define a number of distinct features that are discernable in most ABS: Precepts are the information flows between the environment (e.g. other agents) and the agent; actuator/effector elements provide an information-sending ability by which precepts can be sent to exterior agents; sensors are the precept-sensing faculty of an agent; processing elements make up the internal anatomy of an agent and is highly dependant on the overall ABS architecture as well as the intended functionality of the system - it may well include memory and learning faculties for intelligence capability, for example.

ABS can be extremely varied with types focussing on differences in the way agents behave, taking into account differences in anatomy (sensors, actuators, etc.), goals (function) and rule base types, to the environments (contexts) that they interact in and inhabit. In the case of this book the type of particular interest are those that exist within simulation systems, i.e. the computational analogue of the agents discussed earlier with reference to complex systems. These agents are different from the kind that one, for example, might encounter in optimisation methods where the agent is goal-oriented and perhaps communicative under a specialised architecture. The simulation system, in the case of real-world simulations, would constitute a computational representation of a limited 3D space in which components, i.e. agents,

<sup>13</sup> The associated mathematics has in recent years become extremely well defined but quite extensive and complex and so will not be discussed here specifically. Interested readers are directed to Succi (2002) [174] and Vesely (2001) [186].

**Table 3.1** Similarities and differences between ABS and CA. These are only very abstract descriptions; there exists a plethora of variants and hybrids. Above,  $d$  denotes a decision function, i.e. intelligent selection of appropriate action, which in the special case reduces to a matching function  $\mu$  that uses a simple state-lookup to perform the appropriate action;  $Env$  and  $Neighbourhood$  represent local interactor-capture functions (i.e. a definition of locality);  $\delta \rightarrow 0$  in the continuous case. All other symbols are as in main text (refer to the next chapter for a more in-depth description of the significance of the symbols used here).

Property	ABS	CA
Interactor	$Agent = \alpha$	$Cell = \chi$
Continuity	Discrete/Continuous	Discrete
Position	$xyz \begin{cases} = f(\Delta t) & \text{if spatial} \\ \neq f(\Delta t) & \text{otherwise} \end{cases}$	$xyz \neq f(\Delta t)$
Environment	$E = Env(\alpha) \leftrightarrow \{e_{1..n}\}$	$E = Neighbourhood(\chi) \leftrightarrow \{e_{1..n}\}$
State	$S = \{s_{1..m}\}$	$S = \{s_{1..m}\}$
Rule Base	$R = \{r_{1..p}\}$	$R = \{r_{1..p}\}$
	Action : $d(R, S, E)$	Action : $\mu(R, S, E)$
State Transition	$S_{t+\delta} \wedge E_{t+\delta} : Action_t$	$S_{t+\delta} : Action_t$

interact according to preset logical rules (note that rules do not have to be static - they can be learned, for example) and states. Presumably most agents would represent real-world objects and would therefore need to behave in a similar way in terms of state, space and time.

Of course there is always precedent for more complex abstractions. Take the typical case of a large tumour that can contain millions of tumour and normal cells [152], and would be too difficult to model on a strict individual basis. A more abstract approach will see each agent within a coarser-grained simulation representing a number of cells acting collectively instead. Hybrid systems (mixtures of the individual-based ideas described here) are also commonplace and provide a powerful means by which the most desirable parts each particular mechanism can be utilised. The model integration solution described in the next chapter is such a hybrid system where CA and ABS are combined to form a representation of the real-world complex system of interest.

### 3.4.2 Simulation Software

Formalising a platform for simulations of complex systems is a nontrivial task. Simulations of complex systems provide the key to future experimentation, hypothesis generation and intelligent prediction and hence such a formalisation is very important.

The engineering field in particular has taken the modelling and simulation of complex systems seriously and a number of platforms now exist<sup>[14]</sup>, many of which

<sup>14</sup> A comprehensive list of generalised software can be found at <http://www.idsia.ch/~andrea/simtools.html>

are focussed on one or a particular class of system [197]. Most platforms are commercial. Many applications also work complementarily to other popular packages such as Mathematica<sup>15</sup> and MATLAB. A number of platforms make use of parallel computing, such as the Parsec system [14]. Other packages include ColSim<sup>16</sup> and the Discrete Event Simulation (DEVS) formalism, upon which some older simulation frameworks are based [205]. None of these however has been specifically created for the simulation of biological systems.

Two relatively new packages are Modelica<sup>17</sup> and Ptolemy II<sup>18</sup>, which have already proved to be quite popular. Modelica is quickly gaining ground in becoming a new standard in the open source market as a viable alternative to commercial simulation and mathematical packages such as Mathworks Simulink. Modelica actually constitutes two layered components that are continually being improved and built on: (i) a field-independent model representation language, and (ii) a software infrastructure that runs models written in that language. Many applications have sprouted from this effort including complex aeronautical simulations and other multi-physics systems [138]. Modelica is of particular interest since it explicitly makes use of model/sub-model relationships and facilitates model-building in this way to promote the idea of model reuse. A GUI can be used to create specific modelling components, which can then be integrated with other models using a specifically defined mathematical relationship. This is supported by the idea of partial model generation, which allows the user to create modelling components with missing mathematical relationships that can be later realised (i.e. integrated with another partial model). As will be explained in the next chapter, this idea constitutes a simple linear integration akin to many previous model integration schemes [62, 81]. However there are limitations: so far the software only supports model integration with a restricted set of mathematical tools, namely differential, algebraic and discrete equations. The creators of Ptolemy II, which works along a similar principle using an OOP-like architecture to build and improve models, have made a concerted effort to include a wide variety of models that can be used within the system including finite state machines, event-driven simulations (discrete event) and a number of hybrids [26]. However, an appreciable amount of recoding is required if a new model type that cannot be reconstructed using the basic component types in the standard library is to be integrated into the system and represented in Ptolemy's model language (MoML). Additionally both Modelica and Ptolemy II have been specifically created for (engineering-oriented) physical systems rather than complex systems though the creators claim that given enough computational power the software can cope with massive amounts of calculations efficiently (hundreds of thousands of equations in the case of Modelica, which is in the same order of equations that was used for the Cardiome Project described below).

Two of the most prominent efforts in complex systems simulation on the cellular level are the E-Cell and Virtual Cell projects. E-Cell was one of the first

---

<sup>15</sup> <http://www.wolfram.com/products/mathematica/index.html>

<sup>16</sup> <http://www.colsim.org/>

<sup>17</sup> <http://www.modelica.org/>

<sup>18</sup> <http://ptolemy.eecs.berkeley.edu/ptolemyII/>

serious efforts to materialise from systems biology in 1996 attempting to simulate the behaviour of a single cell. Takahashi (2004) [177] describe recent efforts in the E-Cell project, including the multiscale (time and space) integration of submodules to build up the simulation. The group report some success with the integration of relatively simple implementations of ODE, Gillespie and ODE/Gillespie composite models. They are still actively working on different modules of the cellular system - for example, a model of the mitochondrion has been recently reported by Yugi (2004) [202] - as well as moving into microbial simulations.

The Virtual Cell project is a comparatively new venture with the same ultimate goal of whole cell simulation. Simulations have a heavy reliance on the description of interactions that are controlled by defined differential equations (communicated in Virtual Cell Mathematics Description Language), and the package is equipped with numerical solvers as well as a user-friendly interface [115]. Virtual Cell explicitly accommodates structural information (though non-spatial modelling is also possible), which is not as apparent in E-Cell, and is based on an ontological geometry framework. The framework defines the morphology of the cell upon which the inputted mathematical models operate. The morphology throughout the simulation, however, is restricted to being completely static whilst the spatiotemporal allocation of molecular networks within the cell (e.g. pulse of current through a neuron) is dynamic. Simulations can run in three dimensions and provide quantitative results within the formalised framework. The simulation engine and database is centralised at a high-performance server, making it available via the web, and implemented in Java [164].

There are many other cell-based simulations including the CyberCell Project<sup>19</sup>, which is focusing on modelling of *E.Coli* cells in a purely bottom-up agent-based approach. Discrete agents, i.e. molecules, are simulated to obey biophysical laws so as to mimic cellular pathways. Of course this method is very computationally expensive. Others include MCell<sup>20</sup>, a Monte Carlo modelling-based platform for microphysiology, and a number of efforts coordinated by the International E.Coli Association (IECA) [89], WebCell [112] and various commercial ventures, e.g. Genomatica<sup>21</sup> and Physiomics Plc<sup>22</sup>, who also attempt more macroscale simulations. A relatively recent development is that of the Silicon Cell group [166], powered by JWS Online<sup>23</sup>, which is a data-driven approach to pathway modelling.

The Physiome Project is currently the most well known biologically related complex systems simulation project with an impressive simulation of the heart, its flagship project, already completed [92], though improvements are ongoing. This project is an international collaborative effort that includes similar biological simulations for the lung, musculo-skeletal system, digestive system and liver. Interestingly an e-Science project has started quite recently headed by David Gavaghan<sup>24</sup> at

---

<sup>19</sup> <http://www.projectcybercell.com/>

<sup>20</sup> <http://www.mcell.cnl.salk.edu/>

<sup>21</sup> <http://www.genomatica.com/index.shtml>

<sup>22</sup> <http://www.physiomics-plc.com/>

<sup>23</sup> <http://www.jjj.bio.vu.nl/index.html>

<sup>24</sup> <http://web.comlab.ox.ac.uk/oucl/research/areas/biocomp/tumour.html>

Oxford University. The vision of the Integrative Biology<sup>25</sup> group is the initiation of a Grid-based platform on which multiscale simulations can be performed, including visualisation and data sharing on a global collaborative scale. Latest reports<sup>26</sup> indicate good infrastructure progress, however the platform itself is not yet ready for public release.

A subsidiary of the Integrative Biology group is the Epitheliome Project<sup>27</sup>, focuses on epithelial tissue and is still ongoing. In contrast to the arguments presented in Section 3.3 where it was stated that systems biology, with respect to pathways modelling, has too often been capped to the cell level, the Epitheliome Project strives to describe multi-cell behaviour and does so by crossing genetic, pathway, cell and tissue levels<sup>28</sup> by agent-based modelling. Interestingly, Smallwood (2006) [165] describe a hierarchy of models to cope with complexity, using a metamodelling approach to describe lower-level dynamics (metamodelling will be discussed in more detail in the next chapter). This work has led to the development of the Flexible Agent Modelling Environment<sup>29</sup> (FLAME), a parallelisable software package for modelling biological systems as well as other systems (recent developments have included modelling of the European economy).

A recent and prominent tumour simulation effort, though not formalised into any software platform like FLAME, Virtual Cell or E-Cell, is proposed by Alarcon (2004) [3]. The model presented focuses on blood flow and heterogeneity of the tumour. This paper was, until recently, the only one in the literature to the best of the authors knowledge to describe tumour dynamics (in terms of growth) on such a wide scale, spanning from a molecular activity (including p27, cyc, cdh for control of cell cycle) to blood flow and response to chemotherapy. CViT<sup>30</sup> is an emerging community of systems biologists that aim to create a virtual tumour. However, to the authors knowledge, this group are still in the preliminary stages of providing an online modelling archive and have not yet reached a whole tumour simulation stage. One of the latest modelling efforts is described by Zhang (2007) [207], which is a cross-scale ABM. Specific software engineering details or new modelling methodologies have not yet been disclosed.

### 3.5 Summary

Complex systems have been briefly described as systems in which there are often very large numbers of interacting agents that collectively display both linear and non-linear behaviours, as well as self-organisation. These behaviours can often be described as emergent. At least in biological systems these behaviours, in conjunction with environmental interactions, can be described in the context of local

---

<sup>25</sup> This is in fact an offshoot of the older Physiome Project.

<sup>26</sup> <http://www.integrativebiology.ac.uk/progress06.html>

<sup>27</sup> <http://www.epitheliome.com/>

<sup>28</sup> [http://www.dcs.shef.ac.uk/~rod/Integrative\\_Systems\\_Biology.html](http://www.dcs.shef.ac.uk/~rod/Integrative_Systems_Biology.html)

<sup>29</sup> <http://www.flame.ac.uk/>

<sup>30</sup> <https://www.cvit.org/>

interactions of the agents, e.g. cell-to-cell/intra-cellular space interactions, and their internal states. In recognition of this fact several efforts have arisen in systems biology, such as the Physiome Project, which aim to model biology at various levels of granularity. However not all have taken this paradigm to the individual (agent) level, favouring DE-based techniques instead. A number of simulation software projects and modelling standards have arisen that describe biology at the level of signalling pathways, which has driven systems biology to its current state.

### 3.6 Conclusion

The vast majority of software and modelling projects are working towards a system-wide view but the extent to which such efforts have been successful is as yet arguable. The true yardstick by which success can be measured with respect to tumour modelling will be in how well models will be able to provide mechanistic explanations to complex behaviour and thereby present a viable protocol for therapy development. Many of the methodologies described above, in combination, can conceivably be used to this end however one must bear in mind the sheer complexity of the system in question. Additionally the agent-based nature of biological systems, which are very difficult to model with traditional techniques, must also be considered in serious model efforts. Whilst the system can be broken down into functional components which can often be modelled separately, these models then need to be integrated together to form a computational whole and it is this need that is the prime motivation for this project. The next chapter will discuss what this process is, how it can be done and ultimately how this knowledge can aide the systems biology field in the search for extensive, viable, systems-level *in silico* simulations (and most pertinently to this project, what role model integration will play in realising these goals).



## Chapter 4

# Complexity and Model Integration: Formalisations

**Abstract.** In this chapter model integration will be defined and reviewed with respect to the current technologies. Based on the previous literature review, models will be redefined in terms of paradigm (formalism), what the model actually represents, its associated semantics, scale and behaviour. This will in turn help define the core challenges of model integration.

### 4.1 Preamble

The previous two chapters clearly demonstrate the predictive and analytical power of mathematical and computational techniques. Many published models have demonstrated the ability to predict non-linear behaviour to a limited extent and to this degree it can be argued that these techniques can gain ground as viable substitutes in at least the preliminary steps of the research cycle as has been the case in the engineering fields. However it is vitally important to recognise the fact that models in the field of systems biology are as yet simplifications. Even the more complex models among those reviewed have been simplified to focus on a select number of processes and subsystems, limiting their use to the corresponding biological mechanisms and therefore restricting their clinical viability. To build more representative and detailed simulations one must start to integrate models together wherein the individual models represent different or overlapping components of the system.

**Definition 16 (Model).** (Corollary to Definition [11](#)) An abstraction of one or more observable system of entities, states and/or behaviours such that the artificial abstraction reflects reality to some predefined degree of accuracy (if such a metric is known). Models can arise in a diverse number of forms (explained shortly). Mathematical, computational and logical models use mathematical, computational and logical language respectively, or a mixture (hybrid model), to describe the system in question. Logical varieties are often categorised under the computational category since a set of logical arguments that compose such a model is often translatable into a computational one. The term model is used abstractly throughout this text to mean

any of these types unless otherwise stated. A more detailed definition of models of these types is given in the main text.

**Definition 17 (Model Integration).** The process by which mathematical, computational and/or logical constructs can be incorporated into a single framework such that the resultant model is representative, to some reasonable degree of accuracy, of all of the original constructs. In this case representative means that the resulting model integrates the behaviour and semantics, but not necessarily the precise formalisms, of the original constructs. The integrated model need not run (be solved) within the same framework. The three different views on model integration can be summarised as below:

1. **Isolative Communication (Model Sharing):** Model integration is limited to translation into a lingua franca, which can then be imported into different frameworks [63]. The responsibility of integration (with respect to types 2 and 3) and model solving is then shouldered by the importing application.
  - a. E.g. SBML (explained later in this chapter) is classified as isolative since it only describes a model mathematically, as well as related semantics, but passes off the responsibility of integration to the application(s) that import and interpret it.
2. **Integrative Communication:** Models run independently with middleware that communicates states so that models can coordinate globally. The software components and languages the middleware uses become the focus of model integration research.
  - a. E.g. Systems Biology Workbench (SBW) enables live applications to share objects, such as simulation states. This is different from isolative communication where the communication policy is static, such as an XML file.
3. **Decomposite:** Models are decomposed into component parts, or collections of components and are converted into other runnable components, and are solved under a single framework. This can often reduce to integrative communication on a finer scale, i.e. model to submodel decomposition and communication. There are various degrees of decomposition, discussed later in this chapter.
  - a. To the best of the authors knowledge only the integration strategy developed in this project comes under this category (see next chapter).

## 4.2 The Nature of Models

An understanding of the very nature of models and their modes of use is critical in developing a formal platform on which integration can take place. Any model integration technology would need to be robust to the diversity of modelling techniques and so it becomes essential to develop an understanding of this diversity, from where it originates and on what levels it occurs so that:

1. The complexity of the model integration problem can be stated clearly.
2. A formal methodology for model integration can be explicitly defined.
3. Software design of a model integration platform can be defined<sup>1</sup>.

### 4.2.1 Abstract Views of Models

It is clear that a complete low-level classification of models for the purposes of model integration is futile due to the sheer number of model types, hybrids and contexts in which models are used. Even if such a task is undertaken, by the time one could finish such classification it is likely that concepts have evolved and that new concepts and hybrids of previous methodologies would have been developed. Therefore a more abstract approach is needed.

Giavitto (2002) [82] describe formalised models in terms of their roles:

- *Pedagogical* or *heuristic* models are used to illustrate complex relationships between system components and can be used as a teaching tool (e.g. simulation visualisations, causal networks and illustrative diagrams).
- *Normative* models are reference models, which are used as comparisons between scientists or as blueprints for generic illustration of frameworks (e.g. UML models).
- *Constructive* models are used to build upon pre-existing formalisms or approaches. They are sometimes termed composite models and involve a certain level of (sub)model integration.

Such an abstraction does not inform one of any notion of model content. Ideker (2003) [95] go further and compare the most popular formalisms in computational cell biology (see Figure 4.1) and classify them by granularity. Epidemiological, i.e. phenomenological, models are placed on one end of the spectrum and mechanistic models on the other.

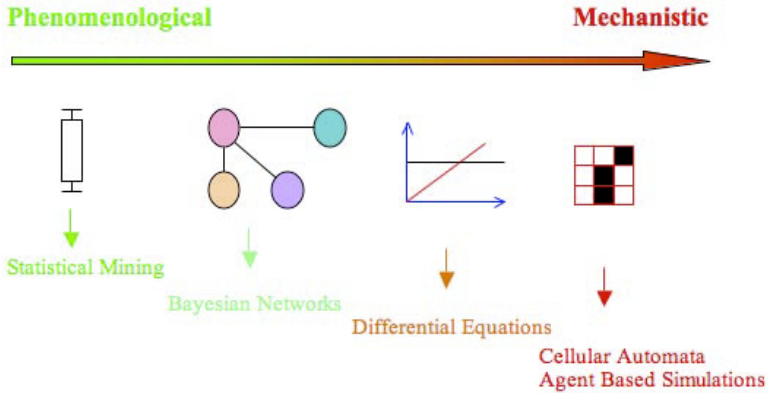
**Definition 18 (Granularity, of a model).** The level of detail that is presented in any model. A fully fine-grained, or low-level, model is defined as the model of which the components cannot become any more fine-grained, either because the model incorporates a full understanding of microscopic (mechanistic) knowledge, which is of course theoretical, or the model has reached the limit of mechanistic knowledge. Conversely a model is termed course-grained (more phenomenological) when mechanistic knowledge is omitted and behaviour is determined at a higher level(s).

By itself this crude classification does not shed any light on the model integration problem, except that different models require and contain vastly different volumes of data and data types based on scale (discussed later).

The points made by Giavitto (2002) [82] and Ideker (2003) [95] based on role and granularity-related information content allude to a rough abstraction of model

---

<sup>1</sup> Exact implementation of the proof-of-concept software is not detailed in this book.



**Fig. 4.1** Different formalisms pertain to different fields of view. Ideally models and simulations of cancer from which one can ascertain causal and emergent phenomena must come from detailed mechanistic models.

formalisms without actually defining those formalisms (e.g. P/ODE-based, statistical, ABM, etc). There is, however, a need to describe models on a more fundamental basis without restricting oneself to a rigid and formalised classification that cannot be later evolved. Such a description needs to include all of the most salient points of the model in light of the possibility of future integrations with other models: the mathematical, computational and logical formalisms that it uses, how the model is run within a software environment and associated issues<sup>2</sup>, what the model actually represents and how it is to be used (i.e. its goals and limitations). The following criteria has been developed to examine the models:

- *Model paradigm*: the abstract mathematical and/or computational basis (formalism) of the model.
- *Model focus*: the real-world subsystems of interest (e.g. angiogenesis, immune system, etc.), including physical components as well as associated behaviour (defined directly or indirectly by the model).
- *Model scope*: the state space of the model defined by allowable variable values, context and underlying assumptions.
- *Model scale*: the spatial and temporal level on which the model operates.

Analysis of models will involve determining the constitution and form of these four points. The results of this analysis are presented in the following sections. These points that define the nature of models, are then further defined from the perspective of model integration and refined into the core challenges that any formalised framework should address. Plausible solutions to these challenges are given in later sections and tested in Chapter 7.

<sup>2</sup> This is in fact a matter of solver technology, which will be discussed in more detail later in the chapter.

### 4.2.1.1 Model Paradigms

**Definition 19 (Model Paradigm and Paradigm Profile).** The fundamental mathematical, computational and/or logical principle(s) upon which a model is based. Models will exhibit more than one paradigm and hence a paradigm profile can be defined. The paradigm(s) dictates not only the nature and behaviour of the model but also the form of any solver that can be attached to it, i.e. how the model is manipulated for use. Note that solvers and models are from hereon treated separately: paradigm, focus, scope and scale relate only to the model in this discussion, though the same theory can be extended to solvers.

**Definition 20 (Solver).** The general definition of a solver in the literature is a procedure, or set of instructions, that a computer can execute to solve an instantiated model [179]. However this does not go far enough to include abstract constructs, e.g. ODE models that contain uninstantiated variables but can be solved by a series of mathematical techniques that would include curve-fitting and discretisation methods. Therefore the definition is extended here to any enabling mathematical technique or software application that manipulates a model, whether it is instantiated or not, such that the resulting representation is usable in at least partial fulfilment of its potential. Note that not all models are designed to be solvable.

Previous work specifically on paradigms is scant; it seems that a paradigmatic angle of understanding has not been developed though many authors have merely commented on the fact that integrating models across qualitative/quantitative, stochastic/deterministic and discrete/continuous paradigms (sometimes called dimensions) is a difficult undertaking [182]. Villa (2001) [187] has been the closest to addressing this issue albeit in the environmental modelling field with respect to model integration. The approach taken, implemented in the so-called Integrating Modelling Architecture (IMA), makes integration of submodels explicit with respect to paradigm (paradigm in IMA being defined quite loosely as mathematical or computational formalisms, e.g. ABM and ODE). In close correspondence with this view CellML, a formal and accepted standard in systems biology that addresses model exchange standards, models are viewed as compositions of subcomponents (synonymous with submodels) each of which must be described by some metadata<sup>3</sup> that details three fundamental paradigms: time, space and behaviour [114]. Whilst the explication of time and space is quite straightforward, behaviour is more difficult to describe. Villa takes an iterative approach in which a new integration strategy is developed as and when a new paradigm profile is encountered. A semi-autonomous approach to model integration can then be sought in which the IMA software generates translation functions between variables in each model component dynamically (i.e. as they are connected together) based on that metadata. Essentially this type of model integration can be termed linear integration, which will be discussed in more detail in later sections. The advantage of such a decomposition is that behaviour encapsulated in individual models need not be lost in the integration process as subcomponents remain as computing subunits within the integrated model. Additionally all

<sup>3</sup> Grammars are expressed in XML.

variables and models are treated as data sources, which makes the process easier to understand for non-technical users. However it is not entirely clear how complex semantics relating focus, assumptions, etc., can be incorporated into the integration process and, of course, the existence of a translation function is not always guaranteed.

To redefine and interpret model paradigm more formally and clearly an in-depth analysis of the cancer systems biology field has been carried out. The results are as follows.

*Functionality* is the most abstract paradigm. *Functional* models are referenced within the context of executable mathematical or computational constructs, i.e. models that perform some kind of processing function, but can also refer to more qualitative constructs such as executable semantic networks. Examples of functional models include differential equation models, Bayesian networks, CA and ABM. Though an UML diagram does not perform a specific executable task it is still considered a type of model since it characterises an abstraction of a real-world system, e.g. the realisation of a software system. The same can be said of schemas, which precisely and formally define, for example, the structure of XML documents. These models are herein defined as *non-functional*. Note that an unsolved model, i.e. a theoretical construct, can also be viewed as being non-functional with the given definition. However the concept of solvability is an essential part of the goal of such models, e.g. ODE equations in contrast to schema. One could counter that a UML diagram similarly has solvability at the core of its goal. Essentially a clear dividing line can only be established if one accepts all models as non-functional, whether solvability is a part of the goal or not, and solved forms of models as functional. This entirely supports Definition [II](#).

One can distinguish models further on the basis of numeric content. In other words the model is either *quantitative* or *qualitative*. Generally qualitative models that have little or no concept of solvability exhibit the aspect non-functionality. However, with a surge of interest in fuzzy systems wherein qualitative models can be quantified the boundaries of this particular paradigm are becoming increasingly blurred. Fuzzification and defuzzification processes, now commonplace in control system models, have closed the gap between qualitative and quantitative aspects of modelling [\[198\]](#). However it must be said that such processes are dependant on human and/or optimised input, e.g. generation of membership functions.

Models also have a *continuity* aspect, i.e. the model is either *discrete* or *continuous*, which more specifically pertains to the actual variables used in the model. In specific modelling paradigm cases this aspect can also pertain to actual components used in the model, e.g. agents within ABS are considered discrete entities (which could concurrently operate on continuous state variables). The previous chapter has highlighted the use of these paradigms in the context of systems biology. Moreover it has become apparent that these extremes are interchangeable when moving from an unsolved to solved form. For example, Anderson (1998) [\[7\]](#) use a model that is fundamentally continuous but solve it discretely via the Euler method. Discrete CA and finite element/mesh methods are common solver technologies for continuous systems [\[195\]](#).

*Stochasticity* is another prominent paradigm, i.e. models have *stochastic* or *non-stochastic* (deterministic) characteristics. A stochastic model implies both functionality and quantitative numeric content. Stochasticity is not strongly linked with continuity even though the traditional probability scale and associated calculi are continuous.

The *centrality* paradigm outlines a fundamental goal-oriented concept within the model-building process and is best understood when considering the difference between simulations and goal-oriented approaches such as optimisation algorithms. For example there is a clear paradigmatic difference between (i) a simulation that is run over time and shows the drug dose effects on tumour growth; (ii) an optimisation or classification algorithm, such as a neural network, which takes the existing tumour size and consistency as input and calculates the recommended dose requirements to minimise tumour size. The first algorithm runs over time whereas the second algorithm has no notion of any simulated trajectory of target parameters. The second algorithm can also be likened to data-mining algorithms where no simulation as such is executed. Rather, an ulterior model or construct is generated, e.g. gene expression data-mining algorithms that infer plausible genetic regulatory networks. Therefore this paradigm is embodied by *centralised* and *decentralised* character. A model can have a specified centralised goal, e.g. optimised therapy or establishment of a classifier from data-mining, as opposed to a relatively decentralised goal that is characteristic of time-continuum simulations.

Finally, there is the aspect of *dimensionality* to all models, i.e. spatial and temporal dimensions. As the previous chapter has shown, models and simulations manifest in a variety of types in terms of dimensionality, from non-dimensional to four-dimensional models (and higher for centralised models). As with centrality, dimensionality is often a property of the solved form rather than the abstract model. For example, ODE equations explicitly defining the Laplacian  $\nabla$  can describe dynamics in all three spatial dimensions but may well be solved and simulated in 1 or 2 dimensions for simplicity. Additionally phenomenological state models are often non-dimensional since they describe some observable state (e.g. tumour radius, cell number).

Note that models are characterised by the six paradigms listed above applied to all submodels, i.e. a submodel elicits a paradigm profile. This divided component-based architecture is common to all complex models encountered in the literature, illustrated in Figure 2.1. Formally a model can be defined explicitly as a novel 6-tuple:

### Theorem 1

$$M = \{P_i^m, \dots, P_n^m\} \quad (4.1)$$

$$P_i^m : \begin{pmatrix} F_i^m = \{f_i^n, \neg f_i^n\}, Q_i^m = \{q_i^n, \neg q_i^n\}, \\ C_i^m = \{c_i^n, \neg c_i^n\}, S_i^m = \{s_i^n, \neg s_i^n\}, \\ E_i^m = \{e_i^n, \neg e_i^n\}, D_i^m = \{d_i^n, \neg d_i^n\} \end{pmatrix} \quad (4.2)$$

where  $M$  is a model composed of  $n$  submodels, indexed by  $i$ , with respective paradigm profiles;  $P_i^m$  is the overall paradigm description indexed over  $i$  submodels;

$[F, Q, C, S, E, D]_i^m$  is the paradigm where  $F, Q, C, S, E$  and  $D$  represent functionality, numeric content, continuity, stochasticity, centrality and dimensionality respectively;  $[f, q, c, s, e, d]_i^m$  is the paradigm type for the respective type.<sup>4</sup>

**Definition 21 (Submodel).** An unsaturated submodel is any model component that can be considered a model in its own right as per Definition 16 and that can clearly define each of the six paradigms without mixture (singleton membership) to form a paradigm profile (i.e. the submodel is unsaturated in all paradigms). If the submodel is found to exhibit multiple membership in any one or more paradigms as listed above the submodel is termed saturated. From hereon, for clarity, the prefix sub is dropped in the main text unless the context is unclear.

The ultimate challenge in formulating model integration strategies from this perspective is finding a solution, within a single formal framework, for integrating models that exhibit very different paradigm profiles. For example, how can one integrate stochastic and non-stochastic models? The explication of these paradigms is aimed at capturing the most elementary nature of mathematical and computational models, thereby revealing the difficulty of model integration due to paradigmatic differences. If model integration can be understood to be a type of communicative ability between running models differences in paradigm imply the implementation of a translational tool<sup>5</sup>, e.g. implementation of a (de)fuzzification algorithm to quantify qualitative models. For any two models, a and b, where  $\forall a \wedge b : X_j^a = X_k^b$ , and where  $j$  and  $k$  index submodels that are to be integrated, model integration with respect to paradigms is trivial (excluding the complexities of focus, scope and scale, which are explained later) since paradigms are identical and no translation as such needs to be implemented.

#### 4.2.1.2 Model Focus, Scope and Assumptions

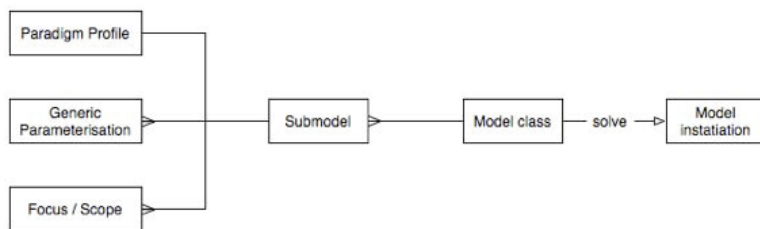
**Definition 22 (Model Focus).** The artificial partition of a system and its associated observable behaviour such that this partition represents subsystem(s) and behaviour(s) of interest and is accordingly the subject of the model. Note that this definition is the abstract equivalent of component (Definition 5) with the extension of including associated behaviour.

**Definition 23 (Subsystem).** A system that is entirely or at least in part encapsulated by a larger system and its behaviour is subject to itself and its environment [18, 197]. In effect a subsystem is a system in its own right; biological entities are made up of many such subsystems, which can be thought of as the real-world equivalent of submodel (the entire CS being equated to a complete model).

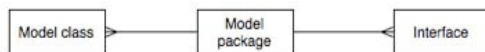
<sup>4</sup> Superscripts are used to distinguish symbols from other formalisations in following sections.

<sup>5</sup> As will be explained in following sections, model integration need not solely take the form of this type of chaining.





**Fig. 4.2** An entity relationship diagram of model anatomy. A submodel consists of a single paradigm profile. In practice a submodel might not be decomposable as such (e.g. a complex black-box), however conceptually a model can be recursively broken down into functional components. Such submodels almost always ascribe focus and scope and in the preliminary stages (especially so for DE models) are formulated in general forms, which then need to be solved. In the case of DE models, for example, parameters are not instantiated until some solver (curve fitting, integration calculus) can be applied to derive an instantiated version. Hence there is a notion of a model class.



**Fig. 4.3** A model class, or a number of classes, is usually packaged into a functional and accessible component. This could represent a solver container, e.g. an ODE package that can read and interpret an equation and solve given inputs, or a dedicated software package. The interface, i.e. the mechanism of communication and interaction, for this package may not be generic or even fully functional in practice, however it forms a critical part of how the model can be integrated with another. Indeed this interface is key to existing model integration strategies, as will be explained later in the chapter. One can see where the model integration types outlined above fit in this view of model anatomy: instantiated, or live models, can communicate their state or be interrogated via the interface(s) in what can be termed integrative communication; the uninstantiated form cannot be interrogated or communicated with in the same way - it is not solved - but, as a schematic, can itself be communicated to importing applications via a lingua franca, i.e. isolative communication. Decomposite communication should be more finely-grained in that it takes the submodels of different models and allows them to communicate under a single framework.

Chapter 2 identified four overarching modelling foci that are prevalent in the cancer modelling literature, namely growth, response, pathway and angiogenesis models. Whilst it is clear that complex systems can be partitioned into subsystems, system variables and behaviours to reduce the complexity of the modelling procedure, the literature is unclear as to how a model would actually communicate its focus to other models. This is critical from a model integration perspective. An integration must take place in the context of focus, otherwise it is unclear what the integration is actually achieving, i.e. how the integrated model relates to the actual system of interest.

Communication of focal, i.e. subsystem and behavioural, information can be understood to be a metadata problem and as such is coupled to model exchange and communication languages when model integration is considered isolative and integrative as per Definition 17. It is even more difficult to describe focus in the decomposite case since its specific solution is highly dependant on focus. Indeed the precise description of focus is essential in identifying interactors. Essentially this challenge can be distilled down to model semantics that includes information about how a model relates to the real-world - in fact focus amounts to an *in silico* description of what a model represents in the real-world. The interrelation *between* foci yields a definition of how subsystems themselves integrate together in reality (or how they are believed to be integrated) and therefore, when applied to CS, focus correspondingly becomes complex. Obviously all of these points would play an important role in any formalisation of model integration.

Concisely, the challenges raised are:

1. *How focus is communicated to a model integration environment or modeller:* if any communication mechanism, such as an exchange language, is to communicate model semantics between modelling and/or simulation environments then a formalisation of focus is required to support model longevity, reuse and extension.
2. *How focus is interpreted and processed by the integration framework:* Traditionally a model focus is implicitly given within the model equations or rules (depending on formalism) and/or explained in plain text. Model integration environments and modellers need to interpret focus so that integration can take place within the context of the real system, i.e. integration of models is reflective of integration of subsystems in the real system. Therefore, to ensure the validity of any integration, the integration environment would require a computable form of focus.

Focus is one of the first considerations when formulating a model. Secondly the limit within which the model is built to operate is hereon referred to as *model scope*. This reduces to boundary conditions on the state space and therefore also defines all allowable values for variables. This incorporates model assumptions, explained below as limitations on how the model relates to reality. Within assumptions one can identify model context, which delimits model use given defined situations. The semantics of model scope can be very complex, even for simple models, and hence it is very difficult to formalise it, which is probably why scope-related concepts are generally written in natural language and rarely tackled formally.

**Definition 24 (Model Scope).** The operational limits of the model imposed by assumptions and context.

**Definition 25 (Model Assumption).** The role of an assumption can abstractly be interpreted as a proposition that is made with the expectation that it will be discharged later<sup>6</sup>. Given in a modelling perspective an assumption, (set of)  $\zeta$ , can then

<sup>6</sup> <http://en.wikipedia.org/wiki/Assumption>

be defined as a proposition(s) relating to paradigm (e.g. assuming a system is discrete), focus (i.e. what the model abstracts as well as the actual system itself) and context (enforced restrictions) that is taken to be true pending such a time when the model can be revised so as to discharge the said proposition(s). Note that not all assumptions are dischargeable - most notably the ones that are related to model formalism, i.e. paradigm.

**Definition 26 (Model Context).** The operational limits of the model imposed by the real-world information with which the model is parameterised [69]. This includes the limits placed by model validations.

Scope, in conjunction with focus, covers what modellers frequently term the application of Occams Razor, i.e. simplification of the system to make the problem more tractable, but this can be abstracted further as follows. Let  $\vartheta$  represent subsystems, sub/system components and sub/system states:

- Assumptions on focus (ex/inclusion of  $\vartheta$  of interest):  $\zeta_\phi$ 
  - *Explicit inclusion of focus*:  $\vartheta$  whose behaviour is included in the model explicitly in the form of variables, constants and expressions.
  - *Explicit exclusion of focus*:  $\vartheta$  whose behaviour is explicitly excluded from the model; usually this is written in plain text or worked into a solver.
  - *Implicit inclusion (encapsulation) of focus*:  $\vartheta$  that are represented in the model are implicitly included to encompass finer-grained behaviours and  $\vartheta$ . This is equivalent to simplifications of  $\vartheta$ , e.g. phenomenological models that implicitly include finer-grained dynamics.
- Assumptions due to paradigm:  $\zeta_P$ 
  - *Functionality, Numeric Content and Continuity*: These three paradigms often embody simplifications that can be reduced to the other assumption abstractions that are listed here. Functional simulations (the solved form of non-functional construct) stack additional assumptions on the original model by extending or limiting scope. Non-functional models often explicitly exclude focus whilst numerically heavy models tend to be more inclusive. Continuity impacts assumptions on scope (see below). Discrete models are in fact very often chosen for that precise reason, i.e. reduction of state space [205].
  - *Stochasticity*:  $\vartheta$  behaviour is condensed to obey probabilistic laws. This, again, is a simplification of  $\vartheta$  and is especially made when finer-grained  $\vartheta$  cannot be simulated due to intractability (e.g. molecular interactions in a large system) or simply because the finer-grained dynamics are not known or cannot be explained by available data.
  - *Centrality*: Global behaviour of  $\vartheta$  is explicitly formalised and mapped to defined outputs. This can often be viewed as a type of black-boxing, which is a parallel situation to that of simplification into a phenomenological model.
  - *Dimensionality*: Spatial restrictions in relation to focus, i.e. spatial structures, are defined and are often simplistic abstractions of structures in the real

system. Such structures, including the concept of time, can also be interpreted as being properties of the solver in some cases.

- Model Scope (of which model assumptions are a subset):  $\zeta_\alpha$ 
  - *Boundary Conditions*: Definition of operational limitations of  $\vartheta$ .
  - *Model Context*: Atomically this can reduce to the definition of behaviour of a particular  $\vartheta$  (which can be approximated to a submodel) in specific conditions.
    - *Variables and constants*: Empirically deduced or estimated (and therefore with associated belief) variables and constants are assumed into the model and dictate behaviour in the same way submodels dictate behaviour of the parental.
    - *Initial Conditions*: The initial values with which the model has been run and subsequently validated. Again, this could be solver-dependant.

**Definition 27 (Black/Grey/White-box).** In the case of modelling, simulations and software systems a black-box can be defined as a functional component wherein calculations and/or modifications are made due to an input, which produces a corresponding output; aside from this, no information whatsoever regarding the functional component is exposed. Similarly, a grey- and white-box allow exposure of a certain degree or all information respectively.

The challenges model scope raises with respect to model integration are:

1. How model scope is communicated, i.e. how is scope metadata encapsulated within a model and how is that metadata extracted and communicated to other models in a standardised fashion?
2. How model scope, specifically model assumptions, is processed.
3. How an integration platform reconciles differences in model scope.
4. When two models of different scope are to be integrated, how differing or contradictory information is reconciled.

**Definition 28 (Scope Processing, in model integration).** The computational administration by which an integration platform can expose the plausibility and rationality of specific model integrations on the basis of comparative scope.

The implication here is that a model integration platform must provide mechanisms to interpret model scope and process them such that the integrated model does not violate either of these concepts. Concurrently, as with model focus, it is necessary that the appropriate communication mechanism can express these constructs in a formal manner.

#### 4.2.1.3 Model Scale

Biological systems operate on a massive scale difference from molecular through to whole-individual dynamics, both temporally and spatially. Modelling at different

granularities gives rise to the concept of so-called phenomenological and mechanistic models as described earlier. Models therefore inherently assume a range on the model focus, where range can be defined as a collection of predefined connected foci that span temporal and spatial scales. Herein lies one of the most difficult challenges for model integration strategies. The ultimate challenge is one of tractability: How does a model integration platform integrate models of vastly different scale without incurring a virtually impossible computational cost?

In terms of simulations the multiscale challenge is a universal one that can be found in fields as diverse as the materials sciences [94], environmental sciences [153] and geophysical sciences [96]. Of course the simplest solution is a brute-force approach where the workload of small-scale models is disseminated in a distributed computing environment. However, as illustrated in Chapter 2 by complex small-scale models, even this approach has its limitations when the chosen mesoscopic scale is comparatively large. So-called model abstraction methods (not to be confused with the abstraction of models described earlier) are the most direct solutions to the multiscale problem where the simulations are manipulated to procure faster output without invalidating the model. The objective is to reduce the complexity of the model by re-conceptualisation [71]. Often this means that sets of functional sub-modules are amalgamated into a simple black-box that approximates the original sets behaviour. Alternatively whole functional modules may well be deleted from the system altogether to reduce load. Of course increasing efficiency by this process comes at the expense of accuracy and mechanistic insight.

Model abstraction has evolved into a relatively recent technique called metamodelling<sup>7</sup> [155] where, essentially, the same principles hold but more formalisation is brought to building models of models (hence the name). Generally a set of equations is sought that transform the inputs to outputs in an efficient calculation. With highly nonlinear systems this is often a difficult task to accomplish, especially when considered in the context of biological complexity where the number of interacting components is enormous and a small number of components, e.g. a small number of gene-regulatory protein molecules, can have a substantial effect on behaviour.

Formally a simulation program or model can be simply expressed as:

$$\mathbf{z} = f(\mathbf{x}, \mathbf{r}) \quad (4.3)$$

where:  $\mathbf{x}$  and  $\mathbf{z}$  represent vectors of inputs and outputs respectively;  $\mathbf{r}$  is a data (or random number, for stochastic models) stream to drive the simulation;  $f$  is the simulation function or model.

Given the above, a metamodel can be expressed as:

$$\mathbf{z} = \mathbf{f}_s(\mathbf{X}, \mathbf{q}) + \xi \quad (4.4)$$

---

<sup>7</sup> This term is also used for other concepts in systems biology. In this book the term metamodelling will only apply to the multiscale solution version of the term unless otherwise stated.

where:  $\mathbf{f}_s$  is a vector of simpler transformation functions whose forms are generally unknown;  $\mathbf{X}$  is the vector of independent variables, e.g. time, and explanatory variables as defined in  $\mathbf{x}$ ;  $\mathbf{q}$  is a vector of new and/or stochastic variables, treated as abstract explanatory variables, that is yet to be determined;  $\xi$  is an error function imposed by simplification and should be equivalent to  $\mathbf{f}_s - f$ .

Since there are so many unknowns various forms of evolutionary algorithms and statistical methods have been prevalent in this area of research. The form of  $f_s$  is often chosen to be some spline, surface response methodology, radial basis function, Cartesian-like correlation model or lookup table [22]. Metamodelling at this stage becomes a case of optimisation and curve-fitting. Essentially, if successful, a metamodel simply presents as a phenomenological model.

With the introduction of variables that do not relate directly to the system,  $\mathbf{q}$ , it has been shown that metamodelling can reduce model complexity and increase simulation efficiency with a satisfactorily small error and has already been tried and tested in fields such as the environmental sciences [103]. However it must be said that none of these applications have tackled natural complexity in any convincing manner. In fact, to the best of the authors knowledge, model abstraction has never been successfully applied to any biological system in any serious way. A strong reason for this simply comes down to accuracy and efficacy. Indeed, Hetherington (2006) [85] confirm that simplifications in this manner with respect to liver systems modelling can potentially lead to quite erroneous output and therefore verification techniques such as sensitivity analysis is key in the validation process.

Another interesting concept that shows promise in the multiscale simulation problem is periodic simulation averaging. Patch Dynamics (PD) was first developed by Kevrekidis (2003) [102] and bridges the scale gap by averaging the effects of pulsed simulation runs and interpolating between pulses (or patches). Note that PD is another form metamodelling, i.e. a less intensive model has replaced a more complex one by way of interpolation.

Cross-scale modelling has been a subject of research for some time in systems biology, although it must be said that all too often integration of molecular data, e.g. microarray experimental output, with macroscale models have been labelled as multiscale models. Whilst such modelling efforts no doubt have their place, in the opinion of the author such labelling is a misnomer.

The momentum with respect to tumour modelling has been relatively gradual. Zhang (2007) [207] investigate the macroscale effects of a molecular-based cellular decision mechanism, which gives rise to a migration/proliferation phenotype, with an ABM formalism. With simple expressions for a whole range of stimulatory and inhibitory factors, e.g. TGF, nutrients levels (including diffusion terms) and space constraints, which are parameterised from experimental data the group achieve a good resolution of behavioural dynamics. Fundamentally it is a bottom-up simulation that bridges molecular/cell scales with application of state dynamics to lattice sites, i.e. instead of modelling individual molecular dynamics the behaviour is modelled with simple state-changing rules that simulate molecules *en masse*.

### 4.3 Model Integration: Previous Work

In the above sections the fundamental nature of models in the tumour modelling and systems biology literature has been discussed to show the challenge model integration really presents. Previous attempts at addressing the problem have been met with varying degrees of success. These attempts will be discussed before an alternative and novel solution is presented.

#### 4.3.1 *Model Integration in Management and Environmental Sciences*

Model Management Systems (MMS) and Integrated Modelling Environments (IME) were seen as extensions to database management technology (DBMS)<sup>8</sup>, and perceived as modelling parallels to the data integration problem [62]. MMS applications were required to enable the modeller to build up models from smaller components in the same way DBMS performed data joins.

**Definition 29 (Model Integration).** (Corollary to Definition [17]) A protocol that outlines how mathematical or computational components, being functional systems in their own right, are structured coherently together by a formal criterion to build a more representative model of the focus system.

Some of the most well developed fields in this regard are operations research and management sciences (OR/MS) and decision support systems (DSS). In his seminal paper on integrated systems modelling and a further two publications a year later Arthur Geoffrion outlined, to the best of the authors knowledge, the first formal theory of model integration, called Structured Modelling (SM) [78, 79, 80]. Although targeted at OR/MS in particular, and more focused on schema-like integration, the theory of models and graph visualisation of components are to some extent transferable. Uptake had initially been slow but recently SM has received at least some renewed attention from outside the OR/MS community, e.g. the environmental sciences [118].

The formal SM theory consists of 28 definitions, pertaining to formalised components of models and components that emerge from integrations, and 8 propositions, pertaining to properties and rules of the said definitions, to establish a semantic depiction of a model on three descriptive levels (elemental, generic and modular), which is summarised in Figure 4.4. SM is based on graph theory and therefore captures mathematical modelling by means of components (nodes) interacting (arcs) with each other. Nodes typically either represent entry points for other graphs or a possible link to an alternate model and graphs are further associated with a calling sequence. Model integration thus becomes a task of identification of synonymous nodes between graphs, or possible intermediate-node conversions, and determination of subgraphs that subsequently become obsolete.

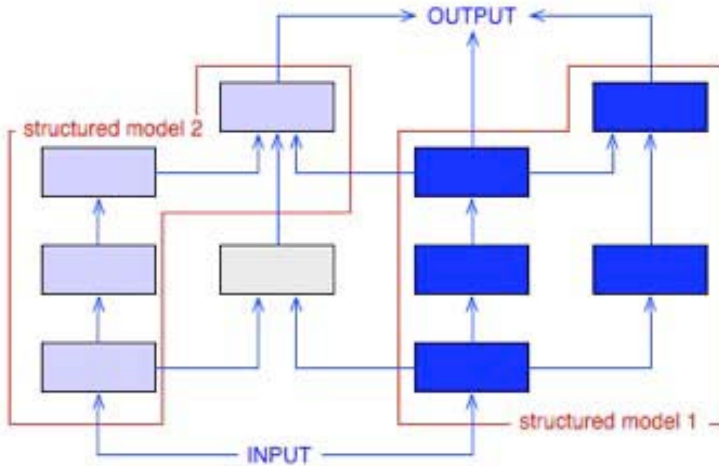
---

<sup>8</sup> For a good review of the parallels between MMS and DBMS the reader is referred to [179].

Geoffrion identifies four types of model integration, corresponding to four classifications of model entities:

- Model Consolidation or Aggregation (integration of model instances)
- Model Class integration
- Model Paradigm Integration
- Model Tradition Integration

A model class is synonymous with the abstraction made earlier wherein a model can exist in an unsolved form. Correspondingly the solved form is defined as a model instantiation, where the constants and variables (except for parametric and dependant variables) are specifically instantiated. The concepts of class and instantiation here have been imported from Object Oriented Programming (OOP). Model paradigms in SM are defined as being a collection of similar model classes, e.g. exponential growth models, and model traditions are defined to be a collection of model paradigms that tend to be associated with one another in a particular field of research. As would be expected, the integration process becomes progressively more difficult as one proceeds through these layers of abstraction listed above.



**Fig. 4.4** In the SM paradigm structured models can be decomposed into elemental structures. This includes O/P/SDE models. There are 5 basic elements: primitive, compound, attribute, function and test (see [79] for full explanations). Elements can be connected by relation via the calling sequence, which is the sequence of calculations given inputs, and may well require custom-made glues (grey box) or translational elements for each integration. This yields an acyclic graph model. Critically, these individual elements in the IME must correspond to some ontology. An intelligent IME can then begin to integrate given models by relation. Once these elements have been defined, it is possible to view the model in three conceptual levels of abstraction: elemental, generic and modular (progressively through these layers of abstraction the model graph becomes simpler as nodes are collectively grouped).



Note that the input/output connectivity of the graph conforms to the original abstract definition of model integration since all four integration types listed above can clearly be classified as either isolative or communicative integrations. In fact SM represents a linear integration strategy, which is the most prominent form of model integration found in the literature. Though the definitions of model integration as listed above are useful as a rudimentary classification a cross-paradigm, focus and scale integration is not forthcoming from the SM strategy. Indeed all subsequently published papers on SM use, most notably restricted to the fields of management and environmental sciences [37,62,117,118,148], fail to show SM as an integration framework that can tackle anything more complex than linear systems. Although SM can produce complicated mathematical models a formal theory of model integration with respect to complex, i.e. non-linear, systems remains as yet underdeveloped [107].

Dolk (1993) [64] identifies further integration types:

- Schema
- Solver
- Environments and modelling systems (i.e. model/simulation software)

Schema, and indeed data, integration are problems that have been extensively researched by the database community and is still ongoing. The fact that solver integration has received some attention in the literature is actually very revealing. Most of this research was conducted when affordable computing and accompanying software was only just emerging and procedural thinking was prevalent over modular OOP-like design. Therefore software packages that performed specific model executions came integrated with custom-made solvers that were quite inseparable from the original application. Good software design in OOP nowadays allows one to logically separate components, i.e. solver from model, so that a more appropriate solver can be used for the integrated model.

IME systems take solver integration further, treating them as modelling or simulation engines that can be used interchangeably within a single or multiple models, but not necessarily at the same time [72]. Abstracting away from these practical issues, solvers can be further defined as algorithms that address the solution of a particular model class (defined as in SM). The third point above, integration of modelling systems, will be discussed shortly in terms of software integration. Again, note that all these integration types, as with SM, can be classified neatly into the earlier definitions of isolative or integrative communication.

Mathematical modelling and simulations have played a crucial role in the environmental and ecological sciences especially due to strong public, political and economic interests in climate change and as such notable attempts have been made at model integration. Argent (2004) reviews the impact of MMS in the environmental sciences in great detail and outlines a number of desirable features that are conducive to model integration, which amongst others include a framework that supports component-based model building, automated service discovery, a well-developed, intuitive GUI and a rich execution (i.e. solver) library. Metadata that should be encapsulated with model subcomponents, as it addresses the fundamental

problem of semantic compatibility of models as well as the technical aspect (such metadata could, for example, include information about model focus and scope). As a result of this research a number of IME systems have been developed specifically for the environmental sciences, one of the most recent being COINS [153]. The COINS methodology is quite invasive: a model or simulation is considered to be a package of code and with an appropriate translation layer between the integration platform and all other models and simulations COINS attempts to logically separate mathematical equations from the running code to produce a more compact and efficient single package. Of course such a platform requires an appreciable amount of human guidance since complex packages may well be impossible to integrate by this methodology.

Similarly Villa (2000) [188] propose a model integration solution based on a server/client architecture where control of simulations and independent models are distributed over a number of machines and coordination is handled by a common simulation interface via a shared language (communicative integration in Definition 17). Because a common interface and language is used the integration solution is screened from model implementation and formalism (i.e. how the model is being run or solved in its original programming language and therefore this can be considered a kind of cross-paradigm integration). The core of the model integration problem then becomes one of generating bindings for every desired integration.

Environmental sciences model integration strategies therefore essentially reduce to a software integration problem [58, 187]. In this respect a number of solutions to model integration exist that vary in depth and automation. Multi-Simulation Interface (MSI) is an open-source middleware for simulation communication (i.e. integration by synchronisation) and even message passing interface (MPI) standards can be regarded as an integrating principle in this respect. Taking examples from systems biology, SBW is a service-oriented architecture (SOA) in which multiple applications can communicate and coordinate service requests. CellML and SBML are established model exchange languages in systems biology, something that OR/MS lacks, and provide a pipeline for communicative integration between live software components.

### 4.3.2 *Model Integration in Systems Biology*

The Physiome and Integrative Biology projects, described in the previous chapter, and various other *in silico* biology efforts, have yielded some encouraging results with respect to model integration technologies. One of the challenges stated earlier was that of model representation, which still poses a hindrance in fields such as OR/MS. Fortunately the systems biology community has been eager to accept formalisations, as seen with the success of SBML and CellML<sup>9</sup>. Although SBML

---

<sup>9</sup> The CellML repository, <http://www.cellml.org/examples/repository/>, contains over 200 models and a number of SBML repositories, including EBIs BioModel, have been developed.

can theoretically be used as a generalised exchange language it has been primarily used as a representational language for pathways along with other formalisms such as BioPAX [173], whereas use of CellML has been more macro-systems oriented. CellML addresses some of the communication problems listed in previously: model semantics in the form of RDF<sup>10</sup> and mathematical components in the form of MathML<sup>11</sup> can be incorporated within it to describe mathematical relationships between variables [114]. However the mathematical capability is thus restricted to the MathML standard, which does not clearly delineate a definitional representation for many methodologies such as ABMs, CA, neural networks, etc., though the creators are acutely aware of that deficiency and are working towards addressing it (Poul Nielsen, University of Auckland, personal correspondence). It is however well suited to representation of differential equation-based models. The hierarchical structure of CellML is modular so as to specifically expose components that are essentially exchangeable or connectable and is thus very similar to the graph-based SM methodology.

Finkelstein (2004) [69] take a similar approach in describing models that addresses the challenges described in this chapter. The nature of models is taken a step further with a tentative delineation of an ontology of characteristics<sup>12</sup>. This ontology provides the beginnings of a broad formal framework for model description embodied by the development of Composite Model Description Language (CMDL) [121]. Crucially, the framework defines model integration consistency on the basis of constraints placed by assumptions, context, focus and interpretation. This will be discussed in more detail in Chapter 7.

Takahashi (2004) [177] provide a more mechanistic approach to model integration, packaged into a software bundle called E-Cell. A meta-algorithm for an integrated model is generated by three components: a *emphdata* structure, which defines informational organisation of a model; a *driver algorithm*, synonymous with SMs calling sequence, defines interactions between submodules; a *numerical integration* algorithm (if appropriate), which defines how state variables are updated (and as the name suggests it is usually a differential equation solver). In the meta-model algorithm mathematical models are defined as vectors of state variables and a set of steppers are defined. Steppers further define processes, which can update variable references via a transition function. Steppers additionally reference local and global times steps, a step function and a vector of interruption methods, which can interrupt step functions of other dependent steppers. A stepper is essentially a computational subunit of a model; therefore an integrated model is ultimately defined as a novel combination of dependant steppers akin to the graph structure of subcomponents represented by CellML and SM. The models are executed in time-steps in a discrete-event simulation. This implementation is elegant but falls short of some important attributes that would define it as a framework for model integration. It is

---

<sup>10</sup> Resource Description Framework, <http://www.w3.org/RDF/>

<sup>11</sup> <http://www.w3.org/Math/>

<sup>12</sup> The term *aspect* in this ontology is synonymous with the term *focus* in this text. Similarly *engine* is synonymous with *solver* in this text. All other terms in the ontology are directly mapped here.

firstly not clear how models of different paradigms, except for discrete/continuous models, can be integrated. Time and spatial scales are dealt with explicitly by a brute-force method (i.e. merely increasing computational power), though some optimisations are possible. Importantly, the semantics imposed by model focus and scope is not explicitly addressed.

## 4.4 Summary

The challenge of model integration is clearly defined by highlighting the diversity of models. The origin of this diversity can be found in the differences in formalism, i.e. model paradigm, the processes and behaviours that connect the model to reality (focus) along with associated limitations (scope, context) and complex assumption semantics. Moreover models operate on a wide range of spatial and temporal scales.

The state of the art in terms of model integration is embodied by linear strategies in which the outputs of one model are fed as inputs into another. Structured modelling goes further to classify nodes of the chaining graph so as to encourage model growth by providing, for example, translational functions. In effect current techniques in systems biology attempt to do the very same, however it is apparent that there are at least two views of model integration in this regard - isolative and integrative communications - i.e. models are either communicated by description languages such as SBML or models are run (solved) with states communicated by a middleware component.

## 4.5 Conclusion

Model paradigm can be described as a novel 6-tuple. The definitions of focus and explication of scope, context and scale yield an indication of how difficult a model integration exercise can become when these semantics vary from model to model. The implication is that the model integration strategy sought in this project will need to tackle these challenges completely. A decomposite form of model integration is therefore described in the next chapter.

## Chapter 5

# Novel Strategies for Integrating Models into Systems-Level Simulations

**Abstract.** In this chapter the motivation and need for model integration in the cancer systems biology field is explained. We seek to re-formalise linear integration (in contrast to previous formalisations in the literature) to give a context to the novel model integration strategy developed in this project. A novel agent-based model integration strategy, which strives to solve the cross-paradigm integration challenge, is formalised. Furthermore, a novel knowledge-based approach, which strives to solve the cross-focus, scope and context challenges, is formalised.

### 5.1 Motivation

Cancer has been dissected from molecular components to clinical phenomenology and so there exists a wealth of data to support the development of understanding of the system at a broad range of levels. Though this has provided the impetus to develop more intricate and complex models in virtually all aspects of the disease there remains a need to produce simulations that provide tangible clinical value, i.e. models that can actually be used as regular components of clinical practice [93,185]. There are many models that can predict various subsystem behaviours, or individual processes such as signalling networks, but there is a lack of models that can accurately predict tumour behaviour as a whole system. Crucially, whilst in the past biologists have focussed on reductionist views of biological systems, though sub-modules of the system are inextricably connected, a paradigm shift is occurring such that component interactions on the microscopic scale within the context of the entire system are becoming the focus. This shift provides the impetus for generation of technologies wherein mathematical and computational models of submodule behaviour and properties are aggregated to function collectively, i.e. model integration. The following points outline why an integrative approach to constructing whole-system models is useful, indeed critical, to cancer systems biology:

1. Whilst a subsystem model can predict the behaviour of its respective focus there is currently no standard formalisation that defines how such models can be integrated aside from linear integration efforts (discussed shortly).

2. A cross-paradigm, cross-scale integrative approach could yield models that combine the strengths of the individual parts, e.g. Stephanou (2005) [170] illustrate a custom algorithm that combines the expressive power of an ODE model describing diffusion of TAF and fibronectin uptake with the autonomic advantages of an individual-based model describing EC movement.
3. A cross-focus integrative approach will yield a far more representative model of the system, which is needed to gain a better understanding of the system as a whole, rather than depending solely on a reductive understanding.
4. Whilst models have traditionally been built on the wealth of empirical data available, a more representative cross-focus model will present the opportunity to discover new mechanisms *in silico*, i.e. reverse engineering.
5. In the wider scope, validated integrated models could provide the predictive power needed to eliminate, or at the very least assist, the preliminary steps of the research cycle.
  - a. From a pharmaceutical perspective this means that preliminary experimentation that cost time and resources can be substituted or assisted, saving on such costs. The overwhelming advantage in this respect is the opportunity to accurately predict, and indeed discover, possible drug targets and dosimetry strategies and could even move therapy closer to individualised medicine.
  - b. From a research perspective this affords the opportunity to test theoretical models of mechanisms. Such *in silico* experimentation will aid in advancing the understanding of the system.
  - c. The tumour system is particularly difficult to observe *in situ*, i.e. in the clinical case, hence much of the data comes from alternate sources, e.g. *in vitro* vascular tissue [120, 10]. It has been shown that mathematical models and simulations can provide a substitute in this kind of situation [170]. An integrative approach aims at establishing models that present such a substitute to close the gap that technological constraints have imposed.

Moreover such an integrative approach is a universal requirement to virtual physiology. These motivations by themselves justify a very in-depth look at the feasibility of model integration technologies.

One can see from the arguments in previous sections that the first two types of model integration outlined in Definition 17 are prevalent over the third in all fields of modelling, which is not a surprising finding when considering the fact that integration has been tackled from a cross-paradigm, focus, scope and scale perspective to only a limited degree (this is also very much the case for systems biology). Isolative integration does not require one to be concerned about cross-scale or any other solver/running-related problems since an exchange language defers this duty to the software that will be performing specific integrations; an exchange language does, however, need to be syntactically rich enough to cope with diversity of model types and the complexity of focus and scope. Similarly integrative communication strategies require a translational tool between running models to cope with this complexity and are therefore not particularly concerned with how individual models are solved; however, cross-scale problems become a pertinent issue and the integration

platform must provide a specific mechanism to cope with this as well as focus and scope.

Any integration scheme needs to be able to address the challenges and questions stated previously. With respect to the three forms of model integration given in Definition 17, isolative communication is tantamount to a lingua franca and to a large extent CellML, SBML and CMDL suitably cover this and so this type of integration will not be addressed directly in this project. Similarly integrative communication takes model integration into a more middleware development setting and has been suitably tackled by software infrastructures such as SBW.

A decomposite model integration strategy strives to take a truly mechanistic stance on model integration wherein all of the challenges must be tackled directly. The following sections detail a novel formalism called Agent-Based Integration (ABI), which embodies such a strategy. Firstly the linear integration strategy, which forms a major component of most if not all current model integration techniques, is reformalised from previous efforts to provide a clear context to ABI. Both the linear and ABI formalisms will be augmented with a novel view of focus and scope that addresses these challenges directly. Associated software engineering, validations and simulations are described in subsequent chapters.

## 5.2 Novel Model Integration Formalisations

### 5.2.1 Linear Integration Strategy

Following an analysis of the existing integration technologies, especially those of SM and CellML/E-Cell approaches, it becomes apparent that there is a common integration feature between them. It can be shown that all models are composed of smaller components that are interrelated by a calling sequence. Formally:

#### Theorem 2

$$M_\phi : \{m, c\} \quad (5.1)$$

$$m_k = f_k(\mathbf{x}) \quad (5.2)$$

$$\phi \in \Phi \quad (5.3)$$

where  $M_\phi$  is a model with focus  $\phi$ ;  $m$  is a set of  $n$  submodels indexed as  $k$ . Importantly, submodels also describe subfoci, which is encapsulated by  $\phi$ ;  $c$  is a calling sequence on the submodels; when  $f_k(\mathbf{x})$  represents a non-functional component,  $c$  becomes redundant;  $\mathbf{x}$  is a vector<sup>1</sup> of  $v$  dependant and non-dependant variables and constants in  $M$ ;  $f_k(\mathbf{x})$  is a function of  $\mathbf{x}$  yielding an output vector  $\mathbf{z}$ . Any  $f(\mathbf{x})$  need not incorporate  $\mathbf{x}$  entirely;  $\Phi$  is an (as yet) ambiguous superset defining all foci;

This expression means that model integration as discussed so far reduces to a linear integration solution where inputs from one functional unit are provided by the

<sup>1</sup> A vector is hereafter treated as an ordered set when appropriate in subsequent explanations.

outputs of another, which can be represented simply as a directed graph, the arcs of which denote the calling sequence. Note that the calling sequence need not be serial. In fact for complex integrations the implementation will almost certainly be parallel, e.g. see [121]). It is likely that such complex integrations, especially those that cross large spatial and/or temporal scales, would be run on clusters of computers amongst which whole sections of the calling sequence has been divided - linear integration readily lends itself to this type of parallelisation. This is the mathematical counterpart to software integration, which is the traditional solution to model integration and remains to be so. Indeed looking to SOAs such as SBW,  $M_\phi$  can be compared to services that are chained together. Essentially the solution to the model integration problem has distilled to one of communication between services. The applicability of this integration has already proved to be computationally efficient [63] and black-boxing of components in this way is conceptually very simple.

Linear integration is a partitioning of components in  $\Phi$ , and subsequent reconnection, such that the integrated model minimises an error function  $\xi$ , which has the relation:

**Theorem 3**

$$M_\phi \pm \xi_M \mapsto \psi \quad (5.4)$$

where  $\psi$  represents the real system of interest.  $M$  therefore represents an abstraction of the real system given a defined certainty or belief as per Definition 16. For clarity the error function will be left out in subsequent expressions.

The goal of model integration is to achieve better *in silico* representations of the real system:

**Theorem 4**

$$\Xi : \{M_\phi, C\} \quad (5.5)$$

where  $\Xi$  is a fully integrated model defined by the supersets  $M_\phi$  and  $C$ ;  $M_\phi$  is the superset of all subcomponents connected by  $C$ ;  $C$  is the superset of all calling sequences on  $M_\phi$ ;

And ideally:

**Theorem 5**

$$\phi \rightarrow \Phi : \xi_M \rightarrow 0 \quad (5.6)$$

$$\Xi \mapsto \psi \quad (5.7)$$

$\Xi$  represents a model that can perfectly predict the dynamics of the real system. Clearly this is not a practical formulation. Even if the said supersets are known the underlying assumption is that the original models can be decomposed into functional components and the integration of all calling sequences can connect these subcomponents together. Whilst algebraic formulae, e.g. O/PDE, may well be decomposable in this fashion, it is not entirely clear how other formalisms such as CA can be decomposed. Effectively linear integration does not fully address cross-paradigm integration. Moreover the integration is constrained by the number



of connectable components. Though functional units can phenomenologically and efficiently represent microscale models, multiscale integration is not specifically addressed.

### 5.2.2 *Agent-Based Integration*

The specific requirement in tumour systems modelling is a model integration strategy that does not restrict paradigm use and is able to perform cross-paradigm and cross-scale integrations without violating the semantics of focus and scope. Additionally an ideal solution will be able to map integrations into a visualisation, such as a 3D simulation visualisation, since analysis from a biologist's perspective is not just numerical but also very much visual. Such information can greatly aid the research process. A 3D visualisation provides a real-world perspective of what the model represents and predicts and therefore renders the user with a greater understanding of the results. The Agent-Based Integration (ABI) strategy, along with other mechanisms described here, is designed to address these needs.

**Definition 30 (Computational Common Ground, for models).** A mutual shared policy under which models can communicate. Isolative and integrative communication strategies use the communicative language, e.g. CellML or MPI, as a common ground for models.

Thus far model integration strategies have tackled the integration problem by directly, and sometimes intrusively, interacting with various components to achieve some level of assimilation under a formalised framework. In particular, the extent to which linear integration can be performed is largely dependant on the model paradigm, i.e. whether the model can be compartmentalised and subsequently linked. Here it is argued that it is not the model as much as what it represents and its behaviour that is important and therefore integration should take place on this level. Whilst in OR/MS, where the nature of modelling is sequential and to which linear integration is very much suited, the cancer systems biology field presents models that represent processes that run in parallel. In such a situation where paradigms can be very different one must look beyond integration at the level of paradigms. Since it is the behaviour of the models that is ultimately of value, at least in the case of cancer research, a model integration strategy could instead look to using behaviour as the subject of integration. Before one can do this a common ground must be established, i.e. a formalism in which all models can interact regardless of paradigm, focus and scope to which models can be decomposed. Zeigler (1993) [205] in his DEVS application endorses such decompositions when models can be converted through some standardised translation to alternate formalisms that can be interrogated in a discrete event simulation. However such a method presupposes that such a translation is possible and there is a potentially substantial loss of information when many such translations are chained together. Instead the ABI strategy takes the common ground to the level of an *in silico* metaphor of the system in the form an ABS. Individuals within an ABS can reflectively exist as their real-world counterparts where the common ground is defined as their interactions. To clarify, it is

proposed here that individual models can superimpose their behaviour in the form of local interactions onto a single ABS simulation such that the ABS coherently expresses the combined behaviour of the original models. Note that such a superimposition breaks the barrier of paradigms and distils it to a discrete space-time setting in addition to dealing with focus intrinsically (in the form of interactors). This method will be explained in detail below. Also note that linear integration cannot offer such a common ground that is blind to the nature of the model whilst sensitive to its behaviour.

Individual-based systems provide an excellent computational and controllable metaphor for the real system to which other formalisms can be made to adapt to make this grounding possible. Abstractly:

**Theorem 6**

$$\Lambda : \{A, R_A, S\} \quad (5.8)$$

$$\Xi_A \mapsto \psi \quad (5.9)$$

where  $\Lambda$  is a local interaction-based system (e.g. ABM, CA or a hybrid);  $A$  is the set of all local interactors, e.g. agents or CA cells, and does not have to be static;  $R_A$  is the set of all rules on  $A$ , where  $\alpha \in A$  and  $r \subseteq R$  and  $r$  is an ordered set on  $\alpha$ , hereon denoted as  $r_\alpha$ ;  $S$  is the superset of all states (in/finite, discrete/continuous) and state transition functions hold.  $\Xi_A$  is the integrated model as represented by  $\Lambda$ .

$\Lambda$  can be considered a suitable grounding for another mathematical or computational model  $M_\phi$  under formalism/paradigm  $\lambda$ , assuming  $\Lambda$  simulation behavioural dynamics is considered over discrete time ( $\tau \in t$ ), if and only if the following is computationally viable:

1. **Theorem 7**

$$\forall \tau \in t : M_{\phi,x} \xrightarrow{\nu} \Lambda_\tau \quad (5.10)$$

- a. Within an acceptable predefined bound of error  $\forall \tau$ .
- b.  $\nu$  is a translational mapping function from  $M_\phi$  to  $\Lambda$  and is ideally symmetric.

2. **Theorem 8**

$$\forall \tau \in t : \lambda_{tau} \xrightarrow{Y} \Lambda_\tau \quad (5.11)$$

- a.  $Y$  is a (set of) translational mapping functions to map  $\lambda$  to  $\Lambda$  and is ideally symmetric.

$h$  is defined as the mapping function that enables Formalisation [7](#), and more generally Formalisation [2](#), where:

**Theorem 9**

$$A_{t+\delta} = \left( \forall s_i \in S : s_{i,t} \xrightarrow{r_i} s_{i,t+\delta} \mapsto \Lambda_{t+\delta} \right) \quad (5.12)$$

$$M_{\phi,t+\delta} = \forall m_k \in M : \mathbf{x}_t \xrightarrow{\partial f_k(\mathbf{x})} \mathbf{x}_{t+\delta} \quad (5.13)$$

$$(\Lambda_{t+\delta} \wedge M_{\phi,t+\delta}) \mapsto \Psi_{t+\delta} \quad (5.14)$$

$$\partial f_k^{global}(\mathbf{x}) \equiv R_{t+\delta}^{local}(S) \quad (5.15)$$

$$M_{\phi,t+\delta} \xrightarrow{h} A_{t+\delta} \quad (5.16)$$

This means that every incremental change in interactors with respect to state changes (including spatial orientation), indexed as  $i$  above, is due to the rules associated with the interactors, as per the definition of individual-based systems, and that this change in state(s) contributes collectively to the overall behaviour of  $\Lambda$  (Formalisation 7). Moreover, this state change must be reflective of the corresponding state changes given by the original model  $M_\phi$  and, more atomically, the submodels that are included in the integration (Formalisation 2). Note that essentially this formalism is a discrete solution. Physically this is but a natural implication since all agents within the real system, e.g. biological cells, molecules, etc., are indeed discrete, though discretisation need not be absolute (e.g. molecules can still be measured as volumes or concentrations, which is stored simply as a state). Though time discretisation is not an absolute necessity it does make the formalisation simpler. Additionally, discrete simulations are known to be more efficient than continuous<sup>2</sup> varieties since, generally, they require less computation [182]. The global<sup>3</sup> functions that  $M_\phi$  defines in discrete time (or continuous time as  $\delta \rightarrow 0$ ) map ambiguously via  $h$  to local interaction rules  $R$ , which in turn invoke state transitions. Indeed if  $\Lambda \rightarrow \Psi$  and  $M_\phi \mapsto \Psi$  (Formalisation 5.8iii), assuming the formal definition of  $\Lambda$ , then there must be some  $R^{local}$  that satisfies the mapping, which is reflective of the original system. To create grounding  $\Lambda$  then becomes a matter of finding  $h$ . Fundamentally this formalisation is only legitimate if  $M_\phi$  is indeed considered to be faithful to the real-world system, i.e. it is validated. This assumption reduces the complexity of dealing with contradictory dynamics contributed by different models, though this issue is still resolvable to some extent (explained shortly).

Any formalism, i.e. any mathematical model or simulation that is to be integrated within this framework, must be able to map at least approximately (i.e. within given error bounds) in incremental time and space to  $\Lambda$  whether or not the original model even includes the concepts of time or space. This in turn will inherently depend on the formalism of  $\lambda$ , i.e. the model paradigm. The properties of these formalisations with respect to (sub)model paradigms are summarised below.

*Functionality:* To the exclusion of those models that are specifically solvable by some solver technology, e.g. an ODE that can be solved by the Euler method, in the case of non-functional models  $\partial f_k^{global}$  becomes obsolete since there is no operation

<sup>2</sup> In reality, however, computational representations of numbers are actually always discrete.

<sup>3</sup> Note that though these functions are labelled *global* they need not necessarily be so if  $\lambda$  happens to represent a local interaction system (or other mechanistic system) already, in which case  $h$  represents a translation of rules (given spatial and temporal dimensions) from the original system to the current framework.

to execute that readily maps to  $\Lambda$  and therefore  $h = \emptyset$ . However non-functional models do contain some information about the target system. For example, even a simple hierarchical model of biological structures in the form a directed graph still directly maps structural information about the system. Such information is indeed translatable to  $\Lambda$ , if not to determine any  $R$  then at least to identify interactor types. This kind of information is applicable to the knowledge-driven approach described in the next section.

*Numeric Content:* Quantitative models obviously hold promise, assuming they are functions, as there is likely to be some transformation function between inputs and outputs by which  $h$  can be defined. Qualitative models would be more difficult, especially if the purpose of the model is meant to be purely abstract. Otherwise there may well be precedent for quantifying these models, e.g. application of fuzzy membership functions [198].

*Continuity:* Assuming functionality and numeric content, a continuous model can be considered a special case of discrete mathematics where  $\delta \rightarrow 0$ . A continuous system should therefore have the capacity to be queried at distinct variable (e.g. time) points. In the case of P/ODEs, for example, one can assume the solution given boundary/initial conditions along a mesh (finite element method). In the decomposite ABI form any intermediate values then would have to be determined by interpolation if they are needed.

*Stochasticity:* This paradigm presents a greater difficulty than others with respect to ABI. If stochasticity is considered mechanistically, i.e. on a fine scale, the degrees of freedom can be intractably large. Predicting the form of local interactions from inconsistent global behaviour presents another problem. The only way to narrow the number of possible solutions is by constraining parameters through knowledge of the system that does not come from the model directly. This could be achieved by the knowledge-directed approach discussed shortly. When global behaviour is consistent, i.e. in the deterministic case,  $h$  can be reliably tested iteratively and its derivation reduces to that of the other paradigms.

*Centrality:* Assuming functionality and quantitative numeric content a centralised method can be considered to model a defined focus at specific parametric instances, e.g. neural network-based gene expressional diagnostic classifiers or therapy optimisation methods that define (sub)system states at specific time-points. Often centralised methods as such take instantiated global behaviour and map it to another, sometimes more subtle, global behaviour or end-trajectory and therefore contain very little information pertaining to mechanistic dynamics; one certainly would not begin the integration process with a centralised model, but it may well be possible to direct behaviour so that such a model is satisfied, e.g. attenuation of local rules so that state dependencies become concordant according to a centralised model. The derivation of  $h$  therefore requires far more knowledge than what the model readily presents.

*Dimensionality:* One particular issue that must be addressed is whether any  $R$  obtained from a model via  $h$  in an  $x$ -D setting is at all applicable to a  $y$ -D setting, where

$y$  represents the dimensionality of the target simulation  $\Lambda$ . The answer to that question obviously depends on what  $R$  actually defines. When  $x > y$  the derivation of  $h$  is conceivable without the need for other information. When  $y > x$ , however, alternate sources of information may be needed to validate the integrated model. Derivations of  $h$  in non-dimensional models reduce to paradigm cases listed above.

The derivation of  $h$  is not directly related to focus, scope and scale though there are some considerations that have already been touched upon above. Firstly, the focus of a model will define the specific interactors in  $\Lambda$ , whether they are abstract or actual. This provides the basis on which models can be mapped to, i.e. where  $R^{local}$  is applied. In relative terms this is of course associated with scale and how  $\Lambda$  can be run efficiently. For example, to integrate a molecule-level model with a tissue-level model could prove to be extremely difficult since the interactors are defined on vastly different scales. Assumptions and scope impose themselves onto the rules, which will be explained further in following sections where such model semantics are incorporated into the interactors directly.

Clearly there is no generic or direct mathematical derivation for  $h$ , which is heavily dependant on the paradigm. When no direct translation is feasible it seems that an iterative process is necessary wherein local interaction rules are generated and optimised by comparing the behaviours  $\Lambda$  to  $M_\phi$  at each iterative step. Evolutionary techniques lend themselves to this approach in which an initial set of rules,  $R_i$ , which represent educated guesses of the true rule set  $R$ , can be enriched. Genetic algorithms (GA), pioneered by Holland (1975) [90], are a proven optimisation technology that has the ability to find global maxima/minima in large search spaces. In fact GA have previously been applied to CA to optimise rules to achieve global behaviour [101]. Mitchell (1996) [127] investigate the feasibility of evolving two-state 2D automata to achieve specified global behaviour and report some success though the sheer number of possible solutions, which exponentially increased as the size of the CA and number of rules increased, admittedly exceeded even the capacity of their GA implementation. This difficulty is well established in the literature however when the solution space is constrained by restricting rule numbers and forms, neighbourhood radii and state magnitudes global behaviour optimisation is indeed achievable. Colonna (1998) [48] demonstrate this by learning CA rules that simulate coordinated social and vehicle movements although any CA learning on the level of biological complexity, to the best of the authors knowledge, has not yet materialised in the literature.

What is obvious from relating  $h$  to the paradigms listed above is that there is an explicit need for some knowledge about focus ( $\phi$ ), interactors ( $t$ ), and states ( $S$ ) to arrive at a solution as well as to limit on the number of plausible solutions. Therefore an explicit knowledge-driven approach is now proposed, which not only drives a search for a solution to  $h$  but simultaneously attempts to manage model focus and scope problems that arise in model integration strategies.

### 5.2.3 *A Knowledge-Driven Approach (KDA) for Addressing Model Focus and Scope*

A framework in which focus can be expressed will greatly aid the modeller in both understanding what models represent as well as how they should be integrated. Such a method must be able to describe the relation *between* all foci as well as each and every model focus itself. None of the model integration platforms described earlier tackle focus directly even though it forms a fundamental part of the model. In essence it is the connection the model has with reality and so development of a method that can express  $\Phi$  in a model integration strategy is highly desirable.

The ideal properties of such a method can be defined:

- High level of capture of the semantics of models (includes model focus, assumptions and scope).
- Direct and dynamic mapping of actual system components.
  - The representation should fully reflect the original modellers knowledge of the system, i.e. what is termed expert knowledge in the literature.
  - The representation should be robust to change in or introduction of new knowledge.
- High level of expression with a relatively simple representation to avoid error.
  - Flexibility in what can actually be represented is vital as complex models of complex systems will presumably go hand-in-hand with complex semantics. Moreover these semantics may well incur human error when the (graphical or exchange language) representation is difficult to understand or generate.
- Portability.
  - Preferably the knowledge that  $\Phi$  represents should be portable to alternate applications, e.g. via some shared standard.

In this proof-of-concept the representation format<sup>4</sup> itself is not as important as its underlying theoretical basis; since focus has thus far been discussed in terms of partitions of the system a set theoretic approach is taken here.

Given the current state of the tumour modelling literature  $\Phi$  is evidently non-existent and cannot be explicitly sought because of a lack of complete knowledge

---

<sup>4</sup> There are a handful of techniques that can manage focus with the said requirements such as semantic networks (SN) [122], sometimes called topic/concept maps in the literature, and ontologies realised as, for example, DAML-OIL (<http://www.daml.org/>) or OWL (<http://www.w3.org/TR/owl-features/>) (and associated knowledge-base technologies). Even though the latter are technically designed for semantic web applications, they provide a neat and well-defined, logical representational format and are widely accepted ontological technologies albeit difficult to understand syntactically. On the other hand SN, though far more versatile due to reduced formalisation, are not as portable via any widely accepted format but are relatively easy to understand and build. Purely for simplicity and versatility, as well as representing the modellers beliefs about the system, SN is the representation of choice but for portability OWL is superior.

(if indeed one can assume *complete* knowledge is even attainable). One of the major reasons for this is technological limitations, i.e. many of the fine details of tumours *in vivo* cannot be measured accurately. Moreover a realisation of  $\Phi$  would ideally require the agreement of a number of experts for a satisfactory level of accuracy as well as acceptance by the fields wider community. To the best of the authors knowledge no such organisation exists although the Integrative Biology Consortium has recently addressed that very issue. Besides this the current situation and requirements suggest that realising  $\Phi$  is therefore an unfeasible undertaking at present.

To make a solution to this challenge practical it is proposed that  $\Phi$  be expressed iteratively as  $\Phi_I$ , where  $\forall \phi \in \Phi_I \subset \Phi$ .  $\Phi_I$  can be viewed as a theoretical partitioning of the system based on individual expert (i.e. the user, modeller or panel) knowledge given a set of individual models. Simply put, as one integrates a model into another independent model (in the case of the first integration) or an existing integrated model,  $\Phi_I$  iteratively develops (tending to  $\Phi$ ) and encapsulates all foci encountered in an ongoing integration exercise.

For ease of expression it is noted that all complex models inherently contain submodels and so the sub is dropped here and any reference to model can be taken to be equally applicable to submodels in all formalisations that follow. Additionally note here that constants and variables can be treated equivalently if and only if the basic requirement of conceptualisation of a defined behaviour is fulfilled since, as will be explained below, focus will be described abstractly from a behavioural viewpoint.

Models relate content to a specific set of real-world behaviours and real-world system components, which completes the abstract definition of focus. It then naturally follows that  $I$  should be expressed in terms of real-world behaviours and how they relate to real-world objects. The anatomy of  $\Phi_I$  consists of a system behaviour, which the associated mathematical component conceptualises, and the many entities that collectively yield the entire system (or partition thereof) and contribute to the said behaviour.

Let  $\phi$  represent the focus of a model or variable, which relates a behaviour to a real-world component  $\kappa$ , or a group of real-world entities that can be coherently categorised as an individual (e.g. an organ, though made up of many components, can be categorised as an individual when given in the context of a whole-body simulation). It follows that  $\Phi_I$  iteratively builds up a hitherto faithful *in silico* metaphor of the system. Furthermore an interesting property of complex bottom-up systems is that behaviours can be further dissected into a number of more fine-grained behaviours, which themselves will be associated with a finer-grained set  $\kappa$ , etc. Concurrently, it has already been demonstrated that this is indeed an inherent assumption of simulations and models that, sometimes implicitly, metamodel finer-grained behaviour into a simple black-box. As a matter of fact this *behavioural inclusion* assumption is part and parcel of the corresponding  $\phi$  if  $\phi \mapsto \psi$  holds true. For example, a Gompertz model ( $M_\phi$ ) of tumour growth ( $\phi$ ) is a phenomenological model of a relatively macroscale behaviour (i.e. tumour growth dynamics, *phi* if  $\phi \mapsto \psi$  holds for MTS) and as such implicitly includes finer-grained behaviours such as pathway dynamics, cellular dynamics, molecular dynamics and so on. This can be intuitively

visualised as a directed acyclic graph of behavioural inclusion where nodes of the graph are specific behaviours and directed arcs denote inclusion; a node is a superset if there exists a tail arc that leads to another behaviour, i.e. a subset, and if not is simply defined as a singleton (where the one member is an instantiation, explained shortly). In essence a behavioural Inclusion Tree (BIT) can be developed.

**Definition 31 (Behavioural Inclusion).** Consider foci to be viewed in a set-theoretic manner; behavioural inclusion can be defined as identification of behaviours (sets or singletons) that are wholly (crisp) or partially (crisp/fuzzy) subsets of more macroscale behaviours that wholly or partially represent the behaviour of the subsets. Presently, for clarity, only crisp sets will be considered. Note that a behaviour (set or singleton) can be a member of a number of sets in which case it partially accounts for those behaviours without having to consider it fuzzy.

As seen in Chapter 2 there are two important concepts of focus and associated behaviour that one must define: the first is that of behavioural inclusion as explained above; secondly there exists an abstraction of behaviours, in terms of an ontological perspective, from model to model that must be dealt with directly to wholly capture the semantics. For example, there are various types of dynamics such as molecular flux, EC and other cellular motility, blood flow, and even growth itself, and within these concepts there are subtypes of dynamics that can be defined. The aim is to enable one to distil the most abstract behaviours of a biological system, and idealise even further to complex systems in general, so that the resulting framework need not be recreated for new integration exercises that may involve systems other than those thus far encountered. This facilitates the iterative building of  $\phi_I$  as explained earlier. Simultaneously the integration platform needs to define the most specific behaviours as in the BIT. A set-theoretic/OOP<sup>5</sup> approach for model focus affords this power of abstraction so that one can deal with these two concepts concurrently.

Let  $U^a$  and  $U^b$  be partitions over the universe of discourse that represents the behavioural class hierarchy, i.e. abstractions of behaviour, and the BIT hierarchy respectively. This partition is also made to highlight the difference in semantics of the subset property. In the case of abstraction the subset property is equivalent to inheritance, i.e. the usual is-a OOP relationship, whereas in BIT it denotes behavioural inclusion. Behaviours in the abstracted sense can be considered OOP-like classes whereas behaviours that models actually exhibit can be considered OOP-like instantiations,  $\theta_k$ . In this way both behavioural abstraction and BIT can be represented on the same graph. Formally:

$$U = U^a \cup U^b \quad (5.17)$$

$$\phi : \{\theta_0.. \theta_n\} \text{ where } \forall k \theta_k \in U^a \wedge U^b \quad (5.18)$$

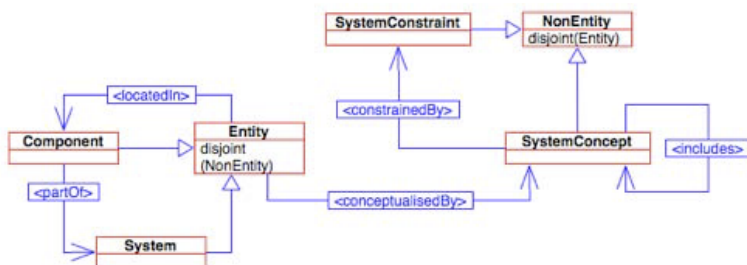
The complete model focus ( $\phi$ ) is defined in this framework as the set of all instantiated behaviours that are both part of the abstraction tree and BIT. Behaviours can

---

<sup>5</sup> Only the concepts of inheritance, classes and instantiations are considered here since these concepts are proven and powerful tools to describe abstraction and are integral concepts in the knowledge representation technologies such as SN [175] and OWL.



be collectively defined as system concepts (most abstract  $\phi$ ), which involve system entities (most abstract  $\kappa$ ). Systems concepts themselves are non-entities. System concepts subsume (include, as defined above) other system concepts to create a BIT. Additionally, to include model scope system concepts are restricted by system constraints. These abstractions are summarised in Figure 5.1 in which the set of all core classes,  $c$ , and related properties is shown. This is a novel construct in which behaviour and system components can simultaneously be described, which will be illustrated and validated in the next chapter.



**Fig. 5.1** Abstraction of the real system and relational mapping of model focus (core classes). Taking an OOP approach, focus itself is conceptualised by a SystemConcept class from which other concepts inherit and/or form a BIT. Note that these relationships also form a type of system ontology. All solid arrows without a label denote inheritance, in terms of abstraction, from parent (head) to child (tail).

Core classes are defined as:

- *SystemConcept*: Any NonEntity concept of a real-world system that describes some behaviour that can be conceptualised by a model, e.g. a biological process.
- *SystemConstraint*: Any NonEntity concept of a real-world system that constrains the behaviour of a SystemConcept. This extends the definition  $U = \Phi_I$  to include scope and context that can be symbolised by any derivation of SystemConstraint. Hence a SystemConstraint can represent logical or mathematical (but not physical) constraints. Mathematical constraints in particular are important since these include parameterisation, e.g. curve fitting, which encapsulates assumptions about the model that is otherwise difficult to formalise in a computational construct.
- *Entity*: Any physically observable element of the real-world.
- *Non-Entity*: Anything that is not an Entity.
- *Component*: Any Entity that is part of a real-world system.
- *System*: The observable real-world system composed of Components and whose dynamics are defined by behaviours, i.e. SystemConcepts.

Core properties<sup>6</sup> are defined as:

- $\langle includes \rangle$  (both domain and range is SystemConcept)
  - Equivalent to Definition 31
  - Transitive.
- $\langle partOf \rangle$  (domain is Component, range is System) and  $\langle locatedIn_i \rangle$  (domain is Entity, range is Component)
  - Refers to physical composition of a system. Biological systems in particular are dynamic and any entity may well be compartmentalised within different structures in different contexts. Component class instantiations must therefore exist for multiple contexts.
  - Transitive.
- $\langle conceptualisedBy \rangle$  (domain is Entity, range is SystemConcept)
  - Refers to a realisation of process by which a physical entity can be conceptualised by a model, i.e. a SystemConcept.
  - Transitive.
- $\langle conceptualises \rangle$  (domain is SystemConcept, range is Component)
  - Inverse of  $\langle conceptualisedBy \rangle$ .
- $\langle constrainedBy \rangle$  (domain is SystemConcept, range is SystemConstraint)
  - Refers to the scope and context of a SystemConcept.
  - Transitive.

Inheritance enables desirable properties when used in conjunction with other relations as listed above. For example, since a child class, *child*, inherits all attributes from all parental classes, instantiations of *child* will also inherit all transitive (and other) relations that parental classes possess. Due to the integration of these OOP concepts the said properties that are defined on instantiations are enforced. In other words, an instantiation of any of the said concepts *must* define the properties imposed by the abstract framework (Figure 5.1) with other instantiations of the correct type. Therefore this method, herein referred to as the knowledge-driven approach (KDA), enables description of an *in silico* metaphor of a real system to any level of abstraction within the limits of (inputted) knowledge.

### 5.2.4 Knowledge-Driven ABI: A Novel Formal Protocol for Model Integration

ABI and KDA have been formalised separately but these two methods can be used together to form a specific solution to model integration. Whilst  $\Lambda$ , together with

---

<sup>6</sup> Note that all properties are enforced on defined domains and ranges.

$h$ , represents a framework that can tackle the cross-paradigm challenge, KDA directly addresses the model focus challenge, including any scope and context that can be symbolised as a constraint, and model assumptions imposed by focus,  $\zeta_\phi$ .  $\zeta_\sigma$  can also be integrated into  $\Lambda$  in the form of rules; boundary conditions impose themselves as limitations in rules that regulate behaviour and context is similarly treated as explained previously. Since  $h$  incorporates behaviour due to paradigm,  $\zeta_P$  is also covered by  $\Lambda$ . A fusion of these two approaches is described in which the full range of challenges can be addressed. As described earlier, the key to addressing the multiscale problem is either application of the brute-force approach or some meta-modelling scheme such as patch dynamics and is therefore a problem associated more with the enabling software.

ABI establishes  $\kappa \equiv \iota$  where  $\iota \in \{\alpha_{all} \vee \chi_{all}\}$  and  $\alpha_{all}$  and  $\chi_{all}$  are supersets of all agent/cell individuals ( $R$  equally applies to  $\kappa$ ). Let  $R$ , representing the rules derived via  $h$ , be partitioned into rule subsets,  $\rho_m$ , that individually represents the rules that are direct derivations due to submodel  $m \in M$ . Simply speaking,  $\rho_m$  defines the behaviour of  $\iota$ , and therefore  $\kappa$ , as local interaction rules due to  $m$ , in the time step  $t + \delta$  to achieve a global dynamic that is consistent with (i) the original model in the same or similar time step and (ii) the real-world component  $\kappa$ , to some predefined degree of accuracy.

The  $d$  and  $\mu$  functions (as given in Table 3.1) make a choice of which rules to fire for any  $\iota$  given a local and/or global situation<sup>7</sup>.  $d$  and  $\mu$  can therefore be formally extended such that:

**Theorem 10**

$$\text{Action} : d(R, S, E, G) = \rho_m^{\text{action}} \quad (5.19)$$

$$\zeta : \{S, E, G\} \quad (5.20)$$

$$\therefore \rho_m^{\text{action}} = d(R, \zeta_\iota) \quad (5.21)$$

where  $G$  is the set of all global states of  $\Lambda$ ;  $\zeta_\iota$  is the context of  $\iota$  (embodied by  $S$ ,  $E$  and  $G$  as per the formal definition).

Model scope,  $m_\sigma$ , can be incorporated into this methodology by integrating boundary knowledge of  $m$  into  $\rho$ ,  $d$ ,  $S$  and/or  $G$  depending on what boundaries are defined. When applied to  $d$  or the rule set of  $\iota$ ,  $m_\sigma$  restricts  $\rho$  in such a way that the resulting dynamics in  $\Lambda$  will not violate the boundaries  $\sigma$  represents. Similar is the case for  $S$  and  $G$ , i.e. state dynamics are restricted so as to not violate  $\sigma$ . In fact  $\sigma$  signifies an additional behaviour for  $\iota$ , and consequently  $\Lambda$ , that  $h$  must codify into local interaction rules. In short, an ABM simulation can be built from a model given enough metadata and knowledge about  $M_\phi$  and  $\psi$  respectively. With each progressive integration, the  $\rho$  subsets must be ordered such that they are concordant with respect to focus (they represent different behaviours) and scope (application of  $d$ ), otherwise an integration of the rules is required, which is detailed below.

<sup>7</sup> For clarity only  $d$  will be considered though the following formalisations are equally applicable to  $\mu$ .

For simplicity, let the formal integration be applied to submodel  $m_1$  where only a single paradigm profile exists and  $\kappa$  and  $\phi$  (as a full KDA specification) are known. In the case of the first model that is to be represented in  $\Lambda$ , i.e.  $\Lambda$  is *naive*,  $h$  is simply invoked to find  $\rho_1^m = R_t$ . Note that though subsequent integrations yield the sets  $\bigcup \rho_m^i = R_t$  any  $\rho \subseteq \rho_m^i$  can be shared, i.e. an existing rule in the rule set attributed to a previous model integration is allowed to form any part of the solution of a new integration so far as there is no incompatibility issue.

Let each submodel, with corresponding  $\kappa$  and  $\phi$  (as a full KDA specification), that is to be integrated into  $\Lambda$  undergo this decomposition process separately within a defined bound of error. If one model conceptualises any SystemConcept that is already wholly or partially conceptualised by another model, i.e. there is an *overlap* (intersection in the separate KDA specifications of each model), or contributes any  $\rho$  to an existing  $t$  in  $\Lambda$  then a synergy must be reached between the models such that there is no contradiction in behaviour of  $t$ . In  $\Lambda$  this means that each  $\rho$  contributed by each  $m$  contains rules that contribute to the same behaviour and so this overlap in contribution must be reconciled. The four possible combinations are as follows between two submodels,  $m_1$  and  $m_2$ , hereafter defined as RC1, RC2, RC3 and RC4 (Rule Combination  $n$ ):

1. *Exact overlap*:  $m_1$  and  $m_2$  cover the same behaviours.

- a. This is a model substitution/replacement situation. Ideally, since both models are abstracting the same subsystem assuming  $m_{1,2} \mapsto \psi$ , their behaviours should actually be the same and their respective  $\rho$  sets should not clash in terms of behaviour of  $t$ . However, if  $m_1$  and  $m_2$  represent theoretically alternative models (e.g. two possible models of a pathway whose dynamics can only be hypothesised into  $m_1$  and  $m_2$ ) some reconciliations can be made in terms of corresponding  $\rho$ :
  - i. Reject  $\rho_1 \vee \rho_2$  and perform two separate simulations. This strategy would be invoked if it were determined that  $m_1$  and  $m_2$  are irreconcilable because they each represent two very different and distinct beliefs of the subsystem and  $\kappa$ .
  - ii. Accept  $\rho_1 \wedge \rho_2$  in a decision framework. This strategy would be invoked if it were determined that  $m_1$  and  $m_2$  represent very similar beliefs of the subsystem and  $\kappa$ .  $d$  must now additionally determine under what circumstances  $\rho_1$  and  $\rho_2$  are fired.
  - iii. Accept  $\rho_1 \wedge \rho_2$  in a belief framework. This strategy would be invoked if it were determined that  $m_1$  and  $m_2$  represent very similar beliefs of the subsystem and  $\kappa$ . Essentially  $\rho_1$  and  $\rho_2$  are now fired under some belief mechanism other than a crisp decision process, i.e. both rules are fired to a certain degree. This can be viewed as a fuzzy firing mechanism.

2. *Partial overlap*:  $m_1$  and  $m_2$  partially cover the same behaviours.

- a. Essentially this is the same as point 1 except that  $\rho_1$  and  $\rho_2$  are now partitioned into a shared subset  $\rho_{1\wedge 2} = \rho_1 \cap \rho_2$  where each overlap subset  $\rho_{1\wedge 2}^2 = \rho_{1\wedge 2} - (\forall r \in \rho_2)$  and  $\rho_{1\wedge 2}^1 = \rho_{1\wedge 2} - (\forall r \in \rho_1)$  can be treated as the exact overlaps

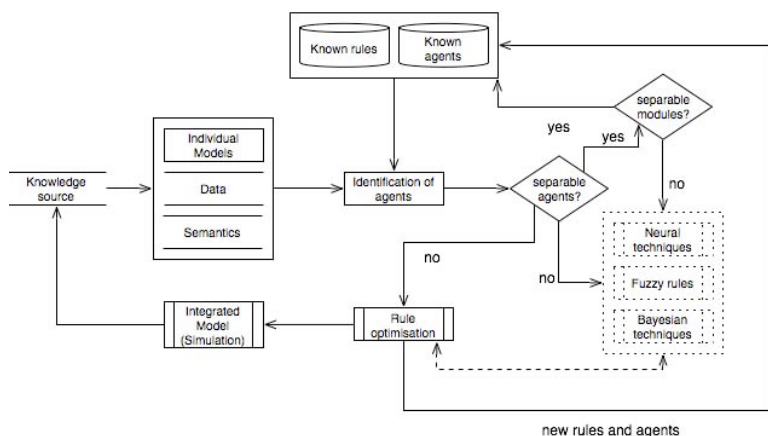
described above so that the same reconciliation strategies listed above can be applied.

3. *Total overlap (subsumption)*:  $m_1$  totally subsumes the behaviour that  $m_2$  represents in addition to other behaviours not included in  $m_2$ .
  - a. This is analogous to the situation described earlier wherein a phenomenological behaviour totally includes, i.e. subsumes, more mechanistic behaviour, however note that the behaviours need not be more/less mechanistic with respect to each other. In the case of an iterative integration (i.e. the process in which  $m_2$  is being integrated with  $m_1$  where  $m_1$  is already represented in  $\Lambda$ ),  $\rho_2$  must perform in  $\Lambda$  such that it does not interfere with the behaviour elicited by  $\rho_1$ . However in the absence of the set  $\neg(\rho_1 \wedge \rho_2)$  this is not always practical to achieve. In such a case a rule merge may well be the best solution wherein the behaviour of the integrated model is as close to  $m_1$  as possible.
4. *Null overlap*:  $m_1$  and  $m_2$  elicit behaviours that are separable.
  - a. If two models elicit behaviours that are separable their respective BITs do not integrate, i.e. there are no possible connections.  $I_1$  that has been derived from  $m_1$  and  $I_2$  that has been derived from  $m_2$ ,  $I_1 \neq I_2$  unless there exists pairs of  $t \in I_1 \vee I_2$  that interact in  $\Lambda$ .

Note that linear integration is also feasible in this formalism in at least three modes:

1.  $\forall m_k \in M_\phi$  are integrated into  $\Lambda$  as described above.
2. An entire  $M_\phi$ , if its behaviour can be encapsulated into any  $t$  in  $\Lambda$  associated with a set of rules (therefore directly able to affect states  $S$  as required).
3. A parallel approach can be adopted wherein the integration is minimal:  $\Lambda$  runs separately from  $M_\phi$ , which simply directly affects  $S$  and/or  $G$ .

Unifying all of these concepts yields the protocol summarised in Figure 5.2. Firstly, as with any ABM simulation, a mesoscopic scale on which  $t$  is defined must be identified. For different models this scale may well be slightly different, within reasonable bounds, depending on the anticipated set  $I$ . The entire simulation framework  $\Lambda$  is built around this mesoscopic scale wherein more global (i.e. more macro) behaviours are considered emergent of the small-scale subset of  $I$  and finer-grained (i.e. more micro) behaviours are either defined explicitly within the framework (brute-force method) or inbuilt (metamodelled) into the rules and/or states (and therefore state transitions) of the larger-scale subset of  $I$ . This is analogous to Brenner's so-called middle-out approach, which is an accepted albeit informal approach in system biology [136]. Secondly, the  $h$  function is applied to the models progressively as explained in previously, which yields the respective rule sets  $\rho_m$ . Every model must also be consecutively encoded by KDA to generate the BIT, which defines the behavioural overlap between the models. When this is known the rule integration strategies, as listed above, can be applied to give an ABM that is representative of the original constructs.



**Fig. 5.2** Knowledge about the model itself as well as external knowledge contribute to the model integration process by KDA, feeding into the ABI process as described in the main text to identify key interactors and behaviours. Rules are optimised and saved for future integrations or further optimisation. Critically, the process preserves the semantics and behaviour of the inputted models. Neural and Bayesian techniques and fuzzy rules could be used in this process as metamodeling techniques, which will be discussed more fully in Chapter 7

### 5.2.5 Expected Results and Validation

An important question is how far an integrated model can be validated mathematically. The nature of the ABI method described above is such that there is bound to cumulative error in decomposition to agent rules (though the rule enrichment process should ensure a satisfactory margin at each stage). For individual models that are integrated into ABI validation is straightforward: there is a direct relation with the original model and hence an analysis can be made as to how well, both quantitatively and qualitatively, the ABM simulates the original model. Integrated models, however, have no such direct relation. As explained in Chapter 3 small changes in the rules of local interactivity can have non-linear effects globally, and hence the effect of rule combinations in the KDA/ABI method could be unpredictable. Therefore the results and validations can only be qualitative in nature, with comparisons that can be drawn from the original models and known knowledge of the real system.

## 5.3 Summary

A formal methodology is proposed based on the literature review presented earlier, wherein an analysis revealed the anatomy of models: abstractions can be made with respect to paradigm, focus, scope, context and scale. Note that though these abstractions are made from the tumour modelling literature, they do in fact pertain to basic mathematical principles. To address cross-paradigm integration the ABI strategy is

proposed whereas focus, scope and context are addressed by the KDA approach. Cross-scale integration is left to the implementing software system. Together these formalisations can be distilled into a protocol in which models can be integrated.

## **5.4 Conclusion**

By using generic classes that include modelling concepts and their corresponding real-world counterparts one begins to build a behavioural inclusion tree that describes behavioural subsumption. Model integration can then begin from a behavioural perspective by additionally optimising agent rules, which effectively takes into account the different types of model focus, scope, context and assumptions and is additionally paradigm-free with respect to the original models. The efficacy of this approach will be tested in the next chapter.

## Chapter 6

# Experiments in Model Integration

**Abstract.** In this chapter the application of KDA/ABI model integration strategies is demonstrated. The design of the experiments performed including rationale is presented and followed by results procured. Simulation experiments were carried out to prove the efficacy of the KDA/ABI approach. The process of generating BITs to identify system/model components and behaviours (i.e. points of integration) is demonstrated. The process of (literature published) model decomposition is demonstrated via a genetic algorithm and subsequent integration is performed by combining BITs and resultant behaviour analysed.

### 6.1 Model Integration Experiments

The aim was to validate the KDA and ABI methodologies. Note that all results and validations described below were obtained with respect to models analysed solely from publication<sup>1</sup>-based information (including situations where further validations required additional information from the original authors).

To validate the KDA approach 39 models were examined in detail, with an extra 5 that were not in the original literature review (so as to introduce a new set of semantics that has not yet been encountered in the development of KDA) to validate the KDA process. These five are detailed below, each yielding a BIT via KDA (including concepts that could be expressed solely from the core sets described in section 5.2.3, which are relatable with respect to each other) - later in the discussion this process will be justified as a technology that enables expression of  $\Phi_I$  (the integrated whole). To keep in line with the cancer-oriented theme of this project only tumour models<sup>2</sup> were considered though it will be demonstrated that the method is

---

<sup>1</sup> Publications in the form of CellML models, such as those presented in EBIs BioModels database, could have been used instead, however this was decided against since (i) it would have required an additional layer of engineering to parse into the existing integration platform and, more importantly, (ii) semantics such as assumptions, etc., are often explained in far more detail in the original publications.

<sup>2</sup> Except for Markus (1999) [123], which is a generic vessel morphogenesis model.



equally applicable across disciplines. To test the limits of KDA, i.e. the volume of semantics that it can incorporate without yielding erroneous implications or creating practical difficulties, the models are used to build a large  $\Phi$ .

Decomposition of models, i.e. the formal process  $h$ , was achieved by means of coupling a custom GA and simulation software. To simultaneously show that the ABI strategy can address the multi-paradigm problem a formal integration was performed with a new set of published models as well as with models described in the literature review. Models are selected on the basis of paradigm, focus, scope, context and assumption diversity and self-validity (i.e. the models are shown to be validated in their own right against the biological system).

## 6.2 Knowledge Driven Approach: BIT-Building

In all BIT diagrams the red components represent the base classes described previously whereas green components represent those derived from the model. BIT diagrams can also become quite complex in terms of number of nodes, especially so for models that incorporate (i) a range of assumptions based on focus and (ii) many submodels, and therefore diagrams below have been condensed by categorising some concepts together. In-depth discussions follow in the next chapter.

The general heuristic used to generate the BITs from plain-text publications is as follows:

- Disambiguation.
  - For every component and system concept (i.e. behaviours and processes), hereafter referred to as entry, a class is created that inherits from System-Concept.
  - For every new entry three checks must be applied:
    - Does the entry already exist? If so, reject the new entry.
    - Does the new entry represent a specialisation of an existing entry? If so it may be appropriate to let the new entry inherit from a more abstract one. In this case one may even granulate the model further by introducing more concepts that make the distinction between the new entry and old ones clearer.
    - If inheritance is applied, do the transitive relations break the logic as described in section 5.2.3 or create any paradox? If affirmative there may be a case to go back to the original core classes (such cases are discussed in Chapter 7).
- Connection.
  - As far as is possible, connect each new entry with other entities using the basic relationships outlined in section 5.2.3.

---

<sup>3</sup> Only sections of this map are shown in this book due to space constraints.

- Further disambiguation can be applied with, for example, Components derivations that appear to be partOf multiple System derivations. In this case one should specialise on the Component in question, i.e. make them separate entries with different names but have them derive from a common parent via inheritance.
- Explicit Assumptions.
  - If guidelines are met one should see that many explicit assumptions. Disambiguation and connection should not violate explicit assumptions, i.e. explicit assumptions take precedence over the BIT-builders knowledge of the system.
- Initial Conditions.
  - These should not be included the BIT unless they represent specific System-Constraints.
  - Parameterisation often can be represented as SystemConstraints.

### 6.2.1 *Chen (2004)*

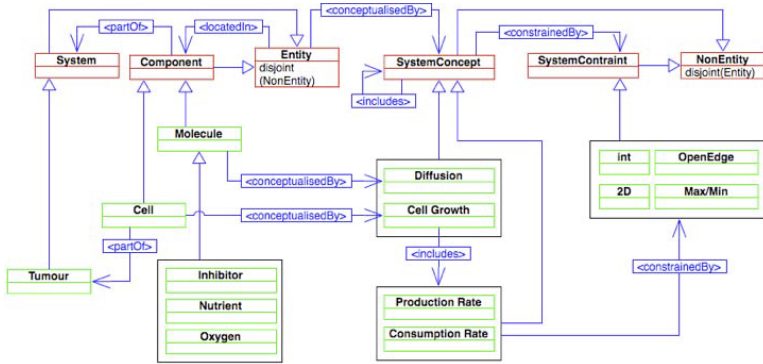
**Summary:** 2D ABM model of spheroid growth with tumour, oxygen, nutrient and inhibitor states measured on 4 different layered grids.

- Focus
  - Cell growth. Oxygen, nutrient, inhibitory agents diffusion.
- Assumptions
  - Cell division dependant on stimulatory (oxygen) and inhibitory (generic) factors. No vascularisation. Probability distributions for entering quiescent/proliferative states.
- Context
  - 2D. States (cell level oxygen consumption, inhibitor sensitivity, diffusion-related values) each with minimum/maximums values. Varying initial conditions.

### 6.2.2 *De Pillis (2005)*

**Summary:** ODE model of tumour growth based on tumour-immune interactions.

- Focus
  - Tumour (generic), immune (NK, CD8+) cell populations and immune cell effectiveness.



**Fig. 6.1** All components of the model described in the original paper can be distilled into this BIT except for logic-based assumptions that have been built into the agents (though these are implicitly included within respective system concepts), e.g. the threshold of nutrients and oxygen at which a cell makes a proliferation decision. Concepts are constrained by mathematical elements of the model, e.g. dimensionality, open boundaries (molecules can escape from the simulation space), capped ranges for certain states.

- Assumptions

- Growth is logistic (used as validation). Immune cell pressure (population) affects growth. NK cells always present, even in absence of tumour cells, whereas CD8+ cells appear only after tumour cells have invaded normal tissue and immune cells die after a certain number of encounters. Immune cells kill tumour cells. Overall effectiveness of immune cells dependant magnitude of immune cell population, i.e. the higher the population the higher the effectiveness on tumour cells (regardless of morphology).

- Context

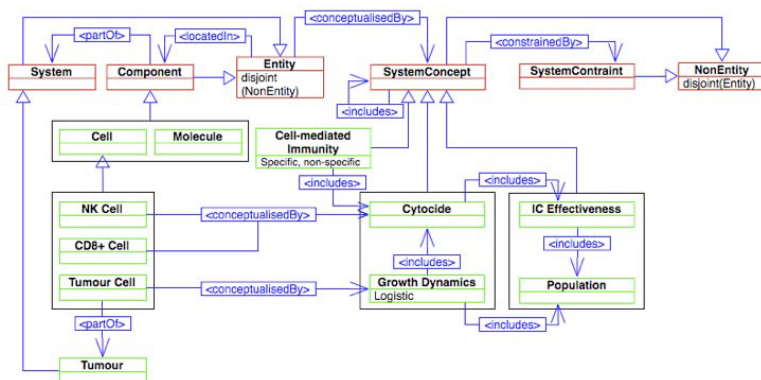
- Logistic curve variables (curve fitting to validation data). Parameterisation of ODEs (30 variables, see Table 1 in the original paper for details).

### 6.2.3 Mallet (2006)

**Summary:** 2D CA/PDE model of tumour growth including nutrient source (abstracted blood vessel) and immune system interaction, in particular cytotoxic T-lymphocytes (CTL) and NK cells.

- Focus

- Host (normal), tumour and immune cell population growth. Nutrient movement.



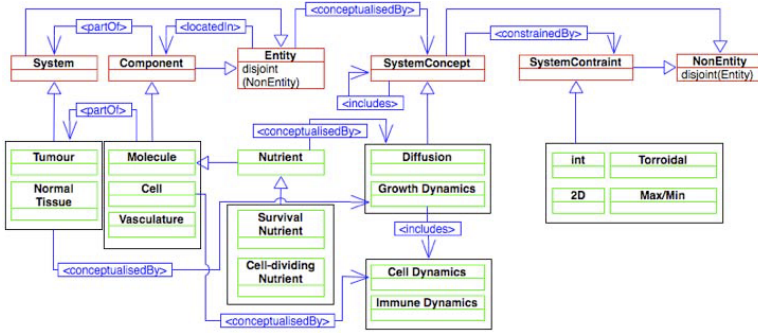
**Fig. 6.2** Components not drawn, but taken account of by abstraction, in the diagram include the SystemConcepts Specific/Non-Specific IC (Immune Cell) Dynamics and NK/CD8+ Life Cycle (including recruitment and maturity). Also, some Growth Dynamics and Population have been condensed (they equally apply to all cell types and the tumour system itself). The conditional presence of CD8+ cells and constant presence of NK Cells can be included by extending Population with the `<includes>` property.

- Assumptions
  - Cell and molecular migration only in 4 directions. Probability distribution that determines cell fate (fixed), i.e. death by necrosis or immune cell interaction, cell division, cell migration, maximum interactions before death (for immune cells only). Survival and mitosis-stimulating nutrients treated as two different entities.
- Context
  - Rates of consumptions and movements (diffusion coefficients) of each type of component. Probability of cell death due to necrosis or immune-cell interaction. Parameterisation of initial cell numbers and typical behaviours (see Table 1 in original paper). Random initialisation.

### 6.2.4 Markus (1999)

**Summary:** CA model of vessel morphogenesis.

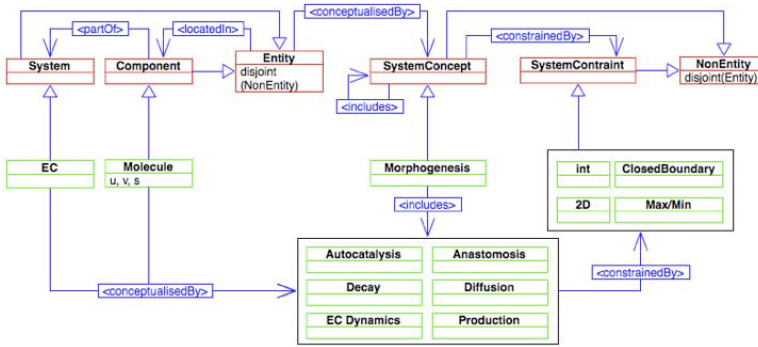
- Focus
  - Diffusion of stimulatory and inhibitory substrates.
- Assumptions
  - Vessel formation includes growth, branching, autocatalysis of activators, tip death, chemotaxis and anastomosis, which are all driven by stimulatory and inhibitory factors only (generic). Rules favour stimulatory substrate depletion.



**Fig. 6.3** Much of the immune system dynamics (e.g. action of NK cells, initiation, death and proliferation) is the same as that of de Pillis (2005) [56] and hence condensed here into the single node Immune Dynamics. Note that due to the transitivity of inheritance and <conceptualisedBy> properties, subclasses of nutrient automatically inherit association to diffusion and growth dynamics, which include cell and immune dynamics (i.e. their respective behaviours).

- Context

- Maximum/minimum values for substrates validated against various systems including angiogenesis and plant vein growth.

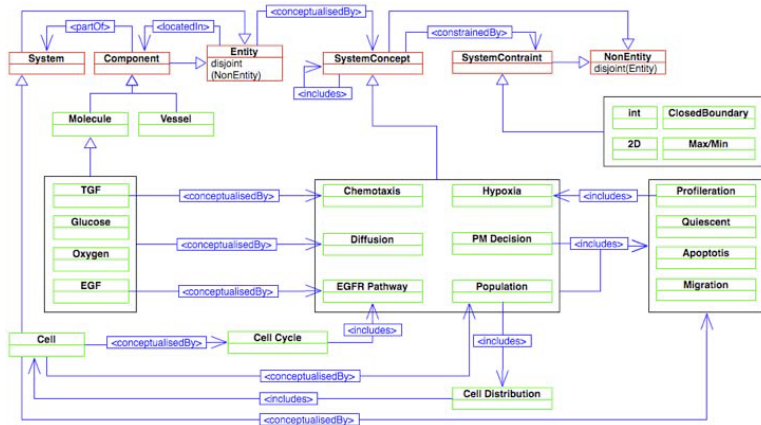


**Fig. 6.4** The behaviour of morphogenesis, treated synonymously to vasculogenesis in this example only, is a function of autocatalysis, anastomosis, production, decay and diffusion of molecules  $u$ ,  $v$  and  $s$  (inhibitory, stimulatory and substrate respectively). These behaviours are embedded as logical rules in the CA.

### 6.2.5 Zhang (2007)

**Summary:** ABM simulation of glioma growth integrated with EGFR pathway ODE model.

- Focus
  - Tumour cells, chemotaxis and diffusion of TGF, glucose and oxygen. EGFR pathways conceptualised by molecular dynamics (series of ODEs), which is triggered by receptor internalisation. Main focus is the switch a tumour cell undergoes from migratory to one of proliferative, apoptotic or quiescent state (and back again).
- Assumptions
  - Migration and proliferation dependant on EGFR pathway state. All agents occupy grid intersections (only one cell can occupy a single grid point). Cells can be in one of four states: proliferative, quiescent, apoptotic or migrating. Hypoxia and availability of glucose affects the state of the cell, which in turn affects distribution of cells within the glioma. Simple cell cycle dynamics are included to determine whether a cell should proliferate.
- Context
  - Diffusion coefficients, rate constants for ODE model.



**Fig. 6.5** The EGFR pathway contains more behaviours than could be integrated into this diagram, e.g. internalisation of EGF and autophosphorylation (not shown for clarity). Other components can also be included that are explicitly described in the model, e.g. membrane, cytoplasm, nucleus and many molecular species but not shown to retain diagram simplicity (their inclusion can easily be incorporated as descendants of Component). Importantly, all model concepts are incorporable into the BIT diagram. The main focus of the model, i.e. proliferation/migration decision making of the cell (PM Decision above) is clearly set in context to the rest of the behaviours that the model includes.

Cell and Tumour entities in all models described above would need to be subclassed to the specific types of cell/tumour that the models represent (to keep the diagrams as simple as possible without reducing their expressive power these child classes were

left out). For example, the BIT diagram for Zhang (2007) [207] should include a subclass of Tumour and Cell to include the glioma types that the model was specified for:



For the most part, however, the state of the modelling literature is such that models tend to focus on generic behaviours rather than specifically on tumour types (though it must be said this is less so for pathway models).

### 6.3 ABI: Model Decompositions

Model decomposition is first tested on simple and testable models. Two such models of particular importance in tumour modelling, as highlighted in Chapter 2, were the diffusion model and the Gompertz function. These models describe dynamics that occur on two very different scales, i.e. molecular and macroscopic respectively, and yield straightforward BITs. They have identical paradigm profiles and are here employed to test same-paradigm integration (multi-paradigm ABI will be addressed in the next section).

#### 6.3.1 Decomposition of the Diffusion Model

Ficks 2<sup>nd</sup> Law in 3 dimensions (the so-called heat equation) was used:

1. To test the first stage of the ABI strategy, i.e. model decomposition.
2. For subsequent integration into other models.
3. As a benchmark for the GA module.

This particular model is well understood and it is relatively simple to find solutions and hence test by a fitness function. The standard diffusion model can be used:

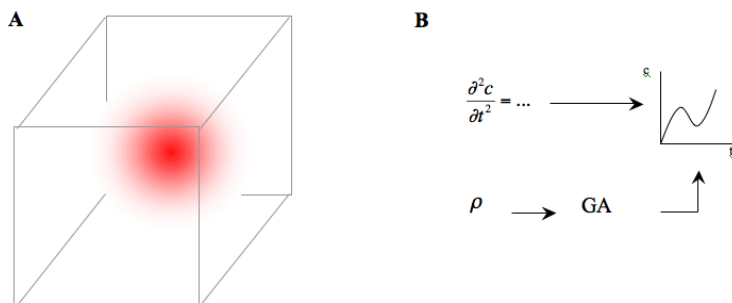
$$\frac{\partial c}{\partial t} = D\nabla^2 c \quad (6.1)$$

where  $c$  is concentration of solute and  $D$  is the diffusion constant.

The ABI method was used to generate rules that mimic the behaviour of the diffusion model. Figure 6.6 illustrates.

Representing molecules on an individual basis is of course not a good choice in a model that intends to simulate diffusion over large spaces, as sought in a

whole-tumour model, due to computational resource limits and output requirements<sup>4</sup>. Instead a lattice approach is adopted here wherein a discretised 3D space contains lattice sites (all of which have identical rule sets) that represent concentration(s) of solute(s) as states upon which rules operate. In effect this is a meta-modelling approach to bridge the scale gap between molecular movement and macroscale behaviour akin to the Lattice-Gas approach. In the first instance 100 units per unit volume of solute was introduced at the midpoint of a 3D closed space of  $10 \times 10 \times 10$  unit dimensions and allowed to diffuse at  $\lambda=1$  (denoting a zero molecular impedance<sup>5</sup>). A von Neumann neighbourhood was assumed on the theoretical basis that the net movement of molecules through a lattice interface is proportional to the area of that interface. This can equally be applied for higher-dimensional neighbourhoods but the von Neumann variety is relatively simple to implement (and hence relate to the said area) and is a typical and well-researched construct in discretised simulations. However the radius of the neighbourhood for which rules can operate on states was left to the GA to optimise on, as explained on the next page.



**Fig. 6.6** A. Introduction of a solute at a single point (or randomly distributed as a secondary initialisation, not shown) in the middle of a 3D space is allowed to diffuse in all directions. B. The non-dimensionalised solution of Ficks 2<sup>nd</sup> Law is used as a fitness function to enrich an initial set of automata rules such that resulting simulation mimics the original equation.

The GA was initialised with a population of 100 individuals with rules (the first generated manually, the rest randomly mutated) and GA operator permutations to procure the results detailed below. The fitness function (explained shortly) was set to eliminate 50% of poorest performing individuals from the population in each generation. These initialisations were chosen on the basis of initialisation setups of previous GA-based rule optimisation experiments, summarised by Bossomaier (1999) [29] and Mitchell (1996) [128].

<sup>4</sup> The theoretical basis of diffusion, i.e. random movement and impedance of particles, could be encoded into an ABM, for example, as rules for individual particles of which there could easily be billions.

<sup>5</sup> The decomposition was later cross-validated with alternate initial conditions and diffusion constants.



A specialised rule was hard-coded on which the GA could operate, which is detailed in pseudocode shortly. Note that the ABI strategy requires additional knowledge about the system for rule initialisation. In this case diffusion is initialised into a static rule set that operates on a state space  $0 \leq s \leq 100$ , which is dependant on the state space of neighbouring lattice sites - this dependency is built into the rule.

The diffusion model was separately solved to perform as a fitness function using the standard Eulers finite difference method<sup>6</sup> (Figure 6.7). The GA took at least 18 generations<sup>7</sup> (depending on initial rules and GA configuration, both explained shortly) to generate a rule set that came within the preset 0.05 distance value, measured at each time and grid point<sup>8</sup> (explanation of this threshold given shortly). Fitness functions employed the standard Euclidean distance metric:

$$\text{total distance} = \sqrt{\sum_{i=0}^{i \leq T} (v_{\text{Observed}}^i - v_{\text{Actual}}^i)^2} \quad (6.2)$$

where  $v$  represents the observed (simulated) and actual (solved from PDE) values of the diffusion model and  $T$  is the upper limit of both simulated and Euler solutions. Unless otherwise stated this distance measure is used in all subsequent GA experiments. The best solution is shown in Figure 6.8

The diffusion model was secondarily used as a test of how well the GA software component performed and functioned<sup>9</sup>. The first test runs of rule optimisation for the diffusion model were coded as an arbitrarily expanding rule set with the default GA configuration<sup>10</sup>. In summary, 100 individuals were initialised with a single chromosome and mutated prior to the GA run. The aforementioned significance threshold was first of all tested by varying between 0 and 0.1 in 0.01 intervals - initial results showed that any threshold below 0.05 rarely yielded positive results, causing a very rapid extinction of the GA population, and a threshold greater than 0.1 would generally lead to a hanging<sup>11</sup> population, though these results were not always reproducible. Though such a determination is of course dependant on the nature of the model that is being decomposed, a significance of 0.05 (i.e. 95%

<sup>6</sup> The solution required a 3D mesh on which finite differences were solved. The software used to perform the simulations in this project was used to also generate this mesh and solve the equation for various values of the diffusion constant. The result was outputted to file prior to GA execution, which was directly accessed by the fitness function model class.

<sup>7</sup> Best result set out of 3 repeats of various configuration changes. Full results not published here.

<sup>8</sup> For clarity not all grid points and time points are shown in subsequent experiments.

<sup>9</sup> Note, however, that this is not a primary goal of this book. The software is tested in this way purely to validate the method as a viable optimisation tool.

<sup>10</sup> Full results not published here

<sup>11</sup> A completed run here is defined as the full execution of a GA through any number of generations terminating in either a positive result or population decrease to a predefined threshold. A hanging population is reached when no discernable increase or decrease in performance amongst individuals is observed, resulting in GA run that never terminates, defined here as an incomplete run. Else a population becomes extinct due to removal of bad performers and lack of good ones.

accuracy) was chosen for all subsequent experiments on the basis that it represented a good balance between yielding positive results in a relatively small number of generations without causing an unstable GA population (at least in the diffusion model case) and it is traditionally the value chosen for biological statistical significance. In two regimes, each with one of the 2 initial rule variations listed in pseudocode below, individuals were initialised into the GA population. This was repeated at least 3 times to verify results.

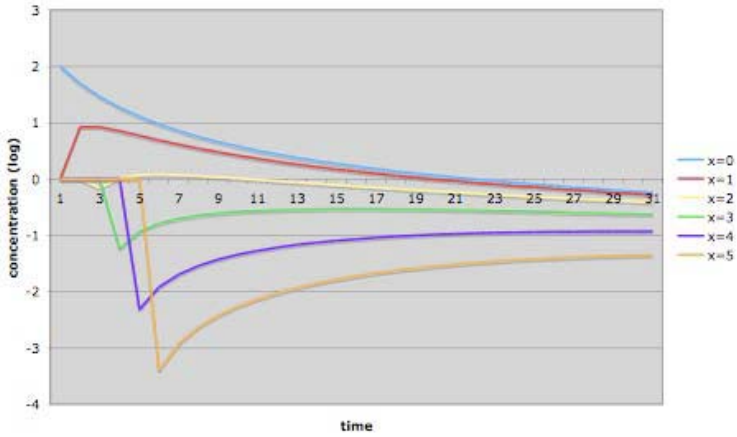
1. If *true* move a fraction  $f$  of concentration in direction  $\pm x, \pm y, \pm z$  over  $n$  surrounding sites (i.e. neighbourhood).
2. If *true* move a fraction of concentration  $\times$  previous rate for this lattice site in direction  $\pm x, \pm y, \pm z$  over  $n$  surrounding sites. (A variation of this rule included the addition of a random fraction of movement).

Shared-state classes ensured the conservation principle, i.e. the total concentration in the entire simulation space remained constant (excepting rounding error). Five variations on  $f$  in the rules above were used as initial values: 0, 0.25, 0.5, 0.75 and 1. In the case of Rule 1 all variations eventually yielded a rule set that procured the required 5% threshold, with 0.5 consistently producing the fastest result, the best of which was 18 generations. Rule 2, however, never produced a result that broke the 5% threshold. The best result was 34% after 58 generations after which there was no improvement even when allowed to run over 150 generations (spanning a runtime of more than 4 days). The addition of a random factor of concentration movement did not procure a positive result either (this was actually only performed once, producing a hanging population; results not shown).

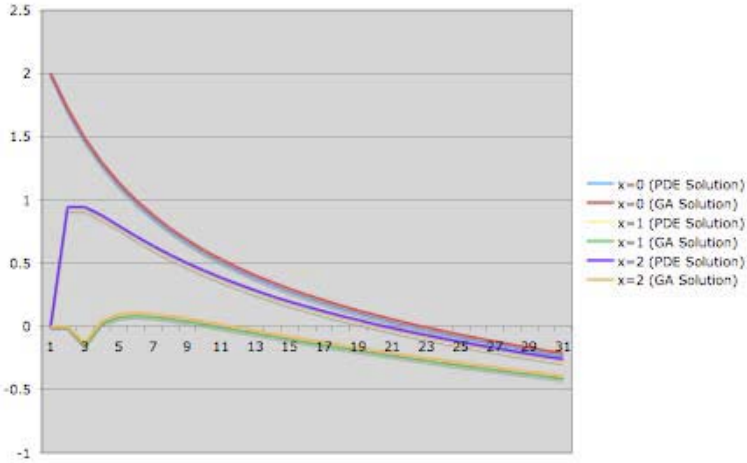
### 6.3.2 *Decomposition of the Gompertz Model*

The same method of decomposition by GA as described above was used to translate the Gompertz model into  $\Lambda$ . The fitness function employed needed an instantiation of the model to compare against. The original experiments done for decomposition of the Gompertz model revealed that initialising the simulation with a low cell number, e.g. 1 to 100, impacted simulation time and hence a single generation could take in the order of hours to complete. Therefore the initial state of the simulation was set to a realistic number of cells, based on initialisation of similar models described by Araujo (2004) [11] and Ribba (2006) [152], described shortly. The first instantiation conformed to the equation below, which yields a smooth curve very much like those procured by some of the growth models described in Chapter 2 and stabilises at a population of  $6 \times 10^5$  cells, which would be approximately consistent with small MTS growths [65]. The paradigm profile is also described as the 6-tuple:

$$y = 5 \times 10^5 + 6 \times 10^7 \times e^\beta \quad (6.3)$$



**Fig. 6.7** PDE solution. Log concentration of a generic solute (10 units per cubic unit of length) introduced at time 0 diffusing over a three dimensional space (only a single axis shown above, from  $x=0$  to 5) over time. The expected exponential decrease at the point of origin is seen as the solute moves outward equally in all directions. Eventually all points in the lattice converge on a concentration of 0.01 (not shown) as the solute is completely dissolved in the available space of  $10 \times 10 \times 10$  closed-boundary space.



**Fig. 6.8** PDE and GA solutions compared at 3 points on the  $x$ -axis. Total distance measured as 3.69 units per unit volume.

$$\beta = e^{-0.005 \times (t-400)} \quad (6.4)$$

$$P_{gompertz}^m : (f^m, q^m, c^m, \neg s^m, \neg e^m, 0d^m) \quad (6.5)$$

Clearly there is no direct relation between the model and any notion of 3D structure or microscopic mechanisms. In the first instance a GA run based on a rate-dependant rule set was performed where the chromosomes, pseudocode below, were used to generate a population of 100 individuals:

- If *true* rate = rate at previous time step for this lattice site  $\pm(0 \cdot x)$
- If *true* increase replicate by  $x \pm$  fraction of  $x$

Note that all mathematical operators and variable names are subject to GA operations, i.e. mutation, crossover and mating (hence the introduction of the term  $0 \cdot x$  in the first rule). A total of 30 separate complete runs (and 15 more incomplete) with varying GA configurations failed to procure a result that achieved the predefined 0.05 threshold. An alternative fitness function strategy was employed where the best 25%, instead of 50%, performing individuals were progressed through each successive generation in a further 10 runs with varying GA configurations, so as to increase the selection pressure. The best performer was found to only achieve 59% accuracy.

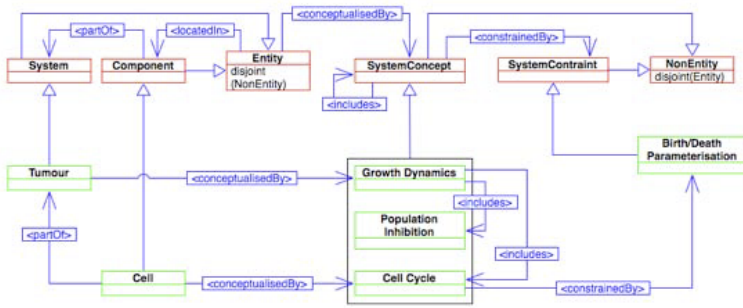
To translate the little information available (i.e. that a variable, in this case number of cells, grows according to constants in the equation, which have no direct relation to actual system components) into  $\Lambda$  external knowledge about the system needed to be added to obtain more meaningful results (none of which are directly extractable from the original model):

- Dependent variable represents cell number (integer).
- Each cell has:
  - A point of birth.
  - A living period.
  - A point of death.
- One cell generates one daughter cell in a single proliferative cycle.
  - Single proliferative cycle defined as a varied number of steps.
- Inhibitory effect due to local population.

A more complex extrapolation can be sought wherein cell cycle phases can be incorporated however this straightforward model is used here for simplicity. The corresponding BIT is shown in Figure [6.9](#).

Firstly, cells were treated individually and initialised with identical rules, i.e. a child cell has the same replication rule as its parent, creating a homogeneous population. However, the cluster quickly reaches its limit after 20 to 25 proliferative

cycles<sup>12</sup>, constituting over 33 million individuals<sup>13</sup>. An approximation strategy was therefore devised wherein lattice sites were assigned separate lattice states representing cell number, as with the diffusion model (including the state the cells are in, i.e. living, birth/death point). Indeed this also constitutes a metamodel as described in the previous chapter. Though this proved to be more difficult with respect to constructing rules, it did eliminate the resource constraint problem completely. In fact all simulations were subsequently run on only 5 nodes (1 master, 4 slaves) satisfactorily, leaving a further two quintuplets to perform parallel GA experiments on.



**Fig. 6.9** BIT diagram for the Gompertz model, with additional assumptions as given in main text.

As before, an appropriate approximation of the rules was initialised into a GA, as a starting point, consisting of 100 individuals with varying chromosome number (chromosome replication and mutation was performed prior to starting the GA to ensure a diverse population is initialised). All GA runs were made with the rule set listed in pseudocode below with varying GA and cell configurations:

Rule Set:

- If {on death time point} die
- If {local population  $> x$ } permissible = true
- If {on birth time point and permissible} replicate
  - Place new daughter cells randomly in local neighbourhood
- If {true} advance point in life and cell cycle

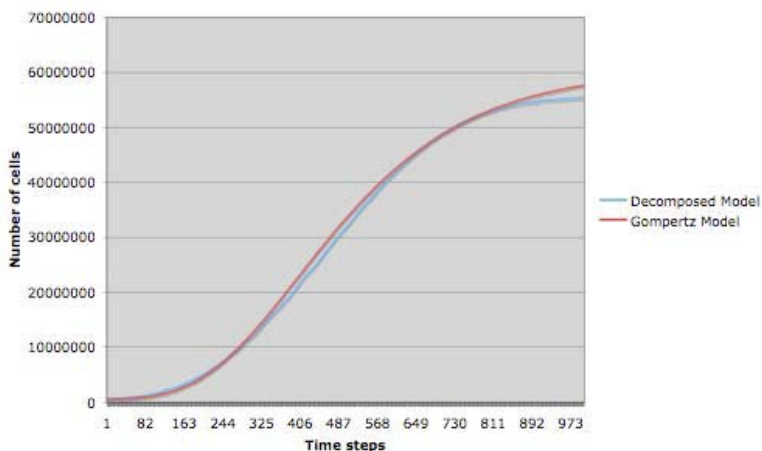
<sup>12</sup> Approximate value. As computational load increased agent and simulation logging became congested in the execution stack and so it was not clear exactly how time steps were incremented.

<sup>13</sup> Some threading and rule execution optimisations were made, however the JVM eventually either runs out of memory (with the upper limit on the number of threads being around 900 threads) or seizes on multiple hard-disk I/O calls (used to record output).

Configurable cell attributes:

- Length of life and cell cycle
  - Point of birth and death set at beginning and end
  - Number of cycles before death
- Stochastic shortening/lengthening of cell cycle and length of life
- Over-population threshold

Out of a total of 48 GA runs with the variations shown (each repeated 3 times) 3 reached the target 5% accuracy. Of those 48 experiments, 12 were run with varying GA initialisation (using GA values of  $Cr_p$ ,  $M_p$  and  $Mt_p$  as 0.75, 0.25 and 0.5 respectively), which produced a single positive result, and a further 3 sets of 12 experiments were similarly generated for variations of the equation, which produced two more positive results. The best result is shown in Figure [6.10](#).



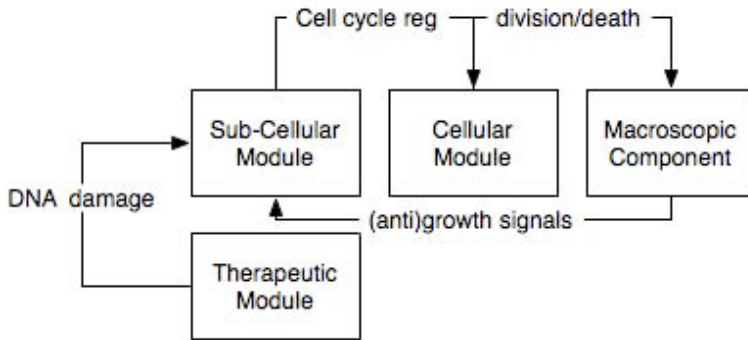
**Fig. 6.10** Best GA result produced a curve with a Euclidean distance of 3.3% from the actual result. The cell cycle length was 21ts (time steps) with a life expectancy of 11 cycles and population threshold of 1253 cells per lattice site. Although after 10000ts the GA-produced model does not converge to the required higher asymptote the bulk of the curve fits the shape of the original.

## 6.4 KDA+ABI: Formal Model Integrations

To demonstrate the formal model integration method developed in this work KDA and ABI analysis was performed on typical tumour models to produce more informative models as per the definition of model integration.

### 6.4.1 Gompertz Model Integration with ODEs and Discrete Models

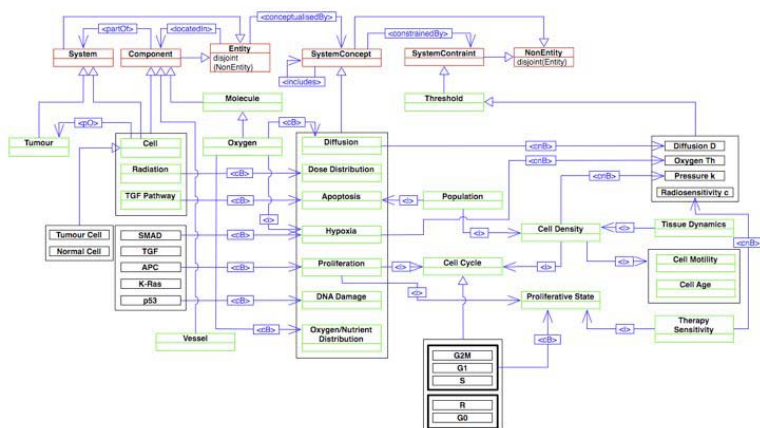
Ribba (2006) [152] describe a 2D multiscale avascular tumour model, integrating the TGF pathway and its role in cell death and proliferation into an ABM. The model is already multi-paradigm, incorporating a DE-approach to growth, diffusion and therapy response with a discrete Boolean network model describing pathway dynamics, all implemented into a discrete CA-like lattice and a parallel DE solver. This model was chosen for model integration with the Gompertz model since (i) it exhibits diverse system behaviours and therefore provides the opportunity to test how well KDA can describe a promising model, and (ii) the model is well-described in terms of modules and therefore components are relatively easy to simplify or extract. The model is described below.



**Fig. 6.11** Summary of multiscale model by Ribba (2006) [152]. A Boolean network of the TGF pathway (Sub-cellular module), incorporating DNA damage (Therapeutic module), apoptosis and proliferation, is coupled with cell cycle regulation on the cell level (Cellular module), which in turn is coupled with cell population pressure and diffusion dynamics in the macroscale (Macroscopic module). Image taken from original paper.

#### 6.4.1.1 Model Implementation

Darcys law is very similar to the diffusion model and includes a pressure term. Since the diffusion model has already been decomposed this module of the original Ribba model was replaced with the diffusion model, the GA (described shortly) optimising on the diffusion constant alongside the other variables. Though this of course does not simulate the original model exactly, it suffices so far as to describe tissue dynamics since they can both describe the biased random movement of entities. The aim of this decomposition is to integrate the ODE and discrete parts of the original avascular model with the Gompertz function, and eventually an angiogenesis model, and therefore the therapy module of the original model need not be extracted. Consequently the Boolean network was not included in the decomposition since it formed



**Fig. 6.12** BIT diagram for Ribba (2006) [152]. Though many of the molecular and parameterisation components have been left out for the sake of clarity (though they are indeed incorporable) the main assumptions and foci the model can be expressed by this approach. Specific instantiations are outlined black, rather than treated as classes as before. This highlights the use of KDA to expose  $h$  as explained previously. Cell is conceptualised by Population (not shown). Properties have been abbreviated: i (includes), cB (conceptualisedBy) and cnB (constrainedBy).

the therapy module (i.e. tumour growth after DNA damage)<sup>14</sup>. All other parts of the model were replicated as explained below.

### 6.4.1.2 Simulation Setup

The original models simulation setup was replicated with a hybrid agent/lattice space (except now in 3D), composed of  $100 \times 100 \times 100$  lattice sites (totalling 1cm each side) with each discrete time step representing 1 hour. In terms of physical scale the original model was a 2D  $8 \times 8$  cm square area. This is reduced here to  $1 \times 1 \times 1$  cm 3D cube to keep computational load low for a speedup in GA optimisation and is later scaled up to the original size.

The BIT diagram exposes states and  $t$  as explained below.

Required lattice states:

- Oxygen level (Float).
  - Hypoxia (Boolean).
- Cell population (Integer).
  - Normal Cells.
  - Cancer Cells.
- Cell phase population (tumour cells only).
- $G_1$  population.

<sup>14</sup> The possibility of further implementing the therapy module is discussed in the next chapter.



- S population.
- G<sub>2</sub>M population.
- G<sub>0</sub> population.
- Apoptotic population.
- Central blood vessel (constant O<sub>2</sub> level).
  - Represented by a set of lattice sites that renew their O<sub>2</sub> status at every time step.

Required rules:

- Oxygen diffusion.
  - Diffusion constant for oxygen.
- Cell populations update.
  - If oxygen level is below threshold or local population is over a threshold, all cells jump to G<sub>0</sub> (hypoxia is declared if oxygen level is below the threshold). If that is the case subsequent rules are not fired.
- Ribba (2006) [152] defines overpopulation as 2000 cells aligned in the area of a single spatial unit. The lattice site is replicated here in 3D, which translates to approximately 13×13×13 cells.
- Take away fraction of O<sub>2</sub> from environment according to total cell number.
  - Oxygen consumption is  $3.6 \times 10^{-13} \text{ mlh}^{-1} \text{ cell}^{-1}$ .
- Update cycle phases.
  - Any cells at the apoptotic phase die (no nutrients are relinquished to the environment since this was not defined in the original model).
  - Increment all cell cycle steps. Reset the newly formed population from mitosis at beginning of cell cycle (G<sub>0</sub>).

Time scale (as from the original model):

- S = 10h.
- G<sub>2</sub>M = 3h.
- G<sub>1</sub> = 20h.
- G<sub>0</sub> = 5h.
- Apoptotic phase = 5h.

### 6.4.1.3 Rule Optimisation

To describe molecular dynamics and cell movement the diffusion model decomposed previously was used. This was implemented as rules that dictate the movement of cell populations (integer state) amongst lattice sites instead of a floating point concentration as before. The original  $k$  constant for Darcys Law ( $2 \times 10^{-5} \text{ cm}^2$ ) was reconstituted here as a starting point for a diffusion constant (representing media permeability) to optimise on.

The fitness function for the GA was modified to measure the scaled increase or decrease in cell population (so as to capture the shape of the curve rather than the Euclidean magnitude) since:

1. Spatial dimensions of the computational domain was increased to 3D.
2. The computational domain itself was decreased from 8mm to 1mm on each side.
3. The macroscale module was changed.

Concisely, the fitness function measured the rate of change at each time step in comparison to the previous time step and measured the Euclidean distance between observed and expected rates:

$$\text{total distance} = \sqrt{\sum_{i=0}^{i \leq T} (v_{\text{Observed}}^i - v_{\text{Actual}}^i)^2} \text{ where } v_x^i = \frac{v^i}{v^{i-1}} \quad (6.6)$$

The original model was replicated, solving the diffusion model as before on a mesh and recreating the 2D lattice with the classes described previously. As in the original model each time step represented 1h of progression. For each initialisation the states for all lattice sites and states were written to file and used by the fitness function to determine fitness of individuals.

#### 6.4.1.4 Integration and Simulation Results

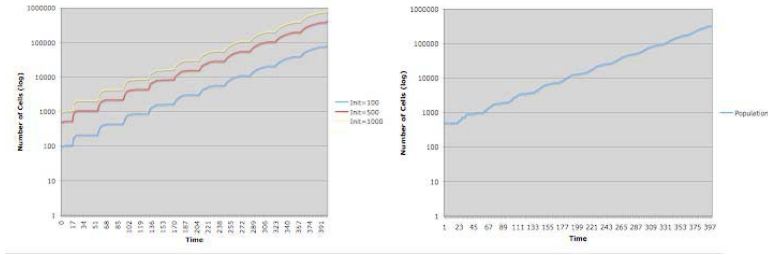
The GA quickly found a solution to oxygen diffusion and cell movement, in 8 and 11 generations respectively. Initial conditions were varied and compared to the original model to verify the rules as follows.

- Number of vessels.
- Population number of each type of cell/phase.
- Phase length.

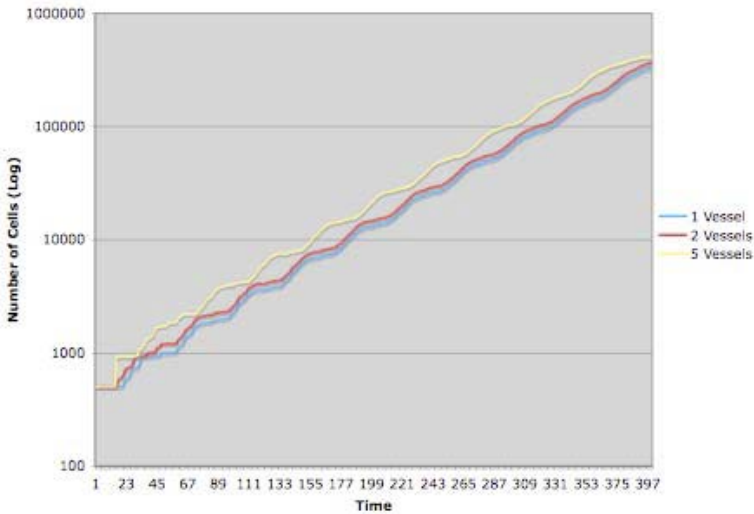
Tumour growth according to the Ribba model is illustrated in Figure 6.13, with variations on cell number (initial conditions) and cell cycles of all cells synchronised. An asynchronous version of the same model, where the cells are initialised into the simulation at a random point of the cell cycle is shown. Increasing vessel number shifts the growth curve upwards as expected.

Figure 6.15 shows the BIT integration between the Ribba et al. model and the Gompertz derivative. Recall that BITs assert behavioural subsumption. In essence phenomenological behaviours subsume finer-grained behaviours. Indeed the integration between the Gompertz function and the Ribba et al. model showed that the finer-grained behaviours of the latter are subsumed by phenomenological behaviours of the former. Concisely:

- Component concordance.
  - The system and main component (tumour and cell) are of course directly relatable, hence those two  $t$  remain.



**Fig. 6.13** (Left) Log plot of cell population against time with a single vessel present. Population growth is exponential in the presence of ample oxygen. Cellular movement ensures that the majority of cells, by the time they have reached the end of  $G_2M$  phase, are clear of other cells and therefore are not affected by overcrowding until approximately  $t=200$ . Note that the original model did not specify the initialisation strategy in terms of synchronisation of cells. The effect of synchronising cell cycles can clearly be seen on the log plot, with steps of population increases at regular intervals equivalent to the cell cycle length. The effect is dissolved with the passage of time as the proportion of cells that enter  $G_0$  increases because of overcrowding. (Right). Log plot of cell population when cell cycles are unsynchronised (starting with 500 cells). Though average population per binned number of time steps is very close to a synchronous population (compare to red curve in a) this curve is far smoother and the effect of overcrowding is observable earlier in the simulation.



**Fig. 6.14** Effect of increasing vessel number shifts the curve upward. The periodic plateau of the curve is concordant with the results obtained by the original the Ribba model.



- Ribba Model

1. If {true} diffuse oxygen
2. If {true} diffuse cell
3. If  $\{O_2 > 8.01 \times 10^{-10} \text{ mlcell}^{-1} \text{ and population in lattice site} < (13 \times 13 \times 13)\}$  permissible = true
4. If {not permissible} go to  $G_0$
5. Else
  - a. If {true} oxygen uptake by cell
  - b. If {at apoptosis point} die
  - c. If {at end of cycle} replicate
  - d. If {true} advance point in life and cell cycle by 1

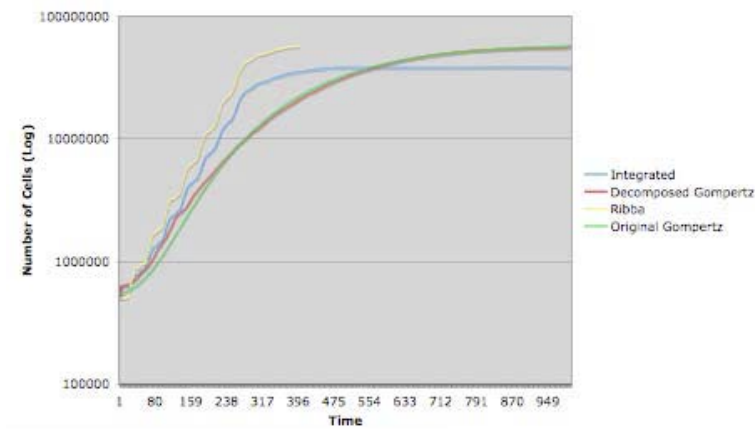
Using the BIT in Figure 6.15 rule integration was defined:

Relation	Rule Combination
$\rho_R^3 \subseteq \rho_G^2$	RC2
$\rho_R^{5iii} \subseteq \rho_G^3$	RC1
$\rho_R^{5iv} \subseteq \rho_G^4$	RC1
$\rho_G^1, \rho_R^1, \rho_R^2, \rho_R^{5i}$	RC4

where symbols G and R denote Gompertz and Ribba models respectively, integers and Roman numerals denote the corresponding rules. Recall that, with the aim of simultaneously achieving a BIT that tends to  $\Phi_I$  and attaining a synergy between different rule sets, RC4 overlaps simply add to the integrated rule set. RC1 and RC2 overlaps must perform in  $\Lambda$  under a belief and/or decision framework. Since a decision framework had not yet been devised or implemented at the time of execution a belief framework was utilised. The rules then became:

1. If {true} diffuse oxygen
2. If {true} diffuse cell
3. If  $\{O_2 > 8.01 \times 10^{-10} \text{ mlcell}^{-1} \text{ and population in lattice site} < [(13 \times 13 \times 13) \text{ or } 1253]\}$  permissible = true
4. If {not permissible} go to  $G_0$
5. Else
  - a. If {true} oxygen uptake by cell
  - b. If {at apoptosis point or life point=241} die
  - c. If {at end of cycle} replicate
  - d. If {true} advance point in life and cell cycle by 1

Since the Gompertz model represented the most macroscopic behaviour, and due to the implications of RC3, its value of 1253 as the population limit was taken as well as a cell cycle length of 21ts. The results are shown in Figure 6.16.



**Fig. 6.16** As Ribba report in their original model the exponential growth of their in silico tumour does reach a point where growth slows (yellow line; population starting at 500000; curve terminates early as the original simulation only continued to 400ts). The introduction of a lower population limit and, perhaps more importantly, a death term and shortening of cell cycle length transforms the shape of the curve dramatically. The characteristic periods of oscillatory growth can still be observed. On the log plot it does seem as if the population is steadied, however a closer look at the raw data reveals that the number is actually falling with periodic spikes in growth. This may well be an artefact of the metamodelling approach (discussed in more detail in next chapter). Though the exact magnitude of the Gompertz curve cannot be achieved the similarity in shape is encouraging.

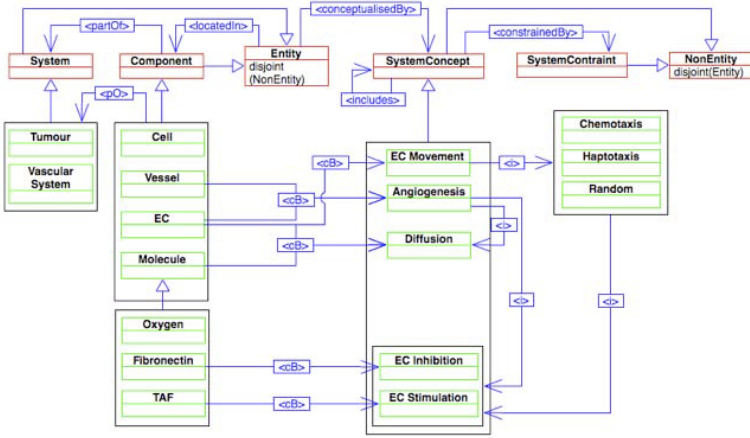
#### 6.4.1.5 Angiogenesis Model Integration

Chapter 2 highlighted the fact that many growth models would have benefited if integrated with a vascular growth component, as this would no longer limit their applicability to MTS. Therefore the integrated model described above was further integrated with a well-known and validated angiogenesis model first described by Anderson (1998) [7] and extended by Chaplain (2006) [36]. Briefly, the model takes a system of PDEs to describe vascular growth, diffusion and interaction with the ECM but solves the equations on an ABM-like platform. The BIT diagram is shown in Figure 6.17.

The simulation required additional states to be assigned to lattice sites as follows:

- Fibronectin concentration.
- TAF concentration.
- EC concentration (agent count).

It was assumed that once an EC (individually modelled as an agent) had moved along a lattice site the trailing vessel that was left was viable, i.e. mature, and able to deliver oxygen. This is concordant with the discrete solution of the original model.



**Fig. 6.17** BIT diagram for [7]. The corresponding diagram for later versions would include terms for chemotherapeutic response, which is not considered here.

In this case the translation of the model into  $\Lambda$  is straightforward. Anderson (2003) described the discrete conversion of systems of PDEs describing organism migration. That solution is adopted here:

$$\frac{\partial n}{\partial t} = D_n \nabla^2 n - \nabla \cdot \left( \frac{\chi_0 k_1}{k_1 + c} n \nabla c \right) - \nabla \cdot (\rho_0 n \nabla f) \quad (6.7)$$

$$\frac{\partial f}{\partial t} = \omega n - \mu n f \quad (6.8)$$

$$\frac{\partial c}{\partial t} = -\lambda n c \quad (6.9)$$

$$\Rightarrow n_{l,m}^{t+1} = n_{l,m}^t P_0 + n_{l+1,m}^t P_1 + n_{l,m+1}^t P_2 + n_{l-1,m}^t P_3 + n_{l,m-1}^t P_4 \quad (6.10)$$

where  $t$  is time;  $f$  and  $c$  are fibronectin and TAF concentrations;  $n$  is EC population.  $\frac{\chi_0 k_1}{k_1 + c}$  is a chemotactic function concatenating TAF decay into TAF diffusion.  $\rho_0$  is the haptotactic coefficient  $\omega$  and  $\mu$  are fibronectin production and uptake constants.  $\lambda$  is the TAF uptake constant.  $n_{l,m}^t P_i$  is the probability distribution for location  $[l, m]$  at time step  $t$ .

The PDE system on the left is converted to the discrete, stochastic agent-based model of EC movement (right) and diffusion was solved on the lattice as described earlier. Anderson (2003) [6] develops a discrete solution to this system of equations, which was adapted by adding a  $z$  term of the same form to Equation 3 in that paper. Concisely:

$$\frac{\partial n}{\partial t} = X + Y + Z \quad (6.11)$$

$$X = \frac{\partial}{\partial x} \left( D_n \frac{\partial n}{\partial x} - n \sum_{i=1}^w \chi_i \frac{\partial m_i}{\partial x} \right) \quad (6.12)$$

$$Y = \frac{\partial}{\partial y} \left( D_n \frac{\partial n}{\partial y} - n \sum_{i=1}^w \chi_i \frac{\partial m_i}{\partial y} \right) \quad (6.13)$$

$$Z = \frac{\partial}{\partial z} \left( D_n \frac{\partial n}{\partial z} - n \sum_{i=1}^w \chi_i \frac{\partial m_i}{\partial z} \right) \quad (6.14)$$

where  $D_n$  is the diffusion constant for migration (in this case, EC),  $m$  is a vector of length  $w$  of molecular stimulants (in this case, TAF and fibronectin) and  $\chi$  is a migration bias for each  $m$ . The probability equation becomes:

$$n_{k,l,m}^t P_0 + n_{k,l+1,m}^t P_1 + n_{k,l,m+1}^t P_2 + n_{k,l-1,m}^t P_3 + n_{k,l,m-1}^t P_4 + n_{k+1,l,m}^t P_5 + n_{k-1,l,m}^t P_6 \quad (6.15)$$

In other words,  $P$  describes the probability of an EC moving in  $\pm x/y/z$  directions (denoted above as  $k, l$  and  $m$  respectively). Using the central difference method each  $P$  has the form:

$$P_r = \frac{kD}{h^2} - \frac{k\gamma}{4h^2} (m_{position}^q) \quad (6.16)$$

The 3D grid on which this solution operates is the same as  $\Lambda$  therefore translation into the simulation is direct. The two modifications made here are to simulate the angiogenesis rules in hours rather than the 1.5 day time step used in the original model and to use a 3D version of the stochastic equation given on the right in Equation 6.15. Diffusion of TAF was implemented by reusing the diffusion model solved earlier.

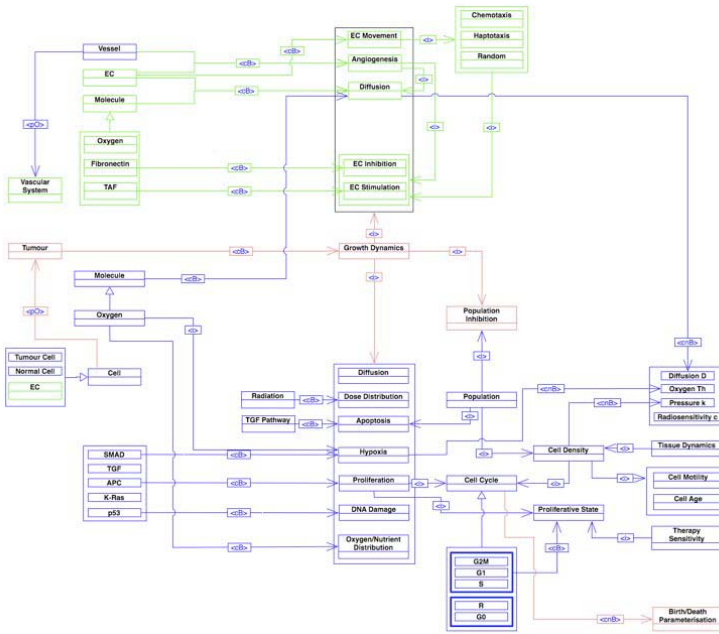
In the original model it was assumed that TAF concentration gradient was constant, using an initialisation that had a downward gradient from cells to vessels and the opposite gradient for fibronectin. This setup was replicated here simply by pre-setting the lattice site states before starting the simulation but was simplified in that it was linear rather than sigmoidal. This was merged with the setup described for the Ribba decomposition except that the parental vessel is moved to the top (i.e.  $y = 0$ ) and ECs allowed to migrate from there (initially 1 cell density).

The integration between the Ribba/Gompertz model and the angiogenesis yielded the BIT shown in Figure 6.18.

Novel  $\iota$  and states added to  $\Lambda$  were determined:

- Endothelial Cell.
  - Implemented as an individual agent within a lattice site.
- Vessel.
  - Implemented as a Boolean state of a lattice site where the state is set to true if an EC traverses through it (this is an extra embedded rule for the EC agent, not shown in pseudocode above).





**Fig. 6.18** BIT diagram of Ribba (blue), Gompertz (red), Anderson (green) integrated model.

- TAF and Fibronectin.
  - Implemented as float states for each all lattice site.

In addition to the rules for Ribba/Gompertz, rules for the Anderson model were implemented as follows (all RC4 with existing Ribba/Gompertz rules):

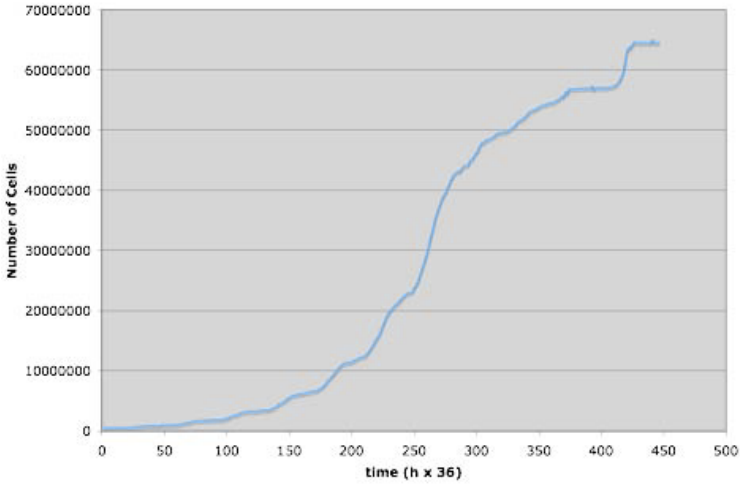
If {true} natural TAF decay (decay=0.6) If {random ;  $P_b$ } create new EC If {true} move EC population according to probability distribution If {true} uptake of fibronectin ( $\mu=0.1$ ) and TAF ( $\gamma=0.1$ )

Where  $P_b$  is the probability of anastomosis occurring based on local TAF concentration (non-dimensionalised as per the original models implementation). For the sake of simplicity of implementation this was adapted to be a discrete step function where:

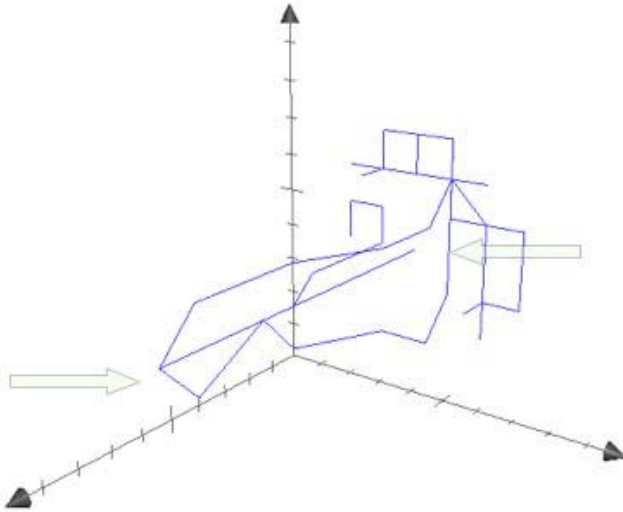
$$0 \leq c < 0.3 \rightarrow P_b = 0 \quad (6.17)$$

$$0.3 \leq c < 0.5 \rightarrow P_b = 0.2 \quad (6.18)$$

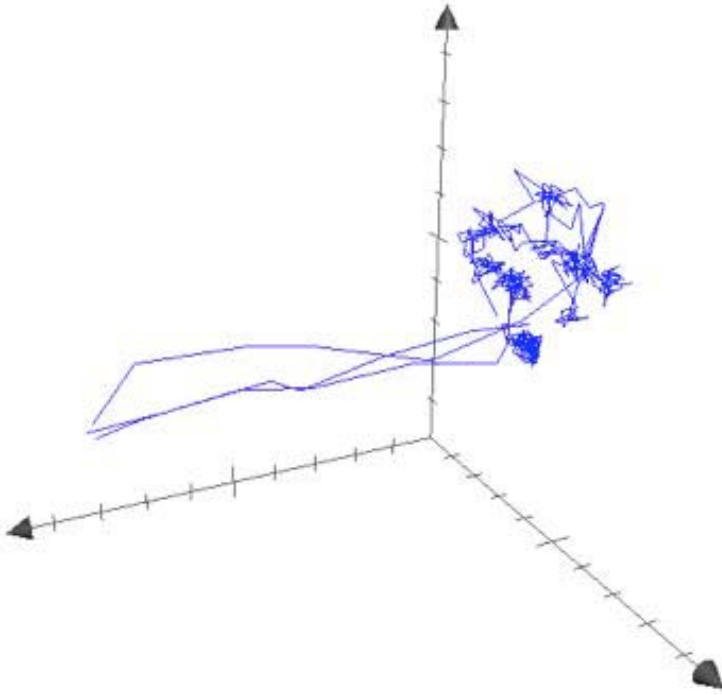
$$0.5 \leq c < 0.7 \rightarrow P_b = 0.3 \quad (6.19)$$



**Fig. 6.19** Growth is erratic as the slower time-steps for angiogenesis progress and deliver oxygen to cells. The shape of the curve is very similar to previous growth curves, however since the tumour is now placed on the side of the simulation space rather than the middle growth appears to be a little more oscillatory since the cells are not free to divide and invade equally in all directions.



**Fig. 6.20** The initial tumour seed is planted (right arrow) at the opposite end of the EC seed (left, arrow). The linear TAF gradient plane ensures a steady directional growth of vessel (blue lines) to tumour site, carrying oxygen with it.



**Fig. 6.21** As the tumour grows outward the vessels back up on themselves (only a small set of separate vessel paths are shown here for clarity). Eventually EC movement becomes entirely random as the TAF bias depletes due to degradation.

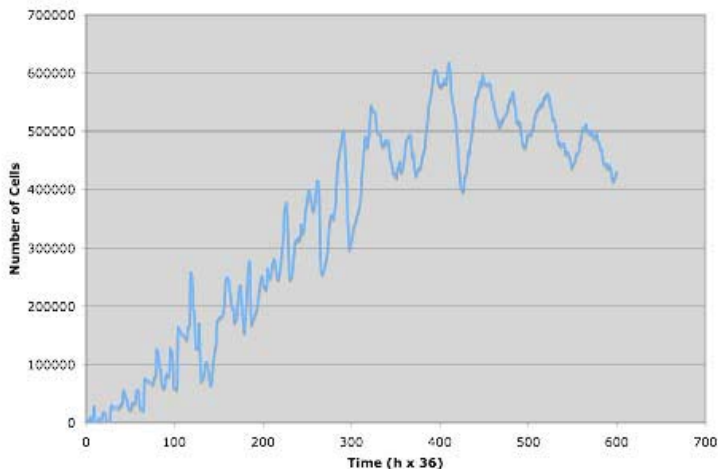
$$0.7 \leq c < 0.8 \rightarrow P_b = 0.4 \quad (6.20)$$

$$0.8 \leq c < 1 \rightarrow P_b = 1 \quad (6.21)$$

With these settings for EC cloning, i.e. splitting of growing vessels, growth behaviour is quite similar to the Ribba curve except stretched over a longer time period (Figure 6.19). Growth is exponential as oxygen supply to cells becomes ever more available. However growth is capped as lattice sites reach maximum capacity.

As the ECs reached the other end of the simulation space, where  $c > 0.8$ , the number of EC duplications per time step increased exponentially. Even so maximum cell number did not reach more than  $7 \times 10^7$ . This is because despite the availability of nutrient the population constraint and death equalise any extra growth. A larger simulation space would of course remedy this. Growth of vessels in shown in Figures 6.20 and 6.21.

The integrated model can now be used for analysis of dependencies between parameters *in silico* that were not possible before with the original models,



**Fig. 6.22** Fluctuations in hypoxia, i.e. total number of cells in a lattice site whose hypoxia state is set to true, whilst vessels grow.

e.g. the distribution and flux of hypoxic cells in a vascularised tumour as illustrated in Figure [6.22](#).

## 6.5 Summary

KDA and ABI have been demonstrated with models from the literature. KDA has been shown to be able to cope with model concepts very well and due to the generic nature of integrative properties BITs from different models can be interlinked whilst maintaining the integrity of the original BITs. This was shown by performing a KDA analysis with 5 models from the literature. ABI was first demonstrated by decomposing the diffusion model and subsequently with the Gompertz model, Ribba (2006) and Anderson (1998).

## 6.6 Conclusion

The ABI strategy is able to decompose models into interaction rules, using BITs to reveal behaviours and components, and hence enabling model integration. Admittedly more experimentation is needed since the KDA/ABI method seems to be capable of more, however for this proof-of-concept the simulations done here suffice to show the methods do work. Critically, agent rules derived from different models can be combined in such a way within an ABS framework to yield a model that is qualitatively similar to the individual components. Since the formalisms, granularity and semantics of the original models are so different an integration like this would otherwise be very difficult to achieve.

# Chapter 7

## Discussion

### 7.1 KDA/ABI Performance

#### 7.1.1 *The Knowledge-Driven Approach*

Concisely, KDA was developed to describe model semantics, i.e. model focus and scope, in a way that was computable (i.e. expressible in a language that can be interpreted computationally and operated on. For example, possibility of scope processing) and in which the process of model integration would be made more explicit by means of providing a computational common ground. This is parallel with the idea of using local interactivity in ABI as the common ground. KDA recognises that the commonality between all models, regardless of focus, scope, scale or paradigm, is that they all relate a set of system concepts, specifically *behaviours*, to a set of foci. This is indeed at least true for the cancer modelling field:

- There is always a focus, i.e. the tumour or related subsystems/components.
- A focus is always attributed with state changes, i.e. behaviour.
- Though not always explicit, there is a persistent notion of associations between behaviours, e.g. inclusion or exclusion of known processes or black-boxing.

Hence the generic classes and properties in KDA were distilled from an in-depth literature review. Note the great advantage of the generic nature of KDA is that its specification is not restricted to biological systems but any system that conforms to the structure and definition of system as described in Chapter 4.

##### 7.1.1.1 Case Studies

The 5 models that have been successfully described in detail with KDA in the previous chapter illustrate the point that there are actually very few concepts about a model that cannot be described by KDAs basic classes (or extensions thereof). One case that was not tested in this book is the non-functionality<sup>1</sup> paradigm.

---

<sup>1</sup> This was due mostly to the fact that there are very few non-functional models in the cancer modelling literature if any at all.

However even in this case, though behaviour and hence behavioural inclusion are not an inherent property of the model, the KDA technique still affords the possibility of describing the model on the basis of its connection with the real system, i.e. its focus and relationship between foci. The few concepts that could not be covered by KDA are as follows:

- Chen (2004) [43]
  - Logical rules pertaining to how oxygen and nutrients are replenished at the edge of simulation space could be interpreted as assumptions. KDA can only take this as far as those concepts exist, i.e. Oxygen and Nutrient Production, Consumption and Diffusion concepts. However these also can be interpreted as a property of the solver and hence have no place in KDA. They do not add any knowledge of the foci and how behaviours are related with each other.
- de Pillis (2005) [56]
  - None
- Mallet (2006) [119]
  - Assumption relating to the equivalence of diffusion constants, to keep in line with an earlier model on which this is based. This was generically inputted into the BIT as a Diffusion concept (for each diffusible molecule).
- Markus (1999) [123]
  - None
- Zhang (2007) [207]
  - None

A consistent consideration whilst building these BITs was whether initial conditions should have been included. Initial conditions do indeed affect behaviour of the model and should be in the inclusion tree on that basis. However it can be argued that they are also part of the solver, from which KDA is supposed to be abstracting away and so were not included. This decision was also based on the fact that the process of model integration was not contingent on any assumptions related to initial conditions. Indeed if the process of decomposition into rules is successful the initial condition setup should be replicable within  $A$ .

### 7.1.1.2 Evaluation

Based on more than 40 decompositions, it can be concluded that the core requirements have been fulfilled:

1. High level of capture of the semantics of models.
  - As described above, there are very few concepts that are not immediately translatable into a BIT. In fact, not only can this approach describe almost all

model concepts, it often exposes relationships that are not made explicit in the model, e.g. the chain of behavioural inclusion.

## 2. Direct and dynamic mapping of actual system components.

- The representation should fully reflect the original modellers knowledge of the system, i.e. what is termed expert knowledge in the literature.
  - KDA requires the user to build relationships between systems concepts and system components. Much of this will come from the model itself. However integration of BITs requires the users knowledge of the system to expose behavioural inclusions and equivalence as well as disambiguating system components.
- The representation should be robust to change in or introduction of new knowledge.
  - There was not a single case in any of the integrations of BITs that required a change in properties or classes (i.e. the structure and therefore semantics of the BIT).
  - A single implementation is robust to change insofar as rearrangement of elements is concerned. However an integrated BIT is hypothetically more difficult to manage in that once a major change is needed one may need to go back to the integration (ABI) stage. That said, however, a change in KDA (i.e. belief in the way the system is behaving) would constitute a change in the model and so reiteration into ABI would be necessary anyway.

## 3. High level of expression with a relatively simple representation to avoid error.

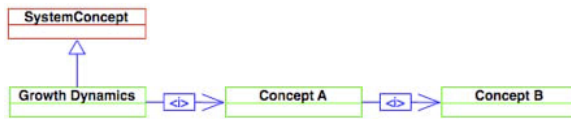
- Flexibility in what can actually be represented is vital as complex models of complex systems will presumably go hand-in-hand with complex semantics. Moreover these semantics may well incur human error when the (graphical or exchange language) representation is difficult to understand or generate.
  - Though BITs become progressively complex and difficult to draw neatly one must weigh this minor disadvantage against what the diagrams represent. Even for simple models there is a wealth of semantics that must be described and the BIT diagrams manage to do this using generic components in a relatively straightforward fashion. Moreover it would seem that there is a point of critical mass where the proportion of concepts in the BIT actually starts to fall as more models are added to it since the existing BIT already contains many of the behaviours already.

## 4. Portability

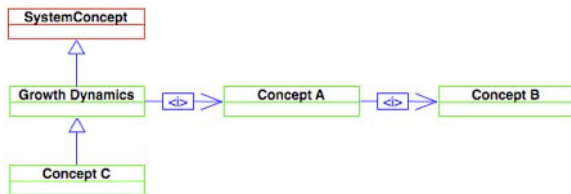
- Preferably the knowledge that  $\Phi$  represents should be portable to alternate applications, e.g. via some shared standard.
  - A BIT including all models in the review chapter was constructed simultaneously with a corresponding ontology in Protg 2000 (not shown here due to space constraint). Though strictly speaking this is not an

ontology-building exercise<sup>2</sup> due to the fact that this is a well-structured technique that is based on a set-theoretic approach it lends itself to a similarly structured language, e.g. an XML derivative.

It can be argued that point 2 above introduces ambiguity into this technique in the sense that association of concepts, i.e. expert knowledge, is subjective to the user who is interpreting the semantics. This ambiguity should not manifest at the individual model stage, i.e. developing a BIT for a single model. The BIT should only contain the concepts that the original model represents (whether or not it is congruent with the users belief of the system). This does, however, present itself when integrating BITs together. Take, for example, the generic nature of system concepts in a BIT from hypothetical Model 1:



And subsequent integration with Model 2, which introduces Concept C:



Concept C now implicitly includes Concepts A and B due to the transitivity of the <includes> property. It is conceivable that conflict can occur if Concept C is indeed a child of Growth Dynamics, i.e. it fulfils the property of inheritance and is validated as such, but does not include Concepts A and B in reality or the association is explicitly negated in Model 2. Though this may at first seem to be a flaw, it does actually highlight one of strengths of the KDA technique with respect to model integration. If such a conflict does occur it would strongly imply that the two models that are being integrated are in fact incompatible as far as current knowledge and belief of the system and the concepts in question are concerned. One would certainly not want to integrate models that have any fundamental differences in belief of the way the system works. Not only does this cause conflict on the level of KDA, such a conflict would manifest at the ABI level in the form of conflicting rules or agents (t). KDA therefore can explicitly highlight conflicts in association of concepts.

<sup>2</sup> Note that the aim is not to explicitly define what behaviours and concepts are (only to relate them with each other using generic properties). Concepts by themselves only attain a meaning once related to other concepts. Moreover there is a clash between the meaning of subsumption in ontologies and the meaning of the <includes> property: OWL and related languages are therefore not a good choice for portability.



### 7.1.1.3 Novelty, Current Works and Future Developments

The idea of describing model semantics is not entirely new. From the outset there were a handful of supporting technologies including:

- Category theory.
  - Abstract and flexible, this method can describe mathematical relationships and morphisms. However it lacks the ability to relate the meaning of the models to the real system.
- Semantic networks (SNs).
  - Shows real promise in describing the relationships between concepts. Indeed KDA shares many ideas with these, however SN representation of concepts and relationships is relatively loose whereas KDA presents a structured and defined approach.
- OWL-supporting technologies (discussed shortly).

Model description has taken shape in as many varied forms as the fields in which they appear. Indeed different fields have had different requirements and hence diverse forms of model representation, as explained in in Chapter 4. Almost all have taken a software approach, e.g. MSI and simulation communication technologies from DARPA<sup>3</sup>.

#### *KDA in Systems Biology*

In the field of systems biology there are at least four standards for model description that are predominant or emerging:

- SBML [91]
- BioPAX [173]
- CellML [114]
- CMDL [121]

SBML was originally meant for general model representation<sup>4</sup> but evolved into the exchange language of choice for molecular pathways. The XML-based language separates two useful representations: pathway components and their mathematical relationships. SBML is however limited in its mathematical representation by MathML, as is CellML, though this does not yet pose a problem as most models in systems biology are P/ODE based for which MathML copes well. Importantly, the planned Level 3 version of SBML will enable submodels to be embedded within larger contexts. This is aligned with the idea of model integration and indeed will encourage integration of sub-models (modules). However, note that such integrations will always be linear. The formalisations in Chapter 4 wholly describe this. Additionally there are no plans to include any description of model semantics though it

<sup>3</sup> Defence Advanced Research Projects Agency (<http://www.darpa.mil/>)

<sup>4</sup> Andrew Finney, personal correspondence.

can be argued that some of these are implicit, e.g. compartmentalisation of reactions and modularisation of (sub)models. Conversely, KDA cannot describe mathematical relationships explicitly but can describe model semantics and hence these two techniques highlight two very different angles from which model integration can be approached:

1. The traditional approach, wherein the mathematical formalisms are dealt with directly. The possibility of model integration is severely limited in this case by paradigm, focus and scale differences.
2. The conceptual approach, wherein the mathematical relationships are considered separately with respect to focus.

Like KDA, BioPAX attempts to connect models to focus. Widely promoted as the next step in describing pathway models after SBML its complexity, in the opinion of the author, has hampered uptake by the systems biology community. BioPAX has gone a step further than SBML in that not only does it allow mathematical relationships to be described, i.e. pathway models, it simultaneously attempts to build an ontology of components. This is in line with the idea of having a computable<sup>5</sup> form of focus as KDA does. However the responsibilities of mathematical *model description* and *model integration* have been delegated separately to KDA and ABI respectively in the work presented here whilst the authors of BioPAX have attempted to do this concurrently (though admittedly model integration is not the primary aim of the BioPAX project). Moreover KDA does not enforce an ontology as such on the modeller whereas BioPAX uses a fixed set of concepts agreed upon by a number of experts<sup>6</sup> and therefore enables individualised model integration exercises. However, this highlights a disadvantage of KDA in that once a number of iterative model integrations have been performed the corresponding BIT diagrams and concepts therein may well vary when compared to other modellers representations - the advantage of disambiguity that BioPAX enforces is lost. On the other hand, there is no reason why future developments of KDA should not adopt a strict approach so that SystemConcepts, Systems and Components in particular are standardised. Indeed, upon closer inspection, the vast majority of concepts in BioPAX already fall neatly under the definition of Component and so a future integration of these projects is highly desirable. Such an integration will allow modellers to use standardised components that can be translated into  $\Lambda$  and reduce the chances of conflict between modelling groups.

Similar is the case for CellML, which can be broken down further into two languages, AnatML and FieldML, which describe anatomy and mathematical properties respectively (though it seems development in both of these has slowed of late). Indeed AnatML seems to be a macroscopic version of what BioPAX aims to do

---

<sup>5</sup> Inference engines can be used to draw relationships between components and hence mathematical relationships can be exposed.

<sup>6</sup> In fact cancer systems biology is taking a leading role in ontology-building and disambiguation of biological knowledge. The drive for therapeutic improvement has yielded a number of projects such as those initiated by the NCI (discussed shortly) and the CViT project for example.

on the molecular level, i.e. description of what is defined as Component in KDA, whilst FieldML was originally designed to express fields such as finite elements or splines (and hence is a good platform on which O/PDE solutions can be described). Again, the explicit mathematical components that CellML can describe are missing in KDA and conversely the expressive power of system concepts is missing in CellML (barring the expression of  $t$ , which AnatML specifically aims to state).

Whilst SBML, BioPAX, CellML and KDA have all focused on the model and system the development of model integration technology by members of the Beacon Project<sup>7</sup> at UCL currently aims to conceptualise the wealth of data that is used, e.g. experiments that yield parameterisation values for models, in the form of a well-structured repository [154]. This is a very important point as it provides the means to explicate where data comes from and how it was derived, and incorporating this with the idea of a lightweight model execution environment (Margoninski et al., 2006) affords the possibility of a full MMS/IME. The supporting language, CMDL, conforms to the linear integration strategy<sup>8</sup>, albeit a more complex one than that of SM and other previous works. This is because CMDL provides multiple interfaces for each submodel, i.e. the solver is separated from the actual model. Note that model interfaces and connections need to be provided, as expected in any linear integration. Though engine wrappers and connectors can be made generic enough to deal with individual models on a paradigm-by-paradigm basis<sup>9</sup> it seems that the development of this book will be contingent on building a rich enough set of wrappers to deal with the huge diversity of model types in systems biology. In contrast, KDA is independent of paradigm since the point of using two steps (KDA/ABI) is to separate the model meaning and model implementation.

Interestingly CMDLs concept of a biological phenomenon (aspect in the original metamodel, see Figure 2 in [69]) appears to be concordant with SystemConcept in KDA although the phenomenon element in CMDL contains textual information. Again, this highlights the fields current stance to that which is described as focus and scope in this work. Much of this metadata is kept in textual form and precedence has been given to the mathematical properties of the model itself. In contrast by defining simple, extendable components KDA provides a set-theoretic way to describe aspects and hence make these computationally meaningful<sup>10</sup>. Additionally the compartment element in CMDL bears close resemblance to the component and <partOf > relationship in KDA. There is therefore a possible point of bridging between these methodologies here. In the opinion of the author CMDL can greatly benefit from a straightforward implementation of KDA in addition to the aforementioned textual information in the phenomenon and compartment elements.

CMDL/KDA as a whole could therefore provide model description, and thus a standard of model exchange as well as a platform on which ABI can be automated (discussed shortly), at a level that is not yet present in the field of systems biology.

<sup>7</sup> <http://grid.ucl.ac.uk/biobeacon/>

<sup>8</sup> Contexts (inputs) are fed into the engine (solver), which provides an interpretation(s) (outputs), which can in turn provide context for another submodel(s).

<sup>9</sup> As far as author is aware, only P/ODE have been tackled so far.

<sup>10</sup> KDA would obviously need to be implemented into a language, e.g. based on XML.

This does however raise the question of controlled vocabularies, i.e. consortium-driven agreement on SystemConcepts, as was the case with BioPAX. The rapid expansion of controlled vocabularies such as the internationally driven open-source EVS<sup>[11]</sup> project from the National Cancer Institute (NCI) is a good start. Though this does not address model foci it does hold promise in terms of internationally coordinated efforts to ontologise cancer-related data. In fact the National Cancer Research Institutes (NCRI<sup>[12]</sup>) Informatics Initiative aims to create coherent and compatible informatics platforms enabling the deposition of a wide range of accessible cancer-related data. This is certainly an important step forward in bringing expert medical knowledge to systems biology, and hopefully in explicating a diverse set of SystemConcepts for use in KDA.

One of the greatest disadvantages of the KDA methodology is the fact that every model, which is usually published either as plain text in a paper or put into a modelling language such as SBML, must be re-conceptualised into SystemConcepts, etc., manually. There is no possibility of automating this since none of the exchange formats provide a way to describe these concepts (and hence there is no obvious mapping) and of course it is challenging to extract this kind of information from plain text (even assuming all the information that is needed is there). However, the benefits of manual generation far outweigh the effort. Along with the mathematical description the connection a model has with reality is concretised into a format that software can then begin to use intelligently. In the opinion of the author this is one of the next great challenges for systems biology. Not only to model biological complexity but to provide intelligent software systems that can infer knowledge about the biological system, extract and present information and hence provide the means to understand the system. Ultimately in the case of cancer research this means determination of therapeutic advantage by (i) enabling model integration and hence a complete simulation of tumour behaviour and (ii) explicit definition of system concepts and their relationships leading to better human understanding. An example of why such intelligent software systems are needed has been alluded to earlier in this section where the transitivity of the <includes> yields relationships between SystemConcepts that were not explicitly stated in the original models. As more and more models are expressed in KDA the associations between elements become massive and hence a software aid becomes necessary. Indeed such an inference technique would represent a truly exciting future development for this book - as  $M_I \rightarrow \Phi_I$  one comes a step closer to possessing a computational model of the system, not in the mathematical sense but in a way that enhances the ABI strategy (discussed shortly). With that in mind considerable effort was put into searching for an appropriate technology to demonstrate this. Two immediate candidates were Protgs inbuilt inference engine (which was subsequently dropped because of irregularities), KAON2<sup>[13]</sup> and Jena<sup>[14]</sup>, which are both open-source projects. There are other packages but most are commercial or lack support.

<sup>11</sup> [http://ncicb.nci.nih.gov/NCICB/infrastructure/cacore\\_overview/vocabulary](http://ncicb.nci.nih.gov/NCICB/infrastructure/cacore_overview/vocabulary)

<sup>12</sup> <http://www.ncri.org.uk/> and <http://www.cancerinformatics.org.uk/>

<sup>13</sup> <http://kaon2.semanticweb.org/>

<sup>14</sup> <http://jena.sourceforge.net/inference/>

Overall KDA has successfully fulfilled all design requirements, enabling both model (and therefore system) description and ABI; but what is especially exciting is its future applications. The integration of ideas across CellML, CMDL and BioPAX would yield an extremely rich language that not only provides the ability to exchange models but also enables model reuse, integration and redevelopment.

## 7.1.2 *Agent-Based Integration*

### 7.1.2.1 **Model Decomposition: Genetic Algorithm Performance**

At the core of the ABI method is the optimisation algorithm that can enrich an initial set of rules to perform in  $\Lambda$ . The reasons for choosing GA were as follows:

1. A large solution space was always expected, even when starting off from a good initialisation set. This is the trademark problem for which GAs in particular have proved to be especially good [127].
2. The structured and compartmentalised form of string-based rules made GA a natural choice when considering how to perform mutations, crossover and mating on rule sets, e.g. mutation on numbers and operators and hetero/homogenous crossover between rules.
3. There is a history, albeit a brief one, of use of GA in rule optimisation for CA (discussed shortly) and hence it was already a proven technology for such use.
4. The evolution of rules can be tracked, i.e. logging the evolution of each individual at each generation which could indicate how best to perform optimisations of a similar kind in future.
  - a. A corollary to this is that GA is not a black-boxed approach.
  - b. This is not yet implemented in the software produced here but represents a very important future enhancement to the software.
5. A number of candidate solutions can theoretically be found.
  - a. Whether this is an advantage or disadvantage depends on how the results are used. If candidate solutions turn out to be very similar (in terms of structure and magnitude of variables) then there is no problem. However if there are hugely varied rule sets then it must be up to the modeller to decide what is biologically feasible based on current knowledge.
6. GAs are very configurable: the GA had 39 configurable attributes based on probabilities for mutation, crossover, mating (and sub-attributes thereof consisting of operator types) as well as choices based on fitness function attributes.
  - a. Whilst this can be a great advantage the reverse argument is that too many choices can lead to a situation where the optimal configuration for a specific experiment is difficult to determine. This is the reason why a benchmark was necessary, i.e. to determine the best conditions for rule enrichment. It must be noted, however, that though the same configuration is not expected to be optimal it does at least give an approximation. In other words, the configuration

for enrichment for the diffusion model rules is unlikely to be the optimal conditions for other more complex models but at the very least indicate an interval in which one can avoid GA individual extinction or a hanging population.

7. Fault tolerance: a small number of good individuals amongst a population of bad performers can eventually come through after a sufficient number of generations. This is levied against the stochastic characteristics of the approach in that good performers may well lose their advantage due to random mutation.
8. Relatively straightforward to program and integrate with the simulation software.

As with any method there are disadvantages:

1. There are potentially a massive number of simulations that must be performed to find the optimal solution, especially since the starting population in most of the experiments performed here consisted of 100 individuals. This can make single run times arbitrarily long. This was combated in three ways:
  - a. Simultaneous application of the fitness function. The original model was run alongside the simulation and the simulation logs were sampled periodically to check for key states, e.g. population number. A cache of best performers at the same time points were stored. If the current simulations state values at those time points had an error larger than that of worst performer in the predefined subset of best performers then it was immediately rejected and simulation was stopped. Otherwise it is added to the best performer list, and displaces the bottom-most individual from that list.
  - b. A cache of the string-based rules so far performed is stored along with the performance of that set. If the same rule set was encountered the simulation was not performed.
  - c. Parallelisation: 15 machines were used and split into 3 quintuplets to perform 3 simultaneous GA instances.
2. Too many configurable components (see earlier point).
3. The model runner embodies the traditional problem in model integration, i.e. the need for a communicative interface.
  - a. This would be the case for any method.
  - b. Integration with SOAs would greatly reduce this problem.
4. Use of GA may well be excessive for some decompositions.
  - a. GAs are especially good at finding solutions when there are many variables.
  - b. Other methods can be used since those parts of the software were designed generically.

Optimisation is an entire science unto itself. There is a plethora of methods that are continuing to be developed. GA was chosen as the most natural and straightforward choice, however there are several alternatives that could be used considering the qualitative as well as quantitative optimisation that is sought:

1. Neural Networks, Bayesian Networks and other learning algorithms.

- a. There is a vast array of neural network and other learning methodologies catering for a multitude of problem sets. However it was not clear how a learning algorithm could have been implemented considering the core elements involved, i.e. an initial rule set, a target behaviour and a required target rule set.
  - b. One way in which these could yield fruitful results is the learning of relationships between rules that have already been optimised and original model attributes, e.g. parameters. This is desirable because the ABI strategy yields a result (rules) for a particular model instantiation, not a model class. The more generic rules can be the more powerful ABI is as an integration tool.
  - c. Another way these can be applied is reinforced learning of rules in agents during simulations. A disadvantage of this method would be that rules would be black-boxed. An outline of the workflow would be:
    - i. Initialise the agents in  $\Lambda$  with naive nets embedded in rules, i.e. inputs would be states/environment, etc. and output would be next state/location for agent.
    - ii. Begin simulation.
    - iii. Each time step is evaluated by fitness function.
    - iv. Good/bad performance back-propagated to nets.
  - d. Effort was put into implementing a GA-modifiable module for this book by adapting Joone<sup>15</sup>, though it was never put to use.
2. Artificial immune systems show some promise in searching large state spaces. The principles that govern the immune systems learning are, however, quite similar to that of evolutionary techniques. They have been especially identified as good candidates for nonlinear pattern recognition and hence may hold more promise in recognising behaviour concordance in the fitness function rather than an optimisation technique itself [54], as well as in a similar role as that described above.

### *Other Improvements*

An alternate line of investigation that should be undertaken is in the way GA actually integrates itself with the simulation environment. A vast improvement was observed simply by reprogramming the fitness function to simultaneously measure fitness during the simulation process.

The idea of more efficient and smaller simulations could be taken further by borrowing from Patch Dynamics theory. For example, let  $\Lambda$  be the simulation space with  $I$  interactors and  $t_i \in I$ . Also let space and time in  $\Lambda$  be discretised,  $\partial_{xyz} \in D$  and  $\tau_k \in T$  with both  $\partial_{xyz}$  and  $\tau_k$  considered intervals. The original model can be abstractly symbolised as  $f(d, t, \mathbf{X})$  where  $\mathbf{X}$  is a vector of inputs (parameters or inner functions). One can now develop a new GA scheme wherein fitness is evaluated at predefined patches in the simulation, both in space and time, such that

<sup>15</sup> <http://www.jooneworld.com/>

$f(\partial_{xyz}, \tau_k, \mathbf{X}_{xyz,k})$  is evaluated for its micro-behaviour in  $\Lambda$  but not in intervals in between. This would greatly reduce simulation and evaluation time as a side effect but more importantly this raises an exciting avenue of investigation. Decomposition of behaviour of a model in its entirety could hinder the progress of a GA; the burden of this complexity is lifted from the GA in this new approach and could lead to more positive results for the most extensive and complicated models.

### 7.1.2.2 Case Studies and Analysis of the KDA/ABI Strategy

#### *Case Studies*

The diffusion model was chosen as the first decomposition since:

1. It is a relatively simple PDE.
2. Many models incorporate it, or variations of it, and hence it is a valuable starting point.
3. It has a strong theoretical basis.

The movement of molecules, modelled by lattice state changes, drove the need to create specialised State-implementing classes that reciprocated molecular movement. This was done in response to initial experimentation where the GA could not develop the right combination of simple rules that would ensure mutual, i.e. equal and opposite, state changes in adjacent sites, hence conserving mass. In other words if there is  $c$  flow from lattice  $l_1$  and  $l_2$ , where the rule being executed belongs to  $l_2$ , there must be a corresponding loss of  $c$  at  $l_1$ . This was an interesting result in itself in that such an optimisation would have required a co-evolutionary strategy, i.e. a rule for incoming and a rule for corresponding negative flow between the same lattice sites. This is a clear indication of the criticality of three points of the optimisation process:

1. The importance of an optimisation algorithm that can co-evolve such rules.
2. Prior knowledge is needed and embedded in initialisation of rules (in this case, conservation of mass).
3. The dependence on software components that can aid the optimisation process: in this case the State-implementing classes that automatically ensure conservation of mass.

The GA traversed more than 1700 simulations<sup>16</sup> before reaching a solution. It is interesting to see the basic form of the resultant rules:

- If{true} move in x,y,z-direction  $[(D/d) \times c]/a/b$

Where  $a$  and  $b$  are constants,  $D/d$  is the given diffusion constant as a ratio of  $d$  and  $c$  is the concentration in the parental lattice site. In actuality all results were more complex than this but could be distilled to this form. For example, due to

<sup>16</sup> An exact number cannot be given since logging was intermittently suspended for efficiency.



chromosome replication and mutation, some rule sets were found in duplicate but their mathematical effect could be distilled to the form shown above<sup>[17]</sup>. The best results for  $a$  were found to be 5.698, 5.744 and 5.898; best results for  $b$  were 1.976, 2.132 and 2.134 (all values given to 3dp, actual values in 14dp). This is a very exciting result as it closely reflects the theory on which Ficks 2nd Law is actually based.  $D/d$  represents a resistance of molecular flux based on the interface between lattice sites. A von Neumann neighbourhood was employed, hence the divisor that tends to 6 ( $a$ ), i.e. the number of neighbours through which there is a 2D area that flux can pass, and correspondingly there is an equal chance of flux in the negative direction for each pair of lattice sites based on the current concentration ( $c/2$ ) and hence  $b$  tends to 2 <sup>[135]</sup>.

The significance of this result, even though this is a relatively simple model, cannot be overstated. Not only has decomposition led to a rule set that solves the original model the rule set actually reflects what is happening on the molecular scale<sup>[18]</sup>. Yet this result should not be surprising. With the right initialisation rules and conservation rules implicit in states it should be natural that such a simple local interaction be evolved (even though the actual forms of the rules in the successful result sets were more complex).

The Gompertz model, being a phenomenological model, could only be decomposed by KDA into a very small set of concepts. Moreover the model does not yield any information about the mechanics of the system and hence there could never be any determination of  $t$  (barring population number and hence cells as agents). Such being the case there was little for the GA to operate on and establish a successful rule. Indeed the initial rules that were used to find a solution were based on the fact that the curve is rate dependant (i.e. it can be expressed as an ODE). It is interesting that none of the 18 GA runs yielded a result. In a separate 2 experiments the population of cells was implemented as a global variable with the same rate-dependant rules. It would seem logical to assume that the GA would be able to find a solution to a relatively simple model however even in the 2 runs with rate rules on the global variables a solution could not be found. This could be a case of a GA being an excessively forceful method seeing as there are relatively few variables to operate on. This is reminiscent of the findings of <sup>[29]</sup> where simple binary CA rules could not be optimised.

A successful result could only be procured once finer grained concepts were introduced. This highlights an aspect of the KDA/ABI strategy that has already been mentioned earlier, which is that the method does require human interaction

---

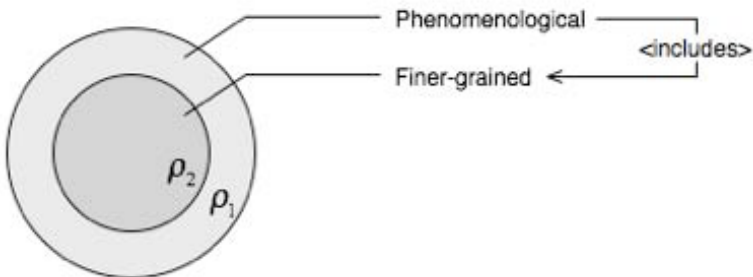
<sup>17</sup> An inbuilt mechanism to reduce rule complexity would be an excellent addition to the GA software. Not only would this simplify rule interpretation by the modeller it would greatly reduce execution time since the number of rules to be executed is potentially cut (certainly in the diffusion case the rules were repetitive, often identical, with mathematical components that could easily have been concatenated by simple arithmetic.

<sup>18</sup> Note that the best results have tended to the theoretical values of 6 and 2 for  $a$  and  $b$  respectively but not reached exactly. This is probably a side-effect of using the Euler difference method, which accumulates error with each progressive step.

and knowledge to complete integrations. This is not however a major disadvantage - human interaction is always expected. Even with the linear integrations described by Geoffrion (1989) [79], Dolk (2000) [62], Villa (2001) [187] and Takahashi (2004) [177] a certain amount of human input has always been needed. In conclusion it has to be said that inputting additional knowledge into the BIT that is not present in the original model is not in KDAs mandate. This is perhaps a limitation of the approach in that purely phenomenological models do not yield enough information.

All subsequent experiments were extremely encouraging. As explained in the previous chapter, in the proof-of-concept implementation, the Ribba model was simplified for ease of implementation, though note that this would not have affected the integration process aside from computational tractability. The pathway model was left out specifically because of the large number of components (i.e. molecular species) that were involved, which would have had to be implemented to each cell (agent) individually. Obviously with the limited resources available (in fact even if there were more nodes) this would have imposed limitations on the integration. This is probably why the original model was implemented in 2D, since the number of possible cells in the simulation space given the population-based growth constraint would have only been a fraction of the number observed in a 3D space. Nevertheless it must be said the pathway model, implemented as a Boolean network, could still have been implemented with a metamodelling approach, such as asserting same-state mechanics within lattice sites. This is exactly what was done with population number with this model, i.e. instead of modelling cells individually population number (separated by cell cycle stage) was implemented as a lattice state. As the results show this actually works very well, especially since the simulation successfully imitates the behaviour of the original model, and is in fact a viable solution to the scale problem.

Integration between the Ribba et al. model and the Gompertz model was relatively straightforward seeing as the Gompertz SystemConcepts, being entirely phenomenological, subsumed the behaviours of the Ribba et al. SystemConcepts. Under the belief framework it must be assumed that phenomenological-derived rules and parameterisations must take precedence over child rules and parameterisations. This is because in KDA the use of the <includes> property asserts total behavioural inclusion. This idea is best illustrated graphically:



KDA asserts the rule set relation  $\rho_1 \leq \rho_2$ , i.e. there exists an unknown rule set of unknown concepts that additionally, along with  $\rho_2$ , contribute to the behaviour elicited by  $\rho_1$ ;  $\rho_2$  can never fully explain phenomenological behaviour and thus in a belief framework the parameterisation and rules of  $\rho_2$  must be superseded by  $\rho_1$ . Hence in the case of Gompertz/Ribba integration preference of population thresholds and cell cycle lengths was given to the Gompertz rules.

As Figure 6.16 shows the integration has had the effect of shifting the original Ribba curve, which was less constrained by population, towards the Gompertz curve (although admittedly the original Ribba et al. model was almost Gompertz-like already). The tighter population constraint also explains why the maximum growth is achieved earlier and lower. Also, the introduction of a death term ensures that cells that fall into  $G_0$  do not simply revive and continue growth. The additional complexity of rules contributed by the Ribba model maintains its original behaviour, which is a significant finding. Had the system been chaotic in the sense described previously one would expect erratic fluctuations due to changes in initialisation or rules. Instead, as the original Ribba model showed when vessel number and other characteristics were changed, the behaviour is not erratic. This is a key property that separates random and chaotic systems from ordered and structured systems.

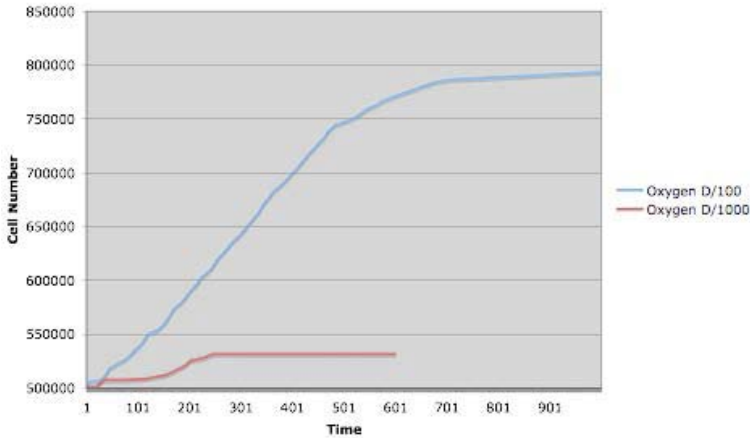
Note that though the Gompertz model is wholly phenomenological, and that one should not expect its behaviour to contribute any value to a finer-grained model that is totally subsumed by it, the integration between the Gompertz curve and the Ribba model has yielded information about the behaviour of both. The dependence on population constraints and cell cycle length in the Ribba model is highlighted and it has now been shown that a more Gompertzian-like growth can be achieved by finding the right balance between these factors, which is provided by the integration.

### *Evaluation Based on Case Studies*

These arguments lead one to the crucial point of model integration: Does integration actually yield a model that is of more value? Does it shed any more light on the behaviours of the original models and the behaviour of the actual system? In this case:

- The resulting model yields a richer behaviour than the original parts.
  - Neither model show similar behaviour, aside from growth in population and an eventual levelling in the Ribba model.
- The Gompertz model does not take into account such things as oxygen levels, but are taken into account macroscopically as population constraint. This can now be divided into both population constraint by neighbouring numbers and constraint by hypoxia. In effect the Gompertzian explanation of the microsystem is richer.
- The Ribba model shows unrestrained growth. Integration has yielded a model that is far more familiar to MTS growth since this is what the Gompertz model was validated against.

- The resulting simulation is now defined on the local level, which is more than what the Gompertz model could describe.
  - Variations on parameterisation, such as vessel number, oxygen diffusion, etc., can now be simulated with relative ease. These only require a change in the corresponding parameter in rules and the initial setup, e.g. see Figure 7.1 below.



**Fig. 7.1** The effect of decreasing oxygen diffusion coefficient (compare to Figure 6.13). Population number is drastically reduced (the same setup with 1 vessel was used). The growth shape is maintained since the result of reducing the coefficient by a factor of 100 effectively reduces the capacity of oxygen to reach its targets in a measure that is related to the coefficient and hence the hypoxia threshold is reached far faster (blue line). A reduction of a factor of 1000 cripples growth almost completely (red line).

The integration between Ribba and the Gompertz function was a proof-of-concept, however one can imagine a scenario wherein the Gompertz or logistic curve has been fitted from actual data, which then needs to be integrated with a pre-generated model. Curve fitting is a regular and necessary step where a model will be developed first and parameterised after experimental results have become available, e.g. Chignola (2000) [44] attempted to parameterise their growth model with MTS data so that it would behave sigmoidally. Viewing data as a model of the system, this can also be viewed as a model integration of a kind. Chignolas model was relatively simple and only involved the optimisation of two parameters to explain growth, i.e. proliferation and death. In the Ribba model case the pre-defined model was a mature description of the system already and the ABI method was able to enhance its behaviour, i.e. fit it to a known growth behaviour. It is quite conceivable that instead of using a model at all in this methodology, one also can integrate (i.e. optimise rules) directly from experimental data. Whilst this is possible theoretically it has not

yet been attempted here but is certainly a line of future study. This highlights the versatility of the ABI method.

The Ribba/Gompertz integration also draws attention to peculiarities of using this discrete method. It was observed that the latter part of the procured curve actually fluctuated in rate after the limit of growth had been reached. Though growth had seemed to steady, upon closer inspection it was found that a lattice site could often exceed its population threshold by almost double. This was in fact because of the metamodelling methodology used to avoid modelling individual cells. Lattice site states denoted population for that lattice (separated by cell cycle), which would halt growth once the population or hypoxia threshold was met. However, on occasion, when the population was below the threshold and for a single time step there was enough oxygen for the hypoxia state to be set to false a small fraction of cells that had migrated from neighbouring sites had the opportunity to multiply despite the population threshold (since that logic step had already passed). If one were to assume that the population threshold was purely a space constraint then of course such an occurrence could never physically take place. However note that the explication of KDA is advantageous in this situation. Indeed these very caveats highlight the need to integrate knowledge together. In this case additional rules need to be set, or a reordering of logic is necessary, that redefine proliferation and cell movement such that space constraint is not violated. This method therefore not only enables model integration but also encourages further integration by highlighting missing behaviours<sup>19</sup>.

Conversely it is also possible to exclude behaviours deliberately. In this example the pathway module was left out as well as the radiotherapy module. Note that both can still be included if desired: the pathway model was discussed earlier; the radiotherapy model is governed stochastically with cell death or cell recovery occurring according to two probability distributions (simulating single and double strand hits) based on the pathway model status. This could have been implemented as follows within the existing rules:

- If {not in  $G_0$ } Update pathway model
  - See Boolean updates, Table 2 in original paper.
- If {true} generate random number.
  - If {within limit  $[0, x]$ } single strand break.
  - If {within limit  $[x, y]$ } double strand break.
  - If {within limit  $[y, 1]$ } no break.

This particular part of the model can now be excluded by simply removing the rules that were derived for the module. Modularised models are relatively straightforward to dissect in this manner. Of course this is also feasible in linear integration, if inputs/outputs can be matched up, however the difference is that now individual rules

---

<sup>19</sup> Of course this is always down to interpretation and flexibility. It might be the case that the inclusion of additional behaviours is not necessary, depending on the application.

(i.e. behaviours for particular agents) can be dissected with more ease and within the context of the real system.

The successful integration of the angiogenesis model was an encouraging development since many models in the literature have fallen short of describing vascularisation. In this particular case the stochastic solution, i.e. a biased random walk, devised by Anderson was completely adaptable to the simulation software and no GA optimisation was needed. It should be noted that a larger simulation space is needed as this was the only limiting factor to growth. However the integration itself was seamless since all rules represented RC4.

The growth curve is reminiscent of other vascular models in the literature (see [120]) where growth is less-than-smooth, unlike the Gompertz curve, but nonetheless exponential. The exponential phase was uneven, as seen before, but this was now additionally due to erratic oxygen supply. After that phase, however, models tend to differ depending on the parameterisation used. For example, Zhang [207] report a termination of growth as constraining factors limit gliomas, as seen here, whereas Arakelyan (2002) [10] report unrestrained growth. This highlights the fact that different models have been validated by different data and this must be taken into account during the integration process. Indeed KDA should reflect this information.

Correspondingly there is a general trend of growth in the number of hypoxic cells in the simulation space. The oscillations seem to be occurring with the course-grained rule updates of vascularisation (every 36h) causing spurts of expansion. The point is that the integrated model now shows vascularisation as well as growth, which would be something not possible to do without a lot of effort (or starting afresh) from the original models. Certainly the resultant model is producing interesting, biologically relevant behaviour - the hypoxia curve in particular is of interest. It is a well-known fact that hypoxic cells are particularly resistant to radiotherapy [168] since they are arrested in the cell cycle and therefore determination of the distribution of both hypoxia and cells by cycle position in a vascularised tumour is an extremely useful metric. This integration affords the possibility to achieve this since the original models separately contribute oxygen distribution, cell growth, cell cycle and vascularisation.

As always, there are improvements that can be made relating to both the integrated model and the simulation software itself:

- TAF production needs to be introduced.
  - The vessels grew up the TAF gradient, however once TAF had depleted EC movement was more random.
- Fibronectin production needs to be introduced.
- Blood flow
  - Blood flow itself can shape vascularisation and have a dramatic effect on overall topology. A number of LB models, as well as PDE models, of this kind exist and so it is well worth investigating integration of these models.

- Vessel maturity
  - Arakelyan (2002) [10] point out the importance of vessel maturity and its effect on oxygen distribution and tumour dynamics. Currently vessels are assumed to be immediately viable, which is not correct in the biological sense.
- A finer-grained lattice is needed
  - A vessel is currently defined by a Boolean state for that lattice site (i.e. the entire lattice site is converted to a vessel). A finer grained approach could be to have more inner lattice sites, however introduction of shaped agents that can traverse the simulation space would be a better option.
- 3D Visualisation and analysis components
  - This would greatly aid the analytical power of the ABI approach. Normoxic and hypoxic distributions *in situ* with vascular distribution would mark a significant step forward when coupled with the fact that the integration process can add even more behaviours.

### 7.1.2.3 KDA and ABI as a Whole: Evaluation and Future Directions

The core aims of the envisioned integration strategy for this project were:

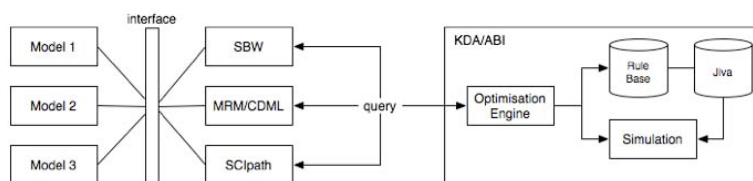
- To perform integrations that are model paradigm independent.
  - This is ensured by the decomposition into local interaction rules.
- Non-violation of focus and scope.
  - This is ensured by KDA.
- Simultaneous means to a 3D visualisation.
  - Though not yet implemented this is possible to achieve.

#### *Analysis and Comparison With Other Methods*

An optimisation method that translates the model into local interaction rules ensures paradigm independence. However one could argue that the job of paradigm integration has been offset to the optimisation algorithm. To some extent this is true - for all integrations described in the previous chapter the implementation of the fitness function worked in conjunction with a specialised model runner implementation with which it was comparing results. Of course writing a specialised implementation for each and every model reduces to the same problem faced by any SOA-like integration strategy such as SBW, i.e. interfaces for every model.

Indeed this very track was explored early in the project with the development of SCIpath [14], formerly MicroCore, which was originally engineered for pathway visualisation and capability for eventual model execution. The need for an integration strategy was quickly identified and a SOA-like architecture was therefore

devised under which different types of data could be exchanged in a structured way. The implementation involved exchange of basic data types in lists or matrices, communicative objects such as ODBC drivers, SOAP interfaces, model language interfaces such as SBML (exchanging nodes and reactions) and a GUI API (for remote GUI invocation). Due to time constraints this could not be developed further, however with SBW and the MMS-like system devised by Saffrey [154] it is now possible to envision a process chain, as illustrated in Figure 7.2 wherein application-level integration tools use their own solvers, or the models solver, to run/simulate the model and the rule optimisation engine operates under the SOA by querying for data. Such a coupling of technologies would mark a state-of-the-art in model integration for systems biology and represents a real solution to a very complex problem.



**Fig. 7.2** Desired integration between SOAs and the software produced in this project. SOAs communicate via interfaces to models and are able to extract data based on a query (implemented as, for example, SQL). This provides the information needed to optimise rules, which are eventually stored with associated agents (“Jiva”).

The questions that arise are: (i) What does ABI offer that these other software (linear) integrations do not already provide by themselves? (ii) What does the coupling of technologies described in Figure 7.2 provide to modellers that has not already been addressed? Concisely:

- The possibility of model integration when there are few or no connecting variables.
  - Linear integration is limited to those models that can be connected by common variables, i.e. matching inputs to outputs and vice versa.
  - Note that this does not negate the value of software infrastructures such as SBW and SClpath: the schema illustrated in Figure 7.2 identifies single model interfaces for the optimisation algorithm.
- Understanding on the level at which the real system is actually operating, i.e. local interactivity and agent-level understanding.
  - Provides a more in-depth understanding on the system.
- Simultaneously relates output to agents.
  - Experimentation can then take place outside the context of the original models, e.g. rules for a certain agent can be modified and the effects observed. This



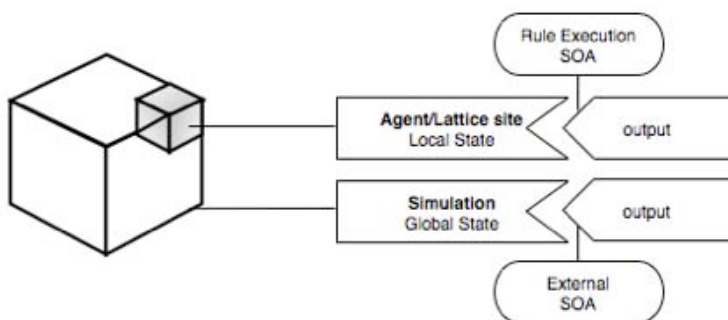
was demonstrated in the previous chapter when varying cell cycles lengths and synchronisation was directly simulated.

- Readily parallelisable.
- Possibility of 3D visualisation output.

Moreover note that ABI does not counteract the freedom of using linear integration. The simulation software, with its own interface to SBW or MRM/CMDI<sup>20</sup>, can run linearly integrated models or form part of a linear integration by exposing agent/global states or run them wholly within rules just the same as any model would do integrated into these SOAs. This is illustrated in Figure 7.3.

The most obvious advantage, and indeed the driving factor that led to ABI development, is the fact that many integrations cannot be achieved by the linear strategy even when paradigm, focus, scale and scope are compatible. There are two predominant views of this type of integration:

1. The mathematical view wherein integration takes place focusing on the mathematical properties of the models.
2. The software oriented view, which takes into account point 1 and additionally provides the infrastructure of communication of software components.



**Fig. 7.3** Linear integration is still achievable in ABI in two ways. Firstly a rule can be embedded with the logic for linear integration. The input would be the local state and output would affect those states. A corollary to this approach would be a parallel linear integration that influences the behaviour of agents in the simulation. Secondly the global state of the simulation can be linked to linear logic to affect both global and defined local states. In both cases note that integrating SOA-like architectures, with MPI or MSI communication, can achieve a chain of simulation spaces. Hence a far larger scale simulation can be envisioned wherein each component represents, for example, a single system such as an organ.

As a validation of ABI it is worth testing whether the Gompertz function, diffusion, and various components from the models by Ribba and Anderson integrate linearly in either view.

<sup>20</sup> Model Run Manager that uses CMDL (a name has not yet been given to the architecture described by [154]).

The Anderson model already incorporates diffusion of TAF. To integrate diffusion of any other molecule, an extra term would need to be put into the system of equations. For it to have any meaning it would need to be related to at least one component of the model. Here is the first hurdle of an unstructured linear approach. In the absence of any obvious connections there is a need for (i) additional knowledge and (ii) translations between connections (or both). Hence the introduction of a KDA-like approach is already of great benefit since it provides a structured, shareable and meaningful way to represent the interactivity of the models. Integration between the Gompertz model and diffusion provides a more apparent connection, i.e. that of concentration of nutrient and cell number and more atomically, B and C. Again a translation component would be needed to model the connection between growth and nutrients and it is not immediately obvious how this could be implemented. Note that the modules in the Ribba model actually were integrated linearly since there was a straightforward connection between the state of the pathway and the decision for growth based on cell cycle (modelled as an equivalence vector of Booleans in the network) and tissue dynamics, which was in part made possible by the particular solver on a 2D lattice. This model is therefore relatively open to a linear strategy in that any one of the components in the chain (molecular, cellular and tissue models) are replaceable so long as the replacement model has the same inputs and output. However even in this case the other models above are not readily open for integration. Hence in the strict mathematical sense none of these models can actually be integrated in any obvious way.

In a software-oriented view the implementation of the models, as well as associated solvers, need to be communicable. In this case the diffusion and Gompertz models, being relatively straightforward to implement in virtually any programming language, should not pose a problem. Moreover they both can be translated into a communicative language such as CellML in the form of MathML components. Such is the case for the angiogenesis model too. These are straightforward because CellML, and indeed most communicative languages and architectures that enable linear integration, have been tested and validated against DE-based models (and rightly so, seeing as most of the models in the literature until recently have been DE-based). It is of note that none of these models have actually been translated into CellML, SBML or CMDL as far as the author is aware (except for diffusion, which forms part of many models).

On this analysis one can see that linear integration is more suited to model-growing situations. A use case would start at a model or set of related models which can be extended by virtue of existing parameters and variables. For example, one can identify key variables that limit the current set and that can be expanded to include other sets of models, and hence the linear approach aids the growth of the model set. One of the great advantages of this type of integration is that models can be kept generic, i.e. in their class form rather than instantiation form. Recall that models can exist in these two forms, e.g. the equations given above are all in class form whereas those used in ABI were in instantiation form. This means that the linear structure along with the calling sequence will work with any instantiation of any of the modules, whereas the ABI methodology does not have this benefit. Instead

rules are optimised for *model instantiations* and hence are only applicable to that instantiation unless a definitive relation between the rule parameterisation/logic and the original model can be found. In the case of diffusion this was straightforward but for more complex models one can expect to find correspondingly more complex relations. This will perhaps be one of the greater challenges ABI faces but a solution represents a great advance in the field of model integration and therefore future research in this particular area is definitely worthwhile.

KDA/ABI provides an abstraction of models such that the modeller can interact with the models regardless of their original basis and implementation. The closest work to KDA/ABI outside of systems biology is that of Vangheluwe (2002) [184] who directly tackles multi-paradigm model integration by providing an abstraction theory involving the iterative deprecation of components to simple ideals. In many ways this is similar to ABI in that a common ground is sought. The theory subscribes to a discretisation of the original model(s) by thresholding outputs into intervals and ultimately into sets of Boolean states (i.e. binary automata). However not all models can be decomposed this way and there is a huge loss of granularity in such a method. As an alternative both Vangheluwe and Zeigler (1993) [205] suggest a chained approach to multi-paradigm integration in which a defined network of decompositions from known formalisms is developed. However such a method also has the pitfalls of loss of granularity as well as being constricted to a network that is woefully short of translation functions (especially with the enormous number of formalisms currently in use that are constantly being redeveloped). ABI on the other hand does not suffer from these setbacks: though also using discretisation, understood to be a necessary step by the modelling community (especially in the case of multi-paradigm modelling in complex systems), the common ground is not found through chains of conversions and loss of granularity; the introduction of knowledge and connection to the system by KDA/ABI ensure maintenance of granularity and behaviour.

### *Impact on Systems Biology*

Systems biology has been a leader in terms of integrating principles, as seen with the developments of the Physiome Project, SBML, CMDL, CellML, NCI initiatives in ontology-building and informational resources and consortia of experts, most notably CancerGrid<sup>21</sup> and CViT in cancer systems biology, yet a fully integrated *in silico* tumour is still lacking. To the best of the authors knowledge, no universal integrating theory or implementation yet exists in the field of systems biology.

It remains a fact that different methodologies suit different problems and that there are distinct experts for every subfield developing models independently. It therefore seems that both linear and ABI strategies together, along with the required resources, will hold the key to making *in silico* experimentation a reality. ABI by itself, purely on the merit of results described here, represents a great step forward: results have shown that models can be decomposed, though the implementation needs extending and optimising and further tests on more complex models is

---

<sup>21</sup> <http://www.cancergrid.org>

necessary. Results have additionally shown that, although model implementation and formalism are important, it is the behaviour that they represent that is the unifying principle, leading to determination of local interaction rules. This presents a shift in the way models are traditionally thought of in the systems biology community, i.e. where conventionally models have been tackled from a mathematical perspective *only* the work presented here stresses the importance of both the mathematical and behavioural perspectives. Taking a behavioural stance does not let the modeller wander outside the realm of what the model should be demonstrating, i.e. providing a connection and context to the real system. It also forces the integrator to think about mathematical or computational components within each model in the context of the whole, and therefore whether some modules even need to be explicated to describe system behaviour ( [198], Chapter 1). The work presented here reinforces ongoing efforts to move towards a true systems-oriented view of biology.

It is of course not claimed here that KDA/ABI is the final solution to the problem. It is claimed, however, that it is an excellent starting point from which modellers can begin to think about integrative measures without impeding freedom of formalism use, as well as address scale and focal issues.

### *Universality of KDA/ABI*

Model integration has been a universal problem and all fields have adapted their own solutions. Note that KDA/ABI, being independent of implementation and using generic components, are impartial to the system that they are simulating. This was a deliberate step in the development of the theories so that their application was not restricted.

The generic nature of KDA ensures that any system that is composed of components and behaviours can be described by it and its properties. Take the classic ant colony example (abstracted from [100]):

- Components
  - Ants
    - Types
- Surrounding objects
  - Predators
  - Food
  - Stimulants/Inhibitions
    - Molecular
- System Concepts
  - Reinforced learning
  - Ant movement
  - Food collection and waste disposal

It is a well-known fact that ants coordinate behaviour of the colony by simple local interaction and hence this is completely amenable to KDA/ABI decomposition.

One could also take the example of traffic, which has received much attention in light of its non-linear behaviour. Here also, components and behaviours can be abstracted from models and implemented in ABI. The very fact that most traffic models are now implemented in ABS anyway reinforces this proposition, in which case further investigation is needed to test whether rule merging can yield plausible integrations.

One limitation however exists where the system cannot be simulated in a closed 3D space. Even if ABI enables linear integrations as explained earlier it fundamentally relies on the fact that components interact locally. There are therefore a subset of complex systems for which KDA/ABI are not immediately suited, most notably financial markets and intelligent social networks, though of course the software system is adaptable. The current system is assumed to be representative of interactions in a spatial context, however interactions between agents and lattice sites could easily be interpreted in a more abstract context.

### *Other Limitations*

Some limitations have already been discussed with respect to the GA and KDA/ABI methods. Judging from the literature review there is a contingent of models for which KDA/ABI will not be able to reconcile in an integration exercise.

Firstly this is the case for stochastic models. Stochasticity is a paradigm as outlined previously however it need not always impede KDA/ABI as illustrated by the solution of the Anderson model. Decomposition of rules becomes difficult for stochastic models for two main reasons:

1. A stochastic approach yields non-replicable results, which impedes the fitness functions ability to consistently find good/bad performers, which of course has a direct impact on selection pressure.
  - a. An alternative method could be to include a fuzzy classifier that assesses fitness qualitatively.
2. Equivalently, if component rules are made stochastic the results could potentially be non-replicable.

As suggested above the best solution may be to find a different optimisation method or fitness function method that can account for diversity in results and assess fitness correctly.

Secondly the combination of centrality and stochasticity could similarly cause problems. A good example is that of Bayesian networks: the simulation must perform such that all state combinations of the network (if that is even computable) are concordant with simulation states.

## 7.2 Conclusion

In this project the major advancements can be summarised as follows:

- The determination of the nature of models based on a review of cancer modelling literature.
  - This has led to the determination of characteristics of models that highlight the challenges inherent to model integration. These characteristics are model paradigm, focus, scale and scope.
- A review of model integration literature has been given to assess the extent to which these challenges have been addressed, i.e. the state of the art. This has been extended outside of the systems biology field since model integration is a universal problem in all fields that make use of models and simulations.
- A set-theoretic approach has been developed to tackle the challenge of model semantics. This takes a knowledge-oriented approach to model description.
- An integration strategy has been developed that re-implements models within an agent-based system using local interaction as the basis on which behaviour occurs.
- A well-developed software infrastructure has been developed to enable the above.

Model integration is a key challenge in systems biology. With such a methodology in hand one can begin to build larger and more representative models of biological systems. The solution presented in this book tackles all of the identified challenges and, with the suggested improvements, provides a powerful approach to developing an *in silico* tumour that can yield behaviour that has real clinical relevance. Moreover the fruitful start that has been made with software development will give these ideas a head start in the field.

Despite millions of pounds of research cancer continues to be the UK's biggest killer. The introduction of informatics and modelling has brought about a shift in the way scientists understand the disease and how to combat it. Thanks to the rapid growth in affordable computing power the goal of *in silico* experimentation is becoming ever closer. It is hoped the technology developed in this project will be of value to that cause.

# References

1. Adam, J.A., Bellomo, N.: *A Survey of Models for Tumor-Immune System Dynamics*. Birkhäuser, Boston (1996)
2. Adiwijaya, B.S., Barton, P.I., Tidor, B.: Biological network design strategies: discovery through dynamic optimization. *Molecular BioSystems* 2, 650–659 (2006)
3. Alarcon, T., Byrne, H.M., Maini, P.K.: Towards whole-organ modelling of tumour growth. *Progress in Biophysics and Molecular Biology* 85, 451–472 (2004)
4. Alber, M., Kiskowski, M., Glazier, J., Jiang, Y.: On cellular automaton approaches to modeling biological cells. *Journal of Applied Mathematics, IMA (Mathematical systems theory in biology, communication and finance)* 134, 12 (2002)
5. Alberghina, L., Colangelo, A.M.: The modular systems biology approach to investigate the control of apoptosis in alzheimer's disease neurodegeneration. *BMC Neuroscience* 7(Suppl.1), S2 (2006)
6. Anderson, A.R.: A hybrid discrete-continuum technique for individual-based migration models. In: Alt, W., Chaplain, M.A., Griebel, M., Lenz, K. (eds.) *Polymer and Cell Dynamics - Multiscale Modeling and Numerical Simulations*, Birkhauser Verlag, Basel (2003)
7. Anderson, A.R.A., Chaplain, M.A.J.: Continuous and discrete mathematical models of tumor-induced angiogenesis. *Bulletin of Mathematical Biology* 60, 857–900 (1998)
8. Anderson, R.A., Quaranta, V.: Integrative mathematical biology. *Nature Reviews Cancer* 8, 227–234 (2008)
9. Antipas, V.P., Stamatakos, G.S., Uzunoglu, N.K., Dionysiou, D.D., Dale, R.G.: A spatio-temporal simulation model of the response of solid tumours to radiotherapy in vivo: parametric validation concerning oxygen enhancement ratio and cell cycle duration. *Physics in Medicine and Biology* 49, 1485–1504 (2004)
10. Arakelyan, L., Vainstein, V., Agur, Z.: A computer algorithm describing the process of vessel formation and maturation, and its use for predicting the effects of anti-angiogenic and anti-maturation therapy on vascular tumor growth. *Angiogenesis* 5, 203–214 (2002)
11. Araujo, R.P., McElwain, D.L.S.: A history of the study of solid tumour growth: The contribution of mathematical modelling. *Bulletin of Mathematical Biology* 66, 1039–1091 (2004)
12. Arciero, J.C., Jackson, T.L., Kirschner, D.E.: A mathematical model of tumour-immune evasion and sirna treatment. *Discrete and Continuous Dynamical Systems Series B* 4(1), 39–58 (2004)

13. Awale, S., Lu, J., Kalauni, S.K., Kurashima, Y., Tezuka, Y., Kadota, S., Esumi, H.: Identification of arctigenin as an antitumor agent having the ability to eliminate the tolerance of cancer cells to nutrient starvation. *Cancer Research* 66, 1751–1757 (2006)
14. Bagrodia, R., Meyer, R., Takai, M., Chen, Y., Zeng, X., Martin, J., Song, H.Y.: Parsec: A parallel simulation environment for complex systems. *IEEE Computer* 31(10), 77–85 (1998)
15. Baish, J., Jain, R.K.: Fractals and cancer. *Cancer Research* 60, 3683–3688 (2000)
16. Baldi, P., Long, A.D.: A bayesian framework for the analysis of microarray expression data: regularised t-test and statistical inferences of gene changes. *Bioinformatics* 17(6), 509–519 (2001)
17. Bar-Yam, Y.: *Dynamics of Complex Systems: Studies in Non-Linearity*. Perseus Press, Oxford (1997)
18. Bar-Yam, Y.: A mathematical theory of strong emergence using multiscale variety. *Complexity* 9(6), 15–24 (2004)
19. Barabasi, A.L., Oltvai, Z.L.: Network biology: Understanding the cell's functional organization. *Nature* 5, 101–113 (2004)
20. Baral, C., Chancellor, K., Tran, N., Tran, N.L., Joy, A., Berens, M.: A knowledge based approach for representing and reasoning about signalling networks. *Bioinformatics* 20(1), i15–i22 (2004)
21. Barazzol, L., Burnet, N.G., Jena, R., Jones, B., Jeffries, S.J., Kirkby, N.F.: A mathematical model of brain tumour response to radiotherapy and chemotherapy considering radiobiological aspects. *Journal of Theoretical Biology* 262(3), 553–565 (2010)
22. Barton, R.: Simulation metamodels. In: Medeiros, D.J., Watson, E.F., Carson, J.S., Manivannan, M.S. (eds.) *Proceedings of the 1998 Winter Simulations Conference*, pp. 167–174 (1998)
23. Bates, R.C., Edwards, N.S., Yates, J.D.: Spheroids and cell survival. *Critical Reviews in Oncology/Hematology* 36, 44–61 (2000)
24. Bedau, M.A.: Weak emergence. In: Tomberlin, J. (ed.) *Philosophical Perspectives: Mind, Causation, and World*, vol. 2, pp. 375–399. Blackwell, Malden (1997)
25. Bernier, F., Poussart, D., Laurendeau, D., Simoneau, M.: Interaction-centric modelling for interactive virtual worlds: The apia approach. In: *16th International Conference on Pattern Recognition (ICPR 2002)*, vol. 3, pp. 31007–31011 (2002)
26. Bhattacharya, S.S., Brooks, C., Cheong, E., Davis, J., Goel, M., Kienhuis, B., Edward, A.L., Liu, J., Liu, X., Muliadi, L., Neuendorffer, S., Reekie, J., Smyth, N., Tsay, J., Vogel, B., Williams, W., Xiong, Y., Zhao, Y., Zheng, H.: *Ptolemy ii: Heterogeneous concurrent modeling and design in java*, November 2006 (2005)
27. Boogerd, F.C., Bruggerman, F.J., Richardson, R.C., Stephan, A., Westerhoff, H.V.: Emergence and its place in nature: A case study of biochemical networks. *Synthese* 145, 131–164 (2005)
28. Borkenstein, K., Levegrun, S., Paschke, P.: Modeling and computer simulations of tumor growth and tumor response to radiotherapy. *Radiation Research* 162, 71–83 (2004)
29. Bossomaier, T., Cranny, T., Schneider, D.: A new paradigm for evolving cellular automata rules. In: *Evolutionary Computation*, Washington DC, USA, vol. 1, p. 176. IEEE, Los Alamitos (1999)
30. Bourguine, P., Johnson, J.: *Living roadmap for complex systems science* (September 28, 2006)
31. Brazma, A., Hingamp, P., Quackenbush, J., Sherlock, G., Spellman, P., Stoeckert, C., Aach, J., Ansorge, W., Ball, C.A., Causton, H.C., Gaasterland, T., Glenisson, P., Holstege, F.C.P., Kim, I.F., Markowitz, V., Matese, J.C., Parkinson, H., Robinson, A.,



- Sarkans, U., Schulze-Kremer, S., Stewart, J., Taylor, R., Vilo, J., Vingron, M.: Minimum information about a microarray experiment (miame) toward standards for microarray data. *Nature Genetics* 29(4), 365–371 (2001) 1061-4036 10.1038/ng1201-365 10.1038/ng1201-365
32. Bruggerman, F.J., Westerhoff, H.V.: The nature of systems biology. *Trends in Microbiology* 15(1), 45–50 (2006)
  33. Byrne, H.M.: Using mathematics to study solid tumour growth. In: *Proceedings of the 9th General Meeting of European Women in Mathematics*, pp. 81–107 (1999)
  34. Catto, J.W., Linkens, D.A., Abbod, M.F., Chen, M., Burton, J.L., Feeley, K.M., Hamdy, F.C.: Artificial intelligence in predicting bladder cancer outcome: a comparison of neuro-fuzzy modeling and artificial neural networks. *Clin Cancer Res.* 9(11), 4172–4177 (2003) 1078-0432 Journal Article
  35. Chalmers, D.: Consciousness and its place in nature. In: Stich, S., Warfield, F. (eds.) *Guide to the Philosophy of Mind*. Blackwell, Oxford (2002)
  36. Chaplain, M.A., McDougall, S.R., Anderson, A.R.: Mathematical modeling of tumor-induced angiogenesis. *The Annual Review of Biomedical Engineering* 8, 233–257 (2006)
  37. Chari, K., Sen, T.K.: An implementation of a graph-based modeling system for structured modeling (gbms/sm). *Decision Support Systems* 22, 103–120 (1998)
  38. Charusanti, P., Hu, X., Chen, L., Neuhauser, D., DiStefano III, J.J.: A mathematical model of bcr-abl autophosphorylation, signaling through the crkl pathway, and gleevec dynamics in chronic myeloid leukemia. *Discrete and Continuous Dynamical Systems Series B* 4(1), 99–114 (2004)
  39. Chen, C.C., Nagl, S., Clack, C.D.: Computational techniques for modeling and simulating biological systems. *ACM Computing Surveys* 34(3), 1–36 (2006)
  40. Chen, C.C., Nagl, S., Clack, C.D.: Complexity and emergence in engineering systems. *Studies in Computational Intelligence Series*, pp. 99–128 (2009)
  41. Chen, C.C., Nagl, S., Clack, C.D.: A formalism for multi-level emergent behaviours in designed component-based systems and agent-based simulations. In: *From System Complexity to Emergent Properties*, *Understanding Complex Systems Series*, pp. 101–114 (2009)
  42. Chen, C.C., Nagl, S., Clack, C.D.: Identifying multi-level emergent behaviors in agent-directed simulations using complex event type specifications. *Simulation* 86(1), 41–51 (2010)
  43. Chen, S., Ganguli, S., Hunt, C.A.: An agent-based computational approach for representing aspects of in vitro multi-cellular tumor spheroid growth. In: *Proceedings of the 26th Annual International Conference of the IEEE EMBS*, San Francisco (2004)
  44. Chignola, R., Schenetti, A., Andrighetto, G., Chiesa, E., Foroni, R., Sartoris, S., Tridente, G., Liberati, D.: Forecasting the growth of multicell tumour spheroids: implications for the dynamic growth of solid tumours. *Cell Proliferation* 33, 219–229 (2000)
  45. Cho, K.H., Wolkenhauer, O.: Analysis and modelling of signal transduction pathways in systems biology. *Biochemical Society Transactions* 31(6), 1503–1509 (2003)
  46. Christopher, R., Dhiman, A., Fox, J., Gendelman, R., Haberichter, T., Kagle, D., Spizz, G., Khalil, I.G., Hill, C.: Data-driven computer simulation of human cancer cell. *Annals in the New York Academy of Sciences* 1020, 132–153 (2004)
  47. Coffey, D.S.: Self-organisation, complexity and chaos: The new biology for medicine. *Nature Medicine* 4(8), 882–885 (1998)
  48. Colonna, A., Di Stefano, V., Lombardo, S., Papini, L., Rabino, G.A.: Learning cellular automata: Modelling urban modelling (1998)

49. Cornish-Bowden, A.: Making systems biology work in the 21st century. *Genome Biology* 6, 317–318 (2005)
50. Cornish-Bowden, A.: Putting the systems back into systems biology. *Perspectives in Biology and Medicine* 49, 475–489 (2006)
51. Coveney, P.V., Fowler, P.W.: Modelling biological complexity: a physical scientists perspective. *Journal of the Royal Society Interface* 2(4), 267–280 (2005)
52. Csete, M., Doyle, J.C.: Reverse engineering of biological complexity. *Science* 295, 1664–1669 (2002)
53. Dasu, A., Toma-Dasu, I., Karlsson, M.: Theoretical simulation of tumour oxygenation and results from acute and chronic hypoxia. *Physics in Medicine and Biology* 48, 2829–2842 (2003)
54. de Castro, L.N., Timmis, J.: Artificial immune systems: A novel paradigm to pattern recognition. In: Corchado, J.M., Alonso, L., Fyfe, C. (eds.) *Artificial Neural Networks in Pattern Recognition*, Paisley, UK, pp. 67–84. SOCO (2002)
55. De Pillis, L.G., Radunskaya, A.: The dynamics of an optimally controlled tumor model: A case study. *Mathematical and Computer Modelling* 37, 1221–1244 (2003)
56. De Pillis, L.G., Radunskaya, A., Wiseman, C.L.: A validated mathematical model of cell-mediated immune response to tumor growth. *Cancer Research* 65(17), 7950–7958 (2005)
57. Deisboeck, T.S., Berens, M.E., Kansal, A.R., Torquato, S., Stemmer-Rachamimov, A.O., Chiocca, E.A.: Pattern of self-organisation in tumour systems: complex growth dynamics in a novel brain tumour spheroid. *Cell Proliferation* 34, 115–134 (2001)
58. Delden, H.V., Luja, P., Engelen, G.: Integration of multi-scale dynamic spatial models of socio-economic and physical processes for river basin management. *Environmental Modelling and Software* 22(2), 1–16 (2005)
59. Desoize, B.: Contribution of three-dimensional culture to cancer research. *Critical Reviews in Oncology/Hematology* 36, 59–60 (2000)
60. Dionysiou, D.D., Stamatakis, G.S.: Applying a 4d multiscale in vivo tumor growth model to the exploration of radiotherapy scheduling: The effects of weekend treatment gaps and p53 gene status on the response of fast growing solid tumors. *Cancer Informatics* 2, 113–121 (2006)
61. Dokoumetzidis, A., Karalis, V., Iliadis, A., Macheras, P.: The heterogeneous course of drug transit through the body. *Trends in Pharmacological Sciences* 25(3), 140–146 (2004)
62. Dolk, D.: Integrated model management in the data warehouse era. *European Journal of Operations Research* 122, 199–218 (2000)
63. Dolk, D., Kottemann, J.: Model integration and a theory of models. *Decision Support Systems* 9(1), 51–63 (1993)
64. Dolk, D.R.: An introduction to model integration and integrated modeling environments. *Decision Support Systems* 10, 249–254 (1993)
65. Dormann, S., Deutsch, A.: Modeling of self-organized avascular tumor growth with a hybrid cellular automaton. *Silico Biology* 2, 0035 (2002)
66. Drew, N., Hendley, R.: Visualising complex interaction systems. In: *Companion to ACM Conf. Human Factors in Computing Systems*, pp. 204–205 (1995)
67. Dubitzky, W.: Understanding the computational methodologies of systems biology. *Briefings in Bioinformatics* 7(4), 315–317 (2006)
68. Duchting, W., Vogelsanger, Y.: Three-dimensional pattern generation applied to spheroidal tumor growth in a nutrient medium. *International Journal of Biomedical Computing* 12, 377–392 (1981)

69. Finkelstein, A., Hetherington, J., Li, L., Margoninski, O., Saffrey, P., Seymour, R., Warner, A.: Computational challenges of systems biology. *IEEE Computer* 37(4), 26–33 (2004)
70. Fisher, M.J., Malcolm, G., Paton, R.C.: Spatio-logical processes in intracellular signalling. *BioSystems* 55, 83–92 (2000)
71. Frantz, F.: A taxonomy of model abstraction techniques. In: Alexopoulos, C., Kang, K., Lilegdon, W.R., Goldsman, D. (eds.) *Proceedings of the 1995 Winter Simulation Conference*, pp. 1413–1420 (1995)
72. Fricks, R., Hirel, C., Wells, S., Trivedi, K.S.: The development of an integrated modeling environment. In: *Proceedings of the 1st World Conference on Systems Simulation, WCCS (1997)*
73. Friedman, A.: A hierarchy of cancer models and their mathematical challenges. *Continuous and Discrete Dynamical Systems - Series B* 4(1), 147–159 (2004)
74. Friedman, N., Linial, M., Nachman, I., Pe'er, D.: Using bayesian networks to analyze expression data. *Journal of Computational Biology* 7(3-4), 601–620 (2000)
75. Frigyesi, A., Gisselsson, D., Mitelman, F., Hoglund, M.: Power law distribution of chromosome aberrations in cancer. *Cancer Research* 63, 7094–7097 (2003)
76. Gatenby, R.A., Maini, P.K.: Mathematical oncology: Cancer summed up. *Nature* 421, 321–324 (2003)
77. Gazit, Y., Baish, J.W., Safabakhsh, N., Leunig, M., Baxter, L.T., Jain, R.K.: Fractal characteristics of tumor vascular architecture during tumor growth and regression. *Microcirculation* 4(4), 395–402 (1997)
78. Geoffrion, A.M.: An introduction to structured modeling. *Management Science* 33, 547–588 (1987)
79. Geoffrion, A.M.: The formal aspects of structured modeling. *Operations Research* 37(1), 30–51 (1989)
80. Geoffrion, A.M.: Integrated modeling systems. *Computer Science in Economics and Management* 2(3-15) (1989)
81. Geoffrion, A.M.: Structured modeling: Survey and future research directions. *ORSA CSTS Newsletter* 15, 1 (1999)
82. Giavitto, J.L., Godin, C.: Computational models for integrative and developmental biology (2002)
83. Giraldi, G.A., da Costa, L.C., Xavier, A.V., Rodrigues, P.S.: Methods for analytical understanding of agent-based modeling of complex systems. *ArXiv Computer Science e-prints* (2005)
84. Harris, A.L.: Antiangiogenesis for cancer therapy. *Lancet* 349(2), 13–15 (1997)
85. Hetherington, J., Warner, A., Seymour, R.: Simplification and its consequences in biological modelling: conclusions from a study of calcium oscillations in hepatocytes. *Journal of the Royal Society Interface* 3, 319–331 (2006)
86. Hill, A.V.: The diffusion of oxygen and lactic acid through tissues. *Proceedings of the Royal Society of London B, Biological Sciences* 104, 39–96 (1928)
87. Hnat, B., Chapman, S.C.: Visualisation of multiscale datasets in numerical models of complex systems. *International Journal of Computer Integrated Design and Construction* 2(4), 235–240 (2000)
88. Holash, J., Wiegand, S.J., Yancopoulos, G.D.: New model of tumor angiogenesis: Dynamic balance between vessel regression and growth mediated by angiopoietins and vegf. *Oncogene* 18(38), 5356–5362 (1999)
89. Holden, C.: Alliance launched to model e.coli. *Science* 297, 1459–1460 (2002)
90. Holland, J.H.: *Adaptation in Natural and Artificial Systems*. University of Michigan Press, Ann Arbor (1975)

91. Hucka, M., Finney, A., Sauro, H.M., Bolouri, H., Doyle, J.C., Kitano, H., Arkin, A.P., Bornstein, B.J., Bray, D., Cornish-Bowden, A., Cuellar, A.A., Dronov, S., Gilles, E.D., Ginkel, M., Gor, V., Goryanin, I.I., Hedley, W.J., Hodgman, T.C., Hofmeyr, J.H., Hunter, P.J., Juty, N.S., Kasberger, J.L., Kremling, A., Kummer, U., Le Novere, N., Loew, L.M., Lucio, D., Mendes, P., Minch, E., Mjolsness, E.D., Nakayama, Y., Nelson, M.R., Nielsen, P.F., Sakurada, T., Schaff, J.C., Shapiro, B.E., Shimizu, T.S., Spence, H.D., Stelling, J., Takahashi, K., Tomita, M., Wagner, J., Wang, J.: The systems biology markup language (sbml): A medium for representation and exchange of biochemical network models. *Bioinformatics* 19(4), 524–531 (2003)
92. Hunter, P.J.: The uips physiome project: a framework for computational physiology. *Progress in Biophysics and Molecular Biology* 85, 551–569 (2004)
93. Hunter, P.J., Borg, T.K.: Integration from proteins to organs: the physiome project. *Nature Reviews Molecular Cell Biology* 4, 237–243 (2003)
94. Hyman, J.M.: Patch dynamics for multiscale problems. *Computing in Science and Engineering* 7(3), 47–53 (2005)
95. Ideker, T., Lauffenburger, D.: Building with a scaffold: emerging strategies for high- to low-level cellular modeling. *Trends in Biotechnology* 21(6), 255–262 (2003)
96. Iskandarani, M., Haidvogel, D., Levin, J.C., Curchitser, E., Edwards, C.A.: Multiscale geophysical modeling using the spectral element method. *Computing in Science and Engineering* 4(5), 42–48 (2002)
97. Jackson, T.L., Byrne, H.M.: A mathematical model to study the effects of drug resistance and vasculature on the response of solid tumors to chemotherapy. *Mathematical Biosciences* 164(1), 17–38 (2000)
98. Jensen, F.V.: *Bayesian Networks and Decision Graphs*. Springer, Germany (2001)
99. Jiang, Y., Pjesivac-Grbovic, J., Cantrell, C., Freyer, J.: A multiscale model for avascular tumor growth. *Biophysical Journal* 89, 3884–3894 (2005)
100. Johnson, S.: *Emergence*. The Penguin Group, USA (2001)
101. Kanoh, H., Wu, Y.: Evolutionary design of rule changing cellular automata. In: Palade, V., Howlett, R.J., Jain, L. (eds.) *KES 2003*. LNCS, vol. 2773, pp. 258–264. Springer, Heidelberg (2003)
102. Kevrekidis, I.G., Gear, C.W., Hyman, J.M., Kevrekidis, P.G., Runborg, O., Theodoropoulos, C.: Equation-free course grained multiscale computation: Enabling microscopic simulators to perform systems-level tasks. *Communications in the Mathematical Sciences* 14, 715–762 (2003)
103. Khu, S.T., Savic, D.A., Liu, Y., Madsen, H.: A fast evolutionary-based meta-modelling approach for the calibration of a rainfall-runoff model. In: *IEMSS conference*, Osnabruck (2004)
104. Kitano, H.: Computational systems biology. *Nature* 420, 206–210 (2002)
105. Kitano, H.: Biological robustness. *Nature Reviews Genetics* 5, 826–837 (2004)
106. Kohn, S., Nagy, J.A., Dvorak, H.F., Dvorak, A.M.: Pathways of macromolecular tracer transport across venules and small veins. structural basis for the hyperpermeability of tumor blood vessels. *Laboratory Investigation* 67, 596–607 (1992)
107. Krishnan, R., Chari, K.: Model management: Survey, future directions and a bibliography. *Interactive Transactions of ORMS* 3(1) (2000)
108. Kunz-Schughart, L.A., Kreutz, M., Knuechel, R.: Multicellular spheroids: a three-dimensional in vitro culture system to study tumour biology. *International Journal of Experimental Pathology* 79, 1–23 (1998)
109. Kyriazis, D., Menychtas, A., Dionysiou, D., Stamatakos, G., Varvarigou, T.: Clinical trial simulation in grid environments. In: *8th International IEEE Conference of Bioinformatics and Bioengineering*, vol. (1), pp. 1–6 (2008)

110. Lakmeche, A., Arino, O.: Nonlinear mathematical model of pulsed-therapy of heterogeneous tumors. *Nonlinear Analysis: Real World Applications* 2, 455–465 (2001)
111. Ledzewicz, U., Liu, Y., Schättler, H.: The effect of pharmacokinetics on optimal protocols for a mathematical model of tumor anti-angiogenic therapy. In: ACC 2009: Proceedings of the 2009 conference on American Control Conference, Piscataway, NJ, USA, pp. 1060–1065. IEEE Press, Los Alamitos (2009)
112. Lee, D.-Y., Yun, C., Cho, A., Hou, B.K., Park, S., Lee, S.Y.: Webcell: a web-based environment for kinetic modeling and dynamic simulation of cellular networks. *Bioinformatics* 22(9), 1150–1151 (2006)
113. Levine, H.A., Sleeman, B.D., Nilsen-Hamilton, M.: Mathematical modeling of the onset of capillary formation initiating angiogenesis. *Journal of Mathematical Biology* 42, 195–238 (2001)
114. Lloyd, C.M., Halstead, M.D.B., Nielsen, P.F.: Cellml: its future, present and past. *Progress in Biophysics and Molecular Biology* 85, 433–450 (2004)
115. Loew, L.M., Schaff, J.C.: The virtual cell: a software environment for computational cell biology. *Trends in Biotechnology* 19(10), 401–406 (2001)
116. Lok, L.: The need for speed in stochastic simulation. *Nature Biotechnology* 22(8), 964–965 (2004)
117. Ma, J., Tian, Q., Zhou, D.: An object-oriented approach to structured modeling. In: IRMA 1998 Proceedings 1998 Information Resources Management Association International Conference, pp. 406–413. Idea Group Publishing, USA (1998)
118. Makowski, M.: A structured modeling technology. *European Journal of Operations Research* 166(3), 615–648 (2005)
119. Mallet, D.G., De Pillis, L.G.: A cellular automata model of tumor-immune system interactions. *Journal of Theoretical Biology* 239, 334–350 (2006)
120. Mantzaris, N.V., Webb, S., Othmer, H.G.: Mathematical modeling of tumor-induced angiogenesis. *Journal of Mathematical Biology* 49, 111–187 (2004)
121. Margoninski, O., Saffrey, P., Hetherington, J., Finkelstein, A., Warner, A.: A specification language and a framework for the execution of composite models in systems biology. In: Priami, C., Ingólfssdóttir, A., Mishra, B., Riis Nielson, H. (eds.) *Transactions on Computational Systems Biology VII*. LNCS (LNBI), vol. 4230, pp. 163–184. Springer, Heidelberg (2006)
122. Marinov, M., Zheliazkova, I.: An interactive tool based on priority semantic networks. *Knowledge-Based Systems* 18, 71–77 (2005)
123. Markus, M., Bohm, D., Schmick, M.: Simulation of vessel morphogenesis using cellular automata. *Mathematical Biosciences* 156, 191–206 (1999)
124. Marusic, M., Bajzer, Z., Vuk-Pavlovic, S., Freyer, J.P.: Tumor growth in vivo and as multicellular spheroids compared by mathematical models. *Bulletin of Mathematical Biology* 56(4), 617–631 (1994)
125. McDougall, S.R., Anderson, A.R., Chaplain, M.A., Sherratt, J.A.: Mathematical modelling of flow through vascular networks: implications for tumour-induced angiogenesis and chemotherapy strategies. *Bulletin of Mathematical Biology* 64(4), 673–702 (2002)
126. Miller, J.H., Zheng, F.: Large scale simulations of eukaryotic cell signaling processes. *Parallel Computing* 30(9–10), 1137–1149 (2004)
127. Mitchell, M.: *An Introduction to Genetic Algorithms*, 2nd edn. MIT Press, Boston (1996)
128. Mitchell, M., Crutchfield, J.P., Das, R.: Evolving cellular automata with genetic algorithms: A review of recent work. In: *EvCA 1996*. Russian Academy of Sciences (1996)
129. Mizraji, E.: The emergence of dynamical complexity: An exploration using elementary cellular automata. *Complexity* 9(6), 33–42 (2004)

130. Moliola, J.L.: An overview of bifurcation, chaos and nonlinear dynamics in control systems. *Journal of the Franklin Institute* 33(6), 819–858 (1994)
131. Mooney, D., Swift, R.: *A Course in Mathematical Modeling*. The Mathematical Association of America, USA (1999)
132. Nagl, S., Williams, M., Williamson, J.: Objective bayesian nets for systems modelling and prognosis in breast cancer. In: Holms, D., Jain, L.C. (eds.) *Innovations in Bayesian Networks: Theory and Applications*. Springer, Heidelberg (2007)
133. Naistat, D.M., Leblanc, R.: Proteomics. *Journal of Environmental Pathology, Toxicology and Oncology* 23(3), 161–178 (2004)
134. Natarajan, L., Berry, C.C., Gasche, C.: Estimation of spontaneous mutation rates. *Biometrics* 59, 555–561 (2003)
135. Nelson, P.: *Biological Physics: Energy, Information, Life*. WH Freeman, New York (2003)
136. Noble, D.: The heart is already working. *Biochemical Society Transactions* 33, 539–542 (2005)
137. O'Connor, T., Wong, H.Y.: Emergent Properties. In: *The Stanford Encyclopedia of Philosophy*, Winter 2002 Edition (2002)
138. Otter, M., Elmqvist, H.: Modelica: Language, libraries, tools, workshop and eu-project realsim (2001)
139. Papetti, M., Herman, I.M.: Mechanisms of normal and tumor-derived angiogenesis. *American Journal of Physiology - Cell Physiology* 282, C947–C970 (2002)
140. Patel, A.A., Gawlinski, E.T., Lemieux, S.K., Gatenby, R.A.: A cellular automaton model of early tumour growth and invasion: The effects of native tissue vascularity and increased anaerobic tumour metabolism. *Journal of Theoretical Biology* 213, 315–331 (2001)
141. Patel, M., Nagl, S.: Microcore: mapping genome expression to cell pathways and networks. *Comparative and Functional Genomics* 5, 75–78 (2004)
142. Pathmanathan, P., Gavaghan, D., Whitley, J.P., Chapman, S.J., Brady, J.M.: Predicting tumor location by modeling the deformation of the breast. *IEEE Transactions on Biomedical Engineering* 55(10), 2471–2480 (2008)
143. Pawaletz, N., Knierim, M.: Tumor-related angiogenesis. *Critical Reviews in Oncology and Hematology* 9, 197–242 (1989)
144. Pienta, K.: Modeling cancer as a complex adaptive system: Genetic instability and evolution. In: Deisboeck, T.S., Kresh, J.Y. (eds.) *Complex Systems Science in Biomedicine*. Springer, Heidelberg (2006)
145. Pinho, S.T.R., Freedman, H.I., Nani, F.: A chemotherapy model for the treatment of cancer with metastasis. *Mathematical and Computer Modelling* 36, 773–803 (2002)
146. Plank, M.J., Sleeman, B.D.: A reinforced random walk model of tumour angiogenesis and anti-angiogenic strategies. *Mathematical Medicine and Biology* 20(2), 135–181 (2003)
147. Please, C.P., Pettet, G.J., McElwain, D.L.S.: A new approach to modelling the formation of necrotic regions in tumours. *Applied Mathematics Letters* 11, 89–94 (1998)
148. Purwasih, W., Nakamori, Y.: Structured modeling technology: A modeling environment for integrated environmental assessment. In: *Complexity and Integrated Resources Management, Transactions of the 2nd Biennial Meeting of the International Environmental Modelling and Software Society, iEMSs 2004* (2004)
149. Qi, A.S., Zheng, X., Du, C.Y., An, B.S.: A cellular automaton model of cancerous growth. *Journal of Theoretical Biology* 161, 1–12 (1993)
150. dos Reis, A.N., Mombach, J.C.M., Walter, M.: The role of decreased cell adhesion in tumor morphology: A simulation study. In: *4th International Eurosim 2001 Congress*, Delft, Netherlands (2001)

151. Resnick, N., Gimbrone, M.A.: Hemodynamic forces are complex regulators of endothelial gene expression. *The FASEB Journal* 9, 874–882 (1995)
152. Ribba, B., Colin, T., Schnell, S.: A multiscale mathematical model of cancer, and it's use in analysing irradiation therapies. *Theoretical Biology and Medical Modelling* 3(7), 1–19 (2006)
153. Roxburgh, S.H., Davies, I.D.: Coins: an integrative modelling shell for carbon accounting and general ecological analysis. *Environmental Modelling and Software* 21, 359–374 (2006)
154. Saffrey, P., Margoninski, O., Hetherington, J., Varela-Rey, M., Yamaji, S., Finkelstein, A., Bogle, D., Warner, A.: End-to-end information management for systems biology. *Transactions on Computational Systems Biology* (8) (2007)
155. dos Santos, I.R., Porta Nova, A.M.O.: The main issues in nonlinear simulation meta-model estimation. In: Farrington, P.A., Nembhard, H.B., Sturrok, D.T., Evan, G.W. (eds.) *Proceedings in the 1999 Winter Simulation Conference*, pp. 502–509 (1999)
156. Sauro, H.M., Hucka, M., Finney, A., Wellock, C., Bolouri, H., Doyle, J., Kitano, H.: Next generation simulation tools: the systems biology workbench and biospice integration. *OMICS* 7(4), 355–372 (2003)
157. Scalerandi, M., Capogrosso Sansone, B., Benati, C., Condat, C.A.: Competition effects in the dynamics of tumor cords. *Physical Review E* 65, 51918 (2002)
158. Scalerandi, M., Peggion, F., Capogrosso Sansone, B., Benati, C.: Avascular and vascular phases in tumor cords growth. *Mathematical and Computer Modelling* 37, 1191–1200 (2003)
159. Scalerandi, M., Sansone, B.C.: Inhibition of vascularisation in tumor growth. *Physical Review Letters* 89(21), 218101 (2002)
160. Schoeberl, B., Eichler-Jonsson, C., Gilles, E.D., Muller, G.: Computational modelling of the dynamics of the map kinase cascade activated by surface and internalised egf receptors. *Nature Biotechnology* 20, 370–375 (2002)
161. Sherrat, J.A., Chaplain, M.A.J.: A new mathematical model for avascular tumour growth. *Journal of Mathematical Biology* 43, 291–312 (2001)
162. Shmulevich, I., Dougherty, E.R., Kim, S., Zhang, W.: Probabilistic boolean networks: a rule-based uncertainty model for gene regulatory networks. *Bioinformatics* 18(2), 261–274 (2002) 1367-4803 Journal Article
163. Siromoney, A., Raghuram, L., Korah, I., Prasad, G.N.S.: Inductive logic programming for knowledge discovery from mri data. *IEEE Engineering in Medicine and Biology Magazine* 19(2), 72–77 (2000)
164. Slepchenko, B.M., Schaff, J.C., Macara, I., Loew, L.M.: Quantitative cell biology with the virtual cell. *Trends in Cell Biology* 13(11), 570–576 (2003)
165. Smallwood, R.H., Holcombe, M.: The epitheliome project: multiscale agent-based modeling of epithelial cells (2006)
166. Snoep, J.L., Bruggeman, F., Olivier, B.G., Westerhoff, H.V.: Towards building the silicon cell: A modular approach. *BioSystems* 83(2-3), 207–216 (2006)
167. Sowa, J.F.: *Processes and causality* (2000)
168. Stamatakos, G.S., Uzunoglu, N.K., Delibasis, K., Makropoulou, M., Mouravliansky, N., Marsh, A.: A simplified simulation model and virtual reality visualization of tumour growth in vitro. *Future Generation Computer Systems* 14, 79–89 (1998)
169. Stamatakos, G.S., Uzunoglu, N.K., Delibasis, K., Makropoulou, M., Mouravliansky, N., Marsh, A., Zacharaki, E.I., Nikita, K.S.: Modeling tumor growth and irradiation response in vitro - a combination of high-performance computing and web-based technologies including vrm1 visualization. *IEEE Transactions of Information Technology in Biomedicine* 5(4), 279–289 (2001)

170. Stephanou, A., McDougall, S.R., Anderson, A.R.A., Chaplain, M.A.J., Sherratt, J.A.: Mathematical modelling of flow in 2d and 3d vascular networks: Applications to anti-angiogenic and chemotherapeutic drug strategies. *Mathematical and Computer Modelling* 41, 1137–1156 (2005)
171. Stewart, R.D., Traub, R.J.: Temporal optimization of radiotherapy treatment fractions. In: *Proceedings of the ANS Topical Meeting on Radiation Protection for our National Priorities, Medicine, the Environment, and the Legacy, RPS 2000* (2000)
172. Stoll, B.R., Migliorini, C., Kadambi, A., Munn, L.L., Jain, R.K.: A mathematical model of the contribution of endothelial progenitor cells to angiogenesis in tumors: implications for antiangiogenic therapy. *Blood* 102(17), 2555–2561 (2003)
173. Stromback, L., Lambrix, P.: Representations of molecular pathways: an evaluation of sbml, psi mi and biopax. *Bioinformatics* 21(24), 4401–4407 (2005)
174. Succi, S., Karlin, I.V., Chen, H.: Colloquium: Role of the h theorem in lattice boltzmann hydrodynamic simulations. *Reviews of Modern Physics* 74, 1203–1220 (2002)
175. Suh, C.K., Suh, E.H., Lee, D.M.: Artificial intelligence approaches in model management systems: a survey. *Computers and Engineering* 28(2), 291–299 (1995)
176. Sutherland, R.M., Durand, R.E.: Growth of multicell spheroids in tissue culture as a model of nodular carcinomas. *J. Natl. Cancer Inst.* 46, 113–120 (1971)
177. Takahashi, K., Kaizu, K., Hu, B., Tomita, M.: A multi-algorithm, multi-timescale method for cell simulation. *Bioinformatics* 20(4), 538–546 (2004)
178. Troyanskaya, O.G., Dolinski, K., Owen, A.B., Altman, R.B., Botstein, D.: A bayesian framework for combining heterogeneous data sources for gene prediction (in *saccharomyces cerevisiae*). *PNAS* 100(14), 8348–8353 (2003)
179. Tsai, Y.C.: Comparative analysis of model management and relational database management. *Omega* 29, 157–170 (2001)
180. Tyson, J.J., Chen, K., Novak, B.: Network dynamics and cell physiology. *Nature Molecule Cell Biology* 2, 908–918 (2001)
181. Ubezio, P.: Unraveling the complexity of cell cycle effects of anticancer drugs in cell populations. *Discrete and Continuous Dynamical Systems Series B* 4(1), 323–335 (2004)
182. Uhrmacher, A.M., Degenring, D., Zeigler, B.: Discrete event multi-level models for systems biology. In: *Transactions on Computational Systems Biology I*. Springer, Heidelberg (2005)
183. Van Dyke Parunak, H.: Industrial and practical applications of dai. In: Weiss, G. (ed.) *Multiagent Systems: a modern approach to distributed artificial intelligence*, p. 643. MIT Press, Cambridge (1999)
184. Vangheluwe, H., de Lara, J., Mosterman, P.J.: An introduction to multi-paradigm modelling and simulation. In: Barros, F., Giambiasi, N. (eds.) *AI, Simulation and Planning in High Autonomy Systems*, Lisbon, Portugal (2002)
185. Versweyveld, L.: Promising but challenging tumour growth simulation experiments at tu dresden. *Virtual Medical Records*, 722X (2006)
186. Vesely, F.J.: Lattice gas cellular automata (2001)
187. Villa, F.: Integrating modelling architecture: a declarative framework for multi-paradigm, multi-scale ecological modelling. *Ecological Modelling* 137(1), 23–42 (2001)
188. Villa, F., Constanza, R.: Design of multi-paradigm integrating modelling tools for ecological research. *Environmental Modelling and Software* 15, 169–177 (2000)
189. Voitikova, M.V.: Cellular automaton model for immunology of tumour growth (1998)
190. Wang, J., Tian, T.: Quantitative model for inferring dynamic regulation of the tumour suppressor gene p53. *BMC Bioinformatics* (11), 36 (2010)



191. Weiss, G.: *Multiagent Systems: A modern approach to distributed artificial intelligence*. MIT Press, Cambridge (1999)
192. Wiley, S.H., Shvartsman, S.Y., Lauffenburger, D.A.: Computational modeling of the egf-receptor system: a paradigm for systems biology. *Trends in Cell Biology* 13(1), 43–50 (2002)
193. Winsor, C.P.: The gompertz curve as a growth curve. *Proceedings in the National Academy of Science USA* 18, 1–7 (1932)
194. Wolfram, S.: *Complex systems theory*. In: *Emerging Synthesis in Science: Proceedings of the Founding Workshops of the Santa Fe Institute*, pp. 183–189. Addison-Wesley, Reading (1988)
195. Wolfram, S.: *Cellular Automata*. Perseus Books Group, Oxford (1994)
196. Wolfram, S.: *A New Kind of Science*. Wolfram Media (2002)
197. Wolkenhauer, O.: *Data Engineering: Fuzzy Mathematics in Systems Theory and Data Analysis*. Wiley Inter-Science, UK (2001)
198. Wolkenhauer, O.: Systems biology: The reincarnation of systems theory applied in biology. *Briefings in Bioinformatics* 2(3), 258–270 (2001)
199. Wolkenhauer, O., Cho, K.H., Kolch, W.: System biology: Towards an understanding of the dynamics of life. *Bioforum Europe* 1, 50–52 (2004)
200. Wolkenhauer, O., Kitano, H., Cho, K.H.: Systems biology. *IEEE Control Systems Magazine* 23(4), 38–48 (2003)
201. Wolkenhauer, O., Ullah, M., Kolch, W., Cho, K.H.: Modelling and simulation of intracellular dynamics: Choosing an appropriate framework. *IEEE Transactions on Nano-Bioscience* 3(3) (2004)
202. Yugi, K., Tomita, M.: A general computational model of mitochondrial metabolism in a whole organelle scale. *Bioinformatics* 20(11), 1795–1796 (2004)
203. Zacharaki, E.I., Stamatakos, G.S., Nikita, K.S., Uzunoglu, N.K.: Simulating growth dynamics and radiation response of avascular tumour spheroids-model validation in the case of an emt6/ro multicellular spheroid. *Comput Methods Programs Biomed* 76(3), 193–206 (2004) 0169-2607 Journal Article Validation Studies
204. Zaider, M., Minerbo, G.N.: Tumour control probability: a formulation applicable to any temporal protocol of dose delivery. *Physics in Medicine and Biology* 45, 279–293 (2000)
205. Zeigler, B., Vahie, S.: Devs formalism and methodology: Unity of conception/diversity of application. In: Evans, G.W., Mollaghasemi, M., Russell, E.C., Biles, W.E. (eds.) *Proceedings of the 1993 Winter Simulation Conference* (1993)
206. Zevedei-Oancea, I., Schuster, S.: Topological analysis of metabolic networks based on petri net theory. *Silico Biology* 3, 0029 (2003)
207. Zhang, L., Athale, C.A., Deisboeck, T.S.: Development of a three-dimensional multiscale agent-based tumour model: Simulating gene-protein interaction profiles, cell phenotypes and multicellular patterns in brain cancer. *Journal of Theoretical Biology* 244(1), 96–107 (2007)
208. Zhu, H., Huang, S., Dhar, P.: The next step in systems biology: simulating the temporospatial dynamics of molecular network. *Bioessays* 26, 68–72 (2003)

# Index

- 6-tuple, 63
- ABI, 79
- adaptive behaviour, 42
- agent, 37, 50
  - cell, 51
  - communication, 50
  - complex system, 37
  - computational, 50
  - continuity, 51
  - local interaction, 42, 55, 82
  - rules, 13, 49
- agent-based integration, 79, 81
- agent-based simulations, 15
- agent-based system, 50
- agent/cell, 91
- AnatML, 132
- angiogenesis, 16
  - physiology, 16
  - random walk, 19
  - subsystem, 60
- angiogenesis models, 16
- Bayesian network, 31
- behaviour, 1, 11, 13, 34, 37, 38, 41, 61
  - adaptive, 42
  - association, 88
  - complex adaptive, 42
  - complexity of, 41
  - diversity of, 46
  - dynamics, 37
  - emergence, 39
  - identification of, 88
  - local interaction, 15, 36, 37, 39, 150
  - model integration, 87
  - optimisation, 85
  - robustness, 42
  - rules, 41
- behavioural inclusion, 87, 88, 129, 140
- behavioural inclusion tree, 89
- biological computing, 43
- biology
  - complex system, 5
- BioPAX, 75, 131, 132
- boolean network, 21, 31
- causal flow, 38
- causality, 28, 36, 38, 39
  - emergence, 38
- cell/agent, 91
- CellML, 61, 74, 75, 79, 97, 131, 132, 148
- cellular automata, 13, 43
  - cell, 49, 51
  - discretisation, 49
  - Lattice Boltzmann, 50
  - Lattice Gas, 50
  - rules, 49
  - vs differential equations, 29
- cellular pathways, 27
- centrality, 84
- CMDL, 75, 79, 131, 133, 147, 148
- COINS, 74
- communication
  - behaviour, 39
  - isolative, 58
  - local interaction, 36, 40, 65
  - model, 58

- model integration, 66, 73
- complex adaptive behaviour, 42
- complex system, 5, 33, 46
  - agent, 46
  - component, 35
  - continuous nature, 49
  - emergence, 39
  - environment, 35
  - linearity, 35
  - local interaction, 36
  - Physiome Project, 53
  - properties, 35, 37
  - randomness, 141
  - subsystem, 37
- complexity, 1, 3, 33, 34, 37, 46, 132
  - behaviour, 13, 34, 36, 39, 65
  - biological, 1, 10, 11, 18, 46
  - local interaction, 43
  - semantics, 32
- component, 34, 35
  - local interaction, 35
  - of system, 34
- computational biology, 43
- context, 67
- continuity, 84
- decomposition
  - genetic algorithm, 135
  - differential equations, 11, 22
    - vs local interaction, 13, 14
- diffusion model, 104
- dimensionality, 84
- discrete
  - aspect of models, 62
- discretisation, 52, 61, 83, 149
  - event simulation, 75
  - event simulations, DEVS, 52
- dynamics
  - angiogenesis, 16
  - complexity, 45
  - growth, 7
  - local interactions, 37
  - molecular pathway, 21
  - patch, 70
  - systems biology, 44
- E-Cell, 46, 52, 75, 79
- emergence, 9, 38
  - properties, 38
  - strong, 38
  - weak, 38
- engineering, 1, 18, 43–45, 51
- environment, 34–36, 40
  - agent interaction, 50
  - as a system, 42
  - interactions, 54
  - local interaction, 13, 40, 44
- environmental sciences, 69–71, 73
- evolutionary algorithm, 2, 70
- FieldML, 132
- fitness function, 104
  - model solver, 106
- focus, 31, 60, 64, 88
  - management, 85
  - mapping, 89
  - solver, 67
- Four R's, 26
- functionality, 83
- genetic algorithm, 85, 97, 135
- Gleevec, 24
- Gompertz, 8
- growth
  - Gompertz, 8
- growth models, 7
- hypoxia, 27, 103, 125, 141
- integrated modelling environment, 71–73
- interactor, 51
  - identification, 66
  - state changes, 83
- John von Neumann, 43, 48
- KDA, 86
- knowledge-driven approach, 86
  - decomposition, 92
- linearity, 34, 35
  - properties, 35
- lingua franca, 65, 79
- local interaction, 13, 15, 44
  - decomposition, 83
  - evolution, 139
  - optimisation, 145
  - rules, 91

- MathML, 75, 131, 148
- mechanistic models, 6, 11, 59, 69
- mesoscale, 44, 46, 69, 93
- metamodelling, 48, 54, 69, 91, 105, 143
- microarray, 1, 28, 32, 43, 45, 70
- middle-out, 93
- model, 5, 57
  - abstraction, 69, 70
  - assumption, 66
  - assumption, processing, 68
  - assumptions, 64
  - behaviour, 11, 57
  - complexity, 3, 70
  - continuous vs discrete, 29
  - decomposition, 58, 61, 97
  - decomposition, ABI, 104
  - decomposition, GA, 107, 112
  - decomposition, local interaction
    - rules, 94
  - discretisation, 10, 13, 18, 67
  - focus, 31, 64
  - granularity, 6, 11, 55, 59, 69
  - integration, 2, 33, 52, 57
  - mechanistic, 11
  - paradigm, 60, 61, 81
  - phenomenological, 11, 13, 47, 48
  - phenomenological, decomposition, 139
  - scope, 60, 64
  - semantics, 44, 62, 75, 86
  - state, 57, 60
- model context, 66, 67
- model integration
  - agent-based, 81
  - behaviour, 81
  - common ground, 81
  - communication, 147
  - communicative strategies, 78
  - complexity, 57, 59
  - decomposite, 58
  - environmental sciences, 71
  - genetic algorithm, 135
  - isolative, 79
  - isolative communication, 58
  - itegrative communication, 58
  - linear, 52, 61, 79
  - lingua franca, 58
  - local rules, 82
  - middle-out approach, 93
  - model paradigm, 72
  - multiscale, 53, 69, 70, 91
  - solver, 73
  - systems biology, 74
- model management, 71
- model management systems, 71, 73, 133, 146
- model paradigm, 81, 83
  - 6-tuple, 63
  - assumptions, 67
  - centrality aspect, 63
  - continuity aspect, 62
  - continuous, 61
  - dimensionality aspect, 63
  - functionality aspect, 62
  - integration, 72
  - model integration, 91
  - numeric content aspect, 62
  - profile, 63, 65
  - stochasticity aspect, 63
- Modelica, 52
- modelling, 1, 5
  - angiogenesis, 16
  - behaviour, 41, 53
  - complex system, 46
  - environmental, 61
  - growth, 7
  - multiscale, 70
  - pathways, 27
  - randomness, 31
  - treatment reponse, 21
- molecular interaction, 16
  - linearity, 29
- multi-scale, 21
- multicellular tumour spheroid, 9
- multiscale model, 31
- multiscale modelling, 54, 70
  - patch dynamics, 70
- neighbourhood
  - Moore, 49
  - von Neumann, 13, 49, 105, 139
- network
  - causal, 59
  - gene, 28
  - metabolic, 28, 29
  - semantic, 62, 86
  - simulation, 30
  - systems biology, 33, 44

- non-linearity, 37
- numeric content, 84
- Object Oriented Programming, 72
- ontology, 72, 75, 86, 89, 130, 132
- OWL, 86, 88, 130, 131
- Patch Dynamics, 70
- pathway models, 27
- Physiome Project, 45, 46, 53, 149
- radiotherapy, 22, 26
- random walk, 14, 20
- reductionism, 39, 43, 77
- reponse models, 21
- rule
  - optimisation, 85
- rule disambiguation, 98, 99, 129
- rules, 49, 83, 84
  - combinations, 92
  - decomposition, 130
- SBML, 28, 45, 58, 74, 79, 131–133, 146, 149
- SCIPath, 44, 146
- scope, 60
- self-organisation, 42
- simulation, 5, 28, 50, 69
  - bottom up, 70
  - complex systems, 51
  - DEVS, 52
  - discrete, 83
  - functional, 67
  - software, 51
  - systems biology, 43
- solver, 5, 60, 61, 133, 146
  - discrete, 62
  - integration, 73
  - numerical, 53
- state, 12, 34, 42, 47
  - agent, 50
  - cellular automata, 49
  - dependencies, 47
  - dynamics, 39
  - environment, 40
  - phenomenological, 47
  - transition function, 47
- state machine, 48
- state space, 67
- stochasticity, 84
- structured modelling, 71, 76
- submodel, 61, 63–65, 92
  - integration, 61
- subsumption, 93, 115
  - ontologies, 130
- subsystem, 20, 64
- system, 9, 11, 12, 33, 35
  - behaviour, 11, 35, 46, 48
  - complexity, 40
  - linearity, 34, 35
  - non-linearity, 37
  - state, 12
- SystemConcept, 89
- systems biology, 6, 11, 24, 33, 43, 57, 74
  - dynamics, 44
  - reductionism, 43
- Systems Biology Workbench, 44, 58, 74, 79, 145, 146
- therapy
  - model, 22, 23, 26
  - resistance, 9
- treatment response, 21
- tumour, 5, 9
  - behaviour, 45
  - dynamics, 6, 54
  - growth, 5
  - modelling, 5
  - simulation, 5, 31, 60
- tumour cord, 9
- tumour growth, 7
- vascularisation, 7, 10, 15, 16
- Virtual Cell, 46, 52