Mrinmoy Majumder
Rabindra Nath Barman *Editors*

Application of Nature Based Algorithm in Natural Resource Management

Springer

Application of Nature Based Algorithm in Natural Resource Management

Mrinmoy Majumder • Rabindra Nath Barman
Editors

Application of Nature Based Algorithm in Natural Resource Management

Springer

Editors
Mrinmoy Majumder
School of Hydroinformatics Engineering
National Institute of Technology Agartala,
Jirania, India

Rabindra Nath Barman
Department of Production Engineering
National Institute of Technology Agartala
Barjala, Tripura, India

ISBN 978-94-007-5151-4 ISBN 978-94-007-5152-1 (eBook)
DOI 10.1007/978-94-007-5152-1
Springer Dordrecht Heidelberg New York London

Library of Congress Control Number: 2013934707

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

The global population surge, along with uncontrolled extraction of natural resources, has increased the concentration of greenhouse gases (GHG) in the atmosphere that has caused global warming nearly everywhere on Earth. Due to global warming, regular climate patterns have been disturbed, and this change in climate has induced abnormalities in the hydrodynamics of many watersheds worldwide.

That is why management of natural resources has become essential in mitigating the impacts of climate change. Activities like afforestation, river-bed restoration, innovative irrigation practices, and utilization of renewable energy are now encouraged.

In this regard the two biggest obstacles faced by scientists and engineers are the unpredictability of the hydroenvironmental response to climate change and the selection or development of optimal, cost-effective, but useful mitigation options. Along with changes in climate in the near future, due to population overgrowth and uneven distribution of natural resources, urbanization will increase rapidly.

The impact of climatic abnormalities on natural resources is unpredictable. The effects can only be imagined, but they cannot be determined precisely. Numerous studies have been conducted to predict the outcome of various climatic abnormalities. Scientists conceptualize climatic scenarios for the future, and simulation platforms are used to predict the outcome of these scenarios so that humankind becomes prepared for the onset of natural extremes.

Climate models of varying dimensions and concepts have been developed to predict future climate with the help of the knowledge of presently occurring events. A climate model is a system of differential equations based on the basic laws of physics, fluid motion, and chemistry. These models can generate future patterns of climate parameters based on time and space. Such models were then combined with hydrologic, hydroenvironmental, hydraulic, or geophysical models to estimate future impacts on the hydrology, environment, and geophysical characteristics of different basins.

In this context *Application of Nature-Based Algorithms on Natural Resource Management* also tries to predict certain parameters by applying different nature-based algorithms. The prediction models described in this book are all short term and mostly meta-heuristic in nature.

Chapters 4, 8, and 16, for example, aim to estimate short-term weather patterns with the help of neural networks, neuro-fuzzy techniques, and a novel cognitive modeling platform called CLIMAGE.

In Chap. 4, advancement of artificial neural network is used to predict short-term rainfall, i.e., the occurrence and intensity of rainfall within the next 5 days. Short-term rainfall is extremely important for scheduling production works and different other time dependant business activities because such activities are directly correlated with rainfall. The study used the nonlinearity mapping and pattern-identification ability of neural networks to estimate the highly uncertain parameters.

In Chap. 8, occurrences of extreme events are predicted with the help of neuro-fuzzy techniques. The advantage of fuzzy logic was applied to identify ideal weightage for the input variables to create an objective function that could differentiate between important and unimportant variables without over- or underestimating the degree of influence of these variables.

In Chap. 16, a novel software platform is developed. This software uses advances in neuro-fuzzy techniques in the estimation of short-term weather patterns with the help of satellite imagery. Satellite images are now widely used in a variety of engineering studies where space is an important parameter. The software uses fuzzy decision making for the identification of importance for the gridded parameters that are varied from the location of interest in an inverse manner. A neurogenetic algorithm is applied to predict the final outcome from the model.

As discussed previously, the climate models are coupled with other conceptual or statistical models for the prediction of climatic impact on related variables as estimated by the latter model. In this monograph, Chaps. 3, 6, 9, 11, 12, 13, 17, and 19 demonstrate the application of climate models coupled with other predictive platforms for estimating the impact of climatic abnormalities on different related, parameters like load factor, water availability, lotus cultivation, ecological and hydrological sensitivity indices, and location selection of a tidal power plant.

In Chap. 3, the interrelationship of rainfall and load factor of a small-scale hydropower plant is analyzed using particle swarm optimization (PSO). The optimal zones for which a minimal rainfall amount will satisfy the maximum amount of demand was identified using the searching ability of PSO.

Chapter 6 compares different nature-based algorithms like neurogenetic, genetic, ant colony (ACO), and bee colony algorithms (ABC) for their efficiency in the prediction of water resource availability for a given series of climatic and hydrologic variables categorized with respect to its impact on water availability. The availability of water is predicted by the nature-based algorithms, and the prediction accuracy is analyzed with the help of Cohen's kappa index of agreement. The impact of climate change is estimated with the help of the given nature-based algorithm where the output of the PRECIS climate model for IPCC A2 and B2 scenarios are used as the future data of the climate parameters.

Chapter 9 predicts the impact of climate change on site selection of lotus cultivation, whereas the climatic influence on location preference for tidal power plants is discussed in Chap. 13. The chapters use the efficacy of neural networks and fuzzy logic to identify suitable locations for these activities and depict the change in location selection due to probable future climatic abnormalities.

Chapter 17 describes a novel software developed for location selection studies with the help of neuro-fuzzy techniques. The software has five different sections: climate, geophysical, ecological, socioeconomic, and electrical inputs. Based on the given inputs of the available locations the model selects the optimal location for tidal power generation. The location suitability of all locations can be compared as a whole or with respect to different input variables. Chapter 19, on the other hand, shows the efficiency of another novel software in estimating impacts from extreme events on the output of a surface water treatment plant (SWTP).

Chapters 11 and 12 describe the methodology for estimating climatic impact on wetlands and estuarine reserve forests. Both studies use an index considering the related ecological (Chap. 11) and hydrological (Chap. 12) parameters. The results of the study will help engineers identify most probable causes of climatic and hydrologic uncertainties existing in the considered locations.

Management of natural resources will be different for pre- and post-climate-change days. The mitigation measures that are implemented early to compensate for weather abnormalities are now becoming useless and insufficient to counter recent extremes. That is why selection of an ideal compensatory measure to either counter or reverse the atrocities of climate-change effects is crucial.

The present collection of research studies also includes some novel investigations in this regard. The application of nature-based algorithms has been found to be a nice tool in simulating the effect of different mitigation scenarios; further, based on the simulation results some educated and knowledgeable decisions can be made so as to withstand the onset of climate related-disasters. This decision-making system will be less error prone and more optimal than the earlier measures due to the advantage of including intuitive nature-based algorithms within the decision-making procedures.

In this book the mitigation measures are divided into three distinct categories: ecology, energy, and land use conservation in location selection studies (Chaps. 1, 2, and 10), innovative mitigation practices (Chap. 7), and rating of existing mitigation measures (Chaps. 15, 18, and 20).

Ecoparks are proposed as one of the better options to mitigate the impact of climate change and as a source of regular income for the dependent population. But the selection of location is important for the success of such options. Chapter 1 of the book highlights a methodology that can be used in the selection of a suitable location for the establishment of ecoparks. The study uses an ant colony algorithm where the concept of ant's food foraging is mimicked to identify the optimal location for ecoparks.

Chapter 2 depicts a methodology for determining a suitable habitat for porcupines. The bee hive theory, or the logic of food-search procedures of a bee colony, was replicated to solve the said problem.

Chapter 10 compares the fuzzy logic and food-search methodology of bats in identifying the best location for installing a small-scale hydropower plant. The data set is represented in a categorized form depicting the degree of influence of each variable on the objective function. For this, popular performance metrics like sensitivity, specificity, precision, and Cohen's kappa index of agreement were utilized to identify the best algorithm among the two.

Chapter 7 depicts a novel procedure to select plant species that would be optimal for cultivation in vertical irrigation systems. This type of irrigation system is here proposed as an alternative method of irrigation for solving the problem of both water and land scarcity that will be common as a result of population overgrowth and large-scale urbanization.

A cognitive index is applied in Chap. 15 to rate capacity of existing irrigation canals to withstand natural extremes. The study proposes a new method for evaluating canals to help engineers engineers in planning the maintenance according to the vulnerabilities of the structures thereby saving a substantial amount of taxpayer's money.

Chapter 18 proposes another new method to rate crop varieties according to their water productivity, which represents the water consumed by a crop. A novel software was developed using this method to help farmers estimate profit from culturing a certain kind of crop. The software enables the user to enter meteorological and hydrological data, along with command area characteristics, to analyze the crop productivity with respect to different aspects like land use, water availability, and others.

Chapter 20 develops a method to rate species for afforestation purposes. The species with the maximum afforestation potential was selected by the application of a fuzzy logic based novel methodology.

Chapters 5 and 14 deal with the impact of urbanization on natural resource management. Chapter 5 demonstrates the application of genetic algorithm to the growth rate of *Bufo* sp. or Asian frog. This species of amphibian is important to maintain the ecological balance of an ecosystem. The species is responsible for controlling insect and worm populations. But due to urbanization, frogs are now seldom spotted in cities, and many species of frog are now under threat of complete extinction. Thus, the growth rate of the *Bufo* sp. was estimated using genetic algorithm to compare the degree of vulnerability of these species to different level of urbanization in order to determine the proper compensatory measures.

Chapter 14 presents a pollution study where groundwater quality is estimated using surface water quality. Urban water supplies often include tube wells or a submersible pump system for the abstraction of groundwater for drinking as well as industrial purposes. This is why the pollution of groundwater is very harmful and can spread health hazards throughout an urban area if not properly monitored. This chapter proposes a neurogenetic model for estimating groundwater quality on the basis of surface water quality. Because surface water is readily accessible for sampling and because surface water is generally contaminated by deadly pollutants from domestic and industrial wastes, the study can help urban water engineers to estimate groundwater quality by measuring surface water quality and adopt mitigation approaches to reduce underground contamination.

In all the chapters, nature-based algorithms are used to predict parameters that are essential for natural resource management. The book represents a collection of novel methodologies and software platforms to be used by the average person as well as by professionals to estimate the impact and devise optimal mitigation measures to counter the onset of climatic abnormalities on management of natural resources.

Acknowledgments

The authors of the book would like to take this opportunity to express their heartfelt gratitude for the kind suggestions, and encouragements, along with the necessary cooperation of Prof. Dr. Probir Kumar Bose, Director, National Institute of Technology Agartala, Tripura West, without which the book might have never become a reality.

The authors would also like to show their appreciation to each of the contributors of the chapters, included in this book and duly acknowledge their perseverance, hard work, and dedication, which facilitated the completion of the manuscript in due time.

The authors offer their sincere thanks to their colleagues, staff, and students for their cooperation and benedictions for which we will be forever indebted.

The authors will like to take this opportunity to acknowledge the editors, publishers and writers of the journals, books and reports from which they have collected and cited various useful data and informations to support their explanations and findings. In this regard, the authors also wish to convey their gratitude towards the institutes like Central Water Commission, Indian Metereological Department, Indian Institute of Tropical Metereology etc., for providing and publishing in public domain the database widely required for the the studies, results of which were depicted throughout this monograph.

Above all, the authors will also like to show their heartiest regards to the support endowed upon them by their parents and family without which this presentation was impossible to prepare.

It is inevitable that there might be errors and omissions in preparing the text for which sincere apologies are being made. The authors will also like to invite the readers of this book for their valuable feedbacks and criticism which will help in modifying the present edition to a great extent.

Agartala, India

Mrinmoy Majumder
Rabindra Nath Barman

Contents

Part I
New Nature Based Algorithms

Chapter 1
Selection of Optimized Location for Ecoparks Using Ant Colony Optimization

Mrinmoy Majumder and Soumya Ghosh

Abstract Ecoparks are an essential component of any ecotourism project. Such parks are developed to protect, preserve, and manage natural land use and land covers that are necessary for the maintenance of the ecological balance of a region. But ecoparks or ecotourism projects have some drawbacks that make the site-selection procedure one of the major determinants of the success of such project. Pollution, the destruction of habitats of wild animals, disturbance to the daily life of indigenous and native people and an increase in the price of local items are some of the major negative contributions of ecotourism. Thus, such projects must be allowed in sites where disturbance and destruction to the local ecosystem will be minimal. But till now no specific and standard methodology has been proposed for the selection of ecopark sites. The selection of sites is now made based on reports received from experts in the related fields; in addition, sometimes the response from native people is also included in the decision-making process. But such decisions often raise controversies, and accusations of favoritism and illegality are rampant. Because there is no quantifiable methodology of site selection, such tumult as discussed above is quite natural. That is why, the present study tries to formulate a methodology of site selection for ecopark projects that is objective and does not depend on the qualitative ratings of experts. Decisions are made using probabilistic optimization methods like ant colony optimization (ACO) algorithm. The ACO algorithm proposed by Dorigo et al. has resolved many critical decision-making problems like the famous problem of the traveling salesman and other within with satisfactory results.

M. Majumder (✉)
School of Hydro-Informatics Engineering, National Institute of Technology Agartala,
Barjala, Jirania 799055, Tripura, India
e-mail: mmajumder15@gmail.com

S. Ghosh
School of Hydro-Informatics Engineering, National Institute of Technology Agartala,
Barjala, Jirania 799055, Tripura, India
e-mail: soumyaee@gmail.com

M. Majumder and R.N. Barman (eds.), *Application of Nature Based Algorithm in Natural Resource Management*, DOI 10.1007/978-94-007-5152-1_1,

Keywords Ant colony optimization • Ecopark • Decision making • Location selection

1.1 Introduction

1.1.1 Ecoparks and Ecotourism

Ecoparks can be defined as parks using ecological landscape features to reduce water use and maintenance while enhancing wildlife and human values. Ecoparks generally give rise to ecotourism. Ecotourism can be defined as "responsible travel to natural areas that conserves the environment and improves the well-being of local people" (TIES 1990).

According to the International Ecotourism Society the principles of ecotourism are as follows:

1. Minimization of the impact of deforestation, urbanization, and climate change;
2. Awareness, culture, and respect for the natural ecosystem;
3. Financial benefits for different conservation mechanisms;
4. Financial benefits for and empowerment of indigenous peoples;
5. Development of positive relationships between visitor and host;
6. Increased sensitivity toward environment, politics, and social equilibrium

Ecotourism projects raises awareness about the importance of various environmental issues and educates the public about the consequence of the uncertainty before it actually happens. Such projects also develops public and financial awareness and monitory supports for various conservation measures and activities.

But many times ecotourism has failed to achieve its objective due to a lack of funding and support from tourists and hosts. If ecotourism projects are not developed in places where they are really needed, the projects will invariably fail. Ecotourism projects can be built in the middle of a large city or where no environmental degradation has been observed for the last 50 years. If people are asked to pay to visit a river, they will be more likely to exclude that place from their itinerary.

The importance of location selection has become one of the crucial factors in the success of projects with respect to economic, social, and environmental goals.

1.1.2 Objective and Methodology Overview

The present investigation proposes a methodology for selecting a suitable location for ecotourism projects. In the study, important parameters for such a selection are first identified based on a review of the relevant literature and with support from experts in related fields.

The data for each parameter must first be classified into five to nine groups based on the data quality and quantity. All the groups are then rated according to ant colony food-search logic.

The parameters are rated in such a way that the higher the value of the resulting function, the more suitable the given location is a place for ecotourism projects.

1.1.3 Ant Colony Food-Search Logic and Algorithms

'The ant colony algorithm is an algorithm for finding optimal paths and is based on the behavior of ants searching for food'.

In their search for food, the ants first wander randomly. When an ant finds a source of food, it walks back to the colony leaving "markers" (pheromones) indicating that the path leads to food. When other ants come across the markers, they will follow the path with a certain probability. They then populate the path with their own markers as they bring the food back. As more ants find the path, it gets stronger until there are several streams of ants traveling to various food sources near the colony.

Because the ants drop pheromones every time they bring food back to the colony, signals associated with shorter paths are more likely to be stronger, hence optimizing the "solution." In the meantime, some ants are still randomly scouting for closer food sources. Once a food source is depleted, the route is no longer populated with pheromones and slowly decays.

Because an ant colony is a very dynamic system, the ant colony algorithm works very well in graphs with changing topologies. Examples of such systems include computer networks and artificial intelligence simulations of workers (Macura 2012).

1.1.3.1 Applications

Ant colony optimization (ACO), i.e., optimization that applies the logic of an ant colony in the search for a food source, has been applied in many optimization and decision-making problems. Table 1.1 depicts a few of the applications of ACO in different fields of science and technology for solving optimization, decision-making, classification, data mining, scheduling, routing, and many other kinds of problems.

1.2 Methodology

The latest research shows that the following seven primary factors must be considered during the site-selection process for an ecopark to ensure its success:

i. Accessibility
ii. Place of Interest/Traditional Tourist Destination
iii. Impact on Native Peoples and Wildlife
iv. Impact on Ecosystem

Table 1.1 Table Showing Some of the Major Studies where ACO is Applied in Optimization and Classification Research

Authors	Type of problem	Subject area	Success rate
Eroğlu and Seçkiner (2012)	Optimization	Windfarm layout optimization	Performance better than other evolutionary problems used for solving similar problems
Hu et al. (2012)	Risk Control	Risk control of energy performance contracting	Solution by ACO was found to have better diversity and convergence
He and Hou (2012)	Optimization	Traffic control (signal timing optimization)	Effective
Yan-hua et al. (2011)	Data routing	Cloud database	Effective in combination with genetic algorithm
Wang et al. (2011)	Multicast routing	Network communication within a multicast tree keeping bandwidth, delay time, and jitters as constraints.	Efficiency of searching, adaptation, and convergence was found to be better than with the other algorithms
Razavi and Jalali-Farahani (2010)	Process optimization	Petroleum engineering	Fast and accurate
Afshar (2010)	Optimization	Storm sewer design	Effective
Zecchin et al. (2006)	Optimization	Water distribution system	Max Min ant system algorithm (local search) was found to perform better than ant system (global search) algorithm
Yang and Kamel (2006)	Clustering	Plant classification	Effective
Rajendran and Ziegler H June (2004)	Scheduling problem	Minimizing the make span and total flow time of jobs	92.22% success rate

v. Availability of Free Land
vi. Hostile Activities
vii. Security Arrangements

These 7 seven factors can be farther subdivided into 18 secondary factors and 53 tertiary factors for an in-depth review of a site for its suitability as a location for a successful ecopark project. All the sub-subfactors were selected based on a review of the latest journal articles and government reports.

The hierarchy of decision-making parameters is given in Table 1.2.

The tertiary factors are first converted into categorical values based on their impact on a decision. Different factors are encoded in different categories with respect to their relationship with the decision variables. The threshold of the different categories of factors is determined using common sense, field reports, and various scientific studies.

Table 1.3 shows the categories of the factors.

Every possible combination between the categories and their factors was generated, and each factor was scored in light of the foregoing logic in terms of ant colony's food searching activities. After the categories were properly rated with respect to the ACO, the scores were added and the sum of scores for each combination was ranked inversely. The probability of each combination was also determined. According to probability analysis, combinations with the highest probabilities were selected as the best for an ecotourism project; a site with such a combination can be recommended for ecotourism.

Table 1.2 Primary, secondary, and tertiary factors considered in the selection of suitable sites for ecopark projects

Primary factors	Secondary factors	Tertiary factors
Accessibility	Road	R_Distance
		R_Count
		R_Type
	Airport	A_Distance
		A_Count
		A_Type
	Rail	Ra_Distance
		Ra_Count
		Ra_Type
	Town	Town_Distance
		Town_Count
		Town_Type
	Hospital	H_Distance
		H_Count
		H_Type

(continued)

Table 1.2 (continued)

Primary factors	Secondary factors	Tertiary factors
Place of interest	River	Ri_Type
		Ri_Characterestics
		Ri_Importance
	Lake	l_Distance
		L_Count
		L_Type
		L_Characterestics
		L_Importance
	Temple	T_Distance
		T_Count
	Popular places	pp_Distance
		pp_Count
Impact on native peoples and wildlife	WL	WL_Distance
		WL_Count
		WL_Type
	Presence of tribes	Tr_Distance
		Tr_Count
		Tr_Type
		Tr_Hostile/Not
	Cultivators	C_Distance
		C_Count
		C_Type
		C_Hostile/Not
Impact on ecosystem	Plant species	PS_Count
		PS_Type
	Number of trees to be removed	Tree_Count
		Tree_Type
Availability of free land	Amount of land that can be acquired	Land_Count
		Land_Type
	Landowners	LO_Count
		LO_Type
		LO_Hostile/Not
Hostile activities	Hostile activities	HA_Distance
		HA_Types
		HA_Frequency
Security arrangements	Police	P_Distance
		P_Count
		P_Type

Notes: Quantitative and qualitative factors are indicated by different colors (cyan and dark green respectively). Prefixes of tertiary factors are given to associate them with the respective secondary factors

Table 1.3 Tertiary factors, categories they are encoded in, and ratings given according to ant colony food-search logic

Factor	Category	Rating
R_Distance	Extremely Far (EF), Very Far (VF), Semi Far (SF), Far (F), Normal (N), Near (Ne), Semi Near (SNe),Very Near (VNe), Extremely Near (ENe)	According to ACO, the shortest path receives the most markers, but if a path road leads to a source where the quantity or quality of food is not optimum, the ants will soon reject the path. The distance of a road from an ecotourism project is a criterion whereby too far and too close are rejectable. If roads are very far, people cannot reach the project area easily. If the road is very near, then the tranquility of the place will be compromised. Therefore, ENe and EF are given a score of 1, whereas M is given a score of 5.
R_Numbers	Extremely High (EH), Very High (VH), Semi High (SH), High (F), Medium (M), Low (L), Semi Low (SL),Very Low (VL), Extremely Low (EL)	According to ACO, a path having a higher concentration of pheromones will receive a higher rating. Thus, EH (highest concentration of roads) is given a score of 9, whereas EL (lowest concentration of roads) is given a score of only 1. In the selection of ecotourism projects a high concentration of roads will disturb the tranquility of a location, which will, of course, compromise the destination's popularity.
R_Type	National Highway, State Highway, Urban Road, Rural Road	According to ACO, a path having a higher quality of pheromone will receive a higher rating. Thus, National Highway, which is the highest-quality road of a country, is given a rating of 4 and a rural road, or lowest-quality road of a country, is given a 1.
A_Distance	Extremely Far (EF), Very Far (VF), Semi Far (SF), Far (F),Normal (N), Near (Ne), Semi Near (SNe),Very Near (VNe), Extremely Near (ENe)	According to ACO, the shortest path receives the most markers, but if the path leads to a source where the quantity or quality of food is not optimum, then the ants will soon reject the path. The distance from an airport must be rated similarly the distance from a road. Therefore, ENe and EF are given a score of 1, whereas M is given a score of 5.
A_Numbers	Extremely High (EH), Very High (VH), Semi High (SH), High (F), Medium (M), Low (L), Semi Low (SL), Very Low (VL), Extremely Low (EL)	According to ACO, a path having a higher concentration of pheromones will receive a higher rating. Thus, EH (or highest concentration of flights) is given a score of 9, whereas EL (or lowest concentration of flights) is given a score of only 1. For the present objective, a large concentration of airports will always increase the probability that the place will be included in tourist itineraries.
A_Type	National Airport, International Airport	According to ACO, a path having a higher-quality pheromone will receive a higher rating. Thus, International Airport, which is the highest-quality airport of a country, is given a rating of 2 and National Airport, or a low-quality airport, is given a 1.

(continued)

Table 1.3 (continued)

Factor	Category	Rating
Ra_Distance	Extremely Far (EF), Very Far (VF),Semi Far (SF), Far (F), Normal (N), Near (Ne), Semi Near (SNe),Very Near (VNe), Extremely Near (ENe)	This factor is similar to the distance from a road or airport. Therefore, both ENe and EF are given a score of 1, whereas M is assigned a rating of 5.
Ra_Numbers	Extremely High (EH), Very High (VH), Semi High (SH), High (F), Medium (M), Low (L), Semi Low (SL), Very Low (VL), Extremely Low (EL)	This factor is similar to the R_Numbers and A_Numbers factors and is rated in a similar fashion.
Ra_Type	Broad Gage, Narrow Gage	According to ACO, a path having a higher-quality pheromone will receive a higher rating. Thus, Broad Gage, which is a higher-quality rail track, is given a rating of 2 and Narrow Gage, or the lowest-quality rail track, is given a 1.
Town_Distance	Extremely Far (EF), Very Far (VF), Semi Far (SF), Far (F), Normal (N), Near (Ne), Semi Near (SNe), Very Near (VNe), Extremely Near (ENe)	If a town is very far away, then it will be difficult for tourists to obtain a steady supply of food, medicine, and other necessary and luxury items. But if a town is very close, then tourists will ignore the place because it will disturb the tranquility of the place, which is why tourists would want to visit a particular location in the first place. Also a nearby town will always try to extend its perimeter and thereby decrease that of the project. Thus, the rating of this factor follows the same method followed in the case of roads, airports, and rail distances.
Town_Numbers	Extremely High (EH), Very High (VH), Semi High (SH), High (F), Medium (M), Low (L), Semi Low (SL), Very Low (VL), Extremely Low (EL)	With this factor, the same methods are followed as were adopted for the roads, airports, and rail numbers .
Town_Type	Extremely High (EH) Population, Very High (VH) Population, Semi High (SH) Population, High (H) Population, Medium (M) Population, Low (L) Population, Semi Low (SL) Population, Very Low (VL) Population, Extremely Low (EL) Population	According to ACO, both the quantity and quality of pheromones must be optimal for attracting ants to the desired path. A very populated town is not conducive to tourism, but a sparsely populated one is not an optimal option, either. Thus, towns with medium-size populations are given the highest rating of 5 and both EH and LH are given the lowest rating of 1.

H_Distance	Extremely Far (EF), Very Far (VF), Semi Far (SF), Far (F), Normal (N), Near (Ne), Semi Near (SNe), Very Near (VNe), Extremely Near (ENe)	According to ACO, the shortest path receives the maximum electability or rating. Therefore, ENe is given a score of 9, whereas EF is given a score of 1. In any ecotourism project, the availability of a good hospital will always make a given location more attractive to tourists. Thus, with this factor the earlier method (R_, Ra_, A_, Town_Distance) is not followed.
H_Numbers	Extremely High (EH), Very High (VH), Semi High (SH), High (F), Medium (M), Low (L), Semi Low (SL), Very Low (VL), Extremely Low (EL)	According to ACO, the higher the concentration of pheromone on a path, the higher will be the path's rating. Thus, EH (highest concentration of hospitals) is given a score of 9, whereas EL (lowest concentration of hospitals) is given a score of only 1.
H_Type	Super Specialty Hospital (SSH), Specialty Hospital (SH), Hospital (H), Public Health Center (PHC)	According to ACO, the higher the quality of pheromones a path has, the higher its rating. Thus, Super Specialty Hospital, which is the highest-quality hospital, is given a rating of 4 and Public Health Center, which represents the lowest-quality health service provider, is given a rating of 1.
Ri_Type	Extremely High (EH) Flow, Very High (VH) Flow, Semi High (SH) Flow, High (H) Flow, Medium (M) Flow, Low (L) Flow, Semi Low (SL) Flow, Very Low (VL) Flow, Extremely Low (EL) Flow	EH flow is given a score of 9 as it increases the chances of participation in adventure sports and other related activities, whereas rivers with EL flow will have no chance of offering such activities. Not even boating will be possible. Thus the EL category is given a score of only 1. In ACO, the higher the quality and quantity of food a path has, the higher the probability it will be selected.
Ri_Characteristics	Extremely Big (EB), Very Big (VB), Semi Big (SB), Big (B), Medium (M), Small (S), Semi Small (SS), Very Small (VS), Extremely Small (ES)	An EB river is given a score of 9 because it increases the natural beauty of a place, whereas Extremely Small rivers will have no such natural attractions. Thus the EL category is given a score of only 1. In ACO, a path having a higher quality and quantity of food will always have a higher probability of being selected.
Ri_Importance	Extremely High (EH), Very High (VH), Semi High (SH), High (H), Medium (M), Low (L), Semi Low (SL), Very Low (VL), Extremely Low (EL)	Rivers are very important. They can be sacred or popular due to their landscape, or they may serve as providers of water to adjacent farms and industries. The importance of a river increases with its usability. The greater the importance of a river, the higher its rating will be.
L_Distance	Extremely Far (EF), Very Far (VF), Semi Far (SF), Far (F), Normal (N), Near (Ne), Semi Near (SNe), Very Near (VNe), Extremely Near (ENe)	According to ACO, the shortest path leading to a large quantity of food will always receive the maximum electability or rating. A lake always raises the suitability of a location as an ecotourism destination. Therefore, ENe is assigned a score of 9, whereas EF is assigned a score of 1.

(continued)

Table 1.3 (continued)

Factor	Category	Rating
L_Number	Extremely High (EH), Very High (VH), Semi High (SH), High (H), Medium (M), Low (L), Semi Low (SL), Very Low (VL), Extremely Low (EL)	According to ACO, a path having a higher concentration of pheromones will receive a higher rating because that will convey to other ants that there is a confirmed source of a large amount of better-quality food at the end of the path. Thus, EH (the highest concentration of lakes) is given a score of 9, whereas EL (the lowest concentration of lakes) is given a score of only 1.
L_Type	Extremely Big (EB), Very Big (VB), Semi Big (SB), Big (B), Medium (M), Small (S), Semi Small (SS), Very Small (VS), Extremely Small (ES)	EB Lake is given a score of 9 because it increases the natural beauty of a place and increases the chance of boating and other related pastimes, whereas Extremely Small lakes or ponds will have no such activities. Thus the EL category is given a score of only 1.
L_Characteristics	Perennial, Rainfed	Perennial lakes attract tourists throughout the year. Thus, the quality of such places as an eco-hotspot is higher than rainfed lakes because the latter do not have water during the nonmonsoon season.
L_Importance	Extremely High (EH), Very High (VH), Semi High (SH), High (H), Medium (M), Low (L), Semi Low (SL), Very Low (VL), Extremely Low (EL)	Lakes are very important. They can be sacred or popular due to their landscape or they may serve as the only source of water to adjacent farms. The importance of a lake increases with its usability. Thus, the more important a lake is, the higher the rating it will have. Similarly, for an ant, the greater the quantity and quality of food source at the end of a path, the larger the deposition of pheromones by returning ants; hence, that path is more likely to be selected.
T_Distance	Extremely Far (EF), Very Far (VF), Semi Far (SF), Far (F), Normal (N), Near (Ne), Semi Near (SNe), Very Near (VNe), Extremely Near (ENe)	Temples alway attract tourists. Therefore, ENe is given a score of 9, whereas EF is given a score of 1.
T_Numbers	Extremely High (EH), Very High (VH), Semi High (SH), High (F), Medium (M), Low (L), Semi Low (SL), Very Low (VL), Extremely Low (EL)	The concentration of temples will only increase the influx of tourists and the probability that a given location will be selected as an ecotourism destination.

pp_Distance	Extremely Far (EF), Very Far (VF), Semi Far (SF), Far (F), Normal (N), Near (Ne), Semi Near (SNe), Very Near (VNe), Extremely Near (ENe)	According to ACO, the shortest path with the highest-quality food always receives the most pheromones to increase its probability of being selected. In the case of the present factors, if a popular location is in very close proximity, then tourists will flock to it due to the dual attraction of the place. Therefore, ENe is given a score of 9, whereas EF is given a score of only 1.
pp_Numbers	Extremely High (EH), Very High (VH), Semi High (SH), High (F), Medium (M), Low (L), Semi Low (SL), Very Low (VL), Extremely Low (EL)	According to ACO, the higher the concentration of pheromones a path has, the higher the rating the path will receive. Thus, EH (highest concentration of pheromones) is given a score of 9, whereas EL (lowest concentration of pheromones) is given a score of only 1.
WL_DISTANCE	Extremely Far (EF), Very Far (VF), Semi Far (SF), Far (F), Normal (N), Near (Ne), Semi Near (SNe), Very Near (VNe), Extremely Near (ENe)	This factor is rated similar to the pheromone factor because, just as with the presence of pheromones, the presence of wildlife increases the desirability of a place as a tourist destination.
WL_Numbers	Extremely High (EH), Very High (VH), Semi High (SH), High (F), Medium (M), Low (L), Semi Low (SL), Very Low (VL), Extremely Low (EL)	This factor receives a rating similar to that of pp_Numbers.
WL_TYPE	Endangered (ED), Rare, Common	A forest with endangered animals will always be more attractive to tourists than other places with rare or common species. Thus, ED is given a rating of 3 and Common a rating of 1. Analogously to ACO, the quality of the place as a tourist destination increases due to the presence of ED animals. Thus, the probability that this place will be selected as an ecotourism project will also increase.
Tr_Distance	Extremely Far (EF), Very Far (VF), Semi Far (SF), Far (F), Normal (N), Near (Ne), Semi Near (SNe), Very Near (VNe), Extremely Near (ENe)	In most cases, sites are not permitted to be too close to tribal peoples so that the uniqueness of the tribe's culture is not affected. But the presence of tribes always increases tourist concentration as visitors want to interact with and learn about the lifestyles of indigenous peoples. Thus, this factor is rated similar to the R_Distance parameter.
Tr_Numbers	Extremely High (EH), Very High (VH), Semi High (SH), High (H), Medium (M), Low (L), Semi Low (SL), Very Low (VL), Extremely Low (EL)	For reasons that are similar to those of the earlier factor and due to the sudden hostile activity by tribal peoples, this factor received a rating similar to Tr_Distance.

(continued)

Table 1.3 (continued)

Factor	Category	Rating
Tr_Type	Endangered (ED), Rare, Common	To ensure proper preservation of ED tribes, places with ED tribes are given a lower rating than other tribe types, Rare (2) or Common (3).
Tr_Hostile/Not	Yes (Y), No (N)	If a tribe is hostile to outsiders, the suitability of the given location as an ecotourism destination becomes 0 because a noncooperating tribal population can harm tourists. Ants also avoid paths that are hostile to travelers (a regular location of disturbance due to predators or any other human activities or geophysical obstacles) even if the path is short and filled with quality markers.
C_Distance	Extremely Far (EF), Very Far (VF), Semi Far (SF), Far (F), Normal (N), Near (Ne), Semi Near (SNe), Very Near (VNe), Extremely Near (ENe)	This factor is rated similar to R_Distance as too close proximity to farms may damage the harvest, but tourists love to observe agro-business in action.
C_Numbers	Extremely High (EH), Very High (VH), Semi High (SH), High (F), Medium (M), Low (L), Semi Low (SL), Very Low(VL), Extremely Low (EL)	This factor is rated similar to R_Distance for similar reasons.
C_Type	Endangered Crops (ED), Rare Crops (RARE), Common Crops (COMMON)	ED crops are important in conservation and also attract tourists. But if a site is too close, Visitors to the park can also damage the cultivation of crops. Thus, ED and Rare are assigned ratings of 1 and 2, respectively, but Common crops receive a rating of 1 since as they are not interesting for tourists because they can be seen anywhere.
C_Hostile/Not	Yes (Y), No (N)	If farmers are hostile to ecotourism projects, the suitability of the given location as an tourism destination becomes 0. Ants also avoid paths that are inhospitable (a regular location of disturbance due to predators or anthropogenic activities) even if the path is short.
PS_Numbers	Extremely High (EH), Very High (VH), Semi High (SH), High (F), Medium (M), Low (L), Semi Low (SL), Very Low (VL), Extremely Low (EL)	Higher the diversity of Plant Species(PS) higher will be the ecological attractiveness of a place as a tourist destination. Thus, EH is given a rating of 9 and EL a rating of 1.

PS_Type	Tree, Shrub, Herb, Ornamental, Medicinal	Tourists are very interested in medicinal plants. The next most popular species of flora are ornamental trees that people use to decorate their homes. Thus, Medicinal, Ornamental, Tree, Herb, and most common species like shrubs (besides medicinal and ornamental ones) were given respective ratings of 5, 4, 3, 2, and 1.
Tree_Number	Extremely High (EH), Very High (VH), Semi High (SH), High (F), Medium (M), Low (L), Semi Low (SL), Very Low (VL), Extremely Low (EL)	An ecotourism project cannot be permitted where density of tree species is high. Accordingly, EH is assigned a rating of 1 and EL a 9.
Tree_Type	Extremely Big (EB), Very Big (VB), Semi Big (SB), Big (B), Medium (M), Small (S), Semi Small (SS), Very Small (VS), Extremely Small (ES)	EB tree species generally belong to older populations and no longer contribute to the carbon balance. Thus, forests containing such trees can be easily cleared for ecoproject structures. Again ES species cannot be removed as they will likely serve as carbon sinks in the near future. Therefore, EB is assigned a rating of 9, whereas EL is rated only a 1.
Land_Number	Extremely High (EH), Very High (VH), Semi High (SH), High (F), Medium (M), Low (L), Semi Low (SL), Very Low (VL), Extremely Low (EL)	Even if a food source exists in large quantities, if it does not have the quality ants are looking for, then sooner or later they will reject the path leading to that source, even if it is short. Similarly, even if the concentration of food is high, this factor will be rated poorly as it impacts the objective in a negative manner. The acquisition of a large number of EH lands would cause political controversy. Thus, EH is assigned a rating of 1 and EL a rating of 9.
Land_Type	Extremely Fertile (EF), Very Fertile (VF), Semi Fertile (SF), Fertile (F), Medium (MF) Fertile, Low (LIF) Infertile, Semi Infertile (SIF), Very Infertile (VIF), Extremely Infertile (EIF)	Fertile lands are used for cultivation, and the acquisition of such lands always causes controversies around the world. Thus, in the case of this factor, EF is given a rating of 1 and EIF a score of 9.
LO_Numbers	Extremely High (EH), Very High (VH), Semi High (SH), High (F), Medium (M), Low (L), Semi Low (SL),Very Low (VL), Extremely Low (EL)	This factor is rated similar to Land_Number.
LO_Type	Extremely Big (EB), Very Big (VB), Semi Big (SB), Big (B), Medium (M), Small (S), Semi Small (SS), Very Small (VS), Extremely Small (ES)	This factor is rated similarly to Land Number and LO_Number, where ratings of 1 and 9 are assigned to EB and ES, respectively.

(continued)

Table 1.3 (continued)

Factor	Category	Rating
LO_Hostile/Not	Yes (Y), No (N)	This factor receives a similar rating to C_Hostile/Not as the impact on the objective is the same.
HA_Distance	Extremely Far (EF), Very Far (VF), Semi Far (SF), Far(F), Normal (N), Near (Ne), Semi Near (SNe), Very Near (VNe), Extremely Near (ENe)	Places with hostile activities like terrorist acts or political unrest are generally avoided as tourist destinations. Thus, EF is given a rating of 8 and ENe a 0.
HA_Types	Extremely Sensitive (ES), Very Sensitive (VS), Semi Sensitive (SS), Sensitive (S), Medium (M), Benign (B), Semi Benign (SB), Very Benign (VB), Extremely Benign (EB)	Similarly, ES is given a rating of 0 and EB an 8.
HA_Frequency	Extremely High (EH), Very High (VH), Semi High (SH), High (F), Medium (M), Low (L), Semi Low (SL), Very Low (VL), Extremely Low (EL)	For this factor the EH group is assigned a rating of 0, whereas the EL group is given a rating of 8.The introduction of 0 as a rating is due to the requirement of representing completely avoidable situations that must have no chance of occurring.
P_Distance	Extremely Far (EF), Very Far (VF), Semi Far (SF), Far (F), Normal (N), Near (Ne), Semi Near (SNe), Very Near (VNe), Extremely Near (ENe)	The presence of police increases the security of a place. Thus, the closer a location is to a police station, the safer it will be. Thus, EF is assigned a rating of 1 and ENe a rating of 9.
P_Numbers	Extremely High (EH), Very High (VH), Semi High (SH), High (F), Medium (M), Low (L), Semi Low (SL), Very Low (VL), Extremely Low (EL)	Similar to the earlier factors, EH and EL are respectively assigned ratings of 9 and 1.
P_Type	State Police Force (SPF), Central Police Force (CPF)	A CPF is always better equipped than an SPF. Thus, a CPF gets a 2 and an SPF a 1.

1.3 Conclusion

This chapter presented an approach to select suitable sites for ecotourism projects. The study used the famous ant colony food-search logic to rate the related factors and applied Dorigo's (1992) objective function derived from the logic of ant's food foraging behaviour to determine the probability that a given option would be selected if many alternatives are available. Probability analysis identified the best combination of categories that an option must have for selection as a site for ecotourism projects. In the next step of the study, raw data from various real-life problems were used for selection of sites for eco-projects. The benefits of such a feasibility study are that it is completely free of bias and is objective, similar to an ant colony, where the colony's only objective is to gather food. Ants show no emotional bias in their search for food. Here also personal bias, which is common in selection requirements performed by human beings, is completely avoided.

References

Afshar MH (2010) A parameter free Continuous Ant Colony Optimization Algorithm for the optimal design of storm sewer networks: constrained and unconstrained approach. Adv Eng Softw 41(2):188–195

Dorigo M (1992) Optimization, learning and natural algorithms. PhD thesis, Politecnico di Milan, Italy

Eroğlu Y, Seçkiner SU (2012) Design of wind farm layout using ant colony algorithm. Renew Energy. Corrected Proof, Available online 29 Jan 2012 (in press)

He J, Hou Z (2012) Ant colony algorithm for traffic signal timing optimization. Adv Eng Softw 43(1):14–18

Hu J, Sun JH, Yan JML, Zhen Y, Shi R (2012) Multi-objective ant colony algorithm in EPC risk control. Procedia Eng 29:1767–1773

Macura WK (2012) Ant colony algorithm. From MathWorld – A Wolfram Web Resource, created by Eric W. Weisstein. http://mathworld.wolfram.com/AntColonyAlgorithm.html

Rajendran C, Ziegler H (2004) Ant-colony algorithms for permutation flowshop scheduling to minimize makespan/total flowtime of jobs. European J Oper Res 155(2):426–438

Razavi S, Jalali-Farahani F (2010) Optimization and parameters estimation in petroleum engineering problems using ant colony algorithm. J Petrol Sci Eng 74(3–4):147–153

Wang H, Xu H, Yi S, Shi Z (2011) A tree-growth based ant colony algorithm for QoS multicast routing problem. Expert Syst Appl 38(9):11787–11795

Yang Y, Kamel MS (2006) An aggregated clustering approach using multi-ant colonies algorithms. Pattern Recogn 39(7):1278–1289

Yan-hua Z, Lei F, Zhi Y (2011) Optimization of cloud database route scheduling based on combination of genetic algorithm and ant colony algorithm. Procedia Eng 15:3341–3345

Zecchin AC, Simpson AR, Maier HR, Leonard M, Roberts AJ, Berrisford MJ (2006) Application of two ant colony optimisation algorithms to water distribution system optimization. Math Comput Model 44(5–6):451–468

Chapter 2
Application of Hive Theory to the Identification of Suitable Porcupine Habitats

Mrinmoy Majumder and Rabindra Nath Barman

Abstract Porcupines are rodents with a coat of sharp spines, or quills, that defend or camouflage them from predators. But the same spines or quills have made them valuable to illegal wildlife dealers who sell these quills to various comb and ornament companies who use them in their products. Also the scale of urbanization and climate change have constrained the area of suitable habitats for porcupines. When an animal loses its habitat, its numbers start to diminish, and if the degradation continues, then that species of animal becomes endangered or extinct. That is why the present study tries to find a methodology to identify suitable habitats for porcupine populations using the artificial bee colony algorithm. This algorithm mimics the food-search procedure of the honey bee. In this study the honey bee algorithm is used to rate different input variables or factors. This rating methodology makes the selection process neutral, independent, and free of personal preferences, which are often encountered when scoring is done on the basis of expert opinion in any selection mechanisms.

Keywords Artificial bee colony algorithm • Porcupine • Location selection

M. Majumder (✉)
School of Hydro-Informatics Engineering, National Institute of Technology Agartala, Barjala, Jirania 799055, Tripura, India
e-mail: mmajumder15@gmail.com

R.N. Barman
Department of Production Engineering, National Institute of Technology Agartala, Barjala, Jirania 799055, Tripura, India
e-mail: rahul.barman@yahoo.co.in

M. Majumder and R.N. Barman (eds.), *Application of Nature Based Algorithm in Natural Resource Management*, DOI 10.1007/978-94-007-5152-1_2,

2.1 Introduction

Porcupines are indigenous to the Americas, South Asia, and Africa. The third largest of the rodents (behind the capibara and the beaver), porcupines are approximately 63–91 cm long, with a 20- to 25-cm-long tail. They weigh between 5.4 and 16 kg, which gives them a rounded look and makes them large and slow. Porcupines can be brown, gray, or, uncommonly, white, and their spiny protection makes them related to otherwise unrelated species of hedgehogs, erinaceomorphs, and monotreme echindas.

Porcupines are eaten in the Americas and Southeast Asia, particularly Vietnam, where the widespread use of them as a food source has contributed to significant declines in their populations. Their quills and guard hairs are used for traditional decorative clothing and ornaments and as comb teeth.

The use of the species as food or ornaments has made them vulnerable to poachers and wildlife smugglers. In certain places in South Asia, porcupines are artificially bred for their quills and meat. However, because of rapid-scale urbanization, suitable habitats of the species are constantly diminishing. Their dwindling habitat, along with the ways in which they are used, has made them increasingly vulnerable. The degradation of the species can be prevented if suitable habitats can be identified and the species is conserved in these selected locations. In this way the porcupines can be prevented from becoming endangered.

The problem with any selection procedure is the methodology by which the identified factors are rated or weighted. In most cases, factors are weighted or ranked based on the opinion of experts, dependent populations, or information obtained from the published literatures. But such approaches are subjective in nature and certainly error prone. Experts can be partial, dependent populations can be motivated by selfish needs, and real-life situations may differ from published cases.

That is why it is better to opt for a neutral, scientific, and data-flexible method of assigning weights to particular factors. In nature, many animals make decisions based on their experience and objectives. Their decisions are always neutral and cannot be manipulated in different scenarios. For example, bees will accomplish their search for food based on their goal – a rich and high-quality source of food – and the experience of previous searches or searches conducted by another bee of the same hive. This sequence of decision making seldomly altered even under uncertain or rare conditions. The present study tries to implement the same search method of bees in determining the weightage of importance of each factors on the decision making.

2.1.1 Importance of Porcupine Species to the Ecosystem

Although the common porcupine is generally considered a nuisance, some of the animal's quills are often used by Native Americans to make quill boxes, jewelry, and other artwork. Some Americans also consider the *Erethizon dorsatum*, one of the 29 species of rodent belonging to the families Erethizontidae, suitable for human consumption.

On the other hand, porcupines are considered to be the most "important mammalian forestry pest" (Raysweb, 2012). Whenever and wherever the salt deposits from human precipitation are available, for example, on woodwork, furniture, tools, saddles, and other objects, porcupines will gnaw them. The common porcupine often damages timber and ornamental trees by eating the bark of the trunk or stripping all of the bark above the snowline. They are also responsible for eating orchards and crops. Due to their quills, porcupines can also injure domestic animals and transmit diseases.

2.1.2 Habitat of Porcupines

Porcupines are common in tropical and temperate parts of Asia, Southern Europe, Africa, and North and South America, where they occupy a small range of habitats. Forests, deserts, rocky outcrops, and hillsides landscapes are preferred by this third largest rodent. Many New World porcupines live in trees, but Old World porcupines stay on rocks within an area as high as 3,700 m. This nocturnal animal is sometimes active during the day. Although both New (Family: Erethizontidae) and Old (Family: Hystricidae) World porcupines are most common in hilly, rocky country, they can also adapt to most habitats except excessively moist forests and the most barren of deserts. "They have even been found on Mt. Kilimanjaro, as high up as 11,480 feet" (AWF, 2012).

2.1.3 Food Habits of Porcupines

The common porcupine (*Erethizontidae dorsatum*) is entirely vegetarian, but a major shift in its food habits takes place between the winter and summer months.

In winter porcupines mainly eat evergreen needles and the cambium layer and inner bark of trees. During spring and summer, they eat buds, tender twigs, roots, stems, leaves, flowers, berries, nuts, and other vegetation.

> On occasion, porcupines have been known to eat insects and small lizards (San Diego Zoo, 2013).

2.1.4 Porcupine Predators

Porcupines are mainly hunted by humans, fishers, martens, coyotes, and bald eagles. Fishers are found to be the most successful predators due to their hunting technique of "flipping the porcupine on its back, exposing the unprotected belly" (Raysweb, 2012).

2.1.5 Objective

The main objective of the present study is to identify suitable habitats for porcupines using an artificial bee colony (ABC) algorithm. The algorithm is used to estimate the weightage of importance for the selected factors with respect to the decision mechanism.

2.1.6 Brief Methodology

In this study the factors in the selection of suitable habitats for porcupines are first made based on a literature survey. According to various studies, the following variables have been found to be influential in the selection of hospitable areas for porcupine populations.

1. Type and concentration of tree available in summer
2. Type and concentration of tree available in winter
3. Availability and concentration of forest cover
4. Type of forest in summer
5. Type of forest in winter
6. Presence of water
7. Temperature
8. Total area available
9. Presence and concentration of predators

All the factors are first categorized and then scored using the food-search logic of the honey bee. The factors are categorized with respect to their type and frequency. For example, type of tree available in the summer as a potential food source is categorized into Herbs, Shrubs, Fruit, Tree, and Buds, whereas the concentration of such types of tree is categorized based on its quantity, i.e., Extremely High, Very High, Semi High, High, Medium, Low, Semi Low, Very Low, and Extremely Low.

After the categorization procedure, the methods adopted by the bee population for selection of food sources is replicated to score the factors such that a decision regarding the suitability of a certain area as a porcupine habitat can be undertaken without any bias and independent of other conditions.

2.2 Artificial Bee Colony Algorithm

The ABC algorithm was proposed by Dervis Karaboga in 2005 and mimics the intelligent behavior of honey bees in searching their food. Only two parameters – colony size and maximum cycle number – need to be defined. The ABC algorithm is a population-based search algorithm in which individuals or solutions like food positions are discovered and used by artificial bees over time. Once the source can no longer be modified, the bees abandon it and sets out to discover the next best food source with large amounts of nectar or a fitness function based on the experience gained by themselves and their nest mates. Finally, the location with the most nectar or best fitness function is chosen.

Once the food source is abandoned, artificial bees, also known as scout bees, fly around in a multidimensional search space and choose food sources randomly without relying on their experience. If the amount of nectar in the new source is higher than that of the previous one in their memory, they memorize the new position and forget the previous one. That is why the ABC system combines both local and global search methods performed by "worker and onlooker bees" and "onlookers and scouts," respectively, balancing exploration and exploitation processes simultaneously.

In short, the ABC algorithm can be explained as:

1. The foraging behavior of honeybees;
2. A global optimization algorithm;
3. A tool initially proposed for numerical optimization;
4. A tool that can be used in combinatorial optimization problems;
5. A tool that can be used for unconstrained and constrained optimization problems;
6. An algorithm whereby only three parameters are required to be entered by the user (population size, maximum cycle number, and limit). Applications of ABC in various practical decision making or optimization or any other many other model making methodology is discussed in Table 2.1.

2.3 Methodology

In the present study the bee colony's food-search methodology is applied in a completely different way than how Karaboga et al. applied it.

In this case the food sources are compared with that region that could become a habitat for porcupines. The factors are compared with subsources of food. It can be thought of as a flowering with lots of flowers. All the flowers are a potential food source. If the nectar of the flowers attracts more bees than any other plant, then that plant is selected as the most suitable place for a porcupine population.

According to the logic of food searching by bees, the worker bees at first randomly select a flower or a factor. Every factor will either encourage or discourage the selection of the place for the porcupines, and the degree of its quantity or quality will determine how intensely it encourages or discourages the selection. If the degree of encouragement or discouragement is compared with the intensity of the nectar, then a source of large amounts of nectar will always be rated by a worker bee higher than a source of low amounts of nectar. The dance of the bee will also indicate the same, and the probability of selecting the factor as a potential food source by the onlooker becomes greater.

That is why, each of the variables are ranked in such a way that the scoring becomes coherrent objective of the decision represented by a rank from ascending to descending manner representing the importance of the variables in taking decision to the current problem. As discussed earlier in ABC goal is to select the alternative with minimum rank. The rank is utilized as a fitness function which help the algorithm to take the correct path both in quality and quantity in adopting the optimal decision out of the many possible alternatives.

Table 2.2 shows the rank of each category of factors. Ranking is done in such a way that the level of importance is set to be directly proportional to the amount of nectar. If the encouragement level is more then nectar is more, also the ranking will be high.

Table 2.1 Recent literature on the application of the artificial bee colony algorithm in different problem-solving and solution-searching objectives

Author	Problem/objective of study	Goal of ABC	Success rate
Fahmy (2012)	Optimize annual power yield and usage time of wind turbine	Selection of optimal speed parameters, i.e., optimization of rated, cut-in, and cut-out speed of wind turbines, which will in turn optimize the power yield and turbine usage time	Compared to PSO and manual optimization of rated speed, ABC found to be superior according to the used benchmarks
Zhang et al. (2012)	Burdening optimization of copper strip production	Minimization of total cost of raw material and maximization of waste materials that can be deposited in furnace	Results found to be more accurate than nondominated sorting genetic algorithm II (NSGAII) and multiobjective particle swarm optimization (MOPSO)
Hsieh et al. (2011)	Stock price forecasting	Optimization of weights of recurrent neural network used to predict stock prices	Results found to be accurate and reliable compared with existing indexes
Karaboga and Ozturk (2011)	Multivariate data clustering	Data clustering of benchmark problems	Compared to particle swarm optimization and nine other classification algorithms, ABC was found to be reliable and accurate enough to search for threshold values of clusters
Taheri et al. (2011)	Minimization of make-span and total data transfer time by Job Scheduling at Computational and Storage nodes of heterogeneous system	Job Scheduling for Computational and Storage nodes.	Job Data Scheduling using bee colony algorithm is better than other similar algorithms
Quijano and Passino (2010)	"Solve a dynamic voltage allocation problem to achieve a maximum uniformly elevated temperature in an interconnected grid of temperature zones"	Dynamic resource allocation problem employing many versions of ABC to solve a single problem	Solution received from ABC found to be accurate
Xu and Duan (2010)	Edge detection/shape matching by aircraft at low altitude	ABC was used in conjunction with Edge Potential Function, which can create a contour of the matching edges and use ABC to create a pattern	Found to be more accurate than genetic algorithms

Table 2.2 Factors, their categories, and the rank assigned

Factors	Category	Rank
Type and concentration of tree available in summer	Herbs	1
	Shrubs	2
	Fruits	3
	Leaves	4
	Buds	5
Concentration of trees (S)	EH	1
	VH	2
	SH	3
	H	4
	M	5
	L	6
	SL	7
	VL	8
	EL	9
Type of tree (winter)	Coniferous	2
	Evergreen	1
Concentration of trees (S)	EH	1
	VH	2
	SH	3
	H	4
	M	5
	L	6
	SL	7
	VL	8
	EL	9
Availability of cover	Logs	1
	Rock	2
	Cave	3
Concentration of cover	EH	1
	VH	2
	SH	3
	H	4
	M	5
	L	6
	SL	7
	VL	8
	EL	9
Cover for reproduction	Burrows	1
	Logs	2
	Foliage	3
	Caves	4
	Rocks	5
	Cliffs	6

(continued)

Table 2.2 (continued)

Factors	Category	Rank
Concentration of cover	EH	1
	VH	2
	SH	3
	H	4
	M	5
	L	6
	SL	7
	VL	8
	EL	9
Presence of water	EH	1
	VH	2
	SH	3
	H	4
	M	5
	L	6
	SL	7
	VL	8
	EL	9
Type of forest (summer)	Meadow	1
	Brushy	2
	Riparian	3
Type of forest (winter)	Coniferous	1
	Evergreen	2
	Mixed	3
Area available	EH	1
	VH	2
	SH	3
	H	4
	M	5
	L	6
	SL	7
	VL	8
	EL	9
Temperature	EH	1
	VH	2
	SH	3
	H	4
	M	5
	L	6
	SL	7
	VL	8
	EL	9
Presence of predatory animals	Fishers	1
	Lions	2
	Bobcats	3
	Wolverines	4

(continued)

Table 2.2 (continued)

Factors	Category	Rank
Concentration of predatory animals	EL	1
	VL	2
	SL	3
	L	4
	M	5
	H	6
	SH	7
	VH	8
	EH	9

After the probability of each of the factors is determined, the value is reversed. The average of the probabilities will determine whether the plant or an alternative can be selected by the bees as a potential source of food. Once the selection is over, the bees that are free from their responsibilities will start a new search for potential food sources or "food plants," and if they find a plant with more nectar than the previous one, they repeat the entire procedure, and if the average probability increases over the earlier probability, then the latter is selected over the former.

This selection of one source over another will be made by comparing the average probability of each site with that of the other; if there are many sites, then the probabilities must be ranked in ascending order, and the highest ranking plant will be the most sought after source or the most suitable site as a porcupine habitat.

In summary, the steps of the ABC food-searching algorithm as applied for selection to sites for porcupines are listed below:

Initialization

1. The factors are compared to flowers and the sites available are compared with flowering plants.

Worker Bees

2. Factors are ranked according to their degree of importance representing about the selectivity of the site for habitat identification of the porcupines. The greater the amount of nectar, the higher will be the ranking.

Onlooker Bees

3. The probability of the factors is also determined using the assigned rank. This is similar to the selection process made by onlooker bees in selecting a food source after the source has been recommended by the dance of the worker bees.
4. As the ranking is made according to suitability, a higher rank will make the option more suitable but its probability will be reduced. That is why all the probabilities are inverted and averaged to determine the overall selectability of the option for porcupine habitation.

Table 2.3 Factors, their categories, and rank assigned to sample study area

Factors	Category	Rank	Inverse probability
Type and concentration of trees available in summer	Buds	5	0.16
Concentration of tree (S)	SH	3	0.70
Type of tree (Winter)	Coniferous	2	0.33
Concentration of tree (W)	H	4	0.60
Types of cover	Logs	1	0.80
Concentration of cover	H	4	0.60
Cover for reproduction	Burrows	1	0.86
Concentration of cover	EH	1	0.90
Presence of water	SH	3	0.70
Type of forest (Summer)	Brushy	2	0.50
Type of forest (Winter)	Coniferous	1	0.75
Area available	SH	3	0.70
Temperature	SH	3	0.70
Presence of predatory animals	Bobcats	3	0.5
Concentration of predatory animals	L	4	0.20
Site suitability (average probability)	**H**	**4**	**0.60**

Scout Bees

5. Once the average inverse probability is determined, the free bees become scouts and start to search for another food source. Another option with its factors is ranked in the same manner as the earlier option. If the average inverse probability becomes more than the earlier one, then this option is selected as the food source and the bees forget about the earlier source. As various bee colony algorithm can be applied under a single problem domain, bee hives which actually represent various bee colony algorithms can cumulatively be referred as hive theory algorithms. (Quissano and Passino 2010).

The logic of this new methodology was applied to a site where porcupines are already present. The site was analyzed for its suitability as a habitat of porcupine populations.

2.4 Results and Discussion

The categories of selected factors with respect to the study area are listed in Table 2.3.

The study area where the porcupines are already living was rated using the ABC algorithm; according to the results, it does contain a large population of porcupines, i.e., it has a high probability of selection as a site for a porcupine population.

The study also showed the justification of applying this methodology, which is sufficiently accurate, as demonstrated by the study results presented in Table 2.3.

2.5 Conclusion

In the present study a new type of bee colony algorithm was proposed; it was used to identify a suitable location for porcupine habitats. These animals are commonly considered pests and are often eaten by predators; as a result, their numbers in the world are decreasing day by day. That is why, the study was conducted to identify an ideal porcupine habitats so that porcupines can be relocated there for reproduction and other activities without fearing about the uncertainties of life as well as food locations. A simple combinatorial data matrix was created that represents all possible combinations of the categories of the 15 input factor and 1 decision output. This data matrix was scored according to the bee colony's foraging behavior, and ultimately a suitable location was selected from the study area based on the ranks achieved by the available options. Although a new kind of ABC algorithm was used to solve the problem considered in this chapter, it was not compared with other algorithms due to the lack of time. In the future, many other statistical and biological algorithms can be compared with the ABC algorithm to find the capability of the same in taking decision.

References

African Wildlife Foundations (AWF), 2012, Porcupines. Retrieved from http://www.awf.org/content/wildlife/detail/porcupine. Feb 2012

Fahmy AA (2012) Using the Bees Algorithm to select the optimal speed parameters for wind turbine generators. J King Saud Univ Comput Inf Sci 24(1):17–26

Hsieh T-J, Hsiao HF, Wei-Chang Y (2011) Forecasting stock markets using wavelet transforms and recurrent neural networks: an integrated system based on artificial bee colony algorithm. Appl Soft Comput 11(2):2510–2525

Karaboga D, Ozturk C (2011) A novel clustering approach: Artificial Bee Colony (ABC) algorithm. Appl Soft Comput 11(1):652–657

Quijano N, Passino KM (2010) Honey bee social foraging algorithms for resource allocation: theory and application. Eng Appl Artif Intel 23(6):845–861

Raysweb (2012) Porcupine. Retrieved from http://raysweb.net/wildlife/pages/12.html. Feb 2012

San Diego Zoo Porcupines (2013) Mammals: Porcupines, Zoological Society of San Diego. Retrieved from http://www.sandiegozoo.org/animalbytes/t-porcupine.html. Feb 2012

Taheri J, Lee YC, Zomaya AY, Siegel HJ (2011) A Bee Colony based optimization approach for simultaneous job scheduling and data replication in grid environments. Comput Oper Res (in press), corrected proof, Available online 25 Nov 2011

Xu C, Duan H (2010) Artificial bee colony (ABC) optimized edge potential function (EPF) approach to target recognition for low-altitude aircraft. Pattern Recogn Lett 31(13):1759–1772

Zhang H, Zhu Y, Zou W, Yan X (2012) A hybrid multi-objective artificial bee colony algorithm for burdening optimization of copper strip production. Appl Math Model 36(6):2578–2591

Chapter 3
Tradeoff Analysis Between Rainfall and Load Factor of a Small-Scale Hydropower Plant by Particle Swarm Optimization

Mrinmoy Majumder, Soumya Ghosh, and Rabindra Nath Barman

Abstract Hydropower is claimed to be one of the least expensive but most reliable sources of renewable energy. The frequency of power generation depends directly on the flow of water on which the power production facility has been constructed. The flow of water depends on the upstream rainfall, which contributes to the surface runoff to create the flow in the channel which rotates the turbine for production of electricity. The utilization factor of a hydropower plant (HPP) is defined as the ratio between the energy actually produced to the energy production capacity of the hydropower plant (HPP). It is synonymous with load factor if the capacity of the HPP and the maximum energy produced become equal. The present study will aim to identify the optimal zones where minimum rainfall and maximum utilization can be achieved by employing particle swarm optimization within the known constraints of small scale hydropower plant. The result of the study will highlight the adjustments required to be followed in the hydropower plants in generating optimal power output even in the days of scarce rainfall.

Keywords Small-scale hydropower • Particle swarm optimization • Tradeoff

---

M. Majumder (✉)
School of Hydro-Informatics Engineering, National Institute of Technology Agartala, Barjala, Jirania 799055, Tripura, India
e-mail: mmajumder15@gmail.com

R.N. Barman
Department of Production Engineering, National Institute of Technology Agartala, Barjala, Jirania 799055, Tripura, India
e-mail: rahul.barman@yahoo.co.in

S. Ghosh
School of Hydro-Informatics Engineering, National Institute of Technology Agartala, Barjala, Jirania 799055, Tripura, India
e-mail: soumyaee@gmail.com

M. Majumder and R.N. Barman (eds.), *Application of Nature Based Algorithm in Natural Resource Management*, DOI 10.1007/978-94-007-5152-1_3,

3.1 Introduction

The utilization factor is the ratio of the maximum load on a power plant to the rated plant capacity, whereas (electrical) load factor can be defined as the average power divided by the peak power over a period of time. If the peak power is equal to the rated plant capacity, then the utilization factor and load factor will be equal.

In the case of a hydropower plant (HPP), the utilization factor depends mainly on the amount of upstream and local rainfall which is the driving force of any hydropower plant. The rated plant capacity will depend on the maximum possible flow that was observed in the flow duration curve of the river on which the power plant was developed. The maximum load on the other hand will depend upon the demand of power from the plant.

The demand for power depends on various factors where change in rainfall is one of the main parameters for which a change in production capacity is observed. During summer, the demand for electricity is high because of the continuous requirement for fans, air-conditioners, and other cooling devices. But during winter, the demand for power is reduced because requirements for electricity diminish. The only major requirement in terms of energy use during this time of year is for running the air-heating devices.

In the case of HPPs, the generation of power also varies with regional climate patterns because the power production capacity of any HPP depends mainly on the velocity of flow, which again is a function of rainfall and head difference. And as demand for power varies with climate also, the average load will vary with change in weather pattern.

That is why, in summary, it can be stated that both the load factor and utilization factor are functions of rainfall. If power demand and rainfall both changes, then the zone in which the load factor reaches its maximum will also change. This study aims to identify the status of the power demand and rainfall when the load factor will become nearly equal to utilization factor. The study used the inherrent searching capability of particle swarm optimization (PSO) to accomplish this objective. There are various types of hydropower plant and this zone where utilization and load factor will become equal also varies with the type of hydropower plant. In the present investigation te small scale hydropower plants are only considered. The next section explains the classification of HPP.

3.1.1 Classification of Hydropower Plants

Based on water head, HPPs can be classified into three types: low-head, medium-head, and high-head HPPs.

3.1.1.1 Low-Head Hydroelectric Power Plants

Low-head HPPs can be defined as power plants where the available water head is less than 30 m. Most of the time, due to the very short head of such power plants, dams are not constructed; instead, a weir is used and the inherent flow of water in the canal/river is used to generate electricity. The low-head types of HPPs are of the nonstorage type and generate electricity only when sufficient flow of water is available. Thus, their relationship to rainfall is directly proportional, i.e., they can produce electricity only during particular seasons like the monsoon season, when abundant flow of water is available. Because the available head of water is directly proportional to electricity production capacity of any HPP, the power-producing capacity or utilization factor of such plants is very low due to lack of reliability and stability in the availability of kinetic energy in the canal/river.

3.1.1.2 Medium-Head Hydroelectric Power Plants

HPPs having a working head of water more than 30 m but less than 300 m are referred to as medium-head HPPs. These HPPs have their own storage systems and are usually developed in mountainous or hilly regions so that the advantage of the height difference can be used for power generation.

3.1.1.3 High-Head Hydroelectric Power Plants

High-head HPPs have a head of water varying from 300 to 1,000 m. Large reservoirs of water in dams that can store water at very high heads are developed, and turbines are connected through penstock so that the water from the dam can be used to rotate it and produce huge amount of power from the generators. Water is mainly stored during the rainy season and can be used in the lean season. Thus, high-head HPPs can generate electricity throughout the year. The total height of the dam will be a function of a number of factors like quantity of available water, power to be generated, surrounding area, natural ecosystem, etc.

In the present study the optimal points were identified only for the low-head HPP that varies directly with rainfall.

3.1.2 Impact of Climate Change on Hydropower Plants

The production of greenhouse gases from the domestic and industrial sectors has increased the average temperature of the planet. Due to this global warming phenomenon rainfall patterns have apparently changed in many places around the world. As rainfall is directly related to the generation of hydropower, the effect of warming will hit these plants as well. Thus, it has been asserted that the relationship between the load factor and rainfall will become a vital concern for the engineers in designing the HPPs.

An increase in temperature will also increase the average load. Thus, the load factor will also change by the change in the climate.

3.1.3 Objective and Scope

The objective of the present study is to determine the tradeoff zones between rainfall, power demand, and load factor. The optimal zones, i.e., minimum rainfall but maximum utilization, will be identified using particle swarm optimization.

These zones can inform engineers about the optimal values of independent variables and highlight the feasibility of a project using the number and magnitude of tradeoff zones.

3.1.4 Brief Methodology

The rainfall data of a low-head river flowing through the northeastern states of India are first collected on a per-month basis. The power demand per month from the adjacent areas of the river is estimated from governmental sources. The discharge of water through the river is estimated from the water balance equation, and the power equation is used for estimation of power production.

After rainfall, demand, and power production are derived, tradeoff zones was generated and identified by the utilization of PSO.

3.2 Particle Swarm Optimization

PSO (Eberhart and Kenedy, 1995) is a population-based search algorithm that follows the behavior of flocks of birds or schools of fish. In this algorithm, random particles are considered as the solution to a given problem space. Each of the particles is a probable solution to the given problems. It gradually converges to the optimal solution based on two criteria:

1. Local best
2. Global best

The iteration starts by selecting random positions for the particles. Each particle has an objective to converge toward the optimal solution in the search domain. The velocity of the particle is updated in the following manner:

$$x_{t+1} = x_t + V_{t+1} \tag{3.1}$$

where

$$V_{t+1} = V_t + c_1 x\text{rand}\left(0,1\right)\times(V_{\text{lb}} - v_t) + c_2 x\text{rand}\left(0,1\right)\times\left(V_{\text{gb}} - V_t\right) \tag{3.2}$$

V_{t+1} = new velocity of the particle as it converges toward the optimal solution; V_t = old velocity acquired by the particle in the previous iteration; x_{t+1} = new position; x_t = old position attained after the last iteration; c_1 and c_2 = learning factors and random functions that generate random numbers between 0 and 1; V_{lb} = velocity attained for optimal value of fitness function when compared within the old and new positions; and V_{gb} = velocity obtained for the best fitness function achieved until the present iteration. The previous velocity can be updated at each iteration based on the value of the fitness function attained after each iteration, but the value of the latter is generally updated once a better output of the fitness function is achieved.

PSO is similar to many evolutionary computation techniques such as genetic algorithms (GAs). But unlike GAs, the efficiency of PSO does not depend on evolutionary parameters like crossover and mutations.

The simplicity in PSO's application and the need to adjust just a few parameters have made PSO a sought-after iteration technique for optimization problems. PSO has been successfully applied in many areas: engineering design (Feng et al. 2010), multiobjective optimization (Mousa et al. 2012), multiobjective planning (Sahoo et al. 2012), artificial neural network training (Chau 2006) and topology selection of neural networks (Mingo et al. 2012), fuzzy systems (Zhao et al. 2010), parameter estimation (Wang et al. 2011), parameter selection (Parsopoulos and Vrahatis 2007), data clustering (Tsai and Kao 2011), and solving high-dimensional problems (Jia et al. 2011).

Many new variants of PSO are also being developed to improve the accuracy and reliability of the algorithm including discrete PSO, constricted constraints, barebones PSO (Zhang et al. 2012), pooled-neighbor swarm optimization (Guo and Zhao 2006), chaotic multihybrid (Mukhopadhyay and Banerjee 2012), hybrid PSO (Shelokar et al. 2007) and chaotic PSO (Khajehzadeh et al. 2011), particle visual modeling analysis considering the degree of particle distribution and dimensional distance (Zhao et al. 2009), grammatical PSO (Lopez et al. 2012), multiswarm cooperative particle swarm optimizer (Zhang et al. 2011), perturbed PSO (involving linear algorithms for position updates to maintain the diversity of the generated data and prevent premature convergence) (Xinchao 2010), and fully informed PSO (Mendes et al. 2004).

To date, there are nearly 7,209 articles and approximately 3,963 book chapters published in various reputable international journals and books about different applications of PSO in practical problem solving.

3.3 Necessity of Hydropower Plans

Power is one of the most essential inputs for sustaining the economic development of a country, but it also invites degradation of the environment and increases in greenhouse gases, which are the main culprit in climate change. "The inevitable increase in the use of fossil fuels to keep pace with the economic growth has associated side effects of threat to energy security of the country and environmental degradation through climate change" (Anonymous). The ever-growing world population and economic development have put pressure on existing resources for power generation. According to the latest reports, the size of the world economy will increase at a rate of 3–5 times by the year 2050 and by 10–15 times by the year 2100 with respect to the present economy. In contrast, the energy requirements of the world will increase 1.5-to 3-fold by 2050 and 2-to 5-fold by 2100.

At present the primary energy consumption of the world is dominated by fossil fuels like oil (36%), natural gas (21%), and coal (24%). Biomass (9%), nuclear fuels (6%), and large hydro and other renewable energy sources (2%) complete the global power consumption scenario. The dependency on fossil fuels has caused several detrimental effects on the environment and ecological balance. Due to the combustion of such finite sources of energy greenhouse gases like sulfur, nitrogen oxides, carbon monoxide, and suspended particulate matter are abundantly present in the atmosphere causing global warming and concomitant climate change. Ozone layer depletion, land degradation, air and water pollution, sea level rise, and loss of biodiversity are other negative impacts of using fossil fuels.

The global consumption of primary energy is increasing at a rate of 2% per year (68 J/capita/year and 1.6 tonnes of oil equivalent/capita). The three primary energy sources are found to have 0.4, 2.3, 1.5, and 0.9% ratio of supply to reserve respectively for coal, oil, natural gas, and total energy sources. Although a minor part of the available reserves are used for consumption, still not all reserves have a 100% utilization factor (Nakicenovic 2012 and HDR 1998).

3.3.1 Global Scenario of Renewable Energy

Energy that can be naturally replenished is referred as renewable energy. The major sources of such energy are sunlight, wind, rain, tides, and geothermal heat. Only 16% of global final energy consumption comes from renewable sources; this figure was only 13% in 1998. The major share of renewable energy is biomass (10%), followed by 3.4% from hydroelectricity. New renewable (small hydro, modern biomass, wind, solar, geothermal, and biofuels) contributes another 2.8% (REN21 2011a). "The share of renewables in electricity generation is around 19%, with 16% of global electricity coming from hydroelectricity and 3% from new renewables" (REN21 2011a).

The contribution of wind power to global renewable energy generation is 30% annually (REN 2011), and the installed capacity of photovoltaic energy is more than 40 GW (REN 2011). Brazil has the largest renewable energy program in the world, where energy from ethanol provides 19% of the total energy requirements of the country.

Renewable energy sources are environmentally friendly and do not produce greenhouse gases, and, although renewable energy is infinite, the cost of conversion and the uncertainty in its availability earlier discouraged the governments of different countries from opting for such energy sources to satisfy their energy needs. But due to the rapid pace of economic development, the growth in energy demands to sustain this trend, and the damaging impacts of fossil fuel, many countries have opted for renewable energy. According to the latest reports, small solar PV systems provide electricity to several million households, and micro-hydro configured into mini-grids serves many more. Over 44 million households use biogas made in household-scale digesters for lighting or cooking, and more than 166 million households rely on a new generation of more-efficient biomass cook stoves (REN21 2011b).

In the case of hydropower, at least 50% of the electricity production in 66 countries and at least 90% in 24 countries is supplied from the energy produced from HPPs.

3.3.2 Classification of Hydropower Plants

HPPs are generally classified based on quantity of water, water head, and nature of load.

3.3.2.1 Classification with Respect to Quantity of Water

HPPs can be classified based on the amount of water used in the following way:

Runoff River Plants Without Pondage

These kinds of plants do are unable to store water and use water as and when available. That is why such plants are dependent on the rate of flow of water; during the rainy season, a high flow rate may mean that some water is wasted, whereas during low run-off periods, due to low flow rates, the generating capacity will be low.

Runoff River Plants with Pondage

In these plants pondage permits storage of water during off-peak periods and use of this water during peak periods. Depending on the size of pondage provided, it may be possible to cope with hour-to-hour fluctuations. This type of plant can be used on parts of the load curve as required and is more useful than a plant without storage or pondage.

This type of plant is comparatively more reliable, and its generating capacity is less dependent on the available rate of water flow.

Reservoir Plants

A reservoir plant is one that has a reservoir of such a size as to permit carrying over storage from the wet season to the next dry season. Water is stored behind a dam and is available to the plant with control, as required. The plant firm capacity can be increased and can be used either as a base load plant or as a peak load plant as required. The majority of hydroelectric plants are of this type.

3.3.2.2 Classification by Availability of Water Head

Based on the availability of the water head, an HPP can be subdivided into Low or Small head (less than 30 m) (SSHP), medium-head (30–300 m), and high-head hydroelectric plants (1,000 m and above). Low-head HPPs can be further subdivided into small-, mini-, and micro-head HPPs.

3.3.2.3 Classification with Respect to Nature of Load

Classification according to the nature of load is as follows:

Base load plants: a base load power plant is one that provides a steady flow of power regardless of total power demand by the grid. These plants run at all times throughout the year except in the case of repairs or scheduled maintenance.

Peak load plants: these are power plants for electricity generation that, due to their operational and economic properties, are used to cover peak loads. Gas turbines and storage and pumped storage power plants are used as peak load power plants. The efficiency of such plants is approximately 60–70%.

The present investigation selected a small-scale HPP for optimization of load factor using PSO algorithms.

3.4 Methodology

The main objective of the present investigation is to optimize the load factor in such a manner that it attains a value of 1 or close to 1 because at that value the average load and peak demand become equal, which means the power plant can function at an efficiency of nearly 100%.

The variables for the present study are the amount of rainfall and demand for power. The constraints of the study were as follows:

$$\text{Precipitation in the Catchment Area of the SSHP}\left(\text{Rain}\right) < 2{,}500\ \text{mm}$$
$$\text{Demand for Power}\ (D) < 20\ \text{MW}$$

The objective function is taken as the resulting value of the subtraction of the load factor from one. Minimization of this resulting value will be the optimal output possible from the Small Scale Hydropower Plant (SSHP) of the present study.

PSO is used to generate different values of rainfall and demand for power in a restricted search domain where the upper limit is determined by the constraints imposed on the variables.

C_1 and C_2, the scale factors, are taken as two, and the fitness function is the objective function itself. If its value is lower than for the previous iteration, then the local best is updated; if the value is the lowest of all iterations up to that point, then the global best is updated.

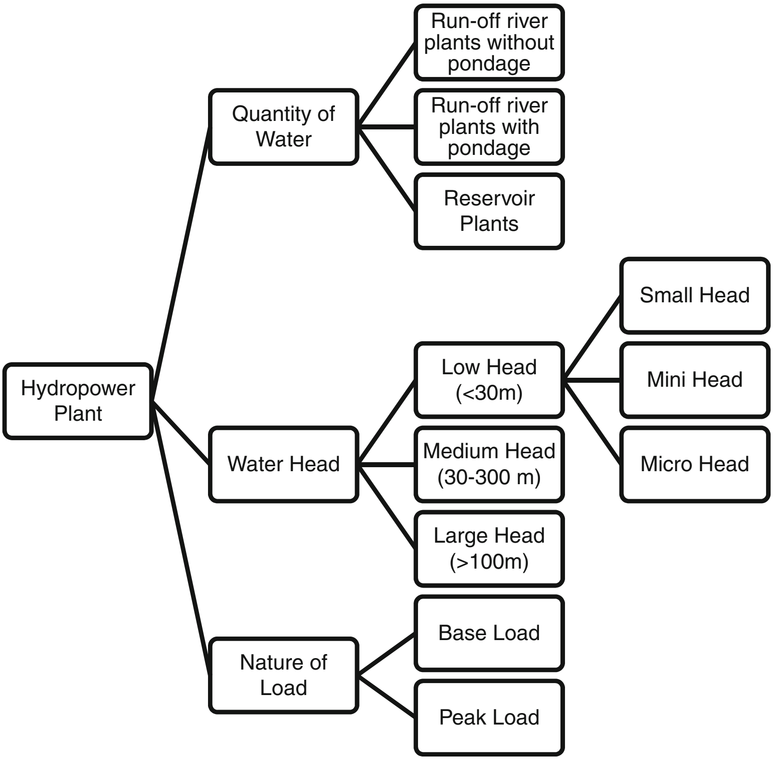


Fig. 3.1 Classification of HPP with respect to different attributes

In this manner 300,000 iterations were performed and the 30 most optimal results were selected from among them. Figure 3.1 depicts the 30 most optimal results of the search and the corresponding normalized values of the variables and objective function L, which was actually derived from

$$L = 1 - (D / P). \tag{3.3}$$

When the difference between the load factor (Eq. 3.3) and unity is at a minimum, the SSHP will operate optimally at that reference point and those conditions of the variables.

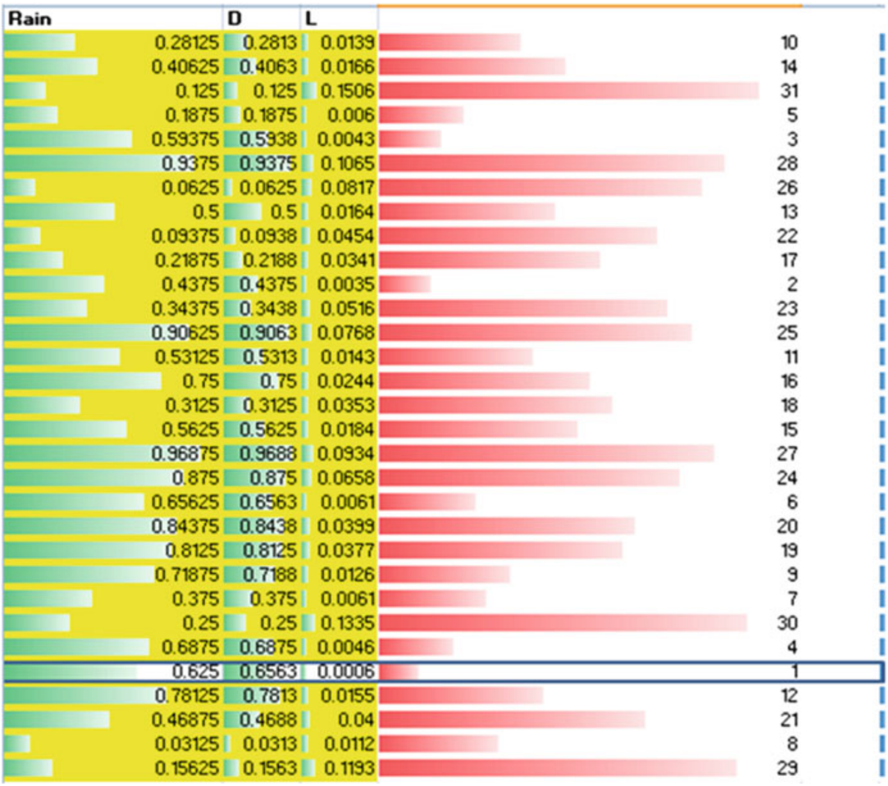
To generate the different values for the selected variables of the present optimization study, the positions generated by the PSO are multiplied by the upper limit of the problem domain so that the regenerated swarm of solutions stays inside the required boundaries of the problem space.

The amount of power that can be generated is derived from the water balance and power equation. Demand values are selected with respect to the common magnitude of demand faced by any SSHP in the world, and rainfall limits are derived from the data of rainfall patterns in tropical countries.

3.5 Results and Discussion

PSO was used to adjust the value of the variables within the given upper limit. The value of the rainfall and demand for power was varied 300,000 times, and each time a new load factor was generated. The load factor was compared with unity and the lower the value of the difference, the more optimal the selection. From the 300,000 load factors 100 values per generation were collected. The values were then ranked and the minimum value obtained (Eq. 3.3) was identified and saved to a data matrix. Similarly, for 30 separate iterations, 100 values were collected and ranked in ascending order. The top-ranked objective function was identified and, along with its variable, saved to the same data matrix. The data matrix was normalized and ranked to identify the minimum. The output of this procedure is shown in Fig. 3.2.

For 60%-plus cases the convergence of PSO toward the optimal solution was observed before 50,000 iterations, but for the remaining case the optimal solution



Rain	D	L	
0.28125	0.2813	0.0139	10
0.40625	0.4063	0.0166	14
0.125	0.125	0.1506	31
0.1875	0.1875	0.006	5
0.59375	0.5938	0.0043	3
0.9375	0.9375	0.1065	28
0.0625	0.0625	0.0817	26
0.5	0.5	0.0164	13
0.09375	0.0938	0.0454	22
0.21875	0.2188	0.0341	17
0.4375	0.4375	0.0035	2
0.34375	0.3438	0.0516	23
0.90625	0.9063	0.0768	25
0.53125	0.5313	0.0143	11
0.75	0.75	0.0244	16
0.3125	0.3125	0.0353	18
0.5625	0.5625	0.0184	15
0.96875	0.9688	0.0934	27
0.875	0.875	0.0658	24
0.65625	0.6563	0.0061	6
0.84375	0.8438	0.0399	20
0.8125	0.8125	0.0377	19
0.71875	0.7188	0.0126	9
0.375	0.375	0.0061	7
0.25	0.25	0.1335	30
0.6875	0.6875	0.0046	4
0.625	0.6563	0.0006	1
0.78125	0.7813	0.0155	12
0.46875	0.4688	0.04	21
0.03125	0.0313	0.0112	8
0.15625	0.1563	0.1193	29

Fig. 3.2 Thirty most optimal outputs from 30,000 generations of solutions using PSO (*rain*: rainfall percentile, *D*: demand for power, percentile; *L*: objective function deducting one from load factor; *red columns*: rank of objective function)

was identified after 50,000 iterations and for 10% of the cases it was observed after 90,000 iterations, although the best solution among all the iterations was observed only at the 270,000th iteration.

PSO is known to be a quick convergent, but the present investigation shows that the best solution is obtained only after 90% of considered iterations were completed.

3.6 Conclusion

The present investigation tried to optimize the load factor of a small-scale HPP (10 m head). A PSO algorithm was used to generate data for the variables considering the upper limit imposed on those variables. After the iteration procedure it was found that a rainfall in the 62.5th percentile and a demand in the 65.65th percentile are required for the SSHP to perform at optimal load factor. The study results can be used in the planning of an SSHP during the feasibility analysis phase. Although PSO is considered a fast convergent, according to the present study, it takes more than 90% model time to identify the optimal solution. Determining the reasons for such results is left to future researchers.

References

Chau KW (2006) Particle swarm optimization training algorithm for ANNs in stage prediction of Shing Mun River. J Hydrol 329(3–4):363–367

Eberhart RC, Kennedy J (1995) A new optimizer using particle swarm theory. In: Proceedings of the 6th international symposium on micro machine and human science, 4–6 October 1995, Nagoya, Japan, pp 39–43

Feng Y, Zheng B, Li Z (2010) Exploratory study of sorting particle swarm optimizer for multi-objective design optimization. Math Comput Model 52(11–12):1966–1975

Guo CX, Zhao B (2006) A pooled-neighbor swarm intelligence approach to optimal reactive power dispatch. J Zhejiang Uni Sci A 7(4):615–622

HDR (1998) Consumption for human development. Retrieved from http://hdr.undp.org/en/reports/global/hdr1998/chapters/

Jia DL, Zheng GX, Qu BY, Khan MK (2011) A hybrid particle swarm optimization algorithm for high-dimensional problems. Comput Ind Eng 61(4):1117–1122

Khajehzadeh M, Taha MR, El-Shafie A (2011) Reliability analysis of earth slopes using hybrid chaotic particle swarm optimization. J Cent South Univ Technol 18(5):1626–1637

López LFM, Blas NG, Arteta A (2012) The optimal combination: grammatical swarm, particle swarm optimization and neural networks. J Comput Sci 3(1–2):46–55

Mendes R, Kennedy J, Neves J (2004) The fully informed particle swarm: simpler, maybe better. IEEE Trans Evolut Comput 8(3):204–210

Mingo LFL, Blas NG, Arteta A (2012) The optimal combination: grammatical swarm, particle swarm optimization and neural networks. J Comput Sci 3(1–2):46–55

Mousa AA, El-Shorbagy MA, Abd-El-Wahed WF (2012) Local search based hybrid particle swarm optimization algorithm for multiobjective optimization. Swarm Evolut Comput 3:1–14

Mukhopadhyay S, Banerjee S (2012) Global optimization of an optical chaotic system by Chaotic Multi Swarm Particle Swarm Optimization. Expert Syst Appl 39(1):917–924

Nakicenovic (2012) World Energy Assessment Report. Retrieved from http://webarchive.iiasa.ac.at/Research/TNT/WEB/Publications/The_World_Energy_Assessment_Re/the_world_energy_assessment_re.html

Parsopoulos KE, Vrahatis MN (2007) Parameter selection and adaptation in Unified Particle Swarm Optimization. Math Comput Model 46(1–2):198–213

REN21 (2011a) Renewables 2011: global status report. Retrieved from http://www.ren21.net/Portals/97/documents/GSR/GSR2011_Master18.pdf

REN21 (2011b) Renewables 2011: global status report, p 18. Retrieved from http://www.ren21.net/Portals/97/documents/GSR/GSR2011_Master18.pdf

Sahoo NC, Ganguly S, Das D (2012) Multi-objective planning of electrical distribution systems incorporating sectionalizing switches and tie-lines using particle swarm optimization. Swarm Evolut Comput 3:15–32

Shelokar PS, Siarry P, Jayaraman VK, Kulkarni BD (2007) Particle swarm and ant colony algorithms hybridized for improved continuous optimization. Appl Math Comput 188(1):129–142

Tsai CY, Kao IW (2011) Particle swarm optimization with selective particle regeneration for data clustering. Expert Syst Appl 38(6):6565–6576

Wang J, Zhu S, Zhao W, Zhu W (2011) Optimal parameters estimation and input subset for grey model based on chaotic particle swarm optimization algorithm. Expert Syst Appl 38(7):8151–8158

Xinchao Z (2010) A perturbed particle swarm algorithm for numerical optimization. Appl Soft Comput 10(1):119–124

Zhang Y, Gong DW, Ding ZH (2011) Handling multi-objective optimization problems with a multi-swarm cooperative particle swarm optimizer. Expert Syst Appl 38(11):13933–13941

Zhang Y, Gong DW, Ding Z (2012) A Bare-bones multi-objective particle swarm optimization algorithm for environmental/economic dispatch. Inform Sci 192:213–227

Zhao Y, Zu W, Zeng H (2009) A modified particle swarm optimization via particle visual modeling analysis. Comput Math Appl 57(11–12):2022–2029

Zhao L, Qian F, Yang Y, Zeng Y, Su H (2010) Automatically extracting T–S fuzzy models using cooperative random learning particle swarm optimization. Appl Soft Comput 10(3):938–944

Chapter 4
Application of Artificial Neural Networks in Short-Term Rainfall Forecasting

Mrinmoy Majumder and Rabindra Nath Barman

Abstract Short-term rainfall is important in agriculture, industry, the energy sector, and any other water-dependent activities where profitability depends on climatic conditions. The scarcity of reliable prediction models encouraged the authors of the present study to develop a modeling platform using a neurogenetic model to estimate rainfall occurrence within a short-term duration. The data on both the quantity and the probability of occurrence of rainfall based on the previous 1–5 days were used to predict the quantity and occurrence of rainfall 1–4 days hence. The potential of neurogenetic models to predict short-term rainfall on the basis of such a small-scale data set was analyzed with the aim of developing a software platform for laypeople and to help related professionals maintain the profitability of their organization by reducing the likelihood of wastage resulting from large-scale prediction errors, which are common with the available linear models. The results indicate that neurogenetic models can reliably predict rainfall 1, 3, and 4 days in advance, but not 2 and 5 days, if the models are trained with a suitable algorithm. The subpar performance of the 2- and 5-day rainfall prediction models was attributed to the choice of training algorithms and length of time, although the reliable prediction of rainfall even 1 day in advance warrants pursuing further development of the present investigation.

M. Majumder (✉)
School of Hydro-Informatics Engineering, National Institute of Technology Agartala, Barjala, Jirania 799055, Tripura, India
e-mail: mmajumder15@gmail.com

R.N. Barman
Department of Production Engineering, National Institute of Technology Agartala, Barjala, Jirania 799055, Tripura, India
e-mail: rahul.barman@yahoo.co.in

M. Majumder and R.N. Barman (eds.), *Application of Nature Based Algorithm in Natural Resource Management*, DOI 10.1007/978-94-007-5152-1_4,

4.1 Introduction

The prediction of short-term rainfall (STR) involves the estimation of the intensity and/or frequency of rainfall events within a span of 5 days. Rainfall impacts the production efficiency of the essential services (like water, electricity, and gas), agriculture, stock exchanges, and various other water-dependent industries. Laypeople also are affected by rainfall events.

Rainfall can impact daily water supplies from the water treatment plant (WTP) of any city because of the increase in suspended solids and pollutants in surface water. The chemical dozing pattern must be adjusted to prevent any excess of toxicity from affecting the quality of the treated water. If a rainfall event could be predicted within the next 24–48 h, then compensatory measures could be taken to maintain the quality of treated water.

Demand for electricity depends on temperature and humidity. After a rainfall event both temperature and humidity decreases, which leads to a reduction in the demand for electricity. Because demand is reduced, production of electricity will need to be adjusted to prevent wastage. If the occurrence of rainfall events within a short span of time could be predicted, then a sufficient amount of electrical energy could be conserved; this would also reduce the release of greenhouse gases into the atmosphere.

The demand on natural gas also varies with the frequency and intensity of rainfall events. Because of the impact of extreme events, electrical transmission or distribution networks can be damaged. The absence of electricity induces increased demand on liquefied petroleum gas and compressed natural gas.

That is why there is a need for a reliable prediction model that can estimate both probability and magnitude of rainfall at least 24–48 h before its occurrence. The absence of a large-scale data set have decreased the availability of reliable prediction models for solving such problems. The absence of a regular pattern in the related parameters within such short time domain has forced modelers to apply stochastic modeling to predict short-term rainfall patterns.

Stochastic models have fewer requirements when it comes to data but are extremely vulnerable to uncertainty due to the unstable nature of the interrelationships of the variables. Due to the complexity and uncertainty involved in such models, linear methods often fail to deliver effective estimations, as can be concluded from the references given in Table 4.1.

4.1.1 Earlier Studies on the Prediction of Short-Term Rainfall

Short-term rainfall is a popular topic of research due to its importance in industry and the agricultural output of a country. Table 4.1 shows the application of different types of mathematical and statistical models in the prediction of short-term rainfall. The accuracy achieved and drawbacks identified are also explained.

Table 4.1 Studies on the estimation of short-duration rainfall

References	Model type	Input requirement	Accuracy level	Remarks
Kisi and Cimen (2012)	Wavelet-support vector machine conjunction model	Daily rainfall data	Combined model was found to perform better than when applied separately	Prediction of rainfall 1 day in advance conducted with two types of model: combined wavelet-support vector machine model and support vector regression used separately. According to performance metrics (wavelet-support vector machine conjunction model) the latter was outperformed by former model's accuracy level
Papalexiou et al. (2011)	Stochastic model developed using scaling behavior of both time and state change of catchments	5–10 s rainfall data	Study recommended use of stochastic models; power-type distribution tails and autocorrelation functions were found to be indicators of better performance from such models	Stochastic models involve assumptions and are highly linear in nature. But in the present study the model uses scaling information of time and state to differentiate domains of solution space
Kottegoda et al. (2003)	Beta and geometric distribution for reducing parameters from daily to hourly values	Daily rainfall	The results from the distribution models were found to predict the statistical extremes satisfactorily	The study tried to disaggregate daily rainfall into hourly values with the help of beta distributions. Wet and dry season daily rainfall was found to follow a geometric distribution. From the disaggregation results it was found that the hourly rainfall also follows a geometric distribution but is conditioned on total daily rainfall values
Sugimoto et al. (2001)	Stochastic model where Kalman filter was used for incrementing time domain	Rainfall from radar data	Developed model was compared with deterministic physical model created for same purpose. Comparison result shows higher level accuracy for stochastic model with respect to deterministic modeling framework	Data from radar were converted and fed as input to a physical mesoscale climate model developed on concept of water balance and thermodynamics. Extended Kalman filter was utilized for state update. Data generated were compared with deterministic model for validation

(continued)

Table 4.1 (continued)

References	Model type	Input requirement	Accuracy level	Remarks
Thielen et al. (2000)	Conceptual, simple mass balance model of water within air columns	Surface rainfall and vertically integrated liquid water content (VIL)	Model was evaluated with numeric data generated from a mesoscale meteorological model and results encouraged further development of the model	Developed model of present study was conceptual in nature and involves simple balancing of water mass in air columns and spatial advection within variables. The result was found to be better than the simple advection routines for specific time frames
Burlando et al. (1993)	Autoregressive moving average (ARMA) model	Hourly rainfall	It was assumed that hourly rainfall follows an autoregressive moving average process; with the help of this assumption hourly rainfall was predicted for both event-based and normal time domain. Predictions were found to be more reliable for the latter with respect to the former	Assumption that hourly rainfall follows an ARMA process was used to predict short-term rainfall. But model was not compared with other models developed for same purpose
French et al. (1992)	Monte Carlo Stochastic State-space model	Hourly rainfall	Generation of hourly data on both time and space scales was found to improve its accuracy when high-density primary data were used as input	This study generates hourly rainfall data from a state-space model iterated by Monte Carlo simulation where effect of sample network density on model reliability was also analyzed. The results clearly show advantage of using high-density sampling network for such forecasting procedures

4.1.2 Neural Network for Short-Term Rainfall Prediction

The ability of neural networks to map the nonlinear and inherently complex interrelationship between a set of input variables and an output variable is well established and supported by many studies on different topics in science and engineering. Table 4.2 shows the application of neural networks in solving various types of problems. Table 4.3 presents earlier applications of neural networks, both individually and in combination with other algorithms in predicting short-term rainfall with respect to different regional conditions.

4.1.3 Neurogenetic Algorithms

In the development of neural network models, network topology, weights assigned to input variables, and the type of activation function are the three important parameters that affect the accuracy and reliability of a neural network model. Because there are no predetermined methodologies for identifying the optimal values of these parameters, various studies have applied different statistical methods, including nature-based algorithms, to determining an ideal value for these three parameters.

When genetic algorithms are used to search for optimal values of these parameters, the models are jointly referred to as neurogenetic models.

From Table 4.3 it is clear that there is a substantial lack of research studies involving stochastic neural network models and short-term rainfall. The table also shows that such models have already been developed to perform effectively in predicting hourly to monthly rainfall intensities and occurrence. The table also demonstrates that in the case of occurrence, neural networks generally prefers categorized data rather than numerical data sets (Olsson et al. 2001).

4.1.4 Objective

The main objective of the present investigation will be to analyze the capability of neurogenetic models in estimating short-duration rainfall patterns. The study involves the prediction of both the quantity and occurrence probability of rainfall within the next 5 days based on the rainfall records of the previous 5 days. A stochastic modeling approach was used, keeping in mind the scarcity of adequate data sets and the level of uncertainty included in such prediction problems.

4.1.5 Brief Methodology

In case of the neural network models the probability of occurrence and amount of rainfall in the previous 5 days were considered as input variables. In total, five

Table 4.2 Examples of application of neural network in different prediction and classification problems

References	Model type	Input-output	Success
Piotrowski et al. (2012)	Noise-injected multilayer perceptron neural networks	Longitudinal dispersion coefficient was predicted to test a novel approach for avoiding "convergence to local minima" and "inability to optimize nondifferentiable data set" errors involved in gradient-based training algorithms by introduction of evolutionary computation	The result of the noise-injected MLP neural network was compared with many evolutionary computation techniques (like distributed DE with explorative–exploitative population families, self-adaptive DE, DE with global and local neighbors, grouping DE, JADE, comprehensive learning particle swarm optimization, efficient population utilization strategy particle swarm optimization, and covariance matrix adaptation –evolutionary strategy) to find the best EC for optimal prediction. The results show that an extended differential evolutionary algorithm performs much better than the other training algorithms in use
Khashei et al. (2012)	Multilinear regression model coupled with multilayered perceptron neural network model	This model was developed to create more accurate but general two classes as well as multiclass classification algorithm. The business credit data of Japan, Indian diabetes data, forensic glass data, and a set of Fisher Iris data were used for classification with the developed model	Yes [the investigation compared the combined model with linear discriminant analysis (LDA), quadratic discriminant analysis (QDA), K-nearest neighbor (KNN), and support vector machines (SVMs)] and concluded that the present model was more effective in classifying the test data sets than the compared models
Lekouch et al. (2012)	Artificial neural network	Air and dew point temperature or relative humidity, wind speed and wind direction, and cloud cover were used to develop the model for prediction of dew yield	Yes (the model was used to extrapolate output for 15 cities in and around the test cities situated in coastal regions of southwest Morocco)

Kisi et al. (2012)	Back propagation neural network coupled with artificial bee colony algorithm (ANN-ABC)	Streamflow from two gauge stations was used to predict sediment load of same	Yes (according to results from performance metrics like mean square error and coefficient s, the ANN-ABC model performed better compared to ANN-fuzzy, neural differential evolution and rating curve models. The logarithm of the inputs was found to increase the model efficiency with respect to the normal data set)
Zhao et al. (2012)	Multilayered (three in the present investigation) neural network model	The first transverse cracking of ice cover was the output and different hydrometric and meteorological data were used as input	Yes (when compared with multilinear regression models the ANN model was found to be more efficient and effective than the former model, at least in cases of river breakup prediction)
Alvisi and Franchini (2012)	Gray neural network (a neural network model whose parameters are represented by gray numbers) (NNs)	The uncertainty level or the difference between the predicted and actual values of the river level was taken as output	Yes (the NNs were compared with a Bayesian neural network and according to the predictions from both models, estimation from NNs was far narrower than the latter neural network model, showing the accuracy level of the NN model)
Kim and Pachepsky (2010)	Regression tree (RT) coupled neural network	Missing precipitation data were the output and available rainfall data from the station were the input	Yes (better than RT when applied separately)
Jain and Kumar (2007)	Conceptual model coupled to neural network	Monthly stream flow data	Yes (four conventional time series models of autoregressive type and four neural network models were used successfully in predicting monthly stream flow of Colorado River, California)
Gautam et al. (2004)	Feedforward neural network	Groundwater level of interconnected wells of a floodplain as input and electrical conductivity was treated as output	Yes (effect of bridge construction was successfully identified by observing changes in electrical conductivity in groundwater compared to preconstruction period benchmarks)
Bodri and Cermak (2000)	Back-propagation-trained neural network	Monthly rainfall data of 38 years were used to predict rainfall for next month and next summer	Yes

Table 4.3 Examples of stochastic neural network for estimation of short-term rainfall

Type of model in prediction of short-term rainfall	Study area	Lead time	Problem solved	References
Neuroclassifier followed by neuropredictor model	Friuli Venezia Giulia region (henceforth FVG, NE Italy)	6 h	The problem of de-classification was mitigated by introduction of a regression-type neural network model	Manzato (2007)
Dynamic recurrent neural network	Wu Tu Watershed, Taiwan	3 h	Absence of link between physical concept and neural networks was overcome with introduction of unit hydrograph model	Pan and Wang (2004)
Serially connected neural network	Chikugo River basin, Kyushu Island, southern Japan	12 h	Problems of zero value, which are generally present in training data set of short time duration	Olsson et al. (2001)
Backpropagation neural networks	Moravia, eastern part of Czech Republic	1 month	Heavy rainfall followed by occurrence of an extreme event cause flooding. Prediction of next-month extreme rainfall performed by neural network; good fit was achieved, which can help professionals in the field with preplanning of compensatory and preventive measures	Bodri and Čermák (2000)

neurogenetic models were prepared, each having two submodels. The occurrence probability and quantity of rainfall for the next 5 days were considered as model outputs for the present study. As stated earlier, each model has two submodels predicting the quantity and occurrence of rainfall in different time domains (one to five).

The models were trained with the help of a data matrix that contains every possible combination between input and output variables if their data values are converted into nine categories representing nine different degrees of intensity of the variables. Table 4.4 shows the input and output variables considered for the five different models. In Table 4.5 are shown the variables and categories into which all are divided based on the intensity of the magnitude.

Because the data matrix contains all possible situations, the model can learn the inherent relationship between the input and output variables for all possible combinations, ensuring the reliability of the model output.

All the developed models were evaluated based on selected performance metrics like the kappa index of agreement, precision, sensitivity, and specificity. The metrics will help to identify the best model among the five models developed. The selection will also indicate the range within which the functionality of the neurogenetics models will be optimized.

4.2 Methodology

In the present investigation the objective is to identify the range within which neurogenetic models will efficiently predict output with the desired level of accuracy and reliability. At first, five neurogenetic models are prepared with the help of common model parameters. Each model has two submodels predicting the quantity and probability of occurrence within different time domains starting from 1 to 5 days in advance of the rainfall.

The input and output variables of all models are given in Table 4.4. Table 4.6 shows the common model parameters adopted for all the neurogenetic models so that a uniform decision can be made.

Before the models are developed, the data to train the models are preprocessed as described in the next section.

4.2.1 Data Preprocessing

All the neurogenetic models considered for the present study have ten input variables. If the data of the input variables are represented as a percentage of the maximum, then all the variables can be encoded in nine categories based on the percentage value of the variables because categorized data were found to perform more efficiently than normal sets. Table 4.5 shows the input variables and the nine categories in which the data values of the variables are grouped.

Table 4.4 Input and output variables of neurogenetic models developed

Input	Output
STRFM1	
Occurrence probability of rainfall, same day (P_t)	Occurrence probability of rainfall, 1 day before (P_{t+1})
Occurrence probability of rainfall, 1 day before (P_{t-1})	Quantity of rainfall, 1 day before (Q_{t+1})
Occurrence probability of rainfall, 2 days before (P_{t-2})	
Occurrence probability of rainfall, 3 days before (P_{t-3})	
Occurrence probability of rainfall, 4 days before (P_{t-4})	
Quantity of rainfall, same day (Q_t)	
Quantity of rainfall, 1 day before (Q_{t-1})	
Quantity of rainfall, 2 days before (Q_{t-2})	
Quantity of rainfall, 3 days before (Q_{t-3})	
Quantity of rainfall, 4 days before (Q_{t-4})	
STRFM2	
Occurrence probability of rainfall, same day (P_t)	Occurrence probability of rainfall, 2 days before (P_{t+2})
Occurrence probability of rainfall, 1 day before (P_{t-1})	Quantity of rainfall, 2 days before (Q_{t+2})
Occurrence probability of rainfall, 2 days before (P_{t-2})	
Occurrence probability of rainfall, 3 days before (P_{t-3})	
Occurrence probability of rainfall, 4 days before (P_{t-4})	
Quantity of rainfall, same day (Q_t)	
Quantity of rainfall, 1 day before (Q_{t-1})	
Quantity of rainfall, 2 days before (Q_{t-2})	
Quantity of rainfall, 3 days before (Q_{t-3})	
Quantity of rainfall, 4 days before (Q_{t-4})	
STRFM3	
Occurrence probability of rainfall, same day (P_t)	Occurrence probability of rainfall, 3 days before (P_{t+3})
Occurrence probability of rainfall, 1 day before (P_{t-1})	Quantity of rainfall, 3 days before (Q_{t+3})
Occurrence probability of rainfall, 2 days before (P_{t-2})	
Occurrence probability of rainfall, 3 days before (P_{t-3})	
Occurrence probability of rainfall, 4 days before (P_{t-4})	
Quantity of rainfall, same day (Q_t)	
Quantity of rainfall, 1 day before (Q_{t-1})	
Quantity of rainfall, 2 days before (Q_{t-2})	
Quantity of rainfall, 3 days before (Q_{t-3})	
Quantity of rainfall, 4 days before (Q_{t-4})	
STRFM4	
Occurrence probability of rainfall, same day (P_t)	Occurrence probability of rainfall, 4 days before (P_{t+4})
Occurrence probability of rainfall, 1 day before (P_{t-1})	Quantity of rainfall, 4 days before (Q_{t+4})
Occurrence probability of rainfall, 2 days before (P_{t-2})	
Occurrence probability of rainfall, 3 days before (P_{t-3})	
Occurrence probability of rainfall, 4 days before (P_{t-4})	
Quantity of rainfall, same day (Q_t)	
Quantity of rainfall, 1 day before (Q_{t-1})	
Quantity of rainfall, 2 days before (Q_{t-2})	

(continued)

Table 4.4 (continued)

Input	Output
Quantity of rainfall, 3 days before (Q_{t-3})	
Quantity of rainfall, 4 days before (Q_{t-4})	
STRFM5	
Occurrence probability of rainfall, same day (P_t)	Occurrence probability of rainfall, 5 day before (P_{t+5})
Occurrence probability of rainfall, 1 day before (P_{t-1})	Quantity of rainfall, 5 day before (Q_{t+5})
Occurrence probability of rainfall, 2 days before (P_{t-2})	
Occurrence probability of rainfall, 3 days before (P_{t-3})	
Occurrence probability of rainfall, 4 days before (P_{t-4})	
Quantity of rainfall, same day (Q_t)	
Quantity of rainfall, 1 day before (Q_{t-1})	
Quantity of rainfall, 2 days before (Q_{t-2})	
Quantity of rainfall, 3 days before (Q_{t-3})	
Quantity of rainfall, 4 days before (Q_{t-4})	

Table 4.5 Input variables and categories of neurogenetic models

Input variable	Category considered
Occurrence probability of rainfall, 1 day before (P_t)	Extremely High (EH), Very High (VH), Semi High (SH), High (H), Normal (N), Low (L), Semi Low (SL), Very Low (VL), and Extremely Low (EL)
Occurrence probability of rainfall, 2 days before (P_{t-1})	Extremely High (EH), Very High (VH), Semi High (SH), High (H), Normal (N), Low (L), Semi Low (SL), Very Low (VL), and Extremely Low (EL)
Occurrence probability of rainfall, 3 days before (P_{t-2})	Extremely High (EH), Very High (VH), Semi High (SH), High (H), Normal (N), Low (L), Semi Low (SL), Very Low (VL), and Extremely Low (EL)
Occurrence probability of rainfall, 4 days before (P_{t-3})	Extremely High (EH), Very High (VH), Semi High (SH), High (H), Normal (N), Low (L), Semi Low (SL), Very Low (VL), and Extremely Low (EL)
Occurrence probability of rainfall, 5 days before (P_{t-4})	Extremely High (EH), Very High (VH), Semi High (SH), High (H), Normal (N), Low (L), Semi Low (SL), Very Low (VL), and Extremely Low (EL)
Quantity of rainfall, 1 day before (Q_t)	Extremely High (EH), Very High (VH), Semi High (SH), High (H), Normal (N), Low (L), Semi Low (SL), Very Low (VL), and Extremely Low (EL)
Quantity of rainfall, 2 days before (Q_{t-1})	Extremely High (EH), Very High (VH), Semi High (SH), High (H), Normal (N), Low (L), Semi Low (SL), Very Low (VL), and Extremely Low (EL)
Quantity of rainfall, 3 days before (Q_{t-2})	Extremely High (EH), Very High (VH), Semi High (SH), High (H), Normal (N), Low (L), Semi Low (SL), Very Low (VL), and Extremely Low (EL)
Quantity of rainfall, 4 days before (Q_{t-3})	Extremely High (EH), Very High (VH), Semi High (SH), High (H), Normal (N), Low (L), Semi Low (SL), Very Low (VL), and Extremely Low (EL)
Quantity of rainfall, 5 days before (Q_{t-4})	Extremely High (EH), Very High (VH), Semi High (SH), High (H), Normal (N), Low (L),Semi Low (SL), Very Low (VL), and Extremely Low (EL)

Table 4.6 Characteristics of model parameters and adopted performance metrics

Model parameter	Value
Genetic algorithm parameters	
Population size	60
Number of generations	50
Crossover rate	0.80
Mutation rate	0.20
Neural network parameters	
Network topology	10-8-2
Network weight	88
Training algorithm selected	Quick propagation (QP) and conjugate gradient Descent
Quick propagation coefficient	1.75
Learning rate	0.20
Generalization loss allowed	50.00%
Correct classification rate (CCR) desired	98.00%
Number of training iterations allowed	1,000,000
Number of retrains	10
Performance metrics of STRM1	
Training CCR of STRM1-QP	99.95%
Testing CCR of STRM1-QP	99.95%
Training CCR of STRM1- CGD	99.95%
Testing CCR of STRM1-CGD	99.95%
Kappa index of agreement of STRM1 -QP	99.95%
Precision of STRM1-QP	99.95%
Sensitivity of STRM1-QP	99.95%
Specificity of STRM1-QP	99.95%
Kappa index of agreement of STRM1 -CGD	99.95%
Precision of STRM1-CGD	99.95%
Sensitivity of STRM1-CGD	99.95%
Specificity of STRM1-CGD	99.95%
Performance metrics of STRM2	
Training CCR of STRM2- QP	99.95%
Testing CCR of STRM2-QP	99.95%
Training CCR of STRM2- CGD	90.91%
Testing CCR of STRM2-CGD	100%
Kappa index of agreement of STRM2 -QP	99.95%
Precision of STRM2-QP	99.95%
Sensitivity of STRM2-QP	99.95%
Specificity of STRM2-QP	99.95%
Kappa index of agreement of STRM2 -CGD	94.76%
Precision of STRM2-CGD	95.55%
Sensitivity of STRM2-CGD	96.04%
Specificity of STRM2-CGD	95.06%
Performance metrics of STRM3	
Training CCR of STRM3- QP	99.95%
Testing CCR of STRM3-QP	99.95%
Training CCR of STRM3-CGD	99.95%

(continued)

Table 4.6 (continued)

Model parameter	Value
Testing CCR of STRM3-CGD	99.95%
Kappa index of agreement of STRM3 -QP	99.95%
Precision of STRM3-QP	99.95%
Sensitivity of STRM3-QP	99.95%
Specificity of STRM3-QP	99.95%
Kappa index of agreement of STRM3 -CGD	99.95%
Precision of STRM3-CGD	99.95%
Sensitivity of STRM3-CGD	99.95%
Specificity of STRM3-CGD	99.95%
Performance metrics of STRM4	
Training CCR of STRM4- QP	99.95%
Testing CCR of STRM4 – QP	99.95%
Training CCR of STRM4- CGD	99.95%
Testing CCR of STRM4 – CGD	99.95%
Kappa index of agreement of STRM4 -QP	99.95%
Precision of STRM4-QP	99.95%
Sensitivity of STRM4-QP	99.95%
Specificity of STRM4-QP	99.95%
Kappa index of agreement of STRM4 -CGD	99.95%
Precision of STRM4-CGD	99.95%
Sensitivity of STRM4-CGD	99.95%
Specificity of STRM4-CGD	99.95%
Performance metrics of STRM5	
Training CCR of STRM5- QP	99.95%
Testing CCR of STRM5-QP	99.95%
Training CCR of STRM5-CGD	99.09%
Testing CCR of STRM5-CGD	100%
Kappa index of agreement of STRM5-QP	99.95%
Precision of STRM5-QP	99.95%
Sensitivity of STRM5-QP	99.95%
Specificity of STRM5-QP	99.95%
Kappa index of agreement of STRM5-CGD	99.37%
Precision of STRM5-CGD	99.41%
Sensitivity of STRM5-CGD	99.44%
Specificity of STRM5-CGD	99.38%

The rule for categorizing the data values of both input and output variables are given below:

```
If {
V < 15%,
Then ( V = EL,
Else,
If (
16% < V < 30%,
```

```
Then ( V = VL,
Else,
If (
31% < V < 40%
Then ( V = SL
Else,
If (
41% < V < 50%,
Then ( V = L,
Else,
If (
51% < V < 60%,
Then ( V = N,
Else,
If (
61% < V < 70%,
Then (V = H,
Else,
If (
71% < V < 80%,
Then (V = SH,
Else,
If (
81% < V < 90%,
Then ( V = VH,
Else,
V = EH )
)
)
)
)
)
)
)
)}
```

4.2.2 Model Training, Testing, and Validation

A combinatorial data matrix is used to train all the neurogenetic models. Quick propagation (QP) and conjugate gradient descent (CGD) are selected as the training algorithms. Performance metrics like the kappa index of agreement, precision, sensitivity, and specificity were used to select the best model from among the five models considered.

4.3 Results and Discussion

As explained in the previous section, Table 4.6 shows the values of the model parameters selected for this study. The models' performance metrics are also shown in the table.

The results from the performance metrics indicate that, except for STRM2-CGD and STRM5-CGD, all the models performed satisfactorily. Because the models were validated using a small set of data randomly selected from the training data set, the metrics results were found to be near 100. Even then, also the kappa index of agreement, precision, sensitivity, and specificity for STRM2-CGD was found to be 94.76, 95.55, 96.04, and 95.06%, respectively, whereas the same for model STRM5-CGD were determined to be equal to 99.37, 99.41, 99.44, and 99.38%, although the STRM5-CGD model when compared to STRM2-CGD was concluded to be preferable to the latter.

4.4 Conclusion

The present investigation tried to estimate short-term rainfall using neurogenetic models and 1–5-days lagged rainfall data. In total, five models were prepared predicting both the occurrence and magnitude of next five days. The network topology of the models was selected using a genetic algorithm, and both QP and CGD were selected for training. The data set of the input and output variables was converted into nine categories, each representing a different level of magnitude and probability. According to the model results, it was found that only the 2- and 5-day rainfall prediction models trained with a CGD algorithm did not perform on par with the other models. This subpar performance can be attributed to the incompatibility of the CGD algorithm and the length of the time domain. If the prediction is made using real-life data of a study area, then the actual reasons for the subpar performance of the STRM2-CGD and STRM5-CGD can be properly analyzed. This approach of predicting short-term rainfall may help to predict extreme events even if data availability is scarce.

References

Alvisi S, Franchini M (2012) Grey neural networks for river stage forecasting with uncertainty. Phys Chem Earth Pt A/B/C 42–44:108–118

Bodri L, Čermák V (2000) Prediction of extreme precipitation using a neural network: application to summer flood occurrence in Moravia. Adv Eng Softw 31(5):311–321

Burlando P, Rosso R, Cadavid LG, Salas JD (1993) Forecasting of short-term rainfall using ARMA models. J Hydrol 144(1–4):193–211

French MN, Bras RL, Krajewski WF (1992) A Monte Carlo study of rainfall forecasting with a stochastic model. Stoch Hydrol Hydraul 6(1):27–45

Gautam MR, Watanabe K, Ohno H (2004) Effect of bridge construction on floodplain hydrology—assessment by using monitored data and artificial neural network models. J Hydrol 292(1–4):182–197

Jain A, Kumar AM (2007) Hybrid neural network models for hydrologic time series forecasting. Appl Soft Comput 7(2):585–592

Khashei M, Hamadani AZ, Bijari M (2012) A novel hybrid classification model of artificial neural networks and multiple linear regression models. Expert Syst Appl 39(3):2606–2620

Kim J-W, Pachepsky YA (2010) Reconstructing missing daily precipitation data using regression trees and artificial neural networks for SWAT streamflow simulation. J Hydrol 394(3–4):305–314

Kisi O, Cimen M (2012) Precipitation forecasting by using wavelet-support vector machine conjunction model. Eng Appl Artif Intel 25(4):783–792

Kisi O, Ozkan C, Akay B (2012) Modeling discharge–sediment relationship using neural networks with artificial bee colony algorithm. J Hydrol 428–429:94–103

Kottegoda NT, Natale L, Raiteri E (2003) A parsimonious approach to stochastic multisite modelling and disaggregation of daily rainfall. J Hydrol 274(1–4):47–61

Lekouch I, Lekouch K, Muselli M, Mongruel A, Kabbachi B, Beysens D (2012) Rooftop dew, fog and rain collection in southwest Morocco and predictive dew modeling using neural networks. J Hydrol (in press), Accepted manuscript, Available online 13 Apr 2012

Manzato A (2007) Sounding-derived indices for neural network based short-term thunderstorm and rainfall forecasts. Atmos Res 83(2–4):349–365

Olsson J, Uvo CB, Jinno K (2001) Statistical atmospheric downscaling of short-term extreme rainfall by neural networks. Phys Chem Earth Pt B Hydrol Ocean Atmos 26(9):695–700

Pan T-y, Wang R-y (2004) State space neural networks for short term rainfall-runoff forecasting. J Hydrol 297(1–4):34–50

Papalexiou S-M, Koutsoyiannis D, Montanari A (2011) Can a simple stochastic model generate rich patterns of rainfall events? J Hydrol 411(3–4):279–289

Piotrowski AP, Rowinski PM, Napiorkowski JJ (2012) Comparison of evolutionary computation techniques for noise injected neural network training to estimate longitudinal dispersion coefficients in rivers. Expert Syst Appl 39(1):1354–1361

Sugimoto S, Nakakita E, Ikebuchi S (2001) A stochastic approach to short-term rainfall prediction using a physically based conceptual rainfall model. J Hydrol 242(1–2):137–155

Thielen J, Boudevillain B, Andrieu H (2000) A radar data based short-term rainfall prediction model for urban areas — a simulation using meso-scale meteorological modeling. J Hydrol 239(1–4):97–114

Zhao L, Hicks FE, Robinson Fayek A (2012) Applicability of multilayer feed-forward neural networks to model the onset of river breakup. Cold Reg Sci Technol 70:32–42

Chapter 5
Application of a Genetic Algorithm to Predict the Growth Rate of *Bufo melanostictus* in Urban Forest

Mrinmoy Majumder and Rabindra Nath Barman

Abstract *Bufo melanostictus*, commonly known as the Asian common toad, is widely found in various urban forests and marshy lands. The animal is not endangered (present conservation status: Least Concern) but is under threat of being so due to various issues. One of the major threats to the toad population is the recent rapid scale of urbanization, which is gradually diminishing the common habitat and reproduction places of the species. Like other species of Bufonidae family, toads breed in still and slow-flowing rivers and temporary and permanent ponds. Many rivers have changed their characteristics due to climate change. For this reason many habitats formerly suitable for breeding are now found to be unsuitable. The present study aims to estimate the growth rate of the toad based on its various habitats and on climate patterns along with food availability. The impact from urbanization and deforestation is also considered. Overall the study tries to analyze the impact of urbanization and changes in climate patterns on the Asian common toad using a genetic algorithm technique.

Keywords Growth rate • Genetic algorithm • Climate impacts

5.1 Introduction

The present scenario of large-scale urbanization and recent changes in climate patterns has impacted on the growth rate of the Asian common toad. Although the species is not under threat of extinction and not even in the class of concern

M. Majumder (✉)
School of Hydro-Informatics Engineering, National Institute of Technology Agartala, Barjala, Jirania 799055, Tripura, India
e-mail: mmajumder15@gmail.com

R.N. Barman
Department of Production Engineering, National Institute of Technology Agartala, Barjala, Jirania 799055, Tripura, India
e-mail: rahul.barman@yahoo.co.in

M. Majumder and R.N. Barman (eds.), *Application of Nature Based Algorithm in Natural Resource Management*, DOI 10.1007/978-94-007-5152-1_5,

according to IUCN Conservation Status Ranking, it may become a member of that class soon if the present trend of habitat loss continues. The role of the toad in ecology is important. Toads actively participate in controlling the insect population. If the population of insects grows out of control, then major losses of crops may occur, creating a threat to the present level of food security of the human population. Unfortunately, human beings are becoming the main threat to toad populations.

This study attempts to predict the growth rate of the toad population with the help of a genetic algorithm (GA). But the main objective of the study is to analyze the impact of urbanization and change in climate on the growth of toad populations. The study may help to visualize future scenarios in which the climate will be different from the current one and the extreme urbanization will have already been implemented. The situation of the toad population at that time will help present-day urban managers to take specific mitigation measures to prevent the decay of the toad population as well as the agro-industry due to their interrelations.

GAs are known to be efficient enough to predict an output variable if a sufficiently large amount of data is fed to the model. The basic methodology of a GA is to create a population of new or mutated traits. The accuracy of a model depends mainly on the efficiency of the GA in replicating new patterns of data.

GAs have been used to predict the growth rate of *Bufo* sp., and a search of the literature revealed that no other algorithm was found to be more satisfactory. Because GAs are problem-independent, data-flexible, nature-based algorithms, there was no problem in applying to the kind of problem considered here.

5.1.1 **Bufo *sp.***

The Asian common toad (*Duttaphrynus melanostictus*; synonym: *Bufo melanostictus*) is a complex of more than one toad species, widely distributed in South Asia, also known as Asian toad, black-spectacled toad, common Sunda toad, and Javanese toad; it grows to approximately 20 cm long. The species breeds during the monsoon season, and the tadpoles are black. Young toads may be seen in large numbers after monsoons.

Asian common toads can be found in northern Pakistan, Nepal, Bangladesh, India, Sri Lanka, Myanmar, Thailand, Laos, Vietnam, Cambodia, southern China, Taiwan, Hong Kong, Macau, Malaysia, Singapore, and the Indonesian islands of Sumatra, Java, Borneo, Anambas, and Natuna Islands. They are also present on the islands of Bali, Sulawesi, Ambon, and Manokwari and dwell in the northeastern portion of the Vogelkop Peninsula in New Guinea as an exotic species. The Asian common toad has been sighted as high as 1,800 m. The common habitats of the toad include lowland, upper beaches, riverbanks, and agricultural and urban areas. They are rare in closed forests (van Dijk et al. 2004).

The Asian common toad can be identified by its warty skin and distinct bulges on the back of its head, known as the parotid glands; it tends to walk not jump and is covered in obvious warts. It has horizontally slit pupils with yellow/golden brown irises, and its dorsal surface and flanks are a fairly uniform brown/greenish gray; it is light sandy brown in warm weather, has a ventral surface that is dirty white/cream with speckles. It varies in length from 8 cm (male) up to 13 cm (female).

5.1.2 Life Cycle of Asian Common Toad

The life cycle of the Asian common toad is divided into six physiologically distinct stages, which are discussed in detail in what follows.

(a) Spawn (egg-mass)

The male frog fertilizes eggs in female frogs as they are laid in the amplexus or mating position. The eggs are laid in long chains known as spawn or egg mass. At this time, temperature and humidity must be favorable enough for the sustenance of the egg. The eggs at this stage are extremely vulnerable and fragile. Extreme climates, predators, or human interference can ruin the entire mass of eggs.

(b) Egg

Toads as well as frogs lay sufficient numbers of eggs so that even if most of the eggs are unable to survive the hazards between fertilization and full-blown frogness, sufficient numbers of them may still exist to carry the traits of the species to the next generation. When the eggs die, they tend to turn white or opaque.

The mitotic division of the central yolk in the egg forms a mass that is similar to raspberry inside a jelly cup, and soon the embryo starts to look more and more like a tadpole, getting longer and moving around in its egg.

In about 6– 21 days after mating, the egg will hatch. At this time eggs need a clear and calm environment withno climatic, predatorial, or human interference.

(c) Tadpole

Shortly after hatching, the tadpole still feeds on the remaining yolk, which is actually in its gut! The tadpole at this point consists of poorly developed gills, a mouth, and a tail and is extremely fragile. It usually sticks itself to floating weeds or grasses in the water using little sticky organs between its mouth and belly area. Then, 7–10 days after the tadpole has hatched, it will begin to swim around and feed on algae.

After about a month, the gills start to acquire skin, until they eventually disappear. The tadpole gets minuscule teeth that help it grate food, turning it into oxygenated, digestible particles. The long coiled guts of the tadpole help it digest as much nutrients from its meager diet as possible.

(d) Tadpole with legs

Tiny legs start to appear after approximately 6–9 weeks, the head becomes more distinct, and the body elongates. The tadpole's diet will now include larger items like dead insects and even plants.

The arms begin to develop to the point where they will eventually pop out, elbow first.

After approximately 2 months and 1 week, the tadpole begins to look like a frog with a tiny tail on its back.

Tadpoles are a delicacy to many fish, reptiles, birds, and even humans; they can easily be killed by the mere force of turbulent water and cannot survive extreme weather conditions and disruptions in their food supply.

(e) Young Toad

After approximately 3 months, the tadpole has only a tiny stub of a tail and looks like a miniature version of the adult frog. All the common habits and external characteristics of the toad start to appear.

Table 5.1 Factors affecting growth rate of *Bufo* sp. in different stages of life cycle

Stage in life of a toad	Factors
Egg	Temperature, rainfall, humidity, type of water body, turbulence in water, presence of predators, anthropogenic impacts
Tadpole	Rainfall, type of water body, flow in water body, presence of food, presence of predators, and anthropogenic impacts
Adult	Temperature, humidity, type of water body, type of landscape, type of forest, presence of food, presence of predators, and anthropogenic impacts
Reproduction	Temperature, humidity, type of water body, type of landscape, type of forest, presence of food, presence of predators, presence of proper number of mates, and anthropogenic impacts

(f) Adult Toad
In approximately 12–16 weeks, depending on water and food supply, the toad can complete its growth cycle and become ready for reproduction.

5.1.3 Factors Affecting Growth Rate of Asian Common Toad

Table 5.1 presents the factors affecting the different stages in the life of a toad. The major contributors to the growth rate of the toad are temperature, rainfall, and type of water body. The presence of predators and anthropogenic impacts play a detrimental role in the life cycle of the Asian common toad.

5.1.4 Objective of Study

The objective of the present study, as already discussed, is to predict the growth rate of the Asian common toad with respect to the climate patterns, habitat, and food sources. The main goal of the study is to analyze the impact of anthropogenic influences and climate change on the growth rate of toad populations, which control insect populations and in turn guarantee our own food sources.

The study employs advances in neurogenetic algorithms in predicting output.

5.1.5 Brief Methodology

First, the input parameters are selected based on their influence on the growth rate of the toad population. A literature search revealed that the following variables largely control toad populations:

1. Type of water body (W_t), depth (h)
2. Climate [mainly humidity (H) and temperature (T)]

3. Plant population (p) and type (p_t)
4. Presence of food source (F), i.e., type and amount of insect population as toads love to eat insects like termites, common flies, and even scorpions
5. Presence of competitive (C) and predatory (P) species
6. Distance from locality (d_L), industry (d_i), and forest/plantations (d_f): toad generally avoid human populations and settlements. But if food sources are inadequate outside urban regions, toads are forced to enter households to capture easy prey like house flies, spiders, mosquitoes, etc.

The input variables are then categorized and scored based on their effect on the growth rate. The score is directly proportional if the variable improves the population growth rate; if the presence of a variable has a detrimental effect on the growth rate, then the scoring becomes inversely proportional.

For example: the presence of predatory/competitive species has a negative impact on a toad population. Thus, the higher the concentration of predatory species, the lower the score for this category. Again, for the plant population a variable, high concentration of plants will have an incremental effect on the growth rate of the toad population due to the availability of a food source, and thus this factor was given a higher score.

Based on these categorized and scored input variables, an objective function (Eqn. 5.1) is developed such that it becomes directly proportional to the growth rate of the toad population.

$$\mathrm{Sum}\left(W_t, h, H, T, p, p_t, F, C, P, d_L . d_i, d_f\right) \tag{5.1}$$

Various situations, both constructive and destructive, are conceptualized and created by categorizing according to the situation. Then the scenario data set is fed to the trained genetic model for output. Based on the value of the output, the impact on population growth rate due to urbanization and climate change is estimated.

5.2 Neural Network and Genetic Algorithm

A neural network is a popular method of nonlinear prediction with hitherto unseen flexibility and simplicity by which it accomplishes its objective with all types of problems from various fields of research ranging from the development of a user-controlled search engine to the prediction of the impact of climate change. The basic methodology of neural networks is to predict output using a function of the weighted sum of all inputs where the weights are used to reduce the difference between the actual and predicted output.

The weights are continuously updated, and the differences between the desired and predicted output are noted until and unless the difference becomes equal to the desired difference or the maximum number of iterations allowed is completed.

The iterations are performed to update the weighting and predicting the output with the new weighted sum. If the error decreases, then the weighting is updated in the earlier pattern or it changes its direction and magnitude. The entire procedure to update the weight for minimization of error is known as training an algorithm. After the model is trained, it is tested with a data set with known output values.

The three most important decisions for improving the accuracy of the network output are as follows.

1. Selection of network topology:
   The methodology of topology selection directly impacts computational power and the duration of the makespan time. A heavy topology will require greater amounts of computational power and time and vice versa. Thus, the selection of the topology is an important aspect for increasing the accuracy of a model. Most commonly the trial-and-error method is used for the selection , but some studies used GAs, particle swarm optimization, or even fuzzy logic to accomplish the same objective. The topology was commonly selected by either exhaustive search method or meta-heuristics like GA.
2. Selection of training algorithm:
   The training algorithm is the iteration method by which an optimal weighting is found by minimization of the difference between the actual and observed output. There are various types of training algorithms like, for example, quick propagation, conjugate gradient descent, back propagation, neurogenetics, genetics, and fuzzy logic.
3. Selection of activation function:
   The activation function is the function of the weighted sum of the inputs. The activation function controls the accuracy of the models and also encourages other researchers to develop new algorithms for templating. The activation function either scales the properties or tries to reduce the noise involved in the signal. An activation function can be, for example, logistic, hyper tan, sinusoidal, or sigmoidal.

Once the model is trained and tested and if satisfactory accuracy can be achieved, it is calibrated and validated with data having known outputs. According to the accuracy of the results, the topology, training algorithm, and activation function can be used for data values that do not have a known output.

5.2.1 Genetic Algorithm

A GA can be defined as a class of adaptive stochastic nature-based heuristic algorithms involving search and optimization. A GA was first used by Holland (1992).

The basic logic behind the algorithm is to mimic a simple picture of natural selection in order to find a good algorithm with the help of mutation or to randomly vary a given collection of sample programs. The second step is to select the optimal population, which is often done by measuring it against a fitness function. The process is repeated until a suitable solution is found.

There are many types of GA. The steps involving mutation, application of crossover techniques, and testing for fitness can be customized based on the objective of the problem. GAs, like neural networks, are prone to getting stuck in a local maximum of the fitness function, but they have the advantage that they do not require the fitness function to be very smooth since a random search is performed instead of following the path of least resistance, although the success of a GA depends on the linear relationship between program parameters and the fitness function (Rowland and Weisstein 2012).

5.2.2 *Recent Application of Genetic Algorithm to Practical Problem Solving*

Table 5.2 shows some recent applications of GAs along with many other algorithms for solving different types of problems from various fields of research. The objective and the success rate achieved are also discussed.

5.3 Methodology

In the first stage of development, variables that could affect the growth rate was identified and categorized into different groups according to their degree of influence on the objective.

Table 5.3 presents the categories into which each of the input variables is divided. Based on the categories, each value of the input variables was rated and the total value of the function was determined. Each category was encoded in such a way that the total rating of the inputs becomes directly proportional to the growth rate.

The threshold values for the division were extracted from various studies and governmental reports. The categorized data was scored with the help of the analytical hierarchy process (AHP), and the total rating of each variable was used to estimate the growth rate function. After categorization of the data set a combinatorial data matrix was developed considering every possible combination that could be created within the input and output categories.

Table 5.3 shows the input variables, categories into which they were divided, and their corresponding rank with respect to their influence on growth rate of Bufo Sp. population. The rating or scoring of the variables was determined using Eq. 5.2:

$$R = \left(\left(1 - \text{Rank}\right)/10\right) \tag{5.2}$$

where R is the score or rating assigned to a variable with respect to its category.

Table 5.2 Recent application of genetic algorithm to different decision-making problems

Authors	Type of GA	Objective	Success rate
Ooka et al. (2008)	Multiobjective genetic algorithm (MOGA)	"Comfortable outdoor thermal urban environment" as a function of landscape, outdoor thermal environment, and economy	Successful at obtaining a pareto-optimal data set representing the optimal outdoor environment
Kardani-Moghaddam et al. (2012)	Hybrid genetic algorithm and variable neighborhood search	"To reduce overall cost of task executions without noticeable increment in system makespan" where task execution cost is inversely proportional to makespan time of market grid	The hybrid GA and variable neighborhood search were found to be better than other algorithms applied to solve similar problems
Galan et al. (2011)	Genetic algorithm	The influence of the forest canopy (tree density, volume of wood, Hart-Becking index, etc.) along with position dilution of precision (PDOP), the signal-to-noise ratio, and the number of satellites on the accuracy of the measurements performed by global positioning systems (GPS) receivers	The influence of forest-canopy-related variables was found to have the greatest influence on the accuracy of the GPS result
Gaafar et al. (2008)	Simple genetic algorithm and modified genetic algorithm by particle swarm optimization (PSO)	To minimize the makespan of a single flexible machine followed by multiple identical assembly stations. The potential of PSO in improving the performance of GAs was also analyzed by comparing the output with other heuristic algorithms	Results show that the regular GA outperforms the heuristic algorithms in many instances of the manufacturing process of shorter schedules, and PSO-optimized GA surpasses the regular GA by 3.6% in many instances
Ines and Honda (2005)	Genetic algorithm	To estimate the number of agriculture and water management practices from low-spatial-resolution remotely sensed imageries under mixed pixel environment based on the sowing dates, area fractions of agricultural land uses in the pixel, and their corresponding water management practices	Results indicate that information about agriculture and water management practices can be extracted from low-spatial-resolution remotely sensed images

Chikumbo and Nicholas (2011)	Multiobjective stand-level optimization island model of genetic algorithm based on selected pareto-optimal data set	To estimate a set of the most efficient thinning regime for Eucalyptus fastigata to maximize value of sawlog harvesting and volume of pulp production	The GA was found to be successfully identify the efficient thinning regime or an pareto-optimal condition
Coillie et al. (2007)	Genetic algorithm connected neural network	To update Flemish Forest Map using IKONOS imagery where GA was used for feature selection and neural network was applied to classifying the imagery	Algorithm that considers feature selection was found to be outperform the performance of models that do not consider feature selection as per the kappa index of agreement calculated from the results

Table 5.3 Categories and corresponding ranks of input variables

Name of input variable	Categories considered and their corresponding rank
Water body type	Small Pond (1), Pond (2), Lake (3), Reservoir (4)
Water body depth	EH (5),VH (4), SH (3), H (2), M (1), L (2), SL (3), VL (4), EL (5)
Temperature	EH (5), VH (4), SH (3), H (2), M (1) , L (2) , SL (3), VL (4), EL (5)
Humidity	EH (5), VH (4), SH (3), H (2), M (1), L (2), SL (3), VL (4), EL (5)
Soil type	Loam (1), Silt (2), Silty-Loam (3), Clay (4), Silty-Clay (5), Loamy-Sand (6), Sandy-Clay (7), Sandy-Loam (9), Sand (9),
Type of flora	Shrubs (1), Herbs (2), Tree (3), Orchids (4)
Concentration	EH (5),VH (4), SH (3), H (2), M (1), L (2), SL (3), VL (4), EL (5)
Presence of predators	EH (9), VH (8), SH (7), H (6), M (5), L (4), SL (3), VL (2), EL (1)
Presence of food	EH (1), VH (2), SH (3), H (4), M (5), L (6), SL (7), VL (8), EL (9)
Presence of competitive species	EH (9), VH (8), SH (7), H (6), M (5), L (4), SL (3), VL (2), EL (1)
Distance from locality	EH (1), VH (2), SH (3), H (4), M (5), L (6), SL (7), VL (8), EL (9)
Dist_Tree cover	EH (9), VH (8), SH (7), H (6), M (5), L (4), SL (3), VL (2), EL (1)
Dist_Industry	EH (1), VH (2), SH (3), H (4), M (5), L (6), SL (7), VL (8), EL (9)

The value of the objective or growth rate function was determined with the help of the scores assigned to each of the variable. The neurogenetic algorithm was now used to understand the relationship and predict the output for a combination not included in the data set. But because the study had considered a combinatorial data matrix, each and every possible combination within the inputs was considered, so the model accuracy for any combination would be predetermined, and thus the reliability of a given prediction would already be estimated.

5.4 Results and Discussion

The parameters for the GA algorithm are presented in Table 5.4. In application of the GA, two different sets of parameters were used and compared. Both of these parameter sets population size and number of generations were varied and the results were compared to determine the better model among the two different models. The topology identified with the two sets of model parameters was named as GA1 and GA2.

Table 5.4 Parameter values selected for search iterations of present neural network model

Parameters	Value 1	Value 2
Population size	40	60
Number of generations	50	40
Network size penalty	5	5
Crossover rate	0.8	0.8
Mutation rate	0.2	0.2

Table 5.5 Performance metrics of GA1 and GA2 models

GA1	GA2
0.388	0.333
0.990	0.995
0.918	0.995

The GA1 model was trained with Levenberg-Merquadart (LM), conjugate gradient descent (CGD), and quick propagation (QP) algorithms. The training precision or correct classification rate (CCR) and testing precision were found to be equal to 96.08, 92.45, and 87.56% and 91.67, 67.02, and 64.57%, respectively, for LM, CGD, and QP, respectively.

As the training and testing CCR was better in the LM algorithm, the neural network trained with LM was selected to predict the growth rate of *Bufo* sp. The network topology as selected by the GA was found to be equal to 14-1-11-1.

The GA2 network was trained only with the LM algorithm following the result of the GA1 model. The training and testing CCR for the latter model was found to be equal to 94.12 and 100%, respectively.

The sensitivity, specificity, and precision values of the GA1 and GA2 models were also calculated to identify the better of the two models.

According to the considered performance metrics, the specificity and precision of the GA2 model clearly indicated that compared to GA1 it had much better accuracy. The abstractness of the sensitivity value might be attributed to the uniformity of the data set with which the model was trained.

The results also clearly highlighted the impact of the model parameters of the GA on the accuracy of a neurogenetic modeling platform. It also indicated that population size must be greater than the number of generations so that the searching algorithm can search through a wider domain.

The low sensitivity value can be attributed to the rarity of the output categories, i.e., within the nine categories of the output SH was present 65 times and SL (Semi Low) three times but all the other categories had a frequency of 2 or 1 out of 74 rows of data.

The nonversatility of the output resulted in a low value of sensitivity for both models (GA1 and GA2). GA1 and GA2 were found to have sensitivities of 38.88 and 33.33%, respectively (Table 5.5).

Table 5.6 shows the categories of input variables assigned to create the scenarios representing climate change and the impact of urbanization.

Table 5.6 Category and corresponding rank of input variables

Name of input variable	Categories considered
Scenario: Climate change: A2 (IPCC)	
Water body type	Small pond
Water body depth	EL
Temperature	VH
Humidity	VH
Soil type	Sand
Type of flora	Shrub
Concentration	VL
Presence of predators	VH
Presence of food	VL
Presence of competitive species	M
Distance from locality	EL
Dist_Tree cover	EH
Dist_Industry	EL
Scenario: Climate change: B2 (IPCC)	
Water body type	Lake
Water body depth	SH
Temperature	M
Humidity	M
Soil type	Loam
Type of flora	Tree
Concentration	VH
Presence of predators	EH
Presence of food	VH
Presence of competitive species	SH
Distance from locality	EH
Dist_Tree cover	EL
Dist_Industry	EH
Scenario: Urbanization: Large-scale urbanization	
Water body type	Small pond
Water body depth	EL
Temperature	SH
Humidity	SH
Soil type	Sand
Type of flora	Shrub
Concentration	EL
Presence of predators	VH
Presence of food	M
Presence of competitive species	M
Distance from locality	EL
Dist_Tree cover	EH
Dist_Industry	EL

(continued)

Table 5.6 (continued)

Name of input variable	Categories considered
Scenario: Urbanization: Mid-scale urbanization	
Water body type	Pond
Water body depth	SL
Temperature	SH
Humidity	SH
Soil type	Silt
Type of flora	Herb
Concentration	M
Presence of predators	VH
Presence of food	H
Presence of competitive species	H
Distance from locality	M
Dist_Tree cover	M
Dist_Industry	M
Scenario: Urbanization: Low-scale urbanization	
Water body type	Lake
Water body depth	SH
Temperature	SH
Humidity	SH
Soil type	Loam
Type of flora	Tree
Concentration	VH
Presence of predators	EH
Presence of food	VH
Presence of competitive species	SH
Distance from locality	EH
Dist_Tree cover	EL
Dist_Industry	EH

The climatic scenarios were simulated with the help of the recommendations taken from the description of the IPCC A2 and B2 scenarios conceptualized to represent two kinds of climatic impacts: in one scenario, industry was prioritized, and in the other environmental sustainability was given a heavier weighting.

Accordingly (Table 5.6), in the present study, temperature and humidity were assigned a category of VH. The category to represent the presence of a body of water and its depth is also presented as per the A2 scenario. Because industrial concerns would be given priority, the number of trees would be close to zero and mostly shrubs would prevail over other types of flora. Also, frogs would have to stay near an industrial plant or a residential complex because these would be present in high concentrations since industry was to be given priority over environmental sustainability. If tree cover were reduced, the presence of predators and competitive species would also decrease.

However, new kinds of predators like birds, city snakes, etc. could harm frog populations. Thus, the category of VH (Very High) was assigned to the factor (Presence of predators). Similarly, species like lizards, large spiders, and some domesticated carnivores would share the frogs' food source. Thus, the category for the presence of competitive species was assigned an M (Medium or Neither High Nor Low).

In the case of large-scale urbanization, the scenario was simulated by assigning the same categories to the input variables except that the climate parameters were put in the SH (Semi High) category representing the typical weather of tropical countries. Also, small ponds are generally frequently found in large urbanized areas. Thus, the type of body of water was given as Small Pond and, like the city ponds, their depth was categorized as EL (Extremely Low).

For B2 and low urbanized cityscape temperature and humidity will be M or Medium i.e. neither high nor low. The type of water body will be lake or reservoir as such geographical features would be protected by strict laws. Predators, food, and competitive species would be pervasive because this type of environment would be conducive to their presence.

As the environment would be conserved, a sufficient amount of protection and cover could be provided to frog species. The category EL was thus assigned to distance from tree cover, but distance from industry and locality has a adverse effect on the toad population so the highest category of the factor is assigned to be EH was given the highest category of the opposite, i.e., EH.

The difference between B2 and low urbanization was the category of climate variable where the latter was assigned a category representing typical tropical climates.

In case of mid-scale urbanization, the climate parameters were given the same category as the other urbanization situations, but all the other variables were adjusted with respect to a mid-scale urbanized city like Nasik or Haridwar of India.

That is why distance from industry, locality, and tree cover was assigned a category of M, which represents the exact situation of a mid-urbanized city. As is characteristic of all semiurbanized landscapes, the presence of predators and competitive species of the common toad was high as the impact of total urbanization was not as high as in a large-scale urbanized city. The soil pattern also changed to silt as such soil is generally observed in semieroded land features common to landscapes likely to be found in full-blown cityscapes.

Table 5.7 shows the categories of output as predicted by the selected neurogenetic models. The output was also explained and compared with real-time scenarios.

5.5 Conclusion

The present investigation tried to predict the growth rate of the Asian common toad using the power of neurogenetic models. The role of GA parameters in enhancing model performance was also examined. After training and testing of the model and selecting the better of two models, predictions were made on the basis of five scenarios representing climate impacts and urbanization situations. According to the model output a semi-low growth rate would be observed if industry and urban extension were given the highest weighting. The opposite situation would hold if environmental laws were strictly enforced; in this scenario, a semi-high growth rate was predicted.

Table 5.7 Scenarios representing different situations due to urbanization and climate change that can effect the Asian common toad population

Scenario	Output	Remarks
Climate change: A2 (IPCC)	SL	The increase in temperature and humidity and impact of industry have resulted in the given output from the model. As the presence of food is VL and the distance from and concentration of tree cover will be scarce, the toad will have no place to reproduce or survive. But the only positive of this scenario is the not-so-high concentration of predators and competitive species. Thus, a semi-high growth rate was predicted instead of extremely low growth rate
Climate change: B2 (IPCC)	SH	The strict conservation of the environment will provide a high concentration of tree cover both in amount and frequency. An adequate amount of water bodies like lakes will be conserved. The species distribution will also be under control. Industry and domesticated spaces will be concentrated in certain locations, and thus the toad will have enough freedom to reproduce and survive. Thus, a growth rate of SH was predicted by the model
Urbanization: large-scale urbanization	SL	The model prediction can be justified by the remarks given for the A2 climate change scenario. The only difference is the temperature and humidity. As far as the model simulation is concerned, neither of the variables has a noticeable influence on output. But if the parameter category were EL or VL, it might have an impact on the growth rate of the toad population
Urbanization: mid-scale urbanization	SH	The situation with large- and mid-scale landscapes was very similar, the only difference was that the impact of urbanization was not as intense as for large-scale urbanization. As the food source, presence of tree cover, water bodies, and their concentration were moderate, the output of the model predicted a semi-high growth rate for the toad population
Urbanization: small-scale urbanization	SH	The impact of urbanization was minimum in this scenario. The presence of predators reduced the growth rate to SH. If predators and competitive species (M) are scarce, the growth rate may be EH or VH

The output from the model indicated that the presence of predators and competitive species had a noticeable influence on the growth of Asian common toad population. The distance from tree cover and industrial plants and residential complexes was also important to the growth rate of the toad. The study highlighted the scope of increasing toad population by identifying the harmful elements that would hinder the rate of growth if proper compensatory measures were not taken. On the other hand, the GA that was used was found to perform better if the number of generations was less than the amount of data population. Although the GAs were found to be very slow with heavy data set.

References

Chikumbo O, Nicholas I (2011) Efficient thinning regimes for Eucalyptus fastigata: multi-objective stand-level optimisation using the island model genetic algorithm. Ecol Model 222(10):1683–1695

Coillie FMBV, Verbeke LPC, Wulf RRD (2007) Feature selection by genetic algorithms in object-based classification of IKONOS imagery for forest mapping in Flanders, Belgium. Remote Sens Environ 110(4):476–487

Gaafar LK, Masoud SA, Nassef AO (2008) A particle swarm-based genetic algorithm for scheduling in an agile environment. Comput Ind Eng 55(3):707–720

Galán CO, Rodríguez-Pérez JR, Martínez Torres J, García Nieto PJ (2011) Analysis of the influence of forest environments on the accuracy of GPS measurements by using genetic algorithms. Math Comput Model 54(7–8):1829–1834

Holland JH (1992) Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence, 2nd edn. MIT Press, Cambridge, MA

Ines AVM, Honda K (2005) On quantifying agricultural and water management practices from low spatial resolution RS data using genetic algorithms: a numerical study for mixed-pixel environment. Adv Water Res 28(8):856–870

Kardani-Moghaddam S, Khodadadi F, Entezari-Maleki R, Movaghar A (2012) Hybrid genetic algorithm and variable neighborhood search for task scheduling problem in grid environment. Procedia Eng 29:3808–3814

Ooka R, Chen H, Kato S (2008) Study on optimum arrangement of trees for design of pleasant outdoor environment using multi-objective genetic algorithm and coupled simulation of convection, radiation and conduction. J Wind Eng Ind Aerodyn 96(10–11):1733–1748

Rowland T, Weisstein EW. Genetic algorithm. Retrieved from MathWorld – A Wolfram Web Resource. http://mathworld.wolfram.com/GeneticAlgorithm.html. Feb 2012

van Dijk PP, Iskandar D, Lau MWN, Huiqing G, Baorong G, Kuangyang L, Wenhao C, Zhigang Y, Chan B, Dutta S, Inger R, Manamendra-Arachchi K, Khan MSK (2004) Duttaphrynus melanostictus. IUCN Red List of Threatened Species. Version 2011.2. International Union for Conservation of Nature, Retrieved from www.iucnredlist.org. Accessed 23 Jan 2013

Part II
Site Selection Algorithms

Chapter 6
Comparison of Nature-Based Algorithms in Impact Analysis of Climate Change on Water Resources

Mrinmoy Majumder and Rabindra Nath Barman

Abstract Predicting the impact of climate change on water availability of many regions, watersheds, or countries is currently a popular subject of research given the different signs of changes in climate. A thorough search of the relevant literature revealed that mainly conceptual, statistical, or stochastic models are used for the prediction of climatic impacts on water availability. In some studies, nature-based algorithms such as neural networks or genetic algorithms were used in predictions. From the performance metrics and according to the authors of such studies, the accuracy of nature-based models is much more consistent than the conceptual or statistical models. That is why the present study tries to analyze and compare the ability of nature-based models to predict climatic impacts on water availability from a data set that represents all possible combinations of input and output variables. Ultimately the goal is to identify the best nature-based model from among the many available ones. The investigation was able to use only four of the most popular nature-based algorithms namely Artificial Neural Network, Genetic Algorithm, Ant Colony Optimization and Artificial Bee Colony algorithm. Readers can employ other meta-heuristics to compare the performance of all such algorithms in prediction of climatic impact on water resources.

Keywords Nature-based algorithms • Climate change • Water availability

---

M. Majumder (✉)
School of Hydro-Informatics Engineering, National Institute of Technology Agartala, Barjala, Jirania 799055, Tripura, India
e-mail: mmajumder15@gmail.com

R.N. Barman
Department of Production Engineering, National Institute of Technology Agartala, Barjala, Jirania 799055, Tripura, India
e-mail: rahul.barman@yahoo.co.in

M. Majumder and R.N. Barman (eds.), *Application of Nature Based Algorithm in Natural Resource Management*, DOI 10.1007/978-94-007-5152-1_6,

6.1 Introduction

In the prediction of the impacts of climate change on water resources of a place, the mechanism of the development of the model is extensive and complex. Generally, in the case of conceptual models, the methodology of development is to follow the inherent relationship between the input and the output variables that are established following governing physical laws. But the problem with conceptual models is that the relationship between parameters and output does not remain constant throughout the data set, nor does it hold linear when spatial domains are changed. The change in the spatial and time domains invites error in the prediction model. Again, in the case of statistical models, the relationship between two variables is found from the data of the variables.

It may happen that a given set of data may or may not follow the empirical equations developed earlier. It is complex and computationally expensive to establish a uniform relationship of a data set with one single equation. That is why statistical equations often are developed for a specific time and space domain.

On the other hand, some models follow or mimic a nature-based logic in identifying the optimal solution within a given time and space domain. In nature, insects, birds, reptiles, human internal organs etc., has their own logic to find the point where optimality can be achieved within a specific domain of time and space. Outside the domain such a search process fails and often an animal or system that attempts to search outside its common territory loses its struggle for survival if it is unable to adapt. This characteristic of adaptation is used by many scientists and engineers to find a solution to their problems within their domain of time and space but with a high level of accuracy and reliability.

This investigation aims to find the better algorithm within an available set of nature-based algorithms to remove the confusion and controversy regarding the selection of a specific meta-heurestic algorithms for application in a specific problem domain. The study aims to identify the better nature-based algorithm for the prediction of water availability. The widely applied artificial neural networks (ANNs), genetic algorithms (GAs), ant colony optimization (ACO), and artificial bee colony (ABC) algorithm have been used for the prediction of water availability with respect to related variables like rainfall, evapotranspiration (ET), runoff, catchment loss, storage capacity, and water demand.

6.1.1 Climate Change and Water Availability

The increase in the atmosphere of greenhouse gases like carbon dioxide, methane, sulfur dioxide, and others has raised the average temperature of the Earth's surface. The impact of increasing temperatures is manifold. Whether in terms of food production or in water availability, changes in climatic patterns due to global warming has caused disruption in every aspect of life.

According to the water balance equation, water availability or storage is a function of rainfall, ET, runoff, and groundwater outflow. Due to climate change, both rainfall and ET have changed their patterns. As both of these parameters also control the floral production of watersheds and as infiltration and water-holding capacities of watersheds also vary with changes in the amount and type of floral population, the runoff from basins is also a function of rainfall and ET. That is why the climate of any region is so important, and a minor change in patterns can change many interrelationships that exist within a natural ecosystem.

Water availability is a direct function of rainfall and ET, and thus changes in the patterns of these two variables can change the water resource situation of countries, states, or regions.

The climatic impacts on water availability were discussed by Parish et al. (2012), Beck and Bernauer (2011), Alcamo et al. (2007), and many other scientists with the help of climatic and hydrologic models along with environmental scenarios as proposed by the Intergovernmental Panel on Climate Change (IPCC). The impact of climate change on food production, crop yield (Kang et al. 2009), water security (Qiu et al. 2012), malnutrition and livelihoods (Jankowska et al. 2012), groundwater (Green et al. 2011) etc. has been widely discussed in many studies published in reputed international journals and governmental reports.

6.1.2 Nature-Based Algorithms

Nature-based algorithms are the most recent addition to the series of optimization algorithms that mimic the decision-making procedures of different natural systems and animals. For example, ANNs follow the signal-transmission methodology of the human nervous system. The efficiency of the output depends on how well the problem was taught to the model. Flexibility in the development and ease with which it simplifies a complex nonlinear problem, has made ANNs among the most sought after modeling methodology [the number of studies related to neural networks published in refereed journals in 2012 (as of this writing) is 11,494 (in Elsevier and Springer 2012)].

In GAs the theory of natural selection at the time of meiotic cell division is followed to find the solution to a given problem within a specific search domain. In such algorithms, crossovers and corresponding mutations are produced by randomization, and the selection of genes is made with respect to the value of fitness functions. GAs are mainly used in optimization problems to search for optimal solutions within a closed domain bounded by constraints. GAs has also been used in the training of neural networks (e.g., Nasseri et al. 2008), interpolation (Ooba et al. 2006), different optimization problems (Srinivas and Ramanjaneyulu 2007), and sometimes for prediction (Hajda et al. 1998). So far, nearly 156,719 articles have been published in different refereed journals of Springer and Elsevier publishers.

The ACO algorithm was proposed by Marco Dorigo in his Ph.D. thesis (1992), and it is now popularly used to search for solutions to various optimization problems. The logic used by an ant colony for identifying food sources is replicated to find the optimal path to a solution. Although new, this algorithm is also applied in various types of problems and has been found to produce satisfactory results in most cases. To date, approximately 22,331articles and 6,157 book chapters about ACO and its application have been published. ACO is used mainly for classification and clusterization (Yu et al. 2011). Analysis of the published studies (Chandra Mohan and Baskaran 2012) shows that the ACO algorithm has been used mainly for bound and unbound optimization (Feng et al. 2007), shortest path problems (Sudholt and Thyssen 2012), image processing (Verma et al. 2012), job scheduling (Guo et al. 2012), flow routing (Euchi and Mraihi 2012), data mining (Weng and Liu 2006), and time series prediction (Gromov and Shulga 2012).

The ABC algorithm is another example of a nature-based algorithm where the foraging behavior of a bee colony is replicated in practical problem-solving methods. ABC was proposed by Karaboga in 2005. It is a swarm-based metaheuristic algorithm used mainly in optimization-related problems. Evaluation of the relevant literature reveals that ABC has not enjoyed much popularity to date. A mere 2,809 articles and 1,130 book chapters were published in 2012. The algorithm is mostly used for optimization problems[(parameter (Samanta and Chakraborty 2011), functional (Kang et al. 2011), or global optimization (Gao et al. 2012)].

Besides the previously discussed algorithms there are more nature-based algorithms like particle swarm optimization, bat and bacterial algorithms, etc.

In the present study an ANN, GA, ACO algorithm, and ABC algorithm were used to predict water availability. The predicting capabilities of all the algorithms are compared with the help of a correct classification rate and the kappa index of agreement. The result of the comparative analysis is a greater ability to identify the best nature-based algorithms in terms of making predictions.

6.2 Methodology

In the present investigation, four models were prepared using the advancement and concepts of ANN, GA, ACO, and ABC. The output amount of available water (WA) and the inputs were of two types:

1. Variables that increase water availability: rainfall (*P*), ET, storage capacity (SC);
2. Variables that reduce the amount of water available: runoff (*Q*), catchment loss (*L*), water demand (WD).

6.2.1 Preparation of Data Set

The basic requirement of any nature-based algorithm is a sufficient amount of training data from which to learn the inherent patterns. That is why an extensive data set is required to properly train the models.

All the variables (input and output) were first categorized into nine groups representing different degrees of magnitude. The groups are as follows:

EH: Extremely High: highest amount possible
VH: Very High
SH: Semi High
H: High
M: Medium: neither high nor low: Moderate
L: Low
SL: Semi Low
VL: Very Low
EL: Extremely Low: lowest amount possible

Once the categorizations of the input variables were completed, a combinatorial data set was created that represented every possible relation within the input and output variables. This combinatorial data set was used to train the ANN and GA models.

6.2.2 Development of Nature-Based Algorithms

In the case of an ANN model, the network topology and weight update were selected respectively by exhaustive nonlinear search and conjugate gradient descent algorithms. The desired correct classification rate (CCR) was kept at 97%. An amount of 3.5% outlier and 5% generalization loss was allowed. After searching for the network topology, 6-2-1(network weight: 14) was found to be an optimal architecture. After training, the model training and testing CCRs of, 99.98 and 99.87% were respectively achieved. CCR can be defined as

$$\mathrm{CCR} = \frac{\text{Number of Correct Classifications}}{\text{Total Number of Classifications}} \tag{6.1}$$

In the case of the GA model, the combinatorial data set was used for training. The root relative square error was selected as the fitness function whose value would determine the direction and number of iterations required to evolve the solution. The relative squared error is relative to what it would have been if a simple predictor had been used. More specifically, this simple predictor is just the average of the actual values. Thus, the relative squared error takes the total squared error and normalizes it by dividing by the total squared error of the simple predictor. By taking the square root of the relative squared error one reduces the error to the same dimensions as the quantity being predicted (Eq. 6.2):

$$\mathrm{RRSE} = \sqrt{\sum_{j=1}^{n} = \frac{(P_{ij} - T_j)}{(T_j - T_m)}} \tag{6.2}$$

where P_{ij} is the predicted value, T_j is the targeted value, and T_m is the mean value of the targets. Before training, the categorized data were converted to their numerical representation. The parameter of the GA was selected as 30 populations with 3 generations having a mutation rate of 0.0051. The training and testing CCR was found to be 98.14%.

The ACO model for the present investigation was developed by dividing the input variables into qualitative and quantitative parameters. The former was compared with the pheromone intensity, and the latter was made similar to the path selection of ants. An objective function was prepared similar to Dorigo's. Variables that increase WA were kept in the numerator and those that are inversely proportional to WA were placed in the denominator. The values of the variables were fed to the objective function and the result from the function was ranked in descending order, where the magnitudically larger objective function would be given higher rankings and vice versa. The inverse probability of the same was calculated and categorized in such a manner that a higher category (like EH) would indicate a high WA and vice versa. The reason for deducing the inverse probability was to maintain the hierarchy of the variables, which must be directly proportional to the availability of water. The CCR and kappa index of agreement were used as performance metrics.

In the case of the ABC algorithm, the three stages of foraging for food by honey bees was replicated in the following manner:

1. Initial Phase
   All the categories of the input variables were ranked such that variables that induce greater availability of water (friendly variables) were scored higher , whereas and variables that reduce WA (unfriendly variables) were assigned a lower rank. This can be compared with the manner in which the bees started their food search. The most likely sources of food are first identified by the scout bees.
2. Worker Bee Phase
   All the variables have now been converted into their respective probability values where friendly variables had a low probability and unfriendly variables were found to have a higher probability due to the manner in which they were ranked in the initial phase. That is why this phase could be compared with the worker bee phase where the scout bees, after finding a food source, become worker bees by returning to the hive with food and indicating the quality of the source by dancing. By this point, the variables have already been ranked, but now their probability was calculated to identify the optimal point properly without the presence of noise and uncertainty.
3. Onlooker Bee Phase
   After deducing their probabilities, the friendly and foe variables were grouped separately and ranked in ascending order so that they could be categorized based on their degree of intensity. This step of the model development can be compared with the onlooker bees' food foraging process where the onlooker bees look for signals from the worker bees and, depending on the signal, fly to the food source. In the present study, the probabilities of the friendly and unfriendly variables can be compared with the signal from worker bees where the good quality food is

signaled in a positive manner and poor quality of food source is indicated by a negative signal.

4. Scout Bee Phase
   Once the resultant value was found by subtracting the rank achieved by unfriendly variables from the rank achieved by friendly variables, a categorization was made and the resultant value was grouped into one of nine categories such as EH, VH, SH, H, M, L, SL, VL, and EL representing different levels of positivity and negativity. Values greater than 0 meant grouping into EH, VH, SH, and H, and the remaining values were encoded as M, EL, VL, SL, and L. After this categorization was complete, the worker bees again looked toward the ranked variables and repeated the entire procedure. At that time the worker bees again turn back into scout bees.

6.2.3 Performance Metrics

After the models were developed, they were used to predict the output for the known values of input variables. The result was compared with the actual values of the input. The best model was selected based on the actual and predicted values by applying the concept of CCR and kappa index of agreement (Cohen 1968).

6.3 Results and Discussion

The values of the performance metrics for all four models are given in Table 6.1. According to the fitness functions considered to evaluate the models, ANN and ACO had the maximum CCR and kappa when actual and predicted clusters were compared. Both models were found to have 0.99 times better CCR and 0.55 times better kappa with respect to the ABC model which had the lowest values of CCR and KAPPA.

From the comparison it could be concluded that ANN and ACO models are better than the GA and ABC models when both are employed to predict availability of water for a certain study area. The simplicity of both the ANN and ACO models along with their ability to map nonlinear, complex relationships may have been the reason behind their reliable and accurate representation of the modeled phenomena.

Table 6.1 Values of performance metrics considered for selected nature-based models

Model name	CCR	Kappa
ANN	0.998	0.998
GA	0.997	0.981
ACO	0.998	0.998
ABC	0.500	0.642

Table 6.2 Categories assigned to each input variable for representing IPCC A2 and B2 scenarios

Name of input variable	A2 scenario	B2 scenario
Rainfall	EH	EH
Evapotranspiration	EH	SL
Runoff	EH	SL
Catchment loss	EH	SL
Storage capacity	VL	VH
Water demand	EH	M

The selected model was used to estimate the availability of water in the case of the IPCC A2 and B2 scenarios. The category assigned to the input variables for representing the A2 and B2 scenarios is given in Table 6.2.

According to the prediction from the selected models, the WA for the A2 scenario would be Very Low (predicted by ANN) and Extremely Low (predicted by ACO) and for B2 scenario it would be Very High (predicted by ANN) and Extremely High (predicted by ACO). The conservative nature of the ANN model could be attributed to the conservative prediction found in the present study where the "Extremely" category was predicted by ACO, and the "Very" category was estimated by ANN model. The inputs were constructed so as to represent the situation of a tropical watershed and all the categories of the input were assigned with respect to the recommendations received from IPCC SRES scenarios.

6.4 Conclusion

The present investigation tried to estimate WA with the help of nature-based algorithms such as ANN, ACO, GA, and ABC. The main objective was to compare the performance of the four most widely used algorithms that mimic natural heuristics. CCR and kappa were used to evaluate the models' efficacy. According to the performance metrics, the ANN and ACO both the models were found to have the larger accuracy as represented by the CCR of 0.998 and kappa of 0.998 than the other two models considered. ABC was found to have the lowest CCR, 0.500, and kappa, 0.642. The selected model was used to estimate the WA of a tropical watershed with respect to SRES A2 and B2 scenarios. The category estimated was found to be in accordance with the prescribed scenarios, which conveys the level of accuracy of the models. The data set of the model was categorized so that the degree of availability, instead of the quantity of water, could be estimated and used to get an idea of the impact of climate change on the availability of water. The model can now be compared with the conceptual models or with some well-known statistical regression techniques to establish the comparative supremacy of the nature-based algorithms.

References

Alcamo J, Dronin N, Endejan M, Golubev G, Kirilenko A (2007) A new assessment of climate change impacts on food production shortfalls and water availability in Russia. Glob Environ Chang 17(3–4):429–444

Beck L, Bernauer T (2011) How will combined changes in water demand and climate affect water availability in the Zambezi River basin? Glob Environ Chang 21(3):1061–1072

Chandra Mohan B, Baskaran R (2012) A survey: Ant Colony Optimization based recent research and implementation on several engineering domain. Expert Syst Appl 39(4):4618–4627

Cohen J (1968) Weighted kappa: nominal scale agreement provision for scaled disagreement or partial credit. Psychol Bull 70:213–220. doi:10.1037/h0026256

Dorigo M (1992) Optimization, learning and natural algorithms. PhD thesis, Politecnico di Milano, Italy

Euchi J, Mraihi R (2012) The urban bus routing problem in the Tunisian case by the hybrid artificial ant colony algorithm. Swarm Evolut Comput 2:15–24

Feng YJ, Yu L, Zhang GL (2007) Ant colony pattern search algorithms for unconstrained and bound constrained optimization. Appl Math Comput 191(1):42–56

Gao W, Liu S, Huang L (2012) A global best artificial bee colony algorithm for global optimization. J Comput Appl Math 236(11):2741–2753

Green TR, Taniguchi M, Kooi H, Gurdak JJ, Allen DM, Hiscock KM, Treidel H, Aureli A (2011) Beneath the surface of global change: impacts of climate change on groundwater. J Hydrol 405(3–4):532–560

Gromov VA, Shulga AN (2012) Chaotic time series prediction with employment of ant colony optimization. Expert Syst Appl 39(9):8474–8478

Guo C, Zhibin J, Zhang H, Li N (2012) Decomposition-based classified ant colony optimization algorithm for scheduling semiconductor wafer fabrication system. Comput Ind Eng 62(1):141–151

Hajda P, Novotny V, Feng X, Yang R (1998) Simple feedback logic, genetic algorithms and artificial neural networks for real-time control of a collection system. Water Sci Technol 38(3):187–195

Jankowska MM, Lopez-Carr D, Funk C, Husak GJ, Chafe ZA (2012) Climate change and human health: spatial modeling of water availability, malnutrition, and livelihoods in Mali, Africa. Appl Geogr 33:4–15

Kang Y, Khan S, Ma X (2009) Climate change impacts on crop yield, crop water productivity and food security – a review. Prog Nat Sci 19(12):1665–1674

Kang F, Li J, Ma Z (2011) Rosenbrock artificial bee colony algorithm for accurate global optimization of numerical functions. Inform Sci 181(16):3508–3531

Karaboga D (2005) An idea based on honey bee swarm for numerical optimization. Technical report TR06, Erciyes University, Engineering Faculty, Computer Engineering Department

Nasseri M, Asghari K, Abedini MJ (2008) Optimized scenario for rainfall forecasting using genetic algorithm coupled with artificial neural network. Expert Syst Appl 35(3):1415–1421

Ooba M, Hirano T, Mogami J-I, Hirata R, Fujinuma Y (2006) Comparisons of gap-filling methods for carbon flux dataset: a combination of a genetic algorithm and an artificial neural network. Ecol Model 198(3–4, 15):473–486

Parish ES, Kodra E, Steinhaeuser K, Ganguly AR (2012) Estimating future global per capita water availability based on changes in climate and population. Comput Geosci 42:79–86

Qiu GY, Yin J, Geng S (2012) Impact of climate and land-use changes on water security for agriculture in Northern China. Agric Sci China 11(1):144–150

Samanta S, Chakraborty S (2011) Parametric optimization of some non-traditional machining processes using artificial bee colony algorithm. Eng Appl Artif Intel 24(6):946–957

Srinivas V, Ramanjaneyulu K (2007) An integrated approach for optimum design of bridge decks using genetic algorithms and artificial neural networks. Adv Eng Softw 38(7):475–487

Sudholt D, Thyssen C (2012) Running time analysis of Ant Colony Optimization for shortest path problems. J Discrete Algorithms 10:165–180

Verma OP, Kumar P, Hanmandlu M, Chhabra S (2012) High dynamic range optimal fuzzy color image enhancement using Artificial Ant Colony System. Appl Soft Comput 12(1):394–404

Weng SS, Liu YH (2006) Mining time series data for segmentation by using Ant Colony Optimization. Eur J Oper Res 173(3):921–937

Yu J, Chen Y, Wu J (2011) Modeling and implementation of classification rule discovery by ant colony optimisation for spatial land-use suitability assessment. Comput Environ Urban Syst 35(4):308–319

Chapter 7
A Neuro-Fuzzy Approach to Selecting Crops in Vertical Irrigation

Mrinmoy Majumder

Abstract Uncontrolled use of land resources and an ever increasing population has led to a scarcity of land in many countries, especially in Asia where population is higher than in other parts of the world. Also, the recent growth in urban populations has induced the use of forest land for agriculture or for residential purposes. In some countries governments are encouraging people to opt for vertical residences (multistoried apartments) where a single area is used to accommodate more than one family. In countries like China and Japan, where land scarcity is acute, people practice agriculture in multistoried structures. But irrigation requirements for this kind of agricultural practice are different from those of conventional procedures. Not all crops can be cultivated inside apartments due to the controlled nature of the inside environment. Thus the present study will try to find a methodology for selecting suitable species of crop for indoor cultivation ensuring the desired level of yield under minimum uncertainty.

Keywords Vertical irrigation • Fuzzy logic • Neuro-genetic model

7.1 Introduction

This investigation aims to formulate a procedure for selecting crops that would be suitable for cultivation under vertical irrigation systems. Advances in neurogenetic models and fuzzy logic are used to identify and predict the suitability of crops for the considered conditions.

The recent demand for land by burgeoning populations and requirements to ensure food security for those populations make the system of vertical crop cultivation the easiest and most useful solution to the current crisis of land availability.

M. Majumder (✉)
School of Hydro-Informatics Engineering, National Institute of Technology Agartala,
Barjala, Jirania 799055, Tripura, India
e-mail: mmajumder15@gmail.com

M. Majumder and R.N. Barman (eds.), *Application of Nature Based Algorithm in Natural Resource Management*, DOI 10.1007/978-94-007-5152-1_7,

The scarcity of land has forced farmers to encroach on forest land and even riverbeds to satisfy the rising demand for food.

To prevent such land use, the cultivation of crops under vertical irrigations systems to maintain the food supply has been found to be the easiest and most reliable solution. But the problem with such a system is that not all crops support vertical irrigation systems. The following factors must be considered for the cultivation of crops under vertical irrigation methods:

1. Length of roots (L_R)
2. Spread of roots (R_S)
3. Nutrient absorption (N)
4. Temperature tolerance (T)
5. Water tolerance (W)
6. Parasite tolerance (Pa)
7. Space requirements (A)
8. Profitability compared with conventional agricultural practices (P)

Short crops whose roots do not spread out very much can absorb nutrients from water and do not depend on the soil for their daily nutrient requirements, and crops that are temperature tolerant and do not attract parasites or other pests can be used for cultivation under vertical irrigation systems.

Because the cultivation will be performed in vertical systems, the weight of the crop is a major factor when selecting crops for such systems. The crops must not be highly water intensive because excess water will increase the weight requirement. That is why drip or sprinkler irrigation is used in such irrigation methods, where water is channelized from a reservoir situated on a roof. A proper drainage system must also be used so that water is not stored anywhere within the irrigation system. Because soil increases the weight requirement and can attract different hazardous species, hydrophilic plants are generally preferred. But plants with minimum requirements for soil or that can be cultivated in sandy soil also are harvested under vertical irrigation systems. Plants must also be odorless and not attract insects or other parasites because this can create disturbances to other residents of the vertical dwelling. Generally fruits like tomatoes, vegetables like lettuce, cabbage, and brinjal, and crops like rice and maize are popular species known to be cultivated under vertical methods of irrigation.

The success of a vertical irrigation system thus mainly depends on the capacity of the crop to tolerate a controlled and closed environment and the efficiency of the farmer in meeting the daily requirements of the crop. Proper selection of crop species also reduces maintenance requirements.

Not only the environment, but the profitability of crops raised in a vertical irrigation system compared to those raised in conventional irrigation systems must also be considered. Only if profit is maximized or loss minimized compared to old irrigation systems can a species of plant be cultured in a vertical irrigation process.

That is why the present investigation uses neuro-fuzzy techniques to estimate the nonlinear interrelationships that exist in the prediction of crop suitability under such specialized irrigation systems.

7.1.1 Brief Methodology

Fuzzy logic, a recent innovation in searching for optimal solutions to complex problems, is used to determine the weighting of each variable according to its importance in determining the suitability of a crop being cultivated under a vertical irrigation systems.

The weighting was used to determine the weighted average of the variables, which is directly related to the suitability of a particular plant species for cultivation under a nonconventional irrigation system.

The weighted average and value of all the input variables are categorized such that each category represents the level of suitability of the crop due to the quantitative or qualitative situation of the variable. For example, the length of the root can be encoded into nine categories from Extremely Small to Extremely Large. The category Extremely Small will have the highest level of suitability with respect to other categories of the variable. Again, for the variable parasite tolerance, Extremely High will represent the higher suitability of the crop compared to the other categories for cultivation in a vertical irrigation system.

The categorized data of the variables will help to create a combinatorial data matrix considering each possible combination that may exist among the input and output variables. Each of the categories was rated according to its impact on the selection mechanism. The rankings were then used to determine the weighted average. This weighted average was also categorized in such a way that Extremely High represents greater suitability of the crop than other categories and Extremely Small represents the opposite characteristic of the sample species.

A neurogenetic model was then used to formulate the relationship between input and output variables in such a way that if the category of the input variables is given, the model predicts the level of suitability of the plant species for cultivation under a vertical irrigation system. Because the neurogenetic model will be trained by the combinatorial data matrix, the model can predict the suitability of any combination of input variables. The model is validated by performance metrics like precision, sensitivity, specificity (CEBM 2012), and kappa index of agreement (GANFYD 2012).

7.1.2 Fuzzy Logic

Zadeh et al. discovered the benefit of following fuzzy logic in solving practical decision-making problems. Fuzzy logic nowadays is applied in various field such as the selection of a reservoir operating mechanism, watershed management, dam break analysis, and other optimization and clusterization problems.

Fuzzy logic was used in this study to determine the weighting that represents the role of variables in identifying species' suitability for cultivation under a vertical irrigation system.

7.1.3 Neurogenetic Models

Neural network models follow the signal-transmission mechanism of the human nervous system. The capability of the human nervous system to identify complex, nonlinear, and uncertain situations is mimicked in solving real-life decision-making problems. Analysis of recent studies shows that neural networks have been applied to various topics in engineering, medicine, science, and the arts (Table 7.1).

Such networks were used either separately or coupled with other sophisticated and high-end linear, nonlinear, and metaheuristic algorithms for prediction, clusterization, and function approximation problems (Table 7.2)

The problem independency, data flexibility, and efficiency in mapping complex collinearity has made the neural network one of the most sought-after modeling techniques, and it was thus applied in the present investigation for the assessment of crop suitability.

Neural networks are often referred to as black boxes and, due to the requirements for large amounts of data and computational infrastructure, applications of neural networks are generally limited to problem domains where a satisfactory database is available and supported by high-end computational facilities. The selection of network topology and activation function along with algorithms for updating the weights is made using either a trial-and-error method or by the application of various search algorithms like genetic or swarm or simulated annealing, as shown in Table 7.2. As was correctly pointed out by Duin (2000) and Jain and Nag (1998), no universally accepted methodology is available yet to guide a developer of such models in the selection of the three most important parameters (network topology, activation function, and input weights), which inherently affect the accuracy level of the developed models.

7.1.4 Genetic Algorithm

A genetic algorithm (GA) is a popular search algorithm that replicates the crossover mechanism of meiotic cell division where traits of the parent are transferred to the new cell. Some traits become dominant, some dormant, but few new traits are formed due to mutation. In practical problem solving, GAs are applied to searching for the optimal solution from among the many available ones. Available solutions are compared with genes that are transferred to the child cell and from the available solution 100 or 1,000 new solutions are randomly generated. Among the newly generated solutions, the five or ten best solutions are selected with the help of fitness functions. The selected optimal sets of solutions are then used in crossing over where traits of one solution are combined with other traits of another solution to generate a new solution. The optimal solution is identified with the help of the same fitness function, but from the new set of generated solutions. The crossover mechanism is not mimicked when optimal solutions are identified with asexual GAs.

Table 7.1 Examples of neural network applications in different fields of science, arts, and engineering

Authors	Subject	Main objective
Xia et al. (2012)	Power system, electrical Engineering	Development of electric load simulator including effects of superfluous torque
Wei et al. (2012)	Water resource management, civil engineering, management science	Decision tree and neural-network-trained decision tree algorithms were compared in decision making with respect to reservoir release during typhoonlike extreme events
Rajagopalan et al. (2012)	Communication theory, electronic engineering,	Optimization of band use of communication channels
Waszczyszyn and Ziemiański (2001)	structural technology, civil engineering	Study was aimed at estimating duration of structural fatigue
Hussain (1999)	manufacturing process technology, chemical engineering	An overview of neural network applications in predictive control, inverse-model-based control, and adaptive control methods in chemical process technologies was summarized where the satisfactory performance from using a neural network for both simulation and online applications was clearly identified.
Abdullayev and Ismaylova (2011)	Radio medicine	Detection of pathologic changes in electrostimulative electromyograms of patients with normal state of neuromuscular system and for muscular syndromes such as carpal tunnel syndrome
Zeng (2012)	Diagnostic medicine	Study aimed to make an efficient EEG monitoring with neonates possible
Mafakheri et al. (2012)	Metallurgy, nanoscience	Estimation of coercivity of cobalt nanowires
Akbari et al. (2012)	Thermal science, HVAC, electrical appliances	Study tries to predict the effectiveness of latent and sensible heat transfer between exhaust and supply air of a run-around membrane energy exchanger (RAMEE) in a high voltage air-conditioning system
Golmohammadi (2011)	Economics, decision making	Identification of the relationship between criteria and alternatives solutions of a decision-making problem and ranking the alternatives based on the learned knowledge
Kumari and Majhi (2012)	Social science	Determination of an optimal classifier for gender categorization through face recognition

Although GAs have been applied with satisfactory results (Park et al. 2012; Pinthong et al. 2009; Reddy and Kumar 2006) in various problem domains where the optimal solution must be found, but due to the randomized nature of the algorithm and the lack of a predetermined methodology, problems are solved using such metaheuristic algorithms when no other solution is available.

Table 7.2 Applications of neural networks in prediction, classification, and regression or functional approximation problems

Authors	Type of application	Type of neural network	Remarks
Liu et al. (2012)	Prediction	Genetic neural network (GNN)	Evapotranspiration and runoff dynamics of two plantation was predicted with the help of GNN, back-propagation neural network and M-Stat statistical model. The results reveal that GNN has better performance efficiency, robustness, and effectiveness than the other two models
Lohani et al. (2012)	Prediction	Adaptive neuro-fuzzy inference system (ANFIS)	Reservoir inflow was predicted with the help of ANFIS, ANN, and an autoregressive technique. Comparison of the model outputs showed that ANFIS outperformed other two models
Moustra et al. (2011)	Prediction	Feed forward neural network	Magnitude of earthquake was predicted with the help of seismic electric signals, which are known indicators of major upcoming earthquakes
Wu and Chau (2011)	Prediction	ANN coupled with singular spectrum analysis (SSA)	Rainfall-runoff transformation model was developed with the help of modular artificial neural network (MANN) and ANN coupled with SSA. As SSA can remove the lag effect of prediction that generally degrades performance of ANN runoff prediction models, the model performance of the latter was found to be better than MANN

Barkhatov and Revunov (2010)	Classification	Kohonen neural network	Classification of discontinuities in space plasma parameter with respect to solar wind parameter
Yamaguchi et al. (2008)	Classification	N-version programming ensemble of artificial neural networks	Land classification of MODIS satellite imagery by an ensemble fault masking neural network classifier
Aquil et al. (2007)	Prediction	Neuro-fuzzy	Study attempted to model daily and hourly runoff patterns in a continuous-time domain. The result demonstrates satisfactory performance of developed model when compared with Levenberg–Marquardt-feedforward neural network (FFNN) and Bayesian regularization FFNN
Teegavarapu (2007)	Regression	Stochastic neural network	Neural network and Kriging interpolation of rain-gauge data were compared; performance metrics made it clear that the former is more efficient than the latter
Tomandl and Schobe (2001)	Regression	Modified general Regression neural network (MGRNN)	MGRNN was used to generate arbitrary data from a database.
Gupta and McAvoy (2000)	Prediction	Simple recurrent neural network (SRNN)	A multiclass prediction problem was solved successfully by introduction of two different SRNNs

7.2 Methodology

The suitability of crops in vertical irrigation systems was identified using fuzzy logic and neurogenetic models. Fuzzy logic was used to determine the weighting of the input variables, and neurogenetic models were used to predict crop suitability. Crop suitability was estimated based on the characteristics of the input variables represented by different categories, which reflected the different levels of intensity of the input variables.

Table 7.3 presents a step-by-step description of the important phases of the study methodology in identifying the suitability of a species in terms of being cultivated under a vertical irrigation system. Table 7.4 gives the rules that were used to categorize variable data sets and corresponding rankings for encoding the influence of variables in the objective or suitability function, which in turn represents the suitability of a species to be cultured in a vertical irrigation scheme.

Table 7.5 shows the rank and corresponding degree of importance of the same, which were later used to determine the weighting of the variables.

The fuzzy categories of the input variables were established using the rule described in Table 7.5. Inference is performed by comparing each variable with the others and assigning the variable to the given fuzzy category according to Table 7.5.

The rule matrix and corresponding membership function are shown respectively in Tables 7.6 and 7.7. The membership function was determined by dividing the row of the rule matrix by the maximum or worst possible rank assigned for that row.

In the rule matrix, the rank of the importance of the variable with respect to the other variables will be shown in each row. If a variable has a higher importance than the variable with which it is compared, then the cell under the compared variable will have a higher rank (1, 2, …) and vice versa. Thus, when the worst rank received by a row is divided, the lowest value of the operation will be assigned to those variables in comparison to which the present variable has received a high rank and the large value of the result will be for those that are more important than the present variable in the context of the present problem.

Thus, the lowest value of the division will have the highest importance received by the variable and the highest value from the same operation will have the lowest importance assigned to the present variable. The lowest value is inverted to determine the weighting of the given input variable so that the influence due to the importance becomes proportional to the objective of the suitability function.

In the present study, the defuzzification procedure was not required as the authors were interested in predicting a categorized output rather than a numerical one so that the generalization aspect of the modeling platform was not affected.

Table 7.3 Important steps in identification of suitability of a species for cultivation under vertical irrigation systems

Step	Method applied	Purpose	Justification	Tools applied
1	Application of fuzzy logic	To determine the weighting of each variable with respect to its influence on the electability of a species for vertical cultivation	Fuzzy logic is well known for its ability to search for the optimal solution from among the many available. In the present problems also no predetermined weighting was available due to the lack of an adequate number of published studies. As fuzzy logic's theory of maximization has been applied to determine the weighting for similar objectives it was later selected to accomplish the present step	Aggregation
2	Categorization of the variable data set according to Table 7.4	To prepare a combinatorial data matrix and to remove any discrepancy that may arise due to the difference in scale of the variables	Categorization of the data set of a variable converts the variable into a dimensionless value where only the degree of intensity or superficiality of the variable was represented. Such a data set was found to be useful in training neural network models as it does not understand the inherent physical relationship of the variables; instead, the objective of such meta-heuristics was to identify the inherent empirical relationships that exist between variables	IF THEN Rule

(continued)

Table 7.3 (continued)

Step	Method applied	Purpose	Justification	Tools applied
3	Conversion of the categorized variable into its numerical form based on ratings	To determine the value of an objective function, the categorized variable was converted into its numerical counterpart based on the ratings given in Table 7.4	This step was found to be necessary to estimate a value of the suitability function that represents the suitability of the species for cultivation under vertical irrigation system	$S = \text{Suitability function}\ \frac{\sum_{n=i}^{i} n_i \times V_i}{\sum n \times \sum V} \quad (7.1)$ where n_i is the weighting that was determined with the help of fuzzy logic in Step 1, V_i is the numerical value of the variable derived from Table 7.4, and i represents the number of variables considered for the study
4	Conversion of the output into categorized form	As the objective of the model was to predict the suitability instead of predicting the output in numerical terms, which would be highly scale dependent, the suitability was estimated in a categorized form so that the relativity of the suitability could be defined in a deterministic manner	As the objective was to estimate suitability, a categorized form will always be more generalized (Lebowitz 1985)	The rule given in Table 7.4 was followed while categorizing the output

5	Development of neurogenetic model	The model was developed for the prediction of the output variable, i.e., suitability function S. The main purpose of the model development was to introduce a uniform platform of prediction that is free from objective bias that may be introduced by fuzzy logic	Neurogenetic model was found to perform better than the statistical and other nonlinear models in the case of prediction with a categorized data set (Kulkarni et al. 2008; Doğan et al. 2007)	Simple feedforward neural network model trained by conjugate gradient descent algorithm and topology identified by genetic algorithm along with a logistic activation function was applied for predicting the interrelationship between the categorized variables
6	Validation of developed model	The model was validated with the help of precisions Sensitivity, specificity and kappa index of agreement by comparing the actual and predicted category of the output variable	The validation of a model depicts the reliability of its performance capability. The selected performance metrics was found to be a better representative of model accuracy (Midgley et al. 1969; Linder et al. 2008)	Precision-sensitivity-specificity (Tan et al. 1999) Kappa index of agreement (Laurent et al. 2008; Van Den Eeckhaut et al. 2006)

(continued)

Table 7.3 (continued)

Step	Method applied	Purpose	Justification	Tools applied
7	Case study with rice (*Oryza sativa*) and maize (*Zea mays*)	Model was tested for a real-life problem where species like rice and maize were evaluated for use under vertical irrigation systems	The case study can represent the reliability of the developed methodology so that the study can be confidently reapplied to solve similar problems	The developed neurogenetic model, along with the required data describing the input variables for rice and maize crops

Table 7.4 Methods of categorization of variables

Level of intensity	Category	Rating (for variables directly proportional to suitability function)	Rating (for variables partially proportional to suitability function)	Rating (for variables inversely proportional to suitability function)
Extremely intense (values between 86 and 100% of max value)	EH	9	5	1
Very intense (values between 76 and 85% of max value)	VH	8	4	2
Highly intense (values between 66 and 75% of max value)	H	7	3	3
Semi highly intense (values between 56 and 65% of max value)	SH	6	2	4
Neither intense nor superficial (values between 46 and 55% of max value)	N	5	1	5
Semi superficial (values between 36 and 45% of max value)	SL	4	2	6
Highly superficial (values between 26 and 35% of max value)	L	3	3	7
Very superficial (values between 16 and 25% of max value)	VL	2	4	8
Extremely superficial (values less than 15% of max value)	EL	1	5	9

Table 7.5 Degree of importance and corresponding ranks assigned to variables based on their importance compare to other variables (fuzzification)

Rank	Degree of importance
1	Highly important
2	Important
3	Neither important nor unimportant
4	Unimportant
5	Highly unimportant

Table 7.6 Rank of importance assigned to different input variables of proposed model (rule matrix)

	L_R	R_S	N	T	W	Pa	A	P
L_R	0	3	2	1	3	1	3	3
R_S	3	0	2	2	4	3	3	3
N	4	4	0	3	4	3	5	5
T	5	4	3	0	3	3	2	5
W	3	2	2	3	0	2	3	4
Pa	5	3	3	3	4	0	3	4
A	3	3	1	4	3	3	0	3
P	3	3	1	1	2	2	3	0

Table 7.7 Weighting assigned to each variable following application of theory of minimization (membership function)

	L_R	R_S	N	T	W	Pa	A	P	Minimum	Wtg
L_R		1.00	0.67	0.33	1.00	0.33	1.00	1.00	0.33	0.67
R_S	0.75		0.50	0.50	1.00	0.75	0.75	0.75	0.50	0.50
N	0.80	0.80		0.60	0.80	0.60	1.00	1.00	0.60	0.40
T	1.00	0.80	0.60		0.60	0.60	0.40	1.00	0.40	0.60
W	0.75	0.50	0.50	0.75		0.50	0.75	1.00	0.50	0.50
Pa	1.00	0.60	0.60	0.60	0.80		0.60	0.80	0.60	0.40
A	0.75	0.75	0.25	1.00	0.75	0.75		0.75	0.25	0.75
P	1.00	1.00	0.33	0.33	0.67	0.67	1.00		0.33	0.67

7.3 Results and Discussion

The steps described in Table 7.4 were followed to develop a suitability function and a model to predict it with the help of fuzzy logic and a neurogenetic model, respectively. Table 7.6 shows the rank of importance assigned to each of the variables with respect to the others. The importance represented by each rank was already shown in Table 7.5. Table 7.7 presents the value of the weighting assigned to each of the variables according to the theory of maximization rule.

The combinatorial data set was used for training the neurogenetic model so that it could predict any possible unknown combination of the input variables. The parameters of the neurogenetic model and performance as represented by the metrics were shown in Table 7.8.

The values of the performance metrics (Table 7.8) show the level of accuracy of the model. As the precision, sensitivity, specificity, and kappa of the model prediction compared to the actual categories of the output were respectively 93.93, 98.04, 99.62 and 96.78%, it can be concluded that the model is robust and effective enough to be used to predict the suitability function (Eq. 7.1).

Table 7.8 Parameters of neurogenetic models and performance metrics used for validation of simulation platform

Model parameter	Value received
Type of network	Feedforward with logistic activation function
Topology selection algorithm	Sexual genetic algorithm (GA)
GA parameters	
Number of people allowed	60
Number of generations allowed	50
Rate of crossover allowed	0.80
Rate of mutation allowed	0.20
Network topology	8-16-1
Network weight	144
Training algorithm	Conjugate gradient descent (CGD)
CGD parameters	
Generalization loss allowed	±50.00%
Outliers allowed	±70.00%
Training accuracy (A) allowed	98.00%
Number of iterations allowed	1,000,000
Number of retrains allowed	10
Category encoding	9
Testing accuracy	98.15%
Training accuracy	98.69%
Model validation parameters	
Precision	93.93%
Sensitivity	98.04%
Specificity	99.62%
Kappa index of agreement	96.78%

Table 7.9 Categories assigned to variables and prediction results from model

Input variables	Rice	Maize
L_R	VL	N
R_S	VL	N
N	H	L
T	H	H
W	VH	N
Pa	H	N
A	L	L
P	H	VH
Output variable: S	**SH**	**VH**

As discussed in the methodology section, the developed model was applied to predict the suitability of rice and maize for cultivation under a vertical irrigation system. Table 7.9 shows the characteristics of rice (Morita and Nemoto 1995; Agrocommerce can be consulted for a thorough description of the cultivation of *Oryza* sp.) and maize (for an in-depth description of the morphology and cultivation characteristics of maize, FAO manuals can be consulted) represented by the model input. The result of the suitability function is also given in Table 7.9.

For example, in the case of rice, the lowland variety was chosen, and in the case of maize, the yellow variety was selected for the present investigation.

In terms of climatic variables, only conditions required in the growing season were considered.

According to the prediction results, *Oryza* sp. had a suitability score of Semi High, whereas that of *Zea* sp. was predicted to be Very High for vertical irrigation schemes. Although rice has a very short root length and spread compared to maize due to its very high water requirements compare to maize (normal water requirement), the degree of suitability of rice was predicted to be lower than that of maize. Also, maize's profitability score was higher than rice's. Both characteristics of the most important input variables (LR and P were assigned the highest weighting among other variables) was found to favor maize, which explains the prediction result from the neurogenetic model.

7.4 Conclusion

The present investigation proposed a methodology for selecting suitable crop species that could be cultivated under vertical irrigation systems. The features that a crop to be cultivated under such systems should or must have were first identified from various studies and government manuals. Once the features were identified, fuzzy logic was used to determine the weighting of each variable. Once the weighting was determined, a neurogenetic model was prepared to predict suitability through a suitability function. Categorized data considering all possible combinations of the input and output variables were used to train the model, and crop suitability was also predicted in a categorized form so as to maintain the generalized nature of the decision support system. Rice and maize were tested with the modeling platform where the model correctly predicted the lower suitability of rice compared to maize for cultivation under a vertical irrigation scheme. The present study tried to highlight the necessity of evaluating crops in terms of their suitability for vertical farming. The conclusions derived from the model could save both money and energy for a farmer planning to initiate indoor agricultural projects. As a further development of the study, various methods of categorization by objective classification can be experimented.

References

Abdullayev NT, Ismaylova KS (2011) Use of neural networks for recognition of pathological changes in stimulative electromyograms. Biomed Eng 45(6):201–206

Akbari S, Hemingson HB, Beriault D, Simonson CJ, Besant RW (2012) Application of neural networks to predict the steady state performance of a run-around membrane energy exchanger. Int J Heat Mass Transf 55(5–6):1628–1641

Aqil M, Kita I, Yano A, Nishiyama S (2007) A comparative study of artificial neural networks and neuro-fuzzy in continuous modeling of the daily and hourly behaviour of runoff. J Hydrol 337(1–2):22–34

Barkhatov NA, Revunov SE (2010) Neural network classification of discontinuities in space plasma parameters. Geomagn Aeron 50(7):894–904

Centre For Evidence Based Medicine (2012) SpPins and SnNouts. Retrieved from http://www.cebm.net/index.aspx?o=1042. 25 May 2012

Doğan E, Yüksel İ, Kişi Ö (2007) Estimation of total sediment load concentration obtained by experimental study using artificial neural networks. Environ Fluid Mech 7(4):271–288. doi:10.1007/s10652-007-9025-8

Duin RPW (2000) Learned from neural network. Retrieved from http://homepage.tudelft.nl/a9p19/papers/. 25 May 2012

GANFYD (2012) Statistical test for agreement. Retrieved from http://www.ganfyd.org/**index**.php?title=Statistical_tests_for_**agreement**. 25 May 2012

Golmohammadi D (2011) Neural network application for fuzzy multi-criteria decision making problems. Int J Prod Econ 131(2):490–504

Gupta L, McAvoy M (2000) Investigating the prediction capabilities of the simple recurrent neural network on real temporal sequences. Pattern Recognit 33(12):2075–2081

Hussain MA (1999) Review of the applications of neural networks in chemical process control – simulation and online implementation. Artif Intell Eng 13(1):55–68

Jain BA, Nag BN (1998) A neural network model to predict long-run operating performance of new ventures. Ann Oper Res 78(0):83–110. doi:10.1023/A:101891040273

Janga Reddy M, Nagesh Kumar D (2006) Optimal reservoir operation using multi-objective evolutionary algorithm. Water Resour Manag 20(6):861–878

Kulkarni MA, Patil S, Rama GV, Sen PN (2008) Wind speed prediction using statistical regression and neural network. J Earth Syst Sci 117(4):457–463

Laurent J-M, François L, Bar-Hen A, Bel L, Cheddadi R (2008) European bioclimatic affinity groups: data-model comparisons. Global Planet Chang 61(1–2):28–40

Lebowitz M (1985) Categorizing numeric information for generalization. Cognit Sci 9(3):285–308

Liu Z, Peng C, Xiang W, Deng X, Tian D, Zhao M, Yu G (2012) Simulations of runoff and evapotranspiration in Chinese fir plantation ecosystems using artificial neural networks. Ecol Model 226:71–76

Lohani AK, Kumar R, Singh RD (2012) Hydrological time series modeling: a comparison between adaptive neuro-fuzzy, neural network and autoregressive techniques. J Hydrol 442–443:23–35

Mafakheri E, Tahmasebi P, Ghanbari D (2012) Application of artificial neural networks for prediction of coercivity of highly ordered cobalt nanowires synthesized by pulse electrodeposition. Measurement 45(6):1387–1395

Midgley AR Jr, Niswender GD, Rebar RW (1969) Principles for the assessment of the reliability of radioimmunoassay methods (precision, accuracy, sensitivity, specificity). Eur J Endocrinol 142:163–184. doi:10.1530/acta.0.062S163

Morita S, Nemoto K (1995) Structure and function of roots. In: Proceedings of the fourth international symposium on structure and function of roots, Kluwer, Stará Lesná, Slovakia, pp 75–86, 20–26 June 1993

Moustra M, Avraamides M, Christodoulou C (2011) Artificial neural networks for earthquake prediction using time series magnitude data or seismic electric signals. Expert Syst Appl 38(12):15032–15039

Park CH, Joo JG, Kim JH (2012) Integrated washland optimization model for flood mitigation using multi-objective genetic algorithm. J Hydro Environ Res 6(2):119–126

Pinthong P, Gupta AD, Babel MS, Weesakul S (2009) Improved reservoir operation using hybrid genetic algorithm and neurofuzzy computing. Water Resour Manag 23(4):697–720

Tan EM, Smolen JS, McDougal JS, Butcher BT, Conn D, Dawkins R, Fritzler MJ, Gordon T, Hardin JA, Kalden JR, Lahita RG, Maini RN, Rothfield NF, Smeenk R, Takasaki Y, Van Venrooij WJ, Wiik A, Wilson M, Koziol JA (1999) A critical evaluation of enzyme immunoassays for detection of antinuclear autoantibodies of defined specificities: I. Precision, sensitivity, and specificity. Arthritis Rheum http://dx.doi.org/10.1002/1529-0131(199904)42:3<455::AID-ANR10>3.0.CO;2-3, doi:10.1002/1529-0131(199904)42:3%3c455::AID-ANR10%3e3.0.CO;2-3

Teegavarapu RSV (2007) Use of universal function approximation in variance-dependent surface interpolation method: an application in hydrology. J Hydrol 332(1–2):16–29

Tomandl D, Schober A (2001) A Modified General Regression Neural Network (MGRNN) with new, efficient training algorithms as a robust 'black box'-tool for data analysis. Neural Netw 14(8):1023–1034

Van Den Eeckhaut M, Vanwalleghem T, Poesen J, Govers G, Verstraeten G, Vandekerckhove L (2006) Prediction of landslide susceptibility using rare events logistic regression: a case-study in the Flemish Ardennes (Belgium). Geomorphology 76(3–4):392–410

Waszczyszyn Z, Ziemiański L (2001) Neural networks in mechanics of structures and materials – new results and prospects of applications. Comput Struct 79(22–25):2261–2276

Wei C-C, Chen L, Hsu H-H (2012) Neural-based decision trees classification techniques: a case study in water resources management. Lecture notes in electrical engineering, 1. Recent Adv Comput Sci Info Eng 124:377–382

Wu CL, Chau KW (2011) Rainfall–runoff modeling using artificial neural network coupled with singular spectrum analysis. J Hydrol 399(3–4):394–409

Xia L, Ruifeng Y, Jianfang J (2012) The control system of electric load simulator based on neural network. In: Advances in future computer and control systems. Advances in intelligent and soft computing, vol 159, pp 681–687

Yamaguchi T, Mackin KJ, Nunohiro E, Park JG, Hara K et al (2008) Artificial neural network ensemble-based land-cover classifiers using MODIS data. Artif Life Robot 13(2):570–574

Zeng Z (2012) Advances in intelligent and soft computing. Advances in computer science and information engineering, vol 169, pp 431–436

Chapter 8
Application of Neuro-Fuzzy Techniques in the Estimation of Extreme Events

Mrinmoy Majumder and Rabindra Nath Barman

Abstract In hydroclimatic science, a hydrologic or climatic event like a flood or rainfall is said to be extreme if its occurrence is rare or the probability of its occurrence is below 5%. Predicting extreme events is a difficult task, and often conceptual models fail to perform optimally while predicting the time and frequency of extreme events. Due to this drawback, scientists are now opting for nature-based algorithms to make predictions about extreme events. The application of neural networks, along with the categorization ability of fuzzy logic, has been found to perform much better than conceptual models. The present study uses the same concept to develop a model that can predict the occurrence and frequency of extreme events with the help of a data set categorized by the application of fuzzy logic.

Keywords Extreme events • Neuro-fuzzy systems • Combinatorial data matrix

8.1 Introduction

Artificial neural networks (ANNs) are a popular method of prediction and categorization that mimics the signal-transmission mechanism in the human nervous system. In such a network, layers of inputs are connected to layers of output, just as axons are connected to dendrites in a nerve cell. All the inputs are attached to a weight layer that is continually updated to attain the desired level of accuracy in

M. Majumder (✉)
School of Hydro-Informatics Engineering, National Institute of Technology Agartala, Barjala, Jirania 799055, Tripura, India
e-mail: mmajumder15@gmail.com

R.N. Barman
Department of Production Engineering, National Institute of Technology Agartala, Barjala, Jirania 799055, Tripura, India
e-mail: rahul.barman@yahoo.co.in

M. Majumder and R.N. Barman (eds.), *Application of Nature Based Algorithm in Natural Resource Management*, DOI 10.1007/978-94-007-5152-1_8,

obtaining the targeted values of the output; this resembles the signal-transmission process that occurs among interconnected nerve cells. Once an input signal, like a change in temperature on a stove burner, is detected by receptor cells, a signal is transmitted to the brain for a desired response. The brain, based on its experience, communicates its response, which will certainly be to remove the hand from the cause of the change. The neural network also has a third layer besides the input and output layers – the hidden layer, which works as a buffer between the two main layers. The hidden layers are just duplicate connections of the input and output layers, so experience gained from repeated iterations becomes replicated.

Thus, with a larger number of hidden layers, the model will gain more experience in a single iteration. But too many hidden layers can make the iteration procedure lengthy and computationally extensive, which is undesirable, making the selection of the number of hidden layers rather confusing and complex. Various decision-making techniques have been implemented to find a logical solution to this problem, but so far, the trial-and-error method is the most widely followed in the selection of network topologies.

On the other hand, the rarity of extreme events has made predicting them extremely difficult and complex. The accuracy of any model, whether conceptual or statistical, will depend on the commonality of the output variable. Neural networks are popular for their prediction accuracy, even in the case of complex problems, such methods are now widely used to predict rare events.

It also has been found that the accuracy of a model developed from a set of clustered or categorized data is more than a model created with the help of a numerical data set. Fuzzy logic is widely used for categorizing sets of data retaining the inherent characteristics of the data. Thus, in this study, neural networks and fuzzy logic were used to predict extreme events.

8.1.1 Prediction of Extreme Events

"Climate is defined not simply as average temperature and precipitation but also by the *type*, *frequency* and *intensity* of weather events" (USEPA). Climate change induced by global warming has the potential to change the probability and severity of extremes such as heat waves, cold waves, storms, floods, and droughts.

As reported in the Intergovernmental Panel on Climate Change Fourth Assessment Report (IPCC 2007), the number of extreme events, especially in the tropics, has greatly increased in magnitude and frequency. For example, the number of heat waves, the area of regions affected by droughts(due to marginal decreases in precipitation and increases in evaporation), the number of heavy daily precipitation, and the intensity, frequency, and duration of tropical storms have increased substantially since the 1950s (USEPA).

But predicting such events under a changing climate is complex and rather impossible, whereas understanding the intensity, probability, and frequency of the changes can help us estimate and prepare for the threats to human health, society, and the environment.

8.1.2 Objective and Scope

The objective model will have the following variables as input:

1. Intensity of rainfall (P)
2. Probability of rainfall (P_p)
3. Frequency of rainfall (P_f)
4. Previous day's evapotranspiration (ET)
5. Probability of evapotranspiration (ET_p)
6. Frequency of evapotranspiration (ET_f)
7. Previous day's humidity (H)
8. Probability of humidity (H_p)
9. Frequency of humidity (H_f)
10. Previous day's wind speed (W)
11. Probability of wind speed (W_p)
12. Frequency of wind speeds (W_f)
13. Average of last 5 day's rainfall (P_5)
14. Average of last 5 day's evapotranspiration (ET_5)
15. Average of last 5 day's humidity (H_5)
16. Average of last 5 day's wind speed (W_5)
17. Types of cloud cover (C_t)
18. Amount of cloud cover (C)
19. Probability of cloud cover (C_p)
20. Frequency of cloud cover (C_f)

There is one output variable:

1. Probability of rainfall (P_p)

8.1.3 Brief Methodology

The entire data set of variables will be categorized with respect to the concept of maximization and minimization under the fuzzy theory of categorization. All the categorized data will be used in predicting the category of the output variable.

The categorical data of the input variables, which can be referred to as a decision matrix, will be fed to the neural network model to predict the output variable.

The decision matrix will have all possible situations that can arise in the near future. Whenever the probability of an extreme event must be estimated, the data value of the input variables will be converted into corresponding groups and then may be fed to the neural network model for prediction. The predicted group will show the occurrence of extreme events in the desired time, space, geophysical, and hydroclimatic domains.

8.2 Artificial Neural Networks and Fuzzy Logic

8.2.1 Artificial Neural Network

An ANN is a pattern identification iteration methodology that mimics the procedures of the human nervous system in responding to a stimulus. The model is flexible and can be applied to any type of problem using available data sets of the input and output variables. Neural networks are applied in various techniques and follow different logic in a wide variety of fields in the arts, science, and engineering. It is widely accepted as a simple but efficient model development methodology with a high level of accuracy.

In neural networks the input layers are multiplied by a weight, and the summation of this weighted sum is converted into a function (logistic, sigmoidal, etc.) to estimate the output. The output is compared with the observed data set for accuracy. If the estimation fails to reach the desired accuracy, then the weights are updated according to a logic known as a training algorithm and the entire process is repeated. In this way, until and unless the accuracy reaches the desired level or a certain number of iterations are conducted, the model continues to minimize the error by optimization of the weights.

But the main distinction of neural networks with respect to nonlinear models is the introduction of hidden layers, which enables a model to replicate its estimation. Hidden layers act as a buffer between the input and output layers. When a hidden layer is introduced, the output becomes the hidden layer and moreover, it becomes the input with respect to the output layer. The estimation work is performed two or three times depending on the number of hidden layers. But embedding too many hidden layers will also increase the requirements for computational power, which is undesirable. Thus, selection of the topology is done in steps that are accomplished either by trial and error or with the help of specialized search algorithms. Methods for updating weights and choosing an activation function are also determined by trial and error. The drawbacks of a neural network lie in these trial-and-error methods, and many studies have been conducted on how to overcome these shortcomings.

8.2.2 Fuzzy Logic

Fuzzy logic is one of the most popular technologies for the development of decision support modules. The capability of fuzzy logic resembles human decision making with its ability to generate precision from approximations. It successfully compensates the gap in engineering design methods produced by “purely mathematical approaches and purely logic-based approaches in system design” (Aziz and Parthiban 1996).

While equations are required to model real-world behaviors, fuzzy design can help to include the ambiguities of real-world human language and logic.

The initial applications of fuzzy theory include process control for cement kilns, the first fuzzy-logic-controlled subway of Sendai in northern Japan (1987), elevators to reduce waiting time, etc. After the initial decade of the 1980s, applications of fuzzy logic in different technologies increased at an alarming rate, affecting the things we use every day.

Some of the noticeable applications of fuzzy logic in essential durable goods include the fuzzy washing machine, which uses fuzzy logic to select the best cycle, the identification of the right time at the proper temperature in a fuzzy microwave, and a fuzzy car with automaneuvering technology.

Fuzzy logic was derived from the fact that most modes of human reasoning are approximate in nature.

The theory of fuzzy logic was first developed by Professor Lofti Zadeh at the University of California in 1965. At that time application of fuzzy logic made the following assumptions:

- In fuzzy logic, exact reasoning is viewed as a limiting case of approximate reasoning.
- In fuzzy logic, everything is a matter of degree.
- Any logical system can be fuzzified.
- In fuzzy logic, knowledge is interpreted as a collection of elastic or, equivalently, fuzzy constraints on a collection of variables.
- Inference is viewed as a process of propagating elastic constraints.

8.2.2.1 Application of Fuzzy Logic

Fuzzy logic has been applied in various fields where decision making with the help of linear or nonlinear conceptual or statistical models was found to be erroneous. For example, the control systems of the 165 MWe Fugen advanced thermal reactor in Tokyo, Japan was embedded with fuzzy logic so that proper decisions could be made by the controller without any manual interference but based on system uncertainty (Iijima et al. 1995). The Steam drum water level had been controlled by proportional-integral control systems, but after the development of fuzzy logic, the water level became more effectively regulated.

Fuzzy logic was applied to control the demand for electric water heater power in such a way that the on-peak demands were shifted to the off-peak period. To achieve that objective, Nehrir and La Meres (2000) divided the entire distribution area into multiple blocks, with each block separately regulated by a fuzzy controller. The use of the fuzzy controller was found to be beneficial in achieving the objectives through demand-side management (DSM).

In the case of a water treatment mechanism, fuzzy logical operators were used to model the time-variant specific fluxes during crossflow microfiltration of several feed suspensions. According to the model output, it was found that fuzzy logic controllers could become an efficient regulator in programmable control systems for improved onsite operation of membrane-based liquid–solid separation (Altunkaynaka and Chellam 2010).

In 1998 fuzzy logic was used to model the advective flux of Atrazine in unsaturated calcareous soil (Freissineta et al. 1998) for dealing with the imprecise estimates that are normally present in water and solute movement in soils. Fuzzy logic was also applied in the development of a strategy to shift the average power demand of residential electric water heaters by selecting a minimum temperature as the control variable (LaMeres et al. 1999); in estimating "environ-metrics," i.e., the interrelations of air, water, and land ecosystems (Astel 2006); in the development of a new water quality index based on fuzzy logic; in a literature review and hydrographic survey for the Ribeira de Iguape River in the southwestern part of São Paulo State, Brazil (Lermontova et al. 2009); in the creation of an index to evaluate surface water quality by converting the traditional, discontinuous classes into continuous forms, where the summed values of the available data (with respect to the latter method of classification) were defuzzified to estimate the actual value of the index (Lcaga 2006). Benlarbi et al. (2004) developed an online fuzzy optimization algorithm that was used to maximize the drive speed and water discharge rate of the coupled centrifugal pump of a photovoltaic water pumping system driven by a separately excited DC motor (DCM), a permanent magnet synchronous motor (PMSM), or an induction motor (IM) coupled to a centrifugal pump. Monthly water consumption was estimated successfully by Yurduseva and Firat (2008) with the help of fuzzy logic with an adaptive neural network [Adaptive Neuro-Fuzzy Inference System (ANFIS)].

Table 8.1 presents more applications of fuzzy logic for solving problems in the field of water research.

8.3 Methodology

The present study will try to highlight a methodology for estimating the probability and frequency of extreme events. As discussed earlier, due to the uncertainty involved in the occurrence of extreme events, it is rather complex to define the exact amount or occurrence of such events. But their zone of certainty can be predicted if the required amount of data is available.

8.3.1 Data Set Preparation

At first the parameters related to the occurrence of extreme events were identified based on the available studies, governmental reports, and expert opinion. In the selection of the input variables the latter was given the maximum weighting due to the experts vast field experience. The variables were then encoded into nine groups representing the different levels of quantity and quality. Figure 8.1 represents the groups in which the variables were distributed.

Table 8.1 Some applications of fuzzy logic to different water-related problems

Authors	Field of water research	Model objective	Success rate
Schulz and Huwe (1998)	One-dimensional, steady-state water flow in unsaturated zone of homogeneous soils	Describe soil water pressures with depth and calculate evapotranspiration rates under steady-state conditions	Impact of different shapes of membership functions of input parameters on resulting membership functions
Ocampo-Duquea et al. (2006)	Water quality	Assess water quality by developing an index	Successful
en and Altunkaynak, (2009)	Water consumption rate (important for planning requirements of drinking water)	Predict water consumption rate using body weight, activity, and temperature of consumers	Classification based on crisp numbers was avoided by introducing fuzzy sets
Saruwatari and Yomota (2000)	Irrigation water management	Predict irrigation water requirements using categorized data generated in consultation with various experts and knowledge base	Successful
Karabogaa et al. (2007)	Reservoir operation	Control operation dynamics of spillway gates	Successful
Khalid (2003)	Operation management	Temperature control of water bath using neuro-fuzzy compensator and back-propagation neural network	Successful

Once the data set was categorized, all possible combinations between the input variables and output were generated and used to prepare a combinatorial data set representing all the possible scenarios that could arise for the present interrelationships between input and output variables. The categorization of any data set must be performed in such a way that the lower categories like L (Low), SL (Semi Low), VL (Very Low), and EL (Extremely Low) will have values that are small but not rare and the higher categories must represent data possessing opposite characteristics.

8.3.2 Development of Scoring Mechanism by Fuzzy Logic

The scoring of the categories to rate the data was performed with the help of fuzzy logic's theory of minimization. In this theory, a fuzzy matrix was prepared by

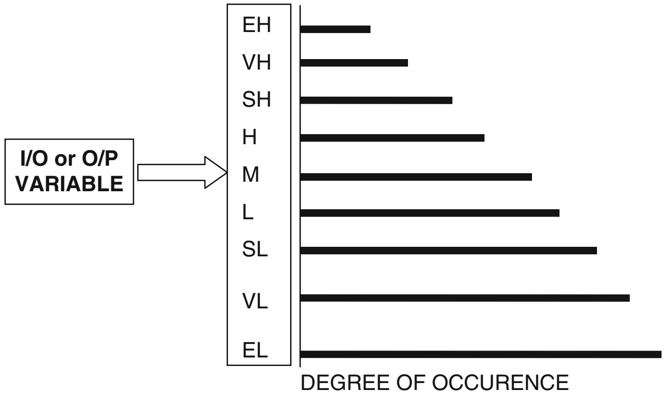


Fig. 8.1 Groups in which the data set must be encoded (*EH* highest degree of quantity or quality, *EL* lowest level of quantity and quality)

comparing each input variable with the others according to the degree of importance with respect to each other. The degree of importance was represented by the following classes:

1. Very Important (VI)
2. Important (I)
3. Neither Important nor Unimportant (NIUI)
4. Unimportant (UI)
5. Completely Unimportant (CUI)

After the input variables were rated, they were all ranked according to their importance in ascending order, i.e., a variable with a VI rating was ranked 1, whereas a variable with a CUI rating was assigned a rank of 5.

After ranking the variables with respect to their importance, the rankings were divided by the worst rank achieved by a given variable. The minimum value of the result from the division represents the highest importance achieved by the variable compared with the other variables. This minimum value was taken as the scale to rate the present variable.

All the variables were assigned a rank with respect to their individual categories such that a higher rank implies a higher probability and frequency of an extreme event and vice versa.

The objective function used to represent the probability of extreme events was developed by summing all the scales and the respective scores based on the scale of the input variables and deducing the percentage of the input variable with respect to the total value of the scale. The percentage will be directly proportional to the probability of the extreme events. After the percentage value of the objective function was derived, it was again categorized, where EH indicates the maximum probability and EL represents the opposite.

With the input and output variables a combinatorial data matrix was used to develop a neural network model with Levenberg-Marquardt as the training algorithm, a logistic activation function, and an exhaustive trial-and-error method for identifying the optimal network topology where all the input variables were used to predict the category of the output.

The results show the probability of extreme events with respect to the given scenarios as represented by the categories of the input variables. The methodology can be replicated in any study area to determine the probability of occurrence of an extreme event.

8.4 Results and Discussion

Table 8.2 shows the fuzzy matrix developed for scoring the input variables with respect to their importance compared to the other input variables. Table 8.3 represents the neural network parameters, and Table 8.4 shows the sensitivity, specificity, and precision of the neural network output.

According to the performance metrics (like sensitivity, specificity, and precision), the model had values of 99.02, 99.98, and 99.51%, respectively, for precision, sensitivity, and specificity, which shows the model's level of accuracy.

8.5 Conclusion

The present study represents an attempt to predict the probability of occurrence of extreme events with the help of 20 selected variables. The variables are encoded into nine categories and rated using fuzzy logic. The rated input variables were used to develop an objective function representing the likelihood of the extreme events. The function was then encoded into nine categories that were similar to the input categories. The categorized input and output variable was used to generate all the possible combinations of the variables and a combinatorial data matrix was produced. This data matrix was used to create a neural network model to predict the outcome of the combinations of the input variables. Once the model is trained to predict the chance of occurrence of extreme events for all possible scenarios, it can be used to predict extreme events based on the input variables in any given study area. Due to a lack of time, the model was not tested for real-time situations, but the authors would be interested in receiving results in real-time settings from the research community that might help to further improve the model's efficiency.

Table 8.2 Fuzzy ratings of each criteria with respect to each other based on the five point fuzzy scale

	1. Intensity of rainfall (P)	2. Probability of the rainfall (P_p)	3. Frequency of the rainfall (P_f)	4. Previous day evapo-transpiration (ET)	5. Probability of the evapo-transpiration (ET_p)	6. Frequency of the evapo-transpiration (ET_f)	7. Previous day humidity (H)	8. Probability of the humidity (H_p)	9. Frequency of the humidity (H_f)
1. Intensity of rainfall (P)	0	I	I	I	I	I	VI	I	I
2. Probability of the rainfall (P_p)	UI	0	NINU	NINU	NINU	NINU	I	NINU	NINU
3. Frequency of the rainfall (P_f)	UI	UI	0	NINU	NINU	NINU	I	NINU	NINU
4. Previous day evapo-transpiration (ET)	UI	NINU	NINU	0	I	I	NINU	NINU	NINU
5. Probability of the evapo-transpiration (ET_p)	UI	NINU	NINU	UI	0	I	I	I	I
6. Frequency of the evapo-transpiration (ET_f)	CUI	UI	UI	NINU	UI	0	I	I	I
7. Previous day humidity (H)	CUI	UI	UI	NINU	UI	UI	0	I	I
8. Probability of the humidity (H_p)	UI	NINU	NINU	NINU	UI	UI	UI	0	I
9. Frequency of the humidity (H_f)	UI	NINU	NINU	NINU	UI	UI	UI	UI	0
10. Previous day wind speed (W)	UI	NINU	NINU	UI	UI	UI	UI	UI	UI
11. Probability of the wind speed (W_p)	UI	NINU	NINU	UI	UI	UI	UI	UI	UI
12. Frequency of the wind speed (W_f)	UI	NINU	NINU	UI	UI	UI	UI	UI	UI
13. Average of last five days rainfall (P_5)	UI	NINU	NINU	UI	UI	UI	UI	UI	UI
14. Average of last five days evapo-transpiration (ET_5)	NINU	I	I	UI	UI	UI	UI	UI	UI
15. Average of last five days humidity (H_5)	NINU	I	I	UI	UI	UI	UI	UI	UI
16. Average of last five days wind-speed (W_5)	NINU	I	I	UI	UI	UI	UI	UI	UI
17. Types of cloud cover (C_t)	NINU	I	I	I	I	NINU	NINU	NINU	NINU
18. Amount of cloud cover(C)	CUI	UI	UI	I	I	NINU	NINU	NINU	NINU
19. Probability of the cloud cover (C_p)	UI	NINU	NINU	I	I	NINU	NINU	NINU	NINU
20. Frequency of the cloud cover (C_f)		UI	NINU	NINU	I	I	NINU	NINU	NINU

10. Previous day wind speed (W)	11. Probability of the wind speed (W_p)	12. Frequency of the wind speed (W_f)	13. Average of last five days rainfall (P_5)	14. Average of last five days evapo-transpiration (ET_5)	15. Average of last five days humidity (H_5)	16. Average of last five days wind-speed (W_5)	17. Types of cloud cover (C_t)	18. Amount of cloud cover(C)	19. Probability of the cloud cover (C_p)	20. Frequency of the cloud cover (C_f)
I	I	I	I	NINU	NINU	NINU	NINU	VI	I	I
NINU	NINU	NINU	NINU	UI	UI	UI	UI	I	NINU	NINU
NINU	NINU	NINU	NINU	UI	UI	UI	UI	I	NINU	NINU
I	I	I	I	I	I	I	UI	UI	UI	UI
I	I	I	I	I	I	I	UI	UI	UI	UI
I	I	I	I	I	I	I	NINU	NINU	NINU	NINU
I	I	I	I	I	I	I	NINU	NINU	NINU	NINU
I	I	I	I	I	I	I	NINU	NINU	NINU	NINU
I	I	I	I	I	I	I	NINU	NINU	NINU	NINU
0	I	I	I	I	I	I	NINU	NINU	NINU	NINU
UI	0	I	I	I	I	I	NINU	NINU	NINU	NINU
UI	UI	0	I	I	I	I	NINU	NINU	NINU	NINU
UI	UI	UI	0	I	I	I	UI	UI	UI	UI
UI	UI	UI	UI	0	I	I	UI	UI	UI	UI
UI	UI	UI	UI	UI	0	UI	UI	UI	UI	UI
UI	UI	UI	UI	UI	UI	0	UI	UI	UI	UI
NINU	NINU	NINU	I	I	I	I	0	NINU	NINU	NINU
NINU	NINU	NINU	I	I	I	I	NINU	0	I	I
NINU	NINU	NINU	I	I	I	I	NINU	UI	0	I
NINU	NINU	NINU	I	I	I	I	NINU	UI	UI	0

Table 8.3 Neural network parameters of the model

Neural network parameters	Model result
Topology search method	Exhaustive nonlinear
Activation function	Logistic
Training algorithm	Levenberg-Marquardt
Network topology	20-6-1
Network weight	126
Training data set	12157665459056928000 data rows
Stop training condition: correct classification rate (CCR)	97%
Stop training condition: minimum improvement in error	0.0000001 during last 10 iterations
Stop training condition: allowed number of iterations	1,000,000
Training correct classification rate (CCR)	98.55%
Testing correct classification rate (CCR)	99.98%

Table 8.4 Performance metrics of neural network model

Performance metric	Model result (%)
Precision	99.02
Sensitivity	99.98
Specificity	99.51

References

Altunkaynaka A, Chellam S (2010) Prediction of specific permeate flux during crossflow microfiltration of polydispersed colloidal suspensions by fuzzy logic models. doi:http://dx.doi.org/10.1016/j.desal.2009.10.018

Astel A (2006) Chemometrics based on fuzzy logic principles in environmental studies. doi:http://dx.doi.org/10.1016/j.talanta.2006.09.026

Aziz SA, Parthiban J (1996) Fuzzy logic. Retrieved from http://www.doc.ic.ac.uk/~nd/surprise_96/journal/vol4/sbaa/report.fuzzysets.html. Feb 2012

Benlarbia K, Mokranib L, Nait-Said MS (2004) A fuzzy global efficiency optimization of a photovoltaic water pumping system. doi:http://dx.doi.org/10.1016/j.solener.2004.03.025

Fayea RM, Sawadogoa S, Lishoua C, Camino FM (2002) Long-term fuzzy management of water resource systems. doi:http://dx.doi.org/10.1016/S0096-3003(02)00151-0

Freissineta C, Erlicha M, Vauclin M (1998) A fuzzy logic-based approach to assess imprecisions of soil water contamination modeling. doi:http://dx.doi.org/10.1016/S0167-1987(98)00067-1

Iijima T, Nakajimaa Y, Nishiwaki Y (1995) Application of fuzzy logic control system for reactor feed-water control. doi:http://dx.doi.org/10.1016/0165-0114(95)00036-K

IPCC (2007) Climate change 2007: impacts, adaptation, and vulnerability. In: Parry ML, Canziani OF, Palutikof JP, van der Linden PJ, Hanson, CE (eds) Contribution of working group II to the fourth assessment report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, 1000 pp

Karabogaa D, Bagisb A, Haktanir T (2007) Controlling spillway gates of dams by using fuzzy logic controller with optimum rule number. doi:http://dx.doi.org/10.1016/j.asoc.2007.01.004

Khalid A (2003) Adaptive fuzzy control of a water bath process with neural networks. doi:http://dx.doi.org/10.1016/0952-1976(94)90041-8

LaMeres BJ, Nehrir MH, Gerez V (1999) Controlling the average residential electric water heater power demand using fuzzy logic. doi:http://dx.doi.org/10.1016/S0378-7796(99)00022-X

Lcaga Y (2006) Fuzzy evaluation of water quality classification. doi:http://dx.doi.org/10.1016/j.ecolind.2006.08.002

Lermontova A, Yokoyamab L, Lermontovc M, Augusta M, Machado S (2009) River quality analysis using fuzzy water quality index: Ribeira do Iguape river watershed, Brazil. doi:http://dx.doi.org/10.1016/j.ecolind.2009.02.006

Nehrir MH, LaMeres BJ (2000) A multiple-block fuzzy logic-based electric water heater demand-side management strategy for leveling distribution feeder demand profile. doi:http://dx.doi.org/10.1016/S0378-7796(00)00124-3

Ocampo-Duquea W, Ferré-Huguetb N, Domingob JL, Schuhmachera M (2006) Assessing water quality in rivers with fuzzy inference systems: a case study. doi:http://dx.doi.org/10.1016/j.envint.2006.03.009

Saruwatari N, Yomota A (2000) Forecasting system of irrigation water on paddy field by fuzzy theory. doi:http://dx.doi.org/10.1016/0378-3774(95)92338-F

Schulz K, Huwe B (1998) Water flow modeling in the unsaturated zone with imprecise parameters using a fuzzy approach. doi:http://dx.doi.org/10.1016/S0022-1694(97)00038-3

Sen Z, Altunkaynak A (2009) Fuzzy system modelling of drinking water consumption prediction. doi:http://dx.doi.org/10.1016/j.eswa.2009.04.028

United States Environmental Protection Agency (2012) Extreme events, climate change-health and environmental effects. Retrieved from http://epa.gov/climatechange/effects/extreme.html. Feb 2012

Yurduseva MA, Firat M (2008) Adaptive neuro fuzzy inference system approach for municipal water consumption modeling: an application to Izmir, Turkey. doi:http://dx.doi.org/10.1016/j.jhydrol.2008.11.036

Chapter 9
Impact of Climate Change on Selection of Sites for Lotus Cultivation

Mrinmoy Majumder and Rabindra Nath Barman

Abstract Lotus (*Nelumbo nucifera*) cultivation provides livelihoods to many people in the tropical and subtropical regions of the world. Lotus is commercially produced in different altitudes. The flower is cultivated in Asia, Australia, North America, and Egypt. Due to demand from the medicinal, fragrance, and culinary industries, as well as from religious sects, cultivation of lotus has become a common source of economic stability due to stable market demand. The harvesting of lotus suffers due to extreme weather patterns, insect attacks, quality degradation of harvesting ponds, and overuse of fertilizers. To ensure optimal production of lotus selection and to control the impact of such inhibitors, the selection of a suitable site is extremely important. That is why farmers invest considerable sums in acquiring ideal ponds for lotus production. Even then, problems still arise: inhabitants of the pond's command area react aggressively to the bad odor produced by rotten tubers, ponds are breeding grounds for mosquitoes, and the rapid spread of lotus stems can prevent local aquatic population from prospering. Thus, the present study was initiated to provide a framework for lotus farmers in which to select a suitable pond to optimize production and maximize profit. The model framework uses neurogenetic models to predict the suitability of a pond for cultivation of the lotus plant.

Keywords Neurogenetic models • Lotus cultivation • Selection mechanism

---

M. Majumder (✉)
School of Hydro-Informatics Engineering, National Institute of Technology Agartala, Barjala, Jirania 799055, Tripura, India
e-mail: mmajumder15@gmail.com

R.N. Barman
Department of Production Engineering, National Institute of Technology Agartala, Barjala, Jirania 799055, Tripura, India
e-mail: rahul.barman@yahoo.co.in

M. Majumder and R.N. Barman (eds.), *Application of Nature Based Algorithm in Natural Resource Management*, DOI 10.1007/978-94-007-5152-1_9,

9.1 Introduction

Lotus (*Nelumbo* sp.) is an aquatic species in order Proteales and family Nelumbonaceae. The plant has wide floating leaves and bright fragrant flowers. The leaves and flowers of the aquatic Plantae (kingdom) have long stems that contain air spaces that float in water. The root functions of the lotus are carried out by a modified root, rhizomes that fan out horizontally through the mud below the water. The leaves are up to 50 cm in diameter, the flowers are rosy pink with little bits of white shade, and the seeds are hard and dark brown. They can vary in shape from round to oval to oblong. The lotus flower is a morning bloom and is a well-known astringent used in treating kidney, liver, spleen, sleeping, and sexual disorders.

The lotus flower is widely used as a raw material in the fragrance industry. It is an important material for different religious sects. The seeds and stems are edible and used in the culinary industry. The medical industry also makes use of the flowers, stems, seeds, and rhizomes to develop medicines for the treatment of various diseases of the gastroenteric and reproductive systems.

The problems with lotus cultivation can be divided into two major classes:

1. Problems with cultivation methodology
2. Hostility from local inhabitants

Problems with the cultivation methodology include extreme temperature (both high and low temperatures inhibit the growth of the plant), abnormal hydrologic conditions like flood or drought, and the depth of the water body (shallow ponds are not suitable for lotus cultivation). The soil must be humid and moist. The water body must be rich in nutrients, and the pH of the water must be maintained in the neutral zone; acidic water is not suitable for lotus cultivation.

The other class of problems includes opposition from local residents because the rotten stems and roots of lotus generate a pungent odor, and the rapid growth of lotus stems and leaves can engulf an entire pond, which may inhibit the growth of other aquatic creatures. The cultivation of lotus will also attract insects like mosquitoes and can cause an outbreak of malaria or dengue.

Thus the location of lotus ponds is important and ensures optimal production of the flower without being affected by local opposition and other factors related to soil, water depth, and temperature. This investigation tries to estimate the suitability of a pond for lotus cultivation with the help of an artificial neural network and genetic algorithm.

9.1.1 Objective and Scope

The present investigation will try to define the suitability of a pond for cultivation of lotus plants in such a way that the major problems associated with the harvesting procedure can be minimized and maximum profit can be ensured. Because all the problems connected with lotus production are location dependent, the selection of a suitable pond is very important.

The scope of the study also includes the determination of a scientific methodology for the selection of a pond for lotus culturing. The framework can help agro-based industries

Table 9.1 Input and output variables considered for present neuro-genetic model for prediction of pond suitability

Input	Output
Temperature (T)	Suitability of pond
Humidity (H)	
Concentration of fish (F_c)	
Existing utility of pond (P_u)	
Size of dependant population (P)	
Pond depth (h)	
Pond area (A)	
Bottom soil (S_o)	
Turbidity (Tu) of pond water	
Salinity (S) of pond water	
Dissolved oxygen (DO) of pond water	
Presence of predatory species (Sp)	
Proximity to residential complexes (d_r)	
Proximity to market (d_m)	
Availability of labour (L)	
Availability of fertilizers (F)	
Availability of electricity (E)	
Price in market (C)	

and individual farmers to adjust their location selection procedure according to the model output. The location of the pond will help to minimize local opposition, bring extremes weather and water conditions under control, and help optimize production of lotus flowers by taking advantage of the appropriate local geophysical properties.

9.1.2 Brief Methodology

Table 9.1 shows the variables selected as input and output. The categorized data set of the variables were fed to the model to predict the ideal location. The categorization was performed based on the degree of influence of the variables upon the output variable. The categories were also rated according to their relationship with the cultivation of lotus plants. The ratings were used to estimate a weighted average, which in turn predicted the suitability of a pond for lotus culture (suitability function).

The weights were determined using nonlinear optimization methods as no previous studies were found from related fields. All the input variables were assigned a random weight, which was varied within a predetermined boundary. The boundary condition was determined based on discussions with experts and feedback from farmers. The objective function was the average weighting itself, which was maximized by varying the weighting of each variable. The scale dependency of the weights were removed by normalizing the same. That is why the value of the weighting does not vary above or below 1 and 0, respectively.

After the optimal weighting was determined, the suitability function or average weighting derived from the different combinations of input variables was also categorized and used as output data for the training data set. The function was categorized in such a way that it represents its relationship with the selection of the pond, i.e., the higher the average weighting, the higher the suitability of a pond for lotus cultivation, and, conversely, the lower the average weighting, the lower the pond's suitability for lotus cultivation.

A combinatorial data matrix was then created where all possible combinations of input variables were considered, and the output was estimated with the help of neurogenetic models, which are known for their efficiency in mapping of highly nonlinear relations. The universal data matrix developed earlier was then used for training the neural nets so that the model could learn the inherent nature of the relationships that exist between the deciders and output variables.

If a lotus farmer has picked out several locations as potential lotus ponds, he or she can evaluate the suitability of the water body using the model. Such knowledge-based decisions can yield fruitful results and reduce the risk of incurring losses from the harvest.

9.1.3 Neurogenetic Models

The ability to map nonlinear relationships between input and output variables and the ease with which problems can be conceptualized and simulated have made neurogenetic models one of the most desired algorithms for data mining and simulation objectives. Table 9.2 presents some of the studies that examine the efficiency of neurogenetic models in solving optimization, functional approximation, simulation, prediction, classification, and clusterization problems.

The data dependency and universality in application of neural networks have often incited critics of neural networks to call them black boxes. Neural nets are also plagued with deficiencies like requirements for large data populations, problems associated with topology selection, and weighting of the update algorithm and activation function by trial and error. In addition, ignorance regarding the physical interrelationships between variables has led some critics to charge that the reliability of models developed with the help of neural network algorithms is somehow degraded.

But the introduction of search algorithms like genetic algorithm, ant colony optimization, artificial bee colony algorithm, and particle swarm optimization for the selection of parameters like model topology, activation function, etc., which are generally selected by stochastic methods, has also enriched the application of combined models in complex nonlinear objectives.

In the present study neural network and genetic models were combined to develop an algorithm for accurate and reliable decision making.

9.1.4 Genetic Algorithm

Genetic algorithms (GAs) are popular search algorithms (Table 9.3) that use the theory of natural selection during crossover of genes undergoing meiotic cell division. Phenomena like the introduction of new traits (mutation) in the reproduced population are also mimicked to increase the efficacy of the algorithm.

Table 9.2 Sample applications of neurogenetic algorithm in different problem-solving approaches

Reference	Comparative model	Purpose	Input	Remark
Monjezi et al. (2012)	Multivariate regression models	Prediction	Characteristics of rocks	Uniaxial compressive strength was predicted with the help of neurogenetic models and multivariate regression analysis to avoid the cost and time of preparing laboratory samples. The prediction results were compared and according to the performance metrics, coefficient of determination and mean square error, the neurogenetic model was better with respect to the multivariate regression model
Ganesan et al. (2011)	Genetic programming	Optimization	Geophysical and operational constraints	Genetic and neurogenetic algorithms were used for optimization of geologic structure mapping. Minimization of the objective function was performed by GA first separately then combined with a neural network model and results clearly showed the superiority of the latter approach over the former
Singh et al. (2007)	Coactive neuro-fuzzy inference system (CANFIS)	Prediction	Temperature and stress	The creep strain of rock was predicted with the help of neurogenetic and CANFIS models for a comparative analysis
Sahoo and Maity (2007)	None	Prediction	Frequency and strain of structure	Location and amount of damage were estimated with the help of neurogenetic models where the topology of the neural network was identified with the help of a GA. The results encourage further use of such models in prediction problems

Table 9.3 Sample applications of GA in different problem-solving approaches

References	Type of genetic algorithm	Purpose	Input	Remark
Ou (2012)	GA coupled GM model	Forecasting	Agricultural output of Taiwan from 1998 to 2010	GA-coupled GM and individual GM were compared with respect to its prediction of Taiwan's agricultural output. Based on the performance metrics like mean absolute percentage error and the root mean square percentage error, it was concluded that GA-coupled GM was only better than GM for in- and out-sample data sets
Cao et al. (2012)	Boundary-based fast genetic algorithm (BFGA)	Reference point optimization or goal programming or searching for an optimal option nearest to the goal from a set of solutions	Economic, environmental, and ecological benefits, social equity including Gross Domestic Product (GDP), conversion cost, geological suitability, ecological suitability, accessibility, "not in my back yard" (NIMBY) influence, compactness, and compatibility	Land use of Tongzhou Newtown in Beijing, China, was optimized with the help of BFGA, which is a modified form of a sexual GA. The GA model helped to search and identifies the near-optimal solution from the available 10,000 land use plans

Panagopoulos et al. (2012)	Soil water assessment tool (SWAT)-coupled GA algorithm (applied with the help of MATLAB-GA toolbox and SWAT software) attached to an economic function model for cost estimation	Optimization	Changes in livestock, crop, soil, and nutrient application management in alfalfa, corn, and pastureland fields	GA, along with SWAT and an economic model, was used to identify the most optimal pair within a location and the best management plan based on constraints like reduction of diffused pollutants and maintaining an optimal cost-benefit ratio. The result yielded an optimal pair that reduces the pollutants total phosphorus by 45% and nitrogen as nitrates by 25% at a cost of 25 euro/person
Shiri et al. (2012)	Gene expression programming (GEP)	Forecasting	Air temperature, relative humidity, wind speed, and solar radiation	Daily reference evapotranspiration was predicted with help of GEP, adaptive neuro-fuzzy inference system (ANFIS), Priestley–Taylor, and Hargreaves–Samani models. On comparing with different performance metrics it was found that GEP, followed by ANFIS, outclassed all other models
Yun et al. (2010)	Noisy GA based on simple GA	Stochastic optimization	Reservoir inflow	Reservoir inflow was optimized with help of NGA and Monte Carlo method. The optimization results identified the former as a better method for stochastic optimization with respect to the latter model
Ines et al. (2006)	Sexual GA coupled to soil water atmosphere plant (SWAP) and surface energy balance algorithm for land (SEBAL)	Data assimilation and water management optimization	Sowing dates, irrigation practices, soil properties, depth to groundwater, and water quality	Different water management options were optimized with the help of the SWAP and SEBAL generated distributed data set of the input variables on which basis the water management options were optimized with the help of a GA

The methodology of GAs is simple where each input variable has a weighting that is varied randomly to find an optimal solution for a given objective. A fitness function is also used to estimate the accuracy of the solutions. The function is used to identify a population of specific size that yielded the best solution to the given problem. The weightings are referred to as genes. There are two types of genetic algorithm: sexual and asexual. In sexual GAs, the weightings of the best solutions are interchanged and the solutions are evaluated based on the accuracy as determined by the fitness function. As exchanges of genes are conducted, new traits are formed along with an increase in the likelihood of introducing new paths for a quick but efficient problem searching approach.

These exchanges of genes are also referred to as crossover, which mimics the crossover of genes in a human cell.

In an asexual algorithm, the weighting of genes of the same solution is varied within the genes themselves; for example, the first one can become the last one, and vice versa. The generation continues until and unless the output from the fitness function does not improve.

The weighting value at which the fitness function for both algorithms does not change can be identified as the optimal solution and used in achieving the present objective.

GAs are also widely used for the identification of optimal parameters of neural nets, for example, the number of hidden layers or the types of training algorithms to update weightings and activation functions which enhance the reliability of such models. Besides improving the neural net performances, GAs have also been used in various optimization and clusterization problems either separately or with other modeling algorithms.

9.2 Methodology

Table 9.4 presents the inputs selected to predict the suitability of a body of water for the cultivation of lotus plants. The table also shows the categories in which the variables were encoded and their corresponding values in scale of importance, were determined in such a way that the objective function becomes directly proportional to the suitability of the location for lotus cultivation.

A GA, along with quick propagation (QP), was used to identify the optimal topology and weighting that enabled the model to learn the inherent relationships between the input and output variables. The activation function was selected with the help of a logistic function.

The sensitivity, specificity, and precision were selected as performance metrics of the model.

The existing utility of the pond variable represents the use of the pond by the dependent population. If the pond is used for fisheries or any other purpose, then that pond is generally avoided due to the probability of hostility from the dependent population. Thus, the EH category for this variable was given a low rating, whereas the EL category was assigned a high rating. Again, proximity to residential complexes can spur opposition from residents due to the hazards associated with lotus

Table 9.4 Model variables and categories they were encoded in and corresponding ratings

Variable	Category	Crisp Counterpart of Encoded Categories
Temperature (T)	EH,VH,SH,H,N,L,SL,VL,EL	1,2,3,4,5,4,3,2,1
Humidity (H)	EH,VH,SH,H,N,L,SL,VL,EL	1,2,3,4,5,6,7,8,9
Concentration of fish (F_c)	EH,VH,SH,H,N,L,SL,VL,EL	1,2,3,4,5,4,3,2,1
Existing utility of the pond (P_u)	EH,VH,SH,H,N,L,SL,VL,EL	1,2,3,4,5,6,7,8,9
No. of dependant population (P)	EH,VH,SH,H,N,L,SL,VL,EL	1,2,3,4,5,6,7,8,9
Pond depth (h)	EH,VH,SH,H,N,L,SL,VL,EL	1,2,3,4,5,4,3,2,1
Pond area (A)	EH,VH,SH,H,N,L,SL,VL,EL	9,8,7,6,5,4,3,2,1
Bottom soil (S_o)	CLAY,CLAY-HUMOUS,HUMOUS,LOAMY, LOAM-SAND, SAND	1,2,4,3,2,1
Turbidity (Tu) of pond water	EH,VH,SH,H,N,L,SL,VL,EL	1,2,3,4,5,6,7,8,9
Salinity (S) of pond water	EH,VH,SH,H,N,L,SL,VL,EL	1,2,3,4,5,6,7,8,9
Dissolved oxygen (DO) of pond water	EH,VH,SH,H,N,L,SL,VL,EL	1,2,3,4,5,4,3,2,1
Presence of predatory species (Sp)	EH,VH,SH,H,N,L,SL,VL,EL	1,2,3,4,5,6,7,8,9
Proximity to residential complexes (d_r)	EH,VH,SH,H,N,L,SL,VL,EL	9,8,7,6,5,4,3,2,1
Proximity to market (d_m)	EH,VH,SH,H,N,L,SL,VL,EL	1,2,3,4,5,6,7,8,9
Availability of labour (L)	EH,VH,SH,H,N,L,SL,VL,EL	9,8,7,6,5,4,3,2,1
Availability of fertilizers (F)	EH,VH,SH,H,N,L,SL,VL,EL	9,8,7,6,5,4,3,2,1
Availability of electricity (E)	EH,VH,SH,H,N,L,SL,VL,EL	9,8,7,6,5,4,3,2,1
Price in the market (C)	EH,VH,SH,H,N,L,SL,VL,EL	9,8,7,6,5,4,3,2,1

cultivation (increased breeding of mosquitoes and other insects, foul odor etc.).Thus the category EH received a low rating, whereas the category EL received a high rating in the case of this variable. As labor, fertilizer, and electricity availability is essential in lotus cultivation the EH, VH, SH, H categories of such variables were given higher ratings than the EL, VL, SL, and L categories. The variable presence of predatory species represents the population of species that were found to threaten the lotus plant (e.g., bacteria, arthropods, and other worms that eat lotus leaves and stems). Naturally, the EH category of this variable was assigned the lowest rating.

The rating of the categories was converted into a percentage by dividing the assigned rating by the maximum rating of the variable. Thus for a certain pond if proximity to market is very high, then the rating of the variable will be the rate assigned to the VH category, which is equal to 2.The percentage of the category will be the assigned rating (i.e., 2) divided by the maximum rating (i.e., 9), or 22.22%. The percentage value was determined to reduce the scale dependency and to introduce uniformity among the ratings of variables.

The objective function or suitability function was mathematically represented as Eq. 9.1, where S is the suitability function of a certain pond, V_i is the ith variable, n_i is the percentage rating of the ith variable, n_{m} is the maximum percentage achieved by a variable among all the considered variables, and i is the total number of variables selected for the present study.

$$S = \frac{\sum_{i=1}^{i} n_i V_i}{n_{\mathrm{m}} \sum_{i=1}^{i} V_i} \tag{9.1}$$

The suitability function was also encoded into nine categories representing the degree of electability of the pond. The higher the suitability function, the higher the electability of the pond for lotus cultivation. That is why the suitability function was categorized into four groups for simplicity [Highly Suitable (HS), Suitable (S), Unsuitable (US), and Highly Unsuitable (HUS)], where the HUS and US group was assigned respectively to suitability functions equal to or less than 0.4 and less than 0.5 but greater than or equal to 0.41. The S and HS category was respectively assigned to suitability values of less than 0.75 but greater than or equal to 0.51 and greater than or equal to 0.76, respectively.

In the second step a combinatorial data matrix was prepared where all possible combinations of the input and output variables were generated. The data were then used to train a neurogenetic model that would predict the suitability of a pond based on the category of the input variables for a given pond location.

9.3 Results and Discussion

The result of the model training and testing along with other performance metrics is given in Table 9.5. According to the model performance the kappa index of agreement between the actual and predicted results was found to be equal to 90.58%, which is commendable in view of the conservative approach of kappa measure of agreement.

Table 9.5 Characteristics of model parameters and adopted performance metrics

Model parameter	Value
Genetic algorithm parameters	
Population size	60
Number of generations	50
Crossover rate	0.80
Mutation rate	0.20
Neural network parameters	
Network topology	18-1-1
Network weight	19
Training algorithm selected	Quick propagation (QP)
QP coefficient	1.75
Learning rate	0.20
Generalization loss allowed	50.00%
Correct classification rate (CCR) desired	98.00%
Number of training iterations allowed	1,000,000
Number of retrains	10
Training CCR	98.34%
Testing CCR	99.46%
KAPPA index of agreement	90.58%

The Levenberg-Merquardt (LM) algorithm was also used, but, due to the high scale loss in generalization, the latter algorithm was rejected in favor of training the model. On the other hand, a QP algorithm was found to have training and testing CCRs of 98.34 and 99.46%, respectively, which were much better than those (72.16 and 72.09%) of LM. The kappa for LM was found to be equal to zero. The value of the agreement index clearly indicates that when trained with LM the neurogenetic model was totally unable to learn the inherent relationship between the input and output variables.

Based on the model performance metrics, the suitability of a pond for lotus cultivation was analyzed with the help of a QP-trained neurogenetic model.

After the model was developed, it was applied to predict some real-life scenarios derived from the case studies conducted on lotus cultivation. The presence and absence of urban populations and changes in the normal climate were found to be the two main factors to impact the suitability of a pond for lotus cultivation. Based on the study results five types of scenario were conceptualized.

Scenario 1: *Large-Scale Increase in Population but Climate Normal*
Scenario 2: *Medium-Scale Increase in Population but Climate Normal*
Scenario 3: *Small-Scale Increase in Population but Climate Normal*
Scenario 4: *Large-Scale Increase in Population and Change in Climate* (*IPCC A2*)
Scenario 5: *Medium-Scale Increase in Population and Change in Climate* (*IPCC B2*)

Scenarios 4 and 5 mimic the climate change scenario as proposed in the IPCC Fourth Assessment Report (2007). The other three scenarios try to represent large-, medium-, and small-scale urbanization that is already taking place in many developing countries.

Table 9.6 shows the categories assigned to the input variables to represent Scenarios 1–5.

Analysis of the model results shows that, except for Scenario 4, the model predicted suitability for all the scenarios. In representations of the scenarios using the input variables, the same category was assigned to the Bottom Soil input variable because the nature of the input mainly depends on location, not on climate or population, although it was assumed that even under population and climate change the pond would not change its geophysical characteristics. Regarding the pond depth and area variables, the category assigned to represent all the scenarios except Scenario 4 is the same because a change in population cannot change the dimensions of a pond, but climatic uncertainty can impose change in the storage capacity of the pond. Thus, in the case of Scenario 4, pond area and depth were assigned a category of VL, which represents the impact of using the water bodies for deposition of waste produced by the uncontrolled population as conceptualized by the IPCC under the A2 scenario.

An increase in population would impose scarcity on basic natural resources like land and water. Thus, most of the land area will be converted to either residential complexes or agricultural zones so that the excess population can be accommodated and a stable food supply ensured. To represent this situation, input variables like existing utility of the pond, dependent population, and proximity to residential complexes, all of which represent the existing demand for land and water resources, were assigned either EH or VH respectively for Scenarios 1 and 4 or Scenario 2. An increase in population increases the demand for land and food, but this change in turn increases requirement for labor and electricity. As demand for food is also increased, fertilizers will mostly be used for field irrigation. Thus, labor, fertilizer, and electricity were all assigned the EL category in Scenario 1.

The scarcity of electricity, labor, or fertilizers will diminish when a medium- or small-scale change in the total population takes place. In the case of a medium-scale change, there will be adequate population to cultivate lotus as well as for construction purposes. The amount of fertilizers will also be properly distributed when demand for it is regulated. The category assigned to the same three variables represents the toned-down situation of Scenarios 2, 3, and 5.

The chemical properties of a water body change according to the use of its catchment area. If the watershed is used mostly for agriculture, then the surface runoff will carry loose soil, dry grains, and inorganic fertilizers that, when deposited in a water body, increase its turbidity and decrease the dissolved oxygen (DO) level. As in the case of Scenarios 1 and 4 the watersheds will be used for agriculture, the turbidity level will certainly be higher than in any other situation, whereas the level of DO will also be reduced. Accordingly, the category of the variables representing the chemical properties of the pond was adjusted as per the scenario.

The model prediction found that only under the A2 and Large-Scale Increase in Population scenarios does the pond become unsuitable. That means, except when both climate and population change in such a way that natural resources become extremely scarce, the suitability decision for a pond with respect to lotus cultivation will not change.

Table 9.6 Representation of considered scenarios and decision predicted by neurogenetic model

Input variable	Category assigned
Scenario 1	
Temperature (*T*)	N
Humidity (*H*)	N
Concentration of fish (F_c)	L
Existing utility of pond (P_u)	EH
Size of dependent population (*P*)	EH
Pond depth (*h*)	N
Pond area (*A*)	N
Bottom soil (S_o)	Humous
Turbidity (Tu) of pond water	VH
Salinity (*S*) of pond water	VH
Dissolved oxygen (DO) of pond water	N
Presence of predatory species (Sp)	H
Proximity to residential complexes (d_r)	EH
Proximity to market (d_m)	VH
Availability of labor (*L*)	EL
Availability of fertilizers (*F*)	EL
Availability of electricity (*E*)	EL
Market price (*C*)	VH
Pond suitability	**S**
Scenario 2	
Temperature (*T*)	N
Humidity (*H*)	N
Concentration of fish (F_c)	N
Existing utility of the pond (P_u)	VH
Size of dependent population (*P*)	H
Pond depth (*h*)	N
Pond area (*A*)	N
Bottom soil (S_o)	Humous
Turbidity (Tu) of pond water	H
Salinity (*S*) of pond water	H
Dissolved oxygen (DO) of pond water	N
Presence of predatory species (Sp)	N
Proximity to residential complexes (d_r)	VH
Proximity to market (d_m)	H
Availability of labor (*L*)	H
Availability of fertilizers (*F*)	VH
Availability of electricity (*E*)	N
Market price (*C*)	H
Pond suitability	**S**
Scenario 3	
Temperature (*T*)	N
Humidity (*H*)	N
Concentration of fish (F_c)	H
Existing utility of the pond (P_u)	N

(continued)

Table 9.6 (continued)

Input variable	Category assigned
Size of dependent population (P)	N
Pond depth (h)	N
Pond area (A)	N
Bottom soil (S_o)	Humous
Turbidity (Tu) of pond water	N
Salinity (S) of pond water	N
Dissolved oxygen (DO) of pond water	H
Presence of predatory species (Sp)	N
Proximity to residential complexes (d_r)	N
Proximity to market (d_m)	N
Availability of labor (L)	N
Availability of fertilizers (F)	VH
Availability of electricity (E)	H
Market price (C)	N
Pond suitability	**S**
Scenario 4	
Temperature (T)	EH
Humidity (H)	EL
Concentration of fish (F_c)	L
Existing utility of the pond (P_u)	EH
Size of dependent population (P)	EH
Pond depth (h)	VL
Pond area (A)	VL
Bottom soil (S_o)	Humous
Turbidity (Tu) of pond water	EH
Salinity (S) of pond water	EH
Dissolved oxygen (DO) of pond water	EL
Presence of predatory species (Sp)	EH
Proximity to residential complexes (d_r)	EH
Proximity to market (d_m)	N
Availability of labor (L)	EL
Availability of fertilizers (F)	EL
Availability of electricity (E)	EL
Market price (C)	N
Pond suitability	**US**
Scenario 5	
Temperature (T)	H
Humidity (H)	L
Concentration of fish (F_c)	N
Existing utility of the pond (P_u)	N
Size of dependent population (P)	H
Pond depth (h)	N
Pond area (A)	N
Bottom soil (S_o)	Humous

(continued)

Table 9.6 (continued)

Input variable	Category assigned
Turbidity (Tu) of pond water	N
Salinity (S) of pond water	N
Dissolved oxygen (DO) of pond water	H
Presence of predatory species (Sp)	L
Proximity to residential complexes (d_r)	N
Proximity to market (d_m)	H
Availability of labor (L)	H
Availability of fertilizers (F)	N
Availability of electricity (E)	H
Market price (C)	H
Pond suitability	**S**

9.4 Conclusion

The present investigation tried to develop a modeling methodology in selecting a pond for lotus cultivation using neurogenetic models. Five scenarios were also conceptualized to analyze the impact of both climate and population change on such decisions. According to the model results, only when both climate and population change in a negative direction is the decision over pond suitability affected. In any other scenario where medium or small changes in population are observed under normal climate conditions, the model prediction for pond suitability remains unchanged. The developed model can become a tool for farmers selecting a pond for lotus cultivation. The results from the software will help farmers select the optimal pond where the probability of disturbance will be minimum. Although the model decision was found to be in agreement with actual scenarios (as found from the kappa index of agreement between the actual and predicted decisions), because the variables are represented as categories, small changes in the variables at levels below the considered range will become inconsequential for the decision. But the categorization was done in such a way that recognizable changes will only be considered in the model decision. Changes with minimum or no impact on the output variables were ignored by considering the ranges of the categories after consulting the latest scientific studies. Pond suitability can also be predicted with the help of neuro-fuzzy or neuro-swarm techniques. A comparative study could be undertaken to determine what models were available to achieve the present objective.

References

Cao K, Huang B, Wang S, Lin H (2012) Sustainable land use optimization using Boundary-based Fast Genetic Algorithm. Comput Environ Urban Syst 36(3):257–269

Ganesan T, Vasant P, Elamvazuthi I (2011) Optimization of nonlinear geological structure mapping using hybrid neuro-genetic techniques. Math Comput Model 54(11–12):2913–2922

Ines AVM, Honda K, Das Gupta A, Droogers P, Clemente RS (2006) Combining remote sensing-simulation modeling and genetic algorithm optimization to explore water management options in irrigated agriculture. Agric Water Manag 83(3):221–232

Li D, Yang HZ, Liang XF (2013) Application of Bayesian networks for diagnosis analysis of modified sequencing batch reactor. Adv Mater Res 610:1139–1145.

Monjezi M, Bahrami A, Yazdani VA (2009) Simultaneous prediction of fragmentation and flyrock in blasting operation using artificial neural network. Int J Rock Mech Min Sci 47:476–480

Monjezi M, Amini Khoshalan H, Razifard M (2012) A neuro-genetic network for predicting uniaxial compressive strength of rocks. Geotech Geol Eng (Online First™, 17 April 2012)

Nourani V, Khanghah TR, Sayyadi M (2013) Application of the artificial neural network to monitor the quality of treated water. Int J Manage Inf Technol 2(2):38–45.

Ou S-L (2012) Forecasting agricultural output with an improved grey forecasting model based on the genetic algorithm. Comput Electron Agric 85:33–39

Pai TY, Hanaki K, Su HC, Yu LF (2013) A 24_h forecast of oxidant concentration in Tokyo using neural network and fuzzy learning approach. CLEAN–Soil, Air, Water.

Panagopoulos Y, Makropoulos C, Mimikou M (2012) Decision support for diffuse pollution management. Environ Model Software 30:57–70

Sahoo B, Maity D (2007) Damage assessment of structures using hybrid neuro-genetic algorithm. Appl Soft Comput 7(1):89–104

Shiri J, Kişi Ö, Landeras G, López JJ, Nazemi AH, Stuy LCPM (2012) Daily reference evapotranspiration modeling by using genetic programming approach in the Basque Country (Northern Spain). J Hydrol 414–415:302–316

Singh TN, Verma AK, Sharma PK (2007) A neuro-genetic approach for prediction of time dependent deformational characteristic of rock and its sensitivity analysis. Geotech Geol Eng 25(4):395–407

Yun R, Singh VP, Dong Z (2010) Long-term stochastic reservoir operation using a noisy genetic algorithm. Water Resour Manag 24(12):3159–3172

Yan Z, Di T, Ye YL, Han WJ (2013) The modeling of petrochemical wastewater activated sludge system and water quality forecast based on neural network. Adv Mater Res 641:219–222.

Chapter 10
Comparison of Bat and Fuzzy Clusterization for Identification of Suitable Locations for a Small-Scale Hydropower Plant

Mrinmoy Majumder

Abstract Hydroelectric plants are an environmentally friendly renewable energy source, but, due to uncertainties in flow patterns, often such energy generation projects fail. Also, deliberations from displaced people and environmental activists (due to large-scale disturbances to the natural ecosystems of adjacent areas) make some highly efficient hydropower projects unfeasible. That is why the success of hydropower projects depends largely on the selection of location. Currently, the efficiency of selecting the ideal locations depends mainly on expert opinion or linear models and other decision-making methodologies where human judgment and opinion play a major role in the reliability of the selection. But, as usual, the error rate in such procedures is generally unsatisfactory. The present study tries to apply clusterization algorithms to identify ideal locations for small hydropower plants in such a way that the need for expert or opinion can be reduced. In the clusterization of a suitable hydropower location, the food foraging behavior of bats and fuzzy-logic-based theory of maximization were applied to a sample population of locations available for hydropower generation including a on where a hydropower plant had already been installed and was operating at rated capacity. The efficiency of the algorithm in identifying this location was analyzed to determine the suitability of the algorithms in estimating the ideal location for hydropower plants. The results showed that both approaches were able to identify the most suitable location, but when the time taken to make the identification was taken into account, fuzzy logic was found to perform better than the bat algorithm as the former took only one iteration to identify the location, whereas the latter needed six iterations but the sensitivity with which the algorithms identified the ideal location was better for bat than fuzzy.

Keywords Clusterization • Hydropower location selection • Fuzzy logic • Bat algorithm

M. Majumder (✉)
School of Hydro-Informatics Engineering, National Institute of Technology Agartala, Barjala, Jirania 799055, Tripura, India
e-mail: mmajumder15@gmail.com

M. Majumder and R.N. Barman (eds.), *Application of Nature Based Algorithm in Natural Resource Management*, DOI 10.1007/978-94-007-5152-1_10,

10.1 Introduction

Hydropower is one of the most reliable sources of energy production and does not require the use of fossil fuels. Although hydropower plants (HPPs) do not cause pollution and hydropower can be found in abundance, the problem of displacement of local populations, changes in land use and land cover of adjacent regions, and, above all, the unreliability of flow in river networks often discourage city planners from moving toward such sources of power.

This problem can be solved if the selection of a location is performed in a logical, scientific, and innovative manner whereby an area with a steady flow throughout the year but minimum displacement in population and minor disturbance to the ecosystem can be selected. But such a location is difficult to find, and generally designers must compromise on one aspect to take advantage of others. But it is essential to identify an optimal solution where wastage will be minimized but utilization of resources will be maximized.

10.1.1 Indian Scenario of Energy Distribution

India, the location of this study, has an installed power generation capacity of approximately 148,000 MW in which thermal power plants powered by coal, gas, naphtha, or oil account for approximately 66% of power generation. Hydropower is by far the single largest renewable energy source in India, accounting for more than 10% of total electricity generation. Most of this energy is from large hydroelectric plants. Renewable sources of energy other than large-scale hydropower have a 7% share, with wind power accounting for the largest share, approximately 5.29% (World Energy Assessment, UNDP (1998) and REN21 (2011a, b)).

The total additional power generation capacity planned for the 11th and 12th 5-year plans (2007–2017) is approximately 150,000 MW, of which the share of renewables such as wind, solar, biomass, and small hydro is slated to reach approximately 10% (i.e., 15,000 MW).

India is ranked fifth with respect to hydropower potential where the total economically exploitable hydropower potential is found to be 148,700 MW of installed capacity, of which the Brahmaputra basin has the largest (66,065 MW) and the Central India River system has the least (4,152 MW) potential to generate hydroelectricity. The Indus basin (33,832 MW) of Punjab and the Ganga basin (20,711 MW) are the two next largest locations with the potential for generation of hydroelectricity.

In addition, 56 pumped storage projects have been identified with a probable installed capacity of 94,000 MW. The hydro potential from small, mini, and micro schemes has been estimated to be equal to 6,782 MW from 1,512 sites. Thus, the total hydropower potential of India is found to be 250,000 MW. The total installed hydropower generation capacity of India is 36,878 MW. That means only 24.80% of the total hydroelectricity potential is being utilized. Bhakra Dam and Nagarjuna are the two major large-scale HPPs in India; they produce nearly 1,100 and 960 MW, respectively, of hydropower annually.

In India, hydropower projects with a station capacity of up to 25 MW each fall under the category of small hydropower (SHP), which is estimated to have a potential of 15,000 MW. The total installed capacity of small hydropower projects (up to 25 MW) as of 31 March 2009 is 2,429.77 MW from 674 projects, and 188 projects with aggregate capacity of 483.23 MW are under construction. That means only 16.2% of the total SHP potential is presently being utilized. The state of Karnataka, with its 83 SHP projects, generates hydroelectricity equal to 563.45 MW. India has 83.8% of SHP potential still to be utilized and has many other areas where the potential of SHP is yet to be identified.

10.1.2 Types of Hydropower Plants

Hydropower plants are generally classified based on quantity of water, water head, and nature of load.

10.1.2.1 Classification by Quantity of Water

HPPs can be classified by the amount of water used as follows:

Runoff River Plants Without Pondage

These plants have no provisions for storing water and use water as and when available. Thus, they are dependent on the rate of flow of water; during the rainy season a high flow rate might mean a certain amount of water goes to waste, while during low runoff periods, due to low flow rates, the generating capacity will be low.

Runoff River Plants with Pondage

In these plants, pondage permits storage of water during off-peak periods and use of this water during peak periods. Depending on the size of pondage provided, it may be possible to cope with hour-to-hour fluctuations. This type of plant can be used on parts of the load curve as required and are more useful than plants without storage or pondage.

This type of plant is comparatively more reliable, and its generating capacity is less dependent on the available rate of water flow (Fig. 10.1).

Reservoir Plants

A reservoir plant is one that has a reservoir large enough to permit carrying over storage from the wet season to the next dry season. Water is stored behind a dam and is available on a regulated basis to the plant as required. The plant firm capacity can be increased and can be used either as a base load plant or as a peak load plant as required. The majority of hydroelectric plants are of this type.

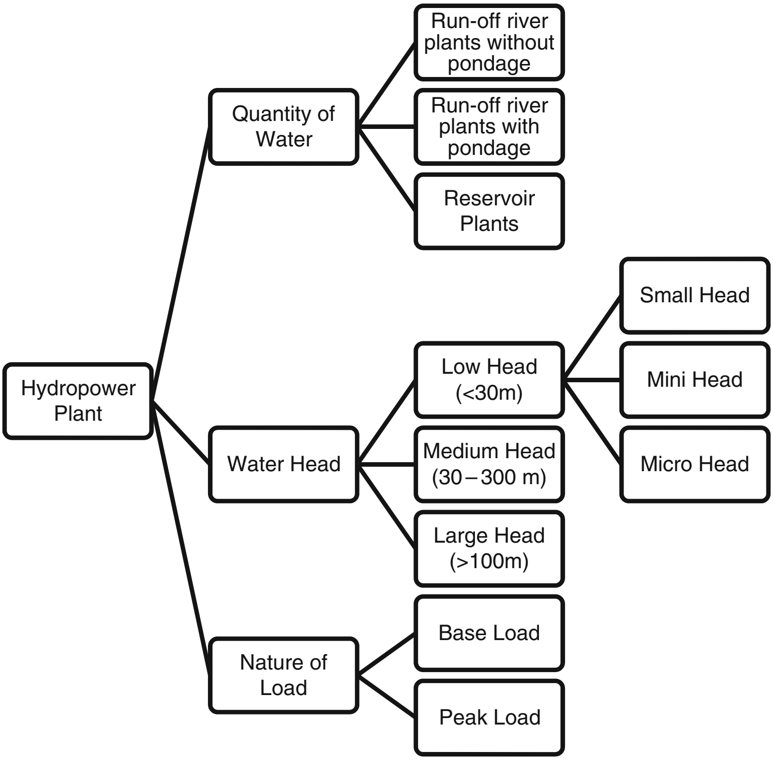


Fig. 10.1 Classification of HPPs with respect to various attributes

10.1.2.2 Classification by Availability of Water Head

Depending on the availability of the water head, HPPs can be subdivided into low-head (less than 30 m), medium-head (30–300 m), and high-head hydroelectric plants (1,000 m and above). Low-head HPPs can be further subdivided into small, mini, and micro-head hydropower plants.

10.1.2.3 Classification by Nature of Load

Classification by nature of load is as follows.

Base Load Plants

A base load power plant is one that provides a steady flow of power regardless of total power demand by the grid. These plants run at all times throughout the year, except in the case of repairs or scheduled maintenance.

Peak Load Plants

Due to their operational and economic properties, peak load plants are generate electricity during peak load periods. Gas turbines, storage and pumped storage power plants are examples of peak load power plants. The efficiency of such plants is around 60–70%.

10.1.3 Importance of Location on Use Factor of Hydropower Plants

Installation of hydropower plants generally starts with the identification of potential locations, followed by a cost-benefit analysis and environmental impact assessment. The task of trying to identify suitable locations in inaccessible areas has often hindered the growth of hydroenergy potential. Thus, GIS and remote sensing are used, as shown by Ghadimi et al. (2011), Cyr et al. (2011), Yi et al. (2010), Kusre et al. (2010), and Larentis et al. (2010). These tools are generally used to determine topographical parameters like head and storage capacity. Image processing is normally performed on satellite imagery because these tools help to extract information about the possible geophysical properties of a location without requiring a visit to the proposed site, which may be located in an inaccessible area. Such visits also incur substantial outlays for logistical support.

Some studies have also used hydrologic models like SWAT or regression equations to determine flow through proposed locations (Cyr et al. 2011; Kusre et al. 2010).

With regard to the parameters used to identify the best option, most studies considered flow and head (Dudhani et al. 2006; Rojanamon et al. 2009; Supriyaslip et al. 2009; Kusre et al. 2010; Yi et al. 2010; Cyr et al. 2011; Fang and Deng 2011). Some scientists, though they are few in number, explicitly included electrical factors like installed capacity, distance from transmission lines, load requirement, etc. Very few studies considered the social acceptance of hydropower projects (Fang and Deng 2011; Supriyaslip et al. 2009; Rojanamon et al. 2009; Nunes and Genta 1996). Environmental parameters like land use and potential pollution were considered in most of the studies, but use of the parameter in decision making was different in different studies. For example, Rojanamon et al. (2009) tried to select locations by ranking them according to their environmental and ecological sensitivity. They also considered social acceptance at the time the final decision was made. Minimum ecological flow, land use, sedimentation, and river bank erosion were considered as environmental factors in the selection of locations for a hydropower plant. Safety of the area, social conflict, legal aspects (Supriyasilp et al. 2009), and overall social acceptance (Rojanamon et al. 2009) were generally considered social factors. Most of the studies collected the necessary data from onsite surveying, focus group discussions, and review of the historical case studies.

In the final step of the decision-making process, an index was created giving weight to the different factors according to their importance in the development of hydroelectric power plants. The importance of the different factors was generally determined on the basis of the local experience and knowledge acquired from discussions with expert personnel.

After the weighting was set, an objective function was calculated using the values of the different factors and their weighting. According to the values of the objective function, suitable locations were identified for hydropower generation. Most studies on the selection of suitable locations for HPPs have used this method.

10.1.4 Cluster Analysis Algorithms

Cluster analysis, or clustering, is the task of assigning a set of objects to groups (called clusters) so that objects in the same cluster are more similar (in one sense or another) to each other than to those in other clusters. Cluster analysis, which is also referred to as data segmentation, has a variety of goals, all of which relate to grouping or segmenting a collection of objects (also called observations, individuals, cases, or data rows) into subsets, or "clusters," such that those within each cluster are more closely related to one another than to objects assigned to different clusters.

Central to all of the goals of cluster analysis is the notion of the degree of similarity (or dissimilarity) between the individual objects being clustered. There are two major methods of clustering – hierarchical clustering and k-means clustering (Table 10.1).

10.1.4.1 Hierarchical Clustering

In hierarchical clustering, data are not partitioned into a particular cluster in a single step. Instead, a series of partitions takes place, which may run from a single cluster containing all objects to *n* clusters, each containing a single object.

Hierarchical clustering is subdivided into *agglomerative* methods, which proceed by series of fusions of the *n* objects into groups, and *divisive* methods, which separate *n* objects into successively finer groupings.

10.1.4.2 K-Means Clustering

A nonhierarchical approach to forming good clusters is to specify a desired number of clusters, say k, then assign each case (object) to one of k clusters so as to minimize the measure of dispersion within the clusters. A very common measure is the sum of distances or sum of squared Euclidean distances from the mean of each cluster. The problem can be set up as an integer programming problem, but because

Table 10.1

Reference	Location	Type of HPP	Tools used	Engineering factors	Environmental factors	Socio-economic factors
Ghadimi et al. (2011)	Lorestan province in Iran	Microhydro power	GIS and local expert opinion	Flow rate and head; existing density of river and canal network	None	Only economic parameters included
Fang and Deng (2011)	Southwestern China	Cascade hydropower plant	Literature review, case studies to develop three reference scales	Per-unit drop of water level and per-unit change in flow (three scales are developed based on these two and minimum ecological flow)	Minimum ecological flow	Social status of different sections of river was considered
Cyr et al. (2011)	New Brunswick, Canada	Small conventional as well as run of river	Digital elevation models of study area to determine elevation and regression model to estimate stream flow	Stream flow, head, and penstock length	None	None
Yi et al. (2010)	Geum River basin, Korea	Small	Geospatial information system	Head, storage capacity, runoff-contributing area	Environmental assessment map was used to extract information related to environment and ecology	None
Kusre et al. (2010)	Kopili River basin in Assam, India	Micro (<0.5 MW)	Geospatial tools and SWAT hydrologic model	Weather and discharge	Soil and land use	None

(continued)

Table 10.1 (continued)

Reference	Location	Type of HPP	Tools used	Engineering factors	Environmental factors	Socio-economic factors
Larentis et al. (2010)	Brazil	Run of river and storage SHP	Remote sensing, available stream flow data, and GIS-based Hydro Spot software	Flow regulation and river basin potential for power generation	None	None
Supriyasilp et al. (2009)	Ping River basin, Thailand	Small (approx. 100 KW)	Data collection	Electrical criteria like installed capacity, annual energy production, length of transmission line, and firm load supply possible; general engineering criteria like head and flow	Flow pattern, amount of flow, loss of habitat, land use, collapse of river bank, and sedimentation	Safety in location, social conflict, water resource issues, land use issues, legal aspects, and infrastructure available
Rojanamon et al. (2009)	Upper Nan River basin, Thailand	Small run-of-river	GIS, questionnaire survey, group discussion	Average annual flow and average annual head	Geomorphological variations and presence of forests and other plantations	Opinion survey followed by focus group discussion
Dudhani et al. (2006)	India	Small, mini, and Micro	GIS and remote sensing (image segmentation based on intensity of RGB pixels)	Geophysical properties for determination of flow	None	None
Nunes and Genta (1996)	Uruguay	Small(1–5 MW), mini < (1 MW), and micro (<100 kW)	New assessment methodology, calculation and simulation tools were developed	Topographic conditions evaluated for estimating amount of power	None	Average demand for power

solving integer programs with a large number of variables is time consuming, clusters are often computed using a fast, heuristic method that generally produces good (but not necessarily optimal) solutions. The k-means algorithm is one such method.

Along with these two major clusterization algorithms, a few other specialized concepts have also been used to cluster data sets. Principal component analysis, fuzzy logic, and neural network for clusterization are also applied to classify data sets in an unsupervised manner.

10.1.4.3 Principal Component Analysis

In data mining, one often encounters situations where there are a large number of variables in a database. In such situations it is very likely that subsets of variables will be highly correlated with each other. The accuracy and reliability of a classification or prediction model will suffer if one includes highly correlated variables or variables that are unrelated to the outcome of interest. Superfluous variables can increase the data-collection and data-processing costs of deploying a model on a large database. The dimensionality of a model is the number of independent or input variables used by the model. One of the key steps in data mining is finding ways to reduce dimensionality without sacrificing accuracy.

Performance of clusterization depends largely on the characteristics of the data to be classified. There is no single classifier that works best on all problems (a phenomenon that may be explained by the no-free-lunch theorem). Various empirical tests have been performed to compare classifier performance and to find the characteristics of data that determine classifier performance. Determining a suitable classifier for a given problem is, however, still more an art than a science.

10.1.5 Objective and Scope

The identification of an optimal site for an HPP is complex and involves rigorous collection of data, field surveys, and impact analysis. That is why selection parameters are first generalized in such a way that decision making can be as easy as possible. Both traditional and nonconventional methods of site selection mainly involve prioritization of parameters, giving more weight to and ultimately making decisions based on the output value from a weighted-average formula or an activation function (Fig. 10.2).

Many studies have also considered the opinions of people affected by HPPs. The value of a given weight is defined by relevant experts and survey results. Such weighting generally ignores regional characteristics and is decided based on the experience gained from previous works. But overdependence on experience and ignorance of local features may result in an erroneous estimate of storage capacity for the proposed hydraulic structures, turbines, and generators. In this regard, the present study will propose two clusterization algorithms for selection of the most suitable sites for an HPP based on selected parameters.

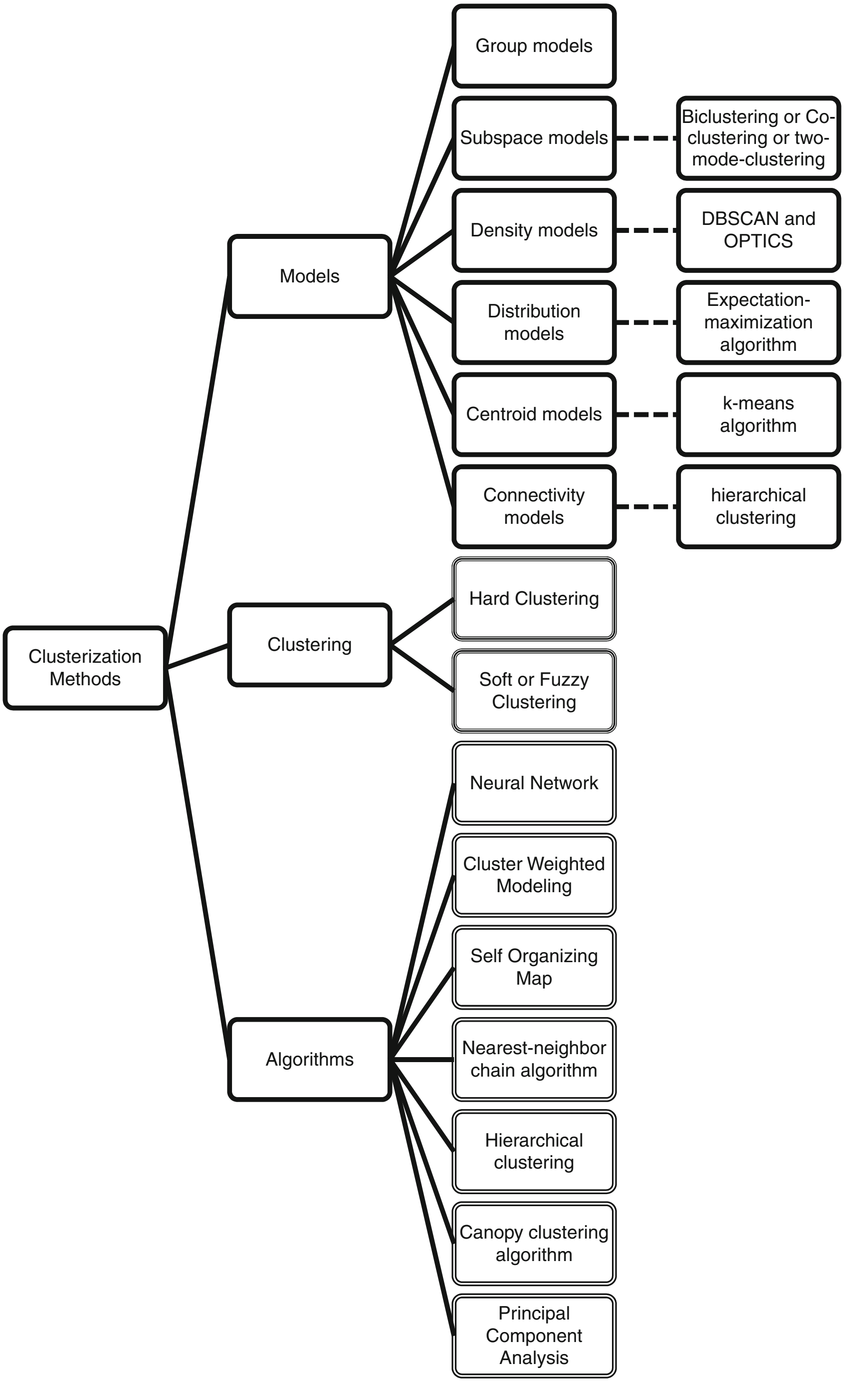


Fig. 10.2 Clusterization models and clustering methods

The main objective of the present study is to analyze the potential of automated clusterization algorithms by a bat algorithm and fuzzy logic to identify the optimal site for an SHP plant. The objective of identifying a suitable methodology for site selection is accomplished simultaneously.

10.1.6 Proposed Methodology

Some studies have also used hydrologic models like SWAT or regression equations to determine flow through proposed locations (Cyr et al. 2011; Kusre et al. 2010).

With regard to the parameters used to identify the best option, most studies considered flow and head (Dudhani et al. 2006; Rojanamon et al. 2009; Supriyaslip et al. 2009; Kusre et al. 2010; Yi et al. 2010; Cyr et al. 2011; Fang and Deng 2011). Some scientists, though they are few in number, explicitly included electrical factors like installed capacity, distance from transmission lines, load requirement, etc. Very few studies considered the social acceptance of hydropower projects (Fang and Deng 2011; Supriyaslip et al. 2009; Rojanamon et al. 2009; Nunes and Genta 1996). Environmental parameters like land use and potential pollution were considered in most of the studies, but use of the parameter in decision making was different in different studies. For example, Rojanamon et al. (2009) tried to select locations by ranking them according to their environmental and ecological sensitivity. They also considered social acceptance at the time the final decision was made. Minimum ecological flow, land use, sedimentation, and river bank erosion were considered as environmental factors in the selection of locations for a HPP. The safety of the area, social conflict, legal aspects (Supriyasilp et al. 2009), and overall social acceptance (Rojanamon et al. 2009) were generally considered social factors. Most of the studies collected the necessary data from onsite surveying, focus group discussions, and review of the historical case studies.

In the final step of the decision-making process, an index was created giving weight to the different factors according to their importance in the development of hydroelectric power plants. The importance of the different factors was generally determined on the basis of the local experience and knowledge acquired from discussions with expert personnel.

In this regard the present study tries to propose a new methodology for classifying the probable locations of HPPs on a river where dependence on expert opinion was avoided by the introduction of automatic clusterization procedures. The most ideal location for an HPP is selected based on the classification results. Classification will be performed by both bat and fuzzy clusterization methods. The results from both clusterization algorithms will be compared based on performance metrics like sensitivity, specificity, precision, and kappa index of agreement.

10.1.7 Bat Clusterization

A microbats is a type of bat that uses echo location to identify a potential food source. Microbats fly around randomly at varying speeds, emitting auditory signals from their mouths with varying loudness. When they find prey, their velocity as they fly toward the location of the prey increases along with the frequency and loudness of their pulse. The change in pulse rate is a signal to other bats regarding the potential food source and invites them to the spot.

In practical problem solving, each food source location is compared with the location of the optimal solution and the rate of convergence toward the solution increases when the fitness functions that are used to analyze the optimality of the solution acquire higher values. Both the frequency and loudness of the pulse then increases, which helps the search algorithm to possibly identify the optimal solution to the given multisolution problem.

The bat algorithm is also a nature-inspired metaheuristic identified by Yang only in 2010.

Because the algorithm is new, few applications using it are available in the scientific literature.

10.2 Methodology

10.2.1 Selection of Factors

In the selection of factors, first a thorough survey of available studies was carried out. Based on the literature survey the most important factors were selected and grouped into four categories. The following section describes the selected factors and the justification for their selection in identifying the optimal location for installing an SHP plant. The categories are as follows:

1. Hydrologic and geophysical factors
2. Environmental factors
3. Socioeconomic factors

10.2.1.1 Hydrologic and Geophysical Factors

Average Change in Flow

The average flow of a river channel is one of the most important factors in the selection of ideal locations for SHP plants or HPPs. Average flow, or Q, was included as a selection parameter in most earlier investigations. In the present study the average flow was calculated using rainfall (P) and evapotranspiration (ET), along with a loss coefficient (L) that will be described in a later section. Equation 10.1 is used to estimate the maximum possible change in flow per month. After the change in flow is calculated for each month the probability of each flow is estimated to draw a flow duration curve for that location:

$$Q = \langle P - \mathrm{ET} \rangle \times L \tag{10.1}$$

Average Change in Net Head

According to various studies, net head is also included as a factor along with average flow and storage capacity. In this study the average head is represented as the change in net head and is estimated from the difference between upstream and downstream water elevation with respect to the considered location:

$$H = \left(H_{\mathrm{u}} - H_{\mathrm{d}}\right), \tag{10.2}$$

where H is the change in head and H_{u} and H_{d} represent the head upstream and downstream, respectively. Like the previous factor, the change in head was calculated for each month and the average change in head was calculated.

Soil Strength

Soil strength was determined using porosity and soil type, where porosity is inversely proportional to the strength of the soil and the soil type is derived in such way that it becomes directly proportional to the strength of the soil. Equation 10.3 represents the soil strength:

$$s = f\left(\frac{\mathrm{ST}}{p}\right), \tag{10.3}$$

where s is the soil strength, ST the soil type, and p the unit less the porosity value.

Slope

The slope of the location is also estimated. The slope is directly proportional to the flow velocity, and a high velocity will always increase the suitability of a location for installing an HPP. In this study, the slope is calculated using the elevation difference between the maximum and minimum elevation divided by the distance between them.

10.2.1.2 Environmental Factors

Land Use and Land Cover or Loss Coefficient (L_c)

According to the environmental policies of many countries, the removal of forest cover is legal as long as adequate replacement is provided. Thus, HPPs are generally not installed in regions with large concentrations of forests. Again, the distance from a forest dictated whether the profit from the HPP could compensate the displacements of forest and the dependent inhabitants. Forests provide shelter to wildlife and livelihoods to the people living there. If forest cover is severely depleted due to the installation of an HPP and if the resulting gain from the HPP is less than the benefits provided by forest products, then the region is not recommended. In the present study, forest cover was determined using satellite imagery and image processing software like FreeView. The images of the locations were classified and then color for both sparse and dense forest cover was identified. Pixel values at 50 different points ($p_1, p_2, p_3 \cdots p_{50}$) within the same land was first identified and then the area covered by those pixels were estimated using the following procedure.

If the number of pixels with p_n values was C_n, and if the total number of pixels in the image was found to be C_T and the area of the image was A, then the area covered by pixels with a value p_n could be estimated by (where $n = 1–50$):

$$\int_{\mathrm{n}=1}^{50} dA_{p_n} = \int_{\mathrm{n}=1}^{50} \left\{ \frac{dC_n}{C_T} \right\} \times A \tag{10.4}$$

The total area of that land feature was estimated by adding the areas of all 50 pixels. The areas of other land features were calculated in a similar manner. If some pixels overlapped, i.e., same pixel is identified to represent two type of land features, then the area of the overlapping pixel is first determined. Then this area is removed from the total area of the image while calculating the extents of individual land use. The distance from the river was measured using a measurement tool available in GIS packages like MapWindow GIS.

This new method was used to estimate the area of:

1. Vegetation,
2. Irrigation fields,
3. Bodies of water,
4. Barren land.

The infiltration or retention capacity of all the above types of landuse is relatively higher than that of the other land use features like pavement, road, or impervious structures.

That is why calculation of the LULC coefficient of a region entails estimating the ratio of the impervious area to the total area. The area of impervious structures are calculated by deducting the sum total of the pervious areas from the total area.

Frequency of Fish Navigation

One of the major drawbacks of HPPs is that they prevent fish populations from navigating freely. Fish navigate in search of food and to locate a suitable location for spawning. Both of these activities are hindered if the path of navigation is blocked by an HPP. That is why frequency of fish navigation varies inversely to the suitability of HPP installations.

Water Quality

An index similar to the NSA water quality index is calculated to represent the overall quality of river water. But instead of parameters related to waste water; quality parameters like TDS, TSS, and pH are used, all of which represent the overall suitability of the water of a river for prime mover. Water with a high concentration of salinity will corrode the blades of a turbine. Again, high TDS or TSS concentrations will represent a high carrying capacity of a river, which generally causes a large amount of sedimentation, which is not beneficial for the turbines of the HPP.

10.2.1.3 Socioeconomic Factors

Average Energy Potential

The energy potential of a river network can be calculated from the average flow and net head. The power generated is calculated with the help of a power equation.

Potential Profit

The profit from an HPP is generally calculated after the cost of installation is recovered. The main income of any HPP is from selling units of power produced at the plant. The expenditures from an HPP include the cost involved in installation, logistics and transportation, land acquisition, labor, relocation of affected population, purchasing of auxiliary power, and other maintenance costs. It was found that a minimum of 5 years is required to recover the installation costs of an HPP depending on the demand and unit price of the supplying region. After that, profit becomes a function of the available potential energy in the stream and variation in selling price, which is controlled by market forces.

Distance from Grid

The distance from the grid is inversely proportional to the suitability of installing HPPs. The greater the distance, the less suitable the location for an HPP. As resistance is directly proportional to the length of a conductor, a long conductor will have higher resistance. An increase in the resistance will also increase the loss of electrical energy. Also, a longer distance will require longer transmission wires, which will increase the overall cost of the power plant. Although few studies include this parameter, the importance of distance on the loss of electrical energy is well documented.

Distance from Consumers

Similarly, the distance from consumers also varies inversely with the suitability of a location for an HPP. The greater the distance, the greater the loss and costs of conducting electricity to the target consumers.

10.2.2 Development of Clusterization Algorithm

In the present investigation, instead of applying conventional clustering methodologies, the novel concepts of fuzzy logic and bat algorithm were used to achieve the study objective.

10.2.2.1 Application of Fuzzy Clusterization

At first, a weighted average of all the factors according to their magnitude was developed where the weights for each of the factors were estimated with the help of fuzzy logic and a bat algorithm.

Table 10.2 describes the weighting for the factors as determined by fuzzy logic.

10.2.2.2 Application of Bat Algorithm

A search optimization algorithm inspired by the echo-location abilities of microbats was developed by Xin-She Yang in 2010. This bat algorithm mimics the method adopted by microbats for locating prey. The method involves varying the rates of pulse emission and loudness based on nearness to prey.

Each bat flies randomly searching for its prey with varying frequency and loudness of pulse emission. When it locates prey, it changes the frequency and loudness of pulse-rate emission. The nearer it comes to its prospective prey, the emission rate from the bat becomes quicker and the pulse becomes louder.

Table 10.2 Rating assigned to some factors with respect to other factors by degree of importance in selection mechanism

	Average change in flow	Average change in net head	Soil strength	Slope	L_c	Fish navigation	Water quality	Average energy potential	Potential profit	Distance from consumers	Distance from grid	Weighting
Average of change in flow		3	2	2	2	1	1	3	3	2	2	0.66
Average change in net head	3		2	2	2	1	1	3	3	2	2	0.66
Soil strength	4	4		4	2	3	3	4	4	4	4	0.50
Slope	4	4	2		3	3	3	4	4	3	3	0.50
L_c	4	4	4	3		3	3	4	3	3	3	0.25
Fish navigation	5	5	3	3	3		2	4	4	2	2	0.6
Water quality	5	5	3	3	3	4		4	4	3	3	0.40
Average energy potential	3	3	2	2	2	2	2		3	2	2	0.33
Potential profit	3	3	2	2	3	2	2	3		2	2	0.33
Distance from consumers	4	4	2	3	3	4	3	4	4		4	0.50
Distance from grid	4	4	2	3	3	4	3	4	4	2		0.50

When the food foraging behavior of bats is applied to solve a problem, the location of a prospective solution is compared to the location of prey. A random search for food by a bat is comparable to random numbers. The pulse rate of emission is represented by *r*, and A_i denotes the magnitude. The higher the rate and magnitude of a pulse, the closer the location of the solution.

In the present investigation, the bat algorithm was applied to select the most suitable location for an SHP plant.

Definition of Good and Bad Location for Availability of Food

To mimic the bat algorithm, first the good and bad locations with respect to availability of food were defined and rated. A high magnitude of factors and low magnitude of factors that respectively increase and decrease the chance of selection of a site for installation of an SHP plant were rated highly, i.e., the chances of finding food at these locations are higher.

Assignment of Randomly Flying Bats to Locations

In the next step, the bats were assigned random numbers, which were multiplied by the rating of the factors to come up with a weighted average. This average was calculated for all the available locations and then normalized. An increase in the random number will increase the chances of selection if the factor is rated as good, and vice versa.

Here randomly flying bats are comparable to the random number that is based on the ratings of the factor or quality of the location increases chance of selection.

Food Spotting

This normalized weighted average can be compared with the pulse rate of microbats whose rate and magnitude increase with the spotting of a food source. The higher the pulse rate, the more reliable is the location of the food source. The weighting at which a location has its maximum weighted average will be the most suitable location for installation of an HPP.

Not only would the attainment of the maximum weighting value make the location optimal, but the rate at which it attains optimality would actually determine its electability. Like the bat, this increases its pulse rate when it finds a location where food is most likely to be present.

This three-stage food-foraging activity was thus mimicked in the present study to identify the most suitable location from among the sample locations available for hydropower generation.

10.3 Results and Discussion

Table 10.3 presents the results of the clusterization and identification efficiency of the algorithms for spotting the best location out of the available nine samples. Location 1 was the ideal one and both algorithms were able to identify it as such. But where fuzzy logic took one iteration only to identify the best location, the bat algorithm took six iterations to achieve the objective.

However, the bat algorithm turned out to be more sensitive and more precise than fuzzy logic as it clearly delineated the optimal solution.

Table 10.3 shows the normalized value of the weighted average as determined from the weighting estimated from the fuzzy and bat algorithms. The respective clusters are also shown in the last column of the table. The number of iterations needed to identify the optimal location was one for the fuzzy and six for the bat algorithm, whereas the sensitivity of selection as represented by the magnitude of the normalized weighted average was clearly higher in the case of the bat algorithm (1.000) than with the fuzzy algorithm (0.894).

Thus, if computational power is not a constraint, then bat-based clusterization can be used to identify the optimal location for an HPP; otherwise, a fuzzy algorithm may be preferred.

Table 10.4 shows the weighting determined by both algorithms in estimating the weighted average of the factors by which the optimal location could be identified.

10.4 Conclusion

The present investigation tried to identify a suitable location for installation of an SHP plant with the help of two algorithms: fuzzy and bat. The potential of both in clusterizing decision variables was also analyzed. It was found that if computational power is not a constraint, then the bat algorithm is better for clusterization than

Table 10.3 Normalized weighted average of factors with respect to weights determined by fuzzy and bat algorithms and corresponding clusters assigned to locations

Location	Normalized fuzzy weighted average	Normalized bat weighted average	Clusters
1	*0.894*	*1.000*	*9*
2	0.790	0.875	8
3	0.690	0.750	7
4	0.593	0.625	6
5	0.500	0.500	5
6	0.411	0.375	4
7	0.324	0.250	3
8	0.241	0.125	2
9	0.160	0	1

Table 10.4 Weighting of factors as estimated using fuzzy and bat algorithms

Factors	Fuzzy	Bat
Average of change in flow	0.6600	0.6506
Average change in net head	0.6600	0.8546
Soil strength	0.5000	0.6126
Slope	0.5000	0.3214
L_c	0.2500	0.7673
Fish navigation	0.6000	0.6381
Water quality	0.4000	0.4656
Average energy potential	0.3300	0.4960
Potential profit	0.3300	0.0531
Distance from consumers	0.5000	0.9008
Distance from grid	0.5000	0.6508

fuzzy-based clusterization. The optimal solution of the present study was already known but was included in the sample population so that the potential of both algorithms could be verified. The results show that, although both algorithms successfully identified the optimal solution, the sensitivity with which the bat algorithm identified the solution surpassed the sensitivity shown by fuzzy clusterization. But bat's requirement for a higher number of iteration can create problems when computational capacity is not high enough to carry out large numbers of iterations.

References

Cyr JF, Landry M, Gagnon Y (2011) Methodology for the large-scale assessment of small hydroelectric potential: application to the province of New Brunswick (Canada). Renew Energy 36(11):2940–2950. ISSN 0960-1481, 10.1016/j.renene.2011.04.003. http://www.sciencedirect.com/science/article/pii/S0960148111001753

Dudhani S, Sinha AK, Inamdar SS (2006) Assessment of small hydropower potential using remote sensing data for sustainable development in India. Energy Policy 34(17):3195–3205. ISSN 0301-4215, 10.1016/j.enpol.2005.06.011. http://www.sciencedirect.com/science/article/pii/S0301421505001667

Fang Y, Deng W (2011) The critical scale and section management of cascade hydropower exploitation in Southwestern China. Energy 36(10):5944–5953. ISSN 0360-5442, 10.1016/j.energy.2011.08.022. http://www.sciencedirect.com/science/article/pii/S0360544211005524

Ghadimi AA, Razavi F, Mohammadian B (2011) Determining optimum location and capacity for micro hydropower plants in Lorestan province in Iran. Renew Sustain Energy Rev 15(8):4125–4131. ISSN 1364-0321, 10.1016/j.rser.2011.07.003. http://www.sciencedirect.com/science/article/pii/S1364032111002371

Kusre BC, Baruah DC, Bordoloi PK, Patra SC (2010) Assessment of hydropower potential using GIS and hydrological modeling technique in Kopili River basin in Assam (India). Appl Energy 87(1):298–309. ISSN 0306-2619, 10.1016/j.apenergy.2009.07.019. http://www.sciencedirect.com/science/article/pii/S0306261909003109

Larentis DG, Collischonn W, Olivera F, Tucci CEM (2010) Gis-based procedures for hydropower potential spotting. Energy 35(10):4237–4243. ISSN 0360-5442, 10.1016/j.energy.2010.07.014. http://www.sciencedirect.com/science/article/pii/S0360544210003786

Nunes V, Genta JL (1996) Micro and mini hydroelectric power assessment in Uruguay. Renew Energy 9(1–4):1235–1238. ISSN 0960-1481, 10.1016/0960-1481(96)88499-0. http://www.sciencedirect.com/science/article/pii/0960148196884990

REN21 (2011a) Renewables 2011: global status report. p 17. http://www.ren21.net/Portals/97/documents/GSR/GSR2011_Master18.pdf

REN21 (2011b) Renewables 2011: global status report. p 18. http://www.ren21.net/Portals/97/documents/GSR/GSR2011_Master18.pdf

Rojanamon P, Chaisomphob T, Bureekul T (2009) Application of geographical information system to site selection of small run-of-river hydropower project by considering engineering/economic/environmental criteria and social impact. Renew Sustain Energy Rev 13(9):2336–2348. ISSN 1364-0321, 10.1016/j.rser.2009.07.003. http://www.sciencedirect.com/science/article/pii/S1364032109001373

Supriyasilp T, Pongput K, Boonyasirikul T (2009) Hydropower development priority using MCDM method. Energy Policy 37(5):1866–1875. ISSN 0301-4215, 10.1016/j.enpol.2009.01.023. http://www.sciencedirect.com/science/article/pii/S0301421509000391

World Energy Assessment, UNDP (1998) http://www.urjaglobal.in/indian_energy.html

Yang X-S (2010) A new Metaheuristic Bat-Inspired Algorithm. In: Gonzalez JR et al (eds) Nature inspired cooperative strategies for optimization (NISCO 2010), Studies in computational intelligence, vol 284. Springer, Berlin, pp 65–74. http://arxiv.org/abs/1004.4170

Yi CS, Lee JH, Shim MP (2010) Site location analysis for small hydropower using geo-spatial information system. Renew Energy 35(4):852–861. ISSN 0960-1481, 10.1016/j.renene.2009.08.003. http://www.sciencedirect.com/science/article/pii/S0960148109003462

Part III
Impact Studies

Chapter 11
Impact of Climate Change on Ecological Sensitivity of Wetlands

Mrinmoy Majumder and Rabindra Nath Barman

Abstract The ratio of eco-friendly to eco-foe land use can be defined as the ecological sensitivity of a region. If the area of eco-friendly land use is more within the targeted range, then that area is designated as ecologically suitable, and if most of the region within the observed area has an eco-foe land use, then the area of interest is said to be ecologically sensitive. The degree of sensitivity and suitability varies with the dominance of eco-foe and eco-friendly land use within the investigated area. Currently, the increase in greenhouse gases in the atmosphere has increased the average temperature of the Earth's surface, which in turn has engendered uncertainties in normal climatic patterns, a situation that is aggravated due to the constant and uncontrolled extraction of natural resources, which has impacted the natural ecosystem of many regions of the world. The impact of climatic uncertainty has also severely affected the ecological conditions of wetlands. Thus, there is an urgent need for impact analysis for wetlands with respect to changes in ecological patterns of wetland watersheds. The present study aims to estimate the degree of uncertainty induced by changes in climatic patterns with the help of ecologic sensitivity and NASA's (Climate change may bring big ecosystem changes. Retrieved from http://www.physorg.com/news/2011-12-nasa-climate-big-ecosystem.html, on Feb 2012) prediction of future biome change. The present investigation has used the decision tree algorithm for the prediction of the ecological sensitivity of Tripura wetlands. Tripura is a state in northeast India undergoing rapid urbanization. The results of the study could help city planners to develop compensatory measures specific to those regions that urgently need them.

M. Majumder (✉)
School of Hydro-Informatics Engineering, National Institute of Technology Agartala, Barjala, Jirania 799055, Tripura, India
e-mail: mmajumder15@gmail.com

R.N. Barman
Department of Production Engineering, National Institute of Technology Agartala, Barjala, Jirania 799055, Tripura, India
e-mail: rahul.barman@yahoo.co.in

M. Majumder and R.N. Barman (eds.), *Application of Nature Based Algorithm in Natural Resource Management*, DOI 10.1007/978-94-007-5152-1_11,

Keywords Land use change • Biome • Climate change • Decision tree algorithm

11.1 Introduction

The population explosion, uncontrolled use of natural resources, and global warming have induced large-scale degradation in different regions of the world. Although interrelated, these three causes of environmental deterioration affect the degree of uncertainty not only of the biotic components (e.g., flora, fauna) but also of the abiotic components (e.g., rainfall, evaporation) of ecosystems. In this regard, proper assessment of the ecological scenario of different regions must be conducted so that concerned authorities can take proper and necessary measures to prevent accelerated depletion of natural resources.

Among many other natural landforms, wetlands are responsible for providing water during times of scarcity. They also impact both the type and the amount of agricultural output from neighboring regions. Wetlands have their own ecology, which includes other living (e.g., fish, birds, insects, humans) and nonliving (e.g., soil, geology, weather, water) entities but varies with the characteristics of the water body. The role of wetlands in the sustainability of natural as well as human resources is well known and cannot be ignored. But because of the overexploitation of stored water, polluted inflow, and recharges (generated from adjacent hotels and industries), their attraction as a tourist destination has contributed to the degradation and even the extinction of many wetlands around the world. When a wetland becomes extinct, adjacent and dependent ecosystems and communities are affected in terms of habitat displacement, changes in livelihood, modification of a source of income, land use changes, and acute scarcity of water.

The remote sensing of land features and subsequent processing of images have reduced the need for extensive fieldwork in earth science projects (river monitoring, land use change detection, surveying, groundwater exploration, habitat analysis, delineation of catchments, site selection, etc.). The advent of sophisticated satellite technology and processing software has greatly enhanced the prospect of "drawing conclusions" from the comfort of a laboratory. An inspection of land use areas, structures, textures, extent, relief, and patterns as derived from the processing of captured images helps our understanding of the prevalence of species, water bodies, etc. This technological advantage can also be applied to the assessment of ecological characteristics without having to visit inaccessible areas. The ecology of an area is well represented by its land use. A concentration of forests suggests a certain level of environmental stability of a given region, whereas the presence of industry or residential dwellings implies urbanization, which is a clear sign of ecological degradation. Also, the presence of forests and bodies of water helps life forms grow without uncertainty.

The state of Tripura boasts abundant natural resources and broad biodiversity, which has encouraged Birdlife International to classify the region as an East Himalayan Endemic Bird Area. Fifty-one avian species, of which 21 are restricted species, can be found in the jungles of Tripura. The only ape of India, the hoolock gibbon (*Hylobates hoolock*), can be spotted in the dense forests of the state.

According to Chatterjee et al. (2006), this is the only state in northeastern India that has seen an increase (by 1,033 km^2) in the total area of forest if dense and open regions are considered together. The endangered Phayre's langur, known locally as "Chasma Banar," can be found only in the state's reserve forest. The state has high hills and many small and large wetlands – 408 in all (SAC 1998), mainly generated from seasonal water logging, reservoirs, oxbows, and tanks (Chowdhury 2003). Although there are minor problems in the availability of water, there is a major problem with using the resources because of the high concentration of iron. Rain provides some relief by reducing the concentration of iron by recharging the groundwater. The presence of wetlands in the neighboring region also helps to maintain iron at a lower level. But during the dry season, the wetland receives water from the ground because of the lower water level, which increases the concentration of iron in the groundwater. Presently, most of the wetlands in the state are polluted because of uncontrolled exploitation for tourism and irrigation. They are silted, encroached upon by irrigation fields, and defaced by tourists.

To understand the relation between wetland deterioration and the quality of an ecosystem, the present study tries to determine the ecological sensitivity of regions adjacent to wetlands in the state by considering eco-friendly and eco-foe land use.

Remote sensing was used in this context, and proper processing and classification were done with the help of a knowledge-based decision tree algorithm.

The results of the classification effort show the disparity between the ecologically stable and sensitive regions within the state. The present status of the wetlands was also considered to assist concerned experts in prioritizing the available economic and technical resources in the prevention of total degradation.

11.1.1 Ecology and Ecosystem

Ecology is the study of interactions or relationships between organisms and the environment, the connectedness between living systems and nonliving systems on Earth.

In a broad sense, ecology involves the study of various attributes of life forms – including their life processes, distribution, adaptation, succession, movement, role in energy transfer, etc. Even though ecology is not associated with environmental science directly, various aspects of this discipline have a crucial role to play when it comes to having a strong knowledge base of environmental studies. Ecological organization (i.e., the different levels of organization in ecology,) is one such aspect of this discipline that makes things simple by classifying the environment at different levels, starting from individual organisms on up to the biosphere as a whole.

Ecology also looks at natural resources such as water, air, and soil. Studies are done on an ongoing basis to determine what changes may be taking place, such as the effects of pollution and particulate matter on water and air over time, what impacts these elements, and how changes may affect other living organisms.

An ecosystem is a community of organisms in a habitat together with the nonliving part of the environment. Therefore, an ecosystem is said to be made up of abiotic and biotic factors. Abiotic factors are the nonliving factors such as light, oxygen, carbon dioxide, air, soil, water, stones, etc. Biotic factors are the living part of the ecosystem and include plants, insects, worms, and other animals. Thus, an ecosystem is simply a group of plants and animals that reside in the same area and are dependent upon each other for survival.

There are many different types of ecosystems in the world including deserts, forests, prairies, rivers, oceans, swamps, and marshes. All of the various ecosystems are interdependent on each other. That means if something affects one ecosystem negatively, it will affect the other ecosystems as well. Within an ecosystem all plants and animals are interdependent on each other for survival. That is why an ecosystem is also referred to as a community of organisms and their physical environment interacting as an ecological unit.

Ecosystems include physical and chemical components such as soils, water, and nutrients that support the organisms living within them.

These organisms range from large animals and plants to microscopic bacteria. Ecosystems include the interactions among all organisms in a given habitat. People are a part of ecosystems.

The health and wellbeing of human populations depend upon the services provided by ecosystems and their components – organisms, soil, water, and nutrients.

Species are related to each other through the roles they play in the food chain as producers, consumers, and decomposers. Producers are photosynthesizing plants, consumers are herbivorous or carnivorous animals, and decomposers are organisms (such as bacteria) that break down organic material into minerals, which are eventually used by producers.

11.1.2 Wetland Ecosystem

A wetland is an area that contains soil that remains wet most of the year or throughout particular seasons. This includes freestanding areas of water within a portion of land. Many species of plants and animals make wetlands their home. Wetlands are vital to the feeding methods of many animals and are important for them to sustain life. These areas provide important nutrients and minerals to the animals and plants and promote an overall healthy environment.

Tidal and nontidal fresh emergent wetland habitats are important habitat-use areas for fish and wildlife that depend on marshes and tidal shallows and support several special-status plant species. The loss or degradation of historic fresh emergent wetlands has substantially reduced the habitat area available for associated fish and wildlife species.

Saline emergent wetland habitats, including brackish and saline wetlands, are important habitat-use areas for fish and wildlife dependent on marshes and tidal shallows in the Bay Delta and support several special-status plant species. The loss or degradation of historic saline emergent wetlands, primarily as a result of reclamation of tidally influenced wetlands for agriculture, has substantially reduced the habitat area available for associated fish and wildlife species. Several plant and animal species closely associated with tidal saline emergent wetlands have been listed as endangered under the state and federal endangered species acts, primarily as a result of the extensive loss of this habitat type. Loss of the habitat has reduced nutrient cycling and foodweb support functions.

The health of an aquatic ecosystem is degraded when the ecosystem's ability to absorb a stress has been exceeded. A stress on an aquatic ecosystem can be a result of physical, chemical, or biological alterations of the environment. Physical alterations include changes in water temperature, water flow, and light availability. Chemical alterations include changes in the loading rates of biostimulatory nutrients, oxygen-consuming materials, and toxins. Biological alterations include the introduction of exotic species. Human populations can impose excessive stresses on aquatic ecosystems.

11.1.3 Climatic Impacts on Ecosystems

Anthropogenic as well as natural climatic impacts have changed the frequency, concentration, and amount of biomass available in ecosystems before abnormalities in climatic patterns were identified. Wang et al. (2007) identified the impact of climate change on an alpine ecosystem in the Yangtze and Yellow River source regions of central Qinghai–Tibet based on soil properties. The soil surface layer had become coarser, and a rise in the bulk density, porosity, and saturated hydraulic conductivity was identified along with a decrease in water-holding capacity. Philippart et al. (2011) discussed climatic impacts on sea aquatic ecosystems in Europe. For example, species in the Arctic, Barents, and Nordic Seas will move to the polar region due to an increase in temperature, which will invite species from temperate regions to fill the gap created by the departed polar species. Some seas that receive large amounts of freshwater from river runoff will be dominated by inland brackish water species due to an increase in rainfall. Semienclosed seas, like the Mediterranean and Black Sea, will lose their endemic species, and species from the adjacent water bodies (Suez Canal) will fill the void created by the now-vanished native species. Briceño-Elizondo et al. (2006) highlighted the impact of climate change on the carbon stock and timber yield of two boreal forests with the help of a process-based ecosystem model. It was found that the carbon stock of the forests will increase if the thinning process taking place in the forests is varied even during severe climatic abnormalities. Table 11.1 points to some other relevant scientific studies that discuss the impacts of climate change on different aspects of various ecosystems.

11.1.4 Objective and Scope

In the present study the ecological status of ecosystems that are adjacent to wetlands was classified with respect to the presence of eco-foe and eco-friendly land use and applying the concepts of decision tool algorithms. After classifying the wetlands'

Table 11.1

References	Study area	Type of ecosystem	Objective
Beniston (2012)	Swiss Alps	Agro-ecosystem	Climatic impacts in alpine mountains on socioeconomic output were modeled and showed that tourism, hydropower, agriculture, and the insurance industry would be most severely affected
Merino et al. (2010)	Peru and Chile	Marine ecosystem	Production of fish food and oil was studied with respect to climatic fluctuations and global demand; results indicated that both global- and regional-level management of the uncertainties was required to counter the impacts of climate change
Cohen (1996)	Mackenzie Basin, Canada	Land and water	Impact on land and water resources due to changes in land use and policy decisions
Boix et al. (1995)	Alicante, Spain	Terrestrial	Climatic impacts on soil properties of a mountain ecosystem; it was found that organic matter and Carbon Equivalent Compound are directly related to climate

present status, the scenario of the wetlands in 2100 is also modeled by imposing information on the amount of change that would occur on land masses and applying the same methodology of classification. The status of future ecosystems is compared with the present for an indication of future vulnerabilities due to climate change.

11.2 Study Area

Tripura was once known as “Hill Tippera,” and the very nomenclature suggests its hilly nature of an undulating surface made uneven by interspersed low hills. Tripura lacks mineral resources; it has poor deposits of kaolin, iron ore, limestone, coal, and natural gas. It has a rich cultural heritage of music, fine arts, handicrafts, and dance. Tripura became a union territory without legislature effective as of November 1, 1956, and a popular ministry was installed in Tripura on 1 July 1963. Tripura, which is bordered by Bangladesh on three sides and parts of Assam and Mizoram on one side, was formerly a princely state. Tripura's tourism industry offers tourists the opportunity to explore the rich cultural traditions, religious legacies, and vast reserves of unique flora and fauna of the northeastern part of India. Tripura can be reached by air through Agartala Airport, and Kumarghat is the nearest railway station. Tripura Sundari, the most important temple in the area, stands 5 km outside of Udaipur on a small hillock in front of a holy lake teeming with carp and turtles. Tripura, currently a full-fledged state of northeast India, was once, as already noted, a princely state with a long list of tribal kings stretching back to antiquity. Tripura's proximity to the markets of South Asia makes it a potential hub for the eventual expansion of trade and commerce in the region. The Tripura Government Museum, located at Post Chowmohani in Agartala, is a storehouse of historical and cultural artifacts, paintings, and documents. Tripura is a tourist's paradise with lush green valleys and picturesque mountain ranges.

It is bordered on three sides by Bangladesh and by Mizoram state; it has an area of 10,486 km^2), and its capital is Agartala. It was an independent Hindu kingdom before it became part of the Mughal Empire in the seventeenth century. After 1808 it fell under the influence of the British government. Tripura became a union territory in 1956 and acquired full status as a state in 1972. Its main economic activity is agriculture, with rice and jute the major crops. The state is located in the biogeographic zone of 9B-North-East Hills and possesses an extremely rich biodiversity. The local flora and fauna components of Indo-Malayan and Indo-Chinese subregions. There are 379 species of trees, 320 shrubs, 581 herbs, 165 climbers, 16 climbing shrubs, 35 ferns, and 45 epiphytes.

The northeastern state of Tripura is among the world's 35 biodiversity hotspots; the region has a profusion of freshwater fish species and diverse flora and fauna. All of this is under threat.

The climate of the state has been roughly divided into tropical, temperate, and alpine zones. For most of the year, the climate is cold and humid as rain falls every month. The area experiences heavy rainfall due to its proximity to the Bay of Bengal. The rainfall in the northern district is comparatively less severe than that of the other districts. The general trend toward decreasing temperatures with an increase in altitude holds good everywhere. Premonsoon rain occurs in April to May, and monsoon season (southwest) normally starts in May and continues up to early October.

Jampui Hills, the highest hill range in the Tripura state, is located approximately 200 km from Agartala, Tripura, 90 km from Kailashahar District. Although Tripura has vast potential, the industrial sector of the state is underdeveloped. The state's secondary sector contributes just 5% to the total employment of the state. Tourism was given the formal status of an industry in 1987. Tripura produces important horticultural products such as pineapple, orange, cashew nut, jackfruit, coconut, tea, cotton, mesta, rubber, and others. The agriculture of the state is largely based on a system of Jhum (shifting) cultivation and gives due importance to animal husbandry and fisheries. The resources, policy incentives, infrastructure, and climate in the state support investments in sectors such as natural gas, food processing, rubber, tea, bamboo, handloom and handicrafts, sericulture, tourism, IT, and medicinal plants. Natural gas deposits are among the most important reserves of Tripura's natural-resource base. Natural gas is present in Baramura Hills and in Rokhia, where gas-based power plants have been set up. The other potential sectors of the state are organic spices, medicinal plants, and biofuel. A proposal has been put forward to set up a state biofuel mission under the State Department of Forestry. Favorable agro-climatic conditions, low use of chemicals, and the availability of a variety of spices offer opportunities for the establishment and development of the spice sector.

Another potential sector is rubber cultivation. The state's soil and climate are conducive to quality rubber production. Tripura is the second largest producer of natural rubber in India. The full scope of the industry remains largely untapped. Like tea, rubber plantations do not employ many people. However, processing units and other activities may provide opportunities to reduce unemployment. The state government has initiated actions and brought otherwise noncultivable areas under rubber cultivation and designed policies to modernize the processing units. In addition, it is necessary to develop proper marketing strategies.

Flora and fauna: due to the plentiful and well-distributed rainfall, the state has an ideal composition of land mass and water that houses a large variety of flora and fauna. A wide variety of plants and orchid species are found in the forests of Tripura. Sal (*Shorea robusta*) is an important forest product here. A series of low ranges running in a northwest-to-southwest direction dissects the state, with elevations progressing gradually from 100 to 3,000 ft from southwest to northeast. Tripura has a large population of tribals and, thus, a tradition of different kinds of crafts. The handloom is the most important craft of the state. The main feature of the Tripura handloom is its vertical and horizontal stripes with scattered embroidery in different colors. The handloom is followed by cane and bamboo craft. Popular handicraft items are bamboo screens, lamp stands, tablemats, sitalpati, woodcarving, silver ornaments, and others. Simplicity is the hallmark of brass and bell metal articles made in Tripura.

11.2.1 Wetlands of Tripura

The wetlands of Tripura were selected as the study area. The biodiversity and rich natural resources have made this small state of northeast India a well-known tourist destination. The state has a number of small and large wetlands. Most of them are formed by seasonal water logging (222) followed by 84 oxbow lakes, normal lakes and ponds (74), irrigation tanks (19), reservoirs (5), and, lastly, permanent water-logged depressions (4), all of which are included on the list of prioritized wetlands identified by the Space Application Center (SAC) in 1998.

Fig. 11.1 Location of four most important wetlands of Tripura

11.2.2 *Reasons for Wetland Degradation*

The main causes of environmental damage in the big reservoirs and lakes of Tripura as identified by Chatterjee et al. (June, 2006) are forest degradation, shifting cultivation, influx of people from Bangladesh, grazing, forest fires, smuggling of forest produce, presence of weeds, introduction of exotic species, plantations, use of chemical fertilizers, consequences of developmental projects carried out without environmental clearance, lack of awareness, and lack of coordination among government institutes.

Tripura receives adequate rainfall and seldom faces drought. Thus, study of the wetlands of the state has become a precondition for identifying the situation and making interested parties aware of the state of the local environment. Below is a brief description of the four main lakes/reservoirs of Tripura as given by the Lake Conservation program.

11.2.3 *Important Wetlands of Tripura (Fig. 11.1)*

11.2.3.1 Gumti Reservoir

Gumti Reservoir is located at 23°25′45″N and 91°49′20″E across the river Gumti in the South Tripura district. The 4,500 ha watersheet draws water from Barak, Raima, and Sarma river basins. Completed in the year 1977, the 103 m long, 30 m high, straight gravity dam is made of brick and stone concrete, storing 235.62 million m^3 of water at the FRL. The river at the dam site has a catchment area of 547 km^2, receiving rainfall of 173 cm.

11.2.3.2 Rudrasagar Lake

The Government of India's Ministry of Environment and Forest has identified Rudrasagar as one of the wetlands of national importance for conservation and sustainable use based on its biodiversity and socioeconomic importance.

Rudrasagar Lake, approximately 55 km from Agartala near Melaghar and having a 5.3 km^2 water area, is another big attraction. In the center of the lake stands the famous lake palace of Tripura called Neermahal. The lake witnesses a large number of migratory birds in every winter. Every year a boat race is organized in July/August. Visiting tourists can enjoy the boating facility in the lake.

The Rudrasagar Lake is located in the Melaghar Block under the Sonamura Subdivision in the West Tripura District at a distance of approximately 50 km from the state capital of Tripura. Geographically the lake is situated between 23°29′N and 90°01′E.

11.2.3.3 Sipahijala Reservoirs

Sepahijala Wildlife Sanctuary was constituted on 2 February 1987. The sanctuary has 456 plant species of monocotyledon and dicotyledon. The dominant types of tree include Sal, Chamal, Garjan, and Kanak. The secondary species consist of Pichla, Kurcha, Awla, Bahera, Hargaja, Amlaki, Bamboo, and grasses. The sanctuary has 4,489 cum per hectare of timber biomass.

11.2.3.4 Trishna Reservoirs

Trishna Wildlife Sanctuary was established in November 1988. The total area of the sanctuary is 194,704 km^2. Trishna Sanctuary has diversity in its floral and faunal contents.

The sanctuary is famous for bison, locally known as gaba, and home to several species of primates. The sanctuary has a number of perennial water rivulets, water bodies, and grassland.

One species of bamboo (*Oxtenanthera nigrocilliate*), locally known as Kaillai, is plentiful, and its leaves are liked by bison. The Trishna Reservoir is situated almost in the middle of the sanctuary and is a tourist attraction.

11.3 Methodology

The present study aims to predict climate change impacts on the ecological sensitivity of Tripura's wetlands with the help of an artificial neural network. In this regard the methodology adopted is described in the next section.

11.3.1 Data Collection

Achieving the objective of the present study would require a large data set of land use distribution in and around the considered wetlands. Future land use distribution as predicted by some models with respect to IPCC climate scenarios.

The requirement of identifying the land use distribution was accomplished by the application of image processing on images captured from the command area of the selected wetlands. A grid of 0.38°/0.38° was created starting from the midpoint of the water bodies. The entire command area of the wetlands was divided into grids of equal area and the land use pattern for each grid was determined with the help of the image processing.

The future land use distribution of the study area was derived from the global land use variation of both A2 and B2 scenarios predicted by the GLSS model, which predicts the spatial distribution of geographical features and agroforestry at a scale 1°/1° according to the two scenarios of climate change.

11.3.2 Image Processing

All the grids in and around the lakes were captured from Google Earth Version 5 and the images were processed by FreeView Software Version 10.3.

The following flowchart (Fig. 11.2) depicts the methods adapted for resampling, processing, and classifying the images according to their ecological stability. During image processing every image was first resampled with the help of a nearest-neighbor method and visually enhanced by adaptive processing. Ultimately the preprocessed image was converted to pseudo-color, which is simply classification of the image based on the pixel value and its presence in the image for a single band. For this study the green band was selected for this conversion as that band was found to be suitable for representing the targeted land use types. For example, forest areas display different levels of green, which is represented by different values of pixels within the image. Water bodies on the other hand reflect blue but absorb green and red. That is why lower pixel values in the green band play the role of representing this particular land feature. Residential or industrial areas are generally shown by pixel values of more than 200 within the selected band. In the case of the two other bands, the forest is indirectly represented, and because the main objective was to delineate the ecological status where forest plays the most important role, the green band was selected to avoid disparity due to band mismatch.

Based on the density and variation of pixels within a certain land feature, the area of that land use was determined. When pixels of certain values were found in two types of land use, rectification to mitigate overlapping was conducted. In this method the feature where the value of a pixel was found to occur more often, the area of the pixel was added to the total area of that feature. Before any processing the image

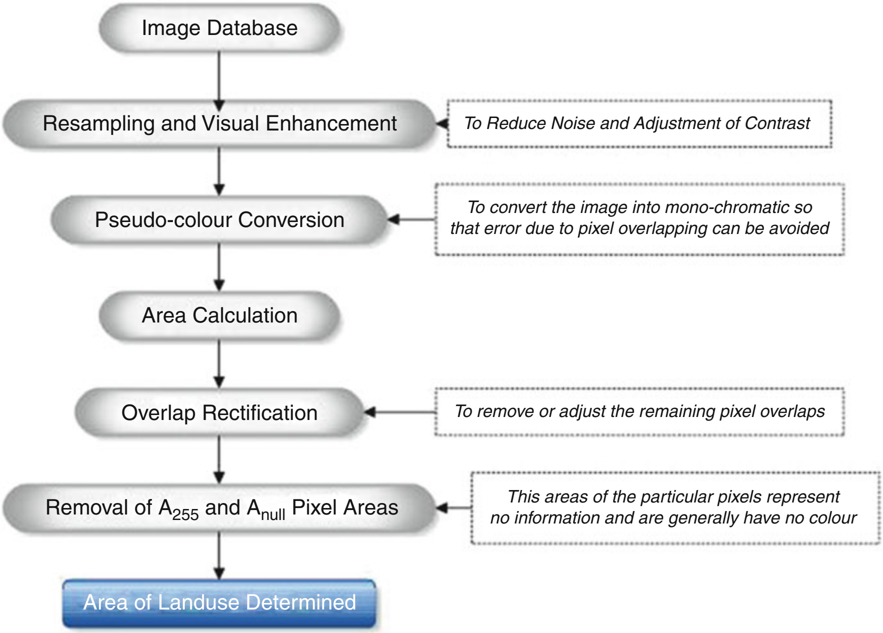


Fig. 11.2 Methodology adopted to process captured images

was filtered linearly to remove noisy areas. If a major part of an image was unrecognizable, then that image was rejected and the data deficiency recorded.

The land use area was calculated by identifying the minimum and maximum valued pixels that represented a specific land use among the four considered (forest, water, residential/industrial/barren land, and road) areas. Then the frequency of those pixel values within the image was determined from the histograms, which are graphs of pixel values, and its rate of occurrence in the image. Because the area of the image was known, the frequency value was multiplied by the total area to find the total area occupied by that pixel value.

First, pixel values of at least 20–50 locations within the same land feature were selected and the area occupied by them was determined. The method can be numerically represented in the following process.

Let the frequency for the minimum valued pixel within the land feature be *pm* and for the maximum valued pixel *pM*; then the area of the land feature can be determined as

$$\mathrm{A}_{\mathrm{lp}} = A_p \int_{pm}^{pM} dp_n - dp_{n-1}, \tag{11.1}$$

where A_p is the total area of the image and A_{lp} is the area of the land feature.

Because the values and location of the pixels were discrete, discontinued or by-parts integration was performed. Generally the result of this numerical operation gave more or less accurate measures of the areas of a given land feature. However, the area of null valued and 255 valued pixels was deducted from the total area so as to avoid pixels with no information, which reduced the possibility of error.

After processing of the image and an area of eco-friendly and eco-foe land use was determined, the procedure described in Fig. 5.2. Once the land use features of each grid were classified into different groups of ecological sensitivity, the frequency of friendly and foe land uses were calculated and the following rules were conceptualized.

11.3.3 Classification Rule

After an area of different geographical features of a grid image was identified, the following classification rule was applied to the image to separate ecologically suitable grids from ecologically sensitive grids.

The classification rule proposed:

1. If the frequency of **Class A** is at the maximum within a grid, then classify it under "Ecologically Suitable Class A (ESuA)"
   Else,
2. If the frequency of **Class B** is at the maximum within a grid, then classify it under "Ecologically Suitable Class B (ESuB)"
   Else,
3. If the frequency of **Class C** is at the maximum within a grid, then classify it under "Ecologically Moderate (m)"
   Else,
4. If the frequency of **Class D** is at a maximum within a grid, then classify it under "Severely Ecologically Sensitive Class B (SSEB)"
   Else,
5. If the frequency of **Class E** is at a maximum within a grid, then classify it under "Severely Ecologically Sensitive Class A (SSEA)."
   Else STOP

After the classification rule was developed, the grid images were classified accordingly with the help of a decision tool mechanism.

The areas with Ecologically Sensitive Class A are highly vulnerable, whereas Class A of Ecologically Suitable was taken as a representative of an ecologically stable region. For areas assigned to Class B is vulnerable but not as vulnerable as Class A, but proper attention is needed to stop the degradation. On the other hand, Class B of the Ecologically Suitable classes represents places that are steadily degrading toward the vulnerable zone. But for the moderate class, the threat is neither very high nor very low.

11.3.4 Prediction of Climatic Impact on Ecological Sensitivity

After the model was trained, it was applied to predict the ecological sensitivity in the year 2100 according to the prediction of land use change by NASA/JPL (2012). The results indicated the level of impact on the wetland watershed due to changes in climatic patterns of the region.

11.4 Results and Discussion

The area of different geographical features in an image was identified with the help of Eq. 11.1. After the area of the features was identified, a decision tree mechanism was used to classify all the images. According to the results of the classification, only 3.24% of the area was found to be severely (class SSEA) and 7.57% area moderately (SSEB) devoid of biodiversity and adequate natural resources to maintain the wetland ecosystems, and 78% of the 370 considered grids were found to be capable of sustaining the present scale of rapid urbanization. These are only 0.54% of grids that had neither good nor bad ecological status and only 4.6% of the grids could not be classified due to the absence of clear delineation between different land features. These areas were thus omitted from the analysis.

The future land use data were extracted from the report made by Asner (2012), who predicted future land use changes due to moderate and high levels of climate change. It was found that, in the Asian subcontinent, climate change will less of influence on deforestation than urbanization and forest logging.

According to the report, regions with sparse urbanization and higher forest cover will be more severely affected than places with sparse forest cover and dense urban populations. In Tripura, urbanization small, but forest cover is high. Thus the impact will be harsher in the state.

The report predicted a 70% change in the land mass of northeast India. The model inputs were thus adjusted to predict the impact of climate change on the ecological stability of the state. According to the model predictions, most of the regions of Tripura would fall within Ecologically Suitable Class A, which implies that there would be no noticeable change observed in this part of India due to climate change.

Comparison of future and present classes of ecological sensitivity revealed that the ESuA class, which had the highest frequency among all other classes under present conditions, would diminish by 96.50%, whereas SSEA would increase by 28.83 times the number of grids having the same class as they presently do. There were 17 grids having ESuB class under present-day climate scenarios, but in the face of climate change, no grids in the ESuB class would be observed, although the SSEB class would also account for zero grids under changed climatic conditions in the state of Tripura. Table 11.2 highlights the comparison between present and future ecological sensitivities of the grids considered in this study.

Table 11.2 Ecological sensitivity classes of study area with respect to present and future climate scenarios

Ecological sensitivity class	Present	Future	Change (%)
ESuA	310	11	−96.50
ESuB	17	0	−1.000
m	0	0	0.000
SSEA	12	358	+2,883.33
SSEB	28	0	−100.00
u	2	0	−100.00

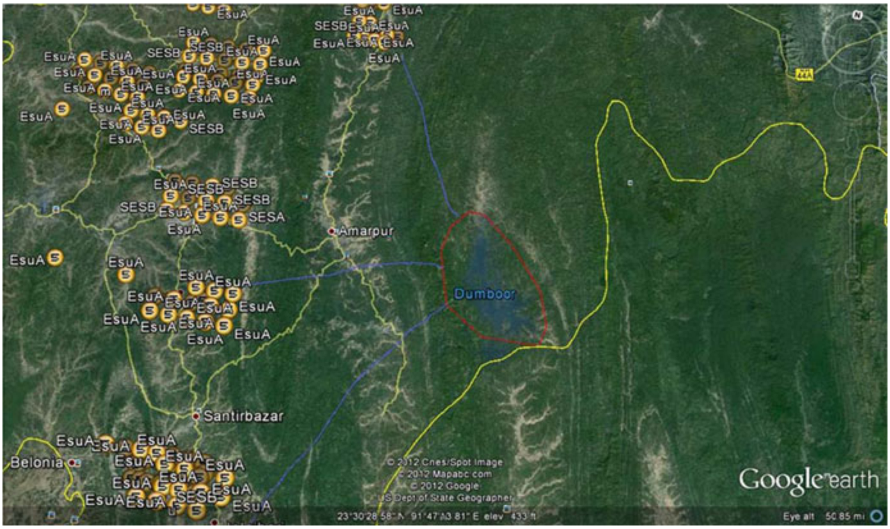


Fig. 11.3 Gomati (Dumbur) Lake and its land use features (grids were delineated from upstream of wetland watershed as indicated by the *yellow bookmarks*). The bookmark titles indicate the ecological sensitivity of the grid. The *red circle* highlights the wetland and the *blue path* indicates the probable path of runoff) (Source: Google Earth)

Figures 11.3, 11.4, 11.5, and 11.6 depict the land use features, delineated grids, and corresponding ecological sensitivity classes (under present climatic conditions) of the four wetlands considered in this study. In the present climate scenario it was found that of the four wetlands, only Rudrasagar Lake was found to have a high concentration of the SSEA class, which indicates a poor ecological state. Thus, in the future this lake will also be more vulnerable than the other wetlands. Regarding future climatic conditions, except for the Sipahijala (where few of the grids in the upstream watershed were classified under EsuA), all other wetland watersheds show signs of extreme ecological degradation, as represented by ecological sensitivity Class A (SSEA).

Fig. 11.4 Rudrasagar Lake and its land use features (grids were delineated from upstream of wetland watershed as indicated by the *yellow bookmarks*). The bookmark titles indicate the ecological sensitivity of the grid. The *red circle* highlights the wetland and the *blue path* indicates the probable path of runoff) (Source: Google Earth)

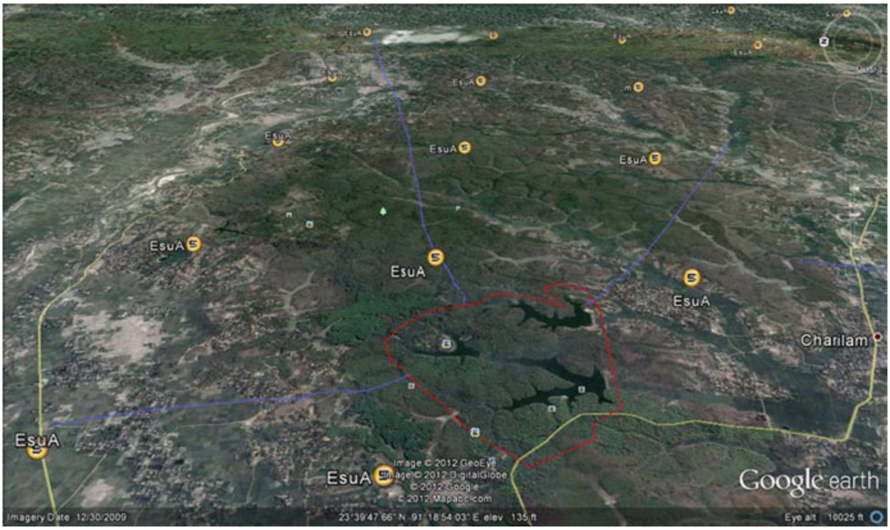


Fig. 11.5 Sipahijala wetlands and its land use features (grids were delineated from upstream of wetland watershed as indicated by the *yellow bookmarks*). The bookmark titles indicate the ecological sensitivity of the grid. The *red circle* highlights the wetland and the *blue path* indicates the probable path of runoff) (Source: Google Earth)

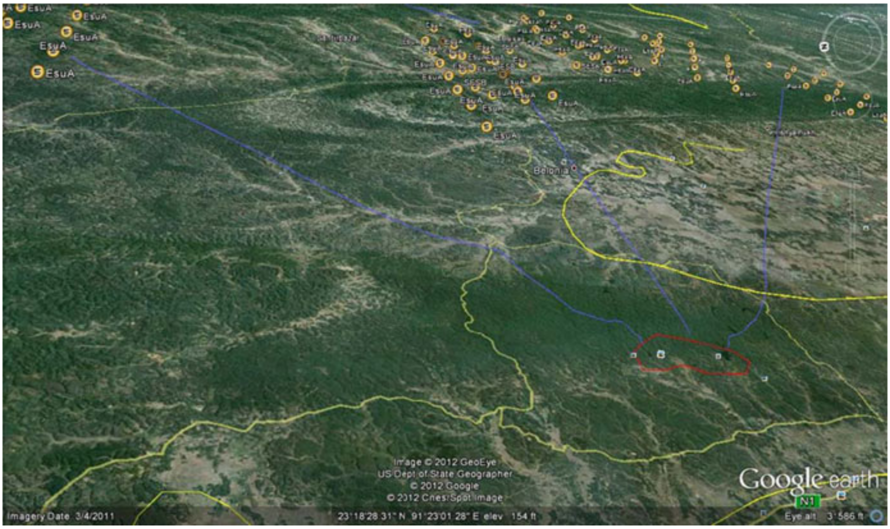


Fig. 11.6 Wetlands inside Trishna reserve forest and its land use features (grids were delineated from upstream of wetland watershed as indicated by *yellow bookmarks*). The bookmark titles indicate the ecological sensitivity of the grid. The *red circle* highlights the wetland and the *blue path* indicates the probable path of runoff) (Source: Google Earth)

11.5 Conclusion

The present study tried to estimate the impact of climate change on ecological conditions of Tripura wetlands. A decision tree algorithm was used to classify the group of grids considered around the wetlands to represent the present ecological status of the wetland watersheds. Changes in land use due to future climatic abnormalities were predicted from the results of the Global Change in Biomass study conducted by NASA (2012). The same decision tree algorithm was again used to classify the grids according to future land use conditions. Comparison of present and future ecological sensitivities showed a 96.50% decrease in ecologically suitable or stable conditions, whereas a massive 2,883.33% increase was observed in the case of the class SSEA, class which represents the ecologically sensitive land characteristics. But the SSEB class, which 28 of the 369 grids considered fell in, was found to be entirely absent under future climate scenarios. Thus, it can be concluded that ecological degradation will be severe under the changed climate scenario in the wetlands of Tripura. The present investigation was conducted on a regional level, but the study may be repeated from a larger perspective. A similar study could be conducted on a global scale so that interrelationships between the climate and biomes would be more clearly represented.

References

Beniston M (2012) Impacts of climatic change on water and associated economic activities in the Swiss Alps. J Hydrol 412–413:291–296

Boix C, Calvo A, Cerdá A, Imeson AC, Soriano MD, Tiemessen IR (1995) Vulnerability of Mediterranean ecosystems to climatic change: study of soil degradation under different climatological conditions in an altitudinal transect in the south east of Spain. Stud Environ Sci 65:763–766

Briceño-Elizondo E, Garcia-Gonzalo J, Peltola H, Kellomäki S (2006) Carbon stocks and timber yield in two boreal forest ecosystems under current and changing climatic conditions subjected to varying management regimes. Environ Sci Policy 9(3):237–252

Chatterjee S, Saikia A, Dutta P, Ghosh D, Worah S (2006) Review of Bio-diversity in North-East India, Retrieved from http://siteresources.worldbank.org/INTSAREGTOPWATRES/Resources/Background_Paper_13.pdf on January 18, 2013

Cohen SJ (1996) Integrated regional assessment of global climatic change: lessons from the Mackenzie Basin Impact Study (MBIS). Glob Planet Chang 11(4):179–185

Merino G, Barange M, Mullon C, Rodwell L (2010) Impacts of global environmental change and aquaculture expansion on marine ecosystems. Glob Environ Chang 20(4):586–596

NASA (2012) Climate change may bring big ecosystem changes. Retrieved from http://www.physorg.com/news/2011-12-nasa-climate-big-ecosystem.html. Feb 2012

Philippart CJM, Anadón R, Danovaro R, Dippner JW, Drinkwater KF, Hawkins SJ, Oguz T, O'Sullivan G, Reid PC (2011) Impacts of climate change on European marine ecosystems: observations, expectations and indicators. J Exp Mar Biol Ecol 400(1–2):52–69

Space Application Center (1998) Watershed Atlas of India, Retrieved from moef.nic.in/downloads/public-information/NWIA_Tripura_Atlas.pdf on 18th January 2013.

Wang G, Wang Y, Li Y, Cheng H (2007) Influences of alpine ecosystem responses to climatic change on soil properties on the Qinghai–Tibet Plateau, China. CATENA 70(3):506–514

Chapter 12
Impact of Climate Change on the Hydrologic Sensitivity of Sundarban Reserve Forest

Mrinmoy Majumder and Rabindra Nath Barman

Abstract The increase in global average temperatures caused by the regular but uncontrolled extraction and degradation of natural resources by either artificial or natural means has initiated noticeable amounts of aberrations in climatic conditions of many regions of the world. Scientists are now busy searching for the causes of rising temperature and different methods of mitigation to counteract the impacts of climate changes that are now "very visible." India is no exception but is at the top of the list of vulnerable countries. The uncertainties imposed on the climate have already displaced and destroyed large numbers of people and their households. Every year the list of casualties due to extreme weather conditions increases. Regions with high levels of regular rainfall now receive less than the average amount of precipitation, which in turn creates a scarcity of water where water was abundant even 20–30 years ago. As the destruction and exploitation of natural resources continue, the situation grows more alarming. The recent hostilities between Kerala and Karnataka over the sharing of Periyar Dam water are a vivid representation of the impact of climatic aberrations. Sundarban Biosphere Reserve is one of the world's largest mangrove forests. Popular for its Royal Bengal tigers, it is situated in the eastern peninsular. The reserve is home to crocodiles, king cobras, and many other rare and endangered species. The impact of climate change has influenced the ecological equilibrium of this biosphere reserve. The saline water intrusion and regular decrease in yearly rainfall have limited the water availability of the islands. The uncontrolled extraction of wooden resources has reduced

M. Majumder (✉)
School of Hydro-Informatics Engineering, National Institute of Technology Agartala, Barjala, Jirania 799055, Tripura, India
e-mail: mmajumder15@gmail.com

R.N. Barman
Department of Production Engineering, National Institute of Technology Agartala, Barjala, Jirania 799055, Tripura, India
e-mail: rahul.barman@yahoo.co.in

M. Majumder and R.N. Barman (eds.), *Application of Nature Based Algorithm in Natural Resource Management*, DOI 10.1007/978-94-007-5152-1_12,

the forest cover that had earlier acted as a protection from cyclonic storms. Alarmingly very few studies have been conducted to measure the level of climatic impacts on the Sundarban Islands or its inhabitants. The availability of fresh water is a matter of concern for the livelihood of the people and the endangered wildlife living on the islands. But no study to identify climatic impacts on the availability of freshwater to local habitants has been undertaken to date. That is why the present project tries to quantify the effects of climate change on freshwater availability with the help of a new index known as hydrologic sensitivity. A combinatorial database considering all possible situations was constructed and applied to predict hydrologic sensitivity, which represents climate change impacts on water resource conditions of the world's only mangrove tiger land.

Keywords Artificial neural network • Freshwater availability • Sundarban biosphere reserve

12.1 Introduction

Sundarbans has a total area of 10,200 km^2 of mangrove forest within which 4,200 km^2 of reserved forest is spread over India and approximately 6,000 km^2 is available in Bangladesh. Another 5,400 km^2 of nonforest area in India, along with the northern and northwestern fringes of mangrove forest, is included in the Sundarban regions, but it is inhabited by humans. The total area of the Sundarban region in India is thus 9,600 km^2, which constitutes the Sundarban Biosphere Reserve. The Sundarbans in India is bound by the Muriganga River and the Harinbhahga and Raimangal rivers in the east and west, respectively. The Matla River divides the Sundarban Reserve Forest into the Sundarban Tiger Reserve and reserve forests in the South 24 Parganas Forest Department.

Sundarban Tiger Reserve was recognized as a World Heritage Site by UNESCO in 1987 and nominated by Government of India for recognition as a Ramsar Site (a wetland of international importance). Sundarban Tiger Reserve was constituted under the Project Tiger scheme in 1973 as the only mangrove forest in the world that is home to the tiger. "Sundarban Tiger Reserve has the highest tiger population in the world" (Sundarban Biosphere Reserve 2011).

12.1.1 Importance of the Sundarban Biosphere Reserve

Sundarban has a rich diversity of flora and fauna enhancing both aquatic and terrestrial ecosystems. The island's highly productive ecosystem acts like a natural fish nursery. Mangroves in the Sundarban also reduce the impact of cyclonic storms and prevent erosion due to tidal waves. Millions of people depend on the Sundarban ecosystem for their livelihood and sustenance through fishing and the collection of honey, firewood, and timber.

12.1.2 *Flora and Fauna*

The Sundarbans at the mouth of the Bay of Bengal in India has a vast area of mangrove forests with a wide variety of indigenous and exotic species of flora and fauna surviving in one of the most unique ecosystems of the world. The natural environment and coastal ecosystem of this biosphere reserve is under threat of physical disaster due to unscientific and excessive human interference and climate change impacts. The region is home to the following flora and fauna, justifying its broad popularity.

12.1.2.1 Flora

The mangrove vegetation of Sundarbans consists 26 species of true, 29 species of associate and old mangroves each along with 64 varieties of Mangal. (Government of West Bengal, 2012, FLORA DIVERSITY, Department of Sundarbans Affairs, Retrieved from http://www.sadepartmentwb.org/Flora.htm on 18th January 2013. namely. Genwa (*Excoecaria agallocha*), Kankra (*Bruguiera gymnorrhiza*), Dhundal (*Cannonball mangrove*), Passur (*Xylocarpus mekongensis*), Garjan (*Rhizophora* spp.), Sundari (*Heritiera fomes*), and Goran (*Ceriops decandra*).

12.1.2.2 Fauna

The forests in the Indian Sundarbans are home to more than 250 tigers, fishing cats, leopard cats, macaques, wild boar, Indian grey mongoose, fox, jungle cat, flying fox, pangolin, chital, and king cobras.

In 1878, Hunter's Statistical Report about the Sundarbans stated that tigers, leopards, rhinoceros, wild buffaloes, wild hogs, wild cats, barasinga, spotted deer, hog deer, barking deer, and monkeys were the principal varieties of wild animals found in the Sundarbans, but in the last century, because of "habitat degradation and ecological changes, the faunal compositions in Indian Sundarbans have undergone changes" (SBR 2012) and animals like Javan rhino, wild buffalo, swamp deer, and barking deer went extinct.

According to the 2004 census, the tiger population in the Indian Sundarban is around 274, of which Sundarban Tiger Reserve and South 24-Parganas Forest Division have 249 and 25 tigers, respectively. Other than tigers, there are 58 species of mammals, 55 species of reptiles, and approximately 248 bird species.

Endangered animals like estuarine crocodile (*Crocodilus porosus*), fishing cat (*Felis viverrina*), common otter (*Lutra lutra*), water monitor lizard (*Varanus salvator*), Gangetic dolphin (*Platinista gangetica*), snubfin dolphin (*Orcella brevirostris*), river terrapin (*Batagur baska*), marine turtles like olive ridley (*Lepidochelys olivacea*), green sea turtle (*Chelonia mydas*), and hawksbill turtle (*Eritmochelys imbricata*) are also found in the Sundarbans.

But the reserve is most popular for its tiger population, which is indigenous to the island and attracts the bulk of national and international tourists every winter season. The steady increase in tourism has raised socioeconomic conditions but also threatened the tranquility and neutrality of these historical islands.

12.1.3 Hydroclimatic Conditions of Sundarbans

The region is situated south of the Tropic of Cancer. However, the temperature of the place is equable due to its proximity to the sea. Average annual maximum temperature is around 35°C. Average annual rainfall is 1,920 mm. Average humidity is about 82%, which is more or less uniform throughout the year.

The islands face frequent cyclonic storms and tidal waves. Availability of freshwater is a problem in these regions. In addition, saline water intrusion due to the adjacent sea water and low groundwater level has decreased the options for fresh and potable water. The inhabitants of these islands depend on the rain water pond and harvesting structures for their daily supply of drinking water. Due to the low groundwater level problems of arsenic and iron contamination are also acute.

Cyclone Ayla, which struck the largest mangrove forest on the island, wrecked the drinking water resources of the region and also reduced the fertility of the adjacent lands. Thus, people nowadays have started shifting from cultivation to fish nurseries, which more profitable and easier to cultivate in these regions.

12.1.4 Objective and Scope

The main objective of this study is to predict the hydrologic sensitivity (HS) of the Sundarban Biosphere Reserve (SBR) with respect to different scenarios of climate change due to global warming and anthropogenic impacts.

The HS index is inversely proportional to freshwater availability, i.e., the higher the sensitivity, the less freshwater is available per capita.

Freshwater availability in regions like SBR is directly related to the amount of saline water intrusion, average rainfall, evapotranspiration, and catchment retentivity or water holding capacity.

The index can be estimated using the following equation:

$$\begin{aligned} \mathrm{HS} = {} & (((P-ET)\times A)\times(1-WHC))\times SW)\,|_n/\,\mathrm{p}_n)\,/ \\ & (\text{Summation of }((((P-ET)\times A)\times(1-WHC))\times SW)\,|_n/\,\mathrm{p}_n) \end{aligned} \tag{12.1}$$

where

HS = hydrologic sensitivity (unit less),
P = precipitation in millimeters per day,
ET = evapotranspiration in centimeters per day,
A = area of interest,
WHC = water holding capacity = (Area of pervious land/Total area) (unitless),
SW = ratio of salinity in groundwater extracted per unit area of interest (unitless),
p = total number of inhabitants present on grid,
n = number of grids.

12.1.5 Brief Methodology

As can be seen from Eq. 12.1 the HS index is a function of *P*, *ET*, *A*, *WHC*, *SW*, and p. At first the data set for the combination matrix was prepared by converting each of the input variables into nine different categories. The resulting HS was also categorized into nine groups by degree of intensity/quality/quantity.

A set of combinations was generated from these grouped input and output variables to prepare the entire combination matrix, which would represent every situation that might arise due to climatic uncertainty.

The combinatorial matrix was then fed to a neurogenetic model for estimation of HS every combination of input variables.

The original data set was then categorized to determine HS for different grids of SBR. The average HS of SBR would show the impact of different climate change scenarios on HS.

12.2 Artificial Neural Network

Artificial neural network (ANN) models are prepared with the help of input variables and weighted synaptic connections between the input layer and hidden layer and between the hidden layer and output layer. The weighted sum of all the input variables is converted into a function using various types of transformation functions like tanh, sigmoidal, sinusoidal, etc., which are referred to as activation functions. The functions of summation of weighted inputs are taken as the input to the hidden layers. The output from the hidden layers is the summed input to the same layers transformed by the different activation functions. After the summation and conversion operations the converted summation is equalized with the original values of the output. The weights of the weighted inputs are constantly updated to reduce the difference between the converted summations and the desired output value.

A weighted update is performed by different training algorithms like gradient descent, back propagation, and various types of supervised and unsupervised algorithms.

When the difference between the estimated and observed value becomes zero or satisfies the targeted difference, then the algorithm stops the weighting upgrade and retrieves the optimal weighting for predictions from input variables that have no observed output.

The ANN was found to be highly successful compared to nature-based algorithms and conceptual/statistical models. But the drawback of ANN is that no studies have identified the optimal method of weighting update, i.e., training algorithms and topology selection (whether one hidden layer is enough or more than two are necessary) for optimal performance.

Many scientists have used ANNs to predict the output of complex relationships. Many relationships have been clusterized using neural networks with a higher accuracy than that of comparable models.

Table 12.1 Relevant studies that achieved their objectives using neural networks and their different subforms

References	Field of study	Type of neural network used	Success rate
Alvisi and Franchini (2011)	Water level or discharge estimation under uncertainty	A fuzzy neural network where weights and biases of network were represented by fuzzy rather than crisp numbers	Compared to Bayesian neural network and local uncertainty estimation model, fuzzy neural networks show better performance
Xiong et al. (2011)	Multidimensional flow dynamics	Multilayer perceptron, modular neural network with self-organizing map for soil-moisture contour generation	Successful
Ceyhun and Yalçın (2010)	Deriving bathymetry map of shallow coastal water using quick bird pan sharpened images and some in situ depth measurements	Simple neural network model	Successful
Zou et al. (2010)	Water conductivity and water content of silty-loam soil was predicted using experimental data covered a 6 year period	Time series neural network applied in conjunction with Box–Jenkins model	According to performance metrics like mean absolute error, coefficient of determination, and sum of squared errors, the neural network model was found to be better than autoregressive integrated moving average (ARIMA)
Chaves and Kojiri (2007)	Reservoir optimization and water quality control	Stochastic fuzzy neural network trained with genetic algorithm	Successful
Ha and Stenstrom (2003)	Identification of land use from water quality data of stormwater	Bayesian neural network with water quality parameters as input, four hidden layers, and four target five land use classes	92.3% correct classification rate
Bazartseren et al. (2003)	Short-term water level prediction of river reaches with very small and incomplete data set	Artificial neural network and neuro-fuzzy systems	Both neural network and neuro-fuzzy models were found to perform better than linear statistical models

12.2.1 Applications

ANNs have been widely used to attain objectives in different fields and specializations. Data flexibility, ease of development, and minimum data requirements have made neural networks a modeling methodology of choice for many researchers. Table 12.1 shows some of the relevant applications of neural networks in accomplishing different water-related objectives. But despite the popularity and success of neural networks, a lack of proper and established specifications for development methodology and data requirements (Maier et al. 2010), along with their dependence on trial and error in selecting network topologies, activation functions and training algorithms, has raised questions among many neural network skeptics.

12.3 Methodology

A three-phase approach was followed for the neural network model developed for this study. First, however, the data were preprocessed and grouped by degree of intensity/quality/quantity. A combinatorial data set containing all possible combinations of input variables that might arise was prepared.

Based on the groups, the entire data set was scored based on a review of the literature. The rated or scored data set was used to estimate the HS in numerical terms. The HS was then broken down into nine groups based on the scoring.

Thus the preprocessed data set was fed to the neural network models and prediction work was performed using the same data but with the help of neural network software.

12.3.1 Selection of Network Topology

The first task in the development of a neural network model is to identify those input variables that are not a function of any other input variables. In the present investigation four search methods for selecting the network topology were applied:

1. Linear search method,
2. Nonlinear search method,
3. Other search methods
4. Genetic algorithm (GA)/other nature-based heuristics.

According to the evaluation criteria, the GA was found to be effective enough to find the best possible network topology.

12.3.2 Selection of Training Algorithms

Training algorithms are used to update the weighting of the input synapse such that the difference between actual and predicted is minimized or at the desired value. In this study, a training algorithm that developed a model with better accuracy, correlation, and standard deviation emerged as the preferred model; training was also done to further develop the model, and updated weightings were used to predict output.

12.3.3 Selection of Activation Function

The selection of activation function was also performed with the help of statistical fitness functions that made it possible to identify the best model. The activation functions used for the model development were:

1. Sigmoidal,
2. Tanh,
3. Sinusoidal,
4. Gaussian transformation and many other functions.

12.3.4 Fitness Functions

The fitness functions used to evaluate the performance of the neural network models predicting a numerical output are as follows:

1. Mean square error,
2. Correlation coefficient,
3. Coefficient of efficiency,
4. Absolute error,
5. Relative error.

In the case of classification problems, sensitivity, specificity, and precision were the most commonly used performance metrics (Ranawana and Palade 2006) for evaluating the accuracy and reliability of the classification algorithms.

Sensitivity can be defined as the proportion of positive patterns correctly recognized as positive, whereas specificity is the proportion of negative patterns identified as negative. Precision is the percentage of samples correctly classified.

In the present study the best network architecture for the given categorized input and output variables was determined using genetic searching algorithms. After the optimal network was identified, conjugate gradient descent was applied for updating the weights. The prediction for the known inputs was then compared with the observed output data for validation. The model was named CGDNNGA. In the next step, the exhaustive network searching algorithms were used to find the best network topology. The optimal network was now trained with the help of a quick propagation algorithm, and the output of the known inputs was compared with the observed data set. The models were compared with respect to the different performance metrics (sensitivity, specificity, precision), and the better of the two models was selected for prediction tasks. The latter model was named QPNNEA.

In both cases the network was trained for 1,000,000 iterations in 10 continuous runs. A 5% generalization loss and 3.5% of outliers were allowed to be included in the training data set. The desired correct classification rate was fixed at 97%, whereas the minimum error improvement required to continue training was set at 0.0001 for the last ten iterations.

The sensitivity, specificity, and precision performance metrics were used to select the better model.

12.4 Result and Discussion

The results of the topology searching step are given in Table 12.2, and the performance metrics of the two models are depicted in Table 12.3. Because the training data set was grouped into nine categories, it was preprocessed and converted into a binary data set of 1 and −1. The 6 input variables were converted into 51 numerical columns, and 1 output variable was transformed into 8 numerical columns. After training the converted data were reverted to their original form, and the output variable was predicted accordingly.

According to the performance metrics of the four models, precision was found to be higher in the case of the QPNNEA model than in the CGDNNGA model. Sensitivity was higher in the QPNNEA model (0.697) than in the other neural networks, whereas CGDNNGA was found to have the lowest rating of 0.577. In the case of specificity was identical in QPNNEA and CGDNNGA (0.987). Because the QPNNEA model was better than the other model with respect to precision (0.910), training error (93.23%), and sensitivity, that model was selected for prediction of the output variable with respect to the various scenarios of the input variables.

Following selection of the ideal model, four scenarios of climate change were created with the help of the input variables. The model was used to predict the output with respect to the given input scenarios. The scenarios were created based on the IPCC scenarios. The following section describes the considered scenarios.

12.4.1 IPCC Climate Uncertainty Scenarios

Scenario 1

According to the IPCC A1 scenario, the world will be united (“homogenous”) and rapid industrial and economic growth will be observed at the cost of environmental sustainability.

Table 12.2 Network training parameters of neural network models

Model name	Network topology	Network weight	Activation function
MODEL: QPNNEA	6-11-1	77	Logistic
MODEL: CGDNNGA	6-3-1	21	Logistic

Table 12.3 Model parameters and performance metrics of four neural network models

SLNO	Model name	Training error	Testing error	Sensitivity	Specificity	Precision
1	MODEL: QPNNEA	93.23	83.56	0.697	0.987	0.910
2	MODEL: CGDNNGA	90.97	90.41	0.577	0.987	0.905

Table 12.4 Situation of input variables in SBR with respect to different future scenarios as proposed by IPCC

Scenario	Input variable	Category assigned	Reason
A1	P	EL	In A1 scenario the world will respect no environmental restrictions, which will sacrifice natural ecosystems for the development of industrial zones. Increase in population will be rampant due to economic stability and urbanization will be at its maximum limit. Thus, global warming and subsequent changes in the climate will reverse the seasonal variation of many countries. SBR is under the Tropic of Cancer, which has frequent, heavy rainfall. But due to climate change, rainfall will decrease sharply. Due to the loss in land cover, erosion will increase, and due to urbanization, mountainous regions will be leveled to accommodate the increasing number of households, which will also increase the area of the watersheds
	ET	VH	
	WHC	VL	
	SW	VH	
	p	VH	
	A	H	
A2	P	EL	The situation in the A2 scenario will be similar to the situation under A1. The only difference will be that in the former scenario the world will be regionally divided and in the latter scenario the world will be united. But the environmental degradation and climatic uncertainties will increase similarly to the A1 scenario. Thus, the groups assigned to the input variables will also be similar to the earlier scenario
	ET	EH	
	WHC	EL	
	SW	EH	
	p	EH	
	A	EH	
B1	P	EH	In the B1 scenario the world will be global and strict enforcement of environmental laws will be maintained. Thus, industrial growth will be sharply curtailed and environmental sustainability will be maintained. Precipitation will increase, temperatures will decrease, erosion will be reduced, and population will be under control
	ET	EL	
	WHC	EH	
	SW	M	
	p	M	
	A	M	
B2	P	EH	In the B2 scenario the world will be divided but strict enforcement of environmental laws will be maintained. Thus, industrial growth will be sharply curtailed and environmental sustainability will be maintained. Precipitation will increase, temperatures will decrease, erosion will be reduced, and population will be under control. Population may reduce by more than under the B1 scenario due to the regionalism that will exist at that time
	ET	EL	
	WHC	EH	
	SW	M	
	p	L	
	A	M	

Table 12.5 Prediction of hydrologic sensitivity from selected network for four IPCC climate change scenarios

Scenario	P	ET	A	WHC	SW	p	HS
1(A1)	EH	EL	M	EH	M	L	VH
2(A2)	EH	EL	M	EH	M	M	SH
3(B1)	EL	EH	EH	EL	VH	EH	SH
4(B2)	EL	EH	H	VL	VH	VH	SH

Scenario 2
The IPCC A2 scenario considers that the world will not be united (“heterogenous”) but there will be mutual agreements within countries. Like the A1 scenario, in A2 industrial development will overwhelm environmental conservation.
Scenario 3
The IPCC B1 scenario depicts the world as global (“homogenous”), but environmental conservation will be given priority over industrial development. Global environmental sustainability will be achieved in this scenario.

Scenario 4
This scenario mimics the IPCC B2 scenario where the world is regional but environmental stability is observed in all countries. Industrial development will be minimum (Table 12.4).

The predictions by the selected neural network for the climate change scenarios as conceptualized by IPCC A1, B1, A2, B2 are presented in Table 12.5. According to the resulting data set, it can be easily concluded that although in Scenarios 2–4 the HS index was found to be Semi High, which implies that enough freshwater would be available for consumption, only Scenario 1 had an HS score of Very High, which means that freshwater for consumption will be scarce.

12.5 Conclusion

The present investigation tried to estimate freshwater availability in the SBR. In this regard an index was conceptualized and used to rate the different input variables for estimating the hydrologic sensitivity of the region. A prediction was made for each of the future scenarios, and except for Scenario 1, all the other scenarios predicted a semi-high, HS, i.e., the model predicts an availability of small amount of water may plague the human in the near future when rapid industrial development will come at the cost of environmental sustainability. In Scenario 1 the model predicts that there will be very high (VH) stress on freshwater availability. All the categories used for the different input variables to predict the output under future scenarios were based on a PRECIS simulation of future climate in India and Bangladesh. The entire SBR falls within these two countries’ eastern and western borders respectively. In light of the combinatorial data set, it may be noted that it was not fed into the model in its entirety. Only the patterns that had not been duplicated were fed into the model for training. In the future studies, people may just categorize their data and feed the data set into the model for an indication of the freshwater scenario.

References

Alvisi S, Franchini M (2011) Fuzzy neural networks for water level and discharge forecasting with uncertainty. Environ Model Softw 26(4):523–537

Bazartseren B, Hildebrandt G, Holz P (2003) Short-term water level prediction using neural networks and neuro-fuzzy approach. Neurocomput 55(3–4):439–450

Ceyhun Ö, Yalçın A (2010) Remote sensing of water depths in shallow waters via artificial neural networks. Estuar Coast Shelf Sci 89(1):89–96

Chaves P, Kojiri T (2007) Deriving reservoir operational strategies considering water quantity and quality objectives by stochastic fuzzy neural networks. Adv Water Res 30(5):1329–1341

Ha H, Stenstrom MK (2003) Identification of land use with water quality data in stormwater using a neural network. Water Res 37(17):4222–4230

Maier HR, Jain A, Dandy GC, Sudheer KP (2010) Methods used for the development of neural networks for the prediction of water resource variables in river systems: current status and future directions. Environ Model Softw 25(8):891–909

Ranawana R, Palade V (2006) Optimized precision – a new measure for classifier performance evaluation. In: Proceedings of the IEEE congress on evolutionary computation, pp 2254–2261

Sundarban Biosphere Reserve (2011) Mangrove - West Bengal Tourism, Retrieved from http://www.westbengaltourism.gov.in/web/guest/mangrove. Accessed 18 Jan 2013

Xiong Y, Wallach R, Furman A (2011) Modeling multidimensional flow in wettable and water-repellent soils using artificial neural networks. J Hydrol 410(1–2):92–104

Zou P, Yang J, Fu J, Liu G, Li D (2010) Artificial neural network and time series models for predicting soil salt and water content. Agric Water Manag 97(12):2009–2019

Chapter 13
Fuzzy-Based Impact Analysis Study on Site Selection of Tidal Power Plants

Soumya Ghosh, Mrinmoy Majumder, and Debasri Roy

Abstract According to the current scenario of electricity production, thermal power is the largest supplier to the ever increasing demand for electricity. But as the amount of coal stored is finite and thermal power plants are one of the major sources of greenhouse gas emissions, presently such power plants are not preferred by engineers. Scientists are now looking for alternative sources of energy like solar, wind, geothermal, or hydro power. The biggest advantage of such sources is they are infinite, but the inability to store such energy has prevented the wide-scale use of renewable energy for meeting the ever-rising demand. The potential of hydropower for the generation of electricity is immense, but, due to the irregularity in frequency of discharge in rivers, such sources of electricity production are unable to replace conventional energy sources. In the case of tidal power, power produced from tidal waves, the selection of a suitable site can alter the stability of power production. The present study tries to identify suitable sites for tidal power generation in the Sundarban region of West Bengal. The ***C**limate-**O**ptimized **B**asic fuzzy-**A**lgo for identification of **L**ocation for **T**idal power* (COBALT) algorithm was developed to identify the optimal location for tidal power plants among available options.

Keywords Site selection • Climate change • Tidal power

S. Ghosh (✉)
School of Hydro-Informatics Engineering, National Institute of Technology Agartala,
Barjala, Jirania 799055, Tripura, India
e-mail: soumyaee@gmail.com

M. Majumder
School of Hydro-Informatics Engineering, National Institute of Technology Agartala,
Barjala, Jirania 799055, Tripura, India
e-mail: mmajumder15@gmail.com

D. Roy
School of Water Resources Engineering, Jadavpur University,
Kolkata 700032, West Bengal, India
e-mail: debasri_roy1@yahoo.co.in

M. Majumder and R.N. Barman (eds.), *Application of Nature Based Algorithm in Natural Resource Management*, DOI 10.1007/978-94-007-5152-1_13,

13.1 Introduction

The rising global population has created a large demand for electricity, which has put thermal power plants under stress. Although many alternative sources of energy are available, because of their limited availability and the complexity of storage, use of such energy sources is expensive. For example, hydropower is a renewable energy source that converts kinetic energy of streamflow into electrical energy. But because of the infrequent availability of discharge, the production of electricity is irregular during most of the year. That is why hydropower plants are not profitable for all the seasons. Tidal power, however, which generates electricity by utilizing the kinetic energy of tidal waves, is a regular phenomenon, but the energy of tidal waves varies with time and location.

In this study a suitable location for tidal power plant was identified in the Sundarban region of the state of West Bengal. The novel COBALT algorithm was developed which with the help of GIS and Neuro-genetic algorithms identify the better location for tidal power plant within the many available options. Sundarban region was selected as there are large number of channels or “khari’s” which are suitable to produce tidal power.

13.1.1 Selection of Sites for Tidal Power

Tidal power is a function of the difference between the depth inside and outside of a channel. The greater the difference, the greater the potential of tidal power. In the feasibility analysis of a site for the installation of tidal power plants, the following criteria were analyzed. According to various scientific studies, for example, Blunden et al. (2013), Hammar et al. (2012), Abundo et al. (2012), Defne et al. (2011), Bryden and Couch (2006), Blunden and Bahaj (2006), and Mettam (1978), the following parameters are found to be influential in the selection of a location for tidal power generation:

1. Socioeconomic factors: the profitability of a project, amount of rehabilitation required, possible cost of power, earning of carbon credits, labor costs, logistics, impact on navigation capability of channel, etc.;
2. Physical factors: in-stream potential, turbulence, interconnection distance, soil stability;
3. Environmental factors: disturbance to flora and fauna populations within and outside the stream.

Accordingly, this study selects the most important parameters from among those listed above for the selection of locations of tidal power plants. Environmental constraints not considered within the study and channels within the reserve forest area were totally ignored. The channel with the highest concentration of local population was removed from the sample population. After applying these two constraints, the remaining number of channels was considered for selection of most suitable location for a tidal power plant.

13.1.2 Study Objective and Scope

In the present study the potential of tidal power generation in the Sunderban Biosphere Reserve was identified with the help of the COBALT algorithm. Bays inside the Sunderban National Park were not included in the study as those places could not be used to generate tidal power due to the presence of wild and endangered animals.

To date there have been very few studies that consider the impact of climate uncertainty on selection of locations for tidal power plants. That is why the present study was conceptualized and conducted for the analysis of the impact of climate change on tidal power resources.

The COBALT algorithm uses a simple logic of weighted average of input variables or factors based on fuzzy logic.

13.1.3 Brief Methodology

Satellite imagery of the southern Sunderbans was retrieved first and the bay area was delineated with the help of existing toposheets of the study area. Once the bay area was identified, bathymetric maps of the channel openings were prepared. The input variables or factors considered for the selection of the most suitable region for tidal power generation are as follows:

Power Potential (P)

Average power (P_{avg})

Maximum power (P_{max})

Minimum power (P_{min})

Turbulence (T)

Interconnection point (IP)

Net profit (NP)

The weighting for each of the variables was determined with the help of fuzzy logic by comparing each input with all other variables and analyzing them with respect to their importance in the selection of a site for tidal power plants.

The output of the weighted average is developed in such a way that its value would show the suitability of the bay for the generation of tidal power. Because the suitability of sites for tidal power plant varies inversely with turbulence in the channel and length of the interconnection point, a negative weighting was considered for these two variables.

According to the value produced by COBALT, all the locations were rated and the channel with maximum ratings was selected as the better channel for generation of tidal power among all the options selected for analysis.

13.1.4 Study Area: Sundarbans

Sundarbans (21°30′–22°30′N, 89°12′–90°18′E) is a World Heritage Site that consists of three wildlife sanctuaries (Sundarbans West, East, and South) lying on disjunct deltaic islands in the Sundarbans Forest Division of Khulna District, close to the border with India and just west of the main outflow of the Ganges, Brahmaputra, and Meghna rivers. The Sundarbans belongs to the Bengalian Rainforest biogeographical province.

The total area of the Bangladesh section of Sundarbans is 595,000 ha of which 139,699 ha are protected as follows: Sundarbans West Wildlife Sanctuary with 71,502 ha, Sundarbans East Wildlife Sanctuary with 31,226 ha, and Sundarbans South Wildlife Sanctuary with 36,970 ha. Sundarbans National Park (133,010 ha), a World Heritage Site, lies to the west in India.

13.2 Methodology

Tidal power plants use change in water heads during tide and ebb and can generate power in both flow directions.

A number of authors have developed equations for converting energy provided by the tides in estuaries. Bernshtein et al. established a simplified theory for estimating the magnitude of tidal power possible to generate with the help of change in level data. This theory is based on changes in tidal level on both sides of the barrage and the discharge through the barrage and relates to the generated power. The difference in level H_b can be calculated as the difference between the level of the reservoir and the estuary. The water level in the reservoir is denoted by $H_r(t)$. The water flow used for generating electricity can be calculated from the variation of the reservoir level H_r with respect to time:

$$\mathrm{Q}(t) = \frac{\delta H_r}{\delta t} \times A_r(t), \tag{13.1}$$

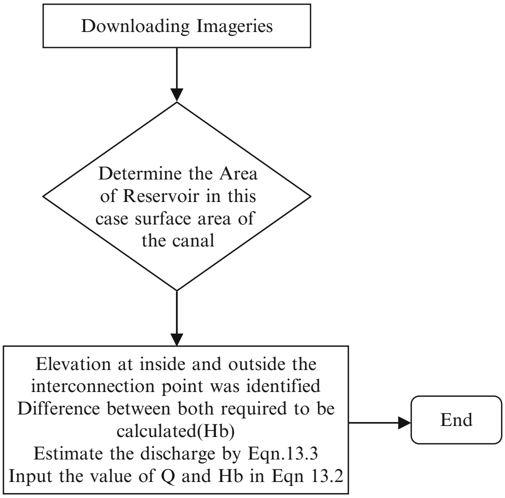
where Q is the flow in cubic meters per second and A_r is the area of the reservoir in square meters.

The power generated in a full tidal cycle can be obtained through the equation

$$\rho \times g \times Q \times H_b, \tag{13.2}$$

where ρ is the density of water, g the gravitational force, Q the area of the image, and H_b the difference between the inner and outer depths.

Fig. 13.1 Basic Methodology of Determination of Tidal Energy from Satellite Imageries



13.2.1 Determination of Stream Resources

The difference in level H_b can be calculated as the difference between the level of the reservoir and the estuary, which was estimated from the highest point inside the channel but closer to the interconnection point and the minimum water level at the exterior of the same (interconnection point). The value of water level difference changes with time and thus represented by $H_{r(t)}$.The water flow used for generating electricity can be calculated from the variation in the reservoir level H_r with respect to time:

$$\mathrm{Q}(t) = \frac{H_r}{t} \times A_{r(t)}, \tag{13.3}$$

where Q is the flow in cubic meters per second and A_r is the area of the reservoir in square meters.

The power generated in a full tidal cycle can be obtained through the equation

$$\mathrm{P} = \rho \times g \times Q \times H_b, \tag{13.4}$$

where ρ is the density of water, g the gravitational force, Q (Eq. 13.3), and H_b is the difference between the inner and outer water level. The aim of the energy model is to estimate the amount of energy generated. After the potential power was determined, the average, maximum, and minimum power generated throughout the year was determined and channels were ranked with respect to the value achieved. Figure 13.1 depicts a graphical explanation of the methods implemented to estimate in-stream resource and power potential of the canals.

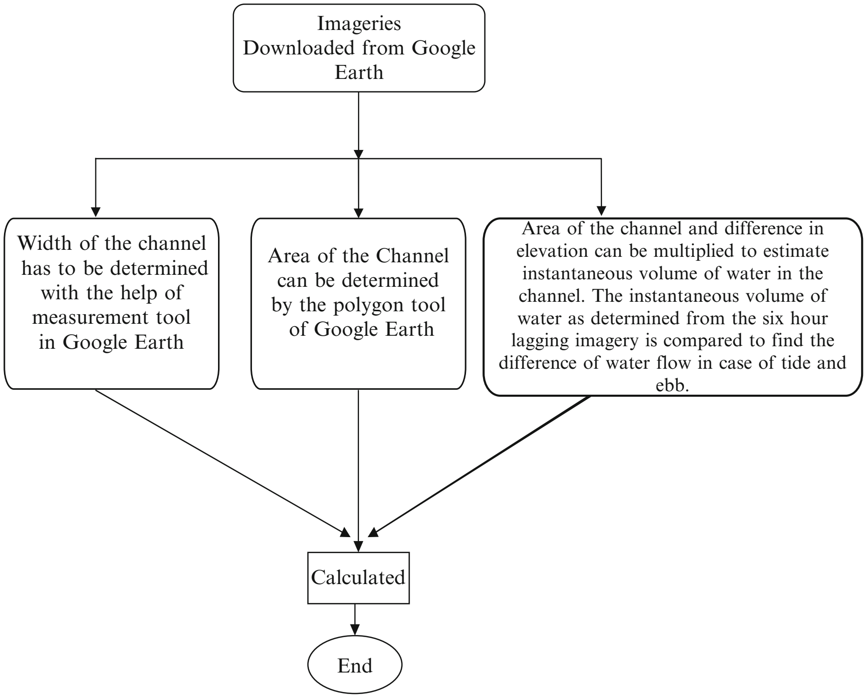


Fig. 13.2 The process of calculating Instream resources

13.2.2 Determination of Turbulence

The images of the bay area were captured at 6-h intervals. Latitutde (*X*), longitude (*Y*), and difference in water level (*Z*) were determined with the help of a topographic map of Google Earth.

The main equation used in describing the open channel flow is Bernoulli's equation. The equation can be followed by

$$\mathrm{H} = \frac{V^2}{2g}, \tag{13.5}$$

where H is the total flow energy, *V* the flow velocity, and *g* the gravitational acceleration. The change of water level was identified by comparing the six hour interval MODIS imageries. The change of head along with the width from the bank to the point was also calculated with the measuring tools in GIS Software. As discharge (Q) (Eqn.13.3) is known; the value of area, found by multiplying the change in head and width,was divided to find the flow velocity at strategic points.

The next step was to determine the pixel value of the same strategic points.

Once the relationship between velocity and their corresponding pixel values were determined the turbulence at the strategic points which are near to the interconnection can be determined map of the bay area can be prepared showing variation of velocity. Equation 13.7 represents the equation which can estimate the velocity from the value of pixels at the strategic points.

$$\mathrm{Y} = 0.0025X^2 - 0.0789X + 0.616 \tag{13.6}$$

as a P1 value, where X is the pixel. Different imageries has different equations to estimate velocity for different values of pixels.

$$\mathrm{Y} = a + \frac{b}{\left(1 + ((x-c)/d)^2\right)} + \frac{g}{\left(1 + \left(\frac{x-h}{k}\right)^2\right)} \tag{13.7}$$

as a P9 value, where $a = -52.196409$, $b = 29214.454$, $c = 16.002186$, $d = 0.7353794$, $g = 29216.038$, $h = 61.368768$, and $k = 0.35908545$.

13.2.3 Low-Cost Interconnection Point

As was stated previously, images were downloaded with the help of Google Earth. The chosen images of the canal were identified and their width at the mouth was determined with the help of measurement tools of Google Earth.

13.2.4 Determination of Net Profit

Net profit can be determined as net income minus expenses.

$$\text{Net Profit} = \text{Income}\,(I) - \text{Expenditure}\,(E) \tag{13.8}$$

$$\begin{aligned} \text{Income} &= 1.\ \text{Rate of electricity} \\ \text{Expenditure} &= 1.\ \text{Labor auxiliary power cost} \\ &\quad\ 2.\ \text{Rate of cable per unit length.} \\ &\quad\ 3.\ \text{Distance from grid} \\ &\quad\ 4.\ \text{Interconnection point.} \end{aligned}$$

13.2.5 COBALT Formulation

Finally, after all the values of all the locations for the input variables were calculated, the COBALT formula was applied for the selection of the optimal channel for tidal power generation. The COBALT formula can be defined as

$$
\begin{aligned}
&+\left(w_{p_{\text{avg}}}\times P_{\text{avg}}\right)+\left(w_{p_{\max}}\times P_{\max}\right)+\left(w_{p_{\min}}\times P_{\min}\right)+\left(w_t\times T\right)\\
&+\left(w_{\text{ip}}\times \text{IP}\right)+\left(w_{\text{np}}\times \text{NP}\right)\sum\left(w_p+w_{p_{\text{avg}}}+w_{p_{\max}}+w_{p_{\min}}+w_t+w_{\text{ip}}\right)\\
&\times\left(P+P_{\text{avg}}+P_{\max}+P_{\min}+T+IP+NP\right)
\end{aligned}
\tag{13.9}
$$

where
P_{avg} = average power,
$P_{\max}$ = maximum power,
$P_{\min}$ = minimum power,
T = turbulence,
IP = interconnection point,
NP = net profit.

The value of the COBALT function is normalized and compared with the values obtained from the other channels for the purpose of identifying the most suitable channels for tidal power plants.

The magnitude of the weighting for each variable was determined with the help of fuzzy logic.

The following rule was utilized for ranking the input variable with respect to every other input variable:

If Input N is much more important than Input M, then
give Input N a rank of 1.
If Input N is more important than Input M, then
give Input N a rank of 2.
If Input N is equally important as *Input M, then*
give Input N a rank of 3.
If Input N is less important than Input M, then
give Input N a rank of 4.
If Input N is much less important than Input M, then
give Input N a rank of 5.

Table 13.1 shows the ranking of all variables with respect to each of the input variables. The weighting is also shown in the last column. The weighting was determined in the following way (Eq. 13.10).

Let the worst rank received by any input variable (*M*) compared with another input variable (*N*) be *R*.

Then the weight of each variable will be determined by

$$\text{w} = 1 - \text{Minimum of}\left(r/R\right). \tag{13.10}$$

Table 13.1 The rank of importance received by each input variable with respect to every other variable for normal climate scenario

	Power	Turbulence	Interconnection point	Net profit	Weighting
Power		2	1	2	0.5
Turbulence	4		2	3	0.5
Interconnection point	5	4		4	0.2
Net profit	4	3	2		0.5

13.3 Results and Discussion

There are many bays, or "khari" as they are known to the locals, in the southern part of the Sunderban region. But not all the bays are suitable for the installation of a tidal power plant. Based on the amount of power it would be possible to generate, the width of the channel, and utilization of the channel for navigation, some of the many available locations were ignored initially. Ultimately seven channels (referred to as P1, P3, P4, P5, P6, P9, and P12) were found to be suitable enough for installation of a tidal power plant. Following the methods described in Sects. 13.2.1, 13.2.2, 13.2.3, 13.2.4, and 13.2.5 the channels were rated and the value of the objective function was determined.

Tables 13.1 and 13.2 present the weighting assigned to the input variables following the fuzzy logic theory of maximization principle for normal and changed climate scenarios, respectively. Table 13.3 shows the magnitude of each of the input variables for the selected locations.

According to the results, the following salient points were observed:

- The most power is generated at channel P1 (42.734 MW) and the least at channel P6 (10.256 MW).
- The maximum value of "average power" is obtained at channel P5, and the minimum is obtained at channel P6.
- The highest value of "maximum power" is obtained at channel P5, and the lowest value is obtained at P6.
- The highest value of "minimum power" is obtained at channel P9, and the lowest value of "minimum power" is obtained at channel P6.
- The maximum turbulence occurs at channel P6 (8.1651), and the lowest turbulence is observed at channel P4 (1.8371).
- The maximum interconnection point is at channel P1 (4.75 km), and the lowest interconnection point is at channel P4 (0.66 km).
- The maximum net profit is at channel P1 (Rs.1.52), and the lowest net profit is at channel P6 (–Rs.18.73).

Tables 13.4 show the values of COBALT. According to the results, channel P4 was found to have the maximum and channel P6 the minimum value of the COBALT function under a normal climate scenario, whereas the same channels P4 and P5 were found to have, respectively, the highest and lowest values of the COBALT function. The second highest value of the function was achieved by P1 in the case of

Table 13.2 The rank of importance obtained for each input variable with respect to all other variables under changed climate scenario

	Power	Average power under transiency	Maximum power under transiency	Minimum power under transiency	Turbulence	Interconnection point	Net profit	Weighting
Power		3	2	4	3	3	2	0.5
Average power under transiency	3		2	1	4	4	4	0.75
Maximum power under transiency	4	4		2	3	4	4	0.5
Minimum power under transiency	2	5	4		4	4	4	0.6
Turbulence	3	2	3	2		3	2	0.33
Interconnection point	3	2	2	2	3		4	0.5
Net profit	4	2	2	2	4	2		0.5

Table 13.3 Value of input variables for each of the selected channels

Channel name	Power (MW)	Average power under transiency	Maximum power under transiency	Minimum power under transiency	Turbulence	Inter connection point	Net profit
P1	42.734	1637.515	3082.107	392.037	3.544	4.750	1.519
P3	12.307	589.5053	1109.558	141.133	4.593	2.330	−18.478
P4	36.922	1339.5853	2219.117	282.266	1.837	0.660	−2.266
P5	19.230	81211.132	966845.752	176.417	7.176	0.660	−7.712
P6	10.256	393.004	739.706	94.089	8.165	0.790	−18.732
P9	32.820	1572.015	2958.823	376.356	3.266	2.330	−3.2651
P12	16.410	693.826	1479.411	122.242	8.165	0.790	−11.607

Table 13.4 Value calculated with COBALT under a normal climate scenario

	P	T	I	N	C_1
P1	42.734	3.544	4.750	1.519	0.996
P3	12.307	4.593	2.330	−18.478	0.547
P4	36.922	1.837	0.660	−2.266	1.000
P5	19.230	7.176	0.660	−7.712	0.982
P6	10.256	8.165	0.790	−18.732	0.000
P9	32.820	3.266	2.330	−3.266	0.996
P12	16.410	8.165	0.790	−11.607	0.968

normal climatic conditions and P9 in the case of a changed climate scenario. Hence, channel P4 was found to be most suitable for tidal power generation among the considered channels of this study. As the interconnection point and turbulence in P4 are at their minimum and the in-stream resource of the channel is next to maximum (P1), the function selects that channel as the most suitable, although the net loss from the channel is Rs.2.26/unit/day. But as the cost was calculated considering the installation cost only and tidal power plants have low maintenance costs, the rated loss could be converted into reasonable profit after 5 years by which time the setup cost will have been recovered.

13.4 Conclusion

This study tried to select suitable tidal power generation sites in West Bengal coastal regions under normal climate scenarios. The objective was to select the best possible channel from among the many channels that exist in the Sundarban region, for the generation of tidal power. In this regard, GIS and remote sensing tools were utilized and an algorithm, COBALT, was developed that identified areas in the Sundarban Biosphere Reserve what would be suitable as sites of tidal power plants based on a weighted average value in which the weights of the variables were determined by fuzzy logic. The study considered net profit, turbulence, length of interconnection point, and in-stream resource for estimation in COBALT under normal climatic conditions. According to the results, channel P4 was selected due to its lower interconnection point and second best potential of tidal power generation. The limitation of the study was that the marine population and soil strength were not considered due to a lack of financial support, but in the near future the study could be repeated to include those variables. The analysis could be repeated to identify the suitability of hybrid power plants (solar and tidal because SBR has high solar insulation and barren lands that could be utilized for the installation of solar panels around the tidal power plant).

References

Abundo M, Nerves A, Paringit E, Villanoy C (2012) A combined multi-site and multi-device decision support system for tidal in-stream energy. Energy Procedia 14:812–817

Blunden LS, Bahaj AS (2006) Initial evaluation of tidal stream energy resources at Portland Bill, UK. Renew Energy 31(2):121–132

Blunden LS, Bahaj AS, Aziz NS (2013) Tidal current power for Indonesia? An initial resource estimation for the Alas strait. Renew Energy 49:137–142

Bryden IG, Couch SJ (2006) ME1—marine energy extraction: tidal resource analysis. Renew Energy 31(2):133–139

Defne Z, Haas KA, Fritz HM (2011) GIS based multi-criteria assessment of tidal stream power potential: a case study for Georgia, USA. Renew Sustain Energy Rev 15(5):2310–2321

Hammar L, Ehnberg J, Mavume A, Francisco F, Molander S (2012) Simplified site-screening method for micro tidal current turbines applied in Mozambique. Renew Energy 44: 414–422

Mettam C (1978) Environmental effects of tidal power generating schemes. Aquat Ecol 12(3–4): 307–321

Part IV
New Computer Models

Chapter 14
Estimation of Groundwater Quality from Surface Water Quality Variables of a Tropical River Basin by Neurogenetic Models

Mrinmoy Majumder and Bipal Kr. Jana

Abstract According to the hydrological cycle, after rainfall infiltration becomes high and once the soil pores become saturated, surface runoff begins. The infiltrated water is added to the groundwater and depressions and canals are utilized to store or drain out excess water. Because surface water and groundwater have the same source, their quality is related, but the physiochemical properties of the soil layers and geological characteristics of the catchments also influence the quality of water in the surface and ground. Many scientific studies have established that surface water is not as pure and fit for drinking as groundwater. Groundwater is free of turbidity, suspended impurities, and organic and inorganic micropollutants. This reactive nature of water is almost neutral. Although groundwater is affected by dissolved metals (like arsenic, iron, etc.), volatile organic compounds and toxic gases, but the intensity of groundwater pollutants varies with location and surrounding geophysical and ecological structures. In most of the places people use groundwater for drinking without adopting any means of purification. If the source is free of organic and inorganic pollutants and if the metal and gaseous concentrations are low, then the ground/surface water can easily be used for drinking or washing purposes without much threat to human health. But if the surface water contaminates the source through leakage or accidental removal of the impervious layers, then it may contaminate the source, and use of the contaminated groundwater could cause affect public health. The present study attempts to predict the quality of groundwater with the help of surface water quality parameters along with some climatic and geophysical parameters. The study utilized neurogenetic models for

M. Majumder (✉)
School of Hydro-Informatics Engineering, National Institute of Technology Agartala, Barjala, Jirania 799055, Tripura, India
e-mail: mmajumder15@gmail.com

B.Kr. Jana
Consulting Engineering Services, "AKARIK," East Tentulberia, P.O. – Panchpota (Near Five Star Club), Kolkata 700152, West Bengal, India
e-mail: bipalkjana@gmail.com

M. Majumder and R.N. Barman (eds.), *Application of Nature Based Algorithm in Natural Resource Management*, DOI 10.1007/978-94-007-5152-1_14,

predicting the quality of groundwater. The results show that predictions of pH and chlorine levels based on the parameters was found to be more accurate and reliable than the prediction of any other quality variables. Thus it can be concluded that if surface water and groundwater are mixed, the pH and turbidity will undergo the most dramatic change among all other quality variables.

Keywords Groundwater quality • Neurogenetic models • Damodar Basin

14.1 Introduction

“Groundwater and surface water are fundamentally interconnected. In fact, it is often difficult to separate the two because they ‘feed’ each other” (CTIC 2008). The quality of surface water thus impacts the quality of groundwater. Because groundwater from unconfined aquifers are extracted for domestic, agricultural, and industrial purposes, showing the interrelationship between the two in a spatial scale is necessary. Reay et al. (1992) established the impact of groundwater discharge on surface water quality of Chesapeake Bay Inlet. The conclusion supports the notion of interactivity between the two types of water and suggests the inclusion of this relationship in water quality management strategies. The impact of surface and groundwater interactivity on herbicide concentration was measured by Verstraeten et al. (1998), and according to the results, the concentrations of one-half to one-fifth of herbicides decrease in groundwater. Gallagher et al. (2007) investigated the transport of land-applied nutrients and pesticides from unconfined aquifers to the tidal surface waters of Virginia’s coastal plain; the study revealed that the levels of nitrogen contaminating the surface water was significant and overall levels of pesticide movement through groundwater, although generally quite low, represented a transport route that is commonly neglected in watershed management. An integrated numerical model was developed to estimate the flow and chemical transport between an integrated surface-subsurface hydrologic system (Van der kwaak 1999). The study concluded that hydrograph separation theory is fundamentally flawed if diffusive modification of tracer concentrations in surface water is prevalent in nature. The aforementioned studies explain the importance of determining the relationship between the main chemical parameters of surface water and groundwater. These studies mainly used field measurements and linear numerical models to estimate that relationship. But linear models were often found to deviate from the actual relationship because of their failure to replicate the temporally or spatially abrupt changes. Also, linear models are unable to separate episodic changes from normal but large changes. In recent years artificial neural networks have been often and successfully applied to the modeling and forecasting of time series. Artificial neural networks offer a relatively quick and flexible means of modeling.

14.1.1 Objective and Scope

This investigation will try to estimate the quality variables of groundwater that has been contaminated by surface water. Each of the quality parameters was examined separately by considering 11 inputs involving climatic, geophysical, and quality parameters of the surface water. Because neurogenetic models can map nonlinearity and the highly complex interrelationship between the input and output variables of the present problem, such models were used to identify the interrelationships between the surface and groundwater quality parameters and other climatic and geophysical properties of the catchment.

The successful development of the model could help engineers to predict the probability of various health hazards if surface water contaminates groundwater. Necessary mitigation measures could be adopted once contamination is detected without having to verify the quality of the groundwater. Thus the compensatory action to recover from the disequilibrium would be fast and control of the situation would be efficient enough to check the spread of the contaminant.

A case study was selected for verification and analysis of the model output where groundwater is often contaminated by surface water. The next section describes the area of investigation and the model development methodology.

14.1.2 Study Area

The Damodar River, which lies between the latitudes 23°30′N and 24°19′N and longitudes 85°31′E and 87°21′E, originates from the Palamu Hills of Chota Nagpur at an elevation of approximately 610 m above mean sea level. It flows in a southeasterly direction, entering the deltaic plains below Raniganj in the Burdwan district of West Bengal, India. Near Burdwan the river abruptly changes course and starts flowing in a southerly direction to join the Hoogli River approximately 48 km below Kolkata. The slope of the river bed during the first 241 km is approximately 1.89 m/km. During the next 161 km the slope is approximately 0.57 m/km and approximately 0.19 m/km over the next 145 km. The river is fed by six streams, of which the principal tributary, Barakar, joins it where the Damodar River emerges from the Palamu Hills. Four main multipurpose reservoirs are located at Tilaiya, Konar, Maithon, and Panchet, and a barrage at Durgapur was commissioned during the period 1953–1959. Another tributary, Khudia, whose catchment is not intercepted by either the Maithon or the Panchet reservoir, joins Damodar near its confluence with Barakar. In the plains, the river splits into several channels and ultimately joins the Roopnarayan and Hoogli rivers. The total length of the river is approximately 541 km, and its total catchment area is 28,015 km^2, of which 10,985 km^2 lies under Panchet (Konar –997 km^2, Tenughat –4,500 km^2, and Panchet 5,488 km^2) and 6,293 km^2 under Maithon (Tilaiya –984 km^2 and Maithon –5,309 km^2). Figure 14.1 shows a schematic diagram of the Damodar River and the location of the Panchet reservoir.

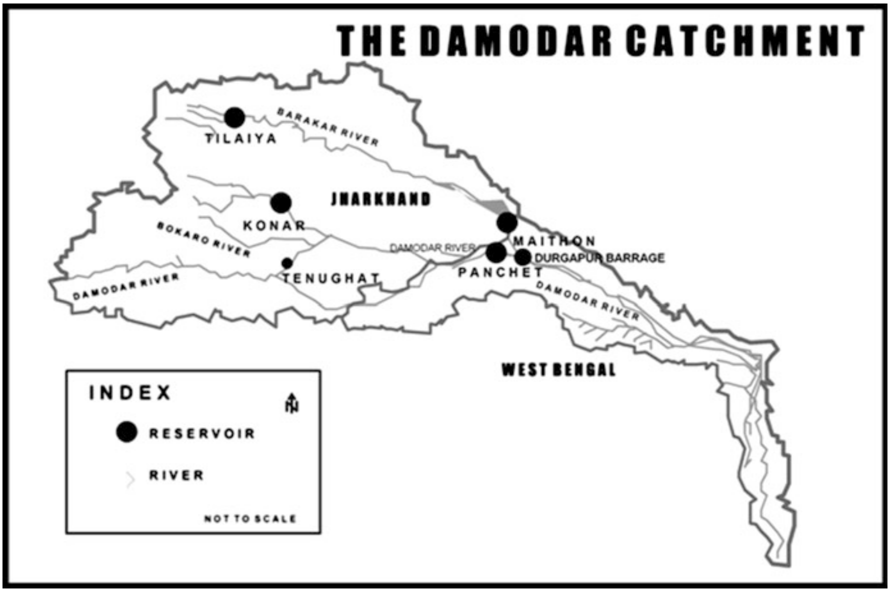


Fig. 14.1 Location of Panchet and Maithon Reservoir in the Damodar Watershed (Majumder et al. 2010)

14.1.3 Brief Methodology

The present study tried to apply the searching capabilities of a neural network and a genetic algorithm to estimate the relationship between surface and groundwater quality of four different locations on the catchments of the Damodar River. The water quality values, as found from field study of surface and groundwater sampling points, along with soil type, surface water discharge, temperature, relative humidity, and rainfall, served as input. Different models were developed to predict the five selected water quality parameters. Even if the same algorithm and training settings were used for all five models, the mean square error (MSE) of the models was different. The reason behind such varied MSEs lies in the degree of the inherent relationship that exists between the values of the output parameter found in surface water and groundwater.

The selection of quality parameters was based on availability of data for both types of water and impact on public health.

14.1.4 Data Description

The average yearly values of maximum temperature (T), maximum and minimum relative humidity (h_{max} *and* h_{min}), discharge (Q), rainfall (P), and concentration of surface water quality parameters like conductivity (cond), turbidity (Turb), chloride (Cl), total hardness (TH), and pH measured from surface water and groundwater at

23 monitoring locations of different areas were considered. Soil type (ST) of the selected locations was also included. The output was quality variables cond*,* Turb*,* Cl*,* TH, *and* pH. One model was developed for each of the output variables. Thus five models were prepared. All the models were trained with three different training algorithms and validated by the three performance metrics MSE, coefficient of relationship (*r*), and standard deviation (STDDEV).

The selected sites were located in and around Tandwa on the bank of the rivers Garhi, Chandrapura, and Maithon, which were located respectively on the banks of the Barakar River, a tributary of the Damodar River, which are respectively in the upstream and downstream of the Damodar River.

The relationships between surface and groundwater quality parameters as measured from the field locations are compared in Fig. 14.2a–e. The correlation coefficient between surface and groundwater Turb, cond, Cl, TH, and pH of the three study areas according to field survey results was found to be equal to 0.5–0.55, 0.02–0.91, 0.6–0.95, 0.57–0.97, and 0.44–0.62, respectively. The relationship between chloride concentration of surface and groundwater was found to be the most prominent and for conductivity it was found to have deviated significantly as the gap between minimum and maximum correlation was nearly equal to 89%. The gap between the maximum and minimum values was a lowly 35%, but the lowest gap was observed in Turb correlation, which was equal to 5%. pH and TH had a correlation gap percentage (CGP) equal to 18 and 40%, respectively. A large CGP indicates the unpredictability of a parameter with respect to the variable, which in this case is a spatial reference number.

14.2 Methodology

14.2.1 Mathematical Modeling of Neural Networks

Artificial Neural Networks are nowadays widely applied (Gaur et al. 2013; Shiri et al. 2013; Ranjithan et al. 1993; Coulibaly et al. 2001; Kuo et al. 2004; Coppola Jr et al. 2003; Nayak et al. 2006; El Tabach 2007)in various field of science, engineering and management. The present study has utilized the advancements of the neural network modeling to predict the impact of surface water contamination of groundwater quality parameters. Simulating an artificial functional model from the biological neuron has three basic components. First, the synapses of the biological neuron are modeled as weights. The synapse of the biological neuron is the one which interconnects the neural network and gives the strength of the connection. For an artificial neuron, the weight is a number, and represents the synapse.

A negative weight reflects an inhibitory connection, while positive values designate excitatory connections. All inputs are added and modified by the weights. This activity is referred as a linear combination.

Finally, an activation function controls the amplitude of the output. For example, an acceptable range of output is usually between 0 and 1, or it could be −1 and 1. The activation functions are of three types:

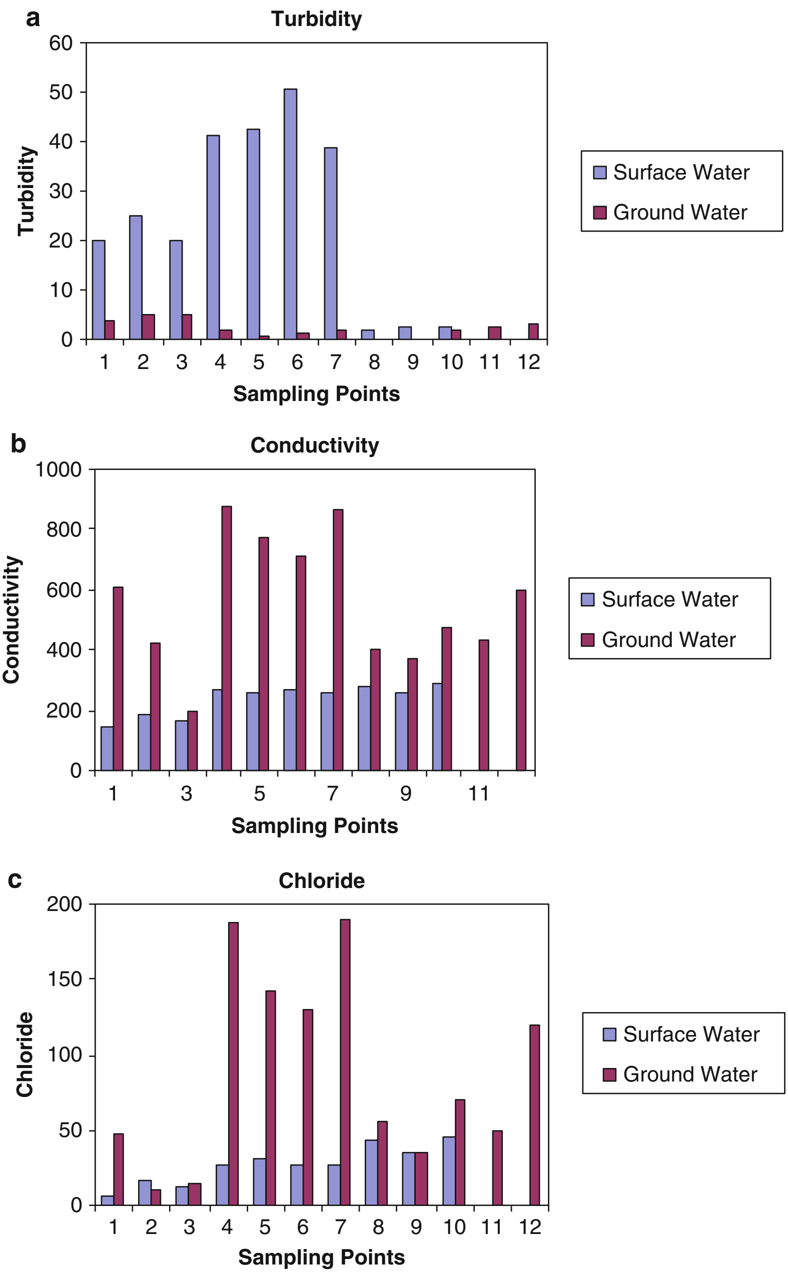
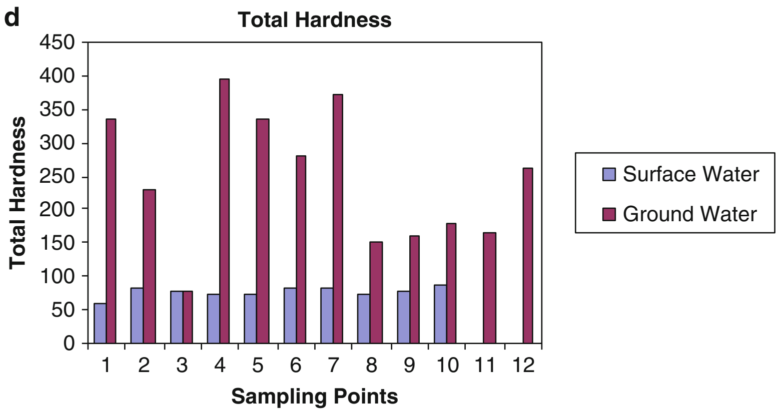


Fig. 14.2 (**a**) Variation in turbidity with respect to sampling points. (**b**) Variation in conductivity with respect to sampling points. (**c**) Variation in chloride with respect to sampling points. (**d**) Variation in total hardness with respect to sampling points. (**e**) Variation in pH with respect to sampling points



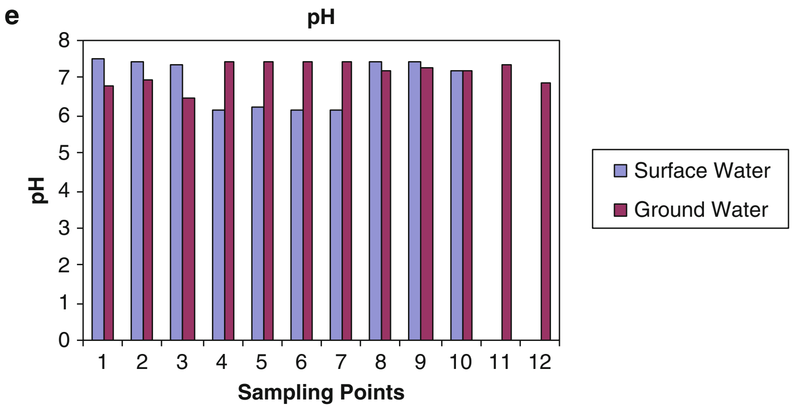


Fig. 14.2 (continued)

The **Threshold Function** takes on a value of 0 if the summed input is less than a certain threshold value (v), and the value 1 if the summed input is greater than or equal to the threshold value.

$$\varphi(v) = \begin{cases} 1 & \text{if } v \geq 0 \\ 0 & \text{if } v < 0 \end{cases}$$

The **Piecewise-Linear function** can also take on the values of 0 or 1, but intermediate values between that depending on the amplification factor in a certain region of linear operation is also possible.

$$\varphi(v)=\begin{cases} 1 & v \geq \frac{1}{2} \\ v & -\frac{1}{2} > v > \frac{1}{2} \\ 0 & v \leq -\frac{1}{2} \end{cases}$$

The **Sigmoid function** can range between 0 and 1, or −1 to 1 range. There are various types of sigmoid function. For example the hyperbolic tangent function can be represented by:

$$\varphi(v)=\tanh\left(\frac{v}{2}\right)=\frac{1-\exp(-v)}{1+\exp(-v)}$$

The Fig. 14.3 describes the basic architecture of an Artificial Neural Networks.

From this model the interval activity of the neuron can be shown to be:

The output of the neuron, y_k, would therefore be the outcome of some activation function on the value of v_k.

$$v_k=\sum_{j=1}^{p} w_{kj}\, x_j \tag{14.1}$$

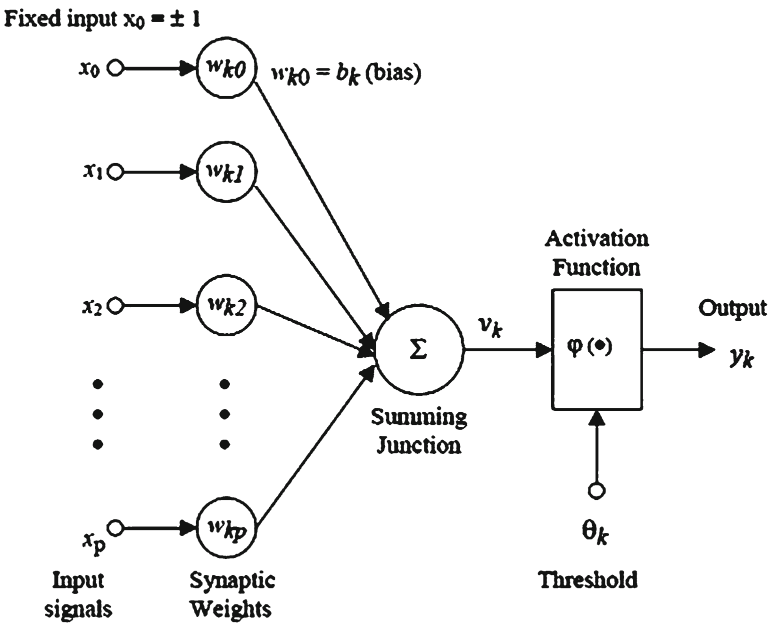


Fig. 14.3

14.2.1.1 Selection of Network Topology

The network topology or number of hidden layers is another parameter which is needed to be determined with the help of trial and error or any other heuristic search algorithms like Genetic Algorithms due to its influence on network accuracy. As many studies (Beniwal et al. 2013; Xin and Zhang 2002; White and Ligomenides 1993; Angeline et al. 1994; Leung et al. 2003; Blanco et al. 2000; Yen and Lu 2000) nowadays are utilizing genetic algorithm for identifying optimal topology of neural networks the present study applied the efficacy of GAs in searching for the optimal solution to estimate the best network topology for the present problem.

After the model topology was selected the network is trained with a set of dataset with known value of the output variables and by applying different training algorithms to identify the optimal weightage (Eq. 14.1) that can result in a neural network model for the present problem which can accurately predict the value of the output variables for unknown situations which are not included in the training dataset.

14.2.1.2 Training Algorithms

In the present study following training algorithms were utilized to find the optimal value of weightage at which the model yields the best results.

Quick Propagation

Quick Propagation algorithm (eg: Constantin et al. 2013) is a batch technique which exploits the advantages of locally adaptive techniques that adjust the magnitude of the steps based on local parameters (for instance, the non-global learning rate). Second, knowledge about the higher-order derivative is used (such as Newton's mathematical methods). In general, this allows a better prediction of the slope of the curve and where the minima lies. (This assumption is satisfactory in most cases.)

Conjugate Gradient Descent

The basic backpropagation algorithm adjusts the weights in the steepest descent direction (negative of the gradient). This is the direction in which the performance function is decreasing most rapidly. It turns out that, although the function decreases most rapidly along the negative of the gradient, this does not necessarily produce the fastest convergence. In the conjugate gradient algorithms a search is performed along conjugate directions, which produces generally faster convergence than steepest descent directions. (e.g., Benli 2013)

Levenberg Merquadart

The Levenberg-Marquardt algorithm was designed to approach second-order training speed without having to compute the Hessian matrix. When the performance function has the form of a sum of squares (as is typical in training feedforward networks), then the Hessian matrix can be approximated as

$$\mathbf{H} = \mathbf{J}^T \mathbf{J}$$

and the gradient can be computed as

$$\mathbf{g} = \mathbf{J}^T \mathbf{e}$$

where **J** is the Jacobian matrix that contains first derivatives of the network errors with respect to the weights and biases, and **e** is a vector of network errors. The Jacobian matrix can be computed through a standard backpropagation technique that is much less complex than computing the Hessian matrix (e.g., Adib and Hadis 2013)

14.2.2 Performance Metrics

Although there are many varieties of metrics to analyze performance of a numerical model Root Mean Square Error; which represent the degree of matching achieved by the model; Correlation Coefficient; represents the magnitude and direction of relationship that exist between the observed and predicted dataset and Covariance; which depicts the deviation of the modeled data from the target dataset; are the three commonly used metrics supported by majority of the scientific community.

14.2.2.1 Root Mean Square Error

The Root Mean Square Error (RMSE) of an estimator $\hat{\theta}$ with respect to the estimated parameter θ is defined as

$$\mathrm{RMSE}(\hat{\theta}) = \sqrt{\mathrm{E}[(\hat{\theta} - \theta)^2]}.$$

The MSE thus assesses the quality of an estimator in terms of its variation and un-biasedness. Note that the MSE is not equivalent to the expected value of the absolute error.

14.2.2.2 Correlation Coefficient

The correlation coefficient (often denoted by Greek rho) is given by the following important equation:

$$Correlation = \rho = \frac{\mathrm{cov}(X,Y)}{\sigma_x \sigma_y}$$

14.2.2.3 Covariance

The covariance between two jointly distributed real-valued random variables x and y with finite second moments is defined as:

$$\sigma(x,y) = \mathrm{E}[(x-\mathrm{E}[x])(y-\mathrm{E}[y])]$$

where E[x] is the expected value of x, also known as the mean of x. By using the linearity property of expectations, this can be simplified to:

$$\sigma(x,y) = \mathrm{E}[xy] - \mathrm{E}[x]\mathrm{E}[y]$$

where E[xy] is the correlation between x and y; if this term is zero then the random variables are orthogonal. Covariance represents the independency of two variables with respect to each other. If Covariance is zero then both the variables is dependent on each other.

14.3 Results and Discussion

Table 14.1 show the neural network topology, GA search settings, MSE, STDDEV, and r achieved by the 15 models that were developed to establish the objective of the present study. For each of the water quality parameters, three algorithms, QP, CGD, and LM, were used. MSE, STDDEV, and r helped to select the best performing algorithm among them. The GA was applied to search for the ideal network architecture with the settings shown in Table 14.1. The search was followed by training of the network with the architecture selected. The three algorithms discussed earlier were now applied to train the models to learn the encoded pattern or relationship between the output and input parameters. As the data set was small (23×11), QP, CGD, and LM algorithms were selected. For training purposes, 75% of the data set was used, 15% was used for validation, and the remaining 15% was applied for testing the model performance.

All 15 models were named by adding the type of training algorithm in the prefix and the parameter considered output in the suffix. Hence, the model trained with the QP algorithm to find the chloride concentration was named QP-Cl.

QP-Turb, QP-pH, CGD-Cl, LM-TH, and LM-Cond were selected as the best performing models respectively for predicting Turb, pH, Cl, TH, and Cond. Among the selected models, the best MSE was obtained by QP-pH followed by CGD-Cl, and the worst MSE was found using LM-Cond followed by LM-TH.

As all the models were trained with the same parameters and the number of iterations for all the models was fixed at an equal value, the accuracy of the models may be influenced by the correlation of the input and output quality variables. Thus it can also be concluded that pH and Cl are largely influenced by the surface water quality. In the case of the other variables, it was found that prediction of conductivity and hardness is relatively complex. Conductivity was affected by the dissolved ions whose concentration depends on the chemical impurities present in the water.

Table 14.1 Attributes and Performance Evaluation of Neuro-genetic Models Considered in the study

Model	Neural network architecture	GA search settings (P-G-Pe-CR-MC)	MSE	STDDEV	r
QP-Cl	11-6-1	40-50-5-0.8-0.2	0.32	6.73	0.97
QP-Turb	**11-16-1**	**40-50-5-0.8-0.2**	**0.37**	**1.33**	**0.99**
QP-TH	11-10-1	40-50-5-0.8-0.2	18.60	18.83	0.92
QP-pH	**11-5-1**	**40-50-5-0.8-0.2**	**0.04**	**0.16**	**0.85**
QP-Cond	11-10-1	40-50-5-0.8-0.2	3.55	2.80	0.99
CGD-Cl	**11-3-2-1**	**40-50-5-0.8-0.2**	**0.05**	**6.38**	**0.98**
CGD -Turb	11-2-1-1	40-50-5-0.8-0.2	3.22	2.47	0.94
CGD -TH	11-4-3-1	40-50-5-0.8-0.2	7.25	10.84	0.95
CGD -pH	11-5-1	40-50-5-0.8-0.2	0.01	0.18	0.68
CGD -Cond	11-1-5-1	40-50-5-0.8-0.2	9.38	13.14	0.94
LM-Cl	11-4-6-1	40-50-5-0.8-0.2	4.87	7.08	0.96
LM -Turb	11-2-8-1	40-50-5-0.8-0.2	0.75	1.74	0.97
LM -TH	**11-1-2-1**	**40-50-5-0.8-0.2**	**1.54**	**9.47**	**0.97**
LM -pH	11-2-8-1	40-50-5-0.8-0.2	0.04	0.30	0.22
LM -Cond	**11-2-3-1**	**40-50-5-0.8-0.2**	**3.09**	**10.61**	**0.99**

The effluent from industrial infrastructure and waste water treatment plants contaminates surface water bodies with such impurities. Again, the presence of certain kinds of geophysical properties (presence of stalagmite or sandstone) can increase the conductivity of the groundwater. The geophysical properties of the study area were favorable for increasing the conductivity, and thus the groundwater displayed higher conductivity than surface water where chemical impurities were present at a lower level. This explains the indifferent relationship between the conductivities of surface and groundwater. The hardness was also higher compared with that of the groundwater due to the same geophysical properties of the catchment. Calcium and magnesium ions were present in much higher concentrations than in the surface water. The presence of hardness in the surface water was also due to the presence of sandstone and stalagmites in the region and because most surface water bodies had been created in depressions like mountain caves and funnels that arose as a result of landslides or earthquakes in those regions millions of years ago. But as the seepage of surface water to the aquifer is minimal due to the presence of impervious layers of sandstone or stalagmites, the hardness of the groundwater was unable to influence that of the surface water. Also, the reaction time of hardness and conductivity is higher than that of Cl or pH.

14.4 Conclusion

This study tries to establish the interactivity of surface water and groundwater with the help of 15 neurogenetic simulation models. The accuracy of the estimation was determined as the criteria of the relationship between surface and groundwater values of the quality parameters. According to the simulation results, pH and Cl models had the smallest MSEs and groundwater values, whereas Cond and TH had the lowest MSEs. It was found from the model that although it would be easier to predict the concentration of chlorine and pH in groundwater due to contamination by surface water, it would be difficult to predict the results in the case of hardness and conductivity which is actually a property of the groundwater itself and is no way related to surface water contamination. That is why there is no significant interrelationship between the surface and groundwater hardness and conductivity. Also, the parameters of surface and groundwater quality may not have a direct relationship to each other in the local areas due to the soil texture, soil characteristics, geology, and geohydrology of the area and the depth of the aquifers. But in larger areas there may be some relationship between different parameters of surface and groundwater quality. In this study, very few sampling points were considered compared to the enormous size of the selected catchment. The same model could be developed with more values from increased sampling points; then the relationship given by the model trained with a larger data set could provide a more practical scenario with greater accuracy. But overall, from the values of the performance metrics it can be concluded that with proper attention to the temporal scale, it would be possible to make an accurate representation of the level of degradation and the threat to human health posed by the contamination of groundwater by surface water.

References

Adib A, Hadis J (2013) Stochastic approach to determination of suspended sediment concentration in tidal rivers by ANN and GA. Canadian J Civil Eng ja

Angeline PJ, Saunders GM, Pollack JB (1994) An evolutionary algorithm that constructs recurrent neural networks. IEEE Trans Neural Netw 5(1):54–65

APHA, AWWA & WPCF (1981) Standard methods for examination of water and wastewater, 15th edn. American Public Health Association, Washington, DC

Beniwal R, Beniwal NS, Gupta HO (2013) Optimization of multicarrier transmission using genetic algorithm. Int J Adv Res Electron Commun Eng 2(1):070

Benli H (2013) Determination of thermal performance calculation of two different types solar air collectors with the use of artificial neural networks. Int J Heat Mass Transfer 60:1–7

Blanco A, Delgado M, Pegalajar MC (2000) A genetic algorithm to obtain the optimal recurrent neural network. Int J Approx Reason 23(1):67–83

Constantin ILIE, Constanta C, Margareta ILIE, Lucia MELNIC (2013) Simulating the number of romanian's pupils and student using artificial intelligence techniques and country's economic estate. revista economică: 102

Coppola E Jr et al (2003) Artificial neural network approach for predicting transient water levels in a multilayered groundwater system under variable state, pumping, and climate conditions. J Hydrol Eng 8(6):348–360

Coulibaly P et al (2001) Artificial neural network modeling of water table depth fluctuations. Water Resour Res 37(4):885–896

CTIC (2008) Groundwater & surface water: Understanding the interaction, know your watershed. Retrieved from http://www.ctic.purdue.edu/media/files/Ground%20Water%20and%20Surface%20Water.pdf. Accessed 23 Jan 2013

El Tabach E (2007) Use of artificial neural network simulation metamodelling to assess groundwater contamination in a road project. Math Comput Model 45(7):766–776
Gallagher DL, Dietrich AM, Reay WG, Hayes MC, Simmons GM Jr (2007) Ground water discharge of agricultural pesticides and nutrients to estuarine surface water. Groundw Monit Remediat. doi:10.1111/j.1745-6592.1996.tb00579.x
Gaur S, Ch S, Graillot D, Nagesh Kumar D (2013) Application of artificial neural networks and particle swarm optimization for the management of groundwater resources. Water Resour Manag 27(3):927–941
IS:2296-1982; 2nd revision, 1st reprint March, 1986: Tolerance limit for inland surface water subject to pollution; Bureau of Indian Standard
IS:10500-1991; 1st revision, 1st reprint October, 1992: Indian standard for drinking water specification; Bureau of Indian Standard
Kuo Y-M, Liu C-W, Lin K-H (2004) Evaluation of the ability of an artificial neural network model to assess the variation of groundwater quality in an area of blackfoot disease in Taiwan. Water Res 38(1):148–158
Leung FHF et al (2003) Tuning of the structure and parameters of a neural network using an improved genetic algorithm. IEEE Trans Neural Netw 14(1):79–88
Majumder M, Roy PK, Mazumdar A (2010) Watershed modeling of river Damodar with the help of artificial neural network and genetic algorithm. Phd thesis, Jadavpur University
Method of water analysis: analysis of water samples were conducted as per IS: 2296, IS:10500 and Standard Methods for Examination of Water and Wastewater (APHA, AWWA & WPCF)
Nayak PC, Satyaji Rao YR, Sudheer KP (2006) Groundwater level forecasting in a shallow aquifer using artificial neural network approach. Water Resour Manag 20(1):77–90
Ranjithan S, Eheart JW, Garrett JH Jr (1993) Neural network-based screening for groundwater reclamation under uncertainty. Water Resour Res 29(3):563–574
Reay WG, Gallagher DL, Simmons GM Jr (1992) Groundwater discharge and its impact on surface water quality in a Chesapeake Bay inlet. Water Resour Bull 28(6):1121–1134
Shiri J, Kisi O, Yoon H, Lee K-K, Nazemi AH (2013) Predicting groundwater level fluctuations with meteorological effect implications-A comparative study among soft computing techniques. Comput Geosci
Van der Kwaak JE (1999) Numerical simulation of flow and chemical transport in integrated surface-subsurface hydrologic systems. Ph.D. dissertation, Department of Earth Sciences, University of Waterloo, Ontario, Canada, 217 p
Verstraeten IM, Carr JD, Steele GV, Thurman EM, Bastian KC, Dormedy DF (1998) Surface water-ground water interaction: herbicide transport into municipal collector wells. J Environ Qual 28(5):1396–1405
White D, Ligomenides P (1993) GANNet: a genetic algorithm for optimizing topology and weights in neural network design. New trends in neural computation. Springer, Berlin/Heidelberg
Xin Z-h, Zhang H-j (2002) Neural Network and Genetic Algorithms for Topology Optimization of the CCS7 Network. Int Trans Operat Res 9:427–436. doi:10.1111/1475-3995.00365
Yen GG, Lu H (2000) Hierarchical genetic algorithm based neural network design. Combinations of evolutionary computation and neural networks, 2000 IEEE symposium on. IEEE

Chapter 15
Rating Irrigation Canals Using Cognitive Indexes

Mrinmoy Majumder

Abstract The main objective of this investigation was to identify an optimal configuration to enable irrigation canals to withstand future uncertainties from climate change and uncontrolled urbanization. To accomplish this objective, a set of factors was selected based on their influence on the stability and functionality of irrigation canals. The selected factors were of two types: conducive, which increase the efficiency of irrigation canals, and deductive, which decrease it. All the variables were rated with respect to their capacity to increase efficiency on a scale of one to nine, where nine is assigned for efficiency-increasing ability and one is assigned to efficiency-decreasing abilities. All possible combinations one the nine-point scale rating of the factors were created to make a combinatorial data matrix that represents every possible situation that might arise in an irrigation canal. The data set was then clusterized with the help of guided neuroclustering methods (GNCM) and an agglomerative decision tree algorithm (DTA). According to the clusterization and comparison of the two methods, the sample with the optimal configuration that both clustering algorithms had selected within their optimal clusters was identified. The selected combination was recommended in the construction of new canals to increase the canals' longevity. According to the clusterization method, flow volume in the canal can be semihigh, but variation in the flow must be very low. Channel loss and demand from farmers must be semi and extremely low, respectively, and there should be as many buffer ponds as possible and the contribution from groundwater must be maximized. The amount of sedimentation must be minimized. That is, an irrigation canal must be developed in such a way that demand from farmers is highly regulated. A large number of buffer ponds must be created in and around the canal. Preventive measures must be strictly imposed to control inflow volume, channel loss, flow turbulence, and sedimentation. Infrastructure to store excess water must be available so that excess water from extreme events can be stored for

M. Majumder (✉)
School of Hydro-Informatics Engineering, National Institute of Technology Agartala, Barjala, Jirania 799055, Tripura, India
e-mail: mmajumder15@gmail.com

M. Majumder and R.N. Barman (eds.), *Application of Nature Based Algorithm in Natural Resource Management*, DOI 10.1007/978-94-007-5152-1_15,

use in times of high demand. Only canals with the above recommended configurations will be able to withstand the vulnerabilities that will arise in the near future from abrupt changes in climate and uncontrolled growth in urban populations.

Keywords Irrigation canals • Decision tree algorithms • Neuro-clustering

15.1 Introduction

The demand for water from the domestic, agricultural and industrial consumers is increasing due to the rapid scale of urbanization observed in most of the major cities of the World. Keeping in view of the impact of climate change due to global warming and uncontrolled growth in population the quality and quantity of the available water resource is shrinking. As a result of population overgrowth, demand for food is rising which in turn has increased the use of water for irrigation. The large scale industrialization to satisfy the demand from the rising population has also increased the demand for water from the industrial sector. That is why rivers all over the world, especially tributaries and dis-tributaries, which are one of the major sources of water, are under threat of extinction from this ever-rising demand. Large scale industrilization and rapid increase in the urban population has also incremented the pollution content of the water bodies.

Indexes are now widely used for qualitative as well as quantitative, but logical, decision making. For example, Sheng et al. (1997) used a geographic information system (GIS) for classification of watersheds in developing countries. Hajeka and Boyd (1994) attempted to design an index to select suitable sites for aquaculture ponds. The methodology was similar to that developed for systems used in the evaluation of soil for irrigation, road construction, waste disposal, and residential development. The potential impact of agriculture drainage on quality of water of receiving streams was evaluated by Brenner and Mondok (1995) with the help of a watershed delivery factor, animal nutrient factor, management factors that were actually indices with fecal coliform and phosphorus as the influential parameters, and a groundwater delivery factor where nitrate concentration of the streams was found to be most the influential determinant. Heathwaite et al. (2000) developed an index to identify the sources and transport pathways that control phosphorus and nitrogen transport. According to the index, P loss was found to be maximum in well-defined areas of watershed, whereas nitrate loss was observed mainly upstream of the watershed. Wanqa et al. (1997) conceptualized the index of biotic integrity (IBI) to determine the relationship between biotic richness and urbanization of a watershed. The index was found to be directly correlated with forest cover but indirectly related to agricultural land, which is again proportional to the amount of urbanization. From the study it was concluded that more than 10–20% urbanization is not good for biotic integrity. Indices also were used in site selection for shrimp farming. Slope, land-use type, soil thickness, elevation, soil type, soil texture, soil pH, distance to sea, distance to roads, local markets, and hatcheries were selected as deciding parameters. Areas that were not allowed to be used for shrimp farming were excluded. A series of GIS models was employed to identify and prioritize the suitable areas for shrimp farming. Only 31% of the land of the study area (in Hiphong) was found to be optimal for shrimp farming. Karthika et al. (2005) used a land-use index to identify suitable sites for brackish water aquaculture in Thane districts of Maharashtra, India. Stagnitti and Austin (1998) developed an independent software tool for the selection of new aquaculture facilities. Principal component clusterization was performed to classify watersheds for environmental flow predictions based on spatial characteristics. The method included 56 parameters to form 5 representative groups of watersheds. The study was conducted by Alcázar and Palau (2010) for the Mediterranean countries. Shaw and Cooper (2008) developed an index that represents the relationship between watershed, stream reaches, and plant type. A study was conducted by Falcone et al. (2010) to compare indices representing the relative severity of human disturbance in watersheds. According to the results, indices composed of many variables performed better than those with single variables. A "threshold method of scoring using six uncorrelated variables: housing unit density, road density, pesticide application, dam storage, land cover along a mainstream buffer, and distance to nearest canal/pipeline" was found to be best representative of anthropogenic disturbances of the watersheds. Jacobs et al. (2010) developed an index of wetland conditions with the help of hydrogeomorphic variables. The index was applied to classify the wetlands of Nanticoke River watersheds. The variables were scored with respect to "range check, responsiveness and metric redundancy." Water quality, evapotranspiration, runoff, species diversity, species health, and stakeholder participation were used by He et al. (2000) to develop an ecological indicator for the assessment of the conditions of altered watersheds. The fuzzy set models were used to prioritize watersheds with respect to the scope of fisheries and abatement of non-point-source pollution by Wenger et al. (1990). Both stream use and stream conditions were utilized to develop the indices for application in the Kewaunee River Basin in Wisconsin, USA. Wang et al. (2010) proposed a framework to evaluate the impacts on watershed ecosystems caused by hydropower development. Watershed ecosystem services were classified into four categories – provisioning, regulating, cultural services, and supporting services – with the help of 21 indicators. Various evaluation techniques (market value method, opportunity cost approach, project restoration method, travel cost method, and contingent valuation method) were used for the calibration of the models. The models helped to identify the key impacts of hydropower development on watershed ecology, like on biodiversity; water quality was found to be negatively impacted by development; the average environmental cost per unit of electricity was derived to be three times the on-grid power tariff, and overall the degradation in watersheds was found to be compensated adequately by extracting maximum utility from the hydropower plant. Zhang and Barten developed an information system for the analysis of the impact of watershed degradation on water yield. The Watershed Forest Management Information System has three submodules. The first module has to do with the prioritization of watersheds based on conservation and restoration requirement. The second module, Forest Road Evaluation System, is concerned with impact analysis of road networks on forest cover. The last module, Harvest Schedule Review System, was developed for the evaluation of multiyear and multiunit forest harvesting, which will assist in reducing the impact of these factors on water yield and associated changes in water quality.

The success of the indices in the identification, delineation, or representation of a decision with respect to related parameters encouraged the authors to create indices for the irrigation canals to rate them according to their ability to suppress ever-increasing demand and number of extreme events.

15.2 Artificial Neural Network

An artificial neural network (ANN) can be defined as a network of complex interactive signal processing networks that mimics the human nervous system while transporting an error signal. The methodology for the development of neural networks for practical problem solving comprises three major steps: building, training, and testing the network. The neural network topology i.e. number of hidden layers necessary to yield accurate results are first identified by the applications of trial and error or any search algorithms. The next step is to train the model for identification of optimal values of weightage with the help of available dataset (referred as training dataset) of the present problem. This goal is achieved by the use of various training algorithms like Quick Propagation, Conjugate Gradient Descent, Batch Back Propagation etc which continuously generates values of weightage and the corresponding predictions are compared with the outputs given in the training data set. In the last phase, the model is tested with a set of dataset which are not used in the training dataset but output of them is known. This step is necessary for verifying the model for its accuracy with datasets that are not utilized in the training data (i.e., dataset with which the model has learned the problem). The model predictions are verified with the help of common performance metrics like Root Mean Square Error, Correlation, Covariance etc. and until and unless the desired accuracy is achieved the model is not allowed for prediction of unknown situations.

Neural networks can be classified based on their topology, error path followed, and activation function used. A network can contain input, output, and hidden layers. Hidden layers are simply a set of pseudo-input layers that creates an additional set of input layers so that the entire error reduction procedure is followed twice before the prediction of the output. The objective problem determines the configuration of the network topology along with the error achieved. Generally, trial and error is followed, but scientists nowadays use specialized search algorithms to find the optimal configuration of neural networks.

The neural network is a topology that the error signals of a problem follow to reduce errors. According to the path followed by the error signal, neural networks can be grouped into feedforward and feedback network systems. The names followed in a feedforward neural network error always propagate to the forward layer of the network, and in the case of feedback the same signal will follow the back path based on a comparative analysis to evaluate the existing accuracy.

In the case of training algorithms, perceptron algorithms (Rosenblatt 1957), gradient descent (Snyman 2005), and batch back-propagation (Bryson and Ho 1969) algorithms are the most popular. During a training session the main objective of the network is to reduce error by applying different algorithms to solve the problem. Many authors have tried to achieve different objectives and solve complex problems through simple or modified versions of neural models. Below is an overview (Table 15.1):

Table 15.1 Relevant studies related to application of neural networks in clusterizing a representative data set

References	Type of neural network applied	Study objective	Remarks
Raju et al. (2006)		Sustainable irrigation planning	
Rodriguez and Martos (2010)		Surface irrigation parameter identification	
Chavez and Kojiri (2007)	Stochastic fuzzy neural network	Maximum water use and improvements to water quality	
Yang et al. (2009)	Generalized regression neural network	Prediction of leaf area index, green leaf chlorophyll density of rice based on reflectance, and its three different transformations: first-derivative reflectance, second-derivative reflectance, and log-transformed reflectance	
Gautam et al. (2004)	Back-propagation algorithm	Effect of bridge construction on spatial variation of groundwater level	
Torfs and Wójcik (2001)	Local probabilistic neural networks	Analyzing effectiveness of a new ANN algorithm that tries to prove that when inputs are lagged, accuracy is maximized in the case of inputs near the actual value. A case study of predicting discharge for a small watershed was used to substantiate the above objective	
Filho and Santos (2006)	Three-layer feedforward ANN trained with linear least-squares simplex training algorithm	Flood wave simulation with rainfall, stage level, or stream flow as input	
Yoon et al. (2011)	Back-propagation algorithm	Predicted groundwater level with past groundwater level, tide level, and precipitation as input	
Wu and Chau (2011)	Modular artificial neural network (MANN) and data preprocessing by singular spectrum analysis (SSA)	This paper aims to eliminate the lag effect of ANN models in rainfall–runoff models	
Mougiakakou et al. (2005)	Classic NN modified using genetic algorithms	This paper classified the land use pictures of 106 different locations, divided into common, normal, and distinctive classes, with respect to their scenic beauty	

15.3 Decision Tree Algorithm (DTA)

In statistics and data mining, decision tree learning is used to categorize data sets according to predetermined attributes. The learning algorithm of a DTA uses a decision tree "as a predictive model which maps observations about an item to conclusions about the item's target values". The DTA is also known as classification trees or regression trees where leaves represent classifications and branches represent conjunctions of features that lead to those classifications.

DTAs are known to be useful for classification purposes. Although they have been used rarely, they are popular for their well-known classification ability. Some examples are listed below:

(a) Release rule configuration of a flood-control reservoir (Wei and Hsu 2009)
(b) Water quality analysis (Litaor et al. 2010)
(c) Reconstruction of missing daily data (Kim and Pachepsky 2010)
(d) Change detection analysis (Pilloni et al. 2010)
(e) Image classification (Yang et al. 2003)
(f) Topological classification (Simon et al. 2007) and in many other decision-making approaches.

15.4 Methodology

The problem of developing an index lies in the selection of its independent parameters upon which a decision about the ratings could be undertaken. Volume of flow, annual variation of flow, channel loss, storage capacity, groundwater contribution, demand from farms, presence of buffer ponds, and sedimentation were considered as independent parameters of incident that are also merely correlated with each other. The justification for the selection of these parameters is given below:

1. *Volume of flow (Q)*: one of the most important parameters of irrigation canal design is volume of flow. The Dimensional properties of the hydraulic structure is estimated based on the volume of flow.
2. *Annual variation of flow (Q_v)*: the distribution pattern of the annual variation in monthly water discharge represents the maximum and minimum amount of flow in the canal. Design of canals for highly varied flow is prone to errors and often is the main cause of overflow and submergence of agricultural lands at a buffer. But for a steady flow the canal faces less uncertainty, making the task for designers easy.
3. *Channel loss (L)*: the loss of water from a canal can be determined as the amount of water that has either infiltrated or evaporated during the time of flow through the canal. This parameter also represents the canal efficiency when loss and water withdrawn are deducted from the volume of inflow and divided by only inflow. The more water is lost, the less able it will be to withstand uncertainty and stress from extreme events.

4. *Storage capacity (S)*: the capacity of a canal to store water. The higher the storage, the lower will be the chance of overflow. The hydraulic structure of the canal will determine its storage capacity. The storage capacity of a canal can be calculated with the help of the hydraulic radius and depth along with the length of the river canal.
5. *Groundwater contribution (G)*: Irrigation canals are generally fed by groundwater during the summer season, and groundwater is recharged by the canal during the monsoon season. This effluence and affluence relationship between the canal and groundwater helps to maintain the water level so that adequate water can be available for harvesting.
6. *Demand from farms (D)*: The demand of water from adjacent agricultural land depends upon the type and frequency of crop harvested, where hydrophilic crops create more demand and hydrophobic crops have lesser requirements for water. But as most cash crops are of the hydrophilic type, demand for water is generally high from the buffering irrigation fields, which again increases the stress on the canal and, often due to the overuse of the canal, creates water scarcity. Again multicrop practices will entail higher demand than monocrop agro lands. The method of water withdrawal from agro-fields will also influence the demand for water from the canal. A sprinkler-irrigated field will demand more water but lose less water than drip irrigation, which will have lower demand but greater loss of water. The type of irrigation practice will also depend on the type of crops being harvested. Thus, such parameters were not considered explicitly. The method of withdrawal of water will also influence the amount of water required. If pumps are used, then large amounts of water will be withdrawn per cycle, but if withdrawal is done by collectors or tube wells, then the amount of water withdrawn per cycle will be less than pump-controlled irrigation.
7. *Presence of buffer ponds (B)*: buffer ponds are generally made to store water in times of scarcity. During summer or in the absence of rain for a long duration such ponds are used by farmers. The presence of buffer ponds greatly reduces the stress on canals. Thus, the more ponds there are, the higher will be the canals' ability to sustain uncertain periods.
8. *Sedimentation (V_s)*: the amount of sedimentation will impact the amount of water that can be carried or stored within the canal. A heavily silted canal will not be able to withstand high volumes of water during positive extreme events like floods.

The classification work was performed by a neural network due to its pattern identification ability as observed and discussed in many scientific studies. The factors were fed to the model as input parameters. The networks' categorization ability was utilized to estimate the cluster of the data sets. At first a training data set was fed to the model so that the characteristics of individual clusters could be identified. Based on the characteristics of clusters the new data set representing situations of uncertainty was fed to the model. This new data set was clusterized, and, because the characteristics of the clusters were already known, the decision-making process about the ability of the canal to sustain uncertainty could easily be estimated.

The same data set was again classified with respect to the DTA. The threshold values were determined with the help of the Euclidean distance from the mean.

Before the data were fed for clustering, the entire data set was rated on a scale of 1–9, where the higher limit was given to the conducive data set and the lower limit was fixed for the deductive data set. That means the rating was inversely proportional to the suitability of canals to withstand abnormal climate conditions.

15.5 Results and Discussion

The clustering of the available sample was performed with the help of neural networks with 11 inputs, 1 output, a log-sigmoid activation function, 1,000 iterations, and 0.3 initial weight, and the clustering radius was kept at 30%. There are a total of 1,138 data samples. The data set for training consisted of 9 rows of data, which clearly portrays the characteristics desired in the nine clusters. Cluster 1 was found to include the most suitable combinations of input factors where the canals could withstand prevailing climate uncertainties. Table 15.2 presents the maximum and minimum values of the factors for each cluster of the training data set. Clearly, Cluster 1 will contain the combination of factors by which a canal will be able to withstand the uncertainties of climate change and urbanization.

Figure 15.1 depicts the weight profile or distribution of clusters within the sample data set, where of 1,138 samples 10.82% were found to be most suitable and 11.43% least suitable to withstand abnormalities in climate and human population changes.

Analysis of the cluster characteristics revealed that Clusters 1 and 9 did not represent the most and least suitable irrigation canals. For example, sample locations that had high sedimentation, low concentration of buffer ponds, low storage capacity, high variation of flow, and high channel loss were also grouped into Cluster 1. But samples with a high storage capacity and moderate groundwater contribution and frequency of buffer ponds were grouped into Cluster 9.

That is why the neuroclustering method was found to be error-prone and did not represent the true situation of the samples.

In addition, a guided neuroclustering method (GNCM) was followed where a summation of all the factor ratings of the conducive factors were divided by the summation of all the deductive factors and the normalized results of the same were added as an input variable and the samples were re-clustered:

$$\text{Objective Function}\left(\text{Obj}\right)=\left(\text{QXSXGXB}\right)/\left(\text{Q}_{\text{v}}\text{XLXDXV}_{\text{s}}\right) \tag{15.1}$$

For each of the samples, *Eq. 15.1 was calculated and normalized according to Eq. 15.2:*

$$\begin{aligned}\text{Obj Normalized} = &\left(\text{Obj of the present sample} - \text{maximum of all Obj}\right)/\\ &\left(\text{maximum of all Obj} - \text{minimum of all Obj}\right)\end{aligned} \tag{15.2}$$

Figure 15.2 depicts the output of the guided clustering procedure.

Table 15.2 Clusterization of training data

Volume of flow	Annual variation of flow	Channel loss	Storage capacity	Groundwater contribution	Demand from agricultural fields	Presence of buffer ponds	Sedimentation	Clusters
9	1	1	9	9	1	9	1	Cluster 1 minimum
8	2	2	8	8	2	8	2	Cluster 2 minimum
7	3	3	7	7	3	7	3	Cluster 3 minimum
6	4	4	6	6	4	6	4	Cluster 4 minimum
5	5	5	5	5	5	5	5	Cluster 5 minimum
4	6	6	4	4	6	4	6	Cluster 6 minimum
3	7	7	3	3	7	3	7	Cluster 7 minimum
2	8	8	2	2	8	2	8	Cluster 8 minimum
1	9	9	1	1	9	1	9	Cluster 9 minimum
9	1	1	9	9	1	9	1	Cluster 1 maximum
8	2	2	8	8	2	8	2	Cluster 2 maximum
7	3	3	7	7	3	7	3	Cluster 3 maximum
6	4	4	6	6	4	6	4	Cluster 4 maximum
5	5	5	5	5	5	5	5	Cluster 5 maximum
4	6	6	4	4	6	4	6	Cluster 6 maximum
3	7	7	3	3	7	3	7	Cluster 7 maximum
2	8	8	2	2	8	2	8	Cluster 8 maximum
1	9	9	1	1	9	1	9	Cluster 9 maximum

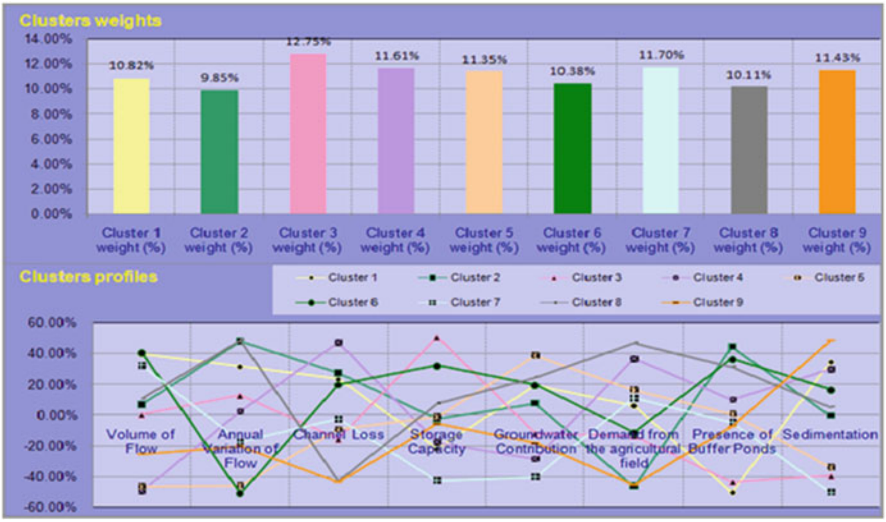


Fig. 15.1 Cluster weights and profiles for available sample data set according to neuroclustering methods

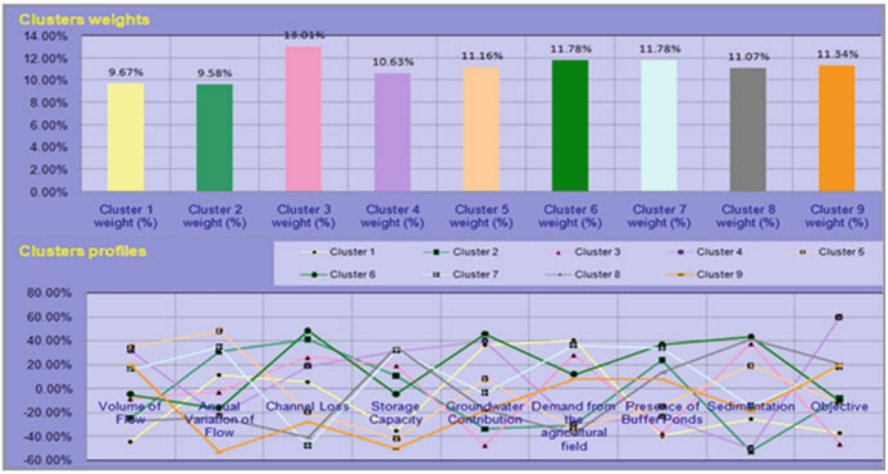


Fig. 15.2 Cluster weights and profiles for available sample data set according to guided neuro-clustering method (GNCM)

In the case of the GNCM method, Cluster 4 was found to represent the most suitable canals with a high volume of flow, low annual flow variation, high channel loss, very high storage capacity, moderate concentration of buffer ponds, very low sedimentation, and low demand from farmers. Although the channel loss was higher in the cluster, it was mitigated by the high storage capacity and low demand from farmers.

From the cluster weight it was found that only 10.63% of the sample population had the characteristic of Cluster 4 and was suitable for withstanding uncertainties from climate change and rapid urbanization. Cluster 3 was found to have the least suitable canals, but within the sample nearly 13.01% of the total population had the characteristics of the Cluster 3 canals.

The sample population was prepared considering every possible combination. In the case of climate change, any canal would face high volume and variation in flow along with high loss of water and deposition of sedimentation. The demand from farmers would be greater due to excess demand from the dependent population. The clusterization procedure revealed that only 10.63% of the total samples that include the impact of climate change and urbanization could mitigate the abnormalities. The remaining combinations would be unable to withstand the uncertainties imposed on them. That is why if the characteristics of Cluster 4 canals were followed in developing new canals, then those canals would not be vulnerable to climatic abnormalities as well as urbanization impacts.

After the application of GNCM the DTA was applied to the same sample population with the help of the following rule:

If the objective function is greater than 85% and less than 100%,
Then assign the canal to Cluster 1
Else
If the objective function is greater than 75% and less than 84%,
Then assign the canal to Cluster 2
Else
If the objective function is greater than 65% and less than 74%,
Then assign the canal to Cluster 3
Else
If the objective function is greater than 55% and less than 64%,
Then assign the canal to Cluster 4
Else
If the objective function is greater than 45% and less than 54%,,
Then assign the canal to Cluster 5
Else
If the objective function is greater than 35% and less than 44%,
Then assign the canal to Cluster 6
Else
If the objective function is greater than 25% and less than 34%,,
Then assign the canal to Cluster 7
Else
If the objective function is greater than 15% and less than 24%,,
Then assign the canal to Cluster 8
Else
Assign the canal to Cluster 9

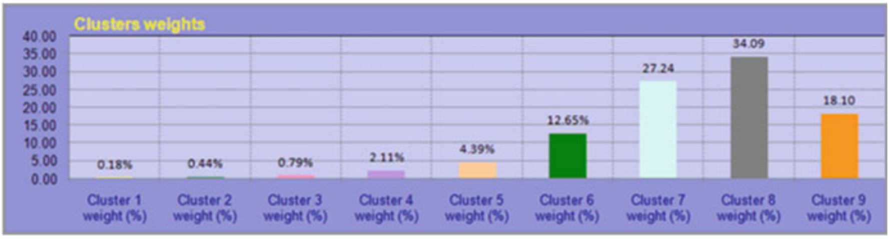


Fig. 15.3 Cluster weights for available sample data set according to guided DTA method

Table 15.3 Optimal configuration of irrigation canals for withstanding future uncertainties (*blue*: conducive variables; *black*: deductive variables)

Volume of flow	Annual variation of flow	Channel loss	Storage capacity	Groundwater contribution	Demand from the agricultural field	Presence of buffer ponds	Sedimentation
7/9	2/9	3/9	8/9	6/9	1/9	9/9	3/9

After the DTA was applied to the sample data set, it was found that Cluster 1 of the DTA was similar to Cluster 9 and 4 of GNCM and Cluster 9 of the DTA was comparable to all clusters of GNCM except Cluster 4.

That is why Cluster 4 is the only cluster that represents channel configurations that can withstand future uncertainties. In the case of the DTA it was found that only 0.17% of the total sample was predicted to have an optimal Cluster 1 whereas 18.10% of the population was found to have rejected Cluster 9 (Fig. 15.3).

As both the DTA and GNCM selected the L191 sample within the optimal cluster [DTA assigned Cluster 1 to the sample whereas GNCM assigned the same sample to Cluster 4; the cluster profile (Fig. 15.2) also supported the selection] the characteristics of L191 were selected as the optimal configuration for a canal to withstand the onslaught of climate and anthropogenic events (Table 15.3).

15.6 Conclusion

The present investigation tried to classify different characteristics of irrigation canals to identify features of a canal that would be able to withstand climatic and urbanization impacts. DTA and GNCM neural networks were applied to clusterize every possible combination of the inputs if all of them were rated on a scale of one to nine according to their impact on canal suitability. The scale was configured in such a way that the suitability of canals was enhanced when the conducive variables increased toward nine and deductive ones decreased towards 1. After the GNCM was applied, it was found that the characteristics of Cluster 4 canals could be identified as a standard for future canal design. But application of DTA revealed that Cluster 1 had the characteristics that engineers should follow to make canals feasible for uncertain situations. It was also found that canals having Cluster 1 in DTA had acquired all the Cluster 4 and 9 characteristics in GNCA but canals having Cluster 9 in DTA had achieved all the Clusters except 4 in GNCA. The sample L191 was found to have configurations that GNCM had assigned to Cluster 4 and DTA had grouped it under Cluster 1. That is why it was concluded that because both GNCM and DTA indicate those canals with the most suitable characteristics to counter future uncertainties, that configuration of the canal may be recommended when new canals are developed in the future. Thus, the ideal irrigation canal in the future must have the characteristics presented in Table 15.3 to make prepare them for the looming changes in climate and human populations.

References

Brenner FJ, Mondok JJ (1995) Nonpoint source pollution potential in an agricultural watershed in northwestern Pennsylvania. Water Resour Bull 31:1101–1112

Bryson AE, Ho YC (1969) Applied optimal control: optimization, estimation, and control. Blaisdell Publishing Company/Xerox College Publishing, Waltham, p 481

Chaves P, Kojiri T (2007) Deriving reservoir operational strategies considering water quantity and quality objectives by stochastic fuzzy neural networks. J Adv Water Res 30(5):1329–1341

Falcone JA, Carlisle DM, Weber LC (2010) Quantifying human disturbance in watersheds: variable selection and performance of a GIS-based disturbance index for predicting the biological condition of perennial streams. Ecol Indic 10(2):264–273

Filho AJP, dos Santos CC (2006) Modeling a densely urbanized watershed with an artificial neural network, weather radar and telemetric data. J Hydrol 317(1–2):31–48

Gautam MR, Watanabe K, Ohno H (2004) Effect of bridge construction on floodplain hydrology—assessment by using monitored data and artificial neural network models. J Hydrol 292(1–4):182–197

Hajek BF, Boyd CE (1994) Rating soil and water information for aquaculture. Aquacult Eng 13:115–128

Heathwaite L, Sharpley A, Gburek W (2000) A conceptual approach for integrating phosphorus and nitrogen management at watershed scales. J Environ Qual 29(1):158–166

He C, Malcolm SB, Dahlberg KA, Fu B (2000) A conceptual framework for integrating hydrological and biological indicators into watershed management. Landsc Urban Plan 49(1):25–34

Jacobs AD, Kentula ME, Herlihy AT (2010) Developing an index of wetland condition from ecological data: an example using HGM functional variables from the Nanticoke watershed, USA. Ecol Indic 10(3):703–712

Karthika M, Surib J, Neelam S, Biradar RS (2005) Brackish water aquaculture site selection in Palghar Taluk, Thane district of Maharashtra, India, using the techniques of remote sensing and geographical information system. Aquacult Eng 32:285–302

Kim JW, Pachepsky YA (2010) Reconstructing missing daily precipitation data using regression trees and artificial neural networks for SWAT streamflow simulation. J Hydrol 394(3–4):305–314

Litaor MI, Brielmann H, Reichmann O, Shenker M (2010) Hydrochemical analysis of groundwater using a tree-based model. J Hydrol 387(3–4):273–282

Mougiakakou SG, Tsouchlaraki AL, Cassios C, Konstantina S, George N, Matsopoulos K, Uzunoglu NK (2005) SCAPEVIEWER: preliminary results of a landscape perception classification system based on neural network technology. Ecol Eng 24(1–2):5–15

Pilloni M, Melis, MT, Marini, A (2010) Analysis and validation of a methodology to evaluate land cover change in the Mediterranean basin using multitemporal modis data. Present Environ Sustain Dev 4:19

Raju KS, Kumar DN, Duckstein L (2006) Artificial neural networks and multicriterion analysis for sustainable irrigation planning. J Comput Oper Res 33(4):1138–1153

Rodríguez JA, Martos JC (2010) SIPAR_ID: freeware for surface irrigation parameter identification. Environ Model Softw 25(11):1487–1488

Rosenblatt F (1957) The perceptron – a perceiving and recognizing automaton. Report 85-460-1, Cornell Aeronautical Laboratory

Simon C, Meessen J, Tzovaras D, De Vleeschouwer C (2007) Using decision trees for knowledge-assisted topologically structured data analysis. In: WIAMIS'07. Eighth international workshop on "Image analysis for multimedia interactive services", June 2007, p 2. doi:10.1109/WIAMIS.2007.84

Sheng TC, Barrett RE, Mitchell TR (1997) Using geographic information systems for watershed classification and rating in developing countries. J Soil Water Conserv 52(2):84–89

Shaw JR, Cooper DJ (2008) Linkages among watersheds, stream reaches, and riparian vegetation in dryland ephemeral stream networks. J Hydrol 350(1):68–82

Snyman JA (2005) Practical mathematical optimization: an introduction to basic optimization theory and classical and new gradient-based algorithms. Springer, New York. ISBN 0-387-24348-8

Stagnitti F, Austin C (1998) DESTA: a software tool for selecting sites for new aquaculture facilities. Aquacult Eng 18(2):79–93

Torfs P, Wójcik R (2001) Local probabilistic neural networks in hydrology. Phys Chem Earth Part B Hydrol Oceans Atmos 26(1):9–14

Wang G, Fang Q, Zhang L, Chen W, Chen Z, Hong H (2010b) Valuing the effects of hydropower development on watershed ecosystem services: case studies in the Jiulong River Watershed, Fujian Province, China. Estuar Coast Shelf Sci 86(3):363–368

Wang L, Lyons J, Kanehl P, Gatti R (1997) Influences of watershed land use on habitat quality and biotic integrity in Wisconsin streams. Fisheries 22(6):6–12

Wang F, Yin H, Li S (2010a) China's renewable energy policy: commitments and challenges. Energy Policy 38(4):1872–1878

Wenger RB, Rong Y, Harris HJ (1990) A framework for incorporating stream use in the determination of priority watersheds. J Environ Manage 31(4):335–350

Wei CC, Hsu NS (2009) Optimal tree-based release rules for real-time flood control operations on a multipurpose multireservoir system. J Hydrol 365(3–4):213–224

Wu CL, Chau KW (2011) Rainfall–runoff modeling using artificial neural network coupled with singular spectrum analysis. J Hydrol 399(3–4):394–409

Yang CC, Prasher SO, Enright P, Madramootoo C, Burgess M, Goel PK, Callum I (2003) Application of decision tree technology for image classification using remote sensing data. Agric Syst 76(3):1101–1117. Retrieved from http://linkinghub.elsevier.com/retrieve/pii/S0308521X02000513

Yang XH, Wang FM, Huang JF, Wang JW, Wang RC, Shen ZQ, Wang XZ (2009) Comparison between radial basis function neural network and regression model for estimation of rice biophysical parameters using remote sensing. Pedosphere 19(2):176–188

Yoon H, Jun SC, Hyun Y, Bae GO, Lee KK (2011) A comparative study of artificial neural networks and support vector machines for predicting groundwater levels in a coastal aquifer. J Hydrol 396(1–2):128–138

Zhang Y, Barten PK (2009) Watershed forest management information system (WFMIS). Environ Model Software 24(4):569–575

Chapter 16
CLIMAGE: A New Software for the Prediction of Short-Term Weather with the Help of Satellite Data and Neuro-Fuzzy Clustering

Mrinmoy Majumder and Tilottama Chackraborty

Abstract Weather is an essential part of decision making in large-scale manufacturing, agroforestry, electrical, and various other rainfall-dependent industries. The generation and distribution of electricity, water supply, and related essential amenities of any region also depend on daily weather variations. That is why short-range weather forecasting has a direct impact on the day-to-day operations of these industries. But due to the amount of manpower, equipment, and computational capacity involved in the prediction of short-term weather, many scientists are now opting to utilize remote sensing and image processing for the estimation of next-day or 3-day weather. In short-range weather forecasting, due to the complex and nonlinear interrelationship between rainfall and weather parameters, quantification of the forecast can be error-prone. Probabilistic forecasting has proven to be more accurate. For such a level of nonlinearity as one finds in short-range forecasting, normal regression methods fail to achieve the level of accuracy required. That is why nature-based algorithms, which are well known for their ability to map the nonlinearity between dependent and independent variables, are more suitable than statistical algorithms. The present investigation tries to estimate the probability of occurrence of rainfall 1 day in advance based on the processing of satellite images and neuro-fuzzy clustering algorithms. A software module is also developed so that the model can be utilized to achieve given objectives.

Keywords Short-range weather forecasting • Neuro-fuzzy clustering • Image processing

M. Majumder (✉) • T. Chackraborty
School of Hydro-Informatics Engineering, National Institute of Technology Agartala, Barjala, Jirania 799055, Tripura, India
e-mail: mmajumder15@gmail.com; tilottama86@gmail.com

M. Majumder and R.N. Barman (eds.), *Application of Nature Based Algorithm in Natural Resource Management*, DOI 10.1007/978-94-007-5152-1_16,

16.1 Introduction

There are four types of weather forecast: short-range, extended-range, medium-range, and long-range. Short-range weather forecasts are forecasts of weather 48–72 h in advance. Weather forecasting involving 3–5, 5–7, and more than 10 days in advance is referred to as extended-range, mid-range, and long-range weather forecasting, respectively. The range for short-range weather forecasting (SRWF) is mainly concerned with weather systems observed in the latest weather charts and by considering the generation of new systems within a certain time period.

16.1.1 Importance of Short-Term Weather Forecasting

In recent days, because of abnormalities in climate patterns perceptible in many places of the world, SRWF has become very complex but very important. What follows are some of the reasons why SRWF is essential and its accuracy of utmost importance.

- Very SRWF and now casting is included in day-to-day decision making in the aviation and shipping industries.
- The impact of small-scale, short-lived, extreme phenomena like thunderstorms, tornadoes, flash floods, dense fog, and freezing rain on the socioeconomic lifestyle of vulnerable populations and agricultural productivity is high. That is why predictions of such short-duration events is important for the prevention of large-scale damage.
- Tropical storms or hurricanes are phenomena of special importance in short-range forecasting.
- SRWF is complex, and the amount of spatial and temporal detail required for forecasting decreases as the period increases.

Predictions from SRWF are used for the optimization of different managerial decisions. Weather can hamper the normal efficiency of any industry that is directly or indirectly connected to weather. In the distribution of electricity an extreme event lasting for a few minutes can wreck the entire supply framework. The demand for electricity also varies with weather patterns (Bolzern and Fronza 1986). A rise in temperature, along with humidity, often decreases the comfortability index (CI) of a region. The reduction in CI will also increase the demand for artificial wind sources like fans or air-conditioning systems, which in turn will increase the demand for electricity. But a nor'wester or westerlies, which are common after a few days of hot and humid weather, can reduce the temperature for a few days and increase the CI, which will again decrease the demand for electricity. But if a power plant is conditioned for excess output due to rising demand from the previous day, then the excess energy will be wasted because of this sudden reduction of demand.

The harvesting pattern for cash crops also requires knowledge of weather patterns and often use SRWF to decide about harvesting time for which maximum production and profit can be realized (Wilks et al. 1993).

That is why managers in related industries decide on their daily productivity after analyzing the weather forecast of a given day.

In this study a software was developed for the prediction of short-range weather (24 h) with the help of fuzzy logic and a neurogenetic algorithm. The input data were retrieved from cloud composition as observed from the satellite images of the target region. This method of weather forecasting can eliminate the need for expensive instruments and supercomputers or skilled workers for collecting relevant data from the target region.

Simple processing of satellite images showing the shape, pattern, and concentration of clouds of 16 square-shaped equal-area grids (eight grids of the first level and eight of the second) having a length or width of 105 km.

Predicting the probability of rainfall 24 h in advance requires satellite images showing the shape, size, thickness, texture, and type of cloud. The information thus collected must be entered into the software as input where the weight of the input and that of each grid adjacent to the target grid are estimated with the help of fuzzy logic in accordance with their influence on the occurrence of rainfall. Once the input data are submitted, the model will automatically estimate the probability of rainfall in the next 24 h.

The advantage of fuzzy logic (e.g., Niros and Tsekouras 2012; Maharaj and D'Urso 2011) and neural network models (e.g., Mitrofanov 2006; Kim et al. 2001) have been discussed in various scientific studies, and the relevancy of the fuzzy logic in application to the present problem domain is also available through various scientific articles published in related journals (e.g., Awan and Awais 2011; Cobaner 2011; Moustris et al. 2011; Chang et al. 2005; Cao and Chen 1983).

16.1.2 Satellite Imagery

Satellite imagery consists of photographs of Earth or other planets made by means of artificial satellites. Satellites equipped with high-resolution sensors revolve around Earth. During rotation the sensors try to capture images from the Earth's surface. These images are then sent to a receiving station on Earth for processing. Once the images are received by the Earth stations, they are converted into a computer readable format. The processed images are then used for the retrieval of relevant and extractable information. Various types of satellites are deployed to perform multiple jobs. The main apparatus of the satellites are the sensors, which are normally high-resolution cameras, or signal generators, which are sent to observatories on the Earth's surface for further processing.

16.1.2.1 Importance of Satellite Imagery

Images captured by satellites are used in meteorology, agriculture, geology, forestry, landscape, biodiversity conservation, regional planning, education, intelligence, and warfare. Satellite imagery is also used in seismology and oceanography in deducing changes to land formation, water depth and sea beds, by color caused by earthquakes, volcanoes, and tsunamis; the potential of prescription and precision farming can also be monitored from the rich source of data retrievable from satellite images.

The images captured by satellites are further processed by transferring them from some accessible source. For example, one of the earliest two-dimensional Fourier transforms applied to digital image processing of NASA photos as well as national security applications by ESL Incorporated (Atlas and Hoffman 2000).

Due to the real-time observation and direct capture of images, satellite images can also produce real-time weather forecasting with greater accuracy and reliability. The images show the concentration, speed, and direction of storms, which enable weather forecasters to estimate more precise information for the areas in the path of a storm. By merely observing a satellite image, the forecaster can see a storm escalating or breaking apart, changing directions, or speeding up or slowing down.

Satellite images have made weather forecasting much more factual and functional than it was even just a few years ago because of the details and quantity of information obtainable. "Forecasters are now able to give the public a much better image of what to anticipate in the next week, day or hour because of information provided by these satellite images (Anton 2010)."

16.1.3 Clustering

Clustering is an unsupervised learning process to classify a collection of unlabeled data into different groups whose members are similar in some way to other members of the group but unlike members of other groups. A loose definition of clustering could be the process of organizing objects into groups whose members are similar in some way.

Thus cluster analysis can be defined as a collection of statistical methods to identify groups of samples that show similar characteristics.

16.1.3.1 Importance of Clustering

Cluster analysis is used to perform the following functions:

- To describe and to make spatial and temporal comparisons of communities (assemblages) of organisms in heterogeneous environments;
- To generate artificial phylogenies or clusters of organisms (individuals) at the species, genus, or higher level that share a number of attributes;
- To build groups of genes with related expression patterns (also known as coexpressed genes);
- To automatically assign genotypes;
- To infer population structures;
- To classify multivariate data from surveys and test panels in market research;
- To create a more relevant set of search results compared with normal search engines like Google; there currently exist a number of Web-based clustering tools such as Clusty;

- To divide a digital image into distinct regions;
- To locate and characterize extrema in a target distribution;
- To classify chemical properties in different sample locations.

The use of clustering is widespread in different decision-making problems where clusters are developed to identify problematic regions and accordingly solutions are searched to solve the problems. K-means, K-medoid, or hierarchical or principal component analysis are some of the methods for clustering data into groups of similar objects. Neuro-fuzzy clustering is one of the widely used hybrid clusterization methods where the classification simplicity of fuzzy logic and mapping capacity of neural networks are utilized to cluster a sample population based on the importance of input variables.

16.1.4 Neuro-Fuzzy Clustering

When a sample population obtained from experimental analysis or mathematical models is grouped according to expert-specified thresholds, uncertainty may creep in and can be removed only if the relevant information to identify the system is present and clustering is done by the application of fuzzy logic.

For example, a major problem in bioinformatics analysis or medical science is obtaining the correct diagnosis from certain important information. Nowadays computers are used to gather, store, analyze, and integrate patterns and biological data , which can then be applied to find a new, useful diagnosis or information with the help of neuro-fuzzy clustering techniques.

16.2 Description of the Software: CLIMAGE

CLIMAGE, or Climate from Image, is a software that allows users to estimate the probability of rainfall on the next day based on cloud information retrieved from the target grid and nine adjacent grids. Each of the variables is encoded with a weight representing its importance for the objective. All the weights are deduced based on their importance as decided by the implementation of the fuzzy logic.

The determination of weights was performed with the help of fuzzy logic. All the input variables were arranged in rows and columns. Then each input was compared with the other inputs with respect to its importance in estimation of the output. The rule utilized to rate the input variable with respect to every other variable is given below (Rule 1):

If Input N is much more important than Input N-1, Then
give Input N a rating of 1.
If Input N is more important than Input N-1, Then
give *Input N* a rating of *2.*
If Input N is as important as Input N-1, Then

Table 16.1 Rating received by input variables with respect to other variables

	Cloud area	Cloud type	Cloud thickness	Cloud shape	Cloud texture	Rainfall
Cloud area	0	5	5	4	4	5
Cloud type	1	0	3	2	3	3
Cloud thickness	1	3	0	2	3	4
Cloud shape	2	3	4	0	3	4
Clud texture	2	3	3	3	0	4
Rainfall	1	3	2	2	2	0

give Input N a rating of 3.
If Input N is less important than Input N-1, Then
give Input N a rating of 4.
If Input N is much less important than Input N-1, Then
give Input N a rating of *5.*

Applying the same rule to rate each of the input variables with respect to every other input variable will produce a data matrix as shown in Table 16.1.

After rating each of the input variables, each row of the matrix is divided by the maximum rating achieved by the row. The dividend will be minimum for those variables that has the highest rating (Table 16.2).

That is why the complement of the dividend was taken as the weight of the variable. Similarly, the locations of adjacent grids were also rated, and their weight was also derived using the same method. Once the weight was decided, the same weight as for the variable and the grid was multiplied by the magnitude of the input variable for an image. The weighted averages of all the input variables were taken as the objective function. The value of the variables is also rated with the help of their influence on the occurrence of rainfall.

All the values of the input and output variables are then encoded into nine groups, each representing the intensity of the variables. The encoded category and the ratings given to them are presented in Table 16.3.

Similar to the Input variables the output variable or objective function is also grouped into nine categories based on magnitude. The category of the objective function represents the occurrence of rainfall in the target grid.

After determination of the weight for the input variables and the weight for the grids a combinatorial data matrix considering every possible combination among the input variables is prepared and the corresponding objective function is determined and encoded into relevant groups. This categorized data matrix is fed to the feedforward, fully connected neural network for training the network according to the conjugate gradient descent training algorithm, logistic activation function, and a topology as determined by a heuristic search method: 7-13-1.

The rule for encoding the objective function is prescribed as follows (Rule 2):

$$\text{OBJECTIVE FUNCTION}(Q) = \begin{pmatrix} \mathrm{w}_n \times \mathrm{A} + \mathrm{w}_n \times \mathrm{T}_c + \mathrm{w}_n \times \mathrm{d}_c + \mathrm{w}_n \times \mathrm{S}_c \\ + \mathrm{w}_n \times \mathrm{T}\times_c + \mathrm{w}_n \times \mathrm{P}\,/ \end{pmatrix}$$
$$\sum \mathrm{w}_\mathrm{n} \times \left(\mathrm{A} + \mathrm{T}_c + \mathrm{d}_c + \mathrm{S}_c + \mathrm{Tx}_c + \mathrm{P}\right) \tag{16.1}$$

Table 16.2 Rating received by grids with respect to other grids

		1st level								2nd level							
		North	South	East	West	Northeast	Northwest	Southeast	Southwest	North	South	East	West	Northeast	Northwest	Southeast	Southwest
1st level	North		2	2	3	1	4	1	1	2	1	1	2	1	3	1	1
	South	4		3	4	3	5	3	4	3	3	3	4	2	4	2	3
	East	4	3		4	3	5	3	4	3	3	2	3	2	4	2	3
	West	3	2	2		2	3	2	2	1	1	1	2	1	2	1	1
	Northeast	4	3	3	4		5	3	4	4	3	3	4	2	3	2	3
	Northwest	2	1	1	3	1		1	2	2	1	1	2	1	3	1	2
	Southeast	5	3	3	4	3	5		3	4	3	3	4	3	4	3	3
	Southwest	5	2	2	4	2	4	3		4	3	3	4	3	4	3	3
2nd Level	North	4	3	3	5	2	4	2	2		2	2	3	2	3	2	2
	South	5	3	3	5	3	5	3	3	4		2	4	3	4	3	3
	East	5	3	4	5	3	5	3	3	4	4		4	3	4	3	3
	West	4	2	3	4	2	4	2	2	3	2	2		2	3	2	2
	Northeast	5	4	4	5	4	5	3	3	4	3	3	4		4	3	3
	Northwest	3	2	2	4	3	3	2	2	3	2	2	3	2		2	2
	Southeast	5	4	4	5	4	5	3	3	4	3	3	4	3	4		3
	Southwest	5	3	3	5	3	4	3	3	4	3	3	4	3	4	3	

Table 16.3 Groups utilized for encoding input variables

Variable name	Groups encoded	Rating
Cloud area (A)	EH, VH, SH, H, M, L, SL, VL, EL	9–1
Cloud type (T_c)	Cumulonimbus	11
	Nimbostratus	10
	Cirrocumulus	9
	Altocumulus	8
	Stratocumulus	7
	Fog	6
	Cumulus	5
	Cirrus	4
	Cirrostratus	3
	Altostratus	2
	Stratus	1
Cloud thickness (d_c)	EH, VH, SH, H, M, L, SL, VL, EL	9–1
Cloud shape (S_c)	ELa, VLa, SLa, La, M, S, SS, VS, ES	9–1
Cloud texture (Tx_c)	Stormy	18
	Mixed with rain	17
	Light rain	16
	Overcast	15
	Jumpy	14
	Dark rain	13
	Incoming storm	12
	Hazy	11
	Stippled	10
	Fuzzy	9
	Fluffy	8
	Crepuscular	7
	Wispy	6
	Distant	5
	Diffuse	4
	Lumpy	3
	Line	2
	Vapor	1
Rainfall (P)	EH, VH, SH, H, M, L, SL, VL, EL	1–9
Objective function (*Q*)	EH, VH, SH, H, M, L, SL, VL, EL	9–1

If ($Q < 10$*,*
Then "EL",
Else If ($Q < 25$*,*
Then "VL"
Else If ($Q < 35$*,*
Then "SL",
Else If ($Q < 45$*,*
Then "L",

Else If (Ϙ < 55,
Then "M",
Else If (Ϙ < 65,
Then "H",
Else If (Ϙ < 75,
Then "SH",
Else If (Ϙ < 90,
Then "VH",
Or Else "EH"

The training and testing correct classification rate was 99.99%, but the testing CCR was 77.04%. The reduction in testing accuracy can be attributed to the small amount of data available for testing purposes.

The network thus trained is utilized for prediction of rainfall in the target grid.

That is why if the necessary information of the input variables is given for some day, the next day occurrence probability can be predicted from this model.

After the probability of occurrence of rainfall is predicted for the target grid with respect to the other 16 grids surrounding the target grid the category of the objective variable as predicted by the software is rated using the same method as with the ratings shown in Table 16.3.

A weighted average is calculated based on the weight of the grids and the respective ratings of the objective function. The value of this weighted average is again converted into a category of the objective function to find the actual occurrence of rainfall in the target grid with respect to the surrounding 16 grids.

16.3 Benefits of CLIMAGE

The neural network model thus trained and tested was embedded into the software framework and can be distributed to different users for testing purposes. All the user needs to collect is the satellite images of nine grids representing the cloud variation. The necessary information about the cloud and the rainfall magnitude can be collected from the meteorologic stations nearest to the grid.

The relevant data when entered into the software will predict the occurrence probability of rainfall in the center grid. The prediction can be verified by waiting till the next day or by comparing with an already available set of data.

The benefits of the model are manifold. The model can be utilized by practitioners and experts in many fields like farmers, power generation companies, energy efficiency managers, business owners, disaster management groups, town planners, construction engineers, material managers, factory supervisors, etc.

The farmers can schedule their harvesting work according to the outcome from the model. Based on the requirements of the crop and the model prediction farmers can plan their harvesting schedule.

As rainfall directly impacts the demand for electricity by decreasing CI operation managers of generating units can adjust their production according to the outcome predicted by the software. The adjustment can save a lot of energy.

Energy efficiency managers can create an optimal plan for the usage of energy in running High Energy Consuming Appliances. The controlling of such high-energy devices can conserve a noticeable amount of energy. The conservation of energy will also increase income from selling carbon credits.

Rainfall or abnormal weather patterns can also impact the level of demand for a specific product. For example, when the temperature is high, sales of ice-cream are also high. But after a sudden shower, sales of ice-cream declines. Thus, a business owner whose profit mainly depends on the efficiency of his or her inventory will modify the management plan to counter a sudden reduction in demand.

The occurrence of extreme events implies that disaster management groups in the region will be under immense pressure to mitigate the sudden rise in the level of uncertainty. That is why if those groups can know in advance about an upcoming surge in rainfall, then they can prepare for the impact 1 day earlier.

Town planners can also benefit from the software. If they analyze the number of extreme events based on different variations of the input variables that are common in the targeted region, they can propose a plan that will already be optimized to prevent the more dire consequences of extreme events. The specification and maintenance required will also be preplanned. Thus a significant amount of money can be saved by saving energy and additional expenses.

Construction engineers can adjust their construction schedule according to the outcome of the model. They can plan in advance by utilizing the software for different combinations of those input variables that are most common in and around the construction site.

The quality of materials utilized at construction sites or manufacturing plants varies with the occurrence of monsoons. A sudden decrease in water content in the air can alter the quality of the materials themselves. Wide but frequent variation in temperature can either wear or tear materials affected by thermal conduction. That is why material managers can preplan their use of specific materials on a specific day when the chance of extreme events will be at a minimum.

Manufacturing and production units are controlled by weather patterns. Water can contaminate and reduce the quality of cement. If seepage of water remains undetected, then an entire inventory can be wasted. That is why managers in many companies are extremely concerned about the occurrence of rainfall on a given day. Accordingly, they prepare their production schedule and manage output. Thus, CLIMAGE can help them in planning at the postproduction stage for the level and time of output they desire on any given day.

Above all, laypeople can also avoid the hassles they may face due to a sudden influx of rainfall. The software can predict the weather 24 h in advance, making it possible to plan for the next day's activities.

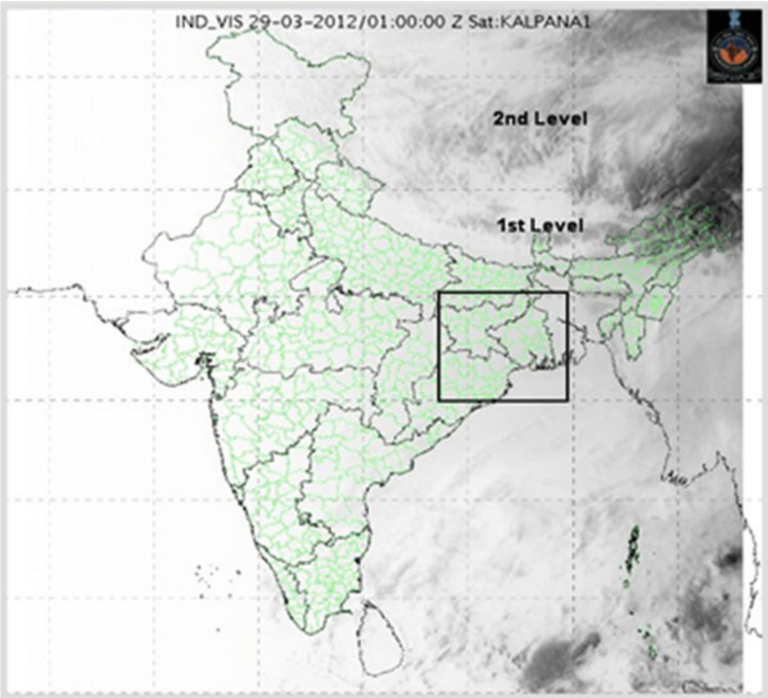


Fig. 16.1 Satellite picture of India as captured by INSAT IIIA

16.4 Case Study

Satellite imagery of India was captured by INSAT satellite Kalpana, where the presence of a cloud with its shape, size, and texture was clearly visible. An image from eastern India was captured and the already gridded image was processed to estimate the type, thickness, area, shape, and texture of the cloud. The data were fed to the software along with the data on present-day rainfall of that region. The software successfully predict rainfall on the following day at the same time of day.

The present study aims to estimate the probability of rainfall in the entire state of Orissa and southern part of West Bengal. The region is highlighted by the white box in Fig. 16.1.

According to the picture, the northeastern part of the target grid in the second level has a sufficient amount of rain clouds engulfing most of the regions of those grids.

But in other areas of the targeted grids no such clouds are observed. Accordingly, the categories of the input variables are configured and shown in Table 16.4.

Table 16.4 Categories assigned to input variables with respect to target grid depicted in Fig. 16.1

Grid	Cloud area	Cloud type	Cloud thickness	Cloud shape	Cloud texture	Rainfall	OBJ
First level							
Southwest	EL	St	EL	ES	Distant	EL	EL
Southeast	EL	St	EL	ES	Overcast	EL	VL
West	EL	St	EL	ES	Distant	EL	EL
South	EL	St	VL	VS	Line	EL	SH
Northwest	EL	St	VL	VS	Fluffy	EL	M
Northeast	VH	Ns	VH	L	Dark rain	EL	VL
North	M	Ns	M	ES	Light rain	EL	M
East	VL	Sc	M	ES	Light rain	EL	VH
Second level							
Southwest	EL	St	EL	ES	Diffuse	EL	EL
Southeast	M	Ns	M	M	Incoming storm	EL	EL
West	EL	St	EL	EL	Wispy	EL	EL
South	VL	St	L	SS	Light rain	EL	EL
Northwest	VL	St	L	SS	Distant	EL	EL
Northeast	EH	Cs	EH	EL	Stormy	EL	VH
North	VH	Ns	H	VL	Dark rain	EL	EH
East	H	Ns	SH	L	Mixed with rain	EL	VH
Next day (24 h in advance) probability of rainfall (according to Rule 2)							EL

After the value of the objective function is calculated with respect to the predicted category of the grids, the value of the function is found to be equal to 5.97% and is thus encoded into the EL group. Thus, the probability of rainfall in the next 24 h for the target grid is extremely low. The weather report from the next day shows that no rainfall had occurred in the 24 h period from the time this picture was taken.

16.5 Conclusion

This research tried to predict the occurrence of rainfall within a grid of interest by the application of neuro-fuzzy clustering and image processing. The study developed a software that can predict the occurrence of rainfall with respect to the previous day's rainfall and cloud type, shape, thickness, and texture. The data of these input variables must be entered for the adjacent 16 grids. The weight according to the importance as determined by fuzzy logic for the grids as well as for the input variables is predetermined. The prediction work is performed by the neurogenetic model developed with the help of a combinatorial data matrix considering every possible situation that might occur among the selected input variables. The study will help a range of people from fields such as meteorology and agriculture as it will allow them to prepare for the onset of weather extremes based on the outcome of the software. Preplanned mitigation measures can save both money and manpower and increase the efficiency of any system. The drawback of the software lies in its inability to predict the quantity of rainfall. In the future, features such as prediction of rainfall 5 days in advance will be introduced, but such a prediction would also require more complexity in the model. Training of neural networks requires a sufficient amount of data, but because the model is fed with a combinatorial data matrix, no other data set can be included in the training data set of the model.

References

Anton (2010) The advantages of satellite images, Retrieved from http://www.stella2000.com/the-advantages-of-satelleite-images/ on 25th Feb 2013

Atlas R, Hoffman RN (2000) Chapter 4: The use of satellite surface wind data to improve weather analysis and forecasting at the NASA Data Assimilation Office. Elsevier Oceanogr Ser 63:57–78

Awan MSK, Awais MM (2011) Predicting weather events using fuzzy rule based system. Appl Soft Comput 11(1):56–63

Bolzern P, Fronza G (1986) Role of weather inputs in short-term forecasting of electric load. Int J Elec Power Energy Syst 8(1):42–46

Cao H, Chen G (1983) Some applications of fuzzy sets to meteorological forecasting. Fuzzy Set Syst 9(1–3):1–12

Chang CL, Lo SL, Yu SL (2005) Applying fuzzy theory and genetic algorithm to interpolate precipitation. J Hydrol 314(1–4):92–104

Cobaner M (2011) Evapotranspiration estimation by two different neuro-fuzzy inference systems. J Hydrol 398(3–4):292–302

Kim JH, Hwang SH, Shin HSA (2001) Neuro-genetic approach for daily water demand forecasting. KSCE J Civ Eng 5(3):281–288

Maharaj EA, D'Urso P (2011) Fuzzy clustering of time series in the frequency domain. Inform Sci 181(7):1187–1211

Mitrofanov DG (2006) Neural network identification of aerial objects from radar images. Meas Tech 49(2):178–186

Moustris KP, Larissi IK, Nastos PT, Paliatsos AG (2011) Precipitation forecast using artificial neural networks in specific regions of Greece. Water Res Manag 25(8):1979–1993

Niros AD, Tsekouras GE (2012) A novel training algorithm for RBF neural network using a hybrid fuzzy clustering approach. Fuzzy Set Syst 193:62–84

Wilks DS, Pitt RE, Fick GW (1993) Modeling optimal alfalfa harvest scheduling using short-range weather forecasts. Agr Syst 42(3):277–305

Chapter 17
OPTIDAL: A New Software for Simulation of Climatic Impacts on Tidal Power

Mrinmoy Majumder, Soumya Ghosh, and Rabindra Nath Barman

Abstract Uncontrolled growth in population has increased the demand for energy, which is reflected in the ever-growing cost of crude oil and electricity. The rise in prices has increased the average cost of basic amenities around the world. That is why alternative, renewable sources of energy are now being explored so that the latter can share the responsibility of mitigating the power demand and reduce the stress on finite energy sources. But problems with renewable energy sources include the frequency of availability and absence of inexpensive methods/devices to store the energy. Available renewable energy sources like solar, wind, and hydro are available in infinite quantities, but to produce energy from these sources requires their availability for a specific amount of time. For example, solar energy can be converted into usable forms only during a specific period of time. Thus, to recover the cost of conversion, the amount of energy that can be produced from renewable sources in a day must either be stored or utilized optimally to mitigate energy requirements. But the availability of inexpensive methods/devices is scarce, and thus solar energy is still a luxury, not a necessity. The second option then becomes more attainable. The time and quantity of hydro energy turns out to be more predictable than other forms of renewable energy sources. The inexpensive conversion mechanism also increases the attractiveness of hydro energy as a possible energy source. But as is common, hydro

M. Majumder (✉)
School of Hydro-Informatics Engineering, National Institute of Technology Agartala, Barjala, Jirania 799055, Tripura, India
e-mail: mmajumder15@gmail.com

R.N. Barman
Department of Production Engineering, National Institute of Technology Agartala, Barjala, Jirania 799055, Tripura, India
e-mail: rahul.barman@yahoo.co.in

S. Ghosh
School of Hydro-Informatics Engineering, National Institute of Technology Agartala, Barjala, Jirania 799055, Tripura, India
e-mail: soumyaee@gmail.com

M. Majumder and R.N. Barman (eds.), *Application of Nature Based Algorithm in Natural Resource Management*, DOI 10.1007/978-94-007-5152-1_17,

power is unreliable with respect to duration and intensity of energy delivered compared to fossil fuels, and thus in the case of utilization, proper levels of intensity must be ensured to maximize energy use. Of the two forms of hydro energy, tidal energy is relatively new but its ease of use and availability favor it over hydro energy. The only requirement to convert such energy into usable form so that it can be used to share some of the responsibility for meeting energy demand lies in maximizing the usability of the available resources. The present study introduces a software that aims to optimize the use of tidal energy resources by maximizing profit. The software is tentatively called OPTIDAL (OPtimization of TIDAL sources).

Keywords Tidal power • Optimization • Decision support systems

17.1 Introduction

The increase in population and the demand for energy have increased the stress on fossil fuels throughout the world. It was found that "energy consumption in developing countries is only one-tenth of that in the developed countries" (Economic Watch 2010). But for developing countries to sustain the growth of their economies, consumption of energy must increase. Oil, coal, natural gas, hydro energy, nuclear energy, renewable combustible wastes, and other energy sources mainly satisfy the needs of energy in the world. It was found that, in 1999, the total supply of primary energy in the world was 9,744.48 million tons of oil equivalent (MTOE), but it was predicted that the total supply of energy in the world in 2010 will be 11,500 MTOE; in 2020 it is expected to be 13,700 MTOE (Tables 17.1 and 17.2).

According to the consumption of fuel, oil is the most important and abundant source of energy in the world, but the price of crude oil is volatile and is a function of market demand. Whereas developed industrialized countries consume, on average,

Table 17.1 Percentage contribution of different sources of energy to total supply of global energy

Fuel	Percentage of total supply
Oil	35.1
Coal	23.5
Natural gas	20.7
Renewable combustible waste	11.1
Nuclear	6.8
Hydro	2.3
Other sources	0.5

Table 17.2 Proportion at which energy from different sources was consumed

Source	Percentage of consumption
Oil	42.7
Coal	8.2
Electricity	15.4
Natural gas	16
Renewable combustible wastes	14.2
Other	3.5

about 43 million barrels daily, developing countries consume only about 22 million barrels per day.

The second most abundant finite source of energy is coal, which is widely used for power generation, although its consumption is not as high as that of oil. According to recent trends, natural gas has the highest rate of consumption growth.

In the case of renewable sources of energy, a 3.7% growth was projected over the 10-year period from 2000 to 2010. Among all other alternative energy sources, utilization of hydro energy was found to be maximum around the globe (Economy Watch 2010). Besides hydro energy, the utilization of tidal waves for generating electricity is now considered to be one of the most promising technologies to mitigate the present energy crisis if used optimally.

17.1.1 Importance of Tidal Energy

Two types of energy can be produced by the oceans: thermal energy from the Sun's heat and mechanical energy from the tides and waves.

Flood tide and ebb tide are daily phenomena that can be observed in coastal bays and river estuaries. Tidal waves are due to the combination of forces exerted by the gravitational pull of the Sun and Moon and the rotation of the Earth. Local effects like shelving, funneling, reflection, and resonance can also increase the intensity of tidal waves. The magnitude of tidal power is entirely site specific and requires mean tidal differences of more than 4 m with favorable topographical conditions, such as estuaries or certain types of bays, to reduce the installation cost.

A major limitation of tidal power stations is that they can produce electricity during the tidal phase only. Besides this major drawback, the following limitations also are observed in the case of tidal power plants:

1. Development of a barrage between estuaries or coastal bays is a costly installation that impacts the environmental stability of the region. According to global estimates, the cost for the generation of tidal power varies between 13 and 15 cents/kWh.
2. It is often observed that installation of a barrage for the generation of tidal power can cause a reduction in flushing, winter icing, and erosion, which can change the vegetation of neighboring areas and disrupt the ecological balance.
3. The minimum eligibility requirement for installation of tidal power plants and production of electricity (at about 85% efficiency) is the availability of a mean amplitude difference of 7 m, and it is desirable that the location have semidiurnal tides. Such eligibility requirements greatly reduces the number of locations for the production of tidal electricity.
4. Barrages across river estuaries can change the flow of water and, consequently, the habitat for birds and other wildlife.
5. Barrages developed in the mouth of a river or coastal bay may affect fish migration and other wildlife. Turbine blades are known to kill fish or aquatic animals. That is why an extra passage or fish ladders are generally provided for the free movement of aquatic fauna.

6. Tidal barrages may induce shoreline flooding and can damage the coastal population if the stored water is not regulated as required.
7. Tidal power is not very useful for people living far from coastlines; the benefits of tidal energy are not enjoyed by the population living inland.

But as tides are totally predictable and their magnitude does not vary, such energy sources are reliable enough to supply electricity at a specific time of a day. The disadvantage of the discontinuous supply can be mitigated by hybridizing the source with a conventional option.

There are many other benefits of tidal power stations like protection against coastline damage, ready-made road bridges, development of recreational zones around stations, and, above all, the reduction in greenhouse gas emissions will automatically force city planners to opt for such a source of energy in meeting daily energy demand.

The largest tidal power station in the world is in the Rance estuary in northern France, near St. Malo. It was built in 1966. About 20% of Britain's coastland can be used for tidal power generation. There are eight main sites around Britain where tidal power stations could usefully be built, including the Severn, Dee, Solway, and Humber estuaries (Clara 2012).

The Gulf of Cambay and the Gulf of Kachchh on the west coast of India where the maximum tidal range is 11 and 8 m, respectively, with an average tidal range of 6.77 and 5.23 m, respectively, has tremendous potential for tidal power production. In eastern India, the Ganges Delta in the Sunderbans also has good locations for small-scale tidal power development. The maximum tidal range in Sunderbans is approximately 5 m, with an average tidal range of 2.97 m (EAI 2011).

"The economic tidal power potential in India is of the order of 8,000–9,000 MW with about 7,000 MW in the Gulf of Cambay, about 1,200 MW in the Gulf of Kachchh, and less than 100 MW in Sundarbans" (MNRE 2012).

17.1.2 Need for Optimization of Tidal Power Production

The magnitude and quality of tidal power depend on the tidal range. The minimum requirement, as discussed in the previous section, is 4 m. Thus a location with such a level head difference is generally selected for the installation of a tidal power plant. Also many other location-specific parameters are important for the success of a tidal power plant. For example, soil strength, flow turbulence, fish navigation, acceptance by the local population, and many other related parameters that vary from one location to another.

Tidal power plants can be optimized or their profitability can be maximized if the production of power can be maximized and the variable cost minimized. (Sullivan and McCombie 1968; Lee and El-Sharkawi 2008; Lee and Dechamps 1978; Swales and Wilson 1968).

Thus the selection of location followed by the optimization of profitability can help city planners decide about the feasibility of installing a tidal power plant. No software, to the authors' best knowledge, is presently available that can assist engineers in the identification of a suitable location and optimization of the production capacity for a tidal power plant.

OPTIDAL aims to identify suitable locations for tidal power plants as well as optimize the production capacity by varying related variables within specific limits. The software may help city planners in the selection of suitable sites for tidal power and also recommend conditions for getting optimal output from the proposed installation.

17.1.3 Software Objective and Scope

The main objective of the software is to identify suitable locations for tidal power development and optimize plant capacity by varying related variables like tidal range, power demand, interconnection points, flow turbulence, and net profit.

17.2 Input, Output, and Working Principle of Software

The input of the software can be divided into five different sections: climate, geophysical, ecological, socioeconomic, and electrical inputs. Figure 17.1 shows the input panel and Figs. 17.2, 17.3, 17.4, 17.5, and 17.6 show the required input variables that the user must enter to estimate the desired outputs. All the input panels of the software can be accessed from the home panel shown in Fig. 17.7. In the home panel, when either of two buttons – LOCATION SELECTOR or LOCATION OPTIMIZER – is clicked, the input panels will be displayed first. In the input panel, all sections of input variables will be displayed, and clicking on the corresponding button will open the data entry window for entering the required data.

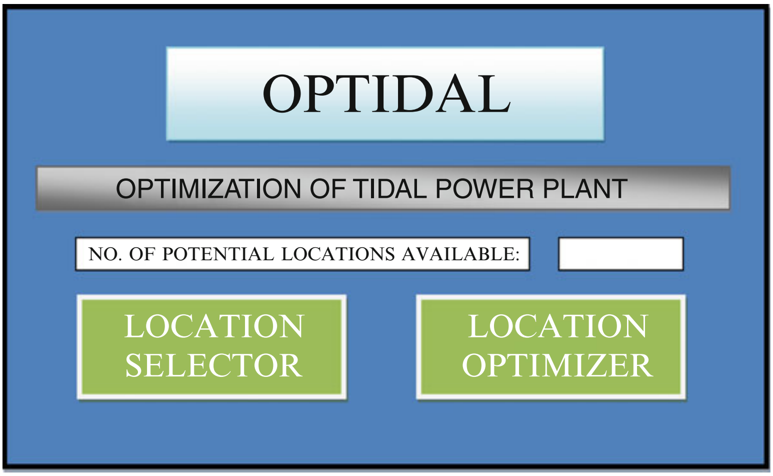


Fig. 17.1 Input panel for location selector and optimizer

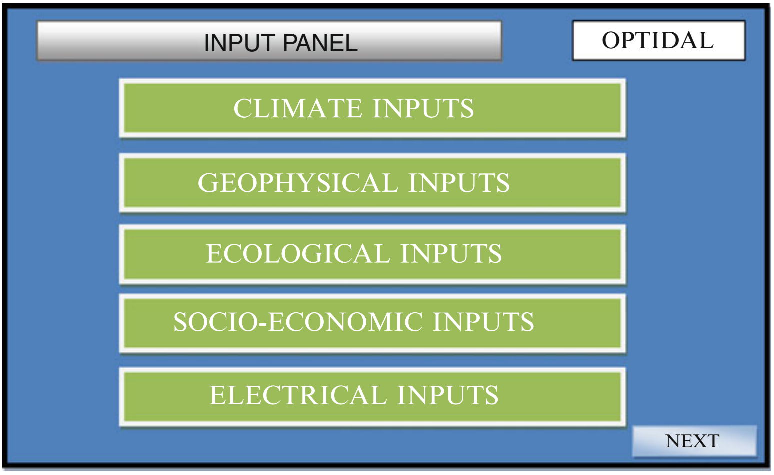


Fig. 17.2 Panel where climatic inputs are entered

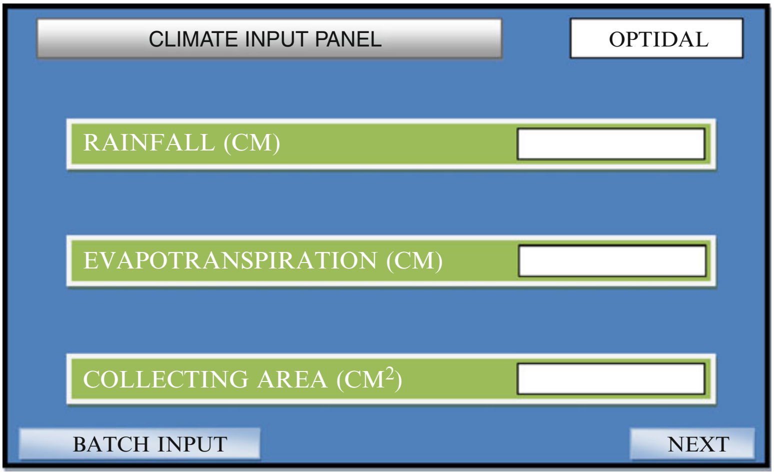


Fig. 17.3 Panel where geophysical inputs are entered

The data can be entered in batch mode (Batch Input button) if more than one options is available and in the single-window model if only one location is available for feasibility analysis.

All the variables are self-explanatory. The watershed loss and channel loss in the geophysical input panel have to be determined in terms of the ratio of the area of the impervious region and the total area of the watershed. Channel loss must be calculated by dividing the outflow volume by the inflow volume. The respective volumes can be measured by measuring the flow in the output and in the input or in the initial region of the channel.

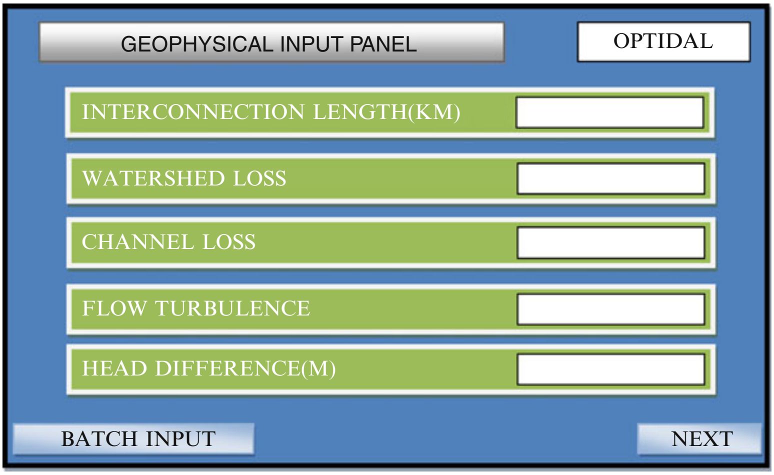


Fig. 17.4 Panel where ecological inputs are entered

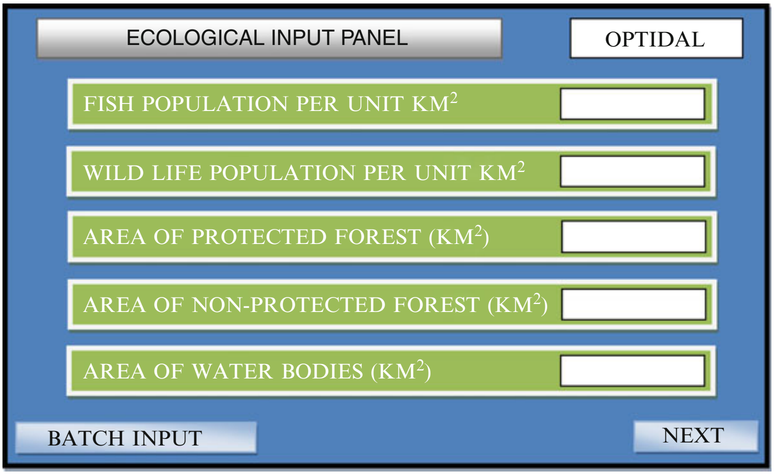


Fig. 17.5 Panel where socioeconomic inputs are entered

Flow turbulence can be determined with the help of the Reynolds number at the mouth of the channel. The value must be determined in numerous locations within the channel output. The average of the numbers must be entered in the input field. The difference between the outside and inside water levels has to be entered in the head-difference field. The difference between the maximum head attained inside the channel during flood tide and the minimum water level that can be observed during ebb tide will determine the maximum tidal range possible in the location, whereas the average of the head difference measured at specific time intervals will estimate the average head difference of the region. In the field of head difference, the user has to enter the latter value so that locations can be compared for a general situation.

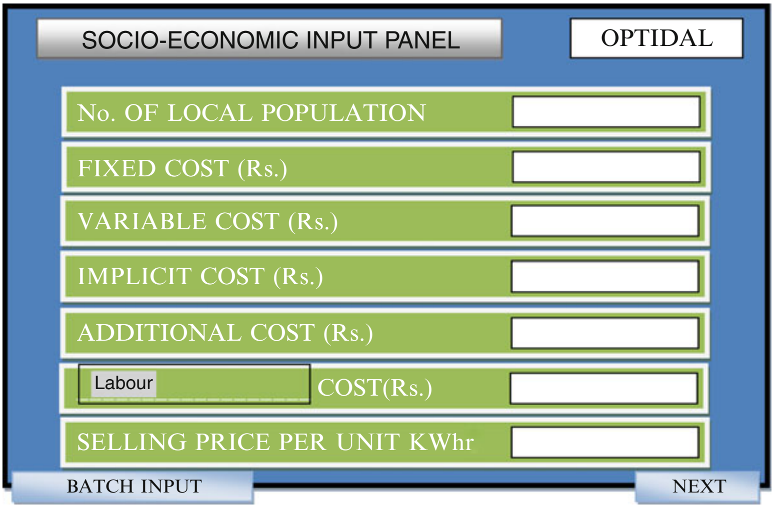


Fig. 17.6 Panel where electrical inputs are entered

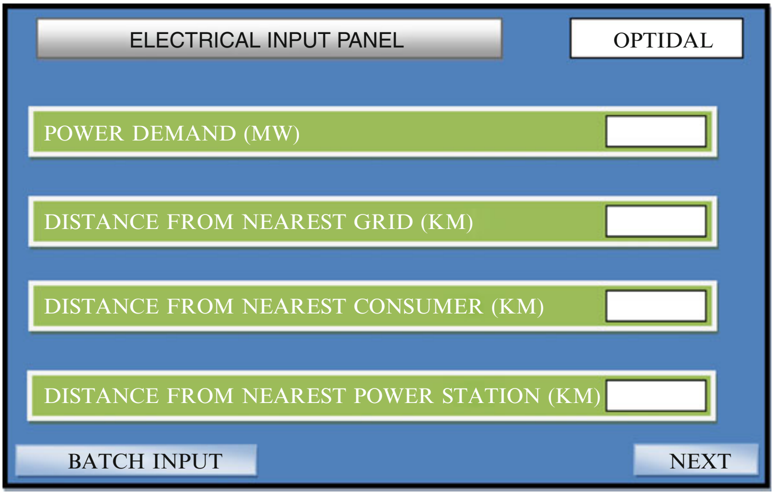


Fig. 17.7 Home panel of OPTIDAL software

The different types of production unit can be classified into four major groups in which all the expenditures incurred by a production unit can be categorized:

1. Explicit costs
   Explicit cost is expenditures that include payments made by the employer or by the owner of the production unit to those factors of production that do not belong to the employer himself. Explicit costs include payments made for raw materials, power, fuel, wages and salaries, land rent payments, and interest on capital. Explicit costs are also referred as accounting costs. These costs are entered in the accountant’s list.

2. Implicit costs
   The implicit or imputed costs arise in the case of those factors that are owned and supplied by the owner him- or herself. The owner of any production unit, including power plants, contributes to increasing the efficiency of those units from the initial phase. The land is generally purchased by the owner unless donated by the government. A certain part of the work load can also be shared by the owner. Thus, the owner receives some remuneration for his or her services in the development of the unit. The cost incurred to remunerate the owner for his or her services is known as implicit cost and is generally not itemized explicitly. All these items would be included under implicit or imputed costs and are payable to the owner. Usually producers ignore these implicit costs when computing total costs. Thus total cost should include both explicit and implicit costs. In this regard, readers may note that in the case of power plants that are included in the essential category of service, the owner may not be an individual or a private trust; it may be the government itself. In that case the remuneration is paid to the government in taxes and dividends if the government becomes a shareholder of the company in which it exercises regulatory power.
3. Fixed costs
   These are costs that are incurred on a unit independently of the volume of production from that unit. Such costs do not change with changes in output. They remain the same regardless of the volume of production. Examples of fixed cost include interest on capital, salaries of permanent staff, insurance premiums, property taxes, and rents. Fixed costs are also known as supplementary costs.
4. Variable costs
   Variable costs are costs that vary with the volume of production. Variable costs are more or less dependent upon increases and decreases in the volume of production. Variable costs are also known as prime costs. Examples of variable costs include costs of labor, raw materials, and chemicals.

Considering the fixed, variable, and implicit costs incurred during installation and in the production phase can estimate the profit of the unit. Or in analyzing the impact of certain types of cost the software has a provision that allows the user to enter a type of cost that may be included in the fixed or variable costs, but if the user wishes to observe the influence of such costs on the profit or optimized variable they can be entered separately. But in that case the user must not include the said cost in the variable, fixed, implicit, and additional costs.

Under additional costs the user may include those costs that cannot be categorized into fixed, variable, or implicit. Explicit costs are not considered as all such costs can be included in either fixed or variable costs.

The cost at which a power plant will sell a unit of electricity produced can be entered in the Selling Cost per unit kWhr field. The selling cost, where the plant attains full capacity, can be estimated so that a comparison can be made that includes the future production capacity of the plant.

In the Electrical Input panel the Distance from the Nearest Power Plant field must contain the distance from the tidal power station to the nearest power station

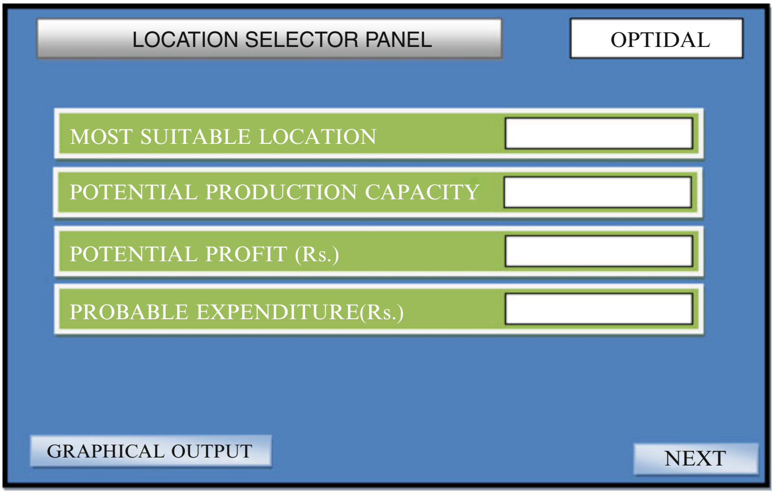


Fig. 17.8 Location selector panel where, based on index, the most suitable location is selected

so that transmission loss for bringing power to the station in the installation phase can be calculated and considered while selecting the most suitable location among the available options.

When the foregoing input data are entered in the software, the data will be processed by OPTIDAL to calculate an index. This index is simply a weighted average of the values of the inputs where the positive variables are placed in the numerator and other variables are naturally placed in the denominator. The weights of the variables are decided with the help of the knowledge gained from different scientific studies and discussions with experts in related fields.

Once the values of the input are entered, the user can either go on to the LOCATION SELECTOR panel (Figs. 17.8 and 17.9) or click to open the LOCATION OPTIMIZER panel (Fig. 17.10). In the Optimization panel, the user can optimize the dependent variables like amount of power produced and net profit. In the first panel to be displayed once the LOCATION OPTIMIZER button is clicked, the name of the variables (included in the drop-down field) and the necessary values of the constraints (lower and upper limit) must be entered. After the values are assigned to the respective fields, the NEXT button must be clicked to see the output (Fig. 17.10) and if the user wants to graphically analyze the impact of the independent variable on the objective, the GRAPHICAL OUTPUT button may be clicked to observe the variation in the LOCATION OPTIMIZER PANEL-GRAPHICAL panel (Fig. 17.11).

Before going to the result panel of the optimizer, the user must enter the serial number of the location for which the user wants to optimize the desired output so that the software can identify the location for which it must optimize the desired variables.

The results of the output from the optimization are displayed in Fig. 17.11, where the numerical optimal values of the independent variables are displayed, and

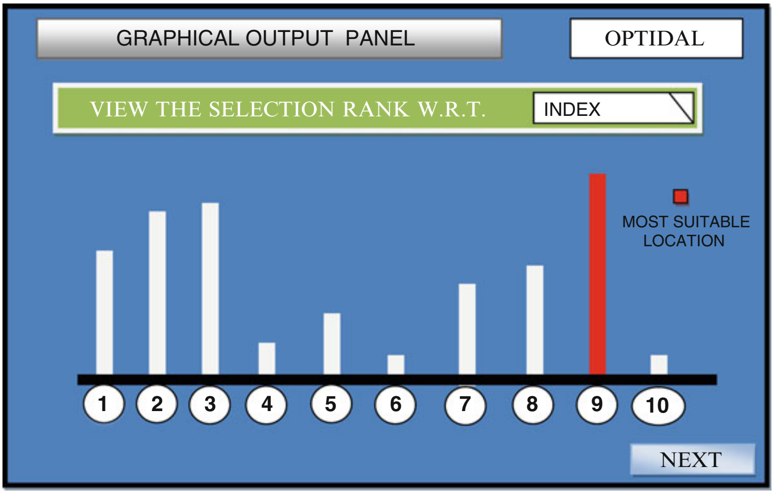


Fig. 17.9 Graphical output panel where graphical comparison of available locations can be made with respect to head difference, flow volume, flow turbulence, net profit, size of local population, wildlife and aquatic faunal population, area of protected and unprotected forest, distance from grid and consumers, and, above all, index value determining location ratings

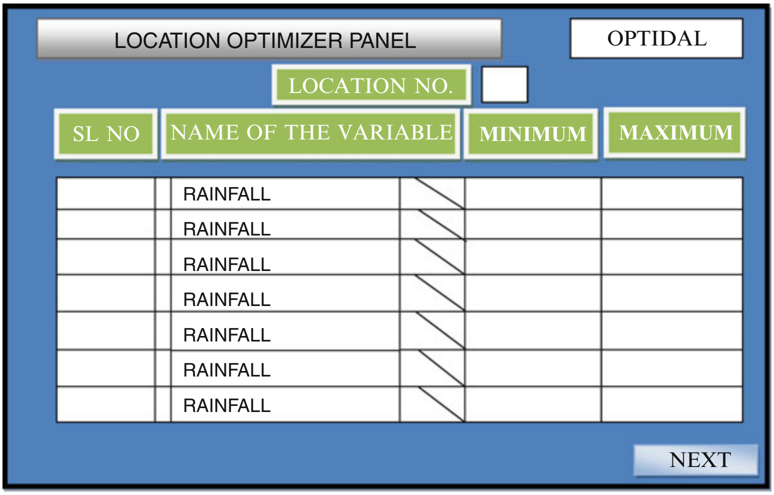


Fig. 17.10 Location optimizer panel where the dependent (optimized) variables and independent variable (variable with which the dependent variable is optimized) can be defined along with their upper and lower limits

Fig. 17.12, where the variation of the numerical optimal values with respect to two independent variables are shown graphically. The user may tweak the maximum and minimum values in the OPTIMIZER panel to identify the impact of constraints on resource availability with respect to the objective variable.

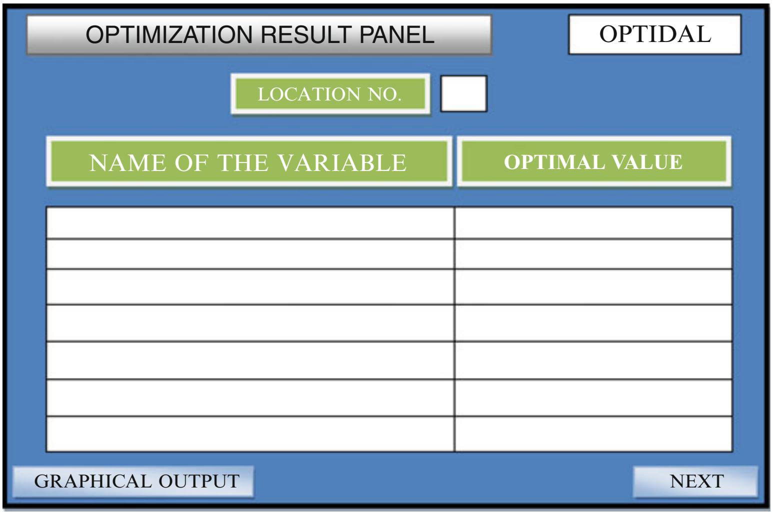


Fig. 17.11 Optimization result panel where optimal values of variables at he point of optimization are displayed

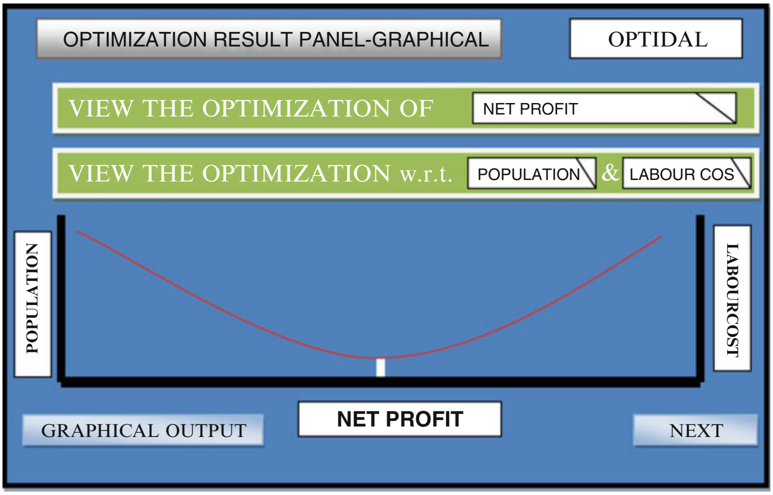


Fig. 17.12 Optimization results panel, where a graphical comparison of dependent and independent variables can be made

17.3 Benefits of OPTIDAL Software

The OPTIDAL software was developed to assist engineers in conducting a comparative feasibility study where different locations within a common geographic region (i.e., different channels within the same coastal zone) can be analyzed for a feasibility study to develop tidal power plants in those locations.

The potential of each of the prospective locations can be analyzed and compared with each other so that the best ones are selected. The values of different necessary inputs can be compared so that a logical and informed decision can be made where the chance of uncertainty will be negligible.

The second benefit of the software is its ability to optimize the two objectives of any tidal power plant. The first one is net profit and the second is the amount of power that can be generated based on the available input variables. The amount of rainfall or area of the protected forest can be varied to display the impact on the objective variables. The GRAPHICAL OUTPUT panel provides a visual display of the influence of two variables on the objective variable.

Based on the output, inferences regarding the impact of different uncertainties (like extreme events, erosion, sedimentation in the channel, urbanization, sudden rise in demand, and cost imbalances in the domestic or international market) can be analyzed and necessary decisions can be taken to mitigate the harmful effects of the abnormalities.

Implementation of the software in practical problem solving may help different people from different fields. For example, business owners will realize immense benefits if they analyze the potential before investing in the development of a tidal power plant.

Government can also use the software to calculate benefits before utilizing coastal areas for the development of tidal-based power. An educated decision in this regard will benefit both the people and the government without risking hostility from the displaced population.

Environmentalists concerned with deforestation activities worldwide will also find benefits in using the software because they will be able to analyze the impact on the environment of the development of a proposed tidal power plant. The impact on aquatic fauna and wildlife can also be compared and analyzed to come to a decision that will benefit everyone.

Energy efficiency managers can also use the software to calculate an overall energy audit of a tidal-based power plant where both expenditures and income in terms of energy can be calculated.

The list of beneficiaries would also include the local people, who can analyze the level of rehabilitation, land loss, and variation in income from pisciculture due to the proposed tidal plant in their watershed.

Because the impact of flow turbulence can also be estimated, the software can be used by designers in selecting suitable turbines for the power plant. The generator and exciters can also be improvised from the output of the OPTIDAL software.

17.4 Drawbacks

The software can be analyzed from the point of view of drawbacks and reliability. The application is still in alpha version, and testing is not yet completed.

Initial observations indicate that having to enter so many data points can become a hindrance for users. In future versions, an automatic data retrieval mechanism may be introduced so that users can automatically determine input values. Another feature that may be introduced is the representation of the spatial variation of the output through Google Earth or Wikimapia. An interface can be made to automatically retrieve data like interconnection length or water level difference from Google Earth images.

A feature for importing data in Microsoft Excel format may also be introduced, although batch data entry is possible in the present version of the product.

17.5 Conclusion

This study introduced OPTIDAL software program, which can be utilized in the selection of suitable regions for tidal power plants so as to maximize output, i.e., maximization of either profit or generation of renewable energy. The location selector capability of the software will help governments invest where needed and avoid wasting money, which will happen if investments are allocated without proper feasibility studies. The Optimization panel of the software will help engineers and designers to vary different related input variables to see their effect on the output. The software thus will help engineers and concerned laypeople to perform risk analysis in the case of newly developed power plants. The reduction of greenhouse gases and the increased income from carbon credits can also be optimized with the help of the software. The effects of deforestation, flow turbulence, and fish navigation can also be visually analyzed with the help of the OPTIDAL software. Besides the drawbacks like lack of automatic retrieval of data from satellite images, the software may be further developed to introduce additional features that will create a state-of-the-art platform for optimization studies in the field of renewable energy.

References

Clara D (2012) Energy resources: tidal power. Retrieved from http://www.darvill.clara.net/altenerg/tidal.htm. Mar 2012

Economy Watch (2010) World consumption of energy, 30th April 2010. Retrieved from http://www.economywatch.com/energy-economy/scenario.html. Feb 2012

Energy Alternatives India (EAI) (2011) Ocean energy in India. Retrieved from http://www.eai.in/ref/ae/oce/oce.html. Feb 2012

Lee ST, Dechamps C (1978) Mathematical model for economic evaluation of tidal power in the Bay of Fundy. IEEE Trans Power Appar Syst 5:1769–1778.

Lee KY, El-Sharkawi MA (eds) (2008) Modern heuristic optimization techniques: theory and applications to power systems, vol 39. Wiley-IEEE press, Hoboken.

Ministry of New and Renewable Energy (MNRE) (2012) Tidal energy. Retrieved from http://www.mnre.gov.in/schemes/new-technologies/tidal-energy/. Feb 2012

Sullivan P, McCombie P (2013) Optimisation of tidal power arrays using a genetic algorithm. Proc Inst Civil Eng-Energy sar:1–10.

Swales MC, Wilson EM (1968) Optimization of tidal power generation. Water Power 20(3): 109–114.

Chapter 18
AGROSIM: A New Model for Predicting Water Productivity from Crop Characteristics

Mrinmoy Majumder, Tilottama Chackraborty, and Rabindra Nath Barman

Abstract The water productivity of any crop can be defined as the profit earned per unit of water expended for the harvesting the crop. This index is a measure of the profitability of a crop based on the amount of water required. Such an index can help farmers select crop species for cultivation. Based on their investment ability and desired profits, farmers want to select species that will maximize their profit. The present study tries to highlight a software developed to help farmers select crops according to their productivity or profitability. In this regard, after entering some related information, the farmer can obtain an estimate of the profit that can be earned from a specific crop over the next 5 years. The software can also optimize the water requirements and amount of crops that must be harvested to earn a level of profit specified by the farmer. Farmers can select species of crop according to the estimated water requirements. Adequate help files are also provided. The main objective of the software is to provide a platform that allows farmers to implement scientific and informed cultivation practices.

Keywords Water productivity • Agro-optimization • Agro-based simulation systems

18.1 Introduction

The software known as AGROSIM, that is, Agriculture Optimization and Simulation System (Fig. 18.1a), was developed to provide a knowledge base that would enable farmers to practice cultivation that is logical and profitable. The main idea behind

M. Majumder (✉) • T. Chackraborty
School of Hydro-Informatics Engineering, National Institute of Technology Agartala, Barjala, Jirania 799055, Tripura, India
e-mail: mmajumder15@gmail.com; tilottama86@gmail.com

R.N. Barman
Department of Production Engineering, National Institute of Technology Agartala, Barjala, Jirania 799055, Tripura, India
e-mail: rahul.barman@yahoo.co.in

M. Majumder and R.N. Barman (eds.), *Application of Nature Based Algorithm in Natural Resource Management*, DOI 10.1007/978-94-007-5152-1_18,

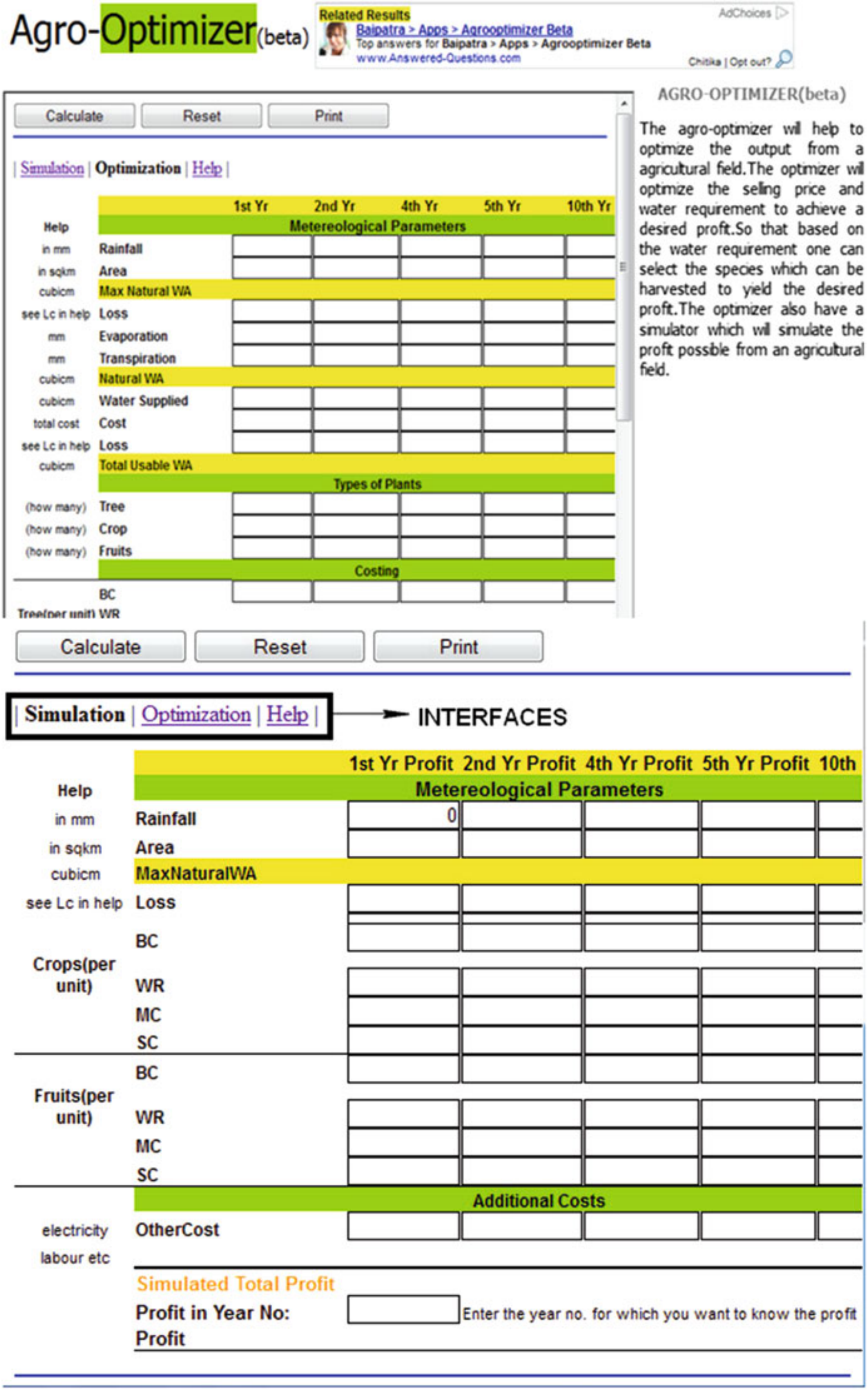


Fig. 18.1 (**a**) Screenshot of AGROSIM software. (**b**) AGROSIM interfaces

the development of the platform was to optimize the crop selection procedure and to help agribusinessmen minimize uncertainty and maximize profits. In this regard, the concept of water productivity was taken as a basis for optimization because this index represents both profit and water requirements. Each crop has its own requirements for water. Thus, this index can estimate profitability based on water availability (in the case of simulation) and requirements (in the case of optimization). The software also ensures that output will tally with the practical needs of farmers. Instead of a simple number or percentage, the exact amount of crop to be harvested or the volume of water required was estimated as output, which enables farmers to select the crops accordingly based on the investment capabilities and availability of water for irrigation on their land.

18.1.1 Justification of the Software Development

Farmers in India mainly face two obstacles in achieving their desired profit level. The first constraint is the availability of water; the second is the amount of rupees they can invest. While the former is a function of location, climate, and size of the dependent population, the latter entirely depends on the farmer's socioeconomic status. Thus, farmers in India are very concerned with the amount of money they can invest, and they select their crops based on water availability and the economy. But if, because of climatic abnormalities, the availability of water becomes scarcer, or if their investable funds are suddenly reduced due to unforeseen events, then the entire harvest will also be jeopardized, which could force the farmer to borrow money from private and unregulated lenders at a high interest. As cultivation goals become hampered, the resulting profit will fail to reach the desired level. Because of this reduction in profit, farmers' capacity to repay loans decreases. Once farmers are unable or less able to repay loans, lenders become land grabbers and eventually the socioeconomic stigma associated with defaulting on one's debt may lead farmers to take extreme measures. That is why farmers must be educated about the uncertainties well in advance, water requirements must be carefully estimated, and potential uncertainties must be considered while estimating actual requirements. Thus a water productivity index (Liu et al. 2013; Niroula and Singh 2013; Kumar et al. 2013) can greatly aid farmers in this regard. Water productivity of the area of interest is huge and high. If the selection of crops is made with respect to the amount of water available and the cops' profitability, then even if uncertain situations do arise, farmers can be prepared well in advance, before the cultivation season begins. Thus, it was for these reasons – to help farmers educate themselves about possible uncertainties and the impact of those uncertainties on the farmers' lives, to help farmers prepare a mitigation plan in advance, and to select crops according to farmers' desired profit level – that AGROSIM software was developed similar to the studies conducted by Hörbe et al. (2013), Wu et al. (2013) or Babbar-Sebens et al. (2013) etc.

18.2 Input, Output, and Working Principle of the Software

The software has two interfaces: simulation and optimization (Fig. 18.1). Along with these two interfaces is a third interface: help. The help interface shows the necessary explanation of the different inputs and output. Because the methodology of the software is self-explanatory, the working procedure is not explained explicitly. The simulation interface has several sections with different inputs. There are four sections in the simulation interface: meteorological parameters, types of plants, costing, and additional costs.

The meteorological parameters section (Fig. 18.2) has eight values: rainfall, area, evaporation and transpiration, water supplied, and cost of water. Loss due to watershed characteristics and supply media are also included. The main aim of this section is to estimate the amount of water available for irrigation. Based on area input, the software requires the user to enter the command area of the irrigation field. The required units in which the input values are desired are also shown beside all the input parameters. After all the required values are entered in the required units, the volume of total usable water is displayed as the output of this section in the *Total Usable Water Availability* (WA) field.

In the next section, i.e., Types of Plants section (Fig. 18.3), the software will ask the user to enter the total number of trees/crops/fruits he or she would like to harvest

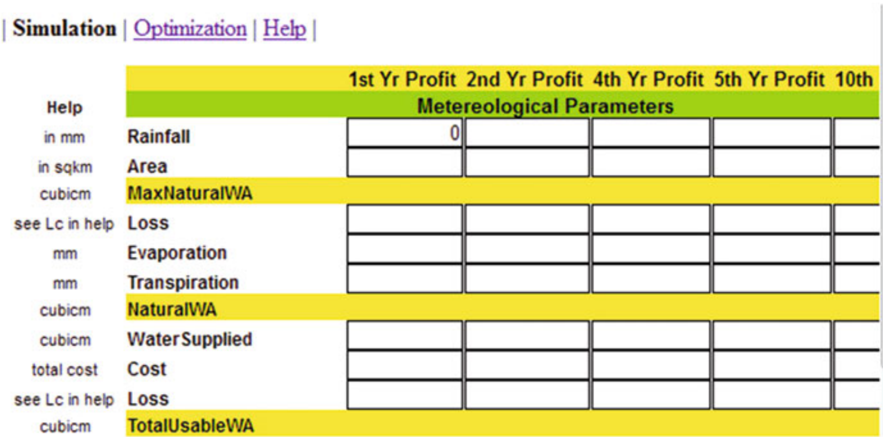


Fig. 18.2 Meteorologic parameter section (MPS) of simulation interface (MPS is the same for both simulation and optimization interfaces)

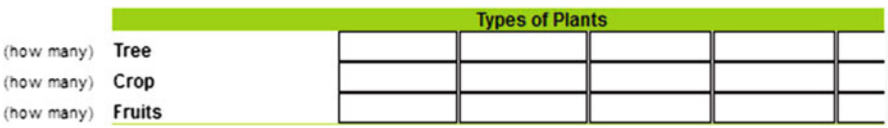


Fig. 18.3 Types of plant section of simulation interface

Fig. 18.4 Simulation interface, costing section

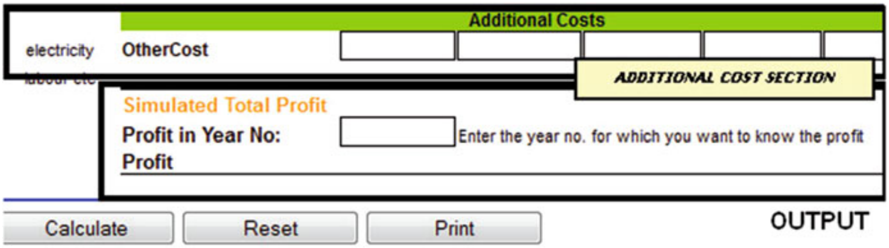


Fig. 18.5 Simulation interface, additional cost and output sections

in the upcoming season. To make the software usable for all agrobusinessmen (farmers, floriculturists, or plantation owners), the software has an added provision for simulating tree and fruit productivity as well.

The next section is the costing section (Fig. 18.4), where the economic features of the crops are entered. Inputs like buying costs, maintenance costs, selling costs, and water required for harvesting the plants must also be entered. If trees are cultivated, the details of tree cultivation are required, and if crops are cultivated, then the economic characteristics of the crops must be entered, and so on. The next and last sections on inputs are the Other Cost section (Fig. 18.5), where the costs of electricity and other related utilities, along with labor costs, are required.

All input values must be given based on a 10-year span if the user wants to estimate profits on the basis of crop yields from the first to the tenth years. If the user wants to predict only the first year's profit, he or she will not be required to enter the details of the inputs for the subsequent 9 years. The software makes it possible to optimize water requirements and the optimal cost to sell water for the first two, fourth, fifth, and tenth years only. After entering all the necessary details of the required inputs, the profit at the end of the first to tenth years will be calculated according to the year

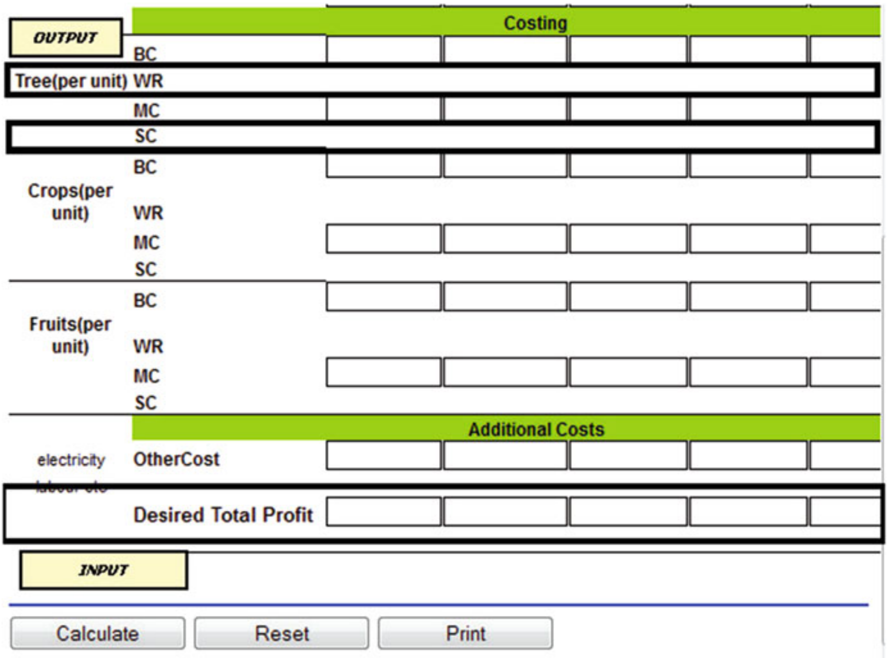


Fig. 18.6 Optimization interface highlighting basic differences with simulation interface

number entered in the *Profit in Year No.* field. Thus, based on the profit, the user can plan his or her harvest such that a desired level of income can be attained with no probability of shortfall.

The user can also simulate uncertain situations by modifying the values entered in the input field. For example, to represent a 100-year rainfall, which is a rainfall that occurs only once every 100 years (extremely rare), the value of such rainfall can be entered in the Rainfall input field. The Evaporation and Transpiration at the time of such a rainfall events would be very low. So the adjusted values of these variables would have to be entered into the respective input boxes. The artificial supply of water would remain the same, but the loss from the watershed would be reduced as the excess water saturated the soil pores. Once the meteorologic variables were adjusted, the user would have to enter the details of all other inputs as he or she would do under normal conditions. Clicking on the Calculate button after entering the number of years after which the user wants to estimate the profit for farming, the output will show the profit based on the impact of the extreme event. That is why with the help of the software farmers can make a plan of the entire crop cycle in a 10-year span considering the risks involved. The impact of urbanization can also be verified with the help of the software by adjusting the command area as urbanization will surely increase the watershed area while trying to accommodate the overgrowing population.

The optimization interface (Fig. 18.6) of the software has the same number of fields. The only difference is the number of inputs and outputs has changed. In the case of optimization, the profit desired by the farmer at the end of the first, second, fourth, fifth, and tenth years must be entered along with the meteorologic and command area characteristics in the *Meteorological Parameter* section and *Buying and Maintenance Costs* in the Costing section. Once the user clicks the Calculate button, the water required to yield the desired amount of profit will be displayed in the WR field of the plant type, which has the *Buying Cost* and *Maintenance Cost* entered in the respective fields. The *Selling Cost* will also be estimated by the software. Thus with the help of the optimization interface farmers can calculate the amount of water required and the selling cost to sell their harvest for attaining the desired goal of profitability within a desired time period.

18.3 Benefits of the Software

The benefits of using the software are manifold. From the farmer's point of view the solution provided by this software will ensure a logical and reliable selection of crops so that the profitability of the user is not compromised under normal as well as uncertain conditions. In addition to helping in the selection of crops, the methodology of cultivation for attaining the desired profit level can also be idealized from the output of the optimization interface of the software. The software makes it possible to help not only farmers but also plantation owners and floriculturist. It can estimate the ideal plant for each type of user. Not only farmers but also agro-based business owners will also benefit from using the software. They will be able to predict their profitability well in advance by changing the meteorological parameters, command area, and number of plants they will need to harvest. Such users can create different scenarios and estimate profit from the simulation interface after the desired number of years. This output will help business owners to plan their investments accordingly. Scientists and engineers with an agricultural background can estimate the water productivity of different crops and the impact on water productivity of both climate change and urbanization using the AGROSIM for their research-based objectives.

18.4 Drawbacks of the Software

The software has two drawbacks. Although it estimates the required amount of water or the profit that can be achieved given a certain volume of water available per plant, which can help users select a plant according to its water requirements, the software has no provision for automatically selecting the ideal species based on the output. But this may mean introducing more complexity to the software and may promote preferential treatment of certain types of plant that may not be acceptable for certain farmers. Another major drawback of the software is that the time domain is fixed at the optimization interface. The software cannot optimize water requirements and selling costs for more than 10 years. This feature may be modified and changed to the system utilized in the Simulation panel where users enter the number of years after which they want to estimate the profit. Perhaps in the next version, the authors will introduce such features in the present framework.

18.5 Conclusion

This study introduced a new software called AGROSIM. The software, developed using JavaScript and MS Excel, simulates profits if the availability of water and number of plants to be harvested are entered as inputs. The software simulates profits for as many years as the user wants. In one of its interfaces, the software helps to optimize the water required for cultivation and selling costs to attain the desired level of profitability from the cultivation. Based on these outputs from the software, the farmer will be able to select the ideal crop species and become knowledgeable about the uncertainty that may be introduced due to climate change and unchecked population growth. The software has two limitations: it cannot estimate crop species directly from the output, and during optimization only a 10-year desired profit level can be entered. Thus, the user cannot optimize water requirements and selling costs to attain a profit level over 10 years or within 1 year. Such features may be introduced in a future version of the software to make it a very useful tool for farmers as well as agroscientists. It is possible that the software system may be encoded into a microprocessor to allow the development of a portable meter for analysis of agricultural requirements to achieve the profitability desired at the beginning of the study, which in turn would help farmers calculate their profitability and requirements before initializing the harvest.

References

Babbar-Sebens M, Barr RC, Tedesco LP, Anderson M (2013) Spatial identification and optimization of upland wetlands in agricultural watersheds. Ecol Eng 52:130–142

Hörbe TAN, Amado TJC, Ferreira AO, Alba PJ (2013) Optimization of corn plant population according to management zones in Southern Brazil. Precis Agric sara:1–16

Liu HJ, Kang Y, Yao SM, Sun ZQ, Liu SP, Wang QG (2013) Field evaluation on water productivity of winter wheat under sprinkler or surface irrigation in the North China plain. Irrig Drain 62:37–49

Kumar P, Pannu RK, Singh A, Jain A (2013) Productivity and water use in wheat (*Triticum aestivum*) under different sources of nutrition and irrigation levels. Ann Agri Bio Res 18(1):23–28

Niroula B, Singh KLB (2013) Seasonal variation in importance value index (IVI), diversity indices and biomass of aquatic macrophytes at Biratnagar and adjoining areas, eastern Nepal. Nepal J Biosci 1:1–7

Wu C, Yang J, Shang J (2013) The particle swarm optimization based calculation research on agricultural mechanization contribution rate. Adv Sci Lett 19(4):1170–1172

Chapter 19
QUALTR: Software for Simulating Output from Water Treatment Plants in the Face of Extreme Events

Mrinmoy Majumder, Paulami De, and Rabindra Nath Barman

Abstract Water treatment plants (WTPs) are responsible for supplying treated water to domestic, industrial, and agricultural users. When water is treated, the quality is upgraded to a predetermined standard. Most WTPs take water in from surface water sources like rivers or wetlands, although the quality of surface water follows a uniform trend. The amount and time of waste influx are predictable and depend upon the domestic, industrial, and agricultural consumers. But during an extreme event like rainfall the surface runoff contaminates the water bodies by washing out both inorganic and organic impurities from the watershed. Such impurities include toxic materials used in fertilizers, cleaning agents, human or animal wastes, etc. In summary, after a storm, surface runoff increases the level of pollutants in water bodies, and WTPs that take in water from those bodies must adjust their chemical dosing to control and maintain the prescribed quality standard. But if the sudden degradation of water quality is not monitored in real time, the contaminated water will be delivered to consumers. The supply of low-quality water can compromise public health and consumer satisfaction and may even create a health hazard that would be expensive and difficult to control. That is why the present investigation will try to introduce a software that can predict the status of treated water with respect to changes occurring in various related quality parameters of the surface water as well as the amount of chemical dosing. If the software is linked with a real-time monitoring system, then it can alert plant managers well in advance of an incoming quality hazard. A neural network was selected as the prediction algorithm due to its ability to identify complex

M. Majumder (✉) • P. De
School of Hydro-Informatics Engineering, National Institute of Technology Agartala, Barjala, Jirania 799055, Tripura, India
e-mail: mmajumder15@gmail.com; depaulami@yahoo.co.in

R.N. Barman
Department of Production Engineering, National Institute of Technology Agartala, Barjala, Jirania 799055, Tripura, India
e-mail: rahul.barman@yahoo.co.in

M. Majumder and R.N. Barman (eds.), *Application of Nature Based Algorithm in Natural Resource Management*, DOI 10.1007/978-94-007-5152-1_19,

relationships between nonlinear variables. The potential of the software was verified with the help of real-time quality data of source water and the corresponding pattern of dosing followed in a small urban WTP.

Keywords Water treatment plants • Simulation software • Neuro-genetic models

19.1 Introduction

The water supplied to domestic, industrial, and agriculture users must at least be satisfactory for drinking purposes from the standpoint of its chemical, physical, and biological characteristics. Drinking water should preferably be obtained from a source free of pollution. The raw water normally sourced from surface water bodies is not suitable for drinking purposes. The objective of any water treatment plant (WTP) is to produce safe and potable drinking water.

19.1.1 Hazard Analysis of Water Treatment Plants

There are four major categories of hazards in any WTP: or workers, treatment, Machinery, Materials, and Methodology of treatment. All kinds of human-imposed errors are classified in the People group. If the dosing of chemicals is performed manually, then, because of absentmindedness or lack of knowledge, the responsible person can over- or underdose, which can cause variations in the quality pattern of the treated waters. The person responsible for cleaning or maintaining the instruments can also induce damages that may result in a decrease in plant efficiency. That is why workers must be trained and supervised by someone who is knowledgeable about these issues. Proper measures must be taken to ensure the physical and mental agility of factory personnel.

The second category of hazard, that is, Machinery, involves errors in the treatment machinery such as dosers, clari-flocculators, and lime injectors. The age of a machine is inversely proportional to its efficiency. The constant work load on a machine can impose precision errors in machine output that might impact the capacity of the WTP. That is why proper monitoring and maintenance is required, and if necessary, old and inefficient machines must be replaced so that the plant efficiency is not compromised.

Material mainly includes variations in the quality of the raw water. Change in the quality of water may take place because of an influx of watershed runoff after a storm, which is generally contaminated by pesticides and fertilizers (if chemical concentrations in irrigated fields are high in the command area), different types of microorganisms and micropollutants [if the concentration of waste water treatment plants (WWTPs) is high in the watershed], and nutrients like carbon, nitrogen, and phosphorus. The effluent from domestic, industrial, and WWTPs can also contaminate raw water. That is why constant monitoring of raw water quality is required to adjust the dosing of chemicals.

Inaccuracies may also be introduced by the inadequacy of the production methodology. The selected treatment procedure and number of phases it has determine the quality of treated water. There are mainly two kinds of inaccuracy involved in the different chemical processes conducted to treat water: degradation of raw water quality and retention/inactivation in the treatment steps. The first type of error can be controlled by real-time monitoring of incoming water and adjustment of the dosing with respect to the raw water quality. The problem of retention or inactivation can be removed if proper coupling can be achieved within the reactor hydraulics and chemical kinematics. Local governments must also be consulted to understand the quality of source water and also to guarantee the quality standards of such water bodies. The methodology chosen to treat raw water will be based on the quality of raw water and inactivation/retention hydraulics.

In summary, it can be concluded that WTPs mainly face four kinds of hazards – Men, Machinery, Material, and Methodology – but it is mainly the presence or absence of micropollutants, microorganisms, hazardous chemicals (trimethylbenzene, EDTA, phenol, etc.), certain nutrients, and suspended materials in raw water and retention hydraulics and their relationship with the chemical kinetics within the different steps of the water treatment procedure that determines the amount of uncertainty and the standards of the production output of WTPs.

19.1.2 Components of Water Treatment Plants

WTPs can be broadly classified into surface and groundwater treatment plants. The water to be treated or raw water can also be collected from industrial or residential waste. A treatment plant used for the purification of wastewater is classified as a WWTP. A treatment plant that purifies spring water is known as a spring water treatment plant.

The components of treatment plants vary depending on purpose of the treated water. The phases through which raw water is passed and treated to produce freshwater and for distribution to consumers are as follows:

1. To prevent suspended materials (e.g., debris, tree leaves) from entering the internal pipeline of the treatment plant, a screen made up of iron and having pore spaces depending on the size of the allowed impurities is generally used . This stage is known as screening.
2. The turbidity of the raw water is reduced by the addition of aluminum or iron salt in the coagulation phase, which is normally followed by a flocculation, sedimentation, and filtration subphase. The purpose of the treated water and condition of the raw water determine the application of the subphases.
3. The coagulated water, following deposition of the froth, is fed to the filtration units. The filter is either a simple slow sand filter or complex activated carbon or reverse osmosis unit. The complexity and efficiency of the unit vary depending on the purpose and condition of the treated and raw water.
4. In the next phase, the water is disinfected to remove pathogens like bacteria, viruses, protozoa, and algal toxins (microcystine). In this phase, generally chlorine, ozone, chloramines, or chlorine dioxide is added to remove the micropollutants.

5. Sometimes to remove hardness and acidity or alkalinity, lime is mixed with the water after raw water is pumped into the treatment plant. But this step is only available when the quality of the water is below the national standard. This step is referred as lime softening. Some other optional steps are available in WWTPs like stabilization (to prevent corrosion and scale formation), activated carbon adsorption (to remove taste- and odor-causing chemicals or synthetic organic contaminants), fluoridation (to increase the concentration of fluoride to the optimum level for the prevention of dental cavities), preozonation (dosing of ozone to disinfect the water before the main ozonation process is initiated).

19.1.3 Impact of Extreme Climate on Water Treatment Plants

Post rainfall, the concentration of organic matter, bacteria, viruses, parasites, and pesticides in streams, rivers, and lakes, along with turbidity, increases. Surface runoff carries pollutants and deposits them into the water bodies. Climatic abnormality is now common in most regions of the world. Change in the amount and frequency of rainfall have forced a redesign of hydraulic structures. Climate change may also affect both surface water and groundwater quality. Increases or decreases in precipitation will impact the concentration of pollutants. The turbidity level will also vary. These changes, coupled with an increasing population in urban areas, will impact the applications of regular materials, machinery, and methodologies of WTPs. The storage requirements for freshwater may increase or decrease depending on the way climate change impacts the supply region of the plant. Uncertain precipitation patterns will impose uncertainty on water supplies. The demand will increase due to increases in population, and so supply will also have to increase. Thus either the capacities of present WTPs must be modified or new supply lines will have to be introduced to meet the rising demand. The demand for maintaining quality will also increase with increased requirements for high-quality machinery, material, and workforce. The uncertainty in weather patterns will create the need for optimization, conservation, reuse, and storage. The result of increased demand will create an increased need for strict regulations to control effluent from consumers.

19.1.4 Objective and Scope of the Present Software

This chapter represents an attempt to introduce a new software framework that can predict the quality of raw and treated water. The output from each of the subcomponents is predicted separately based on the input. The impact of chemical dosing was also considered. Overall the software has one main interface and three subinterfaces. The home interface and all the subinterfaces are interconnected with each other. The user will be required to enter the quality of the raw water in the home panel. All other inputs of the subinterfaces will be automatically populated. Four cascading neurogenetic models with seven subsections in each unit were utilized to predict the overall quality of the output water. The quality of the treated water was predicted with the help of an index developed from the available inputs.

19.1.5 Brief Methodology

The input of the developed model included quality parameters like turbidity, concentration of residual chlorine, pathogens, total and fecal coliform, conductivity and pH; materials that prevents suspended macroscopic impurities from entering the plant like screen pore size, purification agents like aluminum and iron salts and lime concentration; physical parameters like speed of clariflocculator turbines, age of filter beds, pore size of filter beds, and hydraulic properties like flow velocity within the pipelines. In terms of output, the quality index of the treated water was the ultimate variable predicted by the neurogenetic model. As for the submodels, the quality parameters of the incoming water and the hydraulic and physical parameters were the input and a single quality parameter was predicted as output. Hydraulic parameters were also predicted in the submodels as the output of the same inputs. The output of the submodels was then fed to the models of the next stage as input.

The cascading neurogenetic models were then used in the prediction of the water quality index of the treated water.

The dataset of the variables were encoded into 17 categories representing its magnitude or quality and the influence it will create on the output variable. All the combinations that may generate between the input and output variables were produced and fed into the model as training data.

19.2 Development of Neurogenetic Models

The neural networks are not popularly utilized to predict parameters of WTP.But some studies are already conducted to monitor the variation in concentration of impurities like Yan (2013); Nourani (2013); Pai (2013); Li (2013) etc. The neurogenetic models for the software were developed based on the available subphases, purpose, and type of WTP. Based on the type of WTP, quality parameters were either removed or introduced. The purpose of the treated water also determines the number of subsections to be considered in the model. That is why the first step when working with the software is to select the type and purpose of the WTP. Then, in the second window, the user enters all the necessary input data like quality parameters of the raw water, physical properties of the machinery, and patterns of chemical dosing and hydraulic characteristics (flow velocity) of the pipeline.

Table 19.1 shows the inputs and outputs required for all the main and submodels. The output from the input of the submodels is also explicitly highlighted.

All the models are trained by a combinatorial data matrix that includes every possible combination resulting from both normal and uncertain situations within the input variables. As discussed earlier, all the variables were categorized into 17 groups. Table 19.2 highlights the categories and the ratings given to them with respect to their relationship with the water quality of the treated water.

Table 19.1 Input and output of main and submodels considered in proposed software framework

Model name	Model type	Input	Output
QTRWTP	Main	Turbidity (*T*) Total dissolved solids (TDS) Free chlorine (Cl) Pathogens (p) Total and fecal coliform (TC and FC) Conductivity (*K*) pH Dissolved oxygen (DO) Impurities Screen pore size (SPS) Flow velocity (*V*)	Water quality index (QI) of treated water
Intake to Aerator			
WTR011	Sub	Turbidity (*T*) Total dissolved solids (TDS) Free chlorine (Cl) Pathogens (p) Total and fecal coliform (TC and FC) Conductivity (*K*) pH Dissolved oxygen (DO) Impurities Screen pore size (SPS) Flow velocity (*V*)	Turbidity (*T*)

WTR012	Sub	Turbidity (T)	Total dissolved solids (TDS)
		Total dissolved solids (TDS)	
		Free chlorine (Cl)	
		Pathogens (p)	
		Total and fecal coliform (TC and FC)	
		Conductivity (K)	
		pH	
		Dissolved oxygen (DO)	
		Impurities	
		Screen pore size (SPS)	
		Flow velocity (V)	
WTR013	Sub	Turbidity (T)	Free chlorine (Cl)
		Total dissolved solids (TDS)	
		Free chlorine (Cl)	
		Pathogens (p)	
		Total and fecal coliform (TC and FC)	
		Conductivity (K)	
		pH	
		Dissolved oxygen (DO)	
		Impurities	
		Screen pore size (SPS)	
		Flow velocity (V)	

(continued)

Table 19.1 (continued)

Model name	Model type	Input	Output
WTR014	Sub	Turbidity (*T*)	Pathogens (p)
		Total dissolved solids (TDS)	
		Free chlorine (Cl)	
		Pathogens (p)	
		Total and fecal coliform (TC and FC)	
		Conductivity (*K*)	
		pH	
		Dissolved oxygen (DO)	
		Impurities	
		Screen pore size (SPS)	
		Flow velocity (*V*)	
WTR015	Sub	Turbidity (*T*)	Total coliform (TC)
		Total dissolved solids (TDS)	
		Free chlorine (Cl)	
		Pathogens (p)	
		Total and fecal coliform (TC and FC)	
		Conductivity (*K*)	
		pH	
		Dissolved oxygen (DO)	
		Impurities	
		Screen pore size (SPS)	
		Flow velocity (*V*)	

WTR016	Sub	Turbidity (*T*)	Fecal coliform (FC)
		Total dissolved solids (TDS)	
		Free chlorine (Cl)	
		Pathogens (p)	
		Total and fecal coliform (TC and FC)	
		Conductivity (*K*)	
		pH	
		Dissolved oxygen (DO)	
		Impurities	
		Screen pore size (SPS)	
		Flow velocity (*V*)	
WTR017	Sub	Turbidity (*T*)	Conductivity (*K*)
		Total dissolved solids (TDS)	
		Free chlorine (Cl)	
		Pathogens (p)	
		Total and fecal coliform (TC and FC)	
		Conductivity (*K*)	
		pH	
		Dissolved oxygen (DO)	
		Impurities	
		Screen pore size (SPS)	
		Flow velocity (*V*)	

(continued)

Table 19.1 (continued)

Model name	Model type	Input	Output
WTR018	Sub	Turbidity (T)	pH
		Total dissolved solids (TDS)	
		Free chlorine (Cl)	
		Pathogens (p)	
		Total and fecal coliform (TC and FC)	
		Conductivity (K)	
		pH	
		Dissolved oxygen (DO)	
		Impurities	
		Screen pore size (SPS)	
		Flow velocity (V)	
WTR019	Sub	Turbidity (T)	Impurities
		Total dissolved solids (TDS)	
		Free chlorine (Cl)	
		Pathogens (p)	
		Total and fecal coliform (TC and FC)	
		Conductivity (K)	
		pH	
		Dissolved oxygen (DO)	
		Impurities	
		Screen pore size (SPS)	
		Flow velocity (V)	

WTR110	Sub	Turbidity (*T*)	Screen pore size (SPS)
		Total dissolved solids (TDS)	
		Free chlorine (Cl)	
		Pathogens (p)	
		Total and fecal coliform (TC and FC)	
		Conductivity (*K*)	
		pH	
		Dissolved oxygen (DO)	
		Impurities	
		Screen pore size (SPS)	
		Flow velocity (*V*)	
WTR111	Sub	Turbidity (*T*)	Flow velocity (*V*)
		Total dissolved solids (TDS)	
		Free chlorine (Cl)	
		Pathogens (p)	
		Total and fecal coliform (TC and FC)	
		Conductivity (*K*)	
		pH	
		Dissolved oxygen (DO)	
		Impurities	
		Screen pore size (SPS)	
		Flow velocity (*V*)	

(continued)

Table 19.1 (continued)

Model name	Model type	Input	Output
Aerator to clariflocculator			
WTR021	Sub	Turbidity (T)	Turbidity (T)
		Total dissolved solids (TDS)	
		Free chlorine (Cl)	
		Pathogens (p)	
		Total and fecal coliform (TC and FC)	
		Conductivity (K)	
		pH	
		Dissolved oxygen (DO)	
		Impurities	
		Salt dosing (S)	
		Lime dosing (LD)	
		Flow velocity (V)	
WTR022	Sub	Turbidity (T)	Total dissolved solids (TDS)
		Total dissolved solids (TDS)	
		Free chlorine (Cl)	
		Pathogens (p)	
		Total and fecal coliform (TC and FC)	
		Conductivity (K)	
		pH	
		Dissolved oxygen (DO)	
		Impurities	
		Salt dosing (S)	
		Lime dosing (LD)	
		Flow velocity (V)	

WTR023	Sub	Turbidity (*T*) Total dissolved solids (TDS) Free chlorine (Cl) Pathogens (p) Total and fecal coliform (TC and FC) Conductivity (*K*) pH Dissolved oxygen (DO) Impurities Salt dosing (*S*) Lime dosing (LD) Flow velocity (*V*)	Free chlorine (Cl)
WTR024	Sub	Turbidity (*T*) Total dissolved solids (TDS) Free chlorine (Cl) Pathogens (p) Total and fecal coliform (TC and FC) Conductivity (*K*) pH Dissolved oxygen (DO) Impurities Salt dosing (*S*) Lime dosing (LD) Flow velocity (*V*)	Pathogens (p)

(continued)

Table 19.1 (continued)

Model name	Model type	Input	Output
WTR025	Sub	Turbidity (*T*)	Total and fecal coliform (TC and FC)
		Total dissolved solids (TDS)	
		Free chlorine (Cl)	
		Pathogens (p)	
		Total and fecal coliform (TC and FC)	
		Conductivity (*K*)	
		pH	
		Dissolved oxygen (DO)	
		Impurities	
		Salt dosing (*S*)	
		Lime dosing (LD)	
		Flow velocity (*V*)	
WTR026	Sub	Turbidity (*T*)	Conductivity (*K*)
		Total dissolved solids (TDS)	
		Free chlorine (Cl)	
		Pathogens (p)	
		Total and fecal coliform (TC and FC)	
		Conductivity (*K*)	
		pH	
		Dissolved oxygen (DO)	
		Impurities	
		Salt dosing (*S*)	
		Lime dosing (LD)	
		Flow velocity (*V*)	

WTR027	Sub	Turbidity (T) Total dissolved solids (TDS) Free chlorine (Cl) Pathogens (p) Total and fecal coliform (TC and FC) Conductivity (K) pH Dissolved oxygen (DO) Impurities Salt dosing (S) Lime dosing (LD) Flow velocity (V)	pH
WTR028	Sub	Turbidity (T) Total dissolved solids (TDS) Free chlorine (Cl) Pathogens (p) Total and fecal coliform (TC and FC) Conductivity (K) pH Dissolved oxygen (DO) Impurities Salt dosing (S) Lime dosing (LD) Flow velocity (V)	Dissolved oxygen (DO)

(continued)

Table 19.1 (continued)

Model name	Model type	Input	Output
WTR029	Sub	Turbidity (*T*)	Impurities
		Total dissolved solids (TDS)	
		Free chlorine (Cl)	
		Pathogens (p)	
		Total and fecal coliform (TC and FC)	
		Conductivity (*K*)	
		pH	
		Dissolved oxygen (DO)	
		Impurities	
		Salt dosing (*S*)	
		Lime dosing (LD)	
		Flow velocity (*V*)	
WTR210	Sub	Turbidity (*T*)	Flow velocity (*V*)
		Total dissolved solids (TDS)	
		Free chlorine (Cl)	
		Pathogens (p)	
		Total and fecal coliform (TC and FC)	
		Conductivity (*K*)	
		pH	
		Dissolved oxygen (DO)	
		Impurities	
		Salt dosing (*S*)	
		Lime dosing (LD)	
		Flow velocity (*V*)	

Clarifocculator to sedimentation basin			
WTR031	Sub	Turbidity (T)	Turbidity (T)
		Total dissolved solids (TDS)	
		Free chlorine (Cl)	
		Pathogens (p)	
		Total and fecal coliform (TC and FC)	
		Conductivity (K)	
		pH	
		Dissolved oxygen (DO)	
		Impurities	
		Salt traces (ST)	
		Flow velocity (V)	
WTR032	Sub	Turbidity (T)	Total dissolved solids (TDS)
		Total dissolved solids (TDS)	
		Free chlorine (Cl)	
		Pathogens (p)	
		Total and fecal coliform (TC and FC)	
		Conductivity (K)	
		pH	
		Dissolved oxygen (DO)	
		Impurities	
		Salt traces (ST)	
		Flow velocity (V)	

(continued)

Table 19.1 (continued)

Model name	Model type	Input	Output
WTR033	Sub	Turbidity (T)	Free chlorine (Cl)
		Total dissolved solids (TDS)	
		Free chlorine (Cl)	
		Pathogens (p)	
		Total and fecal coliform (TC and FC)	
		Conductivity (K)	
		pH	
		Dissolved oxygen (DO)	
		Impurities	
		Salt traces (ST)	
		Flow velocity (V)	
WTR034	Sub	Turbidity (T)	Pathogens (p)
		Total dissolved solids (TDS)	
		Free chlorine (Cl)	
		Pathogens (p)	
		Total and fecal coliform (TC and FC)	
		Conductivity (K)	
		pH	
		Dissolved oxygen (DO)	
		Impurities	
		Salt traces (ST)	
		Flow velocity (V)	

WTR035	Sub	Turbidity (T) Total dissolved solids (TDS) Free chlorine (Cl) Pathogens (p) Total and fecal coliform (TC and FC) Conductivity (K) pH Dissolved oxygen (DO) Impurities Salt traces (ST) Flow velocity (V)	Total and fecal coliform (TC and FC)
WTR036	Sub	Turbidity (T) Total dissolved solids (TDS) Free chlorine (Cl) Pathogens (p) Total and fecal coliform (TC and FC) Conductivity (K) pH Dissolved oxygen (DO) Impurities Salt traces (ST) Flow velocity (V)	Conductivity (K)

(continued)

Table 19.1 (continued)

Model name	Model type	Input	Output
WTR037	Sub	Turbidity (T)	pH
		Total dissolved solids (TDS)	
		Free chlorine (Cl)	
		Pathogens (p)	
		Total and fecal coliform (TC and FC)	
		Conductivity (K)	
		pH	
		Dissolved oxygen (DO)	
		Impurities	
		Salt traces (ST)	
		Flow velocity (V)	
WTR038	Sub	Turbidity (T)	Dissolved oxygen (DO)
		Total dissolved solids (TDS)	
		Free chlorine (Cl)	
		Pathogens (p)	
		Total and fecal coliform (TC and FC)	
		Conductivity (K)	
		pH	
		Dissolved oxygen (DO)	
		Impurities	
		Salt traces (ST)	
		Flow velocity (V)	

WTR039	Sub	Turbidity (T)	Impurities
		Total dissolved solids (TDS)	
		Free chlorine (Cl)	
		Pathogens (p)	
		Total and fecal coliform (TC and FC)	
		Conductivity (K)	
		pH	
		Dissolved oxygen (DO)	
		Impurities	
		Salt traces (ST)	
		Flow velocity (V)	
WTR310	Sub	Turbidity (T)	Salt traces (ST)
		Total dissolved solids (TDS)	
		Free chlorine (Cl)	
		Pathogens (p)	
		Total and fecal coliform (TC and FC)	
		Conductivity (K)	
		pH	
		Dissolved oxygen (DO)	
		Impurities	
		Salt traces (ST)	
		Flow velocity (V)	

(continued)

Table 19.1 (continued)

Model name	Model type	Input	Output
WTR311	Sub	Turbidity (T)	Flow velocity (V)
		Total dissolved solids (TDS)	
		Free chlorine (Cl)	
		Pathogens (p)	
		Total and fecal coliform (TC and FC)	
		Conductivity (K)	
		pH	
		Dissolved oxygen (DO)	
		Impurities	
		Salt traces (ST)	
		Flow velocity (V)	
Sedimentation basin to filter bed			
WTR041	Sub	Turbidity (T)	Turbidity (T)
		Total dissolved solids (TDS)	
		Free chlorine (Cl)	
		Pathogens (p)	
		Total and fecal coliform (TC and FC)	
		Conductivity (K)	
		pH	
		Dissolved oxygen (DO)	
		Impurities	
		Salt traces (ST)	
		Flow velocity (V)	

WTR042	Sub	Turbidity (T) Total dissolved solids (TDS) Free chlorine (Cl) Pathogens (p) Total and fecal coliform (TC and FC) Conductivity (K) pH Dissolved oxygen (DO) Impurities Salt traces (ST) Flow velocity (V)	Total dissolved solids (TDS)
WTR043	Sub	Turbidity (T) Total dissolved solids (TDS) Free chlorine (Cl) Pathogens (p) Total and fecal coliform (TC and FC) Conductivity (K) pH Dissolved oxygen (DO) Impurities Salt traces (ST) Flow velocity (V)	Free chlorine (Cl)

(continued)

Table 19.1 (continued)

Model name	Model type	Input	Output
WTR044	Sub	Turbidity (*T*) Total dissolved solids (TDS) Free chlorine (Cl) Pathogens (p) Total and fecal coliform (TC and FC) Conductivity (*K*) pH Dissolved oxygen (DO) Impurities Salt traces (ST) Flow velocity (*V*)	Pathogens (p)
WTR045	Sub	Turbidity (*T*) Total dissolved solids (TDS) Free chlorine (Cl) Pathogens (p) Total and fecal coliform (TC and FC) Conductivity (*K*) pH Dissolved oxygen (DO) Impurities Salt traces (ST) Flow velocity (*V*)	Total and fecal coliform (TC and FC)

WTR046	Sub	Turbidity (T)	Conductivity (K)
		Total dissolved solids (TDS)	
		Free chlorine (Cl)	
		Pathogens (p)	
		Total and fecal coliform (TC and FC)	
		Conductivity (K)	
		pH	
		Dissolved oxygen (DO)	
		Impurities	
		Salt traces (ST)	
		Flow velocity (V)	
WTR047	Sub	Turbidity (T)	pH
		Total dissolved solids (TDS)	
		Free chlorine (Cl)	
		Pathogens (p)	
		Total and fecal coliform (TC and FC)	
		Conductivity (K)	
		pH	
		Dissolved oxygen (DO)	
		Impurities	
		Salt traces (ST)	
		Flow velocity (V)	

(continued)

Table 19.1 (continued)

Model name	Model type	Input	Output
WTR048	Sub	Turbidity (*T*)	Dissolved oxygen (DO)
		Total dissolved solids (TDS)	
		Free chlorine (Cl)	
		Pathogens (p)	
		Total and fecal coliform (TC and FC)	
		Conductivity (*K*)	
		pH	
		Dissolved oxygen (DO)	
		Impurities	
		Salt traces (ST)	
		Flow velocity (*V*)	
WTR049	Sub	Turbidity (*T*)	Impurities
		Total dissolved solids (TDS)	
		Free chlorine (Cl)	
		Pathogens (p)	
		Total and fecal coliform (TC and FC)	
		Conductivity (*K*)	
		pH	
		Dissolved oxygen (DO)	
		Impurities	
		Salt traces (ST)	
		Flow velocity (*V*)	

WTR410	Sub	Turbidity (T) Total dissolved solids (TDS) Free chlorine (Cl) Pathogens (p) Total and fecal coliform (TC and FC) Conductivity (K) pH Dissolved oxygen (DO) Impurities Salt traces (ST) Flow velocity (V)	Salt traces (ST)
WTR411	Sub	Turbidity (T) Total dissolved solids (TDS) Free chlorine (Cl) Pathogens (p) Total and fecal coliform (TC and FC) Conductivity (K) pH Dissolved oxygen (DO) Impurities Salt traces (ST) Flow velocity (V)	Flow velocity (V)

(continued)

Table 19.1 (continued)

Model name	Model type	Input	Output
Filter bed to chlorination			
WTR051	Sub	Turbidity (T)	Turbidity (T)
		Total dissolved solids (TDS)	
		Free chlorine (Cl)	
		Pathogens (p)	
		Total and fecal coliform (TC and FC)	
		Conductivity (K)	
		pH	
		Dissolved oxygen (DO)	
		Impurities	
		Salt traces (ST)	
		Flow velocity (V)	
		Age of filter bed (F_a)	
		Pore size of filter bed (F_p)	
WTR052	Sub	Turbidity (T)	Total dissolved solids (TDS)
		Total dissolved solids (TDS)	
		Free chlorine (Cl)	
		Pathogens (p)	
		Total and fecal coliform (TC and FC)	
		Conductivity (K)	
		pH	
		Dissolved oxygen (DO)	
		Impurities	
		Salt traces (ST)	
		Flow velocity (V)	
		Age of filter bed (F_a)	
		Pore size of filter bed (F_p)	

WTR053	Sub	Turbidity (T)	Free chlorine (Cl)
		Total dissolved solids (TDS)	
		Free chlorine (Cl)	
		Pathogens (p)	
		Total and fecal coliform (TC and FC)	
		Conductivity (K)	
		pH	
		Dissolved oxygen (DO)	
		Impurities	
		Salt traces (ST)	
		Flow velocity (V)	
		Age of filter bed (F_a)	
		Pore size of filter bed (F_p)	
WTR054	Sub	Turbidity (T)	Pathogens (p)
		Total dissolved solids (TDS)	
		Free chlorine (Cl)	
		Pathogens (p)	
		Total and fecal coliform (TC and FC)	
		Conductivity (K)	
		pH	
		Dissolved oxygen (DO)	
		Impurities	
		Salt traces (ST)	
		Flow velocity (V)	
		Age of filter bed (F_a)	
		Pore size of filter bed (F_p)	

(continued)

Table 19.1 (continued)

Model name	Model type	Input	Output
WTR055	Sub	Turbidity (T)	Total and fecal coliform (TC and FC)
		Total dissolved solids (TDS)	
		Free chlorine (Cl)	
		Pathogens (p)	
		Total and fecal coliform (TC and FC)	
		Conductivity (K)	
		pH	
		Dissolved oxygen (DO)	
		Impurities	
		Salt traces (ST)	
		Flow velocity (V)	
		Age of filter bed (F_a)	
		Pore size of filter bed (F_p)	
WTR056	Sub	Turbidity (T)	Conductivity (K)
		Total dissolved solids (TDS)	
		Free chlorine (Cl)	
		Pathogens (p)	
		Total and fecal coliform (TC and FC)	
		Conductivity (K)	
		pH	
		Dissolved oxygen (DO)	
		Impurities	
		Salt traces (ST)	
		Flow velocity (V)	
		Age of filter bed (F_a)	
		Pore size of filter bed (F_p)	

WTR057	Sub	Turbidity (T)	pH
		Total dissolved solids (TDS)	
		Free chlorine (Cl)	
		Pathogens (p)	
		Total and fecal coliform (TC and FC)	
		Conductivity (K)	
		pH	
		Dissolved oxygen (DO)	
		Impurities	
		Salt traces (ST)	
		Flow velocity (V)	
		Age of filter bed (F_a)	
		Pore size of filter bed (F_p)	
WTR058	Sub	Turbidity (T)	Dissolved oxygen (DO)
		Total dissolved solids (TDS)	
		Free chlorine (Cl)	
		Pathogens (p)	
		Total and fecal coliform (TC and FC)	
		Conductivity (K)	
		pH	
		Dissolved oxygen (DO)	
		Impurities	
		Salt traces (ST)	
		Flow velocity (V)	
		Age of filter bed (F_a)	
		Pore size of filter bed (F_p)	

(continued)

Table 19.1 (continued)

Model name	Model type	Input	Output
WTR059	Sub	Turbidity (T)	Impurities
		Total dissolved solids (TDS)	
		Free chlorine (Cl)	
		Pathogens (p)	
		Total and fecal coliform (TC and FC)	
		Conductivity (K)	
		pH	
		Dissolved oxygen (DO)	
		Impurities	
		Salt traces (ST)	
		Flow velocity (V)	
		Age of filter bed (F_a)	
		Pore size of filter bed (F_p)	
WTR510	Sub	Turbidity (T)	Salt traces (ST)
		Total dissolved solids (TDS)	
		Free chlorine (Cl)	
		Pathogens (p)	
		Total and fecal coliform (TC and FC)	
		Conductivity (K)	
		pH	
		Dissolved oxygen (DO)	
		Impurities	
		Salt traces (ST)	
		Flow velocity (V)	
		Age of filter bed (F_a)	
		Pore size of filter bed (F_p)	

WTR511	Sub	Turbidity (T) Total dissolved solids (TDS) Free chlorine (Cl) Pathogens (p) Total and fecal coliform (TC and FC) Conductivity (K) pH Dissolved oxygen (DO) Impurities Salt traces (ST) Flow velocity (V) Age of filter bed (F_a) Pore size of filter bed (F_p)	Flow velocity (V)
Chlorination to Storage Tank			
WTR061	Sub	Turbidity (T) Total dissolved solids (TDS) Free chlorine (Cl) Pathogens (p) Total and fecal coliform (TC and FC) Conductivity (K) pH Dissolved oxygen (DO) Impurities Salt traces (ST)	Turbidity (T)

(continued)

Table 19.1 (continued)

Model name	Model type	Input	Output
WTR062	Sub	Turbidity (*T*)	Total dissolved solids (TDS)
		Total dissolved solids (TDS)	
		Free chlorine (Cl)	
		Pathogens (p)	
		Total and fecal coliform (TC and FC)	
		Conductivity (*K*)	
		pH	
		Dissolved oxygen (DO)	
		Impurities	
		Salt traces (ST)	
WTR063	Sub	Turbidity (*T*)	Free chlorine (Cl)
		Total dissolved solids (TDS)	
		Free chlorine (Cl)	
		Pathogens (p)	
		Total and fecal coliform (TC and FC)	
		Conductivity (*K*)	
		pH	
		Dissolved oxygen (DO)	
		Impurities	
		Salt traces (ST)	

WTR064	Sub	Turbidity (T) Total dissolved solids (TDS) Free chlorine (Cl) Pathogens (p) Total and fecal coliform (TC and FC) Conductivity (K) pH Dissolved oxygen (DO) Impurities Salt traces (ST)	Pathogens (p)
WTR065	Sub	Turbidity (T) Total dissolved solids (TDS) Free chlorine (Cl) Pathogens (p) Total and fecal coliform (TC and FC) Conductivity (K) pH Dissolved oxygen (DO) Impurities Salt traces (ST)	Total and fecal coliform (TC and FC)

(continued)

Table 19.1 (continued)

Model name	Model type	Input	Output
WTR066	Sub	Turbidity (T)	Conductivity (K)
		Total dissolved solids (TDS)	
		Free chlorine (Cl)	
		Pathogens (p)	
		Total and fecal coliform (TC and FC)	
		Conductivity (K)	
		pH	
		Dissolved oxygen (DO)	
		Impurities	
		Salt traces (ST)	
WTR067	Sub	Turbidity (T)	pH
		Total dissolved solids (TDS)	
		Free chlorine (Cl)	
		Pathogens (p)	
		Total and fecal coliform (TC and FC)	
		Conductivity (K)	
		pH	
		Dissolved oxygen (DO)	
		Impurities	
		Salt traces (ST)	

WTR068	Sub	Turbidity (T) Total dissolved solids (TDS) Free chlorine (Cl) Pathogens (p) Total and fecal coliform (TC and FC) Conductivity (K) pH Dissolved oxygen (DO) Impurities Salt traces (ST)	Dissolved oxygen (DO)
WTR069	Sub	Turbidity (T) Total dissolved solids (TDS) Free chlorine (Cl) Pathogens (p) Total and fecal coliform (TC and FC) Conductivity (K) pH Dissolved oxygen (DO) Impurities Salt traces (ST)	Impurities

(continued)

Table 19.1 (continued)

Model name	Model type	Input	Output
WTR610	Sub	Turbidity (*T*)	Salt traces (*ST*)
		Total dissolved solids (TDS)	
		Free chlorine (Cl)	
		Pathogens (p)	
		Total and fecal coliform (TC and FC)	
		Conductivity (*K*)	
		pH	
		Dissolved oxygen (DO)	
		Impurities	
		Salt traces (ST)	
Storage tank to distribution			
WTR071	Sub	Turbidity (*T*)	Water quality index (WQI)
		Total dissolved solids (TDS)	
		Free chlorine (Cl)	
		Pathogens (p)	
		Total and fecal coliform (TC and FC)	
		Conductivity (*K*)	
		pH	
		Dissolved oxygen (DO)	
		Impurities	
		Salt traces (ST)	

Table 19.2 Categories and corresponding ratings of model input variables

Name of variable	Category	Rating
Turbidity (T)	L^{8+}	0.056
	L^{7+}	0.111
	L^{6+}	0.167
	L^{5+}	0.222
	L^{4+}	0.278
	L^{3+}	0.333
	L^{2+}	0.389
	L^{+}	0.444
	M	0.500
	L^{-}	0.556
	L^{2-}	0.611
	L^{3-}	0.667
	L^{4-}	0.722
	L^{5-}	0.778
	L^{6-}	0.833
	L^{7-}	0.889
	L^{8-}	0.944
Total dissolved solids (TDS)	L^{8+}	0.056
	L^{7+}	0.111
	L^{6+}	0.167
	L^{5+}	0.222
	L^{4+}	0.278
	L^{3+}	0.333
	L^{2+}	0.389
	L^{+}	0.444
	M	0.500
	L^{-}	0.556
	L^{2-}	0.611
	L^{3-}	0.667
	L^{4-}	0.722
	L^{5-}	0.778
	L^{6-}	0.833
	L^{7-}	0.889
	L^{8-}	0.944

(continued)

Table 19.2 (continued)

Name of variable	Category	Rating
Free chlorine (Cl)	L^{8+}	0.056
	L^{7+}	0.111
	L^{6+}	0.167
	L^{5+}	0.222
	L^{4+}	0.278
	L^{3+}	0.333
	L^{2+}	0.389
	L^{+}	0.444
	M	0.500
	L^{-}	0.556
	L^{2-}	0.611
	L^{3-}	0.667
	L^{4-}	0.722
	L^{5-}	0.778
	L^{6-}	0.833
	L^{7-}	0.889
	L^{8-}	0.944
Pathogens (p)	L^{8+}	0.056
	L^{7+}	0.111
	L^{6+}	0.167
	L^{5+}	0.222
	L^{4+}	0.278
	L^{3+}	0.333
	L^{2+}	0.389
	L^{+}	0.444
	M	0.500
	L^{-}	0.556
	L^{2-}	0.611
	L^{3-}	0.667
	L^{4-}	0.722
	L^{5-}	0.778
	L^{6-}	0.833
	L^{7-}	0.889
	L^{8-}	0.944

(continued)

Table 19.2 (continued)

Name of variable	Category	Rating
Total and fecal coliform (TC and FC)	L^{8+}	0.056
	L^{7+}	0.111
	L^{6+}	0.167
	L^{5+}	0.222
	L^{4+}	0.278
	L^{3+}	0.333
	L^{2+}	0.389
	L^{+}	0.444
	M	0.500
	L^{-}	0.556
	L^{2-}	0.611
	L^{3-}	0.667
	L^{4-}	0.722
	L^{5-}	0.778
	L^{6-}	0.833
	L^{7-}	0.889
	L^{8-}	0.944
Conductivity (K)	L^{8+}	0.056
	L^{7+}	0.111
	L^{6+}	0.167
	L^{5+}	0.222
	L^{4+}	0.278
	L^{3+}	0.333
	L^{2+}	0.389
	L^{+}	0.444
	M	0.500
	L^{-}	0.556
	L^{2-}	0.611
	L^{3-}	0.667
	L^{4-}	0.722
	L^{5-}	0.778
	L^{6-}	0.833
	L^{7-}	0.889
	L^{8-}	0.944

(continued)

Table 19.2 (continued)

Name of variable	Category	Rating
pH	L^{8+}	0.106
	L^{7+}	0.211
	L^{6+}	0.317
	L^{5+}	0.422
	L^{4+}	0.528
	L^{3+}	0.633
	L^{2+}	0.739
	L^{+}	0.844
	M	0.950
	L^{-}	0.844
	L^{2-}	0.739
	L^{3-}	0.633
	L^{4-}	0.528
	L^{5-}	0.422
	L^{6-}	0.317
	L^{7-}	0.211
	L^{8-}	0.106
Dissolved oxygen (DO)	L^{8+}	0.056
	L^{7+}	0.111
	L^{6+}	0.167
	L^{5+}	0.222
	L^{4+}	0.278
	L^{3+}	0.333
	L^{2+}	0.389
	L^{+}	0.444
	M	0.500
	L^{-}	0.556
	L^{2-}	0.611
	L^{3-}	0.667
	L^{4-}	0.722
	L^{5-}	0.778
	L^{6-}	0.833
	L^{7-}	0.889
	L^{8-}	0.944

(continued)

Table 19.2 (continued)

Name of variable	Category	Rating
Impurities	L^{8+}	0.056
	L^{7+}	0.111
	L^{6+}	0.167
	L^{5+}	0.222
	L^{4+}	0.278
	L^{3+}	0.333
	L^{2+}	0.389
	L^{+}	0.444
	M	0.500
	L^{-}	0.556
	L^{2-}	0.611
	L^{3-}	0.667
	L^{4-}	0.722
	L^{5-}	0.778
	L^{6-}	0.833
	L^{7-}	0.889
	L^{8-}	0.944
Salt traces (ST)	L^{8+}	0.056
	L^{7+}	0.111
	L^{6+}	0.167
	L^{5+}	0.222
	L^{4+}	0.278
	L^{3+}	0.333
	L^{2+}	0.389
	L^{+}	0.444
	M	0.500
	L^{-}	0.556
	L^{2-}	0.611
	L^{3-}	0.667
	L^{4-}	0.722
	L^{5-}	0.778
	L^{6-}	0.833
	L^{7-}	0.889
	L^{8-}	0.944

(continued)

Table 19.2 (continued)

Name of variable	Category	Rating
Flow velocity (*V*)	L^{8+}	0.106
	L^{7+}	0.211
	L^{6+}	0.317
	L^{5+}	0.422
	L^{4+}	0.528
	L^{3+}	0.633
	L^{2+}	0.739
	L^{+}	0.844
	M	0.950
	L^{-}	0.844
	L^{2-}	0.739
	L^{3-}	0.633
	L^{4-}	0.528
	L^{5-}	0.422
	L^{6-}	0.317
	L^{7-}	0.211
	L^{8-}	0.106
Salt dosing (*S*)	L^{8+}	0.106
	L^{7+}	0.211
	L^{6+}	0.317
	L^{5+}	0.422
	L^{4+}	0.528
	L^{3+}	0.633
	L^{2+}	0.739
	L^{+}	0.844
	M	0.950
	L^{-}	0.844
	L^{2-}	0.739
	L^{3-}	0.633
	L^{4-}	0.528
	L^{5-}	0.422
	L^{6-}	0.317
	L^{7-}	0.211
	L^{8-}	0.106

(continued)

Table 19.2 (continued)

Name of variable	Category	Rating
Lime dosing (LD)	L^{8+}	0.106
	L^{7+}	0.211
	L^{6+}	0.317
	L^{5+}	0.422
	L^{4+}	0.528
	L^{3+}	0.633
	L^{2+}	0.739
	L^{+}	0.844
	M	0.950
	L^{-}	0.844
	L^{2-}	0.739
	L^{3-}	0.633
	L^{4-}	0.528
	L^{5-}	0.422
	L^{6-}	0.317
	L^{7-}	0.211
	L^{8-}	0.106
Age of filter bed	L^{8+}	0.056
	L^{7+}	0.111
	L^{6+}	0.167
	L^{5+}	0.222
	L^{4+}	0.278
	L^{3+}	0.333
	L^{2+}	0.389
	L^{+}	0.444
	M	0.500
	L^{-}	0.556
	L^{2-}	0.611
	L^{3-}	0.667
	L^{4-}	0.722
	L^{5-}	0.778
	L^{6-}	0.833
	L^{7-}	0.889
	L^{8-}	0.944

(continued)

Table 19.2 (continued)

Name of variable	Category	Rating
Pore size of filter bed	L^{8+}	0.106
	L^{7+}	0.211
	L^{6+}	0.317
	L^{5+}	0.422
	L^{4+}	0.528
	L^{3+}	0.633
	L^{2+}	0.739
	L^{+}	0.844
	M	0.950
	L^{-}	0.844
	L^{2-}	0.739
	L^{3-}	0.633
	L^{4-}	0.528
	L^{5-}	0.422
	L^{6-}	0.317
	L^{7-}	0.211
	L^{8-}	0.106

L large, *M* medium, $^{+}$ increment, $^{-}$ decrement

The maximum and minimum data values of the input variables were first identified, and the difference between the two values was determined. The difference was then divided into 17 equal divisions, and each category was encoded in one uniform division whereby the initial division starting from the minimum value was assigned to the L^{8-} group and the last division near the maximum value was assigned to the L^{8+} group. The probability of each group was then used as the rating of the individual group whereby the relationship with the overall quality of treated water was represented by modifying the probability value such that the highest probability could be assigned to the most ideal value of the variable in regard to the quality of the treated water.

The combinatorial data matrix was fed to the neurogenetic submodels from a water quality transiency representation (WQTR) of 01X to 07X, where X stands for the number of variables out of the total number of variables considered for the study. The main model was called a quality transiency representation of a WTP (QTRWTP).

Feedforward neural network models were prepared. The network topology was determined with the help of a genetic algorithm utilizing the logistic activation function. All the models were trained with the Levenberg-Marquadart (LM) algorithm, and a logistic function was selected as the activation function.

A water quality index was determined with the help of the NSF water quality index. The index was again encoded into 17 groups following a methodology similar to that followed with the input variables.

19.2.1 Development of the QTRWTP

All the sub models were interconnected with the help of Spreadsheet Converter 6, where the backend programming was prepared by simple Microsoft Excel functions.

The model thus prepared were embedded into Excel, where the SC V6 converted the models into a standalone Javascript program.

The standalone program enables users to input the quality parameters based on the raw water and predict the quality of the treated water from one window.

The software can also be used to identify the quality transiency between the different stages of the water treatment process conducted in the WTP.

The number of submodels will vary based on the type of WTP and the use of the treated water. Custom submodels can be added if required.

19.3 Benefits of the Software

This software can be used in real-time prediction of the quality of treated water. The impact of extreme events or changes in climatic patterns, as well as rapid urbanization, could be simulated and the output analyzed in order to make an informed decision about how to mitigate the effects of climate change and population growth.

The software could be used for scenario analysis, variable sensitivity, or identification of optimal values that could help plant engineers to adjust their dosing or maintain machinery so that optimal output could be achieved from the plant.

If the software were attached to a distributed control system (DCS) for monitoring of the different stages of the water treatment process, then an automatic optimization process could be introduced to reduce expenses related to manual error. The real-time input generation and feeding to the software could enable the DCS to predict and act according to the prediction so that an adaptive compensatory mechanism could be implemented to reduce the probability of plant failures.

The software could also be used for the monitoring of raw and treated water so that advance planning of dosing patterns could be undertaken, which would allow for mitigation measurements on the existing machines and a methodology by which the plant may run on a reduced capacity factor.

The load factor or the ability of plants to mitigate demand could also be monitored in real time and adjusted according to variations in demand.

Overall the software could be an essential and useful tool for engineers and workers at WTPs that could help them prepare well in advance of future uncertainties.

19.4 Drawback of the Software

The major drawback of the software is the requirement of a high-end computational facility to allow the seven neurogenetic models to work simultaneously. The neural network models use a noticeable amount of computational power to complete the learning process from the available dataset of the problem. Thus the model will enforce any user to provide adequate computational infrastructure so that it can run and generate results smoothly.

But if the training of the models were performed separately on the client's end, the aforementioned drawback could be compensated.

19.5 Conclusion

The present investigation introduced a software for monitoring and simulating WTPs. The software would enable engineers at a WTP to monitor and optimize the plant's daily operations based on predictions given by the software with the help of its eight cascading neurogenetic models. The software could be used to simulate the impacts of climate change, urbanization, and different hazards that may arise in the day-to-day operations of a WTP. The basic methodology of the software is to predict the outcome of a quality index of treated water based on the input given by the user in the home panel. In the backend of the software new neurogenetic models are running that can be used to predict outputs in all, including the final, stages of the water treatment process. The requirement of a high-end computational facility was found to be a major hindrance for wide-scale distribution of the software. The model could be a useful addition for engineers in the regulation of daily load factors. In the future, additional features like introduction of other nature-based or statistical methods for prediction may be introduced.

Chapter 20
Development of a Neuro-Fuzzy System for Selection of Tree Species for Afforestation Purpose

Mrinmoy Majumder, Tilottama Chackraborty, Santanu Datta, Rajesh Chakraborty, and Rabindra Nath Barman

Abstract Climatic uncertainty due to global warming and continuous deforestation as a result of urbanization has contributed to the degradation of regions once ecologically rich but now becoming arid or desert. In various studies and governmental reports it has been proposed that places that have been desertified or are prone to desertification can be saved and returned their original state if artificial reforestation is undertaken with selected species. Although afforestation has been successful in many desertified regions of the world (India, Philippines, Australia), in many places it has come to represent a waste of money and the environment. This failure has been attributed to the selection of plants of poor quality, poor plant handling, out-of-season planting, insects, and other operational and climatic factors. Thus, selection of ideal plant species is important for the success of an afforestation project. In this study a methodology for the selection of species for afforestation is proposed with the help of neuro-fuzzy techniques. The novelty of the study lies in its attempt to apply neuro-fuzzy techniques in the selection of afforestation species, thus making the selection free of bias and prejudices. By *selection of species* we mean the suitability of a species that can be used in afforestation projects with no chance of failure.

Keywords Afforestation • Neuro-fuzzy systems • Desertification

M. Majumder (✉) • T. Chackraborty
School of Hydro-Informatics Engineering, National Institute of Technology Agartala, Barjala, Jirania 799055, Tripura, India
e-mail: mmajumder15@gmail.com; tilottama86@gmail.com

S. Datta • R. Chakraborty
PG Student
School of Hydro-Informatics Engineering, National Institute of Technology Agartala, Barjala, Jirania 799055, Tripura, India

R.N. Barman
Department of Production Engineering, National Institute of Technology Agartala, Barjala, Jirania 799055, Tripura, India
e-mail: rahul.barman@yahoo.co.in

M. Majumder and R.N. Barman (eds.), *Application of Nature Based Algorithm in Natural Resource Management*, DOI 10.1007/978-94-007-5152-1_20,

20.1 Introduction

Afforestation is defined by the United Nations Framework Convention on Climate Change as "direct human-induced conversion of non-forested land to forested land through planting, seeding and/or the human-induced promotion of natural seed sources." The term *afforestation* is similar to reforestation, but there is a difference between the two. Where afforestation can take place on land that was barren for last 50 years reforestation can be performed on land that was a forest but later converted for some other use.

Selection of a site and species of an afforestation project is an important step for the success of the venture. If a site is selected without a proper feasibility study, then the afforestation project is bound to create problems rather than benefits. Similarly, selection of species for afforestation must be undertaken with proper analysis and literature review. An unsuitable species can ruin an afforestation project, waste money, and provoke grievances among local inhabitants who are an essential part of any afforestation project.

In any afforestation project the local inhabitants should be encouraged to invest money in culturing the afforested species. Once the species reach maturity, the trees were sold to the agency managing the project and new plants were planted for cultivation. Thus a sustainable forest cover and earning opportunity for the local people can be provided.

The main aim of afforestation projects are to enrich the carbon sequestration potential of the region, whether arid or desertified. Other benefits of afforestation projects include an increase in the water-holding capacity of catchments, reduced runoff, and enhancement of biodiversity. If the tree species which are selected for the afforestation for cultivation are not suitable in the regions such as, if they are prone to insects attacks or may used as food for domestic animals or they required substantial amount of care, then the spices may not be recommended for the afforestation activities.

But many tree species, while needing substantial amounts of care and having many other drawbacks such as those just mentioned, have been cultivated in afforestation projects for increased profitability to the local people. Such species may be highly beneficial for the local people for economic growth but may not be suitable for afforestation projects if they cannot act as carbon sinks or withstand the common calamities of a forest. Thus this study will try to identify a methodology to select species for afforestation with respect to climate, soil characteristics, salinity of the project area, water retention capacity, erosion prevention, and carbon sequestration abilities of the species and requirements of the local people.

20.1.1 Desertification

"Desertification is the degradation of land in arid, semi-arid, and dry sub-humid areas. It is caused primarily by human activities and climatic variations. Desertification does not refer to the expansion of existing deserts. It occurs because dryland ecosystems, which cover over one third of the world's land area, are extremely vulnerable to over-exploitation and inappropriate land use" (FAO).

Broadly, the basic causes of desertification of an arid region can be attributed to

1. Poverty;
2. Political instability;
3. Deforestation, which includes

   logging and illegal felling,
   forest fires,
   unsustainable water management, and
   industrial and mining activities;

4. Overgrazing;
5. Bad irrigation practices, including

   extensive cultivation of one crop,
   use of chemical fertilizers and pesticides, and
   shifting cultivation without adequate period of recovery.

According to the United Nations Convention to Combat Desertification, 250 million people, mainly in Africa and Asia, are directly affected by the rapid degradation of dryland ecosystems and some one billion people in more than 100 countries are living in desert-prone areas of the "world's poorest, most marginalized, and politically weak" countries.

In the recent past, the loss of vegetation was very extensive once fertile lands and water became rare commodities. Due to the failure of monsoons, the soil and land in many regions of the world have been adversely affected and in some places have become totally infertile.

There is an urgent need to regenerate once rich ecosystems by taking mitigation measures. For example, in India, many ravines and ridges have been treated by constructing gully plugs and digging out contour trenches, earth check dams with spillways on either side have been built to allow water overflow to run off without damaging dams, and ponds have been constructed to hold larger volumes of water.

Among farmers who have witnessed the impact of soil erosion and lack of water availability on agricultural productivity, "the concept of watershed management gained credence" (Caritas India). Many farmers became convinced that the local environment could be reestablished through a well-managed watershed management program.

Although India's land area is only 2.4% of the world's total, 16.67% of the world's population and 18% of its livestock reside on the subcontinent. The pressures from this rapidly growing population and its requirements for food supplies have played a major role in promoting desertification.

The rapid urbanization and uncontrolled growth of industry have increased the stress on natural resources in many ways, which may lead to permanent loss of vegetation and plant species and will ultimately convert large areas into wastelands, which would contribute to the frequent occurrence of natural disasters.

"Half the land in India is now affected by desertification and this impairs the ability of land to support life" (Caritas India).

The impact of desertification is a multilevel, global phenomenon. The desertification of a land will induce

1. Land degradation,
2. Vulnerability to food shortages,
3. Natural disasters,
4. Depletion of natural resources,
5. Deterioration of the environment.

Vegetation plays an essential role in protecting the soil, and especially trees and shrubs, because trees' and shrubs' long life and capacity to develop powerful root systems assure protection against soil erosion. Their disappearance can considerably increase the vulnerability of the land to turn into a wasteland.

20.1.2 Importance of Afforestation Projects

The benefits of afforestation are immense because rapid urbanization and increased stress on agro products have led to the degradation of much forest land to provide space for agriculture and industry.

Due to deforestation, trees in water catchments and riverside zones become depleted. That is where the importance of afforestation can be demonstrated. Natural forests are heterogeneous, and the supply of wood from such regions varies due to their sensitivity to overusage and slow regeneration. Such forests cannot be used continuously for commercial purposes. The process of artificial forestation in barren or empty lands helps to promote the fast propagation of specific types of trees for the wood industry on one hand and on the other ensures the protection of soil, water, and natural ecosystems of the region.

That is why many countries have introduced the practice of afforestation along with agriculture even on croplands that are beneficial due to the steady demand of wood from the construction industry and local fuel markets.

The benefits of this practice, which is called agroforestry, are as follows:

Socioeconomic benefits:

It provides a supply of timber, fruit, and fodder for cattle as well as crop production and thus ensures the agroproductivity of the region.

Environmental:

It prevents soil erosion, enables better retention of water, and shields crops from excessive wind and sun damage; in addition, it regulates atmospheric carbon dioxide; large-scale afforestation can act as carbon sink and compensate for the release of greenhouse gases from the burning of fossil fuels and industrialization.

20.1.3 Drawbacks of Afforestation Projects

Every method has its drawbacks, and afforestation is no exception. In many regions, afforestation projects failed miserably leading to controversies and large-scale wasting of money. The failure of afforestation projects can be attributed to the following causes:

- Poor plant quality
- Poor plant handling,
- Planting out of season (northwest experience),
- Drought,
- Insect(s),
- Operational factors,
- Error in site selection,
- Specific problems with certain species or stock types.

20.1.4 Suitable Species for Afforestation

Not all plant species can be utilized in artificial forestry projects. The selection of plant species will depend on soil texture, soil moisture, soil salinity, rainfall, temperature, and the elevation of the site selected for afforestation. Tables 20.1 and 20.2 show the plant families that have been found to be the most suitable for given climatic, geophysical, and socioeconomic conditions.

20.1.5 Objective of Present Study

The main objective of the present study is to develop a model to estimate the most suitable species for afforestation for given site conditions and the requirements of the dependent population.

The study will also verify the potential of neuro-fuzzy techniques in solving such prediction requirements, which are extremely important for the success of afforestation projects.

The investigation aims to establish a methodology for natural resource managers to propose an afforestation project based on futuristic but reliable feasibility studies.

20.1.6 Brief Overview of Study Methodology

In the present investigation a neuro-fuzzy model was developed with the help of selected factors and the output decision, which facilitates selection of the most appropriate species for afforestation projects given certain site conditions.

The input variables for the model are as follows:

1. Precipitation (*P*)
2. Temperature (*T*)
3. Elevation (*E*)
4. Slope (*S*)
5. Soil salinity (SS)

Table 20.1 Geophysical and climatic attributes for popular species of afforestation

Plant genus	Soil texture	Soil salinity	Rainfall (mm/year)	Temperature (°C)	Soil moisture	Elevation	Common name
Shorea sp.	Clay (rigid and tough soil).	Low	Low	Moderate	Low	Very high (1,700 m)	*Sal*
Artocarpus sp.	Silty loam (the jackfruit tree flourishes in rich, deep soil or medium or open texture)	None	Moderate to high	Low to moderate (13 ± 1)	Low to moderate	Very low (sea level)	*Jackfruit*
Eucalyptus sp.	Sandy clay (soil texture somewhat loose)	Moderate	Moderate	Moderate (23–26)	Low to moderate	Low to high	*Eucalyptus*
Bambusa sp.	Sandy clay (sand particles are largest clay particles, smallest and silt are somewhat medium sized)	Moderate	Moderate to high	Low to moderate (not below 8)	Moderate	Low to high	*Bamboo*
Tectona sp.	Sandy (pH should not be below 6.5)	Moderate	None to moderate (1,000–1,700)	Moderate (27)	Low to high	Moderate	*Teak*
Mangifera sp.	Sand, loam, or loam with 15.25 % Clay	Low to medium (1–6 days/m)	Low to moderate (not less than 85–140)	Low to moderate (10–12)	High	Low to moderate (below 400 m)	*Mango*
Cocus sp.	Sandy soil	Moderate to high	Moderate to high (1,000–2,250)	Low to moderate (not below –6.66)	Moderate	High to very high(600–900 m)	*Coconut*
Swietenia sp.	Sand (well-drained deep rich sandy soil)	Moderate	Moderate to high (not below 500)	Moderate (16–32)	Low to moderate	Very low to medium (100–500 m)	*Mahogany*

Note: retrieved from World Agro-Forestry Centre

Table 20.2 Attributes of some afforestation genus represented by fuzzy categories

Plant genus	Water-holding capacity	Erosion prevention ability	Income generation potential	Carbon sequestration capacity
Shorea sp.	Low	High	High	High
Artocarpus sp.	Moderate	Moderate	Moderate	Moderate
Eucalyptus sp.	Moderate	High	Moderate	High
Bambusa sp.	High	High	High	Moderate
Tectona sp.	High	High	High	High
Mangifera sp.	Low	Moderate	Moderate	Low
Cocus sp.	Moderate	High	Moderate	Low
Swietenia sp.	Low	Moderate	High	Moderate

Note: retrieved from World Agro-Forestry Centre

6. Soil texture (STe)
7. Required water holding capacity (WHCR)
8. Required soil stability (to prevent erosion) (SStR)
9. Required income generation capability of project (IGCR)
10. Required carbon sequestration capacity (CSCR)

The output variable of the models is the

1. Family of the species for afforestation (*O*).

Each of the input variables was divided into nine categories according to their intensity of quality or quantity. The output variable was divided into seven families of plant species that are suitable for different climatic, geophysical, and socioeconomic demands of the afforestation region.

The weight of each input variable was determined with the help of fuzzy logic, and an objective function is prepared in such a way that its value will determine the most suitable family of plant species for artificial plantation in a region.

20.2 Neuro-Fuzzy Technique

Neuro-fuzzy refers to combinations of artificial neural networks (ANNs) and fuzzy logic and was proposed by J. S. R. Jang.

Neuro-fuzzy hybridization results in a hybrid intelligent system that synergizes these two techniques by combining the humanlike reasoning style of fuzzy systems with the learning and connectionist structure of neural networks.

Neuro-fuzzy systems integrate the humanlike reasoning style of fuzzy systems through the use of fuzzy sets and a linguistic model consisting of a set of IF-THEN fuzzy rules. The main strength of neuro-fuzzy systems is that they are universal approximators with the ability to solicit interpretable IF-THEN rules.

Interpretability versus accuracy is the strength of neuro-fuzzy systems, but in practice, one of the two properties prevails.

The neuro-fuzzy in fuzzy modeling research is divided into two areas: linguistic fuzzy modeling that is focused on interpretability, mainly the Mamdani model, and precise fuzzy modeling that is focused on accuracy, mainly the Takagi-Sugeno-Kang (TSK) model.

20.2.1 Applications of Neuro-Fuzzy Techniques in Related Fields

Neuro-fuzzy systems, whether separately or in conjunction with each other, have been applied in various problem-solving approaches, and the results from the performance metrics recommends their wider application (Table 20.3).

20.3 Methodology

The development of the neuro-fuzzy model starts with the determination of the most influential input variable from the ten considered inputs. This determination was performed by the application of fuzzy theory of maximization.

Table 20.4 shows the degree of ratings received by each input variable with respect to the others. The degrees considered for the rating of a parameter are as follows:

1. Highest degree of influence: Very High Influence (VHI)
2. High degree of influence but not as high as No.1: High Influence (HI)
3. Neither high nor low degree of influence: Neutral (N)
4. Low degree of influence: Low Influence (LI)
5. Lowest degree of influence: Very Low Influence (VLI)

The degree of influence of each of the parameters was assumed from the information received from the various scientific and governmental studies.

After the ratings were assigned, they were replaced by corresponding ranks. For example, the VHI rating received a rank of 1, whereas VLI gets a rank of 5. Accordingly, all the ratings were replaced (Table 20.5).

According to the theory of maximization, ranks received by a variable with respect to the other input variables were divided by the worst rank or the maximum rank and the minimum dividend found from all the calculations was separated out as the weight of the input parameter. The value of the weights is actually the best rank received by an input variable with respect to the others divided by the worst rank received by that variable. If the average value of the variables was determined instead of the minimum value, then a lower value would always indicate higher importance and vice versa. But if weight were represented by smaller numerical values, then the true representation of importance would not be reflected in the objective function. That is why all the values were inversed and used as the weight of a given variable in the objective function.

The variables were summed after being multiplied by the weight and the sum was divided by the sum of the weights to obtain the value of the objective function.

After the weights of all input variables were determined, the values of the objective function for the eight most selected species were calculated for the estimation of the benchmark. The weight of the objective function was estimated in terms of their influence on the selection of a species for afforestation. That is why the greater

Table 20.3 Table depicting some of the recent applications of neuro-fuzzy techniques in different problem solving approaches

References	Study area (name or properties)	Name of neuro-fuzzy model	Objective	Remarks
Srivastava et al. (2010)	India	Fuzzy c-means clustering and artificial neural network (ANN)	Identify significant input variables for rainfall forecasting. Variables were generated from (1) two climate indices, i.e., southern oscillation index (SOI) and Pacific decadal oscillation index (PDOI); (2) sea surface temperature anomalies (SSTa) in the 5°×5° grid points in Indian Ocean; and (3) both climate indices and SSTa	Neural network was utilized to identify the relationship between input variables and rainfall and fuzzy c-means algorithm was used to divide the data into representative groups for training, testing, and validation
Yildirim and Bayramoglu (2006)	Urban landscape	Adaptive neuro-fuzzy inference system	Estimate total suspended particle (TSP) and sulfur dioxide from different related meteorological variables	The model forecasting accuracy was found to be equal to 75–90% (for SO_2) and 69–80% (for TSP)
Cay and Iscan (2011)	Turkey	Fuzzy logic	Compare interview-based and fuzzy-logic-based land reallocation model in land consolidation projects of country	Fuzzy-logic-based land reallocation model was preferred by survey population
Oh and Pradhan (2011)	Tropical hilly area	Adaptive neuro-fuzzy inference system	Create landslide-susceptibility map with help of altitude, slope angle, plan curvature, distance to drainage, distance to road, soil texture, and stream power index (SPI)	In case of triangular, trapezoidal, generalized bell, and polynomial membership functions the ANFIS model was found to produce 84.39% accurate results
Foody (1996)		Maximum likelihood estimation, fuzzy set, and ANN used separately	Map vegetation to soften image clustering to minimize error due to rigid classification	Increases accuracy

(continued)

Table 20.3 (continued)

References	Study area (name or properties)	Name of neuro-fuzzy model	Objective	Remarks
Iliadis and Tachos (2011)	Mountainous watershed of Cyprus	Fuzzy weighted support vector regression with fuzzy partition model	Deduce an efficient water management model for mountainous watershed, i.e., estimate supply of water with respect to different scenarios of demand	The model was compared with an ANN-based support vector regression model, and the former was found to be better than the latter model
Şişman-Yılmaz et al. (2004)		ANFIS-unfolded-in-time (temporal rule based on Takagi-Sugeno-Kang and training by back-propagation algorithm)	Gas furnace data with lagged input; model maintains temporal relationship between input variables	"Experimental results show that the proposed model achieves online learning and prediction on temporal data"

the value of the objective function, the more suitable will a species be for afforestation projects. The benchmarking species determines the degree of suitability. For example, eucalyptus is considered a highly suitable species for afforestation due to its carbon sequestration capacity, income generation ability, water-holding efficiency, and capability to stabilize soil structure. But some species are found to be highly efficient in one parameter and inefficient in others.

The suitability of the eight considered species for afforestation is already known, but which species is better than the others is not known. That is why at first the separation was made with the help of the developed objective function.

With the help of this validation process, any species can now be classified in terms of different degrees of suitability. The prediction of suitability for any species was made through the application of ANNs capable of identifying nonlinear relationships between any variables if properly processed.

A feedforward, Levenberg–Marquardt (LM) algorithm trained with a logistic activation function was used to develop the prediction model. The topology was searched with the help of a genetic algorithm. The model was tested with known species. Once the model was tested successfully, it was applied for the prediction of the suitability of a species for afforestation purposes.

The correct classification rate and kappa coefficient of agreement were used as the performance metrics of the neural network model. The result of the performance metrics represents the accuracy and reliability of the neuro-fuzzy model.

Once the neural network model was developed satisfactorily, the output of the species selected for afforestation would indicate the suitability of the species for afforestation according to the site and requirements of the dependent population.

20.4 Results and Discussion

Tables 20.4, 20.5 and 20.6 showed the degree of influence, rank, and weight of each input variable as derived from the fuzzy theory of maximization.

The values of the performance metrics for the neural network model are given in Table 20.7. According to the metrics, the LM-trained ANN model was trained successfully with 83.62% degree of agreement (Table 20.8).

Table 20.4 Degree of rating assigned to each input parameter with respect to other parameters

	(*P*)	(*T*)	(*E*)	(*S*)	(SS)	(STe)	(WHCR)	(SStR)	(IGCR)	(CSCR)
(*P*)	0	N	VHI	VHI	HI	HI	N	VHI	HI	N
(*T*)	N	0	VHI	VHI	HI	HI	HI	HI	N	LI
(*E*)	VLI	VLI	0	HI	N	N	N	N	LI	LI
(*S*)	VLI	VLI	LI	0	N	N	HI	LI	LI	LI
(SS)	LI	LI	N	N	0	N	HI	N	N	N
(STe)	LI	LI	N	N	N	0	N	LI	LI	LI
(WHCR)	N	LI	N	LI	LI	N	0	N	N	N
(SStR)	VLI	LI	N	HI	N	HI	N	0	LI	N
(IGCR)	HI	N	HI	HI	N	HI	N	HI	0	N
(CSCR)	N	HI	HI	HI	N	HI	N	N	N	0

Table 20.5 Rank assigned to each input parameter with respect to other parameters in light of rating received (Table 20.4)

	(*P*)	(*T*)	(*E*)	(*S*)	(SS)	(STe)	(WHCR)	(SStR)	(IGCR)	(CSCR)
(*P*)	0	3	1	1	2	2	3	1	2	3
(*T*)	3	0	1	1	2	2	2	2	3	4
(*E*)	5	5	0	2	3	3	3	3	4	4
(*S*)	5	5	4	0	3	3	2	4	4	4
(SS)	4	4	3	3	0	3	2	3	3	3
(STe)	4	4	3	3	3	0	3	4	4	4
(WHCR)	3	4	3	4	4	3	0	3	3	3
(SStR)	5	4	3	2	3	2	3	0	4	3
(IGCR)	2	3	2	2	3	2	3	2	0	3
(CSCR)	3	2	2	2	3	2	3	3	3	0

Table 20.6 Weight of each input variable as determined from fuzzy theory of maximization approach (Tables 20.4 and 20.5)

	(*P*)	(*T*)	(*E*)	(*S*)	(SS)	(STe)	(WHCR)	(SStR)	(IGCR)	(CSCR)	Weightage
(*P*)	0.000	1.000	0.333	0.333	0.667	0.667	1.000	0.333	0.667	1.000	0.667
(*T*)	0.750	0.000	0.250	0.250	0.500	0.500	0.500	0.500	0.750	1.000	0.750
(*E*)	1.000	1.000	0.000	0.400	0.600	0.600	0.600	0.600	0.800	0.800	0.600
(*S*)	1.000	1.000	0.800	0.000	0.600	0.600	0.400	0.800	0.800	0.800	0.600
(SS)	1.000	1.000	0.750	0.750	0.000	0.750	0.500	0.750	0.750	0.750	0.500
(STe)	1.000	1.000	0.750	0.750	0.750	0.000	0.750	1.000	1.000	1.000	0.250
(WHCR)	0.750	1.000	0.750	1.000	1.000	0.750	0.000	0.750	0.750	0.750	0.250
(SStR)	1.000	0.800	0.600	0.400	0.600	0.400	0.600	0.000	0.800	0.600	0.600
(IGCR)	0.667	1.000	0.667	0.667	1.000	0.667	1.000	0.667	0.000	1.000	0.333
(CSCR)	1.000	0.667	0.667	0.667	1.000	0.667	1.000	1.000	1.000	0.000	0.333

Table 20.7 Performance metrics achieved from LM-trained ANN model with fuzzy weighted inputs

Performance metric	Result achieved from ANNLM model
Kappa coefficient of agreement	83.62%
Correct classification rate	92.22%
Actual agreement	92.22%
Agreement by chance	52.52%
Training CCR	95.16%
Testing CCR	85.71%
Network topology	10-2-1

Table 20.8 Predicted and actual suitability of species generally selected for afforestation

	P	T	E	S	SS	STe	WHCR	SStR	IGCR	CSCR	Predicted suitability	Actual suitability
Shorea sp.	L	M	M	L	M	Sand	H	H	H	H	SL	M
Artocarpus sp.	VH	SL	L	SL	M	Sandy-clay	H	H	H	M	SL	SL
Eucalyptus sp.	M	M	H	SH	M	Sandy-clay	M	H	M	H	SL	H
Bambusa sp.	H	L	VL	VL	EL	Silty-loam	M	M	M	M	SL	SL
Tectona sp.	L	M	VH	H	L	Clay	L	H	H	H	SL	SL

The suitability of five of the eight selected species was predicted from the model, but two out of five times the model failed to predict the suitability correctly. The actual suitability of the species was derived from earlier studies and governmental reports. The categorization of the inputs and requirements of more intensive training may be the reason why the model was found to have a 60% accuracy with the unknown data (data not included in the training data set).

20.5 Conclusion

The present study developed a neuro-fuzzy model for the determination of the suitability of species for afforestation purposes. The study used fuzzy theory of maximization to determine the weight of the input variables, which were later used to create an objective function or species suitability index. The nonlinear relationship between the ten input and one output variables was trained with the help of a neural network. The kappa coefficient of agreement and correct classification rate were used as the performance validator of the developed model. An agreement of 83.62% and classification accuracy of 92.22% was achieved from the model, but when it was tested with a data set of some species that were widely used for afforestation, the model performance came down to a mere 60%.The lack of training and error in the classification of the variables probably explain why the model failed. But the methodology can be utilized with other species of plants so that only those species that are very suitable for afforestation may be selected, which would thereby reduce the chances of failure of such projects.

References

Caritas India, Desertification in India. Retrieved from http://www.caritas.org/activities/climate_change/www.caritasindia.org/. Feb 2012

Cay T, Iscan F (2011) Fuzzy expert system for land reallocation in land consolidation. Expert Syst Appl 38(9):11055–11071

Food and Agricultural Organizations of United Nations(FAO), Desertification. Retrieved from http://www.fao.org/desertification/intro_txt/en/desert.htm. Feb 2012

Foody GM (1996) Fuzzy modelling of vegetation from remotely sensed imagery. Ecol Model 85(1):3–12

Iliadis L, Tachos S (2011) Soft computing techniques toward modeling the water supplies of Cyprus. Neural Netw 24(8):836–841

Oh HJ, Pradhan B (2011) Application of a neuro-fuzzy model to landslide-susceptibility mapping for shallow landslides in a tropical hilly area. Comput Geosci 37(9):1264–1276

i man-Yılmaz NA, Alpaslan N, Jain L (2004) ANFIS_unfolded_in_time for multivariate time series forecasting. Neurocomputing 61:139–168

Srivastava G, Panda SN, Mondal P (2010) Forecasting of rainfall using ocean-atmospheric indices with a fuzzy neural technique. J Hydrol 395(3–4):190–198

World Agro-Forestry Centre. Retrieved from http://www.worldagroforestry.org. Feb 2012

Yildirim Y, Bayramoglu M (2006) Adaptive neuro-fuzzy based modelling for prediction of air pollution daily levels in city of Zonguldak. Chemosphere 63(9):1575–1582

About the Authors

Principal Authors:

Dr. Mrinmoy Majumder presently works as Assistant Professor in the School of Hydro-Informatics Engineering at the National Institute of Technology, Agartala, India. He received his Ph.D. in 2010 from Jadavpur University. Presently he teaches and conducts research on hydroinformatics, natural resource management, and nature-based algorithms. He has published more than 25 papers in national and international journals and has written three books.
Address: 43/6/1/1, Jheelroad, Kolkata-700031, West Bengal, India. **Email**: mmajumder15@gmail.com

Dr. Rabindra Nath Barman presently works as Assistant Professor in the Department of Production Engineering at the National Institute of Technology, Agartala, India. He received his Ph.D. from Jadavpur University in 2012. He is interested in fluid mechanics and water resource principles. He has published more than 25 papers in national and international journals and has written several book chapters.
Address: Department of Production Engineering, National Institute of Technology Agartala, Barjala, Jirania 799055, Tripura, India. **Email**: rahul.barman@yahoo.co.in

Coauthors:

Miss Tilottoma Chackraborty Tilottoma Chackraborty presently works as Assistant Professor in the School of Hydro-Informatics Engineering at the National Institute of Technology Agartala, India. She completed her M. Tech in 2010 in the School of Water Resource Engineering, Jadavpur University. She currently teaches and conducts research on hydraulic structures, water resource management, and hydrologic modeling. She has published several book chapters.

M. Majumder and R.N. Barman (eds.), *Application of Nature Based Algorithm in Natural Resource Management*, DOI 10.1007/978-94-007-5152-1,

Address: School of Hydro-Informatics Engineering, National Institute of Technology Agartala, Barjala, Jirania 799055, Tripura, India. **Email**: tilottama86@gmail.com

Miss Paulami De presently works as Senior Research Fellow in a project funded by the Department of Science and Technology, Government of India, and was accepted into the School of Hydro-Informatics Engineering in the National Institute of Technology, Agartala, India. She completed her M. Tech in 2011 in the School of Water Resource Engineering, Jadavpur University. She is currently engaged in research on waste and surface water treatment mechanisms.
Address: School of Hydro-Informatics Engineering, National Institute of Technology Agartala, Barjala, Jirania 799055, Tripura, India. **Email**: depaulami@yahoo.co.in

Mr. Soumya Ghosh presently works as a Lecturer at the JAYPEE Institute of Technology and Management, India. He completed his Master of Technology at the School of Water Resource Engineering, Jadavpur University, in 2010. He currently teaches and conducts research on renewable energy, tidal power, and electrical machines.
Address: School of Hydro-Informatics Engineering, National Institute of Technology Agartala, Barjala, Jirania 799055, Tripura, India. **Email**: soumyaee@gmail.com

Mr. Bipal Kr. Jana presently works in consulting engineering services in Kolkata, India. He has published more than ten papers in various national and international journals and has written one book in the field of environmental science and carbon sequestration.
Address: Consulting Engineering Services, “AKARIK”, East Tentulberia, P.O. – Panchpota, (Near Five Star Club), Kolkata 700152, West Bengal, India. **Email**: bipalkjana@gmail.com

Dr. Debasri Roy presently works as Associate Professor in the School of Water Resource Engineering, Jadavpur University, India. She currently teaches and conducts research in the field of water resource and hydrology engineering. She has published more than 15 papers in national and international journals and has supervised more than three Ph.D. candidates.
Address: School of Water Resources Engineering, Jadavpur University, Kolkata 700032, West Bengal, India. **Email**: debasri_roy1@yahoo.co.in

Index

M. Majumder et al., *Application of Nature Based Algorithm in Natural Resource Management*, DOI 10.1007/978-94-007-5152-1,