APPLICATION OF SOFT COMPUTING TECHNIQUES FOR WATER QUANTITY AND QUALITY MODELING

BASANT YADAV

DEPARTMENT OF CIVIL ENGINEERING

INDIAN INSTITUTE OF TECHNOLOGY DELHI

MAY 2017

APPLICATION OF SOFT COMPUTING TECHNIQUES FOR WATER QUANTITY AND QUALITY MODELING

by

BASANT YADAV

Department of Civil Engineering

Submitted

in fulfilment of the requirements of the degree of Doctor of Philosophy

to the

Indian Institute of Technology Delhi

May 2017

To My Family

CERTIFICATE

This is to certify that the thesis entitled "**Application of Soft Computing Techniques for Water Quantity and Quality Modeling**" being submitted by **Mr. Basant Yadav** to the Indian Institute of Technology Delhi for the award of the degree of Doctor of Philosophy is a bonafide record of research work carried out by him under my supervision and guidance. The thesis work, in my opinion, has reached the requisite standard fulfilling the requirements for the degree of **Doctor of Philosophy**. The results contained in this thesis have not been submitted, in part or full, to any other University or Institute for the award of any degree or diploma.

(Sudheer Ch)
Joint Director
Ministry of Environment, Forest &
Climate Change, Jor Bagh, Delhi

(Shashi Mathur)
Professor
Department of Civil Engineering
Indian Institute of Technology Delhi

Acknowledgements

I would like to thank everyone who helped in the completion of this research work. The successful completion of my PhD has been made possible through the practical and professional support and advice from many people, institutions and departments. First and foremost I would like to thank my thesis supervisor Professor Shashi Mathur. Throughout my life, I will remember him for his valuable guidance, constant support and immense patience during my entire research work. His door has always been open for me and my fellow research scholars for insightful technical guidance about the various aspects of research. His deep knowledge in the field and fruitful discussions has brought many ideas in carrying out my research in this very attractive field since the beginning of the work. Professor Shashi Mathur is someone you will instantly love and never forget once you meet him. He is one of the smartest people I know. In future, I hope that I could be as lively and enthusiastic as Professor Mathur.

I would also like to thank my co-supervisor Dr. Sudheer Ch for his remarkable cooperation which he has provided to me throughout this journey of research work. He is one of the most honest and laborious person, I know till now. He is the one who taught me to accept rejection with full heart and try again with more vigour and motivation. I met him during his time of struggle, but I could hardly notice disappointment or frustration on his face, all I could see a motivated, hardworking and immensely patience man. I am so happy that God gave him the gift for his hard work and currently he is working as a Joint Director in Ministry of Environment, Forest & Climate Change, Delhi.

I would like to express my sincere thanks to my SRC members Professor Manoj Datta, Professor G. V. Ramana and Professor S. V. Veeravalli for being a part of my committee member and their suggestions helps me a lot to come through this research work. I also want to thank Professor B. Yadav (IIT Roorkee) and Professor M. Perumal (IIt Roorkee) for helping me out at various occasions.

I want to record my gratitude to the faculty members of Civil Engineering Department for their help and giving me an opportunity to carry out this research work here. I would also like to appreciate the staff of Civil Engineering Department especially Mr. Rajveer Agrawal and N. R Gahlot for providing me helpful co-operation throughout the completion of this research work.

IIT Delhi has always been home away from home. I have made so many friends since I joined here, many more than can be mentioned here. I would like to acknowledge all PhD fellows with whom I have shared social life and experience. The discussion with my colleagues contributed a lot to the progress of this research. Special thanks go to Eliza, Assefa, Vinarsi, Satish, Himanshu and Raktim for their warm friendship. I cannot be thankful enough about my lovely friends Deepak, Amarkant, Shailendra and Sushil for taking me to those wonderful trips.

Special thanks go to Gopinadh, Nitesh, Mohd. Mohsin and Ajay as they were working on their PhDs in parallel with me and with whom I have shared many experiences including suffering and joy during the entire period of our PhD studies. The road was long, with many ups and downs along the way, but together we managed. They helped me in every steps worth more than mentioning here.

Finally, I am very thankful to all my family members and relatives, in particular to my mother and father for their love, support and prayers. It's been a long time we have been living far away from each other but I can't remember a moment where they were not with me. Without your encouragement, support and sacrifice my study would not have been possible. I only hope that someday I could reciprocate all the love and care they provided to me during hard moments. I also want to thank my sister Urmila and sister in law Piyusha Yadav for their support and encouragement. My special thanks go to my brother Hemant Yadav who is my source of confidence and the reason for who I am now. I also want to thank my nephew Rajveer who is just going to be three year old and I hope that he would know how much I missed seeing him growing up if he reads this after many years.

Last but not least I praise the Almighty GOD for his love and blessings along all my ways.

May, 2017 (Basant Yadav)

Abstract

In this study four water quantity and quality problems are studied using advanced simulation and optimization techniques. This study is divided in two parts, the first part deals with water quantity management problems whereas the second part deals with water quality management problems. In the first part, two water quantity management problems are studied that include prediction of discharge in rivers and groundwater level predictions. In the second part of the study, two water quality management problems are studied, that include cost estimation of in-situ bioremediation and modeling of saltwater intrusion in coastal aquifer.

Discharge forecasting in natural rivers is a complicated procedure because of uncertainties involved in the behaviour of the flood wave movement. This further leads to solving complex problems of hydrological modelling using soft computing techniques (data-driven models). In short term flood forecasting where the accuracy of flood peak value and time to peak are critical, frequent model updating becomes unavoidable. In this study, a new technique called the Online Sequential Extreme Learning Machine (OS-ELM) is applied to discharge prediction problem. The main advantage of using this technique is its capability of updating the model equation based on new data entry without additional increase in computational cost. The OS-ELM technique is applied to predict flood forecasting in the Neckar River, Germany. The reach is characterized by significant lateral flow that affects the flood wave formation. Hourly data at an upstream section of is used to forecast flooding at the downstream site with a lead time of one to six hours. Model performance is later assessed by using three statistical measures. The performance of OS-ELM is subsequently compared to the other widely used Artificial Intelligence (AI) techniques such as Artificial Neural Networks (ANN), Support Vector Machines (SVM) and Genetic Programming (GP). The frequent updating of the model equation in OS-ELM gives a closer representation of flood events and peak values coupled with a minimal error compared to ANN, SVM and GP.

Likewise, fluctuation of groundwater levels around the world is an important area in hydrological research. In order to effectively manage the groundwater resources, it is important to obtain accurate readings and forecast of groundwater levels. This study employs two soft computing techniques, namely, ELM and SVM to forecast groundwater levels at two observation wells. A monthly data set of eight years from 2006 to 2014 comprising both the hydrological and meteorological parameters is used to conduct a comparative study of the models. These variables are used in various combinations for an univariate and multivariate analysis of the models. The study demonstrates that the proposed ELM model has better forecasting abilities compared to the SVM model for monthly groundwater level forecasting.

Similarly, a study of in-situ bioremediation of BTEX compound which is one of the most common groundwater remediation procedure used for treating organically contaminated sites is conducted. In-situ bioremediation is a highly complex and a non-linear process, modelling of which is complex and requires significant amount of computational exertion. Soft computing techniques have a flexible mathematical structure which can generalize complex nonlinear processes. In in-situ bioremediation management, a physically-based model is used for simulation and this simulated data is utilized by the optimization model to optimize the remediation cost. The recalling of simulator to satisfy the constraints is an extremely tedious and time consuming process and thus there exists a need to identify an appropriate simulator which can reduce the computational burden. This study presents a simulation-optimization approach to achieve an accurate and cost effective in-situ bioremediation system design for groundwater contaminated with BTEX (Benzene, Toluene, Ethylbenzene, and Xylenes) compounds. In this study, the ELM is used as a proxy simulator to replace BIOPLUME III for the simulations. The selection of ELM is achieved by conducting a comparative analysis using ANN and SVM that have been reported in previous studies of in-situ bioremediation to be effective. Thus, a single-objective optimization problem is solved by a coupled ELM)-Particle Swarm Optimization (PSO) technique to achieve a minimum cost for in-situ bioremediation. The results of this study indicate that ELM as a proxy simulator is much faster and more accurate than ANN and SVM. The total cost obtained by the ELM-PSO approach is assumed to be minimum possible while successfully satisfying all the regulatory constraints of the contaminated site. Further, an application of this approach while studying biological clogging of wells logically provides an additional advantage as it gave the more practical and realistic cost of the in-situ bioremediation system.

Finally, a study is conducted to predict saltwater intrusion in coastal aquifers due to over abstraction of groundwater. The governing equations of saltwater intrusion process is highly nonlinear and complex. The simulation of these equations is time consuming as both the flow and transport processes are density dependent. In this study the data based models like ANN, SVM and ELM are applied to approximate three-dimensional density dependent flow and transport processes in a coastal aquifer. A numerical simulation model SEAWAT yields the data required for training and testing of the data based models. Later, the trained data based models are used to simulate the density dependent saltwater intrusion process for a hypothetical coastal aquifer. A statistical analysis of the simulation results obtained by ANN, SVM and ELM shows that data based models could simulate a complex saltwater intrusion process successfully. Further, a comparative analysis among the three data based models indicates that recent models like SVM and ELM perform better than the well-established ANN model. The selected models are also compared based on their computational ability and the results shows that ELM is the fastest, however SVM is the most accurate proxy simulator.

सार

इस अध्ययन में उन्नत सिमुलेशन और अनुकूलन तकनीकों का उपयोग करके चार जल मात्रा और गुणवत्ता की समस्याओं का अध्ययन किया गया है। यह अध्ययन दो भागों में विभाजित है, पहला भाग जल मात्रा प्रबंधन की समस्याओं से संबंधित है, जबकि दूसरा हिस्सा पानी की गुणवत्ता प्रबंधन समस्याओं से संबंधित है। पहले भाग में, दो जल मात्रा प्रबंधन समस्याओं का अध्ययन किया गया है जिसमें नदियों और भूजल स्तर की भविष्यवाणियों में मुक्ति का पूर्वानुमान शामिल है। अध्ययन के दूसरे भाग में, दो पानी की गुणवत्ता प्रबंधन समस्याओं का अध्ययन किया जाता है, जिसमें तटीय जलभृत में खारे पानी की घुसपैठ के इन-सिटू बायोरेडियोशन और मॉडलिंग के लागत अनुमान शामिल हैं।

बाढ़ लहर आंदोलन के व्यवहार में शामिल अनिश्चितताओं के कारण प्राकृतिक नदियों में छुट्टी का पूर्वानुमान एक जटिल प्रक्रिया है। इससे नरम कंप्यूटिंग तकनीकों (डेटा चालित मॉडल) का उपयोग करके जल विज्ञान मॉडलिंग की जटिल समस्याओं को सुलझाने की ओर अग्रसर होता है। अल्पकालिक बाढ़ की भविष्यवाणी में जहां बाढ़ के शिखर मूल्य और शिखर के समय की सटीकता महत्वपूर्ण है, अक्सर मॉडल अपडेट अपरिहार्य हो जाता है इस अध्ययन में, ऑनलाइन अनुक्रमिक चरम सीखने की मशीन (ओएस-एएलएम) नामक एक नई तकनीक को भविष्यवाणी की समस्या का निर्वाह करने के लिए लागू किया गया है। इस तकनीक का उपयोग करने का मुख्य लाभ कम्प्यूटेशनल लागत में अतिरिक्त वृद्धि के बिना नए डेटा प्रविष्टि के आधार पर मॉडल समीकरण को अपडेट करने की क्षमता है। नेककार नदी, जर्मनी में बाढ़ के पूर्वानुमान की भविष्यवाणी करने के लिए ओएस-एल्म तकनीक लागू की गई है। पहुंच महत्वपूर्ण पार्श्व प्रवाह द्वारा विशेषता है जो बाढ़ लहर गठन को प्रभावित करती है। एक अपस्ट्रीम खंड में घंटेवार डेटा का उपयोग डाउनस्ट्रीम साइट पर एक से छह घंटे के लीड टाइम के साथ बाढ़ के पूर्वानुमान के लिए किया जाता है। बाद में तीन सांख्यिकीय उपायों का उपयोग करके आदर्श प्रदर्शन का आकलन किया जाता है। ओएस-एएलएम का प्रदर्शन बाद में कृत्रिम तंत्रिका नेटवर्क (एएनएन), सपोर्ट वेक्टर मशीन (एसवीएम) और जेनेटिक प्रोग्रामिंग (जीपी) जैसे अन्य

व्यापक रूप से इस्तेमाल किए जाने वाले कृत्रिम इंटेलिजेंस (एआई) तकनीकों की तुलना में है। ओएस-एएलएम में मॉडल समीकरण का लगातार अद्यतन एएनएन, एसवीएम और जीपी की तुलना में कम से कम त्रुटि के साथ बाढ़ की घटनाओं और शिखर मूल्यों के करीब प्रतिनिधित्व देता है।

इसी तरह, दुनिया भर के भूजल के स्तर में उतार-चढ़ाव हाइड्रोलॉजिकल अनुसंधान का एक महत्वपूर्ण क्षेत्र है। भूजल संसाधनों का प्रभावी ढंग से प्रबंधन करने के लिए, भूजल स्तरों के सटीक रीडिंग और पूर्वानुमान प्राप्त करना महत्वपूर्ण है। इस अध्ययन में दो मुलायम कंप्यूटिंग तकनीकों, अर्थात्, एलएम और एसवीएम को दो अवलोकन कुओं पर भूजल स्तर का अनुमान लगाने के लिए नियोजित किया गया है। 2006 से 2014 तक के आठ साल के मासिक डेटा सेट में हाइड्रॉलॉजिकल और मौसम संबंधी दोनों मापदंडों के साथ मॉडल का तुलनात्मक अध्ययन करने के लिए उपयोग किया जाता है। इन चर का प्रयोग मॉडलों के एक विकार और मल्टीवीएट विश्लेषण के लिए विभिन्न संयोजनों में किया जाता है। अध्ययन दर्शाता है कि प्रस्तावित एलएम मॉडल में मासिक भूजल स्तर की भविष्यवाणी के लिए एसवीएम मॉडल की तुलना में बेहतर पूर्वानुमान क्षमता है।

इसी तरह, बीटीईएक्स परिसर के इन-सिटू बायोरेमेडीकरण का एक अध्ययन है जो व्यवस्थित दूषित स्थलों के उपचार के लिए इस्तेमाल किया जाने वाला सबसे आम भूजल उपायों की प्रक्रिया है। इन-सिटू बायोरेमेडियेशन एक अत्यधिक जटिल और एक गैर-रैखिक प्रक्रिया है, जिसमें से मॉडलिंग जटिल है और महत्वपूर्ण मात्रा में कम्प्यूटेशनल श्रम की आवश्यकता है। सॉफ्ट कंप्यूटिंग तकनीकों में एक लचीला गणितीय संरचना होती है जो जटिल गैर-लाइनर प्रक्रियाओं को सामान्य कर सकती है। इन-सिटू बायोरिडायडेशन प्रबंधन में, एक शारीरिक रूप से आधारित मॉडल अनुकरण के लिए उपयोग किया जाता है और यह सिम्युलेटेड डाटा का उपयोग रीमेडिशन लागत को अनुकूलित करने के लिए अनुकूलन मॉडल द्वारा किया जाता है। बाधाओं को पूरा करने के लिए सिम्युलेटर की यादें एक अत्यंत कठिन और समय लेने वाली प्रक्रिया है और इस प्रकार एक उपयुक्त सिम्युलेटर की पहचान करने की आवश्यकता है जो कम्प्यूटेशनल बोझ को कम कर सकता है। यह अध्ययन

बीटीईएक्स (बेंजीन, टोलुएने, एथिलेबेन्जेन, और ज़ेलेनेस) यौगिकों से दूषित भूजल के लिए एक सटीक और लागत प्रभावी सीटू बायोरेमेडेशन सिस्टम डिज़ाइन प्राप्त करने के लिए एक सिमुलेशन-अनुकूलन दृष्टिकोण प्रस्तुत करता है। इस अध्ययन में, एएलएम का उपयोग सिमुलेशन के लिए BIOPLUME III को बदलने के लिए प्रॉक्सी सिम्युलेटर के रूप में किया जाता है। एलएनएम का चयन एएनएन और एसवीएम का उपयोग करके एक तुलनात्मक विश्लेषण करने के द्वारा प्राप्त किया गया है, जो कि इन-सिटू बायोरेमेडियेशन के पिछले अध्ययनों में प्रभावी होने के लिए रिपोर्ट किया गया है। इस प्रकार, एक एकल-अनुकूल अनुकूलन समस्या को युग्मित ELM द्वारा सुलझाया जाता है) -स्वास्थ्यिक बायोरेमेडिशन के लिए न्यूनतम लागत प्राप्त करने के लिए, पार्ट झुंड ऑप्टिमाइजेशन (पीएसओ) तकनीक। इस अध्ययन के परिणाम से संकेत मिलता है कि एएलएम एक प्रॉक्सी सिम्युलेटर के रूप में बहुत तेजी से और एएनएन और एसवीएम की तुलना में अधिक सटीक है। एल्म-पीएसओ दृष्टिकोण से प्राप्त कुल लागत को न्यूनतम संभवतः माना जाता है जबकि दूषित साइट के सभी नियामक बाधाओं को सफलतापूर्वक संतोष व्यक्त किया जाता है। इसके अलावा, कुएं के जैविक घूमने का अध्ययन करते समय इस दृष्टिकोण का एक आवेदन तार्किक रूप से एक अतिरिक्त लाभ प्रदान करता है क्योंकि यह इन-सिटि बायोरेमेडिशन सिस्टम की अधिक व्यावहारिक और वास्तविक लागत प्रदान करता है।

अंत में, भूजल के अमूर्त की वजह से तटीय जलविमानों में समुद्री पानी की घुसपैठ का अनुमान लगाने के लिए एक अध्ययन किया जाता है। खारे पानी की घुसपैठ प्रक्रिया के शासी समीकरण बेहद अलाइनलाइन और जटिल हैं। इन समीकरणों का सिमुलेशन समय लगता है क्योंकि दोनों प्रवाह और परिवहन प्रक्रिया घनत्व निर्भर हैं। इस अध्ययन में एएनएन, एसवीएम और एलएमएम जैसे डेटा आधारित मॉडलों को तटीय जलभृत में लगभग तीन-आयामी घनत्व पर निर्भर प्रवाह और परिवहन प्रक्रियाओं पर लागू किया जाता है। एक संख्यात्मक सिमुलेशन मॉडल SEAWAT डेटा आधारित मॉडल के प्रशिक्षण और परीक्षण के लिए आवश्यक डेटा उपज देता है। बाद में, प्रशिक्षित डाटा आधारित मॉडलों का उपयोग एक काल्पनिक तटीय जलभृत के लिए घनत्व आश्रित समुद्री पानी की घुसपैठ

प्रक्रिया को अनुकरण करने के लिए किया जाता है। एएनएन, एसवीएम और एएलएम द्वारा प्राप्त सिमुलेशन परिणामों के एक सांख्यिकीय विश्लेषण से पता चलता है कि डेटा आधारित मॉडल एक जटिल समुंदर घुसपैठ प्रक्रिया को सफलतापूर्वक अनुकरण कर सकते हैं इसके अलावा, तीन डेटा आधारित मॉडलों के बीच एक तुलनात्मक विश्लेषण यह दर्शाता है कि एसवीएम और एलएलएम जैसे हालिया मॉडल अच्छी तरह से स्थापित एएनएन मॉडल से बेहतर प्रदर्शन करते हैं। चयनित मॉडलों की तुलना उनकी कम्प्यूटेशनल क्षमता के आधार पर की जाती है और परिणाम दिखाते हैं कि एलएल सबसे तेज है, हालांकि एसवीएम सबसे सटीक प्रॉक्सी सिम्युलेटर है।

Table of Contents

List of Figures

List of Tables

Nomenclature

ANN	Artificial neural networks
OS-ELM	Online sequential extreme learning machine
SVM	Support vector machine
GP	Genetic programing
PSO	Particle swarm optimization BTEX
GA	Genetic algorithm
SLFN	Single-layer feedforward neural network
ARIMA	Autoregressive integrated moving average
BNN	Bayesian neural network
ANFIS	Adaptive neuro fuzzy interference
SAR	System seasonal autoregressive model
MLP	Multilayer perceptron
GEP	Gene expression programming
SA	Simulated annealing
NS-GA	Nondominated sorting genetic algorithm
FFBP	Feed forward back propagation
AARE	Average absolute relative error
Q_p	Observed peak discharge (m^3/sec)
q_{er}	Peak error (%)
t_{er}	Error in time to peak (h)
Q_c	Calculated discharge (m^3/sec)
Q_o	Observed discharge (m^3/sec)
$\overline{Q_o}$	Mean observed discharge (m^3/sec)
$\overline{Q_c}$	Mean calculated discharge (m^3/sec)
Q_{pc}	Peak of the simulated discharge hydrograph (m^3/sec)

Q_{po}	Peak of the benchmark discharge hydrograph (m^3/sec)
t_{po}	Time to peak of the observed discharge hydrograph (h)
h	Mean monthly groundwater level (m)
T	Mean monthly temperature (c)
ET	Monthly evapotranspiration (mm)
P	Monthly precipitation (mm)
C	Regularization constant
ε	Insensitive loss function
γ	Parameter of radial basis function
C_s	Concentration of contaminant (M/L^3)
C_o	Concentration of oxygen (M/L^3)
C_s'	Concentration of contaminant in a source or sink fluid (M/L^3);
C_o'	Concentration of oxygen in a source or sink fluid (M/L^3)
q	Volume flux per unit area (L/T)
b	Saturated aquifer thickness (L)
v_i	Average linear velocity in direction i (L/T)
ϑ	Effective aquifer porosity
R_s	Substrate retardation factor for hydrocarbon (dimensionless)
D_{ij}	Hydrodynamic dispersion tensor (L^2/T)
ΔC_R	Change in contaminant concentration due to biodegradation
O	Concentration of oxygen (M/L^3)
f	Ratio of oxygen to contaminant consumed.
F	Total cost of the in-situ bioremediation system ($)
i_r	Discount rate
T	Total time required for the complete remediation (years)
$q(x)$	Pumping rates of injection well and extraction wells at location x
(L^3/T) $C^q(x)$	Cost factor for injection or extraction (($ per L^3/T)

$IP(x)$	Zero-one integer for well existence (injection or extraction)
$C^{IP}(x)$	Well installation cost (injection or extraction) ($ per well)
Nw	Total number of wells
Nw^{i}	Total number of injection wells
Nw^{e}	Total number of extraction wells
q_{ex}^{k}	Pumping rate for extraction well in period k (L^3/T)
q_{ix}^{k}	Pumping rate for injection well in period k (L^3/T);
Φ	Set of all nodes in the selected study area
ψ	Set of monitoring wells in the selected study site.
CL	Well cleaning cost ($)
S_{Sf}	Fresh water specific storage
C_m	Concentration of solute mass per unit volume of fluid (M/L^3).
D	Hydrodynamic dispersion coefficient tensor (L^2/T)