Lal, Alvin Abinesh (2019) ***Development of sustainable groundwater management methodologies to control saltwater intrusion into coastal aquifers with application to a tropical Pacific island country.*** **PhD Thesis, James Cook University.**

Development of Sustainable Groundwater Management Methodologies to Control Saltwater Intrusion into Coastal Aquifers with Application to a Tropical Pacific Island Country

Thesis submitted by

Alvin Abinesh Lal

Bachelor of Science (Mathematics & Physics)

Master of Science (Physics)

University of the South Pacific, Fiji

October 2019

for the degree of Doctor of Philosophy (PhD)

in the College of Science and Engineering

James Cook University

Australia

Dedication

I dedicate this thesis to my parents. I am optimistic that this accomplishment of mine will complete the dream they had for me all those years ago when they chose to provide me with the best education they could.

The journey from a small town in Fiji to doing a PhD in Australia was no fluke. It required hard work, commitment and sacrifice

Acknowledgements

The completion of this PhD research and thesis would not have been possible without the support, motivation and assistance of many people, organisations, family members and friends. First, I must place on record my indebtedness to my supervisor Dr Bithin Datta, who toiled with me at every step of this PhD journey from the day I arrived in Townsville. I am very fortunate to be supervised by Dr Datta and I consider him as a blessing bestowed upon me. His intuitively-driven research ideas, consistent support, motivation and encouragement have been the driving forces of my research journey. I have a deficit of words to express the importance of his role in this thesis-writing endeavour. I will remain ever grateful to him. With my heart and soul, I thank Dr Datta for accepting me as his PhD student and guiding me through the research. Second, I acknowledge my secondary advisor, Associate Professor Dr Siva Sivakugan for his guidance. Special thanks go to other faculty members and staff at the College of Science and Engineering, James Cook University, for their continuous support and motivation. Specifically, I would like to thank Melisa Norton for providing college administrative support during the tenure of my PhD candidature.

Candidature financial support is an important component of a PhD research degree. I am eternally grateful to the Graduate Research School, James Cook University, for granting me the James Cook University Postgraduate Research Scholarship (JCUPRS). The scholarship provided me with financial security during my three and a half years of PhD candidature. Special recognition also goes to the College of Science and Engineering, James Cook University, for providing various levels of financial support for the successful completion of my studies. My sincere appreciation is also extended to my employer, Fiji National University, for granting educational leave for a period of three and a half years. In addition, my special and genuine gratitude goes to the Australian Land and Groundwater Association (ALGA) and International Small Islands Studies Association (ISISA) for providing conference travel grants of $500 AUD and $800 AUD, respectively.

I am also indebted to my research group members and colleagues at James Cook University for their continuous motivation and inspiration. Especially, I would like to acknowledge Dilip, Michael, Muqdad, Geeth, Mehdi, Rabindra and Teddy for their support and encouragement throughout my PhD journey. Special thanks go to my friend, Krishneel Sharma, for his healthy discussions and assistance in the development of various computer codes. I would also like to thank all the members of our Friday Soccer Group. Spending quality time with my friends on the soccer pitch helped me manage my stress levels and stay focused.

Finally, I owe my sincere gratitude to my parents, siblings and other family members, whose understanding, warmth, love and prayers have helped to give me cheer and encouragement. Also, I bow in ovation to my girlfriend and her family for their care and kindness. My girlfriend motivated me throughout and stood with me through the thick and thin of my PhD tenure. Without her support, cooperation and encouragement, the completion of this PhD thesis would not have been possible. I know my words of acknowledgement cannot compensate for the contributions made by all the supporters mentioned above, yet with all humility, I thank each one of them for their noble gestures and splendid support.

Statement of the contribution of others

This thesis was written and compiled under the guidance of my supervisor, Dr Bithin Datta. Dr Datta conceived the idea presented. I developed the computational and technical framework, performed the numerical simulations and interpreted the results. Dr Datta provided constructive suggestions and comments throughout the thesis write-up process. I wrote all the journal manuscripts and conference proceedings and prepared conference oral and poster presentations with input from Dr Datta. Dr Datta proofread the articles, proceedings, presentations and posters, and provided critical feedback. Also, financial support was given by James Cook University in the form of the James Cook University Postgraduate Research Scholarship. The College of Science and Engineering, James Cook University, provided financial support for conference travel and participation. Lastly, the authors are grateful to the Bonriki Inundation Vulnerability Assessment project reports developed by the Secretariat of the Pacific Community in partnership with the Government of Kiribati under the Australian Government Pacific Australia Climate Change Science and Adaptation Planning Program. Historical field data from these reports helped in the calibration and validation of the developed groundwater numerical model.

Abstract

Saltwater intrusion due to the over-exploitation of groundwater in coastal aquifers is a critical challenge facing groundwater-dependent coastal communities throughout the world. Sustainable management of coastal aquifers for maintaining abstracted groundwater quality within permissible salinity limits is regarded as an important groundwater management problem necessitating urgent reliable and optimal management methodologies. This study focuses on the development and evaluation of groundwater salinity prediction tools, coastal aquifer multi-objective management strategies, and adaptive management strategies using new prediction models, coupled simulation-optimization (S/O) models, and monitoring network design, respectively.

Predicting the extent of saltwater intrusion into coastal aquifers in response to existing and changing pumping patterns is a prerequisite of any groundwater management framework. This study investigates the feasibility of using support vector machine regression (SVMR), an innovative artificial intelligence-based machine learning algorithm, to predict salinity at monitoring wells in an illustrative aquifer under variable groundwater pumping conditions. For evaluation purposes, the prediction results of SVMR are compared with well-established genetic programming (GP) based surrogate models. The prediction capabilities of the two learning machines are evaluated using several measures to ensure their practicality and generalisation ability. Also, a sensitivity analysis methodology is proposed for assessing the impact of pumping rates on salt concentrations at monitoring locations. The performance evaluations suggest that the predictive capability of SVMR is superior to that of GP models. The sensitivity analysis identifies a subset of the most influential pumping rates, which is used to construct new SVMR surrogate models with improved predictive capabilities. The improved predictive capability and generalisation ability of SVMR models, together with the ability to improve the accuracy of prediction by refining the dataset used for training, make the use of SVMR models more attractive.

Coupled S/O models are efficient tools that are used for designing multi-objective coastal aquifer management strategies. This study applies a regional-scale coupled S/O methodology with a Pareto front clustering technique to prescribe optimal groundwater withdrawal patterns from the Bonriki aquifer in the Pacific Island of Kiribati. A numerical simulation model is developed, calibrated and validated using field data from the Bonriki aquifer. For computational feasibility, SVMR surrogate models are trained and tested utilizing input-output datasets generated using the flow and transport numerical simulation model. The developed surrogate models were externally coupled with a multi-objective genetic algorithm optimization (MOGA) model, as a substitute for the numerical model. The study area

consisted of freshwater pumping wells for extracting groundwater. Pumping from barrier wells installed along the coastlines is also considered as a management option to hydraulically control saltwater intrusion. The objective of the multi-objective management model was to maximise pumping from production wells and minimize pumping from barrier wells (which provide a hydraulic barrier) to ensure that the water quality at different monitoring locations remains within pre-specified limits. The executed multi-objective coupled S/O model generated 700 Pareto-optimal solutions. Analysing a large set of Pareto-optimal solution is a challenging task for the decision-makers. Hence, the *k*-means clustering technique was utilized to reduce the large Pareto-optimal solution set and help solve the large-scale saltwater intrusion problem in the Bonriki aquifer.

The S/O-based management models have delivered optimal saltwater intrusion management strategies. However, at times, uncertainties in the numerical simulation model due to uncertain aquifer parameters are not incorporated into the management models. The present study explicitly incorporates aquifer parameter uncertainty into a multi-objective management model for the optimal design of groundwater pumping strategies from the unconfined Bonriki aquifer. To achieve computational efficiency and feasibility of the management model, the calibrated numerical simulation model in the S/O model was is replaced with ensembles of SVMR surrogate models. Each SVMR standalone surrogate model in the ensemble is constructed using datasets from different numerical simulation models with different hydraulic conductivity and porosity values. These ensemble SVMR models were coupled to the MOGA model to solve the Bonriki aquifer management problem for ensuring sustainable withdrawal rates that maintain specified salinity limits. The executed optimization model presented a Pareto-front with 600 non-dominated optimal trade-off pumping solutions. The reliability of the management model, established after validation of the optimal solution results, suggests that the implemented constraints of the optimization problem were satisfied; i.e., the salinities at monitoring locations remained within the pre-specified limits.

The correct implementation of a prescribed optimal management strategy based on the coupled S/O model is always a concern for decision-makers. The management strategy actually implemented in the field sometimes deviates from the recommended optimal strategy, resulting in field-level deviations. Monitoring such field-level deviations during actual implementation of the recommended optimal management strategy and sequentially updating the strategy using feedback information is an important step towards adaptive management of coastal groundwater resources. In this study, a three-phase adaptive management framework for a coastal aquifer subjected to saltwater intrusion is applied and evaluated for a regional-scale coastal aquifer study area. The methodology adopted includes three sequential components. First, an optimal management strategy (consisting of groundwater extraction from production and barrier wells) is derived and implemented for the optimal management of the aquifer. The implemented management strategy is obtained by solving a homogeneous ensemble-based coupled S/O model. Second, a regional-scale optimal monitoring network is designed for the aquifer system, which considers possible user noncompliance of a recommended management strategy and uncertainty in aquifer parameter estimates. A new monitoring network design is formulated to ensure that candidate monitoring wells are placed at high risk (highly contaminated) locations. In addition, a *k*-means clustering methodology is utilized to select candidate monitoring wells in areas representative of the entire model domain. Finally, feedback information in the form of salinity measurements at monitoring wells is used to sequentially modify pumping strategies for future time periods in the management horizon. The developed adaptive management framework is evaluated by applying it to the Bonriki aquifer system. Overall, the results of this study suggest that the implemented adaptive management strategy has the potential to address practical implementation issues arising due to user noncompliance, as well as deviations between predicted and actual consequences of implementing a management strategy, and uncertainty in aquifer parameters.

The use of ensemble prediction models is known to be more accurate standalone prediction models. The present study develops and utilises homogeneous and heterogeneous ensemble models based on several standalone evolutionary algorithms, including artificial neural networks (ANN), GP, SVMR and Gaussian process regression (GPR). These models are used to predict groundwater salinity in the Bonriki aquifer. Standalone and ensemble prediction models are trained and validated using identical pumping and salinity concentration datasets generated by solving numerical 3D transient density-dependent coastal aquifer flow and transport numerical simulation models. After validation, the ensemble models are used to predict salinity concentration at selected monitoring wells in the modelled aquifer under variable groundwater pumping conditions. The predictive capabilities of the developed ensemble models are quantified using standard statistical procedures. The performance evaluation results suggest that the predictive capabilities of the standalone prediction models (ANN, GP, SVMR and GPR) are comparable to those of the groundwater variable-density flow and salt transport numerical simulation model. However, GPR standalone models had better predictive capabilities than the other standalone models. Also, SVMR and GPR standalone models were more efficient (in terms of computational training time) than other standalone models. In terms of ensemble models, the performance of the homogeneous GPR ensemble model was found to be superior to that of the other homogeneous and heterogeneous ensemble models.

Employing data-driven predictive models as replacements for complex groundwater flow and transport models enables the prediction of future scenarios and also helps save computational time, effort and requirements when developing optimal coastal aquifer management strategies based on coupled S/O models. In this study, a new data-driven model, namely Group method for data handling (GMDH) approach is developed and utilized to predict salinity concentration in a coastal aquifer and, simultaneously, determine the most influential input predictor variables (pumping rates) that had the most impact onto the outcomes (salinity at monitoring locations). To confirm the importance of variables, three tests are conducted, in which new GMDH models are constructed using subsets of the original datasets. In TEST 1, new GMDH models are constructed using a set of most influential variables only. In TEST 2, a subset of 20 variables (10 most and 10 least influential variables) are used to develop new GMDH models. In TEST 3, a subset of the least influential variables is used to develop GMDH models. A performance evaluation demonstrates that the GMDH models developed using the entire dataset have reasonable predictive accuracy and efficiency. A comparison of the performance evaluations of the three tests highlights the importance of appropriately selecting input pumping rates when developing predictive models. These results suggest that incorporating the least influential variables decreases model accuracy; thus, only considering the most influential variables in salinity prediction models is beneficial and appropriate.

This study also investigated the efficiency and viability of using artificial freshwater recharge (AFR) to increase fresh groundwater pumping rates from production wells. First, the effect of AFR on the inland encroachment of saline water is quantified for existing scenarios. Specifically, groundwater head and salinity differences at monitoring locations before and after artificial recharge are presented. Second, a multi-objective management model incorporating groundwater pumping and AFR is implemented to control groundwater salinization in an illustrative coastal aquifer system. A coupled SVMR-MOGA model is developed for prescribing optimal management strategies that incorporate AFR and groundwater pumping wells. The Pareto-optimal front obtained from the SVMR-MOGA optimization model presents a set of optimal solutions for the sustainable management of the coastal aquifer. The pumping strategies obtained as Pareto-optimal solutions with and without freshwater recharge shows that saltwater intrusion is sensitive to AFR. Also, the hydraulic head lenses created by AFR can be used as one practical option to control saltwater intrusion. The developed 3D saltwater intrusion model, the predictive capabilities of the developed SVMR models, and the feasibility of using the proposed coupled multi-objective

SVMR-MOGA optimization model make the proposed methodology potentially suitable for solving large-scale regional saltwater intrusion management problems.

Overall, the development and evaluation of various groundwater numerical simulation models, predictive models, multi-objective management strategies and adaptive methodologies will provide decision-makers with tools for the sustainable management of coastal aquifers. It is envisioned that the outcomes of this research will provide useful information to groundwater managers and stakeholders, and offer potential resolutions to policy-makers regarding the sustainable management of groundwater resources. The real-life case study of the Bonriki aquifer presented in this study provides the scientific community with a broader understanding of groundwater resource issues in coastal aquifers and establishes the practical utility of the developed management strategies.

Research outputs from this PhD research project

Peer-reviewed journal articles

Lal, A., and Datta, B. (2019). "Optimal groundwater use strategy for saltwater intrusion management in a Pacific island country." *Journal of Water Resources Planning and Management*, 145 (9), 04019032.

Lal, A., and Datta, B. (2019). "Optimal pumping strategies for the management of coastal groundwater resources: Application of Gaussian process regression metamodel based simulation-optimization methodology." *ISH Journal of Hydraulic Engineering,* 1-10.

Lal, A., and Datta, B. (2019). "Multi-objective groundwater management strategy under uncertainties for sustainable control of saltwater intrusion: Solution for an island country in the South Pacific." *Journal of Environmental Management*, 234, 115-130.

Lal, A., and Datta, B. (2018). "Development and implementation of support vector machine regression surrogate models for predicting groundwater pumping-induced saltwater intrusion into coastal aquifers." *Water Resources Management*, 1-15.

Lal, A., and Datta, B. (2018). "Multiple objective management strategies for coastal aquifers utilizing new surrogate models." *International Journal of Geomate*, 15(48), 79-85.

Lal, A., and Datta, B. (2017). "Modelling saltwater intrusion processes and development of a multi-objective strategy for management of coastal aquifers utilizing planned artificial freshwater recharge." *Modelling Earth Systems and Environment*, 1-16.

Articles under review

Lal, A., and Datta, B. (2019). "A comparative performance evaluation of homogenous and heterogeneous ensemble models for groundwater salinity predictions." *Environmental Monitoring and Assessment.*

Lal, A., and Datta, B. (2019). "Application of monitoring network design and feedback information for adaptive management of coastal aquifers subjected to saltwater intrusion." *International Journal of Environmental Research and Public Health*.

Lal, A., and Datta, B. (2019). "Application of the group method of data handling and variable importance analysis for prediction and modelling of saltwater intrusion processes in coastal aquifers." *Neural Computing and Applications.*

Peer-reviewed conference proceeding

Lal, A., and Datta, B. (2019). "Application of data-driven prediction models for modelling seawater intrusion in aquifers." *Proc., 8th International Contaminated Site Remediation Conference***,** Cooperative Research Centre for Contamination Assessment and Remediation of the Environment (CRC-Care), 1-2.

Lal, A., and Datta, B. "Genetic programming and Gaussian process regression models for groundwater salinity prediction: Machine learning for sustainable water resources management." *Proc., 2018 IEEE Conference on Technologies for Sustainability (SusTech)*, Institute of Electrical and Electronics Engineering (IEEE), 1-7.

Oral conference presentations

Lal, A., and Datta, B. (2019). "Application of data-driven prediction models for modelling seawater intrusion in aquifers." *8th International Contaminated Site Remediation Conference – incorporating the 2nd International PFAS Conference,* Cooperative Research Centre for Contamination Assessment and Remediation of the Environment (CRC-Care), Adelaide, Australia.

Lal, A., and Datta, B. (2019). "An ensemble modelling approach for groundwater salinity prediction." *11th World Congress on Water and Environment*, European Water Resources Association (EWRA), Madrid, Spain.

Lal, A., and Datta, B. (2018). "Genetic programming and Gaussian process regression models for groundwater salinity prediction: Machine learning for sustainable water resources management." *6th Annual IEEE Conference on Technologies for Sustainability*, Institute of Electrical and Electronics Engineering (IEEE), Long Beach CA, USA.

Lal, A., and Datta, B. (2017). "The impact of groundwater utilisation on coastal aquifers in the Pacific Island States." *15th Islands of the World Conference*, International Small Islands Studies Association (ISISA), Kangaroo Island, Australia.

Lal, A., and Datta, B. (2017) "Improved efficiency and feasibility in developing regional-scale optimal multiple objective saltwater intrusion management strategies in coastal aquifers utilizing new surrogate models." *Seventh International Conference on Geotechnique, Construction Materials and Environment*, GEOMATE, Tsu City, Japan.

Conference poster presentation

Lal, A., and Datta, B. (2018). "Practical solutions to groundwater withdrawal-induced saltwater intrusion problems: Comparison between artificial recharge and barrier well pumping." *2018 EcoForum Conference & Exhibition*, Australian Land and Groundwater Association (ALGA), Sydney, Australia.

Table of Contents

List of Tables

List of Figures

List of Acronyms

EEA	European Environment Agency
SVMR	Support Vector Machine Regression
MOGA	Multi-objective Genetic Algorithm
S/O	Simulation-optimization
GMDH	Group Method of Data Handling
AFR	Artificial Planned Recharge
ANN	Artificial Neural Network
GP	Genetic Programming
GPR	Gaussian Process Regression
SWI	Saltwater Intrusion
PIDS	Pacific Island Developing States
ANFIS	Adaptive Neuro-fuzzy Inference System
DDM	Density-dependent Model
SIM	Sharp-interface Model
M	Multi-objective
S	Single-objective
MNN	Modular Neural Network
RBFN	Radial Basis Function Network
NSGA II	Non-dominated Sorting Genetic Algorithm
CEMGA	Controlled Elitist Multi-objective Genetic Algorithm
GA	Genetic Algorithm
DE	Differential Evolution
SA	Simulated Annealing
OT	Objective Type

OM	Optimization Method
SM	Surrogate Model
EPR	Evolutionary Polynomial Regression
GMND	Groundwater Monitoring Network Design
SVM	Support vector machine
LHS	Latin Hypercube Sampling
RMSE	Root Mean Square Error
MSE	Mean Square Error
RE	Relative Error
NSE	Nash-Sutcliffe efficiency
CPU	Central Processing Unit
3D	Three-dimensional
HS	Holocene Sediments
PS	Pleistocene Sediments
EC	Electrical Conductivity
MLs	Monitoring Locations
FPWs	Freshwater Pumping Wells
BWs	Barrier Wells
MBE	Mean Bias Error
WI	Willmott's Index of Agreement
PEC	Performance Evaluation Criteria
NM	Numerical Models
NM_Av	Average of Numerical Model
IOA	Index of Agreement
MARS	Multivariate Adaptive Regression Spline
ANN_En	Artificial Neural Network Ensemble Model

GP_En model	Genetic Programming Ensemble Model
SVMR_En	Support Vector Machine Regression Ensemble Model
GPR_En	Gaussian Process Regression Ensemble Model
GMDH	Group Method of Data Handling
MAE	Mean Absolute Error
AFR	Artificial Freshwater Recharge
ADR	Abstraction, Desalination and Recharge
RWs	Recharge Wells
GMS	Groundwater Modelling System

List of Symbols

ρ	Water density
$\rho_{\circ}$	Referenced water density at zero chemical concentration
F	Storage coefficient
h	Pressure head
t	Time-step
∇	Del operator
K	Hydraulic conductivity
z	Potential head
ρ^*	Density of injection water or the withdrawn water
q	Volumetric flow rate per unit volume
ε	Dimensionless density reference ratio
c	Material concentration
c_{max}	Maximum material concentration
θ	Moisture concentration
V	Discharge
D	Dispersion tensor
α'	Compressibility of the medium
c_{in}	Material concentration in the source
$x_n(k)$	Input pumping pattern
$C(k)$	Concentration (output)
f	Function relating inputs and outputs
P	Training dataset
a_i	Vector of real independent variables
b_i	Scaler real dependent variables

w	Weight vector
$\emptyset(a)$	Feature function
C	Cost function
ε	Insensitive loss function
β and β^*	Langrangian multiplier
$K(x_i, x_j)$	Kernel function
γ	Gaussian kernel parameter
C_k^o, c_T , t_i, C_i^o and Sn	Observed (FEMWATER simulated) concentration
C_k^p, c_P, p_i, C_i^p and Pn	Predicted values of saltwater concentration
c^o, $\overline{c_T}$, $\bar{t}$, $\overline{C^o}$ and $\overline{Sn}$	Mean observed saltwater concentration
c^p, $\overline{c_P}$, $\overline{p}$, $\overline{C^p}$ and $\overline{Pn}$	Mean predicted saltwater concentration
K, n, d and N	Total data points
r	Correlation coefficient
R^2	Coefficient of determination
P_w	Production wells
B_w	Barrier wells
M_w	Monitoring wells
G_a	Generalisation ability
G_{av}	Generalisation ability at the validation stage
G_{ap}	Generalisation ability at the prediction stage
$S(\%)$	Sensitivity of input variables
D	Number of datasets for prediction
FPW_n^t	Pumping from n^{th} freshwater pumping well at the t^{th} time
BW_m^t	Pumping from the m^{th} barrier well at the t^{th} time
c_i	Salinity concentration at the i^{th} monitoring location
$\xi(,)$	Surrogate model replacing the numerical simulation model

L	Total number of FPWs
M	Total number of FPWs, BWs
T	Total management time horizon
k	Number of clusters
j	Total number of optimal solution sin the original Pareto-front
E_n	Ensemble model
K_x	Hydraulic conductivity in the x direction
K_y	Hydraulic conductivity in the y direction
K_z	Hydraulic conductivity in the z direction
C_i	Concentration at the i^{th} candidate monitoring well
Y_i	Decision variable indicating ($Y_i = 0$ or 1)
M	Maximum number of monitoring wells permitted
μ	Mean
G	Covariance
a_1	i^{th} input
b^*	Predictor responses
a^*	New input
σ^2	Variance
f^*	Latent function value
$\mu(a^*)$	Gaussian distribution mean
$v(a^*)$	Gaussian distribution variance
P_l^t	Pumping from the n^{th} PW at the t^{th} time
R_q^t	Artificial recharge (injection) from the m^{th} RW at the t^{th} time
Q	Total number of recharge wells
R	Selected solution form the Pareto-front

Chapter 1: Introduction

This chapter provides a theoretical background to the saltwater intrusion problem in coastal aquifers. Section 1.1 of this chapter outlines the various methodologies and approaches used in this thesis. Section 1.2 outlines the research significance and is followed by a description of the original contributions in Section 1.3. Section 1.4 presents the aims and objectives of this thesis. Lastly, Section 1.5 provides details on the thesis organisation and structure.

1.1 Theoretical background and thesis outline

Water is a basic necessity of life. Regions with frequent water stress conditions often use groundwater as an additional water source (Wada et al. 2010). It is estimated that more than two billion people are dependent on groundwater for their daily drinking water needs (Morris et al. 2003). Hence, groundwater is a natural freshwater resource that is treasured by many communities worldwide. Groundwater supplies almost half of the global demand for drinking water, 40% of the water demand of industry and 20% of the water used for irrigation (Foster and Chilton 2003). Lately, overall demands for groundwater have increased greatly due to rapid growths in population, agriculture and economies (Arnell 1999; Hanjra and Qureshi 2010; Rijsberman 2006). Hiscock (2011) suggested that the global demand for groundwater has tripled over the last 50 years, while its supply has declined. Persistently increasing water demands have led to over-exploitation of groundwater resources, resulting in losses of freshwater reserves and degradation of aquifers. Over-exploitation of groundwater reserves has also affected aquifers in and near coastal zones, where a large proportion of the world's population currently resides. Continuously increasing groundwater abstraction rates have had adverse and long-term impacts on coastal aquifers (Ferguson and Gleeson 2012; Vandenbohede et al. 2009), necessitating the implementation of sustainable management methodologies. Given the pressing status of groundwater scarcity and ongoing deterioration in its quality in coastal regions, the development and implementation of a holistic groundwater management framework is urgently needed. To achieve this, the present study aims to facilitate reliable and adaptive management of coastal aquifers subject to saltwater intrusion by the development and evaluation of 1) new salinity predictive models (standalone and ensemble), 2) multi-objective optimal management strategies and 3) monitoring network design methodologies.

Coastal aquifers are hydraulically connected to the sea. Groundwater abstraction from such aquifers disturbs the natural, conventional equilibrium between seawater and freshwater, instigating saltwater intrusion (Godinez and Darnault 2008; Koussis et al. 2003). A basic illustration of the saltwater intrusion phenomenon is presented in Fig. 1.1. Saltwater intrusion refers to the mass transport of saltwater into areas previously occupied by freshwater (Bear et al. 1999; Freeze and Cherry 1979). Overexploitation of groundwater resources has become a common issue in coastal regions around the world, many of which are now experiencing extensive saltwater intrusion and subsequent groundwater contamination. It is estimated that 20% of the world's aquifers are over-exploited, resulting in serious consequences such as saltwater intrusion (Connor 2015). The European Environment Agency (EEA) has also reported that saltwater intrusion due to excessive groundwater abstraction is one of the major threats to groundwater reserves in coastal aquifers (Antonellini et al. 2008). These facts suggest that the future of coastal aquifers does not look optimistic unless comprehensive sustainable management methodologies are developed and implemented. Hence, in this study, the focus is on developing and evaluating multi-objective management strategies for coastal aquifers using numerical simulation models, new prediction models, coupled simulation-optimization models and adaptive management frameworks. Most importantly, for evaluation purposes, the developed multi-objective management strategies and adaptive management strategies are applied to the Bonriki aquifer, a real regional-scale coastal aquifer system situated in the small Pacific island nation of Kiribati.

Figure 1.1: An illustration of the pumping-induced saltwater intrusion phenomenon. a) Saltwater and freshwater in an equilibrium condition (separated by a transition zone), b) saltwater encroaches inwards into the freshwater due to negative hydraulic pressure created by groundwater pumping, c) after a period of time, the freshwater becomes contaminated with saltwater and d) the pumping well is eventually closed due to saltwater contamination.

To achieve its overall aims, this thesis is divided into several components. First, support vector machine regression (SVMR) models, which are a relatively new type of predictive modelling tool, are used to approximate the responses of a complex variable-density-based saltwater intrusion model developed for an illustrative aquifer system. Second, after validating their predictive capabilities, the SVMR models are linked to a multi-objective genetic algorithm (MOGA) model to develop a multi-objective management strategy for the Bonriki aquifer system. The multi-objective management model incorporates conflicting objectives: maximisation of total pumping from production wells (wells for pumping freshwater) and minimisation of total pumping from barrier wells (wells for pumping saline water installed near the coastline). Limiting salinity concentrations at monitoring locations in the aquifer to a pre-specified limit is set as constraints. Third, uncertainties in the numerical simulation model (due to uncertainties in hydraulic conductivity and porosity) are incorporated into the development of a multi-objective management strategy using an ensemble SVMR prediction model based simulation-optimization (S/O) framework. Fourth, an adaptive management framework is developed for the Bonriki aquifer system in which feedback information from a designed monitoring network is used to sequentially modify the future year's optimal pumping rates. This adaptive management framework accounts for both user non-compliance with the recommended management strategy and uncertainties in aquifer parameters. In the later parts of this thesis, a first-ever comparison study concerning homogeneous and heterogeneous ensemble models is made to establish a better-performing ensemble predictive model. In addition, this thesis also introduces group method of data handling (GMDH) models to the field of saltwater intrusion modelling. Specifically, GMDH models are trained and tested to predict groundwater salinities in a coastal aquifer and to determine the most influential pumping rates influencing the groundwater salinity levels. Lastly, this thesis focuses on using artificial planned recharge (AFR) to control saltwater intrusion in coastal aquifer systems. The numerical simulation results and a multi-objective management model incorporating AFR as one of its management options establishes AFR as an efficient and practical solution to the saltwater intrusion crisis currently affecting people residing in coastal areas.

Overall, this thesis is an amalgamation of various models, methodologies and approaches developed and evaluated for the purposes of predicting, controlling and managing pumping-induced saltwater intrusion problems in coastal aquifers. All the groundwater simulation models presented in this study are developed using the FEMWATER computer code, which is a variable-density-based groundwater flow and transport model capable of simulating the multifaceted behaviours of coastal aquifer processes. One of the major features of this study is its application of the developed methodologies to a regional-scale study area: the Bonriki aquifer. Much of the previous research concerning saltwater intrusion modelling and management has focused on illustrative aquifer systems. To fill this research gap, the developed multi-objective management model, the multi-objective management model incorporating aquifer parameter uncertainty, the adaptive management framework, and the heterogeneous and homogeneous ensemble models (comparative study) are evaluated using the Bonriki aquifer system as a case study. It is envisioned that this case study will aid in developing regional-scale coastal aquifer management strategies and, correspondingly, demonstrate the practicality, validity and reliability of the developed methodologies.

1.2 Research Significance

Demands on groundwater reserves in coastal aquifers are at an all-time high and will only continue to intensify in the near future. Saltwater intrusion due to over-exploitation of coastal aquifers is emerging as a critical challenge for coastal communities throughout the world. Fresh groundwater reserves in coastal aquifers will quickly reach their limits (if they have not already), which will adversely affect groundwater-dependent coastal communities. According to Shah et al. (2000), the most serious groundwater challenge facing the world today is not in the development of the resource but in its sustainable management. Sustainable management of coastal aquifers for maintaining abstracted groundwater quality within permissible salinity limits is regarded as an important groundwater management problem (Sreekanth and Datta 2012). Henceforth, this research aims to develop, apply and validate the application of new predictive models, ensemble models, linked S/O models and a monitoring network design, which will aid the development of optimal adaptive management strategies for coastal aquifers.

It is surprising to note that while threats to groundwater reserves are well known, little research has been conducted on the conception and implementation of groundwater management frameworks. This study fills this gap by providing improved techniques for regulating groundwater withdrawals without compromising groundwater reserves in coastal aquifers. Additionally, groundwater withdrawal patterns and the effects of the exploitation of groundwater reserves in coastal aquifers are poorly understood. This study highlights the serious impacts of groundwater abstraction and the subsequent impacts of saltwater intrusion into coastal aquifers.

Furthermore, there is a persistent need for new solutions to coastal aquifer management problems that incorporate new knowledge and data. In this study, new prediction tools such, as SVMR and GMDH, are used to predict saltwater intrusion into coastal aquifers. Also, new standalone and ensemble predictive models are used as approximate simulators and combined with optimization models in a linked S/O framework to determine optimal management solutions for coastal aquifers. In terms of saltwater intrusion modelling, there are several numerical techniques available that are capable of simulating coupled flow and transport processes in coastal aquifers. However, accurate modelling of saltwater intrusion processes in regional-scale aquifer systems is always a complex issue and requires huge sets of field data for calibration and validation. Specific aquifer parameters are needed as inputs and, therefore, are a prerequisite to the development of accurate saltwater intrusion numerical simulation models. Also, the formulation of coastal aquifer management methodologies that consider uncertainties in aquifer parameters merits particular consideration. There is an urgent need to combine these numerical simulation models with optimization techniques to develop optimal yet sustainable saltwater intrusion management methodologies. This study provides these solutions. In addition, this study focuses on saltwater intrusion problems in the coastal aquifers of small developing island developing states. Small island countries are highly susceptible to saltwater intrusion due to their geographic locations and small land areas. However, only a limited number of studies have modelled saltwater intrusion processes and developed management solutions for aquifers in island localities. Improving research in this area is a key feature of this study.

Overall, the development and evaluation of various groundwater numerical simulation models, predictive models, multi-objective management strategies and adaptive methodologies will provide decision-makers with strategies for the optimal, sustainable and long-term use of coastal aquifers. It is envisioned that the outcomes of this research will provide information to groundwater managers and stakeholders and provide policy-makers with potential strategies for the sustainable management of groundwater resources. The Bonriki aquifer case study will provide the scientific community with a broader understanding of coastal aquifer groundwater management strategies and their practical utility. Moreover, this research explores coastal aquifer management efforts, builds research aptitude and provides results that will be of substantial benefit towards maintaining coastal groundwater sustainability. This research also fills a major gap in current coastal aquifer management capabilities by delivering more reliable, robust and practical management solutions. Lastly, the coastal aquifer management methodology that is developed will help solve current saltwater intrusion problems and avoid future unforeseen problems as well. The outcomes of this work will facilitate the long-term regional-scale management, planning and governance of groundwater resources in coastal aquifers. Although much research has already been conducted, much more is needed to ensure the optimal management of groundwater resources in coastal aquifers.

1.3 Original contributions

This thesis makes the following original contributions:

1. It makes the first-ever application of SVMR models to the prediction of groundwater pumping-induced saltwater intrusion in coastal aquifers. Also, a new sensitivity analysis methodology a set of the most influential pumping rates (inputs) that has the most impact on the output salinity concentration. The most influential pumping rates are used to develop new SVMR prediction models with improved predictive capabilities.
2. The first-ever application of SVMR prediction models to accurately approximate the complex groundwater numerical simulation model in the coupled S/O model for developing optimal coastal aquifer management strategies.
3. A new method of analysing huge sets of optimal solutions from a Pareto-front is presented. Specifically, the *k*-means clustering method is applied to provide decision-makers with a reduced set of optimal solutions that can be easily compared and evaluated.
4. A new objective function for designing a monitoring network for the adaptive management of coastal aquifers is presented. Specifically, this study uses maximization of the average logarithmic salinity concentration at candidate monitoring wells as an objective function to ensure candidate monitoring wells are placed in high-risk areas. Also, the first-ever application of *k*-means clustering methodology for determining candidate monitoring locations that are representative of the entire study area model domain.
5. A first-ever application of artificial neural network (ANN), genetic programming (GP), Gaussian process regression (GPR) and SVMR algorithms is made to develop homogeneous and heterogeneous ensemble models using identical datasets from a coastal aquifer system. The prediction performance of two types of ensemble model are compared to establish a better-performing predictive ensemble model.
6. A GMDH algorithm is used for modelling of saltwater intrusion processes in coastal aquifers and to evaluate the importance of input variables (pumping rates).
7. The benefits of using AFR to control saltwater intrusion into coastal aquifers are quantified. Also, AFR is incorporated as a management strategy in a multi-objective management model of coastal aquifers.

1.4 Aims and Objectives

The main aim of this thesis was to develop and evaluate the performance of new computational tools, such as groundwater numerical simulation models, prediction models and multi-objective regional-scale management models, to approximate and control saltwater intrusion in coastal aquifers. The specific objectives of this study are as follows.

Objective 1: Develop and utilize new prediction models to predict salinity concentrations in an aquifer system by approximating the responses of a complex variable-density saltwater intrusion numerical simulation model.

Objective 2: Develop and implement a regional-scale, computationally-feasible, surrogate model-based coupled simulation-optimization framework for the optimal and sustainable utilisation of groundwater from coastal aquifers.

Objective 3: Incorporate aquifer parameter uncertainty in the linked simulation-optimization model based coastal aquifer management model by using ensemble surrogate models.

Objective 4: Develop an adaptive management framework for coastal aquifers utilizing feedback information from the designed monitoring network.

Objective 5: Develop homogeneous and heterogeneous ensemble models for groundwater salinity prediction and compare their performance.

Objective 6: Develop and implement group method of data handling-based predictive models with variable importance analysis to accurately model saltwater intrusion processes in coastal aquifers.

Objective 7: Develop a multi-objective management model to control saltwater intrusion in coastal aquifers that uses planned artificial freshwater recharge as a management strategy.

1.5 Thesis organisation and structure

This thesis is comprised of ten chapters. The main contents of Chapters 3, 4, 5 and 6 have been published in peer-reviewed journals, while the contents of Chapters 8 and 9 are currently under review. References to published and under-review articles are given at the beginning of each of these chapters. A summary of the major contents of each chapter is made below.

Chapter 1 presents the theoretical background to the field of research. It also highlights the key components of this thesis, such as its research significance, objectives and organisation.

Chapter 2 contains a detailed description of saltwater intrusion phenomena and provides a comprehensive review of the various methodologies and approaches used in this study.

Chapter 3 describes the development and implementation of SVMR-based surrogate models for predicting saltwater intrusion in coastal aquifers. It also highlights the use sensitivity analysis to determine the input pumping rates that have the greatest impact on output salinity concentrations. This sensitivity analysis provides a set of the most influential pumping rates, which are used to develop SVMR models with improved prediction accuracy.

In Chapter 4, a multi-objective management model incorporating a SVMR surrogate-based linked S/O model is developed and implemented for the Bonriki aquifer system.

Chapter 5 details the development and implementation of a multi-objective management model for the Bonriki aquifer system that incorporates uncertainty in aquifer parameters. Parameter uncertainty is included by using the SVMR ensemble surrogate model-based linked S/O model.

Chapter 6 describes the application of the linked S/O model and the design of a monitoring network for the adaptive management of the Bonriki aquifer system.

Chapter 7 presents a comparative study of the homogeneous and heterogeneous ensemble models.

Chapter 8 introduces the GMDH modelling algorithm to the field of saltwater intrusion prediction. GMDH models are developed to predict groundwater salinity levels. Also, a variable importance assessment feature of the GMDH algorithm is used to obtain the set of most influential pumping rates that has the most impact on the output salinity concentrations.

Chapter 9 develops and evaluates a multi-objective management strategy for an aquifer system using AFR as one of its management options.

Finally, Chapter 10 presents a summary of the major findings of this thesis and its conclusions.

Chapter 2: Literature review

This chapter provides a detailed description of the saltwater intrusion phenomenon and a comprehensive review of some of the key concepts and methodologies used in this thesis.

2.1 Coastal aquifers

Coastal aquifers are the nexus between the world's oceanic and hydrologic ecosystems and provide a source of freshwater for more than one billion people living in coastal regions (Ferguson and Gleeson 2012). Coastal aquifers are hydraulically connected to the sea and, therefore, susceptible to saltwater intrusion (SWI). Understanding the theoretical background of SWI processes is key to developing efficient SWI preventative measures and dependable coastal aquifer management methodologies. In coastal regions, SWI poses a major threat to aquifers as it causes the depletion and deterioration of fresh groundwater reserves. SWI refers to the mass transport of saline water into zones previously occupied by freshwater (Bear et al. 1999; Freeze and Cherry 1979). Due to density differences, saltwater (which is denser than freshwater) intrudes beneath the freshwater system, creating a saltwater 'wedge' at the coastline (Mohsen et al. 1990). The resulting saltwater-freshwater interface (sometimes called the *zone of dispersion*) is not a firm boundary but is a transition zone with gradual change in salinity. Under equilibrium conditions, this freshwater-saltwater interface remains stationary. However, coastal aquifers are hydraulically connected to the sea and changes in these equilibrium conditions result in changes to the fresh-seawater interface. Such changes in the fresh-seawater interface, and subsequent movement of saltwater into freshwater is known as SWI. The SWI process is exceptionally complex as it is affected by many factors, such as the geological structure of the aquifer, tidal effects, sea-level changes and weather patterns. It is also greatly affected by anthropological activities. Of all these factors, excessive and/or unplanned groundwater withdrawal (Narayan et al. 2003) is considered to be one of the major causes of SWI.

2.1.1 Excessive and/or unplanned groundwater withdrawals

Among all anthropological activities, groundwater pumping has been identified as the main cause of SWI in coastal areas. As stated in Section 2.1, under natural conditions, the pressures between freshwater-saltwater interfaces remain in equilibrium. However, due to groundwater abstraction, the pressure balance turns in favour of the saltwater. This leads to the penetration of saltwater into freshwater, resulting in SWI (Essaid 1990). Specifically, groundwater pumping gives rise to an up-coning effect, which is one of the major causes of SWI (Werner et al. 2009). The extent of saltwater intrusion due to pumping is subject to factors such as the magnitudes of freshwater flow rates from the aquifer to the sea, the total rate of groundwater withdrawal relative to total freshwater recharge to the aquifer, the distance of pumping wells from the sea side boundary, rainfall intensities and frequencies, rates of evapotranspiration, the physical characteristics of aquifer materials, the presence of confining units and tidal effects. The effect of groundwater pumping on SWI rates has been exhaustively debated and reported in many studies (Alley et al. 1999; Cheng et al. 2004; Reilly and Goodman 1987; Sherif and Singh 2002; Zektser et al. 2005). All of these studies have critically examined and demonstrated the negative impacts of groundwater withdrawal on coastal aquifers. Therefore, there is an immediate need to regulate groundwater pumping patterns (Shah et al. 2003) and implement mitigation techniques that control SWI problems in coastal aquifers (Abd-Elhamid and Javadi 2011).

2.1.2 Saltwater up-coning

Saltwater up-coning is a problem of great concern in many coastal aquifers around the world (Cai et al. 2014). *Saltwater up-coning* refers to a vertical rise in the freshwater-saltwater interface beneath abstraction wells where fresh groundwater is underlain by saline water (Jakovovic et al. 2016; Kura et al. 2014; Werner et al. 2009). Zhou et al. (2005) documented a detailed description of saltwater up-coning phenomena, which are not repeated here. Under pumping conditions, the freshwater-saltwater interface moves upwards towards the well, giving rise to a cone-shaped transition zone. The transition zone moves towards the well, resulting in increased salinity levels in the groundwater being pumped. As the pumping rate increases, the salinity level in the well increases. Hence, pumping needs to be regulated and/or controlled so that abstraction wells continue to withdraw freshwater as required.

The consequence of SWI due to up-coning can persist even after groundwater pumping has stopped (Rey et al. 2013). It takes a prolonged period for an aquifer to recuperate to its initial state. This can lead to extensive changes in the saltwater-freshwater interface, resulting in aquifer distortion and groundwater depletion. For this reason, increased salinization of coastal aquifers is regarded as the ultimate hydrogeological threat to coastal communities. Depletion of fresh groundwater source due to increased salinity levels has been extensively documented in the literature (Dagan and Bear 1968; Saeed et al. 2002). Considering the adverse consequences of up-coning, several techniques have been developed and implemented to control its effects. Werner et al. (2009) listed some of these saltwater up-coning controlling measures, which include the rate and/or duration of groundwater abstraction, aquifer hydraulic properties, fluid density differences, groundwater recharge rates and well locations. In addition, to understand the occurrence of saltwater up-coning, many studies have conducted laboratory-based experiments (Koh et al. 2016), theoretical analyses (Rubin and Pinder 1977) and numerical modelling (Jakovovic et al. 2011). However, more studies on the impacts of saltwater up-coning on fresh groundwater resources in regional-scale pumping fields are needed, as better information is crucial to the development of effective groundwater management strategies.

2.2 Saltwater intrusion: Groundwater crises in coastal areas

Groundwater depletion due to SWI is a longstanding issue and numerous studies have reported it to be of great concern. Large-scale SWI problems have been reported in all seven continents of the world, including Australia (Tularam and Krishna 2009; Werner 2010), Europe (Custodio 2010; Scheidleger et al. 2004), South America (Pousa et al. 2007), North America (Andrews 1981; Barlow and Reichard 2010; Newport 1977), Asia (Cheng and Chen 2001; Don et al. 2005; Sherif et al. 2012), Africa (Van Camp et al. 2014) and Antarctica (Hillstrom and Hillstrom 2003).

As highlighted earlier, SWI is a major environmental problem that adversely impacts diverse regions around the world. While developed countries are working rigorously to combat the issue of SWI, little emphasis has been placed on small Pacific Island Developing States (PIDS). The PIDS situated in the Pacific Ocean is one of the regions that are highly vulnerable to SWI (Burns 2002; Van Der Velde et al. 2007; White and Falkland 2010). PIDS are very small in land area and are surrounded by the ocean. In addition to their geological setting, excessive withdrawal of groundwater for local consumption makes SWI relatively rapid. Cases of rampant SWI due to unplanned groundwater withdrawals in the PIDS have been reported in the Cook Islands (Carruthers 2009), Fiji (Dawe 2001), the Federated State of Micronesia (Fletcher and Richmond 2010), Guam (Bendixson et al. 2014), Kiribati (Mourits 1996), the Marshall Islands (Presley 2005), Nauru (Ghassemi et al. 1996), Niue (Mosley and Carpenter 2005), Palau (Füssel 2012), the Solomon Islands (Rasmussen et al. 2009), Samoa (Berthe et al. 2014), Tonga (Hay and Kaluwln 1993), Vanuatu (Singh et al. 2001) and Tuvalu (Webb 2007).

The above-mentioned cases imply the widespread problem of SWI in PIDS. More work in the area of coastal groundwater management is needed in the PIDS. Improved, reliable, robust and practical management solutions are could be instrumental in solving current SWI problems and, potentially, avoiding or exploring future unknown problems as well. There is a need for better long-term regional-scale management, planning and governance solutions for fragile groundwater resources in PIDS.

2.3 Saltwater intrusion prevention techniques

The key to controlling SWI is to maintain a proper balance between the water withdrawn from, and recharged to, aquifers (Barlow 2003). Over the years, numerous methods have been developed and implemented to control SWI into coastal aquifers. Some of these methods may not provide economically-feasible solutions because they are considered long-term solutions and may take many years before any effect is seen. Some of the key methods currently utilised for SWI control into coastal aquifers are summarised in Table 2.1.

Table 2.1: Common saltwater intrusion preventive measures

Method	Brief description	References
Relocation of pumping wells	A more inland location of the pumping well is favourable when compared to a seaward location. When a pumping well is located inland, the thickness of the freshwater lens increases and the danger of saltwater up-coning decreases. A comprehensive description is presented in (Bear et al. 1999) and is not repeated here. Advantage: Helps to decrease the occurrence of saltwater up-coning. Disadvantage: May be costly and difficult to implement. Most commonly, aquifer size and building locations do not allow such positioning.	Abd-Elhamid and Javadi (2011) Javadi et al. (2015) Sherif and Hamza (2001)
Reduction in pumping rates	A reduction in pumping rates ensures lower groundwater withdrawal. This is possible only if water demands are reduced, such as via public awareness and water recycling campaigns. Advantage: Ensures increased freshwater volume in the aquifer, which impedes saltwater intrusion. Disadvantage: Can lead to water shortages that have serious effects on people. May be expensive as water recycling can incur the additional costs of transportation and desalinisation.	Abarca et al. (2006) (Mantoglou 2003)
Increase in natural recharge of the aquifer	This method aims to feed aquifers with additional surface water. The retained water infiltrates into the soil and increases the volume of groundwater storage. This can be achieved through proper land use (natural vegetation and land tillage practices such as contour ploughing). Advantage: Prevents runoff to flow directly into the sea and increases groundwater storage. Disadvantage: Requires highly-permeable soil and can be time-consuming.	Calvache and Pulido-Bosch (1997) Lee and Cheng (1974)
Artificial recharge of the aquifer	Excess surface water is directed into the ground through recharge wells or by altering natural conditions to increase infiltration. Injecting water through recharge wells produces a hydraulic barrier by raising the piezo-metric head of the aquifer and prevents saltwater from encroaching inland (Luyun et al. 2011). Surface water (from rivers and lakes), treated water and desalinated water are potential sources of water for artificial recharge. Advantage: Helps to increase groundwater storage and prevent SWI. Disadvantage: Necessitates other sources of water and cannot be applied in water-scarce regions. It can also be uneconomical as water treatment is costly.	Shammas (2008) Abarca et al. (2006) Bouwer (2002)
Abstraction of saline groundwater	Saline groundwater can be abstracted for cooling and desalting purposes. This will cause the volume of freshwater to increase and subsequently reduce saltwater intrusion rate. Advantage: This method decreases the volume of saline water in the aquifer and protects pumping wells from up-coning. Disadvantage: Disposal of abstracted saltwater can be an issue. Can cause deadly pollution if saltwater is disposed of in the sea.	Soldal et al. (1994) Barrett et al. (2002)
Use of sub-surface barriers	A sub-surface barrier is an underground semi-impervious or impervious structure constructed in a coastal aquifer to impede the inland encroachment of saltwater (Allow 2012). These barriers impede the infiltration of seawater inland while also increasing the groundwater storage capacity. Advantage: Helps to reduce the intrusion of saline water. Disadvantage: Construction, operation and maintenance of the barriers can be costly.	Luyun et al. (2011) Abd-Elhamid and Javadi (2008) Essawy (2013)

Though the above-mentioned SWI-control methodologies cannot entirely prevent SWI, they have proven worthwhile in keeping aquifer salinity levels within acceptable limits. A single method, if implemented correctly, can combat SWI, while a combination of two or more methods can largely halt saline water encroachment into the aquifer. While such methodologies are widely applied, more research is needed in the context of designing inexpensive and reliable techniques for SWI prevention and control. Recent technological advancements provide a good platform on which engineers and stakeholders can collaborate on to control and manage SWI problem, safeguarding already-depleted groundwater reserves in coastal environments.

2.4 Saltwater intrusion modelling approaches

Various modelling approaches have facilitated the understanding of processes relevant to SWI in coastal aquifers. Over the years, a significant amount of research has focused on developing and employing analytical and numerical models for predicting the extent of SWI into coastal aquifers. These have helped confront the challenges facing vulnerable groundwater reserves. The different modelling approaches available for understanding SWI mechanisms are based on two concepts, namely, the *sharp interface* approach and the *diffusive interface* approach (sometimes termed the *dispersive approach*). A brief description of these two concepts is presented below.

Sharp-interface models have been extensively used in solving SWI problems. This approach assumes that saltwater and freshwater are immiscible and, therefore, are separated by an abrupt interface (Bear and Verruijt 2012). The first attempts to model seawater intrusion were made by Ghyben (1889) and Herzberg (1901) and, subsequently, their approaches were combined as the Ghyben-Herzberg approach (Bobba 1993). Specifically, this approach treats the interface between saltwater and freshwater as a sharp and well-defined interface. This approach also established that the saltwater occurs at a depth below sea level that is about 40 times the height of the freshwater above sea level (Essaid 1986). This distribution was attributed to the hydrostatic equilibrium that exists between the two fluids of different densities. Huyakorn et al. (1987) argued that, although it does not give information about the nature of the interface zone, the sharp interface method considers flow dynamics and predicts the responses of the saltwater-freshwater interface to applied stresses. Extensive reviews of the sharp interface approach are given by Reilly and Goodman (1985) and Diersch and Kolditz (2002), and are not repeated herein.

The sharp interface approach has been frequently used in SWI modelling studies because of its simplicity and low computational burden (Kacimov and Obnosov 2001; Kacimov and Sherif 2006; Person et al. 1998; Sakr 1999). However, it is alleged that the sharp interface approach is less accurate and only valid for situations where the saltwater-freshwater mixing zone is very narrow (Zhou et al. 2005). The sharp interface approach also fails to consider the flow of freshwater into the sea. In reality, the saltwater-freshwater interface is not abrupt, as the saltwater merges gradually with the freshwater by the process of mechanical dispersion (Xue et al. 1995). Padilla and Cruz-Sanjulián (1997) argued that modelling SWI using the sharp interface approach is far from rigorous and susceptible to errors. Llopis-Albert and Pulido-Velazquez (2014) similarly clarified that analytical modelling using the sharp interface approach neglects the mixing of saltwater and freshwater and indirectly assumes that seawater remains static. As a result, the authors suggested that resulting analyses may be inaccurate and, in some cases, unacceptable for real-life complex aquifer systems. In addition, Choquet et al. (2016) stated that sharp interface models do not define the behaviour of the actual transition zone but only provide evidence regarding the movement of the saltwater front. These limitations in the sharp interface approach have led to increased use of the diffusive interface approach, which is described next.

The diffusive interface approach considers saltwater and fresh groundwater as miscible liquids separated by a transition zone (Voss and Souza 1987). The transition zone has a finite thickness and the density of the water varies continuously. The diffusive interface approach considers the density dependence of the flow and transport. As a result, solving flow and transport equations simultaneously is essential. Mixing of the two fluids forming the interface at the transition zone occurs due to hydrodynamic dispersion and diffusion. Modelling SWI processes using the diffusive interface approach has provided more realistic results for both homogeneous and heterogeneous aquifer systems (Nguyen 2016). Werner et al. (2013) also stated that only the variable density model can provide an SWI estimate that can be compared to actual field measurements. The characteristics of the transition zone depend on the extent of SWI and aquifer properties. Todd (1974) explained that the thickness of transition zones can vary among different aquifers depending on aquifer properties such as its structure, the rate of groundwater extraction, and variability in recharge, tides and climate. The first attempt to model density-dependent flow in SWI problems was carried out by Henry in 1964 (Dokou and Karatzas 2012). Later, various other studies successfully employed numerical models to simulate SWI problems using the diffusive interface approach. Some of these studies include (Abarca et al. 2007; Putti and Paniconi 1995; Rastogi et al. 2004; Servan-Camas and Tsai 2009).

A wide range of analytical and numerical models have been developed for comprehending and analysing SWI processes in coastal aquifers. As stated earlier, numerical models of density-dependent flow and transport provide more accurate solutions to SWI problems. These numerical techniques include finite difference, finite element, boundary element and finite volume. Common computer codes available for SWI investigation in coastal aquifers includes SUTRA (Voss 1984), FEMWATER (Lin et al. 1997), SEAWAT (Guo and Langevin 2002), HST3D (Kipp 1987), FEFLOW (Trefry and Muffels 2007), 2D/3DFEMFAT (Sorek and Pinder 1999), HYDROGEOSHPERE (Brunner and Simmons 2012), SWIFT (Ward et al. 1984) and SHEMAT (Clauser 2012). Likewise, Werner et al. (2013) and Essink (2003) presented some common simulation codes together with basic features and corresponding practical SWI applications. Selection of the best possible modelling code is user-dependent because of their associated benefits and limitations. Cautious preliminary assessments, relevant input parameters and accurate initial conditions are necessary for precise SWI modelling.

2.5 Simulation-optimisation framework for coastal aquifer management

The simulation-optimisation (S/O) approach is one of the most widely used operative methods for coastal aquifer management purposes. Accurate numerical models simulating non-linear complex flow and transport processes in an aquifer system can generate realistic results if relevant input parameter values are available. Also, numerical simulation models can evaluate the effects of different management strategies on aquifer systems. However, simulation models are descriptive and cannot find optimal solutions instantaneously. Hence, an optimisation algorithm can be coupled with a numerical simulation model in an S/O framework to determine optimal management strategies for coastal aquifers (Roy et al. 2016). The optimisation algorithm is an important part of the S/O framework as it executes an organised search for new and better management strategies. During this search process, the simulation model is assessed numerous times to measure the influence of the proposed management strategy on the saltwater front. Selecting the best possible numerical simulation model and optimisation algorithms from a wide range of options is not an easy task, as both are based on the characteristics of the investigated problem (Hong et al. 2004). In addition, the S/O approach is usually time-consuming as it requires a relatively large number of iterations to identify improved optimal solutions. Comprehensive overviews of the S/O approach to coastal aquifer management are presented in Park and Shi (2015) and Sreekanth and Datta (2015). Some recent S/O-based coastal aquifer management studies are summarised in Table 2.2.

Table 2.2: A summary of recent S/O frameworks designed for coastal aquifer management

MT[a]	OT[b]	Objective functions	Simulation Model	SM[c]	OM[d]	Study Area	Reference
DDM	M	Maximize total pumping from production wells and minimize total pumping from production wells	FEMWATER	ANFIS	CEMGA	Illustrative	Roy and Datta (2017)
DDM	M	Minimize both the total cost of the management process and the total salinity in the aquifer	SUTRA	-	NSGA II	Benchmark aquifer (Henry's Problem)	Javadi et al. (2015)
DDM	M	Minimize economic and environmental costs while satisfying water demand	SEAWAT	MNN	NSGA II	Santorini Island, Greece	Kourakos and Mantoglou (2013)
DDM	M	Minimize total net recharge and minimization of seawater intrusion in the island's freshwater lens	SUTRA	ANN	GA	Kish Island, Persian Gulf	Ataie-Ashtiani et al. (2013)
DDM	M	Minimization of economic and environmental costs	SEAWAT	-	NSGA II	Santorini island, Greece	Kourakos and Mantoglou (2011)
SIM	S	Maximize total groundwater pumping rate	Ghyben-Herzberg relation	-	ECACO	Illustrative	Ataie-Ashtiani and Ketabchi (2011)
DDM	M	Maximize total pumping from production wells and minimize total pumping from production wells	FEMWATER	ANN and GP	NSGA II	Illustrative	Sreekanth and Datta (2010)
SIM	S	Maximize groundwater withdrawals from wells	Ghyben-Herzberg relationship	RBFN	DE	City of Heraklion, Crete	Papadopoulou et al. (2010)
DDM	M	Maximize total pumping from production wells and minimize total pumping from production wells	FEMWATER	ANN	MOGA	Illustrative	Dhar and Datta (2009)
SIM	M	Minimize operational costs, maximize groundwater reserves	SHARP	ANN	SA	Illustrative	Rao et al. (2004)

[a]MT (modelling type). SIM: sharp interface model; DDM: density-dependent model. [b]OT (objective type). M: multi-objective; S: single objective. [c]SM (surrogate model). ANFIS: adaptive neuro-fuzzy inference system; MNN: modular neural network; ANN: artificial neural network; GP: genetic programming; RBFN: radial basis function network. [d]OM (optimisation method). CEMGA: controlled elitist multi-objective genetic algorithm; NSGA II: non-dominated sorting genetic algorithm; GA: genetic algorithm; ECACO: elitist continuous ant colony optimization; DE: differential evolution; MOGA: multi-objective genetic programming; SA: simulated annealing.

2.5.1 Optimisation techniques in saltwater intrusion management studies

Various optimisation techniques have been used in groundwater management studies but only a few have been utilised in S/O-based coastal aquifer management studies. An optimisation model is well-defined in terms of an objective function and a set of constraints. Different objective functions (either single- or multi-objective) and sets of constraints have been used in various studies depending on the problems encountered. Single-objective optimisation problems have a distinctive optimal solution while multi-objective problems have a set of compromised solutions for coastal aquifer management (Singh 2014).

Some of the key optimisation algorithms used in developing groundwater management strategies for coastal areas include linear programming (Nishikawa 1998), non-linear programming (Gorelick et al. 1984), genetic programming (Sreekanth and Datta 2010), ant colony optimisation (Ataie-Ashtiani and Ketabchi 2011), harmony search (Huang and Mayer 1997), simulated annealing (Bhattacharjya and Datta 2005), particle swarm optimisation (Karatzas and Dokou 2015) and structured messy genetic algorithms (Cheng et al. 2004). Other key optimisation algorithms are presented in Table 2.2. Werner et al. (2013), Sreekanth and Datta (2015) and Nouiri et al. (2015) also reviewed some of the common optimisation algorithms used in S/O-based groundwater management studies. In addition, Ketabchi and Ataie-Ashtiani (2015) compared and contrasted traditional optimisation algorithms (linear and non-linear programming techniques) with more recent ones (genetic algorithms, ant colony optimisation, particle swarm optimisation and others) that have been used in the development of optimal management studies concerning coastal aquifers. Similarly, Singh (2014) presented a thorough evaluation of the various optimisation techniques used in coupled S/O-based coastal aquifer management studies.

It is evident from the literature that a wide range of optimisation techniques are available for coastal aquifer management problems. However, choosing an appropriate optimization model necessitates careful analysis of the problem, since many of these techniques are user- and/or problem-dependent. In particular, before choosing an optimization model for any designated coastal aquifer management task, factors such as the optimization model's capabilities, applicability and computational requirements need to be evaluated. Recently, shuffled complex evolution, continuous ant colony optimisation, multi-objective genetic algorithms and particle swarm optimisation have yielded improved, efficient and robust optimisation results when utilised in S/O-based coastal aquifer management studies (Ketabchi and Ataie-Ashtiani 2015). Hence, researchers should focus on these optimization models for developing groundwater remedial solutions concerning coastal aquifers. In addition, more work in the areas of robust optimisation and uncertainty issues in S/O frameworks for coastal aquifer management are desirable. Very few studies have focused on developing robust methodologies (Sreekanth and Datta 2014) or using parameter uncertainties (Benhachmi et al. 2001) to solve complex real-world problems. There is great scope for further research in this context that can facilitate the development of more realistic and applicable methodologies for coastal aquifer management problems.

2.5.2 Surrogate model-based S/O framework

The use of S/O frameworks for coastal aquifer management problems has yielded several key benefits. Ataie-Ashtiani et al. (2013) listed several benefits of S/O frameworks: (1) identification of optimum solutions for coastal aquifer management purposes, (2) accounting for the complexity of coastal aquifer groundwater systems and (3) enabling efficient policy-making before SWI into coastal aquifers begins. However, employing a complex numerical simulation model in an S/O framework is likely to make the process computationally infeasible. This is mainly due to two reasons. Firstly, numerical models are computationally expensive as they require high amounts of data storage and computer memory and, secondly, optimization algorithms require the numerical model to be run several times, making it a time-consuming process (Roy et al. 2016). These two drawbacks are significantly reduced by using a surrogate modelling approach (Emch and Yeh 1998; Johnson and Rogers 2000; Kourakos and Mantoglou 2009). Surrogate models are used as approximate simulators to increase computational efficiency.

A comprehensive description of the surrogate modelling approach in terms of its basic theory, architecture and various applications in engineering problems is given in Forrester et al. (2008). A *surrogate model* (also referred as a *meta-model*) is defined as an approximation of a detailed numerical model used primarily to lessen computational effort with little compromise to the results (Roy et al. 2016). Surrogate models provide comparable but faster models capable of emulating the specified output of a complex numerical model. Before its use, a surrogate model is trained and validated from input-output datasets acquired from the original numerical simulation model. Surrogate models have been established to be useful in various engineering applications and have also gained much consideration from scientific communities dealing with coastal aquifer management problems. Roy et al. (2016) and Sreekanth and Datta (2015) presented summaries of surrogate model-based, S/O framework-based, coastal aquifer management studies. In addition, Luo and Lu (2014) developed and compared various surrogate models employed in groundwater remedial processes. Some of the surrogate models used widely in coastal aquifer management S/O studies are presented in the following sections.

2.5.2.1 Artificial neural networks (ANNs)

One of the most widely-used surrogate models in S/O-based coastal aquifer management studies is ANN. These are biologically-inspired mathematical models that emulate the function of the human brain and obtain knowledge through a learning process (Bhattacharjya et al. 2007). ANNs act as approximate simulators after they are trained using the input and output from a complex numerical simulation model. Once an ANN is trained, the input-output relationship is encoded into a network which is later used to forecast outcomes based on the information fed (input) to the network. Ataie-Ashtiani et al. (2013) presented a thorough explanation of ANN models, including their architecture, generation patterns, training procedures and applicability as surrogate models for SWI simulation. Various saltwater-intruded coastal aquifer management studies have indicated that ANN surrogate model results are highly comparable to those of an original numerical simulation model (Das and Datta 1999; Kourakos and Mantoglou 2009) yet require less computational effort and time (Singh 2014).

The advantages of artificial intelligence-based techniques and their wide acceptance in groundwater management studies have motivated many researchers to use ANNs as potential surrogates to replace computationally-expensive numerical groundwater flow and transport simulation models. Ataie-Ashtiani et al. (2013) presented a list of S/O-based SWI management studies that replaced complex numerical models with ANN models. Also, Bhattacharjya et al. (2007) replaced a three-dimensional flow and transport model (FEMWATER) with an ANN surrogate model to approximate a transient salinity intrusion process in an illustrative study area. The predicted SWI results were highly comparable to those of the original FEMWATER solutions. In a similar study, Bhattacharjya and Datta (2005) established a linked S/O model for the optimal management of salinity intrusion into a coastal aquifer. A simulation model was replaced by an ANN surrogate model, while a genetic algorithm (GA) model was used for optimization. The study revealed that the ANN successfully predicted salinity concentrations, while the linked GA-ANN model presented optimal solutions for managing coastal aquifers subject to SWI. Likewise, Nikolos et al. (2008) developed and utilized ANN surrogate models to simulate groundwater heads and used them in conjunction with an optimization framework to determine optimal pumping rates.

Apart from illustrative study areas, applications of ANN surrogate models to salinity prediction in real-world aquifer systems have also yielded reliable results. For example, Banerjee et al. (2011) replaced a complex mathematical simulation model with an ANN model to predict salinity intrusion into Kavaratti Island aquifer, India. The study confirmed the suitability of ANN models as dependable replacements for numerical models and also suggested that ANNs are simpler and easier to execute. Rao and Manju (2007) replaced a three-dimensional variable density model (SEAWAT) with an ANN surrogate model and linked it with an optimization framework to develop SWI management strategies for aquifers along the banks of the River Yamuna, Delhi (India). Lastly, the use of ANN modelling strategies in S/O-based coastal aquifer management frameworks has been reported in several other studies (Bhattacharjya and Datta 2009; Das and Datta 1999; Grundmann et al. 2012; Grundmann et al. 2012; Johnson and Rogers 2000; Sreekanth and Datta 2011).

A major shortcoming of ANNs, which are data-driven surrogate models, is that as the amounts of input and output increase, the number of patterns required to train them also increases. This increases time consumption which, ultimately, defeats the purpose of using a surrogate modelling approach. The adaptive training of surrogate models (Razavi et al. 2012; Wang et al. 2014) and the development of individual modules for each surrogate model, termed a *modular neural network*; (Kourakos and Mantoglou 2013) is considered an effective way to solve this problem in S/O-based studies. These two methods are thoroughly described in Sreekanth and Datta (2015) and Ketabchi and Ataie-Ashtiani (2015) and are not repeated here.

2.5.2.2 Genetic programming (GP)

The limitations of ANNs have motivated researchers to try other efficient surrogate modelling alternatives. Genetic programming (GP) is one such innovative method that has recently gained huge popularity in the surrogate modelling field. GP is an evolutionary computational method that has been widely employed in the fields of program induction and machine learning (Aler et al. 2002). GP is classed as a powerful learning engine based on the concepts of biology and natural evolution, having the capability to automatically engender working computer models to solve a problem (Carbajal 2007; Kattan and Ong 2015). GP-based surrogate models have been reported to be effective and reliable in achieving computational efficiency in S/O-based coastal aquifer management studies. GP operates by finding a function that approximates the input-output relationship of a complex numerical model. Similar to ANNs, GP-based surrogate models also require training and testing with datasets gathered from the original numerical model. Trained GP models have been used to accurately approximate aquifer simulation processes. Also, the trained GP models are linked to an optimization model to prescribe coastal aquifer management solutions.

The GP-based surrogate modelling method has proven successful in various engineering applications (Esfahani and Datta 2015; Koza et al. 2000; Lew et al. 2006; Muttil and Lee 2005; Sheta and Mahmoud 2001). However, only a few studies have used GP surrogate models as substitutes for numerical simulation models in linked S/O frameworks for coastal aquifer management. Sreekanth and Datta (2010) compared SWI predictive performances of GP and modular neural network (MNN) models and reported GP models to be superior. In addition, Sreekanth and Datta (2011) highlighted the superiority of GP over ANN models in SWI prediction. Moreover, Datta et al. (2014) claimed that appropriately-trained GP models can be employed to decrease computational burdens when compared to numerical simulation model and, at the same time, provide sufficiently accurate prediction results. The authors successfully applied GP surrogate models to two different groundwater modelling applications: 1) for the development of monitoring network designs for optimal identification of indefinite contamination sources and 2) for the development of computationally feasible salinity invasion management strategies for coastal aquifer system.

2.5.2.3 Other surrogate modelling methods

More recently, Hussain et al. (2015) applied evolutionary polynomial regression (EPR) to replicate the behaviour of a non-linear complex aquifer system subject to SWI. The EPR approach was found to have accurate SWI prediction capabilities and was later used in a linked S/O model to develop optimal aquifer management solutions. Additionally, Roy et al. (2016) replaced the computationally-expensive numerical model OpenGeoSys with Gaussian process regression (GPR) surrogate models. The authors compared the GPR models' results with those of ANN models and concluded that both surrogate models efficiently approximated the OpenGeoSys model. However, the study also concluded that the ANN model was faster than the GPR model in approximating non-linear relationships between the input and output datasets. Later, both of these trained surrogate models were used to develop long-term management strategies for a coastal aquifer in Oman. In addition, Papadopoulou et al. (2010) combined radial basis function network (RBFN) surrogates within an optimization algorithm to develop optimal strategies to meet everyday water demands in the coastal region of Heraklion, Crete, without compromising water quality in the area. The MNN algorithm is another popular approach used to develop surrogate models for SWI prediction and its application in the linked S/O framework. MNN has been used in single-objective (Kourakos and Mantoglou 2009) and multi-objective optimization problems (Kourakos and Mantoglou 2013; Sreekanth and Datta 2010) that have helped in the development and implementation of feasible optimal aquifer management methodologies.

2.5.2.4 Ensembles of surrogates

Ensemble methods employ multiple learners (surrogates) to solve problems rather than using a single surrogate model. They are one of the most attractive surrogate modelling approaches as their generalisation ability is significantly better than those of single surrogate models (Wu et al. 2008). A complete explanation of ensemble methods is presented in Zhou (2012). Zhou et al. (2013) presented some of the advantages of using ensemble surrogates over single surrogates (robustness and accuracy) and some common methods for constructing ensembles (weight coefficient selection based on prediction variance, combining surrogates by minimizing cross-validation errors, and combining surrogates by minimizing prediction mean square errors). Ensemble surrogate models have been successfully implemented in various fields (Goel et al. 2007; Sanchez et al. 2008; Wichard and Ogorzalek 2004; Zhou et al. 2011). Likewise, ensemble surrogate models have generated reliable results when used as substitutes for numerical models in S/O frameworks (Jin et al. 2004; Yin et al. 2014). Despite the successful application of ensemble surrogate models in S/O frameworks, they have rarely been used in S/O frameworks for developing coastal aquifer management procedures. Only recently, Sreekanth and Datta (2011) employed an ensemble of GP surrogate models in a multiple-realisation optimisation framework to derive optimal pumping strategies to limit SWI into a coastal aquifer. The study demonstrated that the ensemble surrogate model approach produced dependable results for coastal aquifer management purposes. The predictive uncertainty of these surrogate models was quantified and an ensemble surrogate model was used in the multiple-realization optimization model to determine optimal extraction strategies.

It is apparent from the reviewed literature that a wide range of surrogate modelling tools is available for modelling complex non-linear groundwater flow and transport processes. However, there are a few limitations to such applications. Firstly, the development of an accurate surrogate model requires sufficient input-output data for training and validation purposes, which are sometimes not easily obtainable. Such input data patterns are obtained by executing original numerical models several times. This becomes an issue when the original numerical model takes a long time to converge, making the process time-consuming. Also, evaluation of the predictive capabilities of surrogate models is required before using them as approximate simulators. Statistical methodologies are used to evaluate the reliability and applicability of surrogate models prior to its application in the S/O framework. Secondly, employing surrogate models increases uncertainty in the predicted parameters, which may lead to errors in the optimisation process (Ketabchi and Ataie-Ashtiani 2015). Hence, the intrinsic uncertainties in surrogate models and their application to real-world coastal aquifer management problems need to be discussed.

Regardless of the aforementioned limitations, continuous effort is needed to discover and establish new surrogate modelling methods that produce better results. Further research in the area of surrogate robustness (Lim et al. 2007) and adaptive training of surrogate models (Eason and Cremaschi 2014) for improved performance can lead to more reliable results, which is essential for the development of realistic coastal aquifer management methods. Other renowned regression techniques, such as support vector machine regression (Smola and Vapnik 1997), kriging (Martin and Simpson 2005), multivariate adaptive regression splines (Friedman 1991) and regression tree (Razi and Athappilly 2005), have shown sufficiently good results when used for meta-modelling purposes. These new techniques can be used to approximate non-linear SWI processes and can be linked to S/O frameworks to develop robust solutions to coastal aquifer management problems.

2.6 Monitoring network design for coastal aquifer management

Monitoring is a fundamental part of integrated coastal aquifer management. Monitoring coastal aquifers subject to SWI can be done using geophysical methods for in-situ measurements and water sampling, which can be expensive. All-encompassing reviews of SWI monitoring in coastal aquifers are presented in Cheng and Ouazar (2016) and Custodio (1997). Optimal design of groundwater monitoring networks is seen as a cost-effective technique for monitoring SWI and resulting groundwater contamination. Groundwater monitoring network design (GMND) is principally implemented for one of the following four reasons: (1) detection monitoring, (2) ambient monitoring, (3) research monitoring, and (4) compliance monitoring (Loaiciga et al. 1992). GMNDs differ from case to case, since they are dependent on the monitoring objectives and statistical methods used. Rosen (2009) explains that preeminent network design is attained only when (1) the goal of the monitoring program is well defined, (2) the economic constraints are taken into account and (3) the hydrogeology of the investigated area is appropriately understood and assimilated into the design. A complete review of GMND is presented by Loaiciga et al. (1992), which describes various design monitoring network objectives and summarises the main approaches used in GMND construction.

GMND is widely documented in the literature (Baalousha 2010; Bartram and Ballance 1996; Meyer et al. 1994; Sanders 1983; Strobl et al. 2006). However, GMND for compliance monitoring purposes is very rare. Compliance monitoring enables decision-makers to weigh-up the outcomes of implemented management strategies. Development and implementation of an S/O framework, as described in Section 5, can provide solutions for coastal aquifer management. However, it is highly likely that the field responses to such management strategies could perhaps deviate or differ from what was predicted. Such nonconformity may be due to uncertainties in the characterization of the groundwater system (uncertainties in system parameters) and field-level deviations during the implementation of management strategies (Dhar and Datta 2009; Sreekanth and Datta 2013). These deviations can have unfavourable impacts, rendering management strategies unfeasible and/or dysfunctional. Henceforth, monitoring an aquifer's responses to the impacts of implemented management strategies should be considered crucial. The solution is compliance monitoring (Sreekanth and Datta 2015), which is a stochastically-designed optimal GMND developed to relay information on the real impacts of the implemented management strategy. This is an as an integral part of long-term aquifer management. The information gathered from GMND helps to plan and evaluate strategies to alleviate SWI in coastal aquifers and to append and/or update existing management strategies. Also, the information forms the basis for further hydrological studies and supports the assembly of long-term, dependable, groundwater management policies (Yangxiao 1994).

The main approaches to GMND have three categories: stochastic simulation approaches, variance-based approaches and optimisation-based approaches (Zhang et al. 2005). All of these categories utilise a different framework and all have shown prodigious groundwater monitoring potential. However, optimisation-based approaches have successfully aided the development of optimal, cost-effective, dependable, long-term groundwater management frameworks (Dhar and Datta 2007; Wu et al. 2005; Yeh et al. 2006). A range of optimisation objective functions have been used in GMND, including minimisation of monitoring costs (Mogheir et al. 2009; Mogheir and Singh 2002; Reed et al. 2001), optimal placement of wells (Bashi-Azghadi and Kerachian 2010; Meyer and Brill 1988; Storck et al. 1997), contamination detection (Asefa et al. 2005; Datta et al. 2009; Hudak and Loaiciga 1993; Jha and Datta 2014), and minimisation of variance in estimates (Ben-Jemaa et al. 1994; Carrera et al. 1984; Prakash and Singh 2000). Optimisation techniques also play a vital role in the development of optimal GMNDs. GMND optimisation is considered a non-linear problem and, consequently, is well-suited to heuristic optimisation techniques. A comparison of modern heuristic techniques, conventional heuristic techniques, polytope algorithms and naïve control optimisation techniques used in monitoring network designs is presented in Lee and Ellis (1996). Genetic algorithms (Kollat and Reed 2006; Mugunthan and Shoemaker 2004; Reed and Minsker 2004), simulated annealing (Chadalavada and Datta 2008; Nunes et al. 2004) and ant colony optimisation (Li and Hilton 2005; Li and Hilton 2007) are commonly-used GMND optimisation techniques.

It is evident from the literature that monitoring networks are largely implemented for groundwater quality monitoring and contamination detection. Recently, Bashi-Azghadi and Kerachian (2010) presented a methodology for selecting optimal groundwater monitoring well sites for contamination identification. The authors utilised trained probabilistic support vector machines linked with multi-objective optimisation models for the detection of unknown groundwater contamination sources. The applicability of the developed methodology was validated by using it for GMND implementation in a real case study. Also, Reed and Minsker (2004) applied kriging and non-dominated sorting genetic algorithms (NSGA II) to optimise four objectives (i.e. minimising sampling expenses, minimising estimation uncertainty, maximising precision in contamination maps, and maximizing the precision of contaminant mass approximations) for long-term, cost-effective, groundwater quality monitoring. The study established the application of optimal GMND as a promising tool for groundwater resource management. Other monitoring networks for groundwater quality monitoring and contamination detection have been reported in Preziosi et al. (2013), Baalousha (2010) and Dawoud (2004), among others.

While outcomes and/or information from GMND has benefited many global communities, monitoring network designs for coastal aquifers subjected to SWI are lacking. GMNDs can be implemented for increasing salinity level monitoring and compliance monitoring purposes. A few groundwater quality monitoring networks in saltwater-intruded coastal regions have been reported in Polemio et al. (2009), Hsu (1998) and Lee and Song (2007). In addition, Dhar and Datta (2009) presented a robust multi-objective coastal aquifer management model for SWI control that incorporated an optimal monitoring network design for compliance monitoring purposes. The monitoring network design addressed the impacts of uncertainty in the implemented management strategy for an illustrative aquifer. It was intended that compliance monitoring would provide feedback information that could be used to modify already-implemented management tactics while also developing an altogether new and/or improved strategy for long-term coastal aquifer management. More recently, Sreekanth and Datta (2013) proposed a monitoring network design for compliance monitoring and used it for adaptive management of coastal aquifers subject to SWI. The study employed an S/O framework to develop optimal pumping strategies and simultaneously designed a monitoring network to evaluate the impacts of the implemented optimal strategy on a stressed aquifer. The study demonstrated the benefits of an integrated feedback system, where information from a monitoring network is used to improve the benefits gained from an S/O method.

The several successful implementations of GMNDs imply that they can serve as crucial tools needed for the management of coastal groundwater resources. They can also be integrated into S/O frameworks and used to obtain feedback information for updating long-term management strategies. Furthermore, it is perceived that GMND for compliance monitoring purposes can provide unbiased, useful and reliable results and, hence, facilitate more robust management strategies. There is scope for further research in this area to fully comprehend the advantages of GMND in terms of coastal groundwater quality and compliance monitoring. More studies focusing on GMND in complex aquifer systems are necessary to explore its potential for groundwater monitoring and management.

2.7 Saltwater intrusion in Pacific Island Developing States

Water security is a paramount issue in many small PIDS. Groundwater in these small PIDS offers a secure, sufficient and cost-effective source of fresh water. Groundwater in PIDS is extracted through pumping wells, boreholes and domestic wells. The extracted groundwater is supplied to local communities primarily for domestic work (washing and cooking purposes). In some cases, groundwater is also used for industrial activities and irrigation (Table 2.3).

Table 2.3: Groundwater use in PIDS [Source: Sinclair (2011)]

Country	Groundwater use			
	Domestic		Irrigation	Industrial
	Drinking/cooking	Washing		
Cook Islands	30 %	70 %	NA	NA
FSM	30 %	70 %	NA	NA
Fiji	35 %	50 %	10 %	5 %
Kiribati	40 %	50-55 %	5 %	1-2 %
Marshall Islands	35-40 %	50-55 %	5-10 %	5 %
Nauru	5 %	90 %	5 %	0 %
Niue	25 %	70 %	0 %	5 %
Palau	30 %	70 %	NA	NA
PNG	NA	NA	NA	NA
Samoa	30 %	70 %	NA	NA
Solomon Islands	30 %	70 %	NA	NA
Tonga	20 %	60 %	10 %	NA
Tuvalu	10-15 %	80-85 %	5 %	0 %
Vanuatu	30 %	70 %	NA	NA

NA: data not available

2.7.1 Reported cases of saltwater intrusion in small developing Pacific Island countries

Groundwater in PIDS is limited and highly vulnerable to human activities and/or natural events. The increasing demand for water has put immense pressure on groundwater resources in PIDS. Groundwater availability in the PIDS is threatened by extraction and depletion, rates of which have increased tremendously over the last decade. Increasing salinity levels in aquifers are the most common threat to water security in small PIDS. Specifically, SWI into aquifers caused by excessive, unplanned and/or continuous withdrawal of groundwater is the main cause of escalating salinity in these aquifers. Sea-level rise, storm surges and flooding are other common causes of SWI into coastal aquifers in PIDS. Some reported SWI cases from small PIDS are presented in Table 2.4.

The problem of groundwater salinization in PIDS increased tremendously between 2000 and 2010 (Allen et al. 2014). Such problems are likely to increase as more groundwater is exhausted each day to cater to growing demands. White and Falkland (2012) stated that by the year 2030, freshwater groundwater salinization due to excessive groundwater withdrawal will be the most significant risk to water security in small PIDS. Hence, the sustainability of groundwater resources is crucial for sustaining lives in PIDS. Some of the key recommendations for ensuring sustainability and addressing groundwater management issues in PIDS are described in the next section.

The future of fresh groundwater resources in PIDS does not look promising. It is clear that the problem of water scarcity is real and immense in PIDS. However, every problem has a solution, and so it is for the problem of groundwater contamination in PIDS. The pressure on groundwater resources is increasing and leading to groundwater contamination and deterioration. There is no single solution to the growing groundwater contamination problem in PIDS. However, an integrated approach involving actions at all levels is required (from isolated local communities to law-makers). Understanding and recognising the water crisis in PIDS is essential and should be prioritised. Once the extent of the problem is identified, planning, developing and implementing mitigation actions and groundwater management strategies will take precedence. Lastly, sharing information and developing adaptive solutions and management alternatives is crucial. This can be achieved via collaboration between local communities, stakeholders, government agencies, research organisations and international donor partners.

Table 2.4: Reported cases of saltwater intrusion in small PIDS

Country	Reported cases of saltwater intrusion
Cook Islands	Saltwater has encroached into the Manihiki and Tongareva water catchments, resulting in groundwater contamination (Carruthers 2009).
Fiji	SWI is reported to be responsible for groundwater contamination in Yasawa Island (Ellison and Fiu 2010). Aquifers in Rotuma are also reported to be affected by SWI due to over-extraction (Dawe 2001).
Federated States of Micronesia	Fresh groundwater reserves in Yap State have deteriorated due to salinity intrusion and are no longer suitable for drinking (Fletcher and Richmond 2010).
Guam	High chloride content in water withdrawn from wells is reported in the areas of Finegayan and Agana sub-basin, indicating salinity intrusion (Bendixson et al. 2014).
Kiribati	Bonoriki freshwater lense, which is a source of freshwater for the people of South Tarawa, is contaminated due to salinity intrusion (Terry et al. 2013)
	Excessive pumping at Kiebu village has also led to salinity intrusion, which has contaminated groundwater reserves (Mourits 1996).
	Salinity intrusion by pumping from infiltration galleries on Bonriki and Buota has deteriorated freshwater supply and resulted in decreased coconut production in the area (Kelman and West 2009).
Marshal Islands	The Laura Area of Majuro Atoll is reported to be severely affected by SWI (Presley 2005).
Nauru	Excessive pumping of groundwater in the districts of Aiwo, Anabar, Baitsi, Boe and Nibok has led to SWI into freshwater lenses (Ghassemi et al. 1996).
Niue	Groundwater with high salinity is found towards the coast and near the more heavily-abstracted aquifer areas around Alofi (Mosley and Carpenter 2005)
Palau	Salinity intrusion has affected the low-lying islands of Kayangel, Peleliu and Angaur, and the Southwest Islands of Sonsorol, Hatohobei, Merir, Fanna and Pulo Anna. This is mainly due to excessive groundwater extraction for public water supply and septic tanks (Füssel 2012).
Solomon Islands	Groundwater reserves on the main islands and atolls such as Ontong Java have been affected by salinity intrusion (Rasmussen et al. 2009).
Samoa	In Saoluafata, a village located east of the capital Apia, coastal springs have become saline and unsafe for consumption (Berthe et al. 2014).
Tonga	Pumping of groundwater on the islands of the Ha'apai group has resulted in salinity intrusion (Hay and Kaluwln 1993).
Vanuatu	Groundwater reserves in the areas of Mataso Island in the Shepherds group, east Santo, and Aniwa in the South are affected by salinity intrusion (Singh et al. 2001).
Tuvalu	Over-abstraction of groundwater due to increased population and sea-level rise has resulted in salinity intrusion, which has degraded the fresh groundwater supply in the urban centre of Funafuti. Other areas, such as Niutoa, Vaitupu and Nukulaelae, have also experienced SWI (Webb 2007).

The next chapter describes the development of a predictive model based on a relatively new algorithm—the support vector machine regression algorithm. This is used to predict salinity concentrations in an aquifer system by approximating the responses of a variable density flow and transport numerical simulation model.

Chapter 3: Development and Implementation of Support Vector Machine Regression Surrogate Models for Predicting Groundwater Pumping-induced Saltwater Intrusion into Coastal Aquifers

The main contents of this chapter have been published and copyrighted, as outlined below:

Lal, A., and Datta, B. (2018). "Development and Implementation of Support Vector Machine Regression Surrogate Models for Predicting Groundwater Pumping-Induced Saltwater Intrusion into Coastal Aquifers." *Water Resources Management*, 1-15.

3.1 Summary

Predicting the extent of saltwater intrusion (SWI) into coastal aquifers in response to changing pumping patterns is a prerequisite of any groundwater management framework. This study investigates the feasibility of using support vector machine regression (SVMR), an innovative artificial intelligence-based machine learning algorithm, for predicting salinity concentrations at selected monitoring wells in an illustrative aquifer under variable groundwater pumping conditions. For evaluation purposes, the SVMR predictions are compared with well-established genetic programming (GP)-based surrogate models. SVMR and GP models are trained and validated using identical sets of input (pumping) and output (salinity concentration) datasets. The trained and validated models are then used to predict salinity concentrations at specified monitoring wells in response to new pumping datasets. The predictive capabilities of the two learning machines are evaluated using different proficiency measures to ensure their practicality and generalisation ability. The performance evaluations suggest that the prediction capability of SVMR is superior to that of GP models. Also, a sensitivity analysis methodology is proposed to assess the impact of pumping rates on salt concentrations at monitoring locations. This sensitivity analysis provides a subset of the most influential pumping rates, which is used to construct new SVMR surrogate models with improved predictive capabilities. The improved prediction capability and the generalisation ability of the SVMR models, together with the ability to improve prediction accuracy by refining the input set used for training, makes the use of the proposed SVMR models more attractive. Predictive models with high accuracy are potentially very useful for designing large-scale coastal aquifer management strategies.

3.2 Background

The surrogate modelling approach significantly reduces computational load by substituting complex simulation models with cheaper to run surrogate models (Wang et al. 2014). Commonly-used surrogate models for approximating groundwater transport and flow processes in aquifer include radial basis functions (Christelis and Mantoglou 2016), artificial neural networks (Bhattacharjya and Datta 2009), modular neural networks (Kourakos and Mantoglou 2009), the fuzzy inferences system (Roy and Datta 2016), multivariate adaptive regression splines (Roy and Datta 2017) and genetic programming (Sreekanth and Datta 2011). To propose a more efficient and reliable surrogate model, GP and SVMR models are developed and implemented in a simulated illustrative aquifer to predict SWI processes.

Surrogate models based on GP have recently gained popularity in the context of SWI prediction and have proven to possess accurate predictive capabilities. Sreekanth and Datta (2010) applied and compared GP and MNN models for SWI predictions in a coastal aquifer system. The study reported that GP models are superior to MNN models in terms of predictive accuracy. Also, Sreekanth and Datta (2011) highlighted the superiority of GP over ANN models in predicting aquifer SWI processes. On the other hand, the support vector machine (SVM) is a powerful tool for solving classification and regression problems and has recently gained worldwide popularity in the machine learning field (Gretton et al. 2001; Yang et al. 2002). Farag and Mohamed (2004) reiterated that SVM has become popular due to its attractive features and accurate performance. SVM is a new technique built on structural risk minimisation (SRM) instead of empirical risk minimisation (like ANN); hence, SVMs are robust and accurate (Yoon et al. 2011). Support vector machine regression (SVMR) is a version of SVM used for regression analysis. One of the leading advantages of the SVMR algorithm is that it theoretically minimises the anticipated error in a learning process and lessens the problem of overfitting (Yu et al. 2006). Numerous studies have established that the predictive performance of SVMR models is superior to those of other learning techniques (Chevalier et al. 2011; He et al. 2014; Wen et al. 2009; Wu et al. 2004; Yoon et al. 2011).

A key feature of this study includes investigating the influence of groundwater pumping rates on salinity levels at monitoring wells. Establishing the relative importance of pumping rates at each well is crucial for the development of an efficient surrogate model. Variable (variable in this study refers to groundwater pumping rates) ranking is an important tool in surrogate modelling approaches. It allows the elimination of irrelevant (less impactful) variables, condenses data dimensionality, escalates learning efficiency and improves predictive performance (Liu et al. 2011). In the present study, a sensitivity analysis approach is utilised to estimate the relative importance of input variables. The obtained set of most-influential variables is used to construct new surrogate models with greater prediction capability.

This chapter aims to make two important contributions. First, a comparatively new technique, the SVMR methodology, is used for approximating the responses of a complex saltwater flow and transport numerical model. SVMR models have not yet been utilized for saltwater intrusion process simulation. The second major contribution is in the improvement of predictive accuracy by selecting the most significant input variables based on sensitivity analysis. Performance evaluation of the proposed method is conducted to demonstrate its potential advantages.

3.3 Methodology

3.3.1 FEMWATER: A coastal aquifer simulation modelling tool

The FEMWATER (Lin et al. 1997), a Finite Element Model of Water Flow Through Saturated-Unsaturated Media based computer code licenced from Groundwater Modelling System (GMS; AquaVeo 2011) was used to develop a finite element-based three-dimensional numerical model for simulating saltwater intrusion processes in the Bonriki aquifer. Datta et al. (2009), Roy and Datta (2016) and Sreekanth and Datta (2010) have effectively used FEMWATER code to simulate saltwater intrusion processes in various coastal aquifer systems. The flow and transport equations are coupled by the density coupling coefficient and by Darcy velocities, which makes the saltwater intrusion problem highly non-linear. Therefore, a finite element-based simulation model is utilized to solve these two governing equations concurrently as coupled equations. Equations (3.1) and (3.3) represent the flow and transport processes, respectively.

The governing flow equation is in the form of a modified Richards equation (Lin et al., 1997):

$$\frac{\rho}{\rho_\circ} F \frac{\partial h}{\partial t} = \nabla \cdot \left[K \left(\nabla h + \frac{\rho}{\rho_\circ} \nabla z \right) \right] + \frac{\rho}{\rho^*} q \qquad (3.1)$$

Where, ρ is the water density at chemical concentration c, $\rho_\circ$ is the water density at $c = 0$, F is the storage coefficient, h is pressure head, t represents time, ∇ is a del operator, K is hydraulic conductivity, z is the potential head, ρ^* is the density of injected or withdrawn water and q is the volumetric flow rate per unit volume of the source (recharge) and/or sink (pumping).

In saltwater intrusion problems, the constitutive relationship between fluid density and concentration takes the form:

$$\frac{\rho}{\rho_\circ} = 1 + \varepsilon \frac{c}{c_{max}} \tag{3.2}$$

Where ε is a dimensionless density reference ratio, c is the material concentration in the aqueous phase and c_{max} refers to the maximum material concentration.

The governing equations for transport describe material (dissolved salt) transport through a groundwater system. Some key transport processes that are considered in this study are advection, dispersion/diffusion, and injection/withdrawal. The transport equation (Lin et al., 1997) is given by:

$$\begin{aligned} &\theta \frac{\partial c}{\partial t} + V \cdot \nabla c - \nabla \cdot (\theta D \, \nabla c) \\ &= -\left(\alpha' \theta c \frac{\partial h}{\partial t}\right) + q c_{in} - \frac{\rho^*}{\rho} q c \\ &+ \left(F \frac{\partial h}{\partial t} + \frac{\rho_\circ}{\rho} V \cdot \nabla \left(\frac{\rho}{\rho_\circ}\right) - \frac{\partial \theta}{\partial t}\right) c \end{aligned} \tag{3.3}$$

Where θ refers to moisture concentration, V is discharge, D is a dispersion tensor, α' is the compressibility of the medium and c_{in} is the material concentration in the source.

3.3.2 Coastal aquifer simulation model

A 3D numerical model of the study area, containing a portion of a multi-layered coastal aquifer similar to that in Sreekanth and Datta (2010), was constructed. The length of the coastline was 2.13 km, while the other two boundaries were 2.04 km (*side A*) and 2.79 km (*side B*) in length, respectively. The aquifer (depth = 60 m) was divided equally into three layers. The aquifer was considered vertically heterogeneous based on the different hydraulic conductivity values of the aquifer layers. The study area of 2.53 km^2 incorporated five barrier wells (B_w), eight production wells (P_w) and three monitoring wells (M_w). A 3D view of the study area with specific well locations is given in Fig. 3.1. The P_w were installed for withdrawing fresh groundwater for domestic utilisation, whereas barrier wells were installed near the coastline for SWI prevention. Pumping from B_w enabled SWI prevention by inducing a steeper hydraulic gradient towards the sea, thus averting inward seawater encroachment into the aquifer (Dhar and Datta 2009). M_w was installed for salinity monitoring purposes. The sea-side boundary had a constant head and constant concentration boundary with a concentration of 35 kg/m^3. The other two boundaries of the study area were taken as no-flow boundaries. The modelled aquifer was discretised into finite triangular elements with an average element size of 150 m. The element size near the wells was set to

75 m. Constant groundwater recharge of 0.00054 m/d was specified over the entire study area. The screening interval of all wells was taken from the second and third layers of the aquifer. The compressibility and dynamic velocity of water were taken as 6.7 x 10^{-20} md^2/kg and 131.328 kg.md, respectively. The other key parameters used in the aquifer simulation are listed in Table 3.1.

Table 3.1: Key parameter values for model development

Property		Value
Hydraulic conductivity	*x* direction	15 m/d
	y direction	7.5 m/d
	z direction	1.5 m/d
Bulk density		1600 kg/m^3
Longitudinal dispersivity		50 m
Lateral dispersivity		25 m
Molecular dispersion coefficient		0.69 m^2/d
Density reference ratio		0.025
Soil porosity		0.46
Compressibility		8.5 x 10^{-15} md^2/kg

A 3D transient simulation was commenced from an initially steady-state condition of the aquifer, achieved by constant pumping of 300 m^3/day from only three of the production wells for a period of 20 years. After 20 years, it was noted that the observed heads at different nodes in the model domain became constant. These resultant heads and concentrations were used as initial conditions (initial head and concentration) for aquifer simulation for the specified period of four years (*4^{th}-time step*), where pumping from all production and barrier wells was instigated.

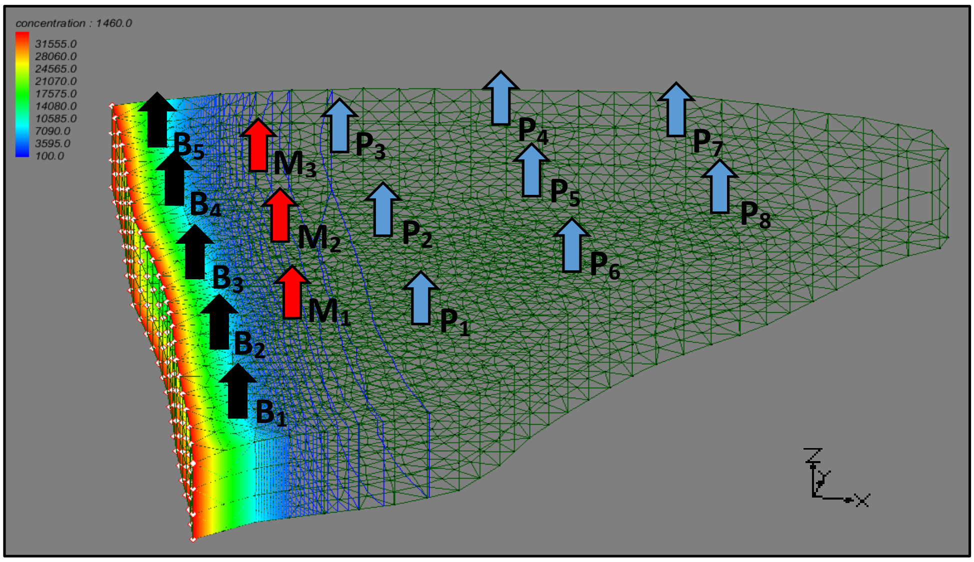


Figure 3.1: Salt concentration contour at the end of 1460 days (4^{th} time step) in response to one set of pumping from all production and barrier wells.

3.3.3 Algorithms for surrogate model construction

3.3.3.1 Genetic programming

The present study built a genetic programming (GP) model that caters for multiple inputs and a single output. The model was trained using inputs (pumping values generated using Latin hypercube sampling) in terms of $x_1(k), x_2(k), \dots\dots, x_n(k)$, and used $C(k)$ as the output (resulting salt concentration attained from the complex numerical simulation model) with (k) denoting the time step. The relationship between these inputs and the output can be expressed as:

$$C(k) = f(x_1(k), x_2(k), \dots\dots, x_n(k)) \tag{3.4}$$

The goal was to find the values of model output $C(k)$ as a function of past outputs. Models produced by GP were employed to estimate the relationship function f.

Discipulus GP software (Foster 2001) was used to develop regression models of salt concentrations prediction at monitoring locations. The GP model learned from the supplied input-output datasets and encoded input-output relationship into the programs. The resultant best-fit model (best program) was chosen for the modelling of salinity intrusion into coastal aquifers. A more detailed description of the GP models is given in Sreekanth and Datta (2011).

3.3.3.2 Support vector machine regression

A brief description of the SVMR modelling algorithm, similar to that of Parveen et al. (2016), is presented here. The regression analysis comprises training dataset $P = \{(a_1, b_1), (a_2, b_2), \dots, (a_N, b_N)\}$, so that a_i is a vector of real independent variables and b_i is the matching scalar of real dependent variables. The regression equation in the feature space can be estimated by:

$$z(a, w) = (w \cdot \emptyset(a) + c) \tag{3.5}$$

Where w is the weight vector, c is a constant, $\emptyset(a)$ represents the feature function and $w \cdot \emptyset(a)$ is the dot product. Support vector works by minimizing the following equations:

$$Q(f) = C\frac{1}{N}L_\varepsilon\big(b, z(a, w)\big) + \frac{1}{2}\|W^2\| \tag{3.6}$$

and

$$L_\varepsilon\big(b, z(a, w)\big) = \begin{cases} 0 & if\ |b - z(a, w)| \le \varepsilon \\ |b - z(a, w)| - \varepsilon & otherwise \end{cases} \tag{3.7}$$

The left-hand side of Eq. (3.6) characterizes the empirical error and the term C provides a measure of the optimisation between the empirical error and the model complexity given by the second term of the said equation. Equation (3.7) describes the loss function, called the ε-insensitive loss function (Vapnik et al. 1997). The optimization problem is transformed into a dual problem by incorporating Langrangian multipliers β and β^*. Only the non-zero coefficients, alongside their input vectors a_i, are termed the support vectors. The final form is as follows:

$$z(a, \beta_i, \beta_i^*) = \sum_{i=1}^{N_{sv}} (\beta_i - \beta_i^*) \left(\emptyset(a_i) \cdot \emptyset(a_j) \right) + c \qquad (3.8)$$

With the assistance of the kernel function $K(x_i, x_j)$, the support vector regression function can be written as:

$$z(a, \beta_i, \beta_i^*) = \sum_{i=1}^{N_{sv}} (\beta_i - \beta_i^*) K(a, a_i) + c \qquad (3.9)$$

The term c is calculated utilizing Karush-Kuhn-Tucker conditions. The most important parameters that control SVMR problems are the cost function C, the radius of the insensitive tube ε and the kernel parameter. MATLAB R2016a software was used to construct the SVMR models. For the present study, a Gaussian kernel was used, with ε, C and ɣ (Gaussian kernel parameters) having values of 0.60, 10 and 0.001, respectively. These values were obtained by carrying out several trial experiments.

3.3.4 Development of surrogate models

3.3.4.1 Generation of input-output patterns

The numerical simulation model was used to generate input-output patterns for training and validating of the surrogate models. Transient pumping (inputs) were obtained from a uniform sampling distribution using Latin hypercube sampling (LHS) (Loh 1996) with an upper bound of 1300 m^3/d and lower bound of 0 m^3/d. The resulting salt concentrations at each monitoring well were obtained from the numerical model after each set of pumping rates from the production and barrier wells were fed to the model. Each numerical model took approximately 4–5 minutes to converge. Seven hundred sets of pumping rates and resulting outputs (concentrations) were assembled by running the simulation model 700 times. These input-output patterns were later used for surrogate model training, validation and prediction.

3.3.4.2 Training, validation and prediction procedure

For cross-validation purposes, generated datasets were partitioned randomly into training, validation and prediction datasets without replacement, similar to the procedure described in Roy and Datta (2016). The training and validation sets were used for surrogate model development, while the prediction set was used to test model performance (Westerhuis et al. 2008). Out of the 700 datasets, 400 were used for training, 100 were used for validation, and 200 were used for prediction. The output (concentration) was only fed into the model during the training and validation stages. No output was fed into the model at the prediction stage. Three different surrogate models were constructed utilising GP and SVMR algorithms for predicting salinity concentrations at the three corresponding monitoring wells.

3.3.4.3 Surrogate model performance evaluation

Root mean square error (RMSE), mean square error (MSE), relative error (RE), correlation coefficient (*r*), and Nash-Sutcliffe efficiency (NSE) were used to appraise the performance of the developed models at the three stages. Mathematical expressions of these error estimates are presented as Eqs. (3.10) to (3.14).

$$RMSE = \sqrt{\frac{1}{K}\sum_{k=1}^{K}(C_k^o - C_k^p)^2} \quad (3.10)$$

$$MSE = \frac{1}{K}\sum_{k=1}^{K}(C_k^o - C_k^p)^2 \quad (3.11)$$

$$RE = \frac{1}{K}\sum_{k=1}^{K}\left|\frac{C_k^o - C_k^p}{C_k^o}\right| \quad (3.12)$$

$$r = \frac{\sum_{k=1}^{K}(C_k^o - c^o)(C_k^p - c^p)}{\sqrt{\sum_{k=1}^{K}(C_k^o - c^o)^2}\sqrt{\sum_{k=1}^{K}(C_k^p - c^p)^2}} \quad (3.13)$$

$$NSE = 1 - \frac{\sum_{k=1}^{K}\left(C_k^o - C_k^p\right)^2}{\sum_{k=1}^{K}(C_k^o - c^o)^2} \quad (3.14)$$

where C_k^o and C_k^p are observed (numerically simulated) and predicted saltwater concentrations, respectively, c^o and c^p are observed and predicted mean saltwater concentrations, respectively, and K represents the number of data points.

The generalisation abilities (G_a) of the developed models at the validation stage (G_{av}) and the prediction stage (G_{ap}) were determined using Eqs. (3.15) and (3.16), respectively.

$$G_{av} = \frac{RMSE\ in\ validation\ stage}{RMSE\ in\ training\ stage} \quad (3.15)$$

$$G_{ap} = \frac{RMSE\ in\ prediction\ stage}{RMSE\ in\ training\ stage} \quad (3.16)$$

3.3.5 Methodology for variable sensitivity analysis

The relative influence of pumping from each well on the salinity level at the respective monitoring wells was achieved *via* sensitivity analysis, as proposed by Liong et al. (2000)

and Kordjazi et al. (2015). Only better-performing SVMR models were used in the sensitivity analysis experiment. The core of this method lies in the fact that the change in concentration at monitoring wells are measured by altering pumping from each pumping well at a constant rate at a time. The sensitivity of the input variable (pumping rate) on the output (concentration at monitoring well) in the developed SVMR surrogate model was calculated using Eq. (3.17).

$$S(\%) = \frac{100}{d}\sum_{k=1}^{d}\left(\frac{\%\,\Delta\, in\ ouput}{\%\,\Delta\, in\ input}\right) \tag{3.17}$$

Where *d* denotes the number of sets of samples used for prediction. To compute the influence of pumping rate on the concentration at a specified monitoring location, the pumping rate was changed by 20% for a given time period with the pumping rates at the other locations held constant. A change in concentration at a monitoring location represents a change in output. The percentage change is computed as a fraction of the change to the input pumping in the corresponding sample set in the input. For output, the percentage change was computed according to the predicted concentration in the sample set of pumping. The investigated aquifer had 13 pumping wells and a constant pumping rate from each well within the four-year management timeframe was set. This gave a total variable of 52 (13 wells × 4 years). Determining the effect of pumping from each well on the salinity concentrations at the monitoring wells was essential in quantifying the influence of pumping stressors on the aquifer.

3.3.6 Dimensionality reduction and performance of SVMR models

For demonstration purposes, new SVMR surrogate models were developed using only the variables indicated by the sensitivity analysis to be the most sensitive. The dimensions of the training, validation and prediction datasets were reduced by eliminating a combination of three less-influential variables. A total of seven new surrogate models were constructed using various combinations (cases A to G) of excluded less-sensitive variables. The SVMR models were trained and validated with only a subset of the most influential variables before being used for prediction. The performance of the new surrogate models was analysed and compared with that of models utilising all variables.

3.4 Results and discussion

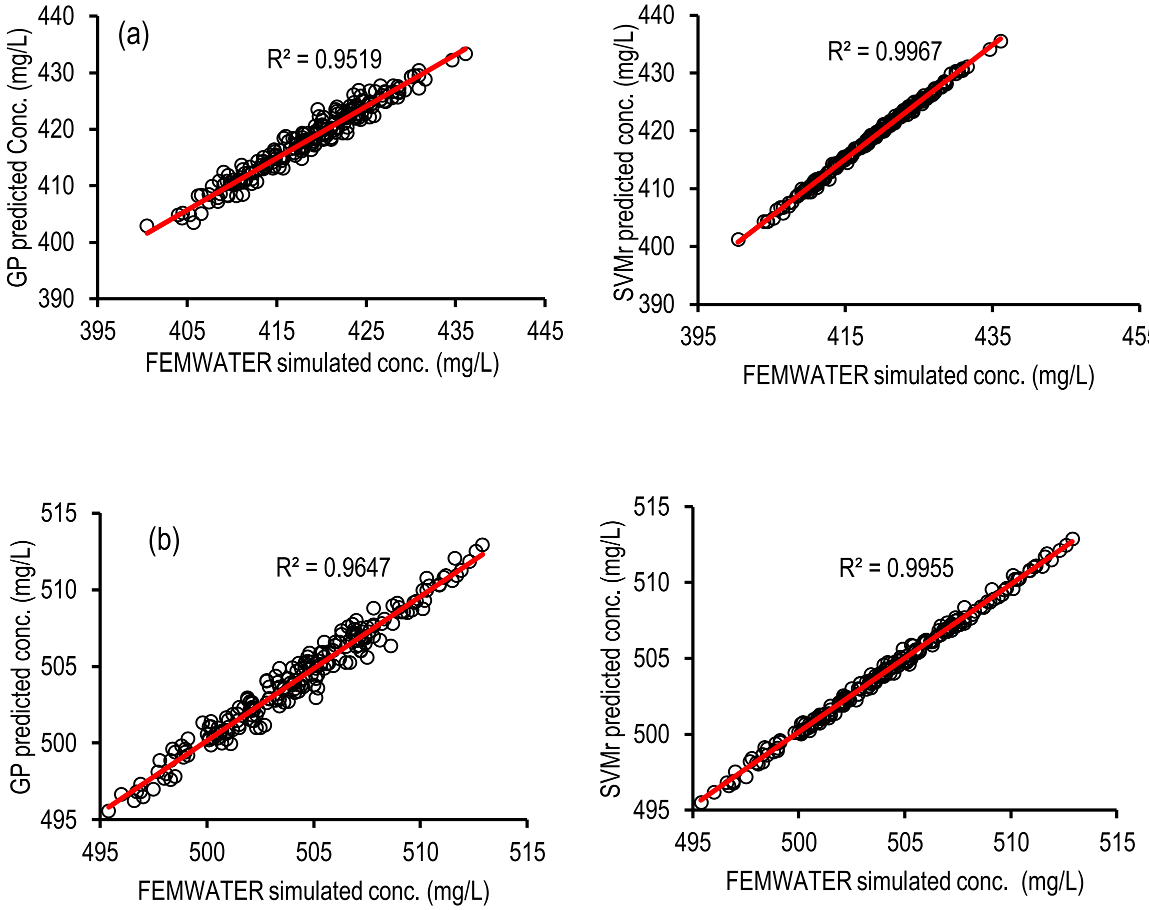
3.4.1 Performance evaluation of the GP and SVMR models

The results of the performance evaluations of the SVMR and GP models at the training, validation and prediction stages are given in Table 3.2. The RMSE was used to measure the average magnitude of error between the target (simulated) values and model output (predicted) values. Similarly, MSE is the mean of the squares of the differences between the target and model output values. The RE expresses the closeness between the target and predicted model output values. The value of *r* presents the degree of correlation between the target and model output values, while NSE denotes the predictive power of a particular surrogate model. The performance of the GP and SVMR models at the training, validation and prediction stages in terms of the five evaluation criteria showed similar trends. The RMSE, MSE and RE values at the three stages were substantially smaller for the SVMR models. During the training stage, SVMR had the smallest RMSE value of 0.238 for M_2 and the highest RMSE of 1.663 for M_1. Comparing the RMSE values of the SVMR models at the validation stage, the lowest value of 0.323 was obtained for M_2, while the highest value of 1.691 was achieved for M_1. Accordingly, the MSE values of the SVMR models were smaller than those of the GP models. Likewise, RE was also comparably smaller for the SVMR model. The *r* and NES values did not significantly differ for the two model types. However, the *r* values for SVMR model were slightly higher than those of the GP models at all monitoring wells. This outcome highlights the superiority of SVMR models over GP models. A model can be considered accurate if the calculated NSE value is greater than 0.8 (Shu and Ouarda 2008). The NSE values obtained using the GP and SVMR models were greater than 0.8, indicating that both types of models had promising results and can be employed for SWI prediction.

Table 3.2: Results of performance evaluation of the GP and SVMR models

Training						
M_w	Model	MSE	RMSE	RE	*r*	NSE
M_1	SVMR	0.164	0.405	5.0×10^{-7}	0.997	0.99
	GP	2.764	1.663	1.4×10^{-6}	0.933	0.95
M_2	SVMR	0.057	0.238	1.6×10^{-7}	0.997	0.99
	GP	0.571	0.756	6.1×10^{-7}	0.963	0.99
M_3	SVMR	0.184	0.428	6.7×10^{-8}	0.989	1.00
	GP	0.407	0.638	2.0×10^{-10}	0.975	1.00
Validation						
M_1	SVMR	0.202	0.449	1.8×10^{-6}	0.994	0.99
	GP	2.859	1.691	4.7×10^{-6}	0.929	0.99
M_2	SVMR	0.105	0.323	2.6×10^{-7}	0.993	1.00
	GP	0.379	0.616	2.4×10^{-6}	0.971	0.99
M_3	SVMR	0.187	0.432	5.4×10^{-7}	0.989	1.00
	GP	0.412	0.642	7.2×10^{-7}	0.976	1.00
Prediction (using a new set of input data)						
M_2	SVMR	0.155	0.394	7.5×10^{-7}	0.997	1.00
	GP	3.284	1.812	5.6×10^{-6}	0.952	0.99
M_2	SVMR	0.073	0.271	2.9×10^{-7}	0.996	0.99
	GP	0.570	0.755	1.5×10^{-6}	0.965	0.99
M_3	SVMR	0.255	0.505	1.7×10^{-7}	0.984	1.00
	GP	0.486	0.697	5.3×10^{-7}	0.957	0.99

The performance of the developed SVMR models in predicting salinity levels at specified monitoring wells was superior to that of GP models (refer Table 3.2). At all the monitoring wells, higher *r* and NSE values and lower RMSE and RE values were observed for the SVMR model compared to the GP models. The values of *r* and NSE were also close to 1 if not 1. The obtained prediction results indicate that the SVMR model had higher prediction accuracy and was able to effectively emulate a complex numerical simulation model. Figure 3.2 shows scatterplots of simulated concentrations (from a numerical model) versus the concentrations at the three monitoring wells predicted by the GP and SVMR models. Overall, the performance evaluation results at all three stages establish that the SVMR model was the most effective in terms of forecasting SWI at specified wells in the modelled coastal aquifer.



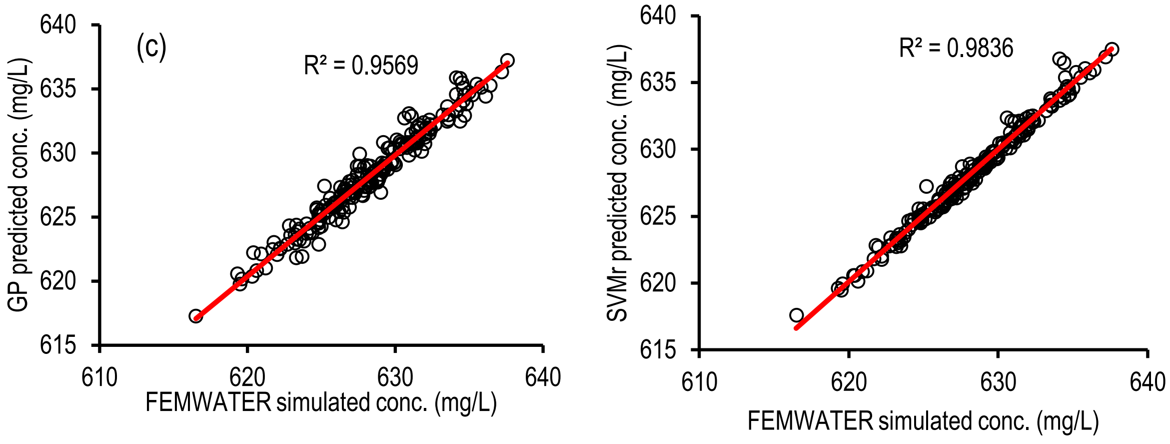
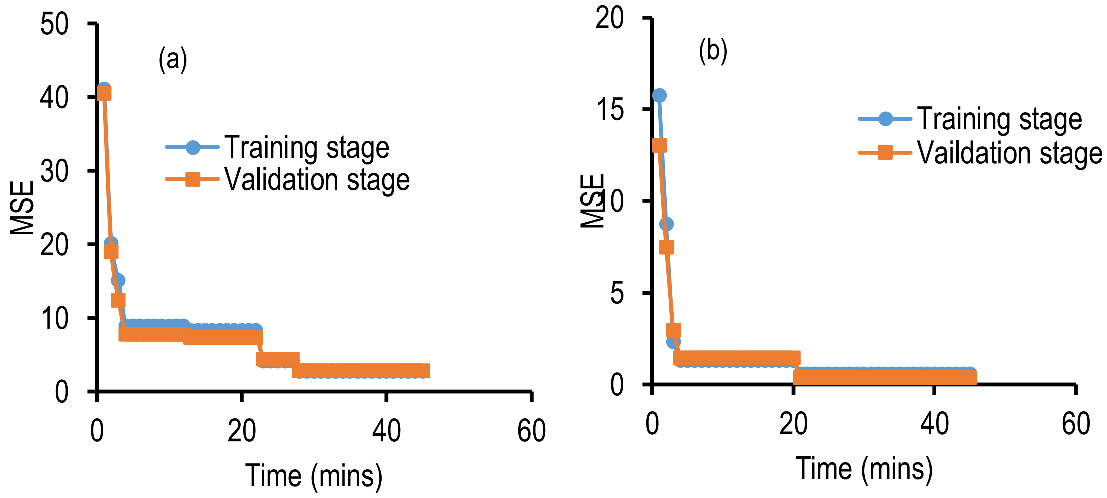
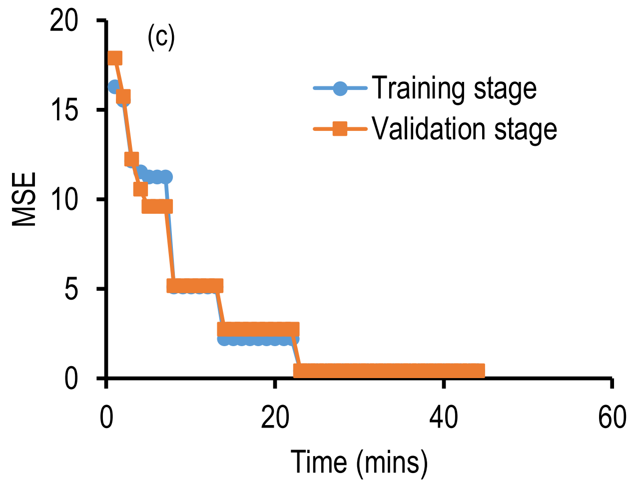


Figure 3.2: Correlations between FEMWATER simulated salt concentrations (conc.) and concentrations predicted by surrogate models for wells (a) M_1, (b) M_2 and (c) M_3

In addition, time is considered a key factor when training and validating surrogate models. In the present study, the time taken to train and validate the GP models was higher than that required for SVMR models. Each GP model took approximately 45 minutes to converge (i.e. to achieve a constant MSE value). However, only a few seconds were needed to train and validate the SVMR models. This is one of the major drawbacks of the GP technique. The time taken to train and validate the GP models is presented in Fig. 3.3. Time constraints are a major factor that favours SVMR models, making them more efficient and applicable.





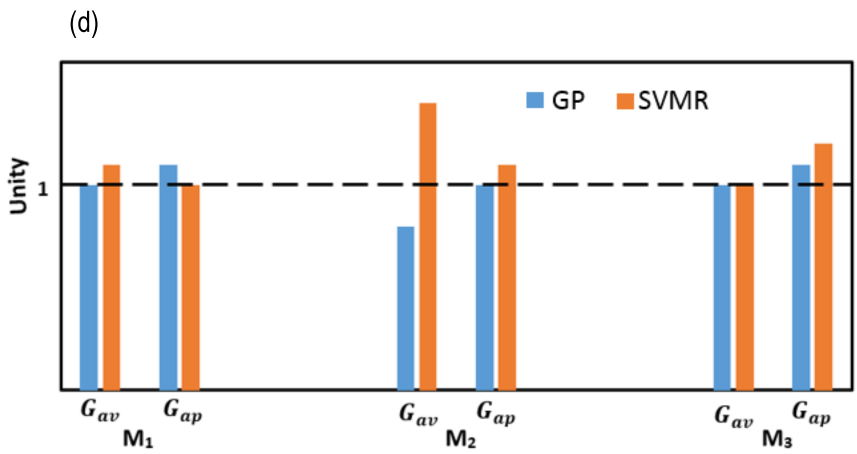


Figure 3.3: MSE convergence for the GP model at wells a) M_1, b) M_2 and c) M_3. d) Generalisation ability of the developed GP and SVMR models.

3.4.2 Generalisation ability of the developed models

The G_a values of the GP and SVMR models are presented in Fig. 3.3 (d). If a surrogate model can learn a given system perfectly (in this case, a FEMWATER simulation), then the G_{av} and G_{ap} values should both reach unity (Yoon et al. 2011). An overtrained model will have G_{av} and G_{ap} values greater than unity, whereas an undertrained model will have G_{av} and G_{ap} values lower than unity. The results presented in Fig. 3.3 (d) show perfectly trained, overtrained and under-trained models at the respective validation and prediction stages. It is noted that there are no specific trends in the G_a evaluations of the models. However, all the G_{av} and G_{ap} values are approximately closer to unity. At well M_1, the G_a ability of both models at both stages is relatively good as compared to those at wells M_2 and M_3. At M_2, the G_{av} value for the GP model indicates undertraining, while that of SVMR represents overtraining. Also, it is noted that the SVMR models were not undertrained in any of the cases. The G_{av} and G_{ap} values of the SVMR models were either 1 or > 1. This establishes that SVMR can provide a reliable surrogate model.

3.4.3 Improvements in surrogate model prediction capability

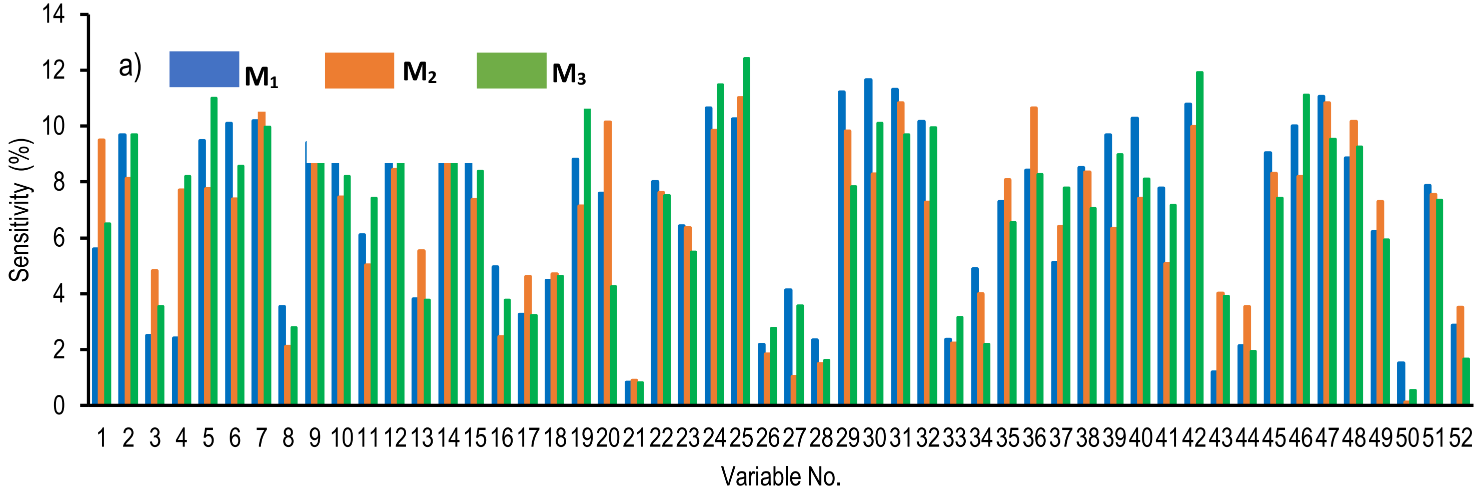
The sensitivity of each input pumping rate (variable) was calculated and the results are presented in Fig. 3.4 (a). Determining the effect of pumping rate at each well on the salinity concentration at monitoring wells is essential in quantifying the influence of pumping rates on the aquifer. This quantification process showed the influence of different variables (pumping rates from beneficial and barrier wells) on the salinity levels at monitoring wells. It can be seen that variables 7, 9, 19, 24, 25, 29, 30, 31, 42, and 47 were of greater importance in predicting salinity concentrations at all three monitoring wells. However, variables 50, 21 and 28 had little contribution and were of less importance in the prediction of salinity intrusion at the respective monitoring wells.

The sensitivity analysis suggests that this information can be to adjust the pumping strategy to achieve desired concentrations and/or rectify deviations from predicted concentrations. It recommends variables which need to be focused on and prioritised. Pumping rates at less influential wells can be altered since they have little influence on the salinity at monitoring wells. Also, training a model using only a subset of the most influential variables can improve its predictive capabilities (Chebrolu et al. 2005). Similarly, the sensitivity analysis undertaken in this study specified a subset of influential variables and redundant variables. The predictive performance of the new SVMR surrogate models developed using only the most influential variables (as indicated by the sensitivity analysis) were evaluated. GP models were also trained using a subset of the most influential variables. However, only a minor improvement (< 0.5 % improvement in RMSE) was shown for case A. Also, GP models took the same amount of time (~45 mins) to train and validate. Minor improvements and time are important factors that do not favour GP models. Hence, training GP models using other subsets of variables was unsuitable.

The performance evaluation of the new SVMR models at the three respective stages is summarised in Table 3.3. These results confirm that a better-performing surrogate model can be obtained by using a subset of the most influential variables only. This is evident from Table 3.3, where decreases in MSE, RMSE and RE are observed in all cases. Consequently, minor increases in r and NSE values were also attained. The evaluation results suggest there were improvements in model performance at the three stages in all seven cases. The percentage improvements in the MSE, RMSE and r results at the prediction stage are shown in Fig. 3.4 (b). Small improvements in the training, validation and prediction stages for the first three cases (A, B and C) were observed. However, when two variables were eliminated (i.e. in cases D, E, and F), a substantial improvement was observed. This improvement decreases slightly in case G, in which all three less sensitive variables were excluded. The improvements in cases D, E and F are greater than those of the other cases in which one or three variables were excluded. The results suggest that models can be improved by reducing the dimensionality of the sample dataset and using the most influential variables only.

Table 3.3: Results of SVMR surrogate model performance evaluation during training, validation and prediction stages

Case (variables eliminated)	M_w	Training					Validation					Prediction				
		MSE	RMSE	RE	r	NSE	MSE	RMSE	RE	r	NSE	MSE	RMSE	RE	r	NSE
A (50)	M_1	0.159	0.405	4.9×10^{-7}	0.998	0.99	0.199	0.446	1.4×10^{-6}	0.995	0.99	0.150	0.387	7.1×10^{-7}	0.998	1.00
	M_2	0.051	0.238	1.5×10^{-7}	0.998	0.99	0.101	0.318	2.5×10^{-7}	0.994	1.00	0.069	0.236	2.6×10^{-7}	0.997	0.99
	M_3	0.179	0.428	6.3×10^{-8}	0.990	1.00	0.179	0.423	5.1×10^{-7}	0.991	1.00	0.251	0.501	1.4×10^{-7}	0.988	1.00
B (21)	M_1	0.162	0.402	4.8×10^{-7}	0.998	0.99	0.200	0.447	1.2×10^{-6}	0.996	1.00	0.148	0.385	7.0×10^{-7}	0.997	1.00
	M_2	0.054	0.232	1.4×10^{-7}	0.998	0.99	0.100	0.316	2.0×10^{-7}	0.995	1.00	0.070	0.265	2.1×10^{-7}	0.997	0.99
	M_3	0.180	0.424	6.2×10^{-8}	0.991	1.00	0.181	0.425	5.1×10^{-7}	0.990	1.00	0.248	0.498	1.5×10^{-7}	0.989	1.00
C (28)	M_1	0.161	0.401	4.8×10^{-7}	0.998	0.99	0.198	0.445	1.3×10^{-6}	0.996	0.99	0.152	0.390	7.2×10^{-7}	0.998	1.00
	M_2	0.053	0.230	1.2×10^{-7}	0.998	0.99	0.102	0.319	2.0×10^{-7}	0.995	1.00	0.071	0.266	2.3×10^{-7}	0.997	1.00
	M_3	0.181	0.425	6.5×10^{-8}	0.992	1.00	0.183	0.428	5.4×10^{-7}	0.993	1.00	0.250	0.500	1.1×10^{-7}	0.986	1.00
D (50, 21)	M_1	0.164	0.405	5.0×10^{-7}	0.998	0.99	0.198	0.445	1.1×10^{-6}	0.997	1.00	0.149	0.386	6.9×10^{-7}	0.998	1.00
	M_2	0.053	0.230	1.4×10^{-7}	0.999	0.99	0.101	0.318	2.0×10^{-7}	0.995	1.00	0.068	0.261	2.1×10^{-7}	0.997	0.99
	M_3	0.180	0.424	6.1×10^{-8}	0.992	1.00	0.182	0.427	5.1×10^{-7}	0.992	1.00	0.250	0.500	1.3×10^{-7}	0.989	1.00
E (50, 28)	M_1	0.158	0.397	4.8×10^{-7}	0.998	0.99	0.197	0.444	1.3×10^{-6}	0.996	0.99	0.148	0.385	6.8×10^{-7}	0.998	1.00
	M_2	0.055	0.235	1.1×10^{-7}	0.998	0.99	0.101	0.318	2.2×10^{-7}	0.995	1.00	0.067	0.259	2.2×10^{-7}	0.998	1.00
	M_3	0.180	0.424	6.2×10^{-8}	0.993	1.00	0.183	0.428	5.1×10^{-7}	0.993	1.00	0.247	0.497	1.1×10^{-7}	0.991	1.00
F (21, 28)	M_1	0.159	0.399	4.6×10^{-7}	0.999	0.99	0.196	0.443	1.3×10^{-6}	0.996	0.99	0.149	0.386	6.8×10^{-7}	0.998	1.00
	M_2	0.057	0.239	1.3×10^{-7}	0.998	1.00	0.103	0.321	2.1×10^{-7}	0.995	1.00	0.068	0.261	2.4×10^{-7}	0.997	0.99
	M_3	0.184	0.429	6.2×10^{-8}	0.990	1.00	0.184	0.429	5.0×10^{-7}	0.993	1.00	0.251	0.501	1.3×10^{-7}	0.992	1.00
G (50, 21, 28)	M_1	0.163	0.404	4.9×10^{-7}	0.998	0.99	0.199	0.446	1.3×10^{-6}	0.997	0.99	0.152	0.390	6.8×10^{-7}	0.998	1.00
	M_2	0.056	0.237	1.5×10^{-7}	0.998	0.99	0.102	0.319	1.9×10^{-7}	0.995	1.00	0.069	0.263	2.2×10^{-7}	0.997	1.00
	M_3	0.182	0.427	6.6×10^{-8}	0.993	1.00	0.181	0.425	5.0×10^{-7}	0.991	1.00	0.248	0.498	1.0×10^{-7}	0.989	1.00



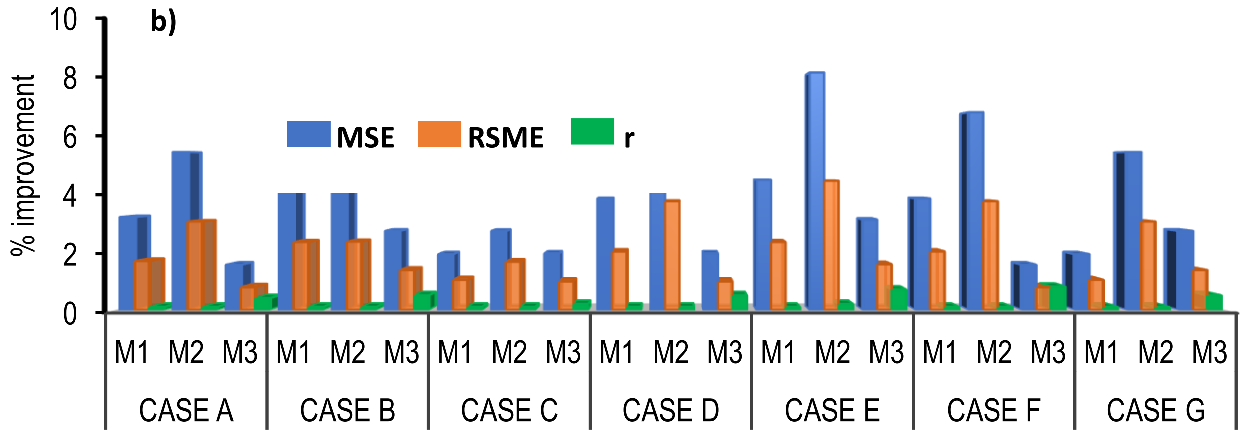


Figure 3.4: a) Influence of each variable on the predicted salinity levels at monitoring wells and b) improvements in SVMR surrogate models after variable elimination

The better-performing SVMR surrogate models were those developed for cases D, E and F. Case E could be classified as the best surrogate model as it had a higher percentage improvement than the other cases. These improved surrogate models can be used for SWI prediction purposes as they yield more accurate and reliable results than the others. Eliminating variables with negligible influences on the outcome reduces the data dimensionality and, consequently, reduces unnecessary data collection and operational costs. These benefits could be of high significance in large-scale investigations where extensive data collection and effort are required. Also, prioritising which data are collected avoids the collection of redundant data, which reduces storage requirements saves time and costs.

3.4 Conclusions

This chapter presented SVMR models as a feasible alternative to the GP method for the problem of SWI prediction. Trained and validated GP and SVMR models were developed for predicting multifaceted SWI processes in coastal aquifers in response to variable pumping patterns at a combination of production and barrier wells. The performance evaluation results revealed that SVMR models are superior to GP models and can be successfully applied to obtain precise and dependable SWI predictions. The evaluation results suggest that SVMR prediction models can be applied in groundwater management studies as computationally-efficient substitutes for the FEMWATER model. Another advantage of utilizing SVMR surrogates is that the time required to train and validate them is significantly less than that required by GP models. Also, the proposed method of ranking the input variables used in the surrogate models presented useful information. This study demonstrates that developing a surrogate model by refining the input dataset and retaining only the most influential variables can yield a superior predictive model with substantial benefits. In summary, this chapter establishes SVMR models as robust tools for predicting SWI into coastal aquifers. Hence, SVMR models can be utilized as proxies for more complex numerical models in coastal aquifer management studies incorporating surrogate models. The use of SVMR surrogates can significantly reduce the computational burdens encountered in optimization problems. The results of the present study establish that SVMR models can be applied to SWI prediction problems and related management studies. In the next chapter, reliable SVMR surrogates are used in a simulation-optimisation framework for developing optimal groundwater pumping strategies for the sustainable management of coastal aquifers.

Chapter 4: Optimal groundwater-use strategy for saltwater intrusion management in a Pacific Island country

The main contents of this chapter have been published and copyrighted as outlined below:

Lal, A., and Datta, B. (2019). "Optimal groundwater use strategy for saltwater intrusion management in a Pacific island country." *Journal of Water Resources Planning and Management*, 145 (9), 04019032.

4.1 Summary

Escalating salinity levels in the Bonriki aquifer due to unplanned groundwater extraction are a major concern for the people of Kiribati. A multi-objective management model capable of providing sustainable optimal groundwater pumping strategies, and simultaneously confining salinity concentrations in the aquifer within specified limits, are needed for the Bonriki aquifer system. This study applies a regional-scale linked simulation-optimization methodology with a Pareto front clustering technique to prescribe optimal groundwater withdrawal patterns from the Bonriki aquifer. A numerical simulation model is calibrated and validated using available field data. For computational feasibility, support vector machine regression (SVMR) surrogate models are trained and tested utilizing input-output datasets generated by a numerical flow and transport simulation model. The developed surrogate models were externally coupled with a multi-objective genetic algorithm (MOGA) optimization model as a substitute for the numerical model. The study area consisted freshwater pumping wells for extracting fresh groundwater. Pumping from barrier wells installed along the coastlines is also considered as a management option to hydraulically control saltwater intrusion. The multi-objective linked simulation-optimization (S/O) model generated 700 Pareto-optimal solutions. Analysing a large set of Pareto-optimal solutions is a challenging task for decision-makers. Hence, the *k*-means clustering technique is utilized to reduce the size of the original Pareto-optimal solution set to solve the large-scale saltwater intrusion management problem in the Bonriki aquifer.

4.2 Background

Kiribati is a small Pacific Island developing country that is heavily reliant on groundwater resources for its freshwater supply. The Bonriki aquifer is regarded as the national groundwater resource for Kiribati and supports the livelihood of Kiribati's urban population. Generally, freshwater aquifers in atoll islands exist as thin lenses (a few tens of meters in depth), contained in highly-permeable aquifers surrounded by the sea (White and Falkland 2010). These freshwater lenses are crucial for the survival of atoll settlements and any degradation in freshwater aquifers has serious consequences for atoll populations. Extracted freshwater from the Bonriki aquifer is mainly used to meet the growing demands of the South Tarawa island community. South Tarawa, the capital of Kiribati, is currently experiencing a huge population influx from smaller islands that have already lost the fight against saltwater intrusion. Excessive groundwater withdrawals from the Bonriki aquifer have resulted in saltwater intrusion (White and Falkland 2010). Saltwater intrusion has increased the salinity of this freshwater aquifer, thereby threatening its suitability for domestic use. The issue of saltwater intrusion caused by groundwater pumping requires a multi-objective management model capable of determining optimal groundwater pumping strategies so that the coastal aquifer remains sustainable.

Linked S/O utilizing trained surrogate models have the potential to provide the sustainable multiple-objective groundwater extraction strategies needed for the Bonriki aquifer system. Linked S/O models have delivered many management alternatives needed for the sustainable optimal management of coastal aquifers (Ataie-Ashtiani et al. 2013; Ketabchi and Ataie-Ashtiani 2015; Kourakos and Mantoglou 2013; Park and Shi 2015; Roy and Datta 2017; Sreekanth and Datta 2010). However, the implementation of an S/O model has proven to be a challenging and multifaceted task. In an S/O model, the objective function (a function that is anticipated to be minimized or maximized) is evaluated many times before an optimal solution (depending on the constraints) is reached. When a complex numerical model is linked to an optimization algorithm, the optimization process becomes time-consuming and computationally demanding. The alternative use of embedded linked S/O with finite element or finite difference equations embedded as constraints to simulate transient three-dimensional density-dependent flow and transport in coastal aquifers may become computationally infeasible for such systems (Das and Datta 1999). Also, externally linking an optimization model algorithm with complex numerical flow and transport simulation models to search for an optimal solution can be very difficult, cumbersome and computationally inefficient. The computational complexity is further increased when multiple objectives are considered (Akhtar and Shoemaker 2016). In a coastal aquifer management case study, the computational time required to generate a single solution on the Pareto front for a small study area when a numerical simulation model (FEMWATER based model) model was linked to an optimization model was several days of CPU time (Dhar and Datta 2009). The solution to this issue is to use a surrogate modelling paradigm instead of a complex numerical simulation model in an S/O framework.

A surrogate model can be understood as a "model of a model", which describes the relationship between the inputs (a model's adjustable parameters) and outputs (Wang et al. 2014). Substantial gains in computational efficiency and computational feasibility for large study areas can be achieved when a surrogate model-assisted S/O method is used. Many saltwater intrusion management studies based on linked S/O studies have used surrogate models and have delivered dependable results. A detailed description of the use of surrogate models in the field of water resources research is presented in Razavi et al. (2012). In the domain of saltwater intrusion simulation and prediction, a range of surrogate model types has been successfully developed and implemented (refer to Section 3.2 of Chapter 3). It is evident from the literature that a range of surrogate modelling techniques is available. However, the selection of a surrogate modelling technique and its accuracy are dependent on the data available and the approach used for surrogate model construction (training and testing). The present study utilises the support vector machine regression (SVMR) surrogate modelling tool for the prediction of salinity concentrations at monitoring locations in response to changing transient groundwater pumping patterns. SVMR is a relatively new technique from the field of artificial intelligence that has been successfully applied to several forecasting applications (Xia et al. 2017). The key advantages of SVMR predictive models are listed in Chapter 3 (Section 3.2).

One concern regarding the application of a multi-objective S/O-based saltwater intrusion management study is the interpretation and implementation of its results. The solution to a multi-objective optimization problem takes the form of a Pareto front, which consists of many solutions known as *non-dominated solutions* or *Pareto-optimal solutions* (Wang and Rangaiah 2017). Several studies in the field of saltwater intrusion management have obtained optimal solutions in the form of Pareto-optimal trade-offs dependent on the respective objective functions used in the proposed management model. All the solutions in the entire non-dominated Pareto-front can be used by the decision maker for selecting a single preferred solution as long as their preference ordering is identified before the multiple objective optimisation problems are solved. However, in various situations, a single optimal solution is not easily identified. Hence, selecting a solution from several hundred or thousands of Pareto-optimal solutions on the Pareto front is a difficult task for the decision maker. One way of alleviating this difficulty is to present the decision maker with a small number of solutions representative of the Pareto front's characteristics (Zio and Bazzo 2010). This allows easy selection of the optimal solution, consequently allowing the decision maker to achieve their optimization goal. A method for selecting a scalar optimum solution from a set of vector optimization solutions allowing decision makers to make informed decisions for conceivable saltwater intrusion control needs to be developed and implemented. A case study with an example of a search for a single compromise solution from a set of Pareto-optimal solutions for water management based on quantified required trade-off and decision maker's acceptable compromises was presented in Datta and Peralta (1986). Formulation of a method for selecting an optimal solution from a large set of Pareto-optimal solutions needs further attention. Such a method would provide decision-maker(s) with fewer representative solutions to choose from and aid the practical implementation of management policy.

Various preference elicitation and ranking methods can be used to help decision-maker(s) find solutions that best meet optimization goals in multiple-objective scenarios. Three different approaches are commonly applied to describe decision-maker preferences for choosing a single solution to vector optimization problems: *apriori*, *aposteriori* and *interactive* approaches (Branke and Deb 2005; Geiger 2006; Jafari et al. 2014). The *apriori* approach necessitates preference articulation in the formulation of the objectives. The *aposteriori* approach, on the other hand, involves identification of an optimal solution on the Pareto-front through analysis of the trade-offs represented by it, which are obtained after the optimization process. In the *interactive* approach, the decision maker can express their preference at various times during an ongoing optimization process. In general, the *aposteriori* approach is more logical, as the final choice of optimal solution is based on weighting of the trade-offs between the objectives (Datta and Peralta 1986). The *aposteriori* approach has been used in many optimization-related studies (Kao 2010; Pawar et al. 2017; Yadollahi et al. 2015). The *aposteriori* approach has been proven to be less subjective than the other two Pareto-optimal solution analysis techniques (Bui and Alam 2008). Also, employing an *aposteriori* approach allows the decision-maker to navigate through the Pareto-front (i.e., when there are changes in the decision-maker's preferences) without repeatedly executing the optimization algorithm (Yu et al. 2017).

The clustering technique is a common *aposteriori* approach used during the post-Pareto analysis stage of various optimization problems. Liong et al. (2004) used the clustering approach to analyse Pareto-optimal solutions in a reservoir optimization problem. Liong et al. (2004) established that the clustering approach aids in reducing the number of feasible solutions by offering a set of representative solutions distributed over the entire Pareto front. Taboada and Coit (2007) also used the clustering technique to prune the size of a Pareto-optimal solution set and obtained a smaller representation of the Pareto front in a system reliability optimization problem. Taboada and Coit (2007) emphasized that clustering enables decision-makers to easily select meaningful solutions for final implementation. In this study, the clustering technique is used to help decision-makers concentrate on a smaller set of optimal solutions representative of the Pareto front. Clustering helps in reducing the number of solutions in the Pareto-front that need to be explicitly evaluated by the decision-maker. Specifically, the *k*-means clustering technique is used to partition the Pareto front solutions that share common features. The *k*-means clustering technique is utilized in the present study because it is known to be efficient (Mattson et al. 2004). Using the *k*-means clustering approach reduces the number of solutions in the Pareto-optimal solution set to be the same as the number of clusters (Aguirre and Taboada 2011). Successful application of *k*-means clustering as a Pareto front analysis tool has been demonstrated in various studies (Aguirre and Taboada 2011; Bandyopadhyay and Maulik 2002; Chaudhari et al. 2010; Taboada and Coit 2008; Zio and Bazzo 2010). In *k*-means clustering, the centroid of each cluster provides the representative solution of that cluster and is used as a reference solution for comparison among all Pareto-optimal solutions, allowing decision-makers to choose from a smaller set of optimal solutions (Cheikh et al. 2010). The post-Pareto analysis allows the decision-maker to select the most preferred optimal solution by easily navigating through the trade-offs of the solutions in the reduced solution set.

This study develops a linked S/O model for prescribing computationally-efficient, regional-scale, multi-objective management strategies for a real-life coastal aquifer affected by saltwater intrusion. The SVMR technique is used for developing efficient surrogate models that approximate density-dependent saltwater intrusion processes in such aquifers. The efficiency and accuracy of SVMR-based surrogate models was evaluated for a hypothetical aquifer in Chapter 3, with the results reported in Section 3.4.1. This chapter also essentially deals with the application of the developed method to a regional-scale coastal aquifer in the Pacific Island of Kiribati. Here, groundwater is the most important source of freshwater and its contamination affects the local economy. This study addresses the goal of the linked S/O methodology while also highlighting the need for the implementation of an adequately calibrated and validated model that can predict the responses of the aquifer system to management strategies. In addition, this work also highlights some of the limitations of the method in terms of data availability and accurate modelling of the complex physical processes involved. The results and evaluations obtained are new and represent an important step in the application of management models to the sustainable management of regional-scale coastal aquifers.

The main objective of this study is to prescribe multiple-objective, optimal, freshwater pumping strategies for the sustainable management of the Bonriki aquifer in Kiribati utilizing a surrogate-assisted S/O model. The main aspects of this study are: (i) development (calibration and validation) of a 3D density-dependent saltwater intrusion model for the Bonriki aquifer, taking into consideration limited field data availability; (ii) application of SVMR-based, trained surrogate models for predicting saltwater intrusion to ensure the computational feasibility of the developed multi-objective management model at a regional scale; (iii) assessment of different management scenarios, such as the use of barrier well pumping to hydraulically limit saltwater intrusion into the Bonriki aquifer and (iv) demonstrate the application of the *k*-means clustering method for the selection of a subset of representative solutions from a potentially large Pareto-optimal solution set.

4.3 Study area

4.3.1 Location of the Bonriki aquifer and groundwater use

The Bonriki aquifer is situated in the Tarawa atoll (1°30′ N, 173°00′ E) in Kiribati. Kiribati consists of chains of small, low-lying, coral atoll islands and is situated in the central Pacific Ocean. Figure 4.1 illustrates the geographic location of Kiribati and the Bonriki aquifer. Kiribati has a tropical climate throughout the year and has a land area of approximately 800 km^2. Tarawa, the capital of the Republic of Kiribati, is divided into South Tarawa and North Tarawa. South Tarawa is an urban area and contains 48% of the nation's population (Duvat et al. 2013). The Bonriki aquifer system is located in South Tarawa and is the main source of reticulated water for the urban population of South Tarawa (White et al. 1999).

4.3.2 Hydrogeology of the Bonriki aquifer

The Bonriki aquifer is a vertically-heterogeneous system typical of atoll island aquifers (Ayers and Vacher 1986; Terry et al. 2013). The fresh groundwater in the Bonriki aquifer is largely contained in unconsolidated Holocene sediments that unconformably overlie older Pleistocene limestone (Bosserelle et al. 2015). Holocene sediments are moderately permeable with hydraulic conductivity values ranging from 5 m/d to 20 m/d, while the hydraulic conductivity of the Pleistocene limestone sediments is much greater (Bosserelle et al. 2015; White et al. 2008). The very high hydraulic conductivity of the Pleistocene sediments enhances the mixing of freshwater and seawater (Bosserelle et al. 2015). The unconformity between these two geological layers is found at 12–20 m below mean sea level (Bosserelle et al. 2015). Such unconformity is very important to the formation of aquifers and is, therefore, regarded as the main feature controlling freshwater reserves in Bonriki (Metai 2002). Fresh groundwater flows into the adjoining and underlying seawater, and a transition zone consisting of brackish water exists at the interface. The salinity levels at the transition zone progressively increase over several metres as the aquifer blends into seawater (Terry et al. 2013). The aquifer is bounded by the water table and the transition zone. The thickness of the transition zone oscillates depending on the rate of groundwater withdrawal and recharge through rainfall (Storey and Hunter 2010). It is estimated that the transition zone in Bonriki is approximately 23 m deep (Bailey et al. 2010; Falkland 1992).

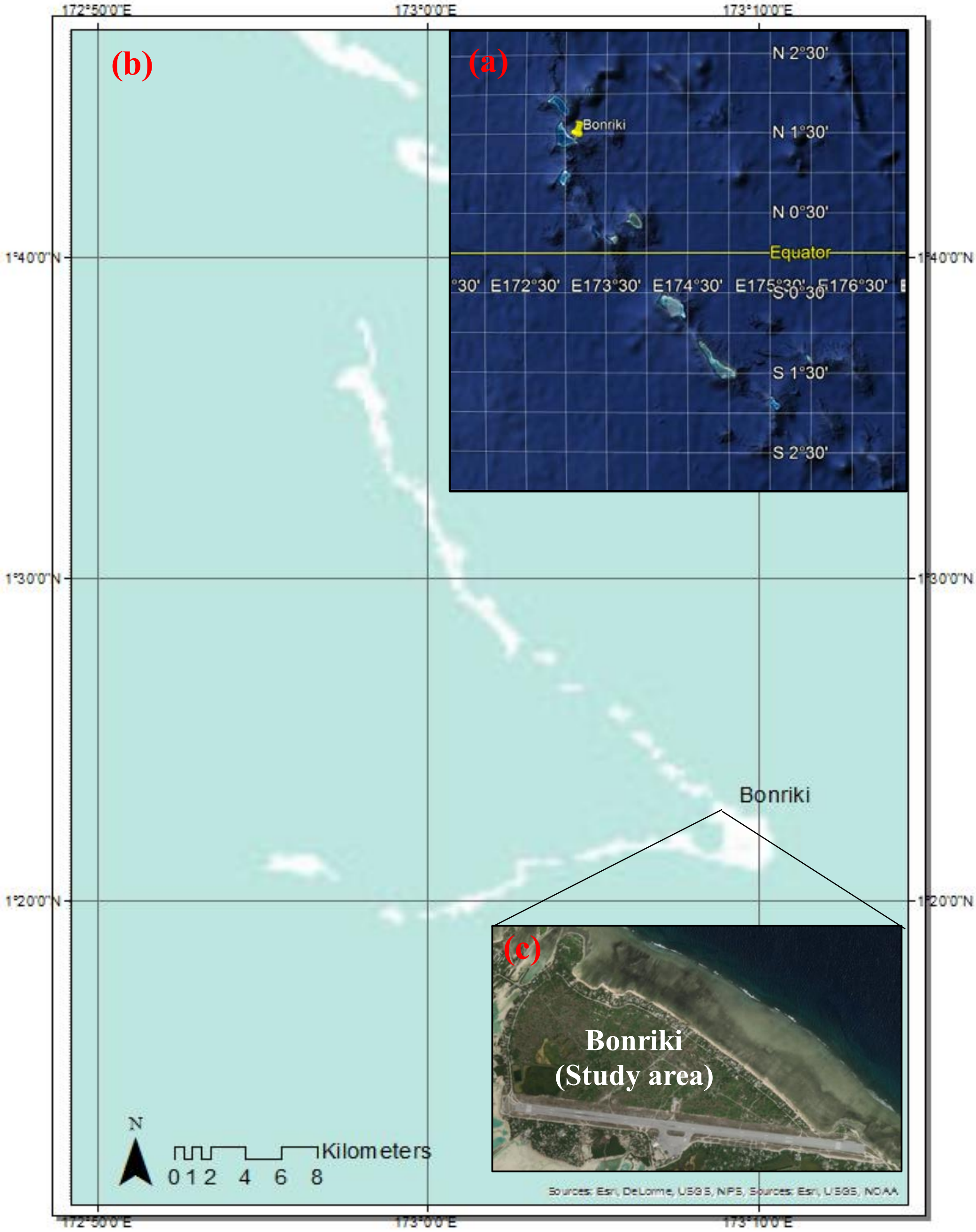


Figure 4.1: a) Geographical location of Kiribati, b) close-up view of the Tarawa Atoll and c) close-up view of the study area (Bonriki aquifer)

4.4 Methods

4.4.1 Saltwater intrusion numerical modelling tool

A 3D numerical model was developed using the FEMWATER computer package (Lin et al. 1997) to simulate pumping-induced salinity intrusion into the aquifer. The FEMWATER package allows simulation of density-dependent coupled groundwater flow and transport processes in aquifer systems. FEMWATER uses the Galerkin finite-element approximation and residual finite-element methods to approximate flow and transport equations, respectively. A detailed description of the FEMWATER model is given in Section 3.3.1 of Chapter 3.

4.4.2 Borehole data and lithology details

Two aquifer layers were considered in the modelling of the island aquifer, similar to that in Bailey et al. (2009). Borehole data from Bosserelle et al. (2015) was used to map the lithology of the study area. Borehole data was imported into Groundwater modelling system (GMS) software and interpolated using the GMS horizon package. The study area was structured with two distinct layers: Holocene sediments (HS) and Pleistocene sediments (PS). The Holocene sediments extended from 5–15 m in depth and overlaid the Pleistocene sediments. Hydrogeological data from 19 boreholes were imported and used in the interpolation of aquifer characteristics. The model domain area was 1.50 km^2 with a depth of 60 m.

4.4.3 Groundwater level, concentration and extraction data

Field-measured data for groundwater level, electrical conductivity (EC) and average groundwater abstraction (annual) from the Bonriki aquifer was obtained from Sinclair et al. (2015). Groundwater level and EC data for a period of 1.5 years were available (from April 2013 to August 2014). Groundwater level and EC data from six monitoring locations (MLs) were included in the calibration and validation processes. Concentration data (in mg/L) was obtained by converting EC (in µS/cm) by a factor of 0.69, similar to that described by Ghassemi et al. (1996) and Ghassemi et al. (1990). The converted concentration and groundwater level data from the six MLs, and groundwater abstraction data from 19 groundwater pumping wells, were used in the calibration and validation of the numerical model.

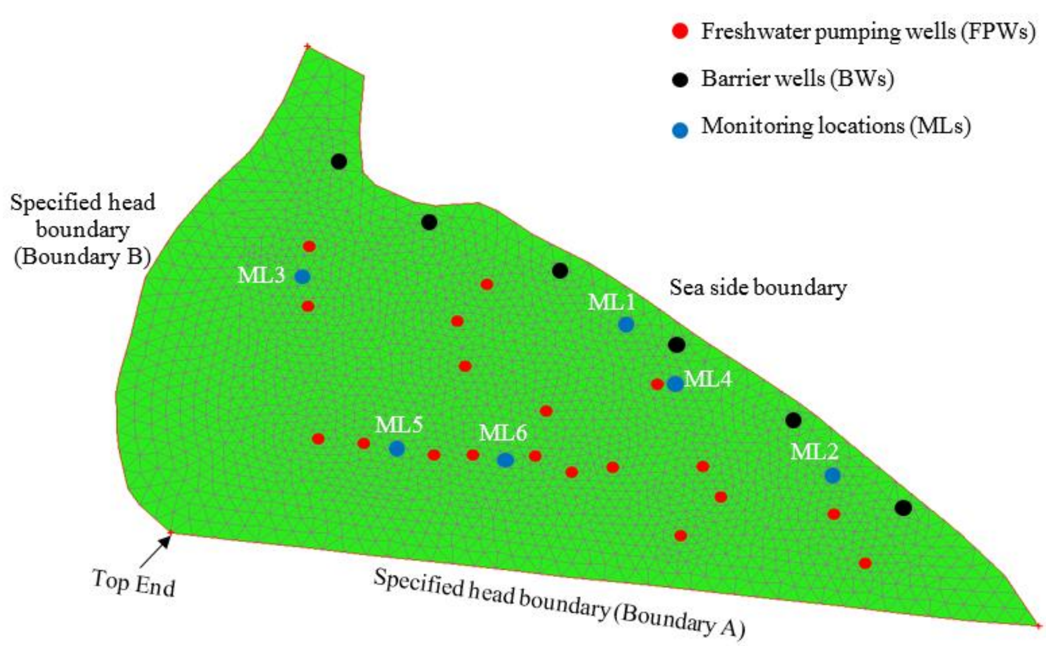
4.4.4 Boundary conditions and key aquifer parameters

The assigned boundary conditions (seaside, boundary A and boundary B) of the model domain are shown in Fig. 4.2 (a). The seaside boundary was in direct contact with the ocean and was defined as a constant head (Dirichlet boundary) and constant concentration boundary (seawater level and concentration). This means that the heads and salinity concentrations were specified and remained the same over time. A constant head of zero and a constant concentration of 35,000 mg/L was assigned to this seaside boundary. The sea level was taken as the zero level, as in many other previous studies. Specified pressure head boundaries were assigned to boundaries A and B because the heads along these two boundaries are not strictly zero. A pressure head of 1 m (at the top end) was assigned to boundaries A and B, and allowed to vary linearly along the boundary until it reached a constant value of 0 m at the seaside boundary. Groundwater recharge was represented by a constant vertical flux across the entire model domain. The rainwater recharging the study area was assigned a salinity of zero. In addition, each pumping well constituted an important source and/or sink for groundwater into the Bonriki aquifer system. In the finite element-based numerical model, the model domain was horizontally discretised into a mesh of triangular elements. The vertical discretisation was based on the lithological structure of the study area. A 3D model of the study area with specific well locations is shown in Fig. 4.2 (b). A recent survey by the World Bank stated that sea-level rise will have no significant impact on the Bonriki aquifer and will slightly increase its volume (Terry et al. 2013). Hence, the impact of sea-level rise was not incorporated in the 3D model. Also, uncertainties in aquifer parameter values and model predictions were not explicitly included in this study.

4.4.5 Calibration and validation of the flow and transport numerical simulation model

Monthly groundwater levels and EC data from six MLs for the period of April 2013 to August 2014 was available. This field data was divided into two sets: Set 1 consisted of 12 months' data (April 2013 to March 2014), which was used for model calibration. Validation was performed using data from Set 2, which consisted of five months' data (April 2014 to August 2014). The calibration process was performed manually based on a trial-and-error approach. For calibration, a transient approach was utilised in which the model was simulated for a period of 334 days (April 2013 to February 2014) in monthly time steps. Initially, during the calibration stage, the model groundwater level and concentration data from the numerical model did not match the field measurements. However, after gradually and iteratively modifying various parameters (hydraulic conductivity, recharge and porosity) within a reasonable range in the numerical simulation model, the correlation between observed and simulated values improved (in terms of R^2 values). The targeted R^2 value for calibration was > 90%. Other parameter values were based on previous field investigations, and some were obtained from the literature (refer to Table 4.1). Once a desired level of accuracy was achieved, the validation stage commenced. The parameter values of the calibrated aquifer model are listed in Table 4.1.

a)



b)

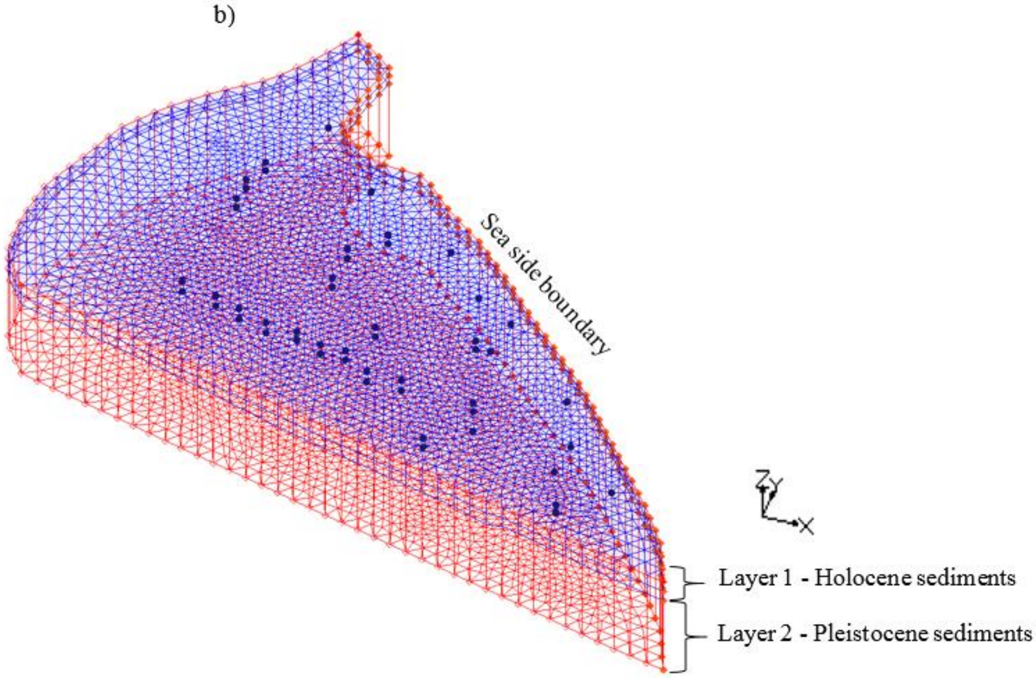


Figure 4.2: a) Study area with freshwater pumping wells, barrier wells and monitoring locations and b) developed finite element based 3D model

Table 4.1: Aquifer hydrogeological parameters.

Hydrogeological parameter		Holocene sediment	Pleistocene sediment	Source
	x	15	450	
Hydraulic conductivity (m/d)	y	7.5	225	Calibrated
	z	1.5	45	
Porosity (%)		20	30	Calibrated
Effective recharge (m/d)		0.0055		Calibrated
Seawater density (kg/m^3)		1025		Oberdorfer et al. (1990)
Freshwater density (kg/m^3)		1000		Oberdorfer et al. (1990)
Molecular diffusivity (m^2/s)		1.5 x 10^{-9}		Ghassemi et al. (1996)
Dynamic viscosity of water (kg/ms)		280985.76		-
Longitudinal dispersivity (m)		1		Bosserelle et al. (2015)
Lateral dispersivity (m)		0.05		Bosserelle et al. (2015)
Compressibility of water (m^2/N)		4.4 x 10^{-10}		Oberdorfer et al. (1990)

4.4.6 Surrogate models as approximate simulators of saltwater intrusion processes

The SVMR prediction algorithm is a promising tool that has been used in various non-linear predictive modelling studies (Al-Anazi and Gates 2012; Gizaw and Gan 2016; Liu et al. 2013). For the present study, the values of C and ε and the kernel function were obtained through trial-and-error in a process similar to that of Suganyadevi and Babulal (2014). Specifically, various combinations of SVMR parameters and kernel functions were used to construct SVMR models. Combinations that provided strong correlations (R^2 > 90 %) between simulated and predicted values were selected.

4.4.7 Generation of input-output training and testing datasets

The calibrated numerical model was used to generate sets of transient pumping (input)-concentration (output) datasets for use in surrogate model development. A total of 700 transient pumping patterns from freshwater pumping wells (FPWs) and barrier wells (BWs) were obtained from a uniform sampling distribution using the LHS methodology. For the illustrative coastal aquifer management problem investigated in the previous chapter (Section 3.3.4.2), 700 pumping and concentration datasets were found to be sufficient for training and validating SVMR surrogate models with reasonable predictive accuracy. Also, the numbers of training and testing datasets required are dependent on the predictive performance of each surrogate model type. The numbers can be increased or decreased depending on the prediction capabilities of the models, which can be deduced from performance evaluation. The 700 pumping patterns generated were fed into the numerical simulation model (one set at a time) and the output concentrations at MLs were obtained for each input pumping pattern. Each numerical simulation model took approximately 4-5 minutes to converge. The 700 input pumping patterns and resulting 700 output ML concentrations were generated by running the simulation 700 times. These input-output patterns were used to train and test the SVMR surrogate models.

4.4.8 Surrogate model development procedure

Out of the 700 data sets assembled, 500 were used for training the models and 200 were used for testing them. Six surrogate models were trained for predicting concentrations at the six corresponding MLs. After training, each of the six surrogate models was tested using the corresponding testing dataset. The SVMR models were trained and tested offline using the MATLAB 2017a platform. An offline training and testing paradigm was employed because it can deliver reliable saltwater intrusion surrogate models with low computational complexity. The surrogate models were named SVMR1, SVMR2, SVMR3, SVMR4, SVMR5 and SVMR6, where the numbers represent the corresponding MLs. In constructing the predictive SVMR models, a Gaussian kernel was used with the values of parameters ε, C and γ set to 0.0004, 10 and 0.05, respectively. These parameter values were obtained after several trial-and-error runs.

4.4.8 Surrogate model performance evaluation

Assessing the performance of the surrogates was decisive in determining the accuracy and reliability of the developed models. The performance evaluation criteria used to quantify the saltwater intrusion prediction capabilities of each SVMR model during the training and testing stages are given by Eqs. 4.1 to 4.4. The root mean square error (RMSE) was calculated to measure the difference between the predicted values (SVMR model) and the simulated values (numerical model). An SVMR surrogate model was considered acceptable when the RMSE was < 5 mg/L. Mean bias error (MBE) can also represent the difference between predicted and simulated values. Coefficients of determination (R^2) were calculated to indicate the degree of association between the predicted and simulated values. Nash-Sutcliffe efficiency coefficients (NSE) were calculated to assess the predictive capabilities of the hydrological models.

$$RMSE = \sqrt{\frac{1}{N}\sum_{n=1}^{N}(Sn - Pn)^2} \tag{4.1}$$

$$MBE = \frac{1}{N}\sum_{n=1}^{N}(Sn - Pn) \tag{4.2}$$

$$R^2 = \frac{\left(\sum_{n=1}^{N}(Sn - \overline{Sn})(Pn - \overline{Pn})\right)^2}{\sum_{n=1}^{N}(Sn - \overline{Sn})^2 \sum_{i=1}^{N}(Pn - \overline{Pn})^2} \tag{4.3}$$

$$NSE = 1 - \frac{\sum_{n=1}^{N}(Sn-Pn)^2}{\sum_{n=1}^{N}(Sn-\overline{Sn})^2} \tag{4.4}$$

In the above equations, Sn is the simulated concentration, Pn is the predicted concentration, $\overline{Sn}$ is the average simulated concentration, $\overline{Pn}$ is the average predicted concentration and N represents the number of data points.

4.4.9 Linked simulation-optimization-based management model

The urban population of South Tarawa is heavily reliant on freshwater from Bonriki aquifer. Thus, the main aim of the proposed management model was to prescribe optimal groundwater pumping rates from the FPWs that meet the growing freshwater demand while simultaneously restricting the salinity levels in the aquifer to specified limits. The decision variables within the approximate simulators (surrogate models) represent the pumping rates at each of the FPWs and BWs. The objective function measures the performance of the different pumping solutions while satisfying the set constraints. A four-year management horizon was chosen for this illustrative evaluation, taking into consideration the number of input patterns of FPW and BW pumping rates to be included for training the SVMR surrogate model. If the aquifer system was considered to be in a steady-state in terms of head and salt concentration, the management period would not need to be restricted. However, in this study, both pumping rates and heads are considered transient. Therefore, training of the SVMR surrogate models (which approximate the response of the aquifer to various pumping stresses) was dependent on the initial conditions, time-varying concentrations and heads during the training period. Therefore, the random patterns generated as inputs, and their corresponding outputs, needed to consider input patterns that covered the entire management time horizon as paired sets. For training or learning purposes, an input pattern constituted a randomized pumping rate at each FPW and BW, with the rate varied for each management period over the entire management time horizon. Hence, to restrict the training process to reasonable computation time, a four-year management time horizon was chosen. While longer horizons could be considered, this would require larger pumping-concentration datasets for SVMR model training, which would likely increase the complexity of the S/O models. A total of 100 decision variables were considered for the management of Bonriki aquifer. These decision variables correspond to pumping rates from the FPWs and BWs during the four-year management period (4 time-steps). Variables P1-P19, P26-P44, P51-P69 and P76-P94 represented freshwater withdrawal rates from the 19 FPWs, while variables P20-P25, P45-P50, P70-P75 and P95-P100 represented pumping rates from the six BWs for the four management periods. The management model was formulated by coupling the approximate simulators to an optimization algorithm. For the present study, a multi-objective genetic algorithm (MOGA) (Deb et al. 2002) was used to solve the multi-objective optimization problem. MOGA has been successfully implemented as an optimisation tool in various saltwater intrusion management studies, including Dhar and Datta (2009), Sreekanth and Datta (2010) and Gad and Khalaf (2013). The working principle of MOGA is extensively discussed in Deb (2001) and Coello et al. (2007). To execute the MOGA model, a population size of 2000, function tolerance of 1×10^{-4}, constraint tolerance of 1×10^{-3} and crossover fraction of 0.8 was used in the developed optimisation problem. These MOGA parameters were obtained after several trial runs. Function tolerance and constraint tolerance were used as the stopping criteria of the optimization problem. The tested SVMR models were externally coupled to the MOGA via the MATLAB 2017a platform. The surrogate models presented candidate groundwater pumping rate (from the FPWs and BWs) solutions to the MOGA. This allowed the optimization algorithm to search for solutions whereby the only way to improve a particular objective was to decrease the performance of another conflicting objective. Mathematical expressions of the objective function, constraints and bounds are given below.

Objective I: Maximize pumping from FPWs

Maximize:

$$F_1(P_{FPW}) = \sum_{n=1}^{L} \sum_{t=1}^{T} FPW_n^t \tag{4.5}$$

Objective II: Minimize pumping from BWs

Minimize:

$$F_2(P_{BW}) = \sum_{m=1}^{M} \sum_{t=1}^{T} BW_m^t \tag{4.6}$$

Subject to:

$$\text{Constraints} \quad c_i = \xi(FPW, BW) \tag{4.7}$$

$$c_i \leq c_{max,i} \; \forall i, T \tag{4.8}$$

$$\text{Bounds} \qquad FPW_{min} \leq FPW_n^t \leq FPW_{max} \tag{4.9}$$

$$BW_{min} \leq BW_m^t \leq BW_{max} \tag{4.10}$$

Where FPW_n^t denotes pumping from the n^{th} freshwater pumping well at time t and BW_m^t denotes pumping from the m^{th} barrier well at time t. c_i represents the salinity concentration at the i^{th} monitoring location at the end of management the time horizon. $\xi(,)$ symbolizes that the surrogate model replaces the numerical simulation model, while constraint (4.7) denotes the coupling of the surrogate model with the optimisation algorithm within the S/O framework. Variables *L, M* and *T* are the total numbers of FPWs and BWs and the length of the management time horizon, respectively. Inequality (4.8) represents the imposed constraints, which ensure that the salinities at the MLs are within specified limits. Inequalities (4.9) and (4.10) represent the upper and lower bounds of pumping rates from FPWs and BWs, respectively. For the present case, salinities at MLs were maintained below specified limits to ensure that the water withdrawn from the 19 FPWs was suitable for domestic consumption. The maximum tolerable salinity for ML1 and ML2 was set to 20,000 mg/L. These locations are closer to the shoreline and the sea-side boundary is assigned a concentration of 35,000 mg/L. Therefore, very low permissible concentrations at ML1 and ML2 may not be feasible if the time horizon considered is limited to a few years. The maximum allowable salinities at ML3 and ML4 were set to 5000 mg/L and 4000 mg/L, respectively. Lastly, the maximum tolerable salinity at ML5 and ML6 was set to 450 mg/L. ML5 and ML6 were located in an area with a dense distribution of FPWs and it was anticipated that water withdrawn from this region was suitable for a range of activities of the South Tarawa community (household chores and irrigation). The upper and lower pumping rate bounds for both the FPWs and BWs were set to 1200 m^3/day and 0 m^3/day, respectively. The search space was restricted by these bounds. The pumping rate of 1200 m^3/day was just an upper bound. The optimal solutions were, however, based on the objective functions and the set constraints and should lie within these bounds. It is also possible that the upper bound of 1200 m^3/day would not be reached in the prescribed optimal solutions. Further justification of the upper bound is presented in the Results and Discussion sections.

4.4.10 Clustering of Pareto-optimal solutions and decision-making

In this study, *k*-means clustering was used to group the Pareto-optimal solutions into clusters with similar characteristics. The clustering was done by calculating the centroid of each cluster and iteratively assigning each solution to the cluster with the closest centroid (Taboada and Coit 2008). Detailed descriptions of the *k*-means clustering algorithm are presented in Bandyopadhyay and Maulik (2002) and Zio and Bazzo (2010). However, a brief outline of the pruning of the Pareto-optimal solution set using the *k*-means clustering technique is presented below.

Step 1: Obtain the Pareto-optimal solution set by executing the multi-objective optimization problem.

Step 2: Apply the *k*-means clustering technique to group the optimal solutions enclosed in the Pareto front into separate clusters.

Step 3: Once the *k*-means stopping criterion is reached, locate the centroid of each cluster.

Step 4: Select the solution in the Pareto-front closest to the centroid. The reduced optimal solution set will contain as many solutions as the number of centroids.

Step 5: The decision-maker is now provided with a reduced number of optimal solutions representative of the original Pareto-optimal solutions. A solution from the reduced Pareto-front is chosen based on the decision-maker's preferences and is implemented to achieve the optimization goal.

The *k*-means clustering code was written in the *R* software package. A fixed number of iterations was taken as the stopping criterion (Zio and Bazzo 2010); in this case, 50. The rule-of-thumb approach (Cui et al. 2014; Kodinariya and Makwana 2013) (Eq. 4.11) was used to decide the number of clusters required.

$$k \approx \sqrt{j/2} \tag{4.11}$$

Where j represents number of optimal solutions in the original Pareto-front.

A flowchart of the implemented management strategy utilizing the linked S/O model and the Pareto-front clustering technique is given in Fig. 4.3.

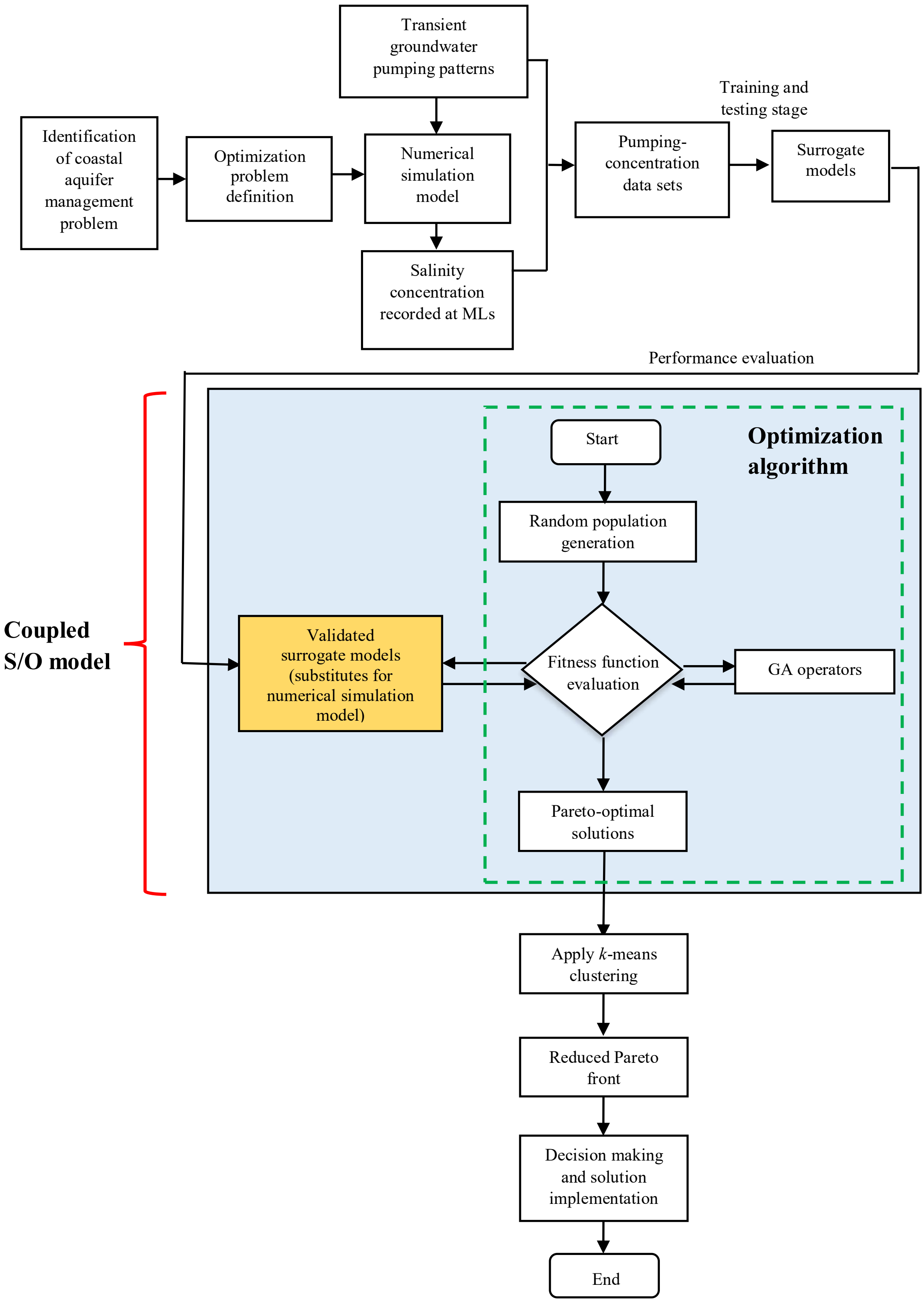


Figure 4.3: Multi-objective management model development framework

4.5 Results and discussion

4.5.1 Calibration and validation results: Observed vs simulated

During both the calibration and validation stages, the simulated groundwater levels and salinity concentration data at the respective MLs were in good agreement with the observed field values. The close agreement between the observed and simulated values indicates that the numerical model accurately replicated saltwater intrusion processes in the Bonriki aquifer. An acceptable level of agreement ($R^2 > 90$ %) between the simulated and field values was obtained. Before calibration, the R^2 values for the correlations between the simulated and observed groundwater levels and salinities were in the ranges of 35–45 % and 40–45 %, respectively. Comparisons of the simulated and observed groundwater levels and concentration at the validation stage are presented in Figs. 4.4 and 4.5, respectively. The R^2 values for the simulated vs. observed groundwater levels at the six MLs were within the range of 93.86–99.2 %, while those for salinities were 98.7–99.9 %. This was perceived to be an acceptable level of accuracy, as the differences between the simulated and observed values were < 10 % of the variability in the field data across the model domain (Khadri and Pande 2016). Also, as per the upheld calibration criteria stated in the methodology, it was observed that the simulated data not only agreed satisfactorily with the observed data in terms of accuracy, they also correlated with the observed patterns of fluctuation in salinity. The observed fluctuations in salinity due to variations in pumping rate were also reflected by the numerical model. While it is obvious that there are some uncertainties involved in the 3D modelling of the saltwater intrusion process in the Bonriki aquifer, the overall results suggest that manual calibration allows the numerical model to reflect the actual regional characteristics of groundwater flow and transport processes in the Bonriki aquifer with reasonable accuracy.

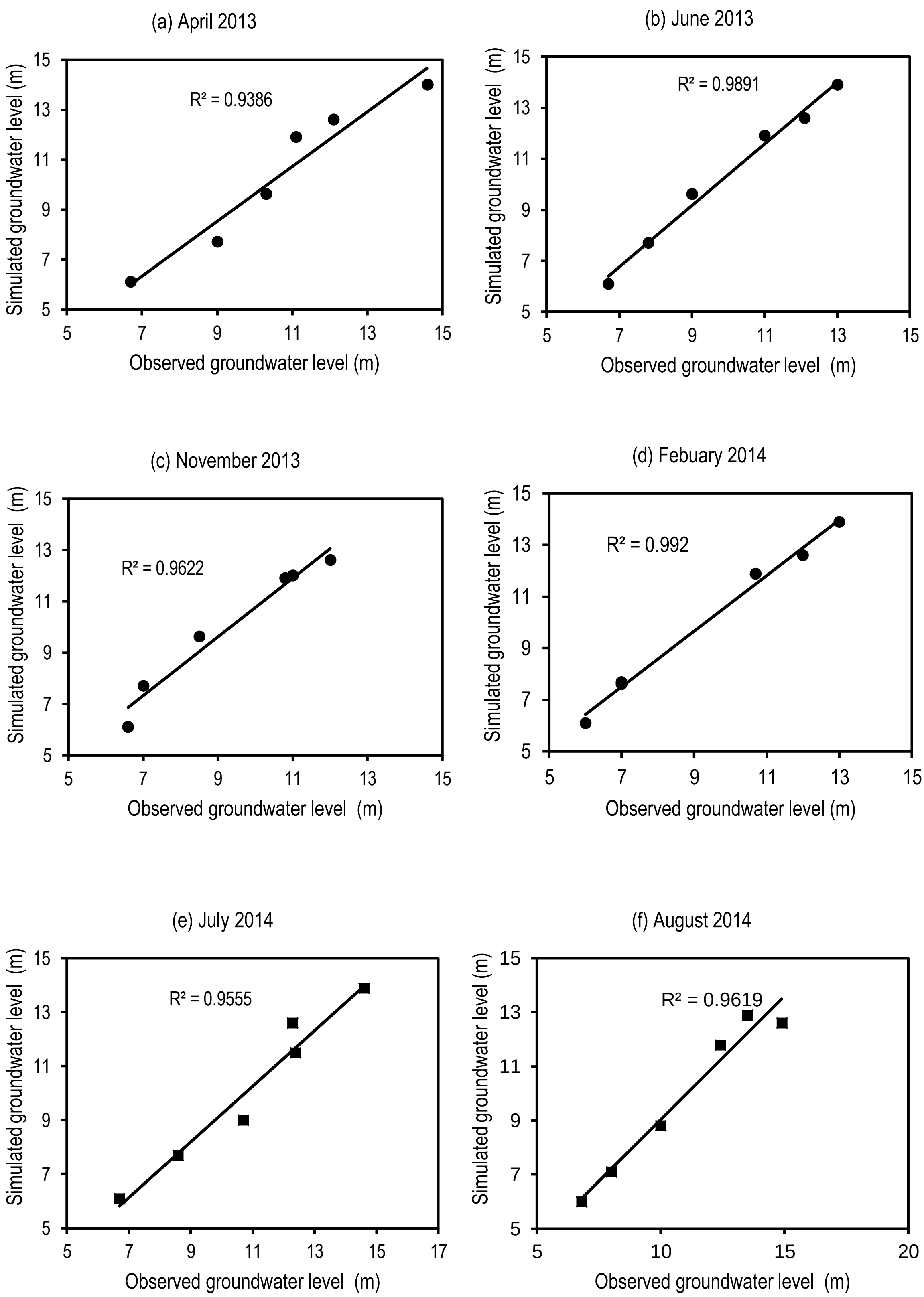


Figure 4.4: Correlations between observed and simulated groundwater levels obtained during the model calibration (a, b, c and d) and validation stages (e and f)

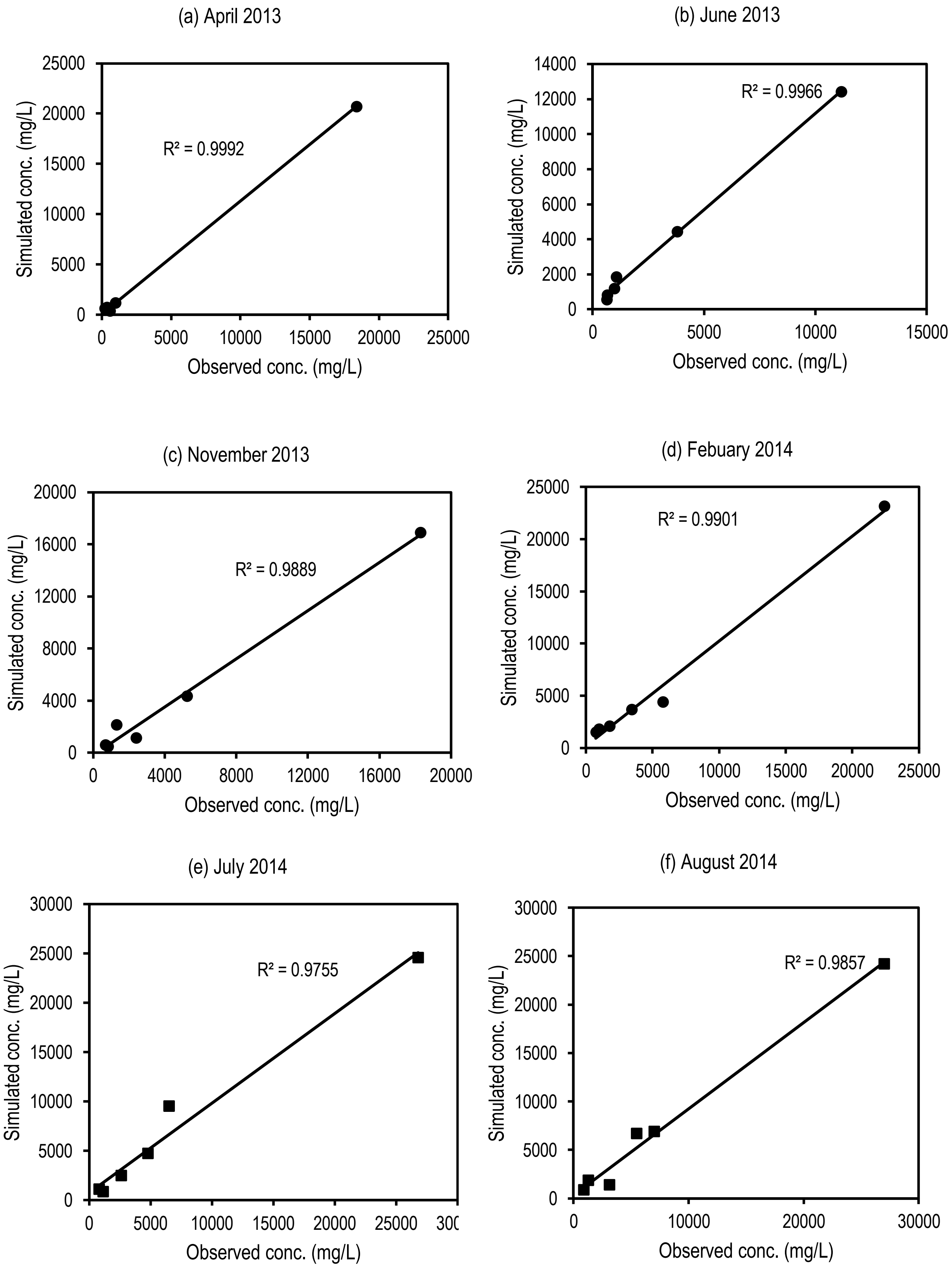


Figure 4.5: Correlations between observed and simulated concentrations (conc.) obtained during the model calibration (a, b, c and d) and validation stages (e and f).

4.5.2 Effect of barrier well pumping on the Bonriki aquifer

Before incorporating BWs into the S/O model, it was important to compute their potential benefit in minimizing saltwater intrusion into the Bonriki aquifer. Pumping from BWs installed near a coastline is a common approach to controlling saltwater encroachment into fresh groundwater and has been applied in several cases worldwide (Kallioras et al. 2013; Sreekanth and Datta 2010). BW pumping creates a trough along the shoreline, causing the seawater to flow inwards and the freshwater to flow in the opposite direction; i.e., towards the sea (Todd 1974). This results in a hydraulic barrier which can reduce saltwater intrusion into the freshwater system. In the present study, a series of BWs were installed to minimize saltwater intrusion into the Bonriki aquifer. To establish and quantify its benefit, five randomized transient pumping rates (before and after initiating BW pumping) obtained by LHS were implemented in the numerical model. The salinities obtained at the respective MLs are presented in Fig. 4.6. For the five pumping rate sets, reductions in salinity concentration were recorded at all six MLs after pumping was initiated at BWs. Analysis of the salinity data, in terms of maximum, upper quartile, median, lower quartile and minimum values (both before and after BW pumping initiation) clearly demonstrates that BW pumping has a positive impact on the Bonriki aquifer.

The evaluated results highlight that BWs can serve as a practical saltwater intrusion control method where feasible. However, it can be noted that any pumping rate at any of the specified BW locations is indeed a decision variable. An optimal solution can also specify a zero pumping from one or, all barrier well pumping locations if the optimization model infers that no further pumping from the BWs can increase the beneficial pumping from the FPWs. Also, in scenarios with two conflicting objectives, decision-makers can choose a Pareto-optimal solution which specifies zero BW pumping.

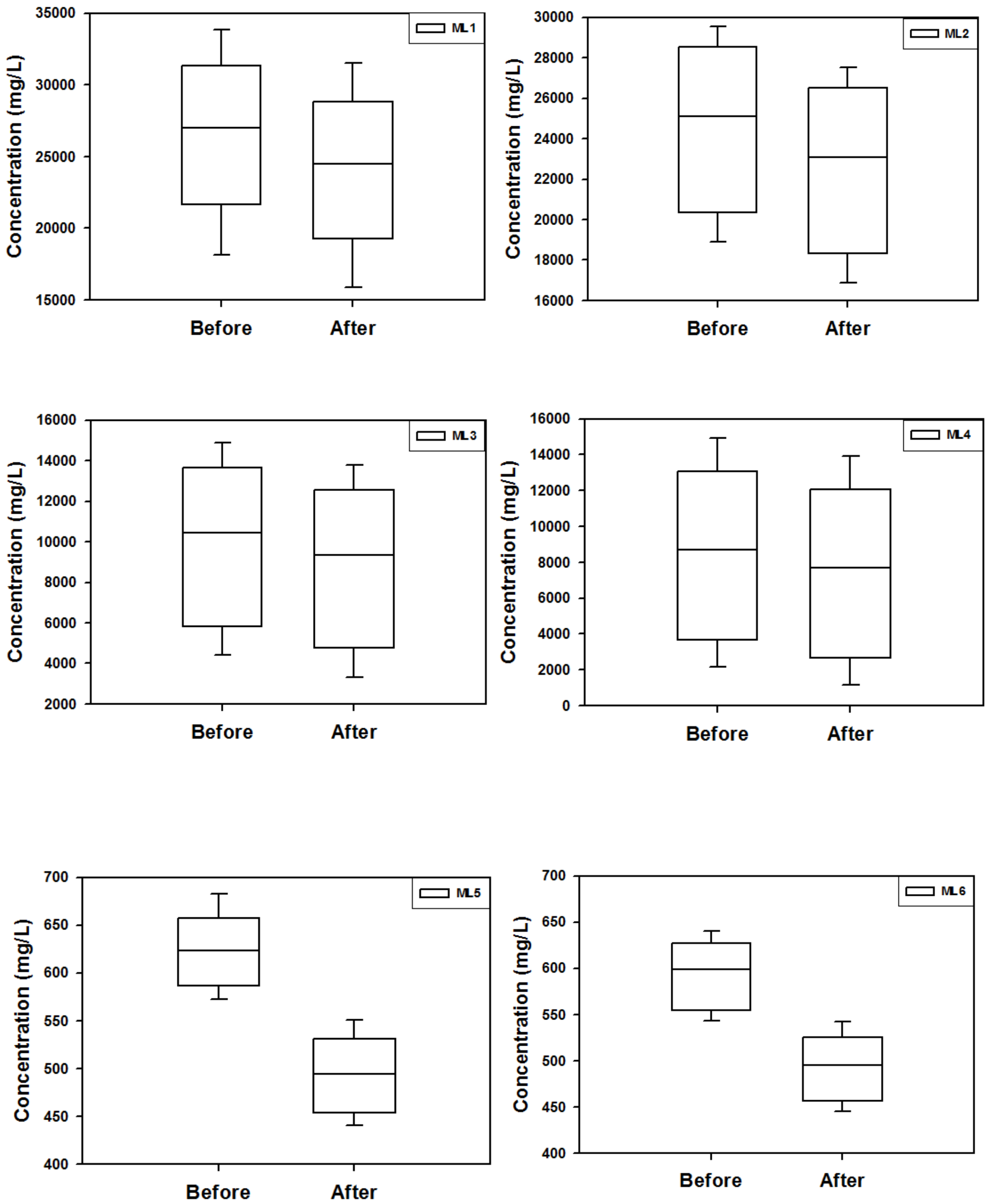


Figure 4.6: Comparisons of concentrations at MLs before and after BW pumping

4.5.3 SVMR surrogate model performance evaluation

The performance of the SVMR models at the training and testing phases is summarized in Table 4.2. The performance evaluation results obtained for both phases show similar trends in terms of the four evaluation criteria. The obtained results indicate that the models were adequately trained to predict salinity at the MLs in response to changing pumping rate patterns at the BWs and FPWs. The RMSE values for the developed models at both phases were substantially smaller and ranged from 2.75–7.14 mg/L). Accordingly, the MBE values of the SVMR models were comparatively smaller. Also, the R values for all six models were close to 1 and ranged from 0.96–0.99. The R values computed for the six SVMR models indicate the strength of the linear relationship between the actual and predicted salinity values. Taylor (1990) reported that an R-value close to 1, regardless of the direction, indicates a strong linear relationship between two variables. Also, the NSE values obtained indicate the models are reliable in their predictions. An NSE value of 1 indicates a perfectly trained predictive model (He et al. 2014). Shu and Ouarda (2008) stated that a model can be considered accurate if its NSE value is > 0.8. In the present case, the NSE values for all six SVMR models were either 0.99 or 1. This indicates an adequately-trained predictive model. Figure 4.7 shows the comparison between actual and predicted salinities at the testing stage. For easy visualization, Fig. 4.7 only shows 20 randomly-selected points instead all 200 testing points. The results at the testing stage indicate that the developed models have good predictive performance.

Table 4.2: Results of performance evaluation of the developed SVMR models

Phase	Model	RMSE	MBE	R	NSE
Training	SVMR 1	4.62	0.21	0.99	1
	SVMR 2	2.75	0.12	0.99	1
	SVMR 3	3.58	0.16	0.99	1
	SVMR 4	5.67	0.24	0.99	1
	SVMR 5	3.59	0.16	0.99	1
	SVMR 6	3.80	0.17	0.99	1
Testing	SVMR 1	5.66	0.40	0.98	1
	SVMR 2	4.95	0.35	0.98	1
	SVMR 3	5.30	0.38	0.96	0.99
	SVMR 4	5.73	0.41	0.96	0.99
	SVMR 5	7.07	0.50	0.98	0.99
	SVMR 6	7.14	0.51	0.98	0.99

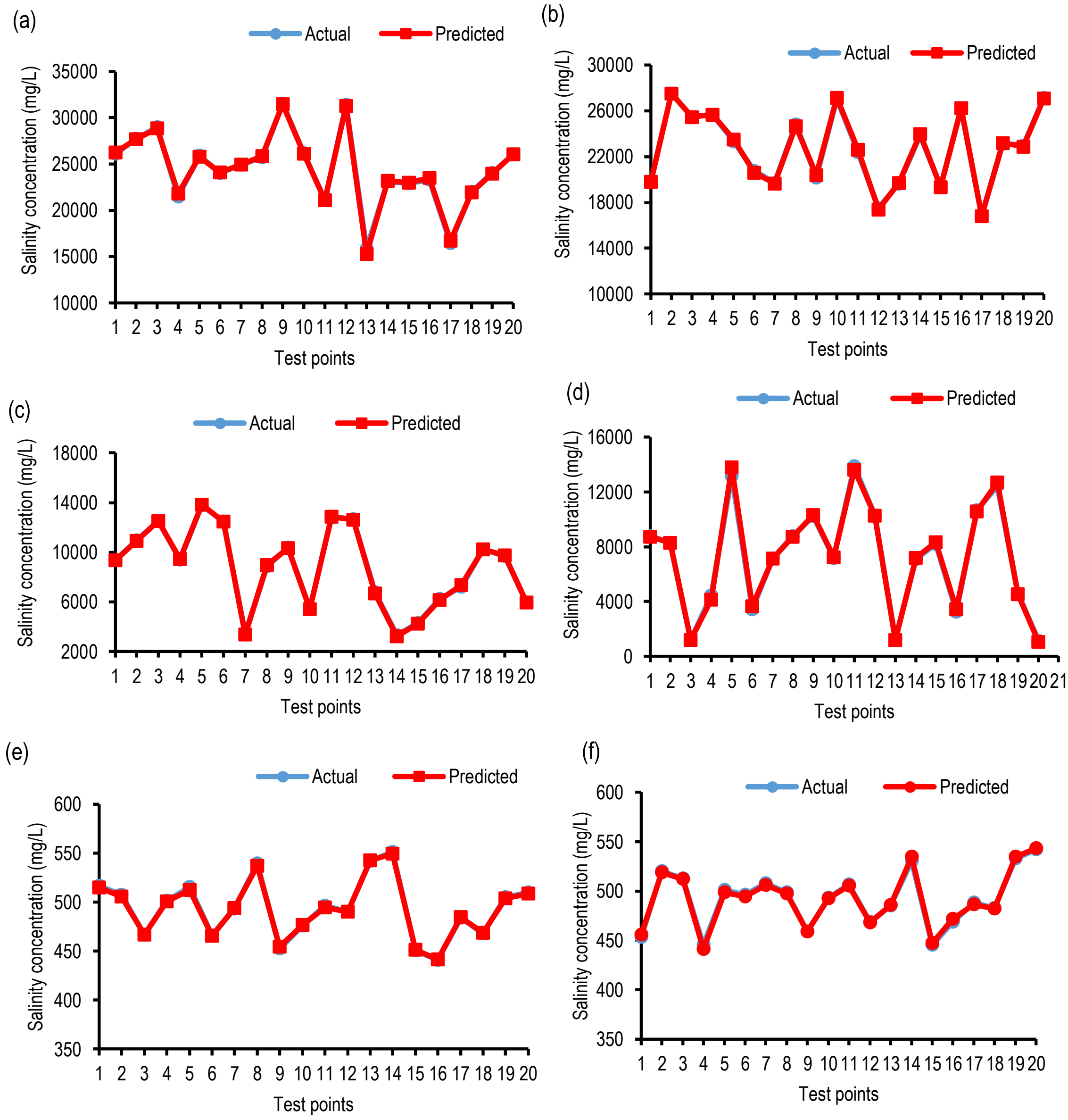


Figure 4.7: Comparisons of actual and predicted salinity values using 20 test points for a) ML1, b) ML2, c) ML3, d) ML4, e) ML5 and f) ML6

Overall, the performance evaluations indicate that the SVMR models effectively learnt the non-linear relationship between the groundwater pumping rate and salinity datasets. This highlights that the developed SVMR models are capable of emulating the numerical simulation model's responses to variable transient groundwater pumping rate patterns. Hence, the SVMR models can be used as an approximation of the more complex numerical model.

4.5.4 Pareto-optimal trade-offs

The execution of the linked S/O model presented a Pareto-optimal solution set comprising various solutions depending on the trade-offs between conflicting objective functions. Use of a population size of 2000 and a Pareto front population factor of 0.35 generated 700 optimal solutions in the Pareto-front. Figure 4.8 (a) presents the Pareto front of the executed S/O model. A clear trade-off is observed between the total FPW and total BW pumping rates. The Pareto-front reveals the rate of BW pumping required to achieve a preferred rate of FPW pumping. The total optimal pumping rates from all FPWs and all BWs for the four-year management horizon ranged between approximately 40,000–47,000 m^3/day and 2000–8000 m^3/day, respectively. These optimal pumping rates are based on the permissible salinity limits specified for the different MLs in the management model. For example, a total optimal pumping rate of 40,000 m^3/day from the FPWs is the total pumping rate from all 19 FPWs for the four-year management horizon. Annually, the pumping rate would be around 10,000 m^3/day (40,000 m^3/day divided by 4 years). This rate implies a total annual withdrawal of water from the aquifer of nearly 3.65 million m^3 per year (10,000 m^3/day × 365 days). The soil layer above the aquifer is quite permeable and consists of sand and gravel. The maximum and minimum annual rainfalls in Tarawa are approximately 4300 mm and 2100 mm, respectively (Bosserelle et al. 2015). With this annual rainfall over a highly-permeable aquifer top cover with a proportionately very small built-up area, it is reasonable to assume an average vertical annual recharge rate of nearly 2000 mm. This vertical recharge amount is itself around 3 million m^3 per year. Therefore, if the BW extraction rate is excluded from the total withdrawal amount computed above (FPWs plus BWs), and as a large proportion of BW extraction is contributed to by the sea face constant head boundary, the total specified withdrawal from the freshwater supply wells nearly matches the vertical recharge estimate. Therefore, the solutions and the boundary conditions imposed appear to be reasonable. To demonstrate specific pumping patterns from each well for each year in the management period, an optimal solution was randomly chosen from the Pareto front (Fig. 4.9). Similar pumping patterns were observed for other solutions on the Pareto-front. In Bonriki, the average withdrawal rate (field pumping rate) averaged over all FPWs for a one-year period was approximately 85 m^3/day per well. There were 19 FPWs; therefore, the annual average pumping rate for the entire aquifer was 1600 m^3/day (85 m^3/day × 19). This amounts to 584,000 m^3/year (1600 m^3/day × 365 days) from the entire aquifer. The average pumping rate from a single well for a year would be 31,000 m^3/year (85 m^3/day × 365 days). The upper bound of 1200 m^3/day was only used as a management upper bound to consider the option of utilising BWs in the search for an optimal solution. It does not reflect actual solutions but makes such an option plausible in the search for acceptable solutions. Essentially, this upper bound with a lower bound of zero defines the restrictions on the feasible decision space.

However, as is evident from Fig. 4.9, a typical average optimal FPW pumping rate is only about 450 m^3/day at the end of the management horizon. The upper bound is never reached in any solution. This increase in pumping rate is possible and also intuitively logical due to the option of utilizing BWs and also because it is based on a planned optimal strategy. Depending on the decision-maker's preference, a solution from the 700 optimal solutions available can be selected and implemented for the Bonriki aquifer. However, choosing a solution from a Pareto-front with a uniform spread and wide coverage is a challenging task for the decision-maker.

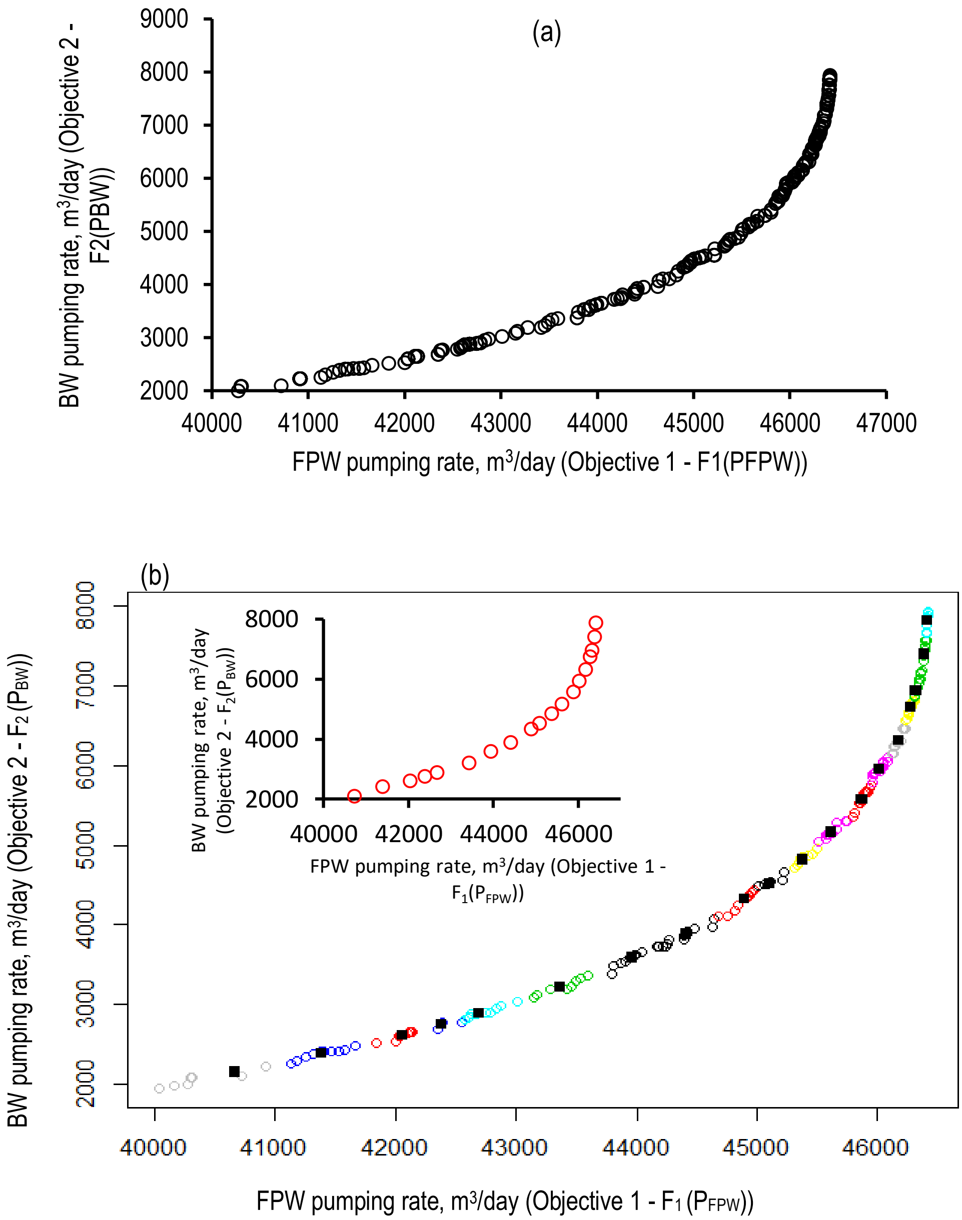


Figure 4.8: a) Pareto front from the SVMR-MOGA management model and b) Pareto front with clusters and centroid locations. Insert: Reduced Pareto front.

4.5.5 Reduced Pareto front

To facilitate decision-making, 700 solutions on the Pareto front were grouped into 19 clusters using *k*-means clustering. These 19 clusters presented 19 centroids, which were used as reference points. The solution closest to these reference points forms part of the reduced Pareto front. Figure 4.8 (b) presents the clustered Pareto front with centroid locations and the reduced Pareto front. The centroids with the closest corresponding Pareto-optimal solutions are presented in Table 4.3. For the present case, 19 centroids from the 19 clusters delivered 19 Pareto-optimal solutions. The decision-maker now has 19 optimal solutions which best represent the Pareto front under analysis. The decision-maker now has the option to choose solutions from the reduced Pareto front and decide on a strategy for optimal extraction of groundwater from the Bonriki aquifer.

Table 4.3: Cluster centroids with corresponding solutions in the reduced Pareto front

Solution	Centroid		Solutions in the reduced Pareto front (closest to centroid)	
	Total FPW pumping (m^3/day)	Total BW pumping (m^3/day)	Total FPW pumping (m^3/day)	Total BW pumping (m^3/day)
1	40664.92	2156.72	40719.89	2101.82
2	41391.71	2392.14	41383.78	2407.41
3	42056.11	2613.22	42035.38	2602.83
4	42386.75	2756.35	42380.53	2764.25
5	42692.82	2888.45	42677.52	2888.64
6	43363.04	3218.48	43420.89	3200.24
7	43959.38	3593.71	43939.32	3588.87
8	44407.63	3886.25	44408.37	3879.08
9	44897.02	4323.91	44891.49	4330.59
10	45104.19	4524.89	45085.43	4517.93
11	45379.91	4821.59	45366.83	4831.36
12	45619.88	5168.69	45617.43	5159.23
13	45871.98	5580.98	45877.21	5572.04
14	46013.52	5962.60	46022.76	5933.92
15	46171.84	6313.68	46168.41	6317.32
16	46277.27	6735.05	46278.19	6745.76
17	46319.39	6943.08	46317.06	6939.55
18	46384.30	7400.16	46386.55	7395.34
19	46410.97	7824.99	46410.37	7859.84

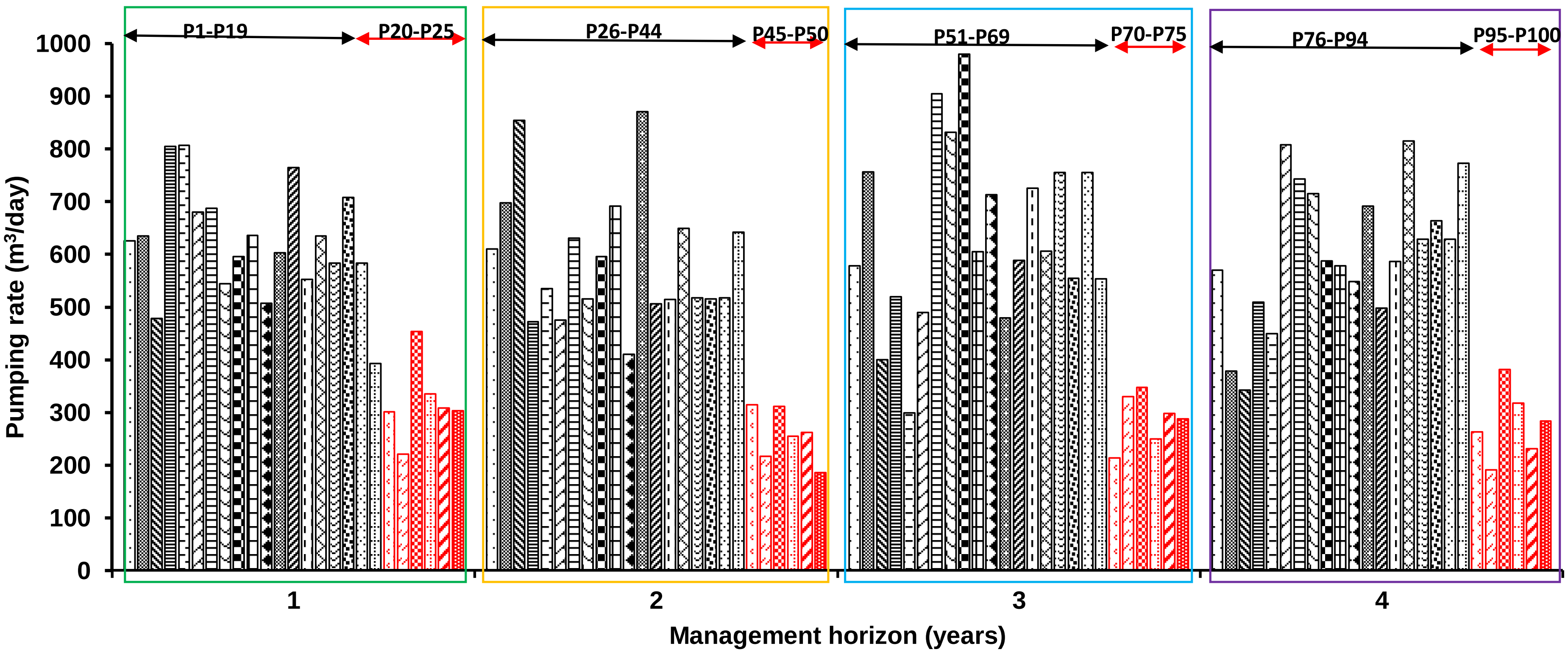


Figure 4.9: Pumping rates over the 4-year management horizon for a randomly-selected optimal solution

4.6 Discussion

This study shows that a decision-maker can be provided with a small number of representative solutions (Table 4.3). The decision-maker can then easily select a focused and informed solution based on the importance of two different objectives and the acceptable level of trade-off between them. The pruned Pareto-front containing 19 optimal solutions makes it relatively easier for decision-makers to weigh up the various trade-offs in the Pareto front. A solution from the reduced Pareto-front can be selected and implemented based on the decision-maker's preferences. The management model proposed in this study has the potential to control aquifer salinity levels and, simultaneously, meet the growing water demands of Kiribati. The Public Utility Board, which looks after the management of the Bonriki aquifer can implement the proposed management model and possibly solve the groundwater salinization issue in Kiribati. If implemented, the groundwater extracted from FPWs can be distributed to local communities and groundwater withdrawn from BWs can be discharged to the sea.

An important point of discussion is the optimal number of clusters to be used in the *k*-means clustering approach. Apart from the rule-of-thumb, other approaches such as the elbow approach, Silhouette approach and cross-validation approach could also be considered to decide on the number of clusters to be used. Also, validation of the optimal solutions obtained from the S/O model is another crucial component of the S/O approach. The optimal solutions are derived from the SVMR-assisted linked-S/O model. Hence, the validation of these solutions in terms of constraint satisfaction is critically important. For the present case study, validation was accomplished by implementing five randomly-selected optimal solutions from the reduced Pareto front into the complex numerical model. The concentrations obtained from the numerical model were compared with those predicted by SVMR. The relative differences between concentrations obtained as solutions of the optimization model and those of the corresponding numerical simulation are presented in Table 4.4. Table 4.4 indicates a very small relative difference (within 5 %) in concentration datasets obtained from the two types of model. This highlights the fact that the SVMR surrogate models accurately predict salinity concentration in response to transient pumping rate patterns. It was observed that the predicted concentrations were below the permitted levels (as per the imposed constraints), as required by the management strategy. In addition, the SVMR-predicted concentrations also converged to the upper limit of the permissible salinity level specified as a constraint. For example, at ML1, the maximum allowable salinity limit was 20,000 mg/L. All the predicted solutions at ML1 (Table 4.4) did not exceeding this limit. Also, as seen in Table 4.4, the SVMR-predicted concentrations converge to the upper limits of the constraints but do not exceed the maximum allowable limit. Similar trends were evident for ML2, ML3, ML4, ML5 and ML6. These results demonstrate the validity of the optimal groundwater extraction patterns determined for the Bonriki aquifer. Use of such optimal pumping patterns will ensure sustainable withdrawal of acceptable-quality groundwater for the South Tarawa community.

Table 4.4: Results of implementing optimal solutions in the numerical model (NM) and SVMR surrogate model

Solution	C_{max} at ML1 ≤ 20,000 mg/L			C_{max} at ML2 ≤ 20,000 mg/L			C_{max} at ML3 ≤ 5000 mg/L			C_{max} at ML4 ≤ 4000 mg/L			C_{max} at ML5 ≤ 450 mg/L			C_{max} at ML6 ≤ 450 mg/L		
	NM	SVMR	RE (%)	NM	SVMR	RE (%)	NM	SVMR	RE (%)	NM	SVMR	RE (%)	NM	SVMR	RE (%)	NM	SVMR	RE (%)
1	20021.5	19995.5	0.13	20019.9	19991.3	0.14	5002.9	4998.2	0.09	4012.3	3998.2	0.35	455.2	444.2	2.42	456.7	448.7	1.75
2	20008.4	19982.3	0.13	20014.6	19989.2	0.13	5010.2	4899.6	2.21	4029.6	3991.8	0.94	459.6	446.8	2.79	461.2	443.9	3.75
3	20011.5	19994.1	0.09	20028.2	19988.5	0.20	4999.8	4995.1	0.09	3998.2	3986.5	0.29	461.2	449.1	2.62	467.9	448.2	4.21
4	20016.1	19998.7	0.09	20016.9	19991.2	0.13	5008.2	4997.3	0.22	4011.2	3997.5	0.34	456.8	448.2	1.88	451.2	441.8	2.08
5	20008.3	19982.2	0.13	20006.7	19993.3	0.07	5011.3	4996.2	0.30	4019.7	3994.2	0.63	449.4	441.9	1.67	449.8	446.8	0.67

It is important to note that the use of BWs is only an option in the decision model. In a decision model based on the linked S/O approach, the options are to use BWs, recharge wells or no wells at all. It is not possible to predict all outcomes in a complex system with temporally- and spatially-varying pumping patterns and multiple interconnected layers. However, different management constraints can be imposed to limit (even to zero) pumping from any or all BWs. The optimal decision models can determine whether there are any better alternatives. Indeed, our solutions for different management scenarios show that the total amount of beneficial pumping will be substantially reduced if BWs are not utilized i.e. restricted to zero withdrawal. The main utility of developing such a complex management decision model is to be able to consider all such scenarios and be able to identify the best strategy for maximizing beneficial and sustainable withdrawal from the aquifer. If the simulation model is calibrated (with reasonable accuracy), all such management options are generated by the optimization algorithm and evaluated as candidate solutions. Then one can be selected that attains the best objective function levels. This is the main motivation for developing such computationally-intensive linked-S/O models. For example, it is also possible to include the option of freshwater injection wells and determine whether their distribution and usage patterns could improve the total amount of water pumped from other locations over time. It is easy to add such management alternatives and search for the best solutions in terms of planning. If the BWs are not useful, the solution will indicate it is better not to use them, as one of the objectives is to minimize the total amount of BW pumping. In the present study, the solutions of the management model show substantial benefits if BW pumping is adopted as an alternative. Beneficial pumping was found to increase from about 40,000 m^3/day to nearly 47,000 m^3/day over the time horizon of four years; i.e., an almost 20 % increase in beneficial pumping was possible using only six potential BWs. However, there are some economic and environmental issues related to the disposal of saline water pumped from BWs. These issues were not within the scope of this study.

4.7 Conclusions

This study applied a linked S/O-based methodology with a Pareto-front clustering technique to prescribe optimal groundwater pumping strategies for the Bonriki aquifer that ensure its salinity levels remain within specified limits. The management model incorporates pumping from FPWs and the option of pumping from potential BWs to maximize the supply of freshwater to the South Tarawa community. To ensure computational feasibility, the groundwater flow and transport numerical simulation model was replaced by trained and tested SVMR surrogate models. The evaluation of the surrogate models' performance indicates that they can be adequately trained to accurately approximate the density-dependent saltwater intrusion dynamics in the Bonriki aquifer in response to groundwater pumping patterns prescribed for FPWs and BWs. The SVMR surrogate models were externally linked to the MOGA optimization model. The multi-objective linked-S/O-based management model generated an optimal Pareto-front that exhibited different levels of trade-off between the total FPWs and BWs pumping solutions. Selection of a single solution from this huge optimal solution set was deemed a difficult task, especially for a group of decision-makers. Therefore, the *k*-means clustering technique was used to group solutions with similar features. A total of 19 clusters were formed, each having a different number of optimal solutions. The centroid of each cluster was used as a reference solution, and the solution in the Pareto front closest to these reference solutions formed part of the reduced Pareto front. The *k*-means clustering technique pruned the Pareto front and presented a workable reduced number of optimal solutions. This will aid decision-makers in decision-making. Overall, our limited performance evaluations show that the suggested method for solving multi-objective aquifer management problems has the potential for application to other islands facing similar saltwater intrusion problems. Hence, it can be further evaluated and utilized to develop feasible and reliable regional-scale coastal aquifer management strategies that ensure the sustainability of fragile groundwater resources. In the next chapter, a similar multi-objective coastal aquifer management model is presented that incorporates aquifer parameter uncertainty.

Chapter 5: A multi-objective groundwater management strategy incorporating aquifer parameter uncertainty: A solution for an island country in the South Pacific

The main contents of this chapter have been published and copyrighted as outlined below:

Lal, A., and Datta, B. (2019). "Multi-objective groundwater management strategy under uncertainties for sustainable control of saltwater intrusion: Solution for an island country in the South Pacific." *Journal of Environmental Management*, 234, 115-130.

5.1 Summary

To date, simulation-optimization (S/O)-based groundwater management models have delivered optimal saltwater intrusion management strategies for coastal aquifer systems. At times, however, uncertainties in numerical simulation models due to uncertainty in aquifer parameters are not incorporated into management models. The present study explicitly incorporates aquifer parameter uncertainty into a multi-objective management model to prescribe optimal strategies of groundwater pumping from the unconfined Bonriki aquifer, which is situated in a small Pacific Island country. The aim of the multi-objective management model was to maximise pumping from production wells and minimize pumping from barrier wells (hydraulic barriers) to ensure that the water quality at monitoring locations (MLs) remained within pre-specified sustainable limits. To achieve the targeted management goal, a coupled flow and transport numerical simulation model of the Bonriki aquifer was developed using the FEMWATER numerical code. This three-dimensional numerical model was calibrated and validated using limited available hydrological data. To make the management model computationally efficiency and feasible, the numerical simulation component of the S/O model was replaced with ensembles of support vector machine regression (SVMR) surrogate models. Each standalone SVMR surrogate model in each ensemble was constructed using datasets produced by numerical simulation models that used different hydraulic conductivity and porosity values as input. These ensemble SVMR models were coupled to a multi-objective genetic algorithm optimization model to solve the Bonriki aquifer management problem. The executed optimization model generated a Pareto-front with 600 non-dominated optimal trade-off pumping solutions. The reliability of the management model, which was established after validation of the optimal solution, suggests that the constraints implemented in the optimization problem were satisfied; i.e., the salinities at MLs were within the specified limits. The results of this study indicate that the developed management model has the potential to address groundwater salinity problems in small island countries.

5.2 Background

As discussed and demonstrated in Chapter 4, the coupled simulation-optimization (S/O) methodology is undoubtedly a promising approach to developing optimal coastal aquifer management and remedial strategies. However, one of the major drawbacks of coupling a complex groundwater flow and transport simulation model to an optimization model within the S/O methodology is that the resulting model may be computationally-demanding and, therefore, infeasible for large-scale study areas (Das and Datta 1999; Dhar and Datta 2009). This drawback can be resolved by substituting the complex simulation model within the S/O framework with efficient approximate groundwater simulators termed *surrogate models* (Sreekanth and Datta 2011). The effectiveness and reliability of employing a surrogate model to approximate physical processes in coastal aquifers have long been topics of concern. In this chapter, ensembles of SVMR models are used to incorporate aquifer parameter uncertainty in a multi-objective coastal aquifer management model that informs the design of an optimal aquifer management strategy.

The ensemble modelling paradigm is a new computational intelligence methodology in which diverse expert opinions are incorporated and integrated strategically to solve a problem (Rokach 2010). In the field of surrogate modelling research, the ensemble methodology combines standalone surrogate models using statistical means to build an "ensemble model", with more accurate and reliable prediction capabilities. The use of ensemble surrogate models within coupled S/O-based coastal aquifer management models is quite rare. However, more recently, efforts have been made to use ensemble surrogate models instead of standalone surrogate models in S/O frameworks to develop computationally-feasible coastal aquifer management models. For example, Sreekanth and Datta (2011) and Roy and Datta (2017) developed optimal coastal aquifer management strategies using ensemble surrogate models to address predictive uncertainties in the surrogate models. However, uncertainties involved in saltwater intrusion numerical simulation models are yet to be investigated and incorporated into a coastal aquifer management model. Uncertainties in saltwater intrusion numerical simulation models due to uncertainties in aquifer parameters present an important research gap that needs to be explored. Therefore, this study aims to incorporate aquifer parameter uncertainty in the development and evaluation of an ensemble surrogate model-based multi-objective coastal aquifer management model. This is a logical extension of previously-developed coastal aquifer management models based on the ensemble surrogate modelling paradigm.

In addition to combining individual models to improve their predictive capabilities, the ensemble modelling paradigm can also be used to integrate divergent views of data, different random selections of data and different diverse decisions (Anifowose et al. 2017). These properties of the ensemble modelling technique were utilized in this study to combine the responses of several numerical simulation models constructed using different values of uncertain coastal aquifer parameters. Liu et al. (2017) stated that structural uncertainties in a hydrological model can be handled using a multi-hydrological model ensemble. This is because a single hydrological model does not usually represent a system's behaviour adequately. Various hydrological studies have addressed structural uncertainty in simulation models by integrating several hydrological models into a single composite model, an approach which can increase accuracy and decrease uncertainty (Corzo et al. 2009; Seiller et al. 2012; Toth 2009). Therefore, combining several coastal aquifer numerical simulation models with the help of ensemble surrogate models as approximate simulators can address the issue of uncertainty in aquifer parameters, thereby ensuring the robustness of optimal solutions obtained from S/O management models.

In S/O models, numerical models are generally used to simulate groundwater flow and transport processes. However, the contaminant fate in the subsurface is considerably affected by uncertainties in natural porous media (Gelhar 1993). Owing to this uncertainty, numerical models are unlikely to precisely represent the real physical processes of hydrological systems. The issue of coastal aquifer parameter uncertainty usually arises due to the inaccurate estimation of model parameters (errors in field observations) due to heterogeneity in the hydrological environment and a scarcity of related data. The present study considers hydraulic conductivity and porosity as two important parameters that cause uncertainty in numerical simulation models. The random distribution of hydraulic conductivity is the most important parameter contributing to uncertainty in simulation models (Luo et al. 2014; Ranjithan et al. 1993). Hydraulic conductivity regulates the transport mechanism in porous media through its effect on velocity. Konikow (2011) argued that the accuracy of a simulation model is dependent on how accurately it can represent the actual distribution of hydraulic conductivity. Porosity is another crucial uncertain parameter of aquifers that affects the seepage velocity, consequently affecting both mechanical dispersion and advection transport (Konikow 2011). In S/O models, disregarding the inherent uncertainty in hydrological model parameters can lead to significant errors in simulation results, resulting in ambiguous, inappropriate and/or infeasible optimal solutions (Chan Hilton and Culver 2005). Hence, aquifer parameter uncertainty cannot be discounted and should be considered a key component in any coastal aquifer management model.

To the best of the authors' knowledge, very few studies in the hydrologic literature have used ensemble surrogate modelling approach (combined several numerical simulation models) to increase the robustness of the simulation models in coupled S/O model-based coastal aquifer management models. The design of a multi-objective coastal aquifer management model and its application to an island setting is a new contribution to the field of water resources management research. Details of the groundwater numerical simulation model, surrogate model, uncertain aquifer parameters, ensemble modelling paradigm, multi-objective management model design and the study area are provided in the Methods section. A comprehensive analysis of the results with key discussions is presented in the Results and Discussion section.

5.3 Methods

5.3.1 Characterisation of aquifer parameter uncertainty

Support vector machine regression (SVMR) models were used as surrogate models to approximate saltwater intrusion processes in the Bonriki aquifer and to accurately predict salinity concentrations at monitoring stations. Details of the SVMR modelling paradigm are presented in Chapter 3. Hydraulic conductivity and porosity are two uncertain aquifer parameters that were considered in the development of the multi-objective coastal aquifer management model. As established by Binley et al. (1997), Chen et al. (2003) and Zhao et al. (2013), hydraulic conductivity and porosity have log-normal and normal distributions, respectively. Different combinations of these two uncertain parameters in different aquifer layers were implemented in the numerical model while keeping the other parameters constant during the simulation period. *N* numerical models were developed using *u* sets of hydraulic conductivity and porosity values. Uniformly-distributed, random, transient, groundwater pumping rate patterns were generated from the decision space using LHS to train and test the SVMR-based surrogate models. The implementation of each set of pumping rate patterns into the different numerical models generated different output salinities at each monitoring location *m*. The input-output (pumping rate-salinity) datasets from each of the numerical models were used to construct different surrogate models capable of approximating a particular numerical model. From each of the input-output datasets, 80% of the data was used to train each surrogate model, while 20% was used to test its performance. Utilizing the training and testing data from *N* numerical models led to the development of *N* surrogate models. An outline for constructing an ensemble surrogate model for each monitoring location is presented in Fig. 5.1. Specifically, the ensemble model responsible for predicting salinity concentration at each *m* had *N* standalone models. Each of these standalone surrogate models was built using the datasets generated from numerical models based on different hydraulic conductivity and porosity values. Ensembles of these *N* surrogate models were developed by combining the predicted output of each surrogate model in the ensemble using the simple averaging method (Shu and Burn 2004). This method is common and has produced excellent results in a range of applications (Bishop 1995; Perrone and Cooper 1995). The formula for constructing an ensemble model (E_n) for predicting salinity concentration at each monitoring location m by combining the predicted output (P) of a standalone surrogate models (n) using the simple averaging method is given by:

$$E_{n,m} = \frac{1}{N}\sum_{n=1}^{N} P^n \qquad n = 1,2,\ldots,N \qquad (5.1)$$

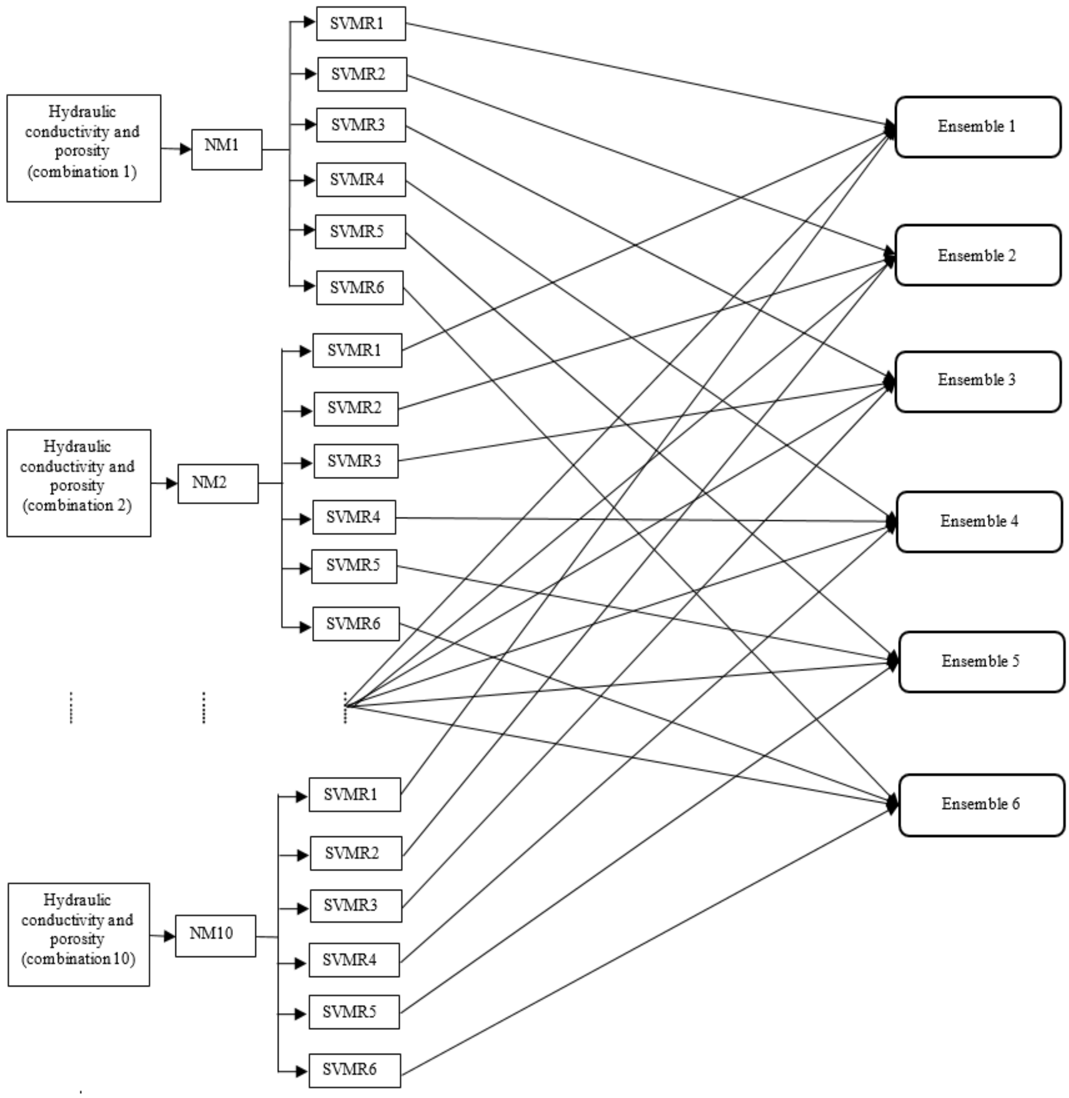


Figure 5.1: Procedure for the development of ensemble surrogate models

5.3.2 Surrogate model performance evaluation criteria

The performance of the standalone SVMR and ensemble models was evaluated using four statistical parameters: root mean square error (RMSE), correlation coefficient (r), Nash-

Schliffe efficiency (NSE) and Willmott's index of agreement (WI). The mathematical expressions of these statistical parameters are given below:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(c_T - c_P)^2} \qquad (5.2)$$

$$r = \frac{\sum_{i=1}^{n}(c_T-\overline{c_T})((c_P-\overline{c_P})}{\sqrt{\sum_{i=1}^{n}(c_T-c_T)^2}\sqrt{\sum_{i=1}^{n}(c_P-c_P)^2}} \qquad (5.3)$$

$$NSE = 1 - \frac{\sum_{i=1}^{n}(c_T-c_P)^2}{\sum_{i=1}^{n}(c_T-\overline{c_P})^2} \qquad (5.4)$$

$$WI = \left|1 - \left(\frac{\sum_{i=1}^{n}(c_T-c_P)^2}{\sum_{i=1}^{n}|c_P-\overline{c_P}|+|c_T-\overline{c_T}|}\right)\right| \qquad (5.5)$$

Where n represents the total number of datasets, c_T is the true salinity from the numerical model, c_P represents the salinity predicted by the surrogate model, $\overline{c_T}$ is the mean true salinity predicted by the numerical model and $\overline{c_P}$ denotes the mean salinity predicted by the surrogate model.

5.3.3 Formulation of the multi-objective management model

The main purpose of this study was to develop a coastal aquifer management model incorporating parameter uncertainty by utilizing an ensemble surrogate model-based simulation-optimization methodology. The main objectives of the multi-objective management model were to prescribe optimal groundwater pumping rates at freshwater pumping wells (FPWs) and barrier wells (BWs) while ensuring that salinities at monitoring locations (MLs) remained within specified limits. The pumping rates at the FPWs and BWs are the decision variables used in the management model. Instead of using a complex numerical model, the SVMR ensemble models constructed for predicting salinity concentration in the aquifer were separately coupled to a multi-objective genetic algorithm (MOGA) optimisation model. The mathematical expressions of the two objective functions and the constraints and bounds of the management model are given in Section 4.4.9 (Chapter 4).

5.3.4 Validation of the optimal solutions

Execution of the multi-objective ensemble surrogate-based S/O model generates optimal solutions in the form of a non-dominated Pareto front. The Pareto front consists of a set of non-dominated optimal solutions dependent on the trade-offs between two conflicting objectives. Validating these optimal solutions by implementing them in the complex numerical model is a crucial step. For this purpose, five optimal solutions on the Pareto-front were chosen randomly and implemented into each numerical model. The salinities obtained from each numerical model were compared with those of their corresponding standalone

SVMR surrogate model. Finally, the ensemble model's predicted salinities were compared with the mean salinity predicted by the numerical simulation. This comparison was performed to validate the role of the ensemble surrogate model in the coupled simulation-optimization-based management model. Figure 5.2 outlines the complete stepwise procedure of the multi-objective coastal aquifer management model incorporating parameter uncertainty.

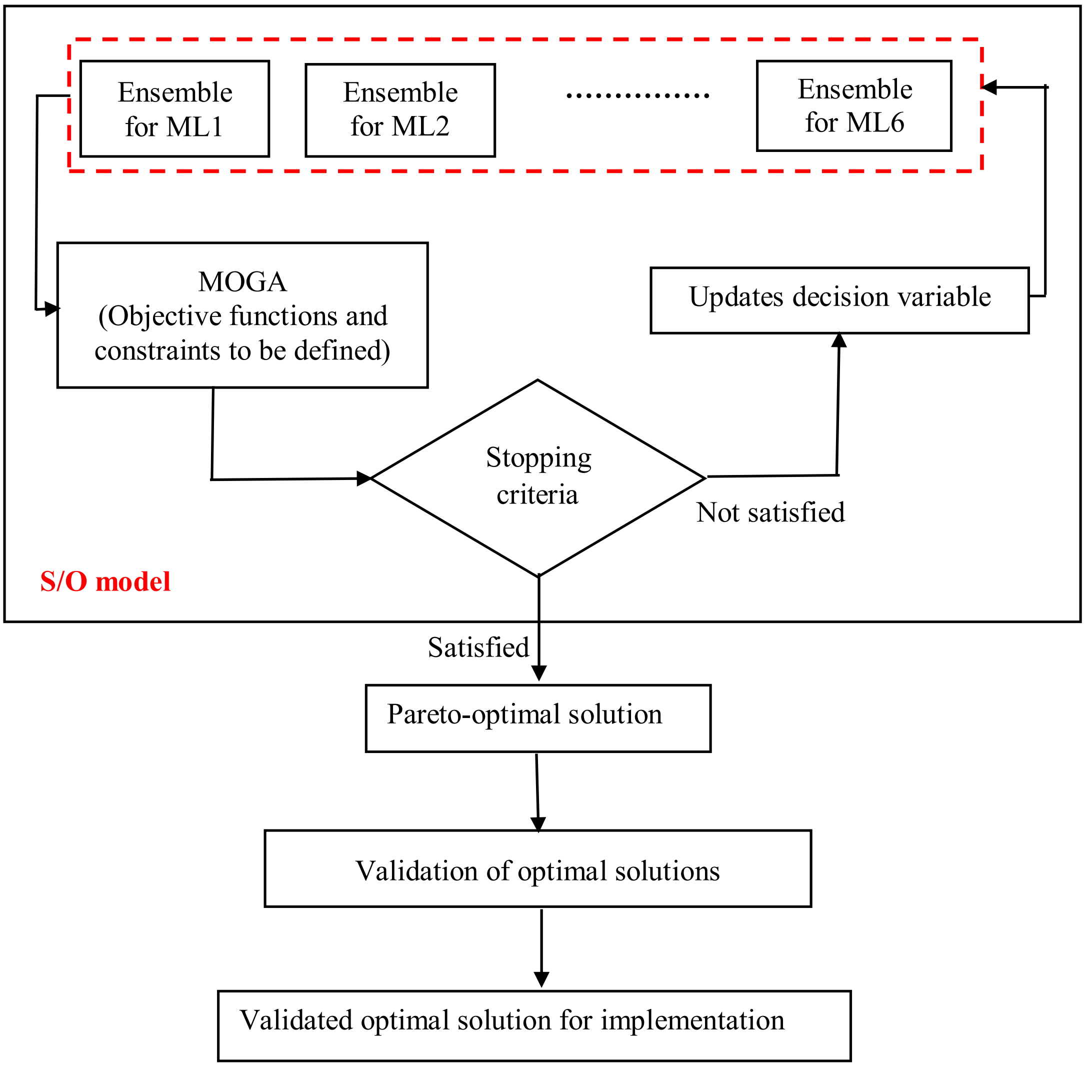


Figure 5.2: Step-wise procedure of the developed S/O-based management framework using ensemble surrogate models

5.3.5 Application of the developed method

The developed method was applied to the Bonriki coastal aquifer, which is situated in the small Pacific Island nation of Kiribati. The aim was to derive an optimal, long-term, sustainable, groundwater management strategy. The study area's location, hydrogeology and field data (groundwater level and concentration) are extensively described in Chapter 4. A 3D model of the Bonriki aquifer was developed using the FEMWATER numerical computer code, a detailed description of which is presented in Chapter 3. The modelling details are also described in Chapter 4. The developed model was calibrated and validated using available field data, as described in Chapter 4. The calibration and validation results are also presented in Chapter 4.

Inclusion of aquifer parameter uncertainty in the management model was made possible with the help of ensemble surrogate models. Ten numerical models were developed using various values of hydraulic conductivity and porosity (Table 5.1). The values of hydraulic conductivity were derived from a lognormal distribution using the calibrated value of hydraulic conductivity as the mean and a variance of 0.4. Similarly, the values of porosity were derived using a normal distribution with a calibrated value of porosity as the mean and a variance of 0.1. Data from these ten numerical models was used to construct ten surrogate models for each monitoring location. Hence, the ensemble model used to predict the salinity at each monitoring location consisted of ten standalone surrogate models constructed using different numerical model solutions. The ten combinations of hydraulic conductivity and porosity values were used to evaluate the proposed methodology only. More combinations of these parameters could be used depending on the availability of high-performance computers. A total of 700 input-output datasets were used to construct each surrogate model in the ensemble (460 for training and 240 for testing). Each standalone SVMR model was constructed offline using MATLAB R2016a software. A Gaussian kernel was used with parameters ε, C and γ having values of 0.60, 10 and 0.001, respectively.

Table 5.1: Values of hydraulic conductivity and porosity used in the numerical models (NM)

Model	Holocene sediment				Pleistocene sediment			
	Hydraulic conductivity			Porosity	Hydraulic conductivity			Porosity
	K_x	K_y	K_z		K_x	K_y	K_z	
NM1	15.62	7.81	1.56	0.171	626.12	313.06	62.61	0.340
NM2	20.87	10.44	2.09	0.213	299.85	149.93	29.99	0.267
NM3	18.59	9.30	1.86	0.216	372.72	186.36	37.27	0.355
NM4	21.48	10.74	2.15	0.171	475.35	237.68	47.54	0.340
NM5	14.23	7.12	1.42	0.169	400.42	200.21	40.04	0.297
NM6	14.14	7.07	1.41	0.182	507.72	253.86	50.77	0.257
NM7	22.45	11.23	2.25	0.206	528.07	264.04	52.81	0.294
NM8	12.26	6.13	1.23	0.194	310.21	155.11	31.02	0.295
NM9	9.02	4.51	0.90	0.211	419.27	209.64	41.93	0.325
NM10	12.87	6.44	1.29	0.231	349.89	174.95	34.99	0.320

The developed ensembles of surrogates were linked externally to the MOGA optimisation model using the MATLAB 2017a platform. The model parameters used were: population size = 2000, function tolerance = 1×10^{-4}, constraint tolerance = 1×10^{-3}, Pareto front population fraction = 0.3 and crossover fraction = 0.8. The permissible salinity levels in the aquifer were imposed as another set of management constraints. These constraints were assigned to the management model to ensure that the salinities at the MLs were restricted to the specified limits. The maximum acceptable salinities at ML1 and ML2 were set to 20,000 mg/L. ML1 and ML2 were closer to the shoreline and confining their salinity levels to an extremely lower level was impractical. The maximum salinities at ML3 and ML4 were set to 5000 mg/L and 4000 mg/L, respectively. Finally, the maximum salinities at ML4 and ML5 were both set to 450 mg/L. ML5 and ML6 were located in an area with many pumping wells and it was anticipated that water extracted from these locations would be suitable for a range of activities (e.g., household use and irrigation) by the local community.

5.4 Results and discussion

5.4.1 Performance evaluation of the SVMR models

Table 5.2 lists the salinity prediction capabilities of the SVMR models during the testing stage in terms of four performance evaluation criteria (RMSE, r, NSE and WI). RMSE is a quadratic scoring criterion that measures the average difference between predicted and actual values (Mehr and Kahya 2017). Hence, RMSE values were used to assess the difference between actual and SVMR-predicted salinities. The RMSE values for all 60 SVMR models were within a reasonable range of 1.33–8.43. These RMSE values are sufficient to classify the developed SVMR models as accurate. Similarly, r is another performance evaluation criterion used to assess the predictive capabilities of the developed SVMR models. The r value describes the degree of collinearity between actual and predicted values (Moriasi et al. 2007). The r values can range from −1 to 1, where 1 indicates a perfect positive relationship and −1 indicates a perfect negative one. No linear relationship exists when $r = 0$. In the present case, the computed values of r were in the range of 0.96–0.99. These results establish that the degrees of collinearity between the actual and predicted salinities were positive and almost perfect.

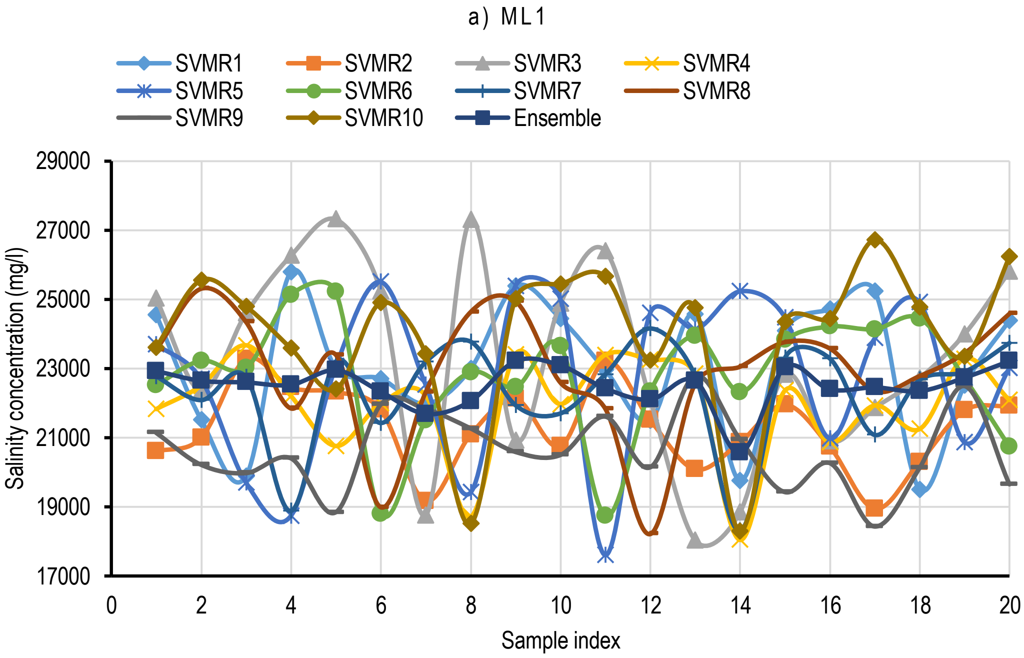
Table 5.2: Performance of the surrogate models (SVMR) relative to the numerical models (NM) in the testing phase

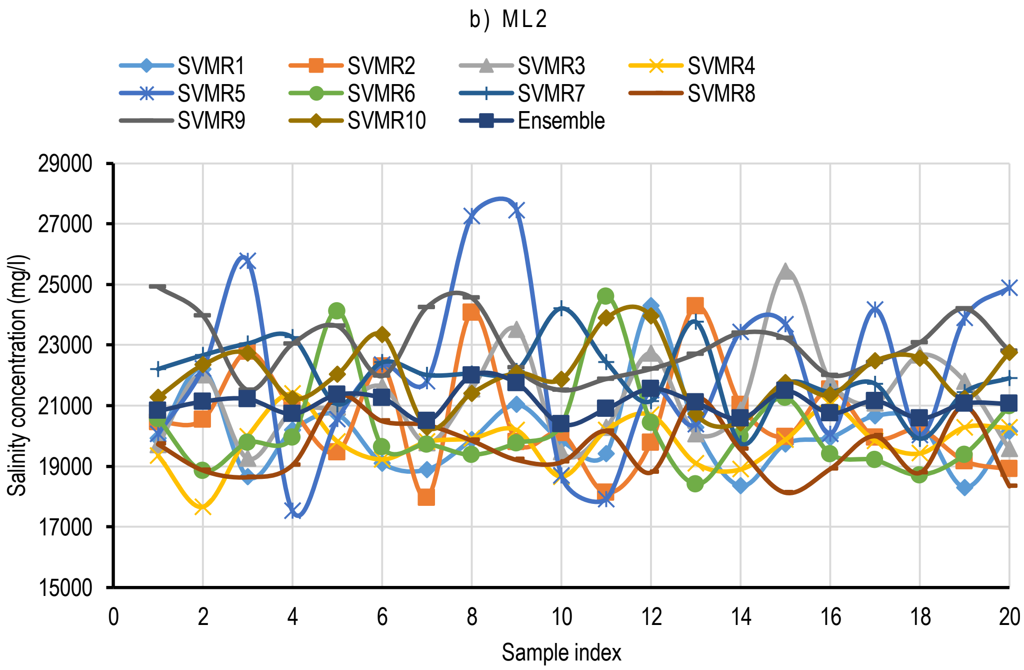
NM	Performance measure	SVMR1	SVMR2	SVMR3	SVMR4	SVMR5	SVMR6
1	RMSE	5.10	6.17	3.74	2.95	2.02	1.89
	r	0.96	0.97	0.97	0.97	0.98	0.98
	NSE	0.97	0.96	0.98	0.97	0.98	0.98
	WI	0.94	0.95	0.95	0.96	0.96	0.96
2	RMSE	5.98	5.62	2.82	2.04	1.59	1.33
	r	0.97	0.97	0.98	0.97	0.98	0.98
	NSE	0.96	0.96	0.96	0.97	0.97	0.97
	WI	0.95	0.94	0.95	0.96	0.96	0.96
3	RMSE	4.16	5.22	3.51	4.86	3.02	2.14
	r	0.97	0.96	0.97	0.96	0.98	0.98
	NSE	0.97	0.97	0.98	0.97	0.98	0.99
	WI	0.94	0.95	0.95	0.94	0.95	0.96
4	RMSE	6.60	5.33	5.27	4.65	3.53	3.05
	r	0.97	0.98	0.96	0.97	0.97	0.97
	NSE	0.97	0.96	0.96	0.97	0.97	0.98
	WI	0.95	0.94	0.93	0.95	0.95	0.96
5	RMSE	6.96	7.13	5.12	5.68	4.25	4.56
	r	0.97	0.97	0.97	0.97	0.98	0.97
	NSE	0.97	0.98	0.98	0.97	0.98	0.97
	WI	0.95	0.94	0.95	0.94	0.96	0.95
6	RMSE	7.63	5.32	5.24	5.69	4.25	3.57
	r	0.97	0.98	0.98	0.97	0.99	0.99
	NSE	0.98	0.98	0.98	0.98	0.98	0.99
	WI	0.96	0.97	0.97	0.97	0.98	0.98
7	RMSE	7.26	6.75	6.03	5.87	5.66	5.12
	r	0.97	0.98	0.98	0.98	0.98	0.99
	NSE	0.97	0.98	0.98	0.98	0.98	0.99
	WI	0.96	0.97	0.97	0.97	0.98	0.98
8	RMSE	6.35	7.16	5.57	5.31	5.26	5.19
	r	0.98	0.97	0.98	0.98	0.98	0.98
	NSE	0.97	0.97	0.97	0.97	0.97	0.97
	WI	0.96	0.95	0.96	0.96	0.96	0.96
9	RMSE	7.37	6.89	8.43	6.22	5.32	4.41
	r	0.98	0.98	0.96	0.97	0.97	0.98
	NSE	0.97	0.97	0.96	0.97	0.97	0.98
	WI	0.95	0.96	0.95	0.96	0.97	0.97
10	RMSE	7.14	6.59	6.91	5.88	4.71	4.28
	r	0.98	0.98	0.98	0.98	0.98	0.99
	NSE	0.98	0.98	0.98	0.98	0.98	0.99
	WI	0.96	0.97	0.97	0.97	0.97	0.98

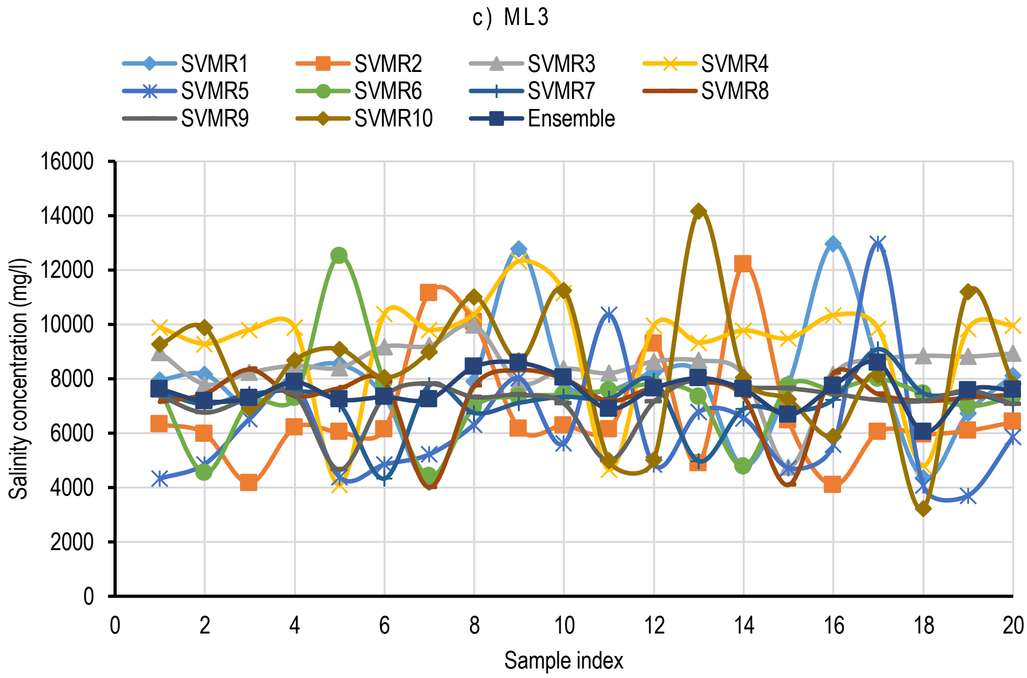
In addition, NSE values were calculated to evaluate the performance of the SVMR models. NSE is a normalised statistic that determines the relative magnitude of the residual variance compared to the variance in the measured data (Nash and Sutcliffe 1970). NSE values range from 0 to 1, with a value of 1 corresponding to a best-fit predictive model. A predictive model can be considered accurate if the NSE value is close to 1 (Pulukuri et al. 2018). The calculated NSE values were found to be within the range 0.96–0.99; hence, all were very close to 1, demonstrating the accuracy and reliability of the developed SVMR predictive models. Lastly, the WI criterion was used to evaluate the agreement between the actual and predicted salinities. The WI measures the closeness of two datasets to a 1:1 line and normally ranges from 0 (complete disagreement) to 1 (perfect agreement) (Willmott 1981). All the computed WI values were close to 1 (range of = 0.93–0.98), which demonstrates the closeness between the actual and predicted salinity datasets. Overall, the performance evaluation results confirm that the developed SVMR models are sufficiently accurate, efficient and reliable. This suggests that they can emulate complex numerical model solutions with reasonable accuracy. The high accuracy of the developed SVMR models is evident in the high correlations between the input and output datasets presented to the models during the training stage. In addition, the accuracy of the SVMR models is dependent on the input data structure (especially during testing), the training/testing method used and the SVMR algorithm parameters. Hence, these features need to be selected carefully.

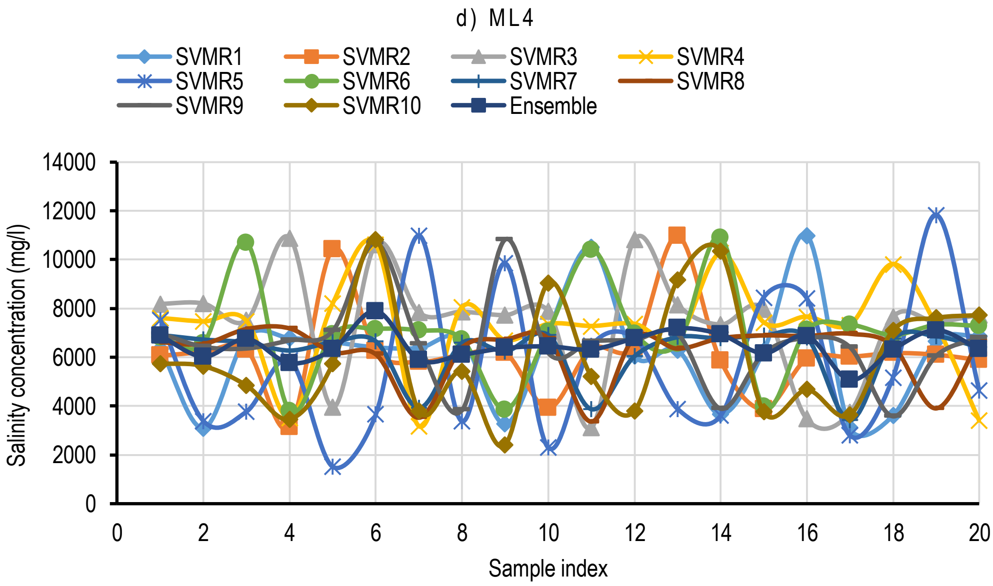
5.4.2 Utilizing ensemble models to incorporate parameter uncertainty

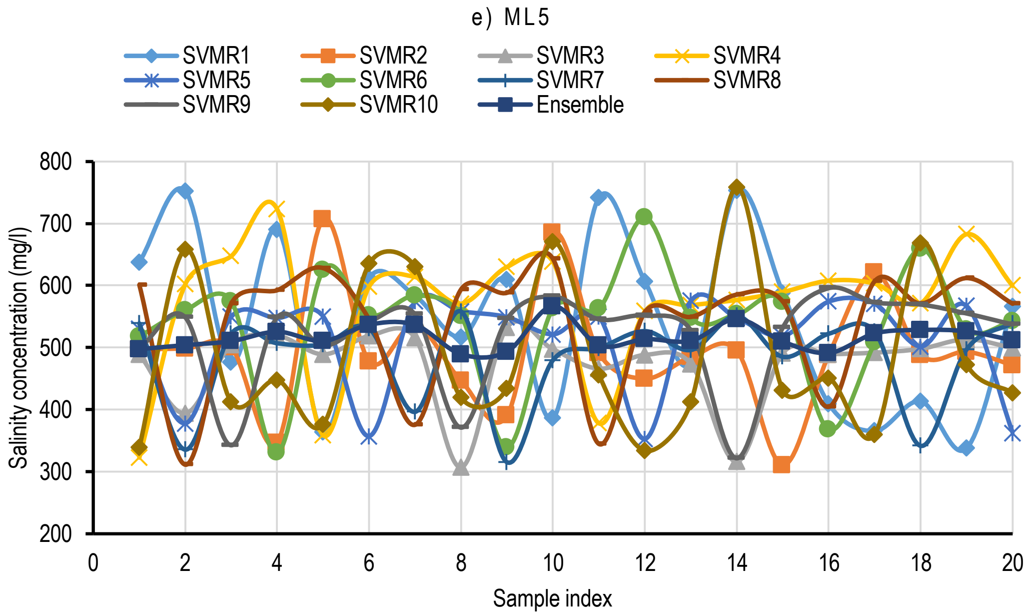
The ensemble of trained and tested surrogate models was used to address the uncertainties associated with the prediction of the saltwater intrusion process. The tested SVMR models were used as an approximation of the saltwater intrusion numerical simulation model in the coupled S/O management model to evaluate and provide optimal solutions based on the objective functions and set constraints. To ensure the robustness of the simulation model in the coupled S/O model, parameter uncertainty in the developed simulation model was incorporated by training and testing each SVMR surrogate model in the ensemble using different sets of hydraulic conductivity and porosity data. Ten sets of uncertain hydraulic conductivity and porosity values were used to train and test ten surrogate models. The ten SVMR surrogate models developed for predicting salinity at each monitoring location were combined using the ensemble modelling paradigm. Figure 5.3 presents the results of the ensemble model prediction using the ten standalone surrogate models. The ensemble SVMR surrogate models coupled in the S/O-based coastal aquifer management model provide a robust and versatile predictive system for the evaluation and selection of optimal groundwater pumping strategies from FPWs and BWs. It was observed that the salinity datasets obtained from the ten numerical models differed. Similarly, as shown in Fig. 5.3, the predictions of each standalone SVMR model also differed. The accuracy of the standalone SVMR models (presented in Table 5.2) and the robustness of the ensemble SVMR models ensures the computational efficiency and reliability of the optimal pumping strategies obtained from the executed coupled S/O-based coastal aquifer management model.


a) ML1
SVMR1
SVMR2
SVMR3
SVMR4
SVMR5
SVMR6
SVMR7
SVMR8
SVMR9
SVMR10
Ensemble
Salinity concentration (mg/l)
Sample index
29000
27000
25000
23000
21000
19000
17000
0
2
4
6
8
10
12
14
16
18
20


b) ML2
SVMR1
SVMR2
SVMR3
SVMR4
SVMR5
SVMR6
SVMR7
SVMR8
SVMR9
SVMR10
Ensemble
Salinity concentration (mg/l)
Sample index
29000
27000
25000
23000
21000
19000
17000
15000
0
2
4
6
8
10
12
14
16
18
20


c) ML3
SVMR1
SVMR2
SVMR3
SVMR4
SVMR5
SVMR6
SVMR7
SVMR8
SVMR9
SVMR10
Ensemble
Salinity concentration (mg/l)
16000
14000
12000
10000
8000
6000
4000
2000
0
0
2
4
6
8
10
12
14
16
18
20
Sample index


d) ML4
SVMR1
SVMR2
SVMR3
SVMR4
SVMR5
SVMR6
SVMR7
SVMR8
SVMR9
SVMR10
Ensemble
Salinity concentration (mg/l)
14000
12000
10000
8000
6000
4000
2000
0
0
2
4
6
8
10
12
14
16
18
20
Sample index



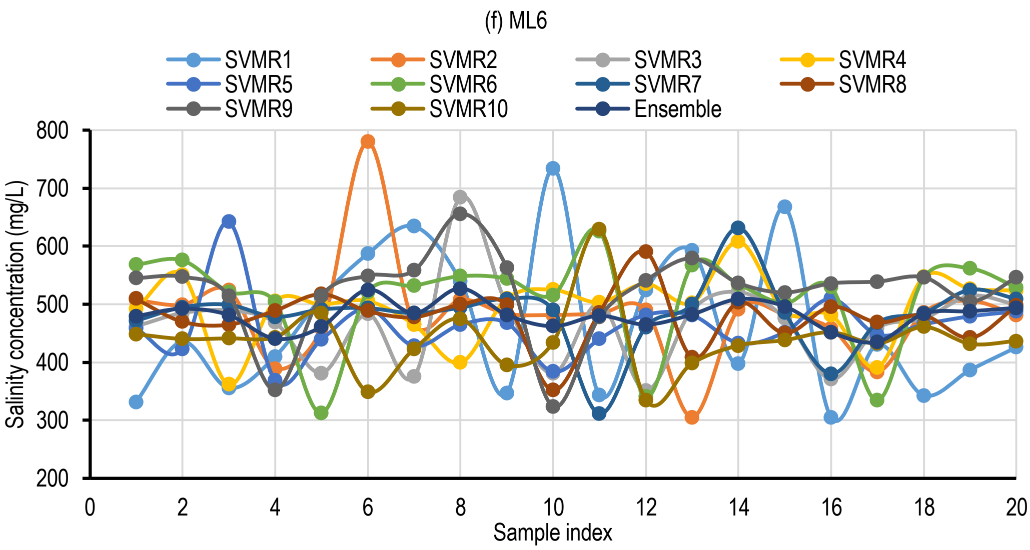


Figure 5.3: Salinities predicted by the standalone and ensemble SVMR models at six monitoring locations (ML1–ML6, shown in figs. (a) –(f), respectively).

5.4.3 Pareto-optimal solutions and trade-offs

The optimal or non-dominated set of solutions from the developed multi-objective coastal aquifer management model are presented in the form of a Pareto front (Fig. 5.4). The Pareto front consists of 600 (population size of 2000 × Pareto-front population fraction of 0.3) different solutions defining the trade-offs between the total FPW and BW pumping rates. When solving a multi-objective optimization problem, the decision-maker is responsible for determining the most preferred optimal solution on the Pareto front based on their management goals and preferences. This process becomes a bit more complicated if multiple decision-makers with conflicting utility functions are involved (Datta and Peralta 1986). The decision-maker needs to understand the various trade-offs existing in the different Pareto-optimal solutions.

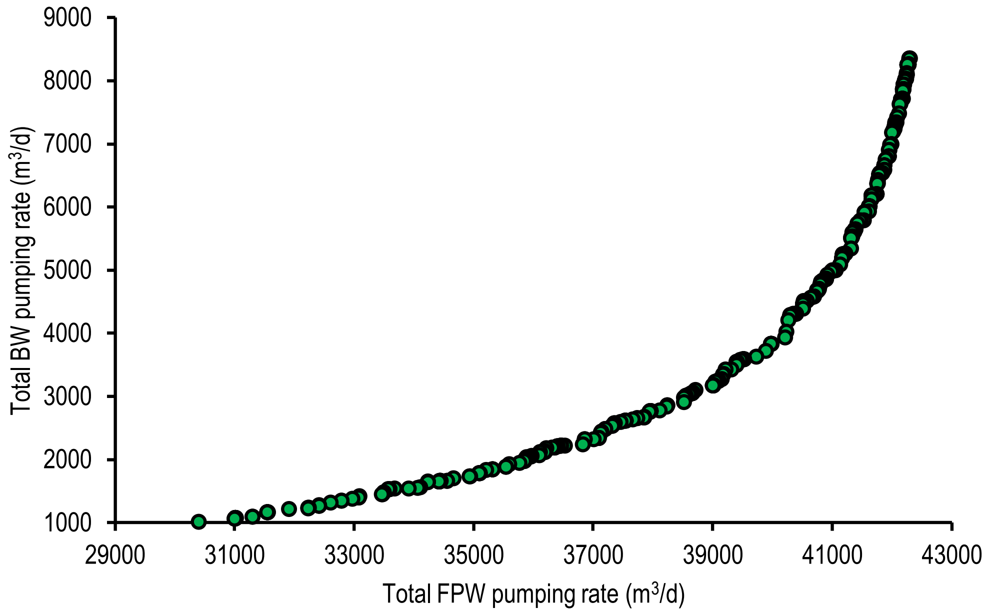


Figure 5.4: Pareto-front defining the trade-off between total FPW and BW pumping rates for the entire four-year management period

In the present case study, 600 optimal solutions provided 600 different FPW and BW pumping strategies to be evaluated by the decision-maker. From the information presented in Fig. 5.4, the maximum total FPW pumping rate for the four-year management period is 42,297.35 m^3/d with a corresponding total BW pumping rate of 8354.58 m^3/d. The Pareto front only shows the total PW and BW pumping rates for the entire four-year management horizon. Therefore, for demonstration and evaluation purposes, four random solutions from different regions of the Pareto front were chosen and the corresponding yearly pumping amounts from the FPWs and BWs over the entire four-year management period (from years 1 to 4, Y1–Y4) are presented in Fig. 5.5 (a). The FPW and BW pumping rates presented in Fig. 5.5 (a) are the total rates at the 19 FPWs and 6 BWs considered in the present study. A further breakdown of solution 1 is presented in Fig. 5.5 (b), where the specific optimal pumping solutions from each of the PWs and BWs are presented.

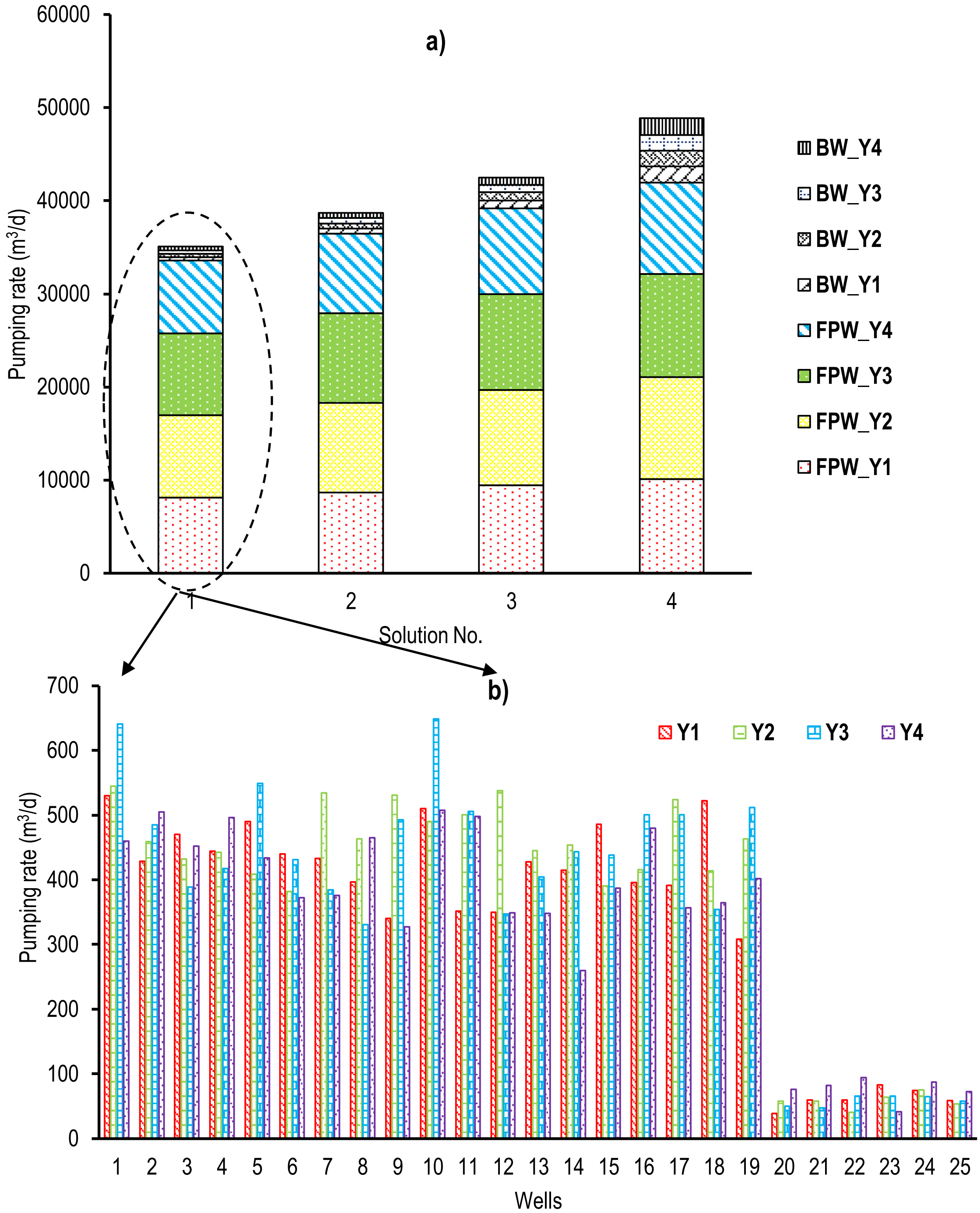


Figure 5.5: a) Annual FPW and BW pumping rates for the four randomly-selected optimal solutions and b) specific pumping rates from each well (wells 1-19 are FPWs and wells 20-25 are BWs)

Choosing the preferred solution from the large solution set comprised of 600 optimal solutions is difficult for a decision-maker. In many cases, solution selection is dependent on the priorities or utility of the decision-maker. For example, if maximising the extraction of freshwater from FPWs is considered important, then the solution prescribing a high total production well pumping rate (solution on the upper-right portion of the Pareto-front, Fig. 5.4) can be selected for implementation. This implementation will also result in a large volume of saline water being pumped from BWs. Disposal of this saline water may be uneconomical for island communities and can lead to other environmental issues. On the other hand, if priority is placed on minimising pumping from BWs, a solution from the lower portion of the Pareto-front can be selected for implementation.

Validation of the optimal solutions is a key step in any S/O-based coastal aquifer management methodology. The optimal pumping solutions on the Pareto-front are the responses of the trained and tested ensemble surrogate models, and are not obtained using the complex numerical simulation model. Hence, validating the optimal solutions by implementing them in the numerical simulation model is necessary. Table 5.3 lists the salinity at ML1 obtained after implementing five randomly-selected optimal solutions in each of the individual numerical models and standalone SVMR surrogate models. The results presented in Table 5.3 show that the standalone SVMR surrogate models in the ensemble models were accurate in approximating the corresponding numerical simulation model's responses. Similar trends were observed for all other MLs. The validation results in Table 5.3 also show that the implemented constraints were satisfied; i.e., the salinity at ML1 was within the pre-specified limit. For example, the salinities at ML1 converged to the upper limit of the maximum allowable concentration of 20,000 mg/L. Similar trends were observed for ML2, ML3, ML4, ML5 and ML6. In addition, a comparison of the ensemble SVMR model's predicted salinities with the average salinities predicted by the numerical model (NM_Av) for each monitoring location is presented in Fig. 5.6. The comparison establishes that the ensemble SVMR surrogate models approximate the numerical simulation model's results with reasonable accuracy. In addition, for each of the optimal solutions, the corresponding salinities obtained from the NM_Av and ensemble surrogate models were not only below the pre-specified limit but also converged to the upper limit of the set constraint.

Table 5.3: Comparison of salinities predicted by the numerical models (NM) and corresponding standalone SVMR surrogate models

	ML1 [$C_{max} \leq$ 20,000 mg/L]																			
	NM1	SVMR1	NM2	SVMR2	NM3	SVMR3	NM4	SVMR4	NM5	SVMR5	NM6	SVMR6	NM7	SVMR7	NM8	SVMR8	NM9	SVMR9	NM10	SVMR10
1	20079.3	19770.7	20036.5	19292.3	19434.8	19575.2	19185.3	19158.3	20446.7	19810.4	19788.5	19075.8	20178.8	19652.4	20032.6	19728.2	19630.7	19573.2	19891.1	19133.4
2	19713.8	19162.2	19953.8	19661.0	19974.4	19901.9	19952.2	19287.0	19127.9	19583.6	19929.7	19285.3	19712.2	19587.3	19893.4	19842.9	19043.9	19993.8	19786.9	19906.6
3	20266.8	19991.5	20290.1	19900.2	19801.2	19929.4	20226.2	19871.9	19984.4	20378.9	19820.8	19589.6	19878.9	19827.9	19720.3	19497.3	19872.9	19012.2	19930.5	19470.4
4	19082.6	19473.0	19503.0	19092.8	19831.5	19025.2	19968.1	19516.6	19308.2	19771.6	20279.4	19612.8	19443.6	19283.1	20344.8	19764.5	19990.0	19739.7	20263.0	19395.0
5	19337.4	19343.6	19654.1	19511.0	20314.8	19745.9	19833.0	19758.8	19877.3	19166.7	19033.7	19854.3	20154.9	19003.1	19018.3	19320.3	20199.4	19340.9	19850.9	19631.6

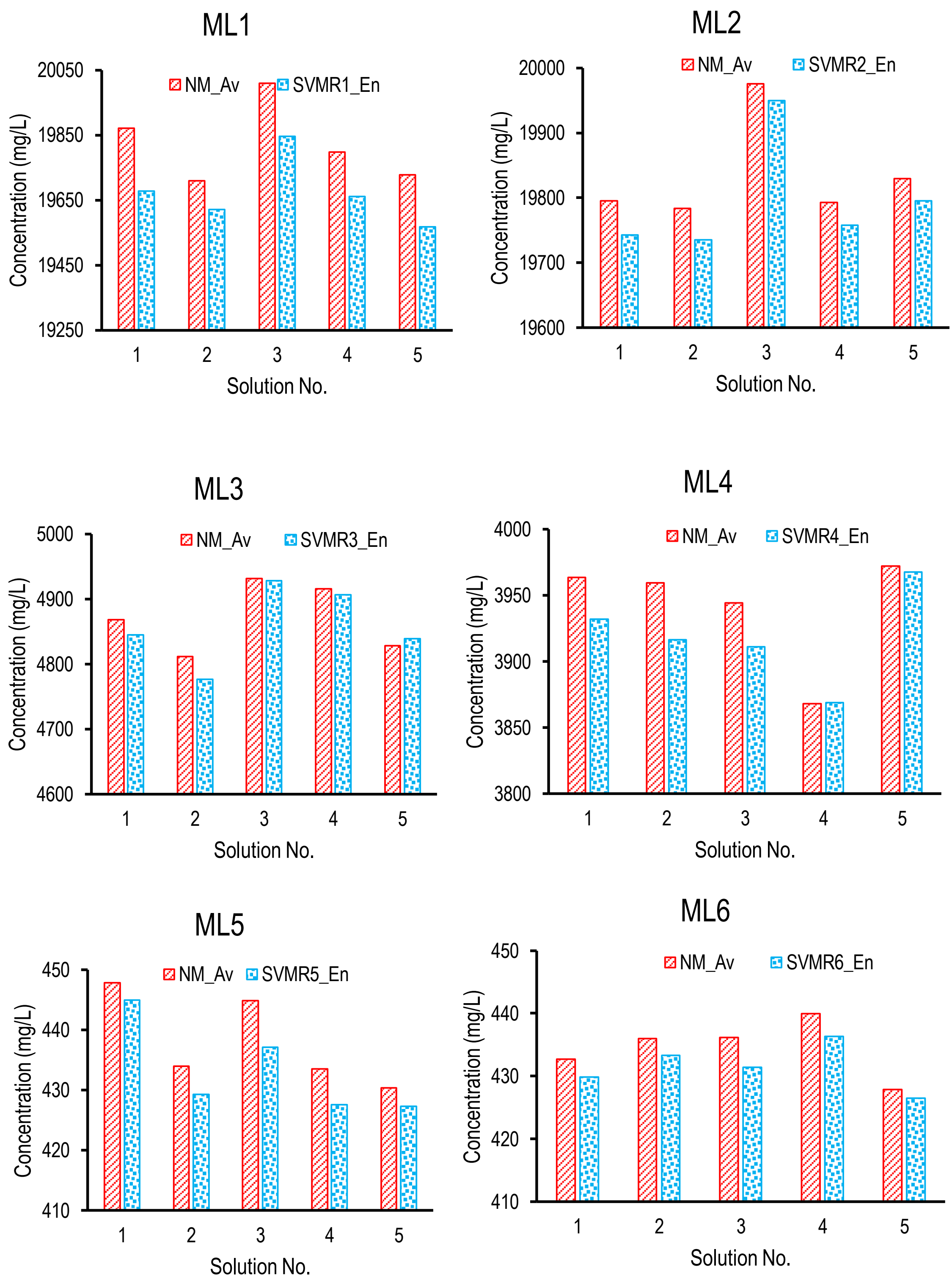


Figure 5.6: Comparison of the concentration results from numerical models (Average) and ensemble SVMR surrogate models

5.5 Conclusions

This study demonstrates the potential feasibility of incorporating uncertainty in aquifer parameters (hydraulic conductivity and porosity) in the development of computationally-feasible, regional-scale, sustainable management strategies for coastal aquifers. This was accomplished utilizing an ensemble surrogate model-based coupled multi-objective S/O model to develop reliable, optimal pumping strategies for the unconfined Bonriki aquifer system. The constraining issue of computational burden encountered during execution of the developed S/O model was resolved using SVMR-based surrogate models, which provided reliable approximations of the complex numerical saltwater intrusion simulation. The SVMR models were trained and tested using the input (pumping rate) and output (salinity) datasets obtained by solving the calibrated 3D saltwater intrusion numerical simulation model. To ensure the robustness of the developed S/O model, uncertainties in the numerical saltwater intrusion simulation model were characterised by developing several standalone SVMR surrogate models based on different combinations of uncertain hydraulic conductivity and porosity values. Therefore, instead of linking a standalone SVMR surrogate model, ensembles of ten standalone SVMR models for each monitoring location were coupled to a MOGA optimisation model.

The performance evaluation criteria show that the standalone SVMR models provided accurate, efficient and reliable approximations of the saltwater intrusion numerical simulation model. The accuracy and reliability of the standalone SVMR models ensured that the ensemble models also generated reliable estimates. Execution of the ensemble surrogate-based multi-objective S/O model produced 600 optimal solutions in the form of a Pareto front. The 600 optimal solutions represent 600 potential pumping strategies that could be evaluated and implemented to ensure the sustainable management of the Bonriki aquifer system. An optimal solution meeting decision-makers' management goals can be selected for implementation. However, this may require an analysis of various other trade-off solutions on the Pareto front.

This study makes three major contributions. First, it provides a methodology for incorporating aquifer parameter uncertainty when developing a computationally-feasible coastal aquifer management model. The incorporation of uncertainty in parameters (hydraulic conductivity and porosity) into the management model ensures that the prescribed optimal solutions are robust and reliable. Secondly, this study utilizes ensembles of SVMR surrogate models, which have not been used previously in the domain of S/O-based coastal aquifer management research. Thirdly, it evaluated the management model by implementing it in a regional-scale coastal aquifer system. Previous management models have mainly been evaluated using illustrative or hypothetical aquifer systems based on assumptions and, therefore, without calibration and validation of the numerical models involved. The calibration and validation process outlined in this chapter, and the implementation and evaluation of the developed method of salinity prediction for the Bonriki aquifer, are advancements in the field of saltwater intrusion management research. Adoption of the proposed management model and its prescribed groundwater pumping patterns are strongly recommended for the Bonriki aquifer. This may help to ensure the sustainable management of the fragile groundwater resources of Kiribati, thus ensuring social and economic stability. In addition, the proposed coastal aquifer management methodology incorporating parameter uncertainty can be further evaluated and applied to other coastal aquifers prone to saltwater intrusion.

The proposed approach also has some limitations. In the present field-scale application study, the Bonriki aquifer system was considered as a two-layer system due to limited data availability and to ensure the convergence of the developed 3D FEMWATER numerical model. The layers (constructed using limited borehole data) were considered vertically heterogeneous, based on the geological stratification of the layers. However, the materials in each layer were the same, albeit anisotropic. The proposed method can be applied to completely heterogeneous coastal aquifer systems in other geological settings. In addition, the current management model only considered two management strategies based on the two different objective functions. Other management objectives, such as assigning pumping well locations, prescribing optimal operating costs and incorporating recharge wells and other environmental risks, could also be considered. In addition, the influences of tidal fluctuations and seasonal variations on the movement of the saltwater front could also be investigated. These two influences are very important in island aquifer models; however, this study has ignored them for simplicity. These influences could be investigated further and incorporated into the proposed management model. Implementing them would present significant challenges as it would require higher 3D model convergence tolerance, mesh tolerance, computing power, and larger hydrological datasets for numerical model calibration and validation. In the next chapter, an adaptive management methodology is presented in which feedback information from a monitoring network design is used to modify prescribed optimal management strategies while considering user non-compliance and aquifer parameter uncertainty.

Chapter 6: Application of monitoring network design and feedback information for adaptive management of coastal aquifers

A journal article based on the main contents of this chapter has been submitted to a journal for publication and is currently under review, as outlined below:

Lal, A., and Datta, B. (2019). "Application of monitoring network design and feedback information for adaptive management of coastal aquifers subjected to saltwater intrusion." *International Journal of Environmental Research and Public Health*.

6.1 Summary

Optimal strategies for the management of coastal groundwater resources can be derived using coupled simulation-optimization-based management models. However, the management strategies actually implemented in the field sometimes deviate from the recommended optimal strategy, resulting in field-level deviations. Monitoring these field-level deviations during the implementation of a recommended optimal management strategy and sequentially updating the management model using feedback information is an important step towards the efficient adaptive management of coastal groundwater resources. In this study, a three-phase adaptive management framework for a coastal aquifer subject to saltwater intrusion is applied and evaluated for a regional-scale coastal aquifer study area. The methodology adopted includes three sequential components. First, an optimal management strategy stipulating groundwater extraction rates from production and barrier wells is derived and implemented in the aquifer. The implemented management strategy is obtained by solving a homogeneous, ensemble-based, coupled, simulation-optimization model. Second, a regional-scale optimal monitoring network is designed for the aquifer system that considers 1) possible user noncompliance with the recommended management strategy and 2) uncertainties in estimating aquifer parameters. A new monitoring network design objective function is formulated to ensure that candidate monitoring wells are placed in high-risk (highly-contaminated) locations. In addition, a new methodology is utilized to select candidate monitoring wells in areas representative of the entire model domain. Finally, feedback information in the form of salinity data measured at optimal monitoring wells is used to sequentially modify pumping strategies for future time periods in the management horizon. The developed adaptive management framework is evaluated by applying it to the Bonriki aquifer in Kiribati, a small developing South Pacific island country. The results of this study suggest that the implemented adaptive management strategy has the potential to address

important practical implementation issues arising due to noncompliance with an optimal management strategy and uncertainty in aquifer parameters.

6.2 Background

This chapter evaluates the application of a three-phase adaptive management framework for the optimal and sustainable control of saltwater intrusion in coastal aquifers. In Phase 1, an optimal management strategy obtained by solving a coupled simulation-optimization (S/O) model is implemented for the optimal management of the aquifer. The next phase (Phase 2) develops a regional-scale monitoring network for the aquifer that considers user noncompliance of a recommended management strategy and uncertainties in estimating aquifer parameters. In the final phase (Phase 3), feedback information from the optimal monitoring wells is used to sequentially modify/update pumping strategies for future time periods in the management horizon.

The coupled S/O model provides optimal strategies for the management of coastal groundwater resources. However, the correct implementation of recommended optimal management strategies on the field is always a concern for decision-makers. To monitor field-level deviations from a recommended management strategy due to uncertainty in aquifer parameters and non-compliant activity, a well-designed, robust and efficient groundwater monitoring network is essential. The major objectives, criteria and procedures for designing reliable groundwater monitoring networks can be found in Yangxiao (1994). Some key reasons for formulating and developing a groundwater monitoring network include groundwater level monitoring (Kumar et al. 2005; Prinos et al. 2002; Yang et al. 2008; Zhou et al. 2013), contamination detection (Dhar and Datta 2007; Hudak and Loaiciga 1992; Mahar and Datta 1997; Meyer et al. 1994; Prakash and Datta 2013; Storck et al. 1997; Zhu et al. 2019), groundwater quality assessment (Ammar et al. 2008; Baalousha 2010; Loaiciga 1989; Masoumi and Kerachian 2010; Mogheir and Singh 2002) and conflicting economical/financial factors (Destandau and Zaiter 2019; Reed et al. 2000; Zhang et al. 2005). In addition, a comprehensive review by Loaiciga et al. (1992) summarised the most important approaches to consider when designing groundwater monitoring networks. In saltwater intrusion management projects, a properly designed groundwater monitoring network helps to collect data on groundwater quality during and after the implementation of an optimal management strategy. Such field data can be used to assess the compliance of an implemented management strategy with the targeted coastal aquifer management objectives.

A monitoring network is important in achieving the goals of an adaptive coastal groundwater management framework. Adaptive management is crucial in solving problems arising from the field-level implementation of a recommended optimal management strategy. In adaptive management structure, the management strategies for future time periods in the management horizon is sequentially updated using feedback information gathered from the optimal monitoring wells. In the domain of saltwater intrusion research, only a few studies have developed and evaluated adaptive management methods for the management of coastal groundwater resources. Recently, Sreekanth and Datta (2013) developed an adaptive management method for saltwater intrusion control in coastal aquifers based on optimal monitoring network design and the sequential modification of management strategies according to feedback information. Also, Dhar and Datta (2009) designed and implemented a monitoring network to monitor compliance with an optimal aquifer management strategy. However, both of these studies were only validated using models of hypothetical/illustrative coastal aquifer systems. In contrast, the present study assesses the performance of the developed adaptive management methods after they were applied to a real, regional-scale, coastal aquifer system on an island.

The application of a three-phase adaptive management framework for the optimal and sustainable control of saltwater intrusion into a coastal aquifer is of great significance. Adaptive management is an iterative process in which groundwater pumping strategies and policies are regularly revised/updated according to changing pumping conditions and uncertainty in aquifer parameters. The main goal of an adaptive management strategy is to ensure that the prescribed strategies and policies are accepted and correctly implemented in the field. An adaptive management strategy ensures the correct execution of the prescribed policies and will suggest amendments to the optimal pumping strategies that are not correctly followed. The proposed approach emphasizes the practical aspects of implementing a realistic coastal aquifer management strategy especially considering the following two issues. First, the recommended management strategy for optimal temporal and spatial groundwater withdrawals may differ from what actual users implement as often it may be very difficult to enforce the prescribed strategy. Second, even if the actually implemented strategy is identical to the optimal recommended withdrawal strategy, its impacts on the aquifer may be different from predicted impacts due to ubiquitous uncertainties in the estimated and modelled parameters, aquifer boundary conditions, errors in measurements including those in initial conditions and hydraulic heads. Therefore, the need arises to regularly revise the strategy based on measurements obtained from an aquifer monitoring network. Then, a revised management strategy can be derived by re-solving the optimal management model using updated information; e.g., hydraulic head and salinity data. The revised management strategy is an updated version of the initial strategy and, accordingly, its impacts will differ. The revised management strategy increases the likelihood that the original management goals can be achieved. This approach also makes it possible to address the practical issues resulting from deviations from the prescribed pumping strategy or errors in predicting their impacts, which can arise even when the strategy is implemented correctly. The practical utility of this approach is enormous as it provides solutions to the practical difficulties in achieving the goals and objectives of the management model. This study applies this approach to the Bonriki aquifer and evaluates its implications in terms of improving the effectiveness of sequentially updating the optimal pumping strategy using salinity concentration measurements from a designed monitoring network. Hence, this integrated adaptive approach, together with the evaluation of its application, is significant and novel.

This study is a logical extension of the author's earlier work, in which groundwater management mythologies were tested using models of illustrative aquifer systems. The Bonriki aquifer considered in this study is situated in Kiribati, a small, developing, island country in the central Pacific Ocean. This study aimed to present a straightforward and step-wise method for the adaptive management of the Bonriki aquifer system. Specifically, the combined use of a multi-objective optimization model, data clustering and integer programming is used to achieve the management goals. A first-ever monitoring network is designed and implemented for the adaptive management of the Bonriki aquifer. The Bonriki aquifer is a crucial life-sustaining resource for the Kiribati community and, as such, requires sustainable management. Hence, the development and application of the methods presented in this study make significant contributions to the field of sustainable water resource management and administration in Pacific Island developing countries. Also, a recent study by Post et al. (2018) concluded that more work is needed that focuses on the management of groundwater pumping from the Bonriki aquifer and re-evaluates its sustainable yield. This work presents a novel adaptive management framework for the Bonriki aquifer system which, if adopted, would be beneficial to the South Tarawa community. The results and evaluations obtained represent an important step in the regional-scale application of adaptive management strategies to the sustainable management of coastal aquifers.

6.3 Methodology

The coastal aquifer adaptive management framework methodology has three phases, which are presented in the following sections.

6.3.1 Phase 1: Prescription and implementation of an optimal management strategy

The first step towards adaptive management of coastal aquifers is the prescription and implementation of an optimal management strategy. In this study, an optimal management strategy was prescribed using a coupled S/O approach. To reduce computational time and to ensure feasibility, support vector machine regression (SVMR)-based homogeneous ensemble models were used as approximates of the simulation model in the S/O framework. Key details of the homogeneous SVMR ensemble models and the management model are described below.

6.3.1.1 Homogeneous support vector machine regression-based ensemble surrogate models

Homogeneous SVMR ensemble models were used in the S/O model as approximate simulators of the saltwater intrusion numerical simulation model. The SVMR surrogate was used because it is a relatively new and popular supervised data-driven methodology for constructing surrogate models (Liu and Zio 2016). SVMR has been used in numerous recent modelling studies. The newly developed SVMR-based surrogate models have been evaluated for efficiency and accuracy in modelling a hypothetical aquifer. The evaluation results are reported in Chapters 3 and 4. In addition, the results presented in Chapter 3 established that SVMR-based prediction performance was better than that of genetic programming-based surrogate models. The main focuses of this study were on the design of an aquifer monitoring network and adaptive management based on feedback information. Hence, a detailed description of the SVMR working principle is not presented here. However, a thorough description of the SVMR algorithm can be found in Chapter 3 (Section 3.3.3.2). To ensure the robustness of the optimal solutions, ensemble SVMR models were used to incorporate aquifer parameter uncertainty into the management model.

Uncertainty in two aquifer parameters i.e., hydraulic conductivity and porosity was considered while developing the coastal aquifer management model. To train each SVMR model in the ensemble, paired sets of input data (hydraulic conductivity, porosity, and random transient groundwater pumping patterns from active wells) and output data (salinity at monitoring wells) were generated. In this study, the aquifer materials within each layer were considered homogeneous. The specific values of hydraulic conductivity and porosity for each homogeneous layer were obtained from log-normal and normal distributions, respectively. Transient groundwater pumping patterns were generated using a uniformly distributed Latin hypercube sampling (LHS) method. Different combinations of the two uncertain parameters in the respective aquifer layers were implemented into a variable density flow and salt transport numerical model, keeping the other parameters constant during the simulation period. Each set of pumping patterns was provided as input to the variable density flow and salt transport numerical model, with different combinations of the two uncertain parameter values yielding different output concentrations at each specified monitoring well. Of the input-output datasets, 80% were used for training the SVMR models while the remaining 20% were used for validating their performance. The validated standalone SVMR models were strategically combined into an ensemble model using the simple averaging method (Shu and Burn 2004). Each homogeneous ensemble SVMR model was constructed to predict salinity at each monitoring well. The simple average methodology for constructing ensemble models is a popular technique and has been used in various research applications (Bishop 1995; Perrone and Cooper 1995). A mathematical expression for constructing an ensemble model (E_n) by integrating the predicted outputs (P) of various standalone models (n) using the simple averaging methodology is given in Eq. 6.1.

$$E_n = \frac{1}{N}\sum_{n=1}^{N} P^n \qquad n = 1,2,\ldots,N \qquad (6.1)$$

Once trained and validated, the predictive capabilities of the standalone SVMR and ensemble models were quantified using various statistical indices, such as root mean square error (RMSE), mean bias error (MBE), correlation coefficient (R), Nash-Sutcliffe efficiency (NSE) and index of agreement (IOA).

6.3.1.2 Formulation of the multi-objective coastal aquifer management model

The main aim of the developed management model was to prescribe optimal pumping strategy for two groups of wells (production and barrier wells) and simultaneously maintain salinity concentration in the aquifer within specified permissible limits. Freshwater pumping wells (FPWs) were designed to extract fresh groundwater for local consumption. Barrier wells (BWs) were installed closer to the sea-side boundary and were used to extract saline water. BWs acted as a hydraulic barrier, thus preventing saltwater intrusion into the aquifer. The mathematical formulation of the two conflicting objectives, constraints and bounds considered in the management model are given in Section 4.4.9 (Chapter 4).

6.3.2 Phase 2: Regional-scale monitoring network design

6.3.2.1 Possible deviations in pumping rates and aquifer parameter uncertainty

One of the key features of a monitoring network is that it should be able to accommodate possible deviations in field-level implementation of the prescribed optimal pumping strategy and also uncertainties associated with the aquifer parameters. In this study, uncertainty due to possible field-level deviations in implementation of a prescribed optimal pumping strategy, and uncertainty in aquifer parameter (hydraulic conductivity and porosity) estimates were considered in the design of an optimal monitoring network. First, to consider field-level deviation in the chosen optimal management pumping solution, slightly perturbed optimal pumping rates were utilized. To achieve this perturbation, a truncated normal distribution of the deviations, ranging from 0-20% of the actual deviations in pumping rates, similar to

Sreekanth and Datta (2013), was considered. Second, uncertainty in aquifer parameter estimates was incorporated by utilizing different realization of the hydraulic conductivity and porosity values in the variable-density flow and salt transport numerical simulation model. Different realization of hydraulic conductivity and porosity were obtained using lognormal and normal distributions, respectively. The perturbed input optimal pumping rates and different realization of the two uncertain aquifer parameters were used in the variable-density flow and salt transport numerical simulation model to obtain different realizations of salinity concentration at all the candidate (potential) monitoring wells.

6.3.2.2 Location of candidate monitoring wells

The location of candidate monitoring wells demands careful consideration. A subset of these candidate monitoring wells will be selected as optimal monitoring wells. Many times, only certain areas of the model domain are randomly chosen and used as locations of candidate monitoring wells. However, in real scenarios, any possible node on the model domain can be considered as a candidate location for a monitoring well. In this study, *k*-means clustering (MacQueen 1967) was utilized to determine the locations of candidate monitoring wells that would be representative of the entire model domain. Clustering of all existing nodes using the *k*-means clustering methodology ensured that candidate monitoring wells were chosen from the entire area of the model domain. The *k*-means clustering is a distinctive clustering algorithm that offers an efficient and simple method of data clustering (Žalik 2008). Detailed explanations of the *k*-means clustering methodology are presented in Bandyopadhyay and Maulik (2002) and (Nazeer and Sebastian 2009). The main idea of using *k*-means clustering is to categorize the set of nodes into *k* disjoint clusters, where *k* is fixed in advance. After convergence, the *k*-means clustering solution offers a centroid for each of the clusters. The node number closest to this centroid is chosen as the node, indicating where a candidate monitoring well is to be installed.

6.3.2.3 Formulation of the optimal monitoring network model

The main goal of designing an optimal monitoring network was to monitor compliance with the management strategy. This is achieved by comparing salinity concentration resulting from an optimal management strategy (prescribed) with the actual concentrations (measured on the field). The salinity measurements from a designed monitoring network also provide feedback on the actual impacts of the field-level implementation of a management strategy. This feedback information can be utilized to redesign the management strategy to better achieve its objectives. In many cases, monitoring data can be collected from numerous locations in an aquifer. However, this may be impractical and inefficient due to restrictions in budget and data redundancy (Dhar and Datta 2009). Therefore, a key feature of a monitoring network design is to locate the permissible number of monitoring wells (within budgetary limits) at locations suitable for collecting useful and reliable monitoring data. For the present study, the mean of the logarithmic salinity concentration at each candidate monitoring well was maximized to ensure that candidate monitoring wells were placed in high-risk areas (with high salinity). The objective function with respective constraints is described below.

$$\text{Maximize } \frac{\sum_{i=1}^{N} log(C_i)}{N} Y_i \tag{6.2}$$

$$\text{subject to; } \sum_{i=1}^{N} Y_i \leq M \tag{6.3}$$

$$Y_i \in (0,1)$$

Where C_i is the concentration at the i^{th} candidate monitoring well and Y_i is the decision variable indicating whether to install a monitoring well $(Y_i = 1)$ or not $(Y_i = 0)$ at the i^{th} location. Variable M represents the maximum number of monitoring wells permitted in the monitoring network (due to budgetary or other management limitations). Phase 2 is designed to obtain optimal monitoring well locations. As an adaptive measure, feedback information in the form of salinity concentration data will be used to sequentially modify future year pumping rates. The monitoring network was designed once; however, the information from this network of wells is used sequentially for the modification of future year pumping rates.

6.3.3 Phase 3: Sequential modification of the management strategy

Optimal coastal aquifer management strategies for the sustainable control of saltwater intrusion are largely developed for longer time horizons, *T*. However, they are implemented in smaller time-steps *t*. With the help of a properly designed monitoring network, it is possible to gather feedback information regarding compliance with a management strategy based on the comparison of measured and predicted salinities. This information can be utilized to sequentially modify and/or update the management strategy at succeeding time steps, thereby improving the prospects of attaining its goals.

In this study, a management time horizon of four years ($T = 4$) was considered. However, the management strategy was implemented in steps of one year; i.e., t = 1, 2, 3 and 4. The implemented four-year optimal pumping strategy (selected from the Pareto-front) was obtained by solving the homogeneous ensemble SVMR surrogate-based coupled S/O model. Yearly salinity concentration data from the monitoring network due to implemented pumping rates can be obtained using the variable-density flow and salt transport numerical model. However, in real field scenarios, it is common practice that the prescribed optimal pumping strategy will not be accurately implemented, and/or, the actual concentrations resulting from an implemented strategy will differ from the predicted impacts due to various uncertainties in prediction. In this study, for performance evaluation purposes only, the field level-deviations between actual and predicted concentrations after first year of the implementation (t=1) of optimal pumping strategy is incorporated by simulating the concentrations taking into account random deviations of actual pumping rates from prescribed pumping rates. The perturbed pumping rates different from optimal prescribed pumping rates are generated by adding random errors (0-20%) to the optimal prescribed pumping rates. The inclusion of such deviations will affect the resulting salinity and may lead to noncompliance with management goals in terms of permissible salinities. Again, for performance evaluation purposes, the actual salinity concentrations at the designed optimal monitoring wells for each time step can be simulated for the perturbed pumping rates, using the variable-density flow and salt transport numerical simulation model. In actual applications, no such artificial perturbation is required, as the actual salinities will be measured in the field using the monitoring network. The monitoring network (designed to gather information about noncompliance with the prescribed management strategy) helps in updating or revising the management strategy for t = 2, 3 and 4 to achieve the original management goals. The multi-objective coupled S/O model was utilized to sequentially update the management strategy using feedback information from the earlier time steps. This was performed by re-running the S/O model for future time steps after updating the initial and boundary conditions according to feedback from the monitoring network. For sequential modification of the pumping rates using feedback information's, the multi-objective management model was solved as a single objective optimization problem. Objective 2 (Eq. 4.6) of the original multi-objective management model (i.e. barrier well pumping for the selected management strategy) was added as an additional constraint. The other constraints (Eqs. 4.7 and 4.8) and bounds (Eqs. 4.9 and 4.10) of the optimization problem remained unchanged.

6.3.4 Case study: Application and evaluation of the developed method

The developed coastal aquifer adaptive management methodology was applied to the Bonriki aquifer. A detailed description of the study area is given in Chapter 4. Groundwater extracted from the Bonriki aquifer is the main source of freshwater for the people of South Tarawa (White et al. 1999). Approximately 60% of the South Tarawa population are dependent on groundwater extracted from the Bonriki aquifer (Metutera 2002). The FEMWATER computer package was used to construct a 3D numerical simulation model of the Bonriki aquifer. A detailed description of the modelling paradigm and boundary conditions, the hydrogeology of the area, and field data are given in Sections 4.3.2, 4.4.2, 4.4.3 and 4.4.5

(Chapter 4). The calibration and validation procedure of the numerical model with the results are also given in Section 4.4.5 (Chapter 4).

The validated model was used to evaluate the adaptive management framework. Firstly, the multi-objective management model was evaluated using the developed variable-density flow and salt transport numerical model. All 19 operational FPWs and 6 BWs were used. A total management horizon of four years was considered. A total of 100 decision variables (25 wells × 4-year management horizon) was implemented into the S/O model. The maximum (1200 m^3/day) and minimum (0 m^3/day) pumping rates for both well types were added as bounds. The maximum allowable salinity concentration at the six MWs at the end of the management horizon were incorporated as optimization constraints. Each SVMR model was trained and validated using 700 input-output datasets. The 700 sets of randomised input pumping rates (from both well types) were generated using the LHS method. The maximum (1200 m^3/day) and minimum (0 m^3/day) pumping rates for both well types were added as bounds. Each set of generated input pumping rates was fed separately into the variable density flow and transport numerical model to obtain salinity concentration data at the respective monitoring wells. Input pumping and output concentration datasets were used to train and test an SVMR surrogate model for each well location. Training and testing of the SVMR models in the ensemble were done on the MATLAB 2017a platform. To achieve satisfactory predictions, a Gaussian kernel was used with parameters ε, C and γ having values of 0.60, 10 and 0.001, respectively. These parameter values were obtained after numerous experimental runs. Each standalone SVMR model could only predict salinity concentration at a specific monitoring well. Ten different combinations of hydraulic conductivity and porosity values were used to develop ten SVMR models for each monitoring well. The predictions of these ten SVMR models were integrated into an ensemble. Thus, six ensemble SVMR models were developed to predict salinity at the six corresponding MWs. The hydraulic conductivity values were derived from a lognormal distribution with the calibrated value of hydraulic conductivity as the mean, and a variance of 0.4 m/d. Similarly, porosity values were derived using a normal distribution with a calibrated value of porosity as the mean and a variance of 0.1.

The validated SVMR ensemble models were externally coupled to the multi-objective genetic algorithm (MOGA) optimisation model using the MATLAB 2017a platform. One of the main reasons for using MOGA is its efficiency. In a single run, MOGA can provide an optimal Pareto-front comprising non-dominated solutions at the end of the stated number of generations. The MOGA model used a population size of 2000, function tolerance of 1×10^{-4}, constraint tolerance of 1×10^{-3}, Pareto front population fraction of 0.3, and crossover fraction of 0.8. The number of generations was fixed to 10,000. This value was obtained after trying different generation sizes for the convergence of the population. The constraint of the optimization (maximum allowable salinity at the 6 MWs) ensured that the salinity at the MWs was restricted to a pre-specified limit. The maximum tolerable salinities for MW1 and MW2 were set to 20,000 mg/L. MW1 and MW2 were closer to the shoreline and restricting the salinity at these locations to lower levels was impractical. The maximum acceptable salinities at MW3 and MW4 were set to 5000 mg/L and 4000 mg/L, respectively. Finally, the maximum salinities at MW5 and MW6 were set to 450 mg/L. MW5 and MW6 were located in an area with many pumping wells. It was anticipated that the water extracted from these locations was of good quality and suitable for consumption by the local South Tarawa communities.

For Phase 2, 100 candidate monitoring well locations were chosen using the *k*-means clustering method. The *k*-means clustering code was written and executed in the *R* platform. A fixed number of iterations was used as the stopping criterion (Zio and Bazzo 2010). In the present case, 50 iterations were considered. Also, before the execution of the *k*-means clustering code, the number of candidate monitoring wells to be used in the monitoring network was specified as the value of *k* (number of clusters). One hundred perturbed pumping rates from the chosen optimal pumping rates were obtained using the LHS strategy. These perturbed pumping rates and 100 combinations of hydraulic conductivity and porosity were used in the variable density flow and salt transport numerical simulation model to obtain 100 different realizations of salinity concentration at the 100 candidate monitoring wells. Ten (M = 10) optimal MWs out of the 100 candidate monitoring well locations were obtained by implementing the designed monitoring network. For Phase 3, the single objective optimization problem used for sequential modification of future year pumping rates was solved using the genetic algorithm optimization solver available in the MATLAB 2017 platform. A flowchart of the proposed adaptive management framework is presented in Fig. 6.1.

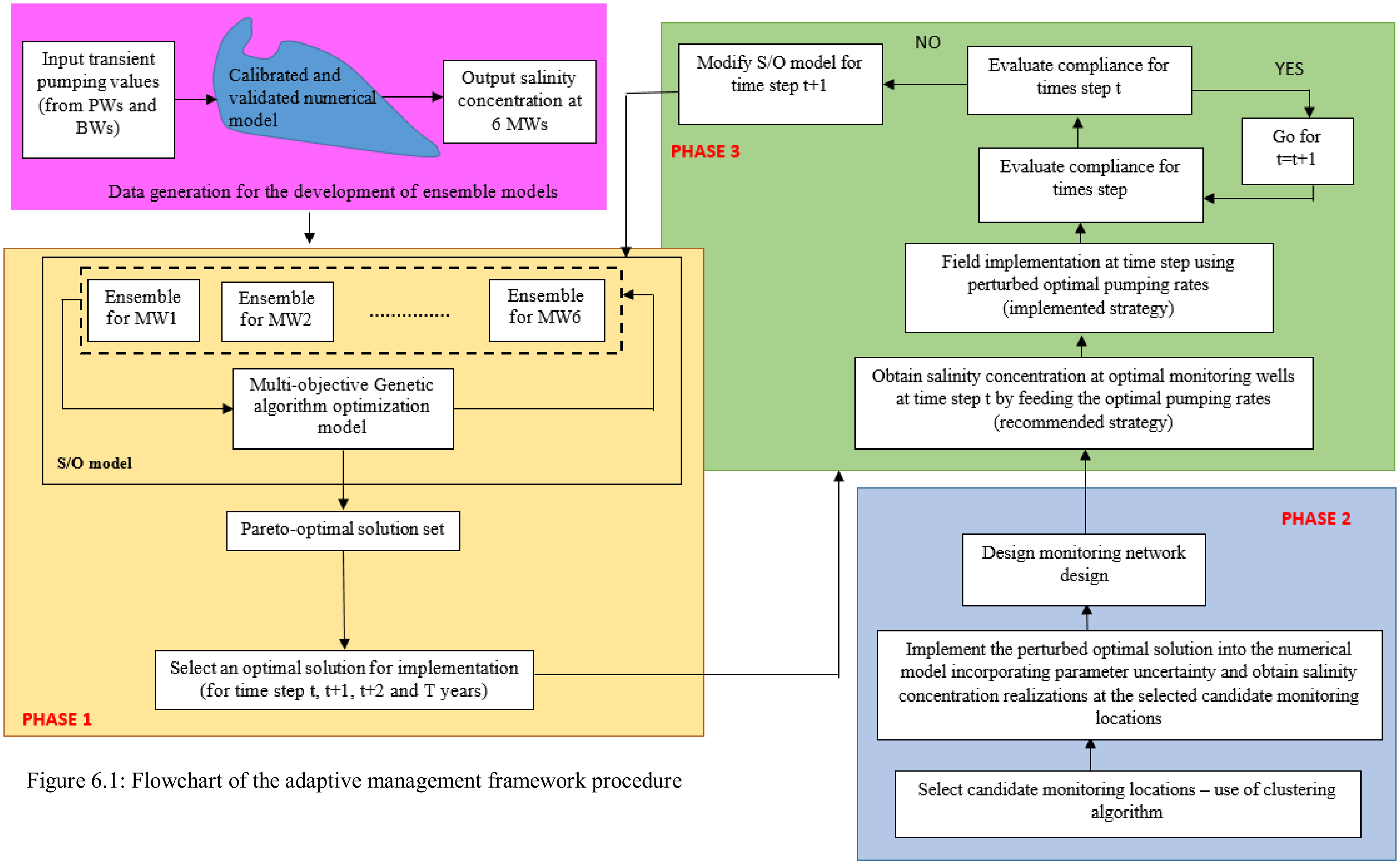


Figure 6.1: Flowchart of the adaptive management framework procedure

6.4 Results and discussion

6.4.1.1 Performance evaluation of the homogeneous ensemble models

The utility of the integrated approach to management strategy development, implementation, and subsequent modification based on feedback measurements obtained from a monitoring network were evaluated to establish its potential applicability to the study site. The homogeneous SVMR ensemble model was used to approximately simulate aquifer responses in the coupled S/O model. The homogeneous SVMR ensemble model consisted of ten standalone SVMR models. Each standalone SVMR model was trained and tested using datasets obtained from different variable-density flow and salt transport numerical models developed using different combinations of hydraulic conductivity and porosity values. The predictive accuracy of each standalone model in the ensemble is shown in Table 6.1. It was observed that all the standalone models in the ensemble predicted salinity at the monitoring wells with reasonable accuracy (quantified in terms of RMSE, MBE, R, NSE and IOA). This accuracy of the standalone models was eventually reflected in the accuracy of the ensemble models used in the coupled S/O model. The performance of the homogeneous SVMR ensemble surrogate model is presented in Table 6.2.

Table 6.1: Performance evaluation results of the standalone SVMR models

Model	Evaluation measure	$SVMR_1$	$SVMR_2$	$SVMR_3$	$SVMR_4$	$SVMR_5$	$SVMR_6$
NM1	RMSE	5.10	6.17	3.74	2.95	2.02	1.89
	MBE	0.41	0.45	0.38	0.41	0.39	0.35
	R	0.96	0.97	0.97	0.97	0.98	0.98
	NSE	0.97	0.96	0.98	0.97	0.98	0.98
	IOA	0.94	0.95	0.95	0.96	0.96	0.96
NM2	RMSE	5.98	5.62	2.82	2.04	1.59	1.33
	MBE	0.47	0.56	0.62	0.52	0.39	0.44
	R	0.97	0.97	0.98	0.97	0.98	0.98
	NSE	0.96	0.96	0.96	0.97	0.97	0.97
	IOA	0.95	0.94	0.95	0.96	0.96	0.96
NM3	RMSE	4.16	5.22	3.51	4.86	3.02	2.14
	MBE	0.71	0.43	0.48	0.47	0.38	0.31
	R	0.97	0.96	0.97	0.96	0.98	0.98
	NSE	0.97	0.97	0.98	0.97	0.98	0.99
	IOA	0.94	0.95	0.95	0.94	0.95	0.96
NM4	RMSE	6.60	5.33	5.27	4.65	3.53	3.05
	MBE	0.52	0.55	0.64	0.64	0.43	0.48
	R	0.97	0.98	0.96	0.97	0.97	0.97
	NSE	0.97	0.96	0.96	0.97	0.97	0.98
	IOA	0.95	0.94	0.93	0.95	0.95	0.96
NM5	RMSE	6.96	7.13	5.12	5.68	4.25	4.56
	MBE	0.59	0.63	0.72	0.52	0.33	0.36
	R	0.97	0.97	0.97	0.97	0.98	0.97
	NSE	0.97	0.98	0.98	0.97	0.98	0.97
	IOA	0.95	0.94	0.95	0.94	0.96	0.95
NM6	RMSE	7.63	5.32	5.24	5.69	4.25	3.57
	MBE	0.44	0.65	0.66	0.46	0.41	0.34
	R	0.97	0.98	0.98	0.97	0.99	0.99
	NSE	0.98	0.98	0.98	0.98	0.98	0.99
	IOA	0.96	0.97	0.97	0.97	0.98	0.98
NM7	RMSE	7.26	6.75	6.03	5.87	5.66	5.12
	MBE	0.58	0.62	0.55	0.47	0.44	0.36
	R	0.97	0.98	0.98	0.98	0.98	0.99
	NSE	0.97	0.98	0.98	0.98	0.98	0.99
	IOA	0.96	0.97	0.97	0.97	0.98	0.98
NM8	RMSE	6.35	7.16	5.57	5.31	5.26	5.19
	MBE	0.54	0.58	0.52	0.49	0.33	0.41
	R	0.98	0.97	0.98	0.98	0.98	0.98
	NSE	0.97	0.97	0.97	0.97	0.97	0.97
	IOA	0.96	0.95	0.96	0.96	0.96	0.96
NM9	RMSE	7.37	6.89	8.43	6.22	5.32	4.41
	MBE	0.59	0.63	0.67	0.56	0.42	0.38
	R	0.98	0.98	0.96	0.97	0.97	0.98
	NSE	0.97	0.97	0.96	0.97	0.97	0.98
	IOA	0.95	0.96	0.95	0.96	0.97	0.97
NM10	RMSE	7.14	6.59	6.91	5.88	4.71	4.28
	MBE	0.55	0.62	0.52	0.55	0.39	0.44
	R	0.98	0.98	0.98	0.98	0.98	0.99
	NSE	0.98	0.98	0.98	0.98	0.98	0.99
	IOA	0.96	0.97	0.97	0.97	0.97	0.98

Table 6.2: Performance evaluation results of the homogeneous ensemble models

Evaluation measure	En_SVMR_1	En_SVMR_2	En_SVMR_3	En_SVMR_4	En_SVMR_5	En_SVMR_6
RMSE	4.70	5.61	3.34	2.99	2.16	1.79
MBE	0.42	0.44	0.36	0.41	0.33	0.31
R	0.97	0.97	0.98	0.97	0.98	0.98
NSE	0.98	0.97	0.98	0.98	0.98	0.98
IOA	0.96	0.96	0.96	0.97	0.97	0.97

En_SVMR_n = homogeneous ensemble surrogate model developed using 10 standalone SVMR surrogate models for monitoring well n

6.4.1.2 Implementation of the optimal aquifer management strategy

The executed homogeneous SVMR ensemble-based coupled S/O model presented a Pareto-front containing several trade-off optimal solutions in a runtime of about three hours. Each optimal solution on the Pareto-front represents an optimal pumping strategy that can be implemented as a management policy. The total optimal pumping rates from all the FPWs and all the BWs for the four-year management horizon ranged from about 30,000–44,000 m^3/day and 1000–9000 m^3/day, respectively. These pumping rates were based on the imposed constraints (permissible concentration limits placed at the different MWs) specified in the management model. The optimal pumping values were also within the specified bounds used in the optimization model. The maximum and minimum annual rainfall in Tarawa is approximately 4300 mm and 2100 mm, respectively (Bosserelle et al. 2015). Based on this annual amount of rainfall over a highly permeable aquifer top cover with the proportionately very small built-up area, it is reasonable to assume a vertical annual recharge rate of nearly 2000 mm. This vertical recharge amount is itself around 3 million m^3 per year. Therefore, if the BW extraction rate is excluded from the total withdrawal amount computed above, as a large proportion of BW extraction is contributed by the sea face constant head boundary, the total specified withdrawal from FPWs nearly matches the estimated vertical recharge. Therefore, the recharge rate imposed appears to be reasonable.

Validation of these optimal solutions is a crucial step in an S/O management framework. Validation of optimal solutions was carried out by randomly selecting a few optimal solutions from the Pareto-front and implementing them in the original variable-density flow and salt transport numerical model. In this study, five random optimal solutions were implemented into each of the ten variable-density flow and salt transport numerical models and ten standalone SVMR surrogate models. The average of the concentration values of the ten variable-density flow and salt transport numerical models were compared with those of the homogeneous SVMR ensemble surrogate models. These comparisons are presented in Table 6.3. The relative errors in this comparison were less than 5% at all MWs. This establishes the fact that the homogeneous ensemble SVMR surrogate model approximated the variable flow and salt transport model with reasonable accuracy. Also, it was observed that the salinity values converged to the upper limit of the set constraints. For example, at MW1, the maximum allowable salt concentration in the optimization model was specified as 20,000 mg/L. In the comparison result presented in Table 6.3, it is seen that for all five selected optimal solutions, the salinities converged to the upper limit (20,000 mg/L). A similar pattern was observed for all other MWs.

The main focus of this study was to design a monitoring network for the Bonriki aquifer. For this purpose, a randomly-selected optimal solution (solution *R* in Fig. 6.2) was selected and implemented as a coastal aquifer management strategy. The total production well and barrier well pumping rates for the selected optimal solution were 39,728.44 m^3/day and 3630.89 m^3/day, respectively. The specific pumping rates for each FPW and BW for the selected management strategy are shown in Fig. 6.3.

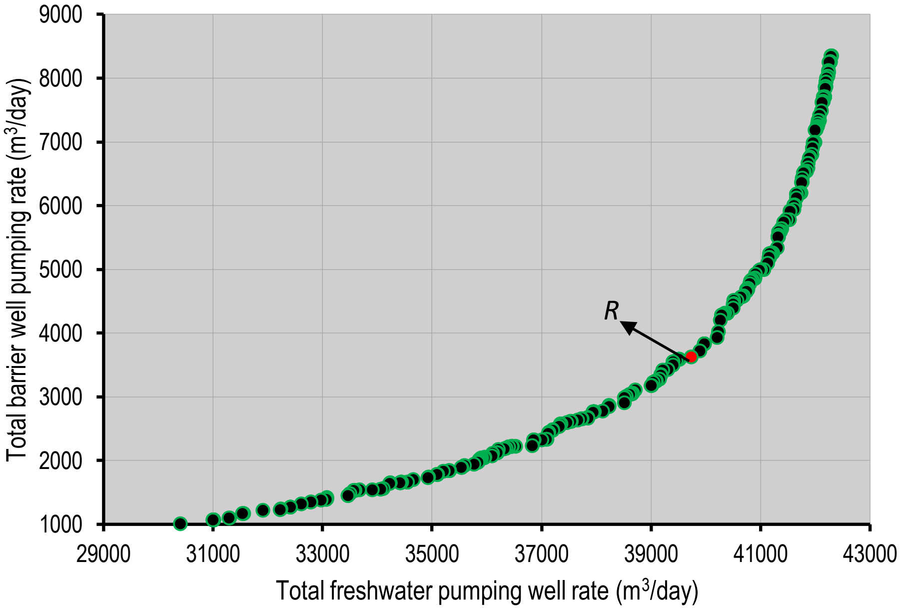


Figure 6.2: Pareto-front displaying various trade-off optimal solutions

Table 6.3: Optimal solution validation results

Solution number	MW1			MW2			MW3			MW4			MW5			MW6		
	NM_{av} (mg/L)	En_SVMR_1 (mg/L)	RE (%)	NM_{av} (mg/L)	En_SVMR_2 (mg/L)	RE (%)	NM_{av} (mg/L)	En_SVMR_3 (mg/L)	RE (%)	NM_{av} (mg/L)	En_SVMR_4 (mg/L)	RE (%)	NM_{av} (mg/L)	En_SVMR_5 (mg/L)	RE (%)	NM_{av} (mg/L)	En_SVMR_6 (mg/L)	RE (%)
1	19870.4	19677.0	1.0	19795.63	19742.53	0.3	4868.18	4844.85	0.5	3963.43	3932.23	0.8	447.82	444.94	0.6	432.72	429.88	0.7
2	19708.8	19621.2	0.4	19783.73	19735.64	0.2	4811.80	4776.57	0.7	3959.52	3916.66	1.1	433.94	429.27	1.1	435.97	433.33	0.6
3	20009.2	19846.9	0.8	19975.89	19949.89	0.1	4931.85	4928.47	0.1	3944.47	3911.19	0.8	444.84	437.12	1.7	436.18	431.45	1.1
4	19798.4	19660.5	0.7	19793.16	19757.80	0.2	4915.77	4906.76	0.2	3868.29	3868.82	0.0	433.54	427.58	1.4	439.93	436.34	0.8
5	19727.4	19567.6	0.8	19829.44	19795.19	0.2	4828.40	4839.35	0.2	3972.34	3967.71	0.1	430.35	427.29	0.7	427.90	426.47	0.3

NM_{AV} = mean salinity values from all 10 variable density flow and salt transport numerical models; RE = relative error

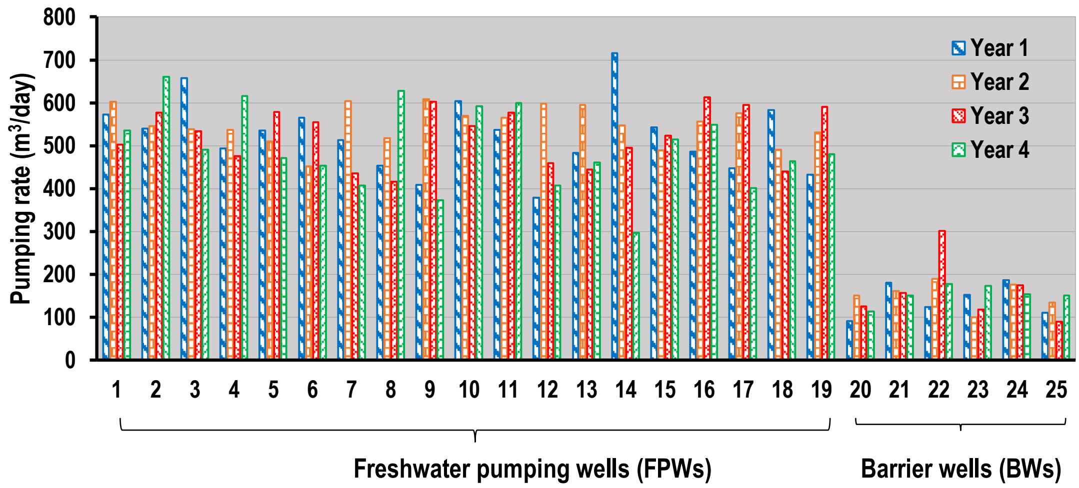


Figure 6.3: Pumping rates specified (recommended) by the management strategy

6.4.2 Optimal monitoring wells

The locations of candidate monitoring wells were obtained using the *k*-means clustering method and are presented in Fig. 6.4. The clustering method ensured that the candidate monitoring wells were scattered over the entire model domain. The truncated optimal pumping patterns (to demonstrate field-level deviations) and uncertain aquifer parameters were used to generate 100 salinity concentration realization at 100 candidate monitoring wells. Out of the 100 candidate monitoring wells, only 10 were selected as optimal monitoring wells. The monitoring network optimization formulation using the LINGO 17 platform (Scharage 1999) presented ten optimal monitoring wells, which are presented in Fig. 6.5. The average of the logarithmic salinity at each candidate monitoring well was maximized to ensure that candidate monitoring wells were placed in high-risk areas (high salinity areas). As seen in Fig. 6.5, the locations of all optimal monitoring wells were close to the seaside boundary, where salinities due to production well pumping were the highest. Also, a good spread of optimal monitoring wells was observed, which avoids redundancy in monitoring well installation. The objective function used in the design of a monitoring network is only one possible objective. Other objectives based on different study area management scenarios can also be considered. The locations of the optimal monitoring wells were dependent on the monitoring network design's objective functions and will differ according to the monitoring network objectives. For example, maximizing the weighted mean salinities at candidate monitoring locations, as an objective function, will result in a different set of optimal monitoring wells compared to the set designed in this study. However, for this particular study, a simple objective function was used to highlight the other aspects of linked S/O models and the use of sequential information to modify management strategies over time.

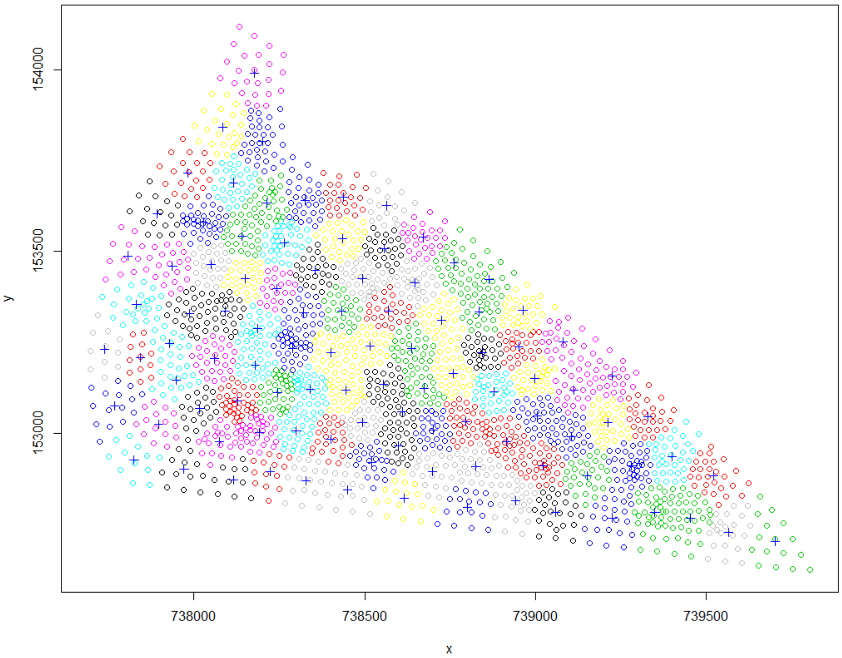


Figure 6.4: Locations of the 100 candidate monitoring wells (+)

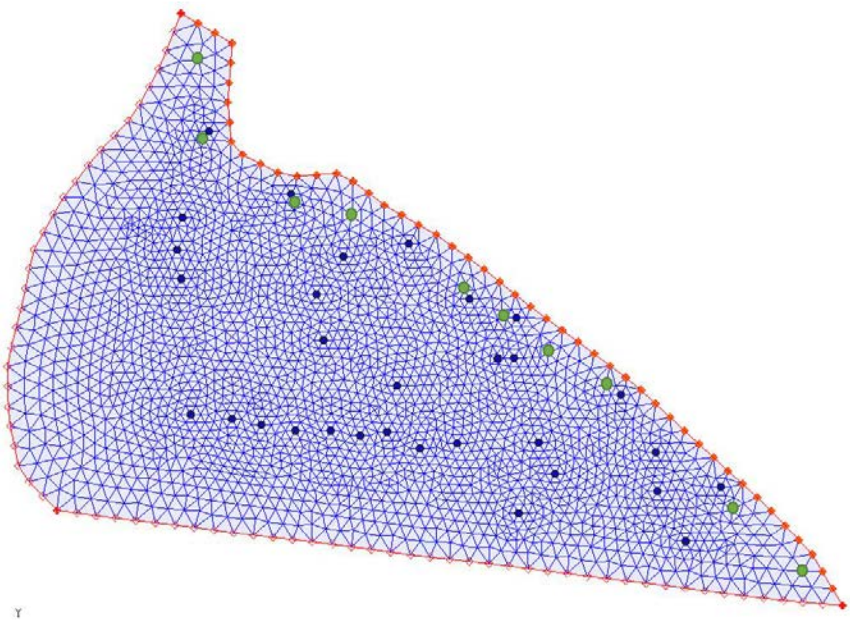


Figure 6.5: Locations of the 10 optimal monitoring wells (green circles)

6.4.3 Modifying pumping rates using feedback information

The management strategy selected from the Pareto-front was used to adaptively modify pumping rates based on deviations in salinity concentration data; i.e., the difference in salinities obtained after using the recommended (optimal) and actually implemented management strategy. The selected recommended four-year management strategy was modified based on feedback information obtained based on the preceding year's implemented pumping rates. For the selected management strategy, the total production well pumping and barrier well pumping rates were 39,728.44 m^3/day and 3630.89 m^3/day, respectively. The resulting salinity values obtained at the optimal monitoring wells at the end of year 1 as a result of the recommended pumping strategy are given in Table 6.4 (Situation A of year 1). It is highly likely that the pumping rates recommended by the management strategy will not be exactly implemented in the field. To cater to this field-level deviations, the recommended production well and barrier well pumping rates were perturbed by 0–20%. This perturbation reflects actual pumping rates implemented in the field. However, this step is only relevant for performance evaluation purposes. In actual field applications, deviations from recommended rates will be measured at monitoring wells in a monitoring network. The salinity concentration at the optimal monitoring wells at the end of year 1 as a result of the perturbed pumping rates are given as Situation B of year 1. It is observed from Table 6.4 that the deviations in pumping rates in year 1 causes a slight deviation in the salinity values. Based on these values, the pumping rates for years 2, 3 and 4 were modified by rerunning the coupled S/O model while keeping the other management constraints unchanged. The modified FPW pumping rates are given in Table 6.5. The modifications were only applied to the pumping rates for years 2, 3 and 4 after gathering feedback on the implementation of the year 1 pumping rates. These three changes were obtained based on the revised solution of the optimization model i.e., using using feedback information from the optimal monitoring wells. The first year's pumping rates remained unchanged.

Table 6.4: Salinities (mg/L) at the optimal monitoring wells

	Year 1		Year 2		Year 3		Year 4	
OML*	Situation A	Situation B	Situation A	Situation B	Situation A	Situation B	Situation A	Situation B
1	24,168.14	24,135.24	25,951.33	25,612.34	27,952.67	27,956.33	30,215.25	30,258.11
2	23,256.87	23,247.65	24,696.11	24,616.56	26,151.85	26,146.51	27,298.55	27,204.65
3	23,055.84	23,016.28	24,856.22	24,843.22	25,871.43	25,886.93	26,598.29	26,577.16
4	24,136.82	24,089.56	24,623.64	24,647.89	25,026.94	25,049.66	26,884.34	26,813.55
5	17,452.31	17,486.33	17,898.23	17,954.62	19,560.05	19,587.79	22,389.71	22,384.02
6	19,585.62	19,546.20	20,114.95	20,168.77	22,895.44	22,905.68	25,468.92	25,423.96
7	23,656.98	23,641.25	24,891.33	24,923.56	26,454.72	26,484.66	28,355.79	28,397.10
8	24,556.28	24,587.34	25,831.25	25,838.51	27,206.84	27,198.23	28,113.95	28046.82
9	23,584.02	23,546.95	24,669.27	24,646.85	26,158.34	26,144.26	27,138.24	27138.16
10	25,136.57	25,136.18	26,882.16	26,876.29	28,654.23	28,679.34	30,218.97	30158.49

Situation A = due to recommended strategy; Situation B = due to implemented strategy; OML = optimal monitoring well

Table 6.5: Modified production well pumping rates (m^3/day) over the entire management period

	Year 1		Year 2		Year 3		Year 4	
	R	I	R	I	R	I	R	I
Year 1	9946.89	9822.57						
Year 2	10,426.51		10,335.12	10,378.77				
Year 3	9957.58		9987.14		9874.23	9904.38		
Year 4	9397.47		9414.87		9369.16		9325.29	-
Total	39,728.44		29,737.13		19,243.39		9325.29	

R^* = recommended; I^* = implemented

The implemented year 1 pumping rates (9288.57 m^3/day) and the future modified year 2, 3 and 4 modified pumping rates (29,737.13 m^3/day) was 39,559.70 m^3/day. This was less than the pumping rates from the originally recommended management strategy. The salinity concentration due to the modified recommended pumping rates at the end of year 2 is given in Table 6.4 (Situation A of year 2). Situation B of year 2 is the salinity concentration due to the implementation of the perturbed pumping rates. Similarly, Situations A and B for years 3 and 4 in Table 6.4 represent the salinities resulting from the modified and perturbed pumping rates, respectively. The pumping rates for future years 3 and 4 were modified using the same procedure as discussed above. The modified objective function value for year 3 was 39,444.73 m^3/day, which is also less than the total pumping rates of the original recommended management strategy. With the adaptive management framework, the total of 4 years' implemented pumping rates were 39,431.01 m^3/day, which is less than the pumping rates of the recommended management strategy (39,728.44 m^3/day). This is intuitively justifiable, as the actually implemented strategy for the initial year was suboptimal.

The results presented in Tables 6.3 and 6.4 demonstrate that optimal pumping solutions recommended from the S/O model will need modification because of the field-level deviations encountered during the implementation process, or noncompliance by the user. As observed, the feedback salinity measurement information can be utilized to modify pumping rates for the remaining future time periods of the management horizon. Therefore, a properly designed optimal monitoring network and feedback information are crucial for the adaptive management of coastal groundwater resources.

In general, the adaptive management method evaluated for the Bonriki aquifer system has given promising results. Solving the multi-objective management model prescribed a set of optimal solutions in the form of a Pareto-front. The obtained solutions were validated to ensure that the constraints were satisfied. Obtaining an optimal solution and then exactly implementing it in the field are two different but critical issues discussed in this study. An optimal solution can be obtained using all the computational powers at hand. The issue of user non-compliance due to incorrect implementation of a selected optimal solution is the main concern addressed. To monitor the possible effects of user non-compliance and to update the subsequent time period's optimal solution in order to rectify it, feedback information in the form of salinity data was obtained from the optimal monitoring wells. The subsequently modified yearly pumping strategies help to achieve the original management goals despite of earlier deviations from the prescribed strategy. All the computational powers can be used to develop an optimal pumping strategy for the aquifer system, but the question of user non-compliance remains the same and is not entirely a computational issue. In practical situations, we cannot guarantee that an optimal solution will be correctly implemented. In such scenarios, the adaptive management framework presented in this study will be useful. Theoretically, it is possible to search for and then identify an optimal solution by enormous enumerations; however, for complex large-scale problems, this is totally impractical. The main contribution of optimization is that it can efficiently search for an optimal solution that is almost impossible to identify by enumeration. When more than one objective is considered, this becomes more critical.

6.5 Conclusions

This study demonstrated the use of an integrated approach to the adaptive management of an island aquifer system. This involved formulating an optimal management strategy, designing an optimal monitoring network, and obtaining feedback information from it. In achieving the targeted management goals, an optimal production well and barrier well pumping strategy was considered as an option for the sustainable control of saltwater intrusion into the Bonriki aquifer system in the South Pacific. Using this prescribed optimal strategy, optimal monitoring wells were identified. A new monitoring objective function was developed to determine the locations of optimal monitoring wells in high-salinity areas. The resulting optimal monitoring wells were then used to monitor compliance with the recommended management strategy. Based on the field-level deviations between actual and planned salinity levels, the pumping rates for future time periods in the management horizon were revised using the updated coupled S/O model. It is noted that field-level deviations observed during the implementation of recommended pumping rates could lead to significant differences between the salinity concentrations measured at optimal monitoring wells. Hence, updating the management model using the feedback information from earlier time periods could be crucial to the management of the Bonriki aquifer system. The practical aspects of the actual implementation of coastal aquifer management strategies are emphasized in this approach. The solutions presented in this study open pathways for similar studies that could be undertaken in other small island countries, where saltwater intrusion due to excessive groundwater withdrawal threatens the sustainability of freshwater resources. The developed and evaluated adaptive management method has the potential to be applied to other regional-scale coastal aquifers subject to saltwater intrusion. However, such applications would require the development of a variable density groundwater flow and transport numerical model specific to the study area. This would necessitate numerical modelling skills, software and computational resources and groundwater datasets (head and salinity). These datasets are not always readily available and may require rigorous field investigations, which can be costly. In the next chapter, the predictive performance of homogeneous and heterogeneous ensemble models is investigated to establish a better-performing type of ensemble model.

Chapter 7: Comparative performance evaluation of homogeneous and heterogeneous ensemble models for groundwater salinity prediction

A manuscript based on the main contents of this chapter has been submitted for publication and is currently under review, as outlined below:

Lal, A., and Datta, B. (2019). "A comparative performance evaluation of homogenous and heterogeneous ensemble models for groundwater salinity predictions." *Environmental Monitoring and Assessment*.

7.1 Summary

Accurate prediction of salinity concentrations in aquifers in response to fluctuating groundwater pumping patterns is an essential component of any coastal groundwater planning and management framework. Ensemble predictive models are known to be more accurate and robust than standalone predictive models. The present study develops and utilises homogeneous and heterogeneous ensemble models of various standalone evolutionary algorithms, such as artificial neural network (ANN), genetic programming (GP), support vector machine regression (SVMR), and Gaussian process regression (GPR), to predict groundwater salinity concentrations in a small Pacific Island coastal aquifer system. Standalone and ensemble predictive models are trained and validated using identical pumping and salinity concentration datasets obtained by solving a numerical 3D transient density-dependent coastal aquifer flow and transport model. After validation, the ensemble models are used to predict salinity concentration at selected monitoring wells in the modelled aquifer under variable groundwater pumping conditions. The predictive capabilities of the developed ensemble models are quantified using standard statistical procedures. The performance evaluations suggest that the predictive capabilities of the developed standalone prediction models (ANN, GP, SVMR and GPR) are comparable to those of the groundwater variable-density flow and salt transport numerical model. However, the GPR standalone models had better predictive capabilities than the other standalone models. Also, the SVMR and GPR standalone models were more efficient (i.e. took less computational training time) than the other standalone models. In terms of ensemble models, the performance of the homogeneous GPR ensemble model was superior to that of the other homogeneous and heterogeneous ensemble models. Overall, the homogeneous GPR ensemble model was better both in terms of accuracy and efficiency. Therefore, it can be utilised as a reliable groundwater salinity prediction tool and as an approximate simulator in coupled simulation-optimization models needed for prescribing optimal groundwater management strategies.

7.2 Background

As evident from the results presented in Chapters 3, 4, 5 and 6, the advantages of using predictive models in saltwater intrusion management are significant. Predictive models constructed using input-output datasets obtained from a numerical simulation model are largely used to predict future groundwater salinity levels in aquifers and as approximate simulators in coupled simulation-optimization models. The latter has attracted considerable attention in water resource engineering, as coupled simulation-simulation models are often used to prescribe optimal management strategies for high-risk coastal aquifers (as discussed in Chapters 4, 5 and 6). In many cases, the predictions of standalone models can be combined into homogeneous or heterogeneous ensemble models. These ensemble models tend to have greater predictive accuracy and reliability. Comparing the predictive capabilities of the two ensemble modelling approaches using identical datasets is vital, as it will help in establishing a more robust predictive model. However, to the best of the authors' knowledge, no such comparisons have been reported.

With advancements in computer systems and measurement techniques, several data-driven modelling algorithms have been applied to develop reliable models for predicting groundwater salinity by emulating the responses of a complex numerical simulation model. This has been well explained in the previous chapters. However, these standalone models have sometimes been statistically combined to develop ensemble predictive models capable of providing more accurate and robust predictions. An ensemble model consists of several standalone models whose predictions are combined statistically to predict new instances (Petrakova et al. 2015). A homogeneous ensemble model is produced when all the standalone models in the ensemble are constructed using the same algorithm. On the other hand, in a heterogeneous ensemble model, all the standalone models in the ensemble are different; i.e., developed using different algorithms. Both of these ensemble modelling approaches have yielded accurate predictions in various hydrological investigations (Duan et al. 2007; Qu et al. 2017; Velázquez et al. 2011). While the ensemble modelling paradigm has successfully contributed to many hydrological studies, its use in saltwater intrusion modelling is very rare. Only recently, Sreekanth and Datta (2011) and Roy and Datta (2017) utilized ensemble GP and multivariate adaptive regression spline (MARS) models, respectively, to predict the responses of a saltwater intrusion numerical model with greater accuracy than that achieved by standalone models. These results establish that ensemble models can be more accurate than standalone models and are better suited to the prediction of salinity in coastal aquifers.

The idea of using ensemble models (both homogeneous and heterogeneous) for groundwater salinity prediction is very attractive but somewhat unexplored. In this paper, homogeneous and heterogeneous ensemble predictive models are developed using several data-driven techniques and utilized to predict salinity concentrations in a small-island coastal aquifer system. Homogeneous ensemble models are used to harness the capability of several standalone models that use a single algorithm. This study also aimed to investigate the performance of heterogeneous ensemble models, since the different standalone models can perform best locally, while merging their outputs can help achieve better prediction accuracy. The novelty of this work lies in comparing the performance of homogeneous and heterogeneous ensemble models based on the identical datasets used for the prediction of groundwater salinity in a real island coastal aquifer system. A first-ever comparative study of homogeneous and heterogeneous ensemble models is made to establish a better-performing ensemble model. It is important to compare the performance of homogeneous and heterogeneous ensemble predictive models developed using identical training datasets generated from a regional-scale variable-density flow and salt transport numerical simulation model. This comparison will allow the establishment of more accurate and reliable prediction tools. Hence, the results of this study can be used as a reference on potential tools for groundwater salinity prediction, and when utilizing it as approximate simulators in coupled simulation-optimization models.

7.3 Methods

7.3.1 Predictive modelling techniques

Of the regression-based predictive modelling tools available in the literature, we selected ANN, GP, GPR and SVMR algorithms for investigation in this study. These tools have been widely used in saltwater intrusion prediction studies. However, their performances have not been compared with each other. The GP and SVMR models are extensively discussed in Chapter 3. However, a basic account of the working principles of the ANN and GPR modelling tools is made in the sections below.

7.3.1.1 Artificial neural networks

The ANN is an artificial intelligence-based approach, which has been used in various domains of research worldwide. ANN has also been the most common modelling tool used in saltwater intrusion prediction studies. The main advantages of ANN models are that they provide an opportunity to retrieve hidden information, which enables decision-makers to solve complex problems, and can generalise and produce both linear and non-linear outputs (Petrakova et al. 2015). An ANN model behaves like a simplified model of brain cells that collaborate with each other to perform a desired function. ANN can be described as a parallel knowledge processing system containing a set of neurons arranged in layers (Angelaki et al. 2018). An ANN model comprises an input layer, hidden layer and target layer. The target layer is the main processing part of an ANN model. The flow of information takes place from layer to layer, serially. ANN models possess a number of connecting weights, which ultimately control the flow of information through the various layers. In the present work, ANN models were developed (trained and tested) using large input (pumping rate)-output (salinity) datasets. The resulting validated ANN models were used to predict salinity concentration at several monitoring wells in a coastal aquifer.

7.3.1.2 Gaussian process regression

Gaussian process regression is a full Bayesian non-parametric machine learning tool used in various field of studies worldwide (Crevillén-García et al. 2017; Grbić et al. 2013; Sun et al. 2014; Zhang et al. 2016). GPR has been applied to a variety of approximation, regression and classification problems (Karbasi 2017). GPR offers a powerful method of accurate function approximation in highly-dimensional space (Nguyen-Tuong et al. 2009). Detailed descriptions of the working principles of GPR models are widely available in the literature (Aye and Heyns 2017; Kong et al. 2018; Richardson et al. 2017; Zhang et al. 2016). When using GPR, we model a finite set of random variables $f = [f(a_1, \ldots, f(a_S)]^T$ as a joint Gaussian distribution with mean μ and covariance G, where a_1 is the i^{th} input. The GPR method directly defines a prior probability distribution over a latent function with assigned values of μ and G (Karbasi 2017). If the function f has a GP prior, it can be denoted as

$$f \sim GP(\mu, G) \qquad (7.1)$$

In many cases, a zero mean and a chosen kernel matrix are assigned as a covariance matrix. In GPR, the objective is to predict the responses b^* of a new input a^*, given a training set $\{(a_i, b_i)\}_{i=1}^{S}$ containing S training points, where a_i is the input variable and b_i is the corresponding response variable. The response variable b_i is modelled as a noise-version of the function value $f(x_i)$:

$$b_i \sim \mathcal{T}((f(a_i), \sigma^2) \qquad (7.2)$$

Where the distribution of noise is Gaussian $\mathcal{T}(0, \sigma^2)$ with variance σ^2. According to the above definition, the combined probability of the response variables and latent function variables $p(b, f) = p(b|f)p(b)$ can be calculated. Then, we can infer that the distribution of the latent function value f^* is a Gaussian distribution with mean $\mu(a^*)$ and variance $v(a^*)$:

$$\mu(a^*) = k_{a^*A}(\sigma^2 I + k_{AA})^{-1} b \qquad (7.3)$$

$$v(a^*) = k_{a^*a^*} - k_{a^*A}(\sigma^2 I + k_{AA})^{-1} k_{Aa^*} \qquad (7.4)$$

Where $k_{a^*A} = k(a^*, A)$ represents an *n*-dimensional row vector of the covariance between a^* and S training points, and $k_{AA} = k(A, A)$ specifies the kernel matrix of the S training points.

7.3.1.3 Homogeneous ensemble models

The standalone models used within the homogeneous ensemble model were developed using the same algorithms. In this case, homogeneous ensemble models of ANN, GP, SVMR and GPR algorithms were developed for comparison with the heterogeneous ensemble model. For easy identification, we name the developed homogeneous ensemble models using ANN, GP, SVMR and GPR as ANN_En model, GP_En model, SVMR_En and GPR_En model, respectively. Each homogeneous model comprised four standalone models trained and testing using different realizations of input-output datasets. These realizations of the datasets were generated using a random sampling without replacement procedure. Figure 7.1 (a) shows the procedure for developing a homogeneous ensemble model. The output of the homogeneous ensemble model ($E_{n,\,m}$) is obtained by combining the outputs of *N* standalone models using the simple average methodology expressed by Eq. 7.5:

$$E_{n,m} = \frac{1}{N}\sum_{n=1}^{N} P^{n} \qquad n = 1,2,\dots,N \tag{7.5}$$

7.3.1.4 Heterogeneous ensemble models

Each standalone model in the heterogeneous ensemble model was developed using a different algorithm; in this study, ANN, GP, GPR and SVMR standalone models were used. For easy identification, we will name this heterogeneous ensemble model the ANN_GP_SVMR_GPR_En model. The four different standalone models were trained and tested using different realizations of training and testing datasets, similar to the process used for the homogeneous ensemble models. The outputs of each standalone model were combined using Eq. 9. Figure 7.1 (b) shows the procedure for developing a heterogeneous ensemble model.

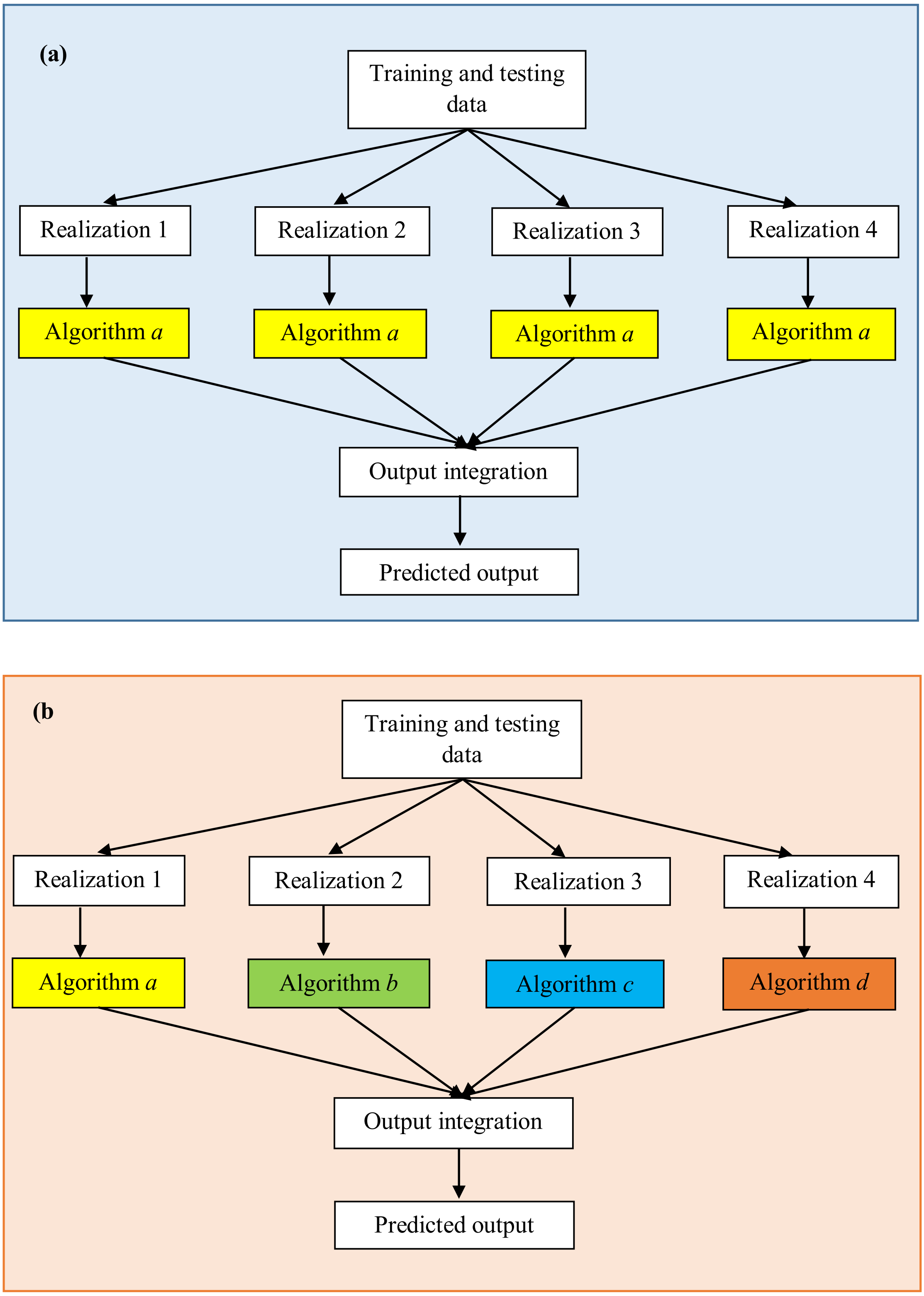


Figure 7.1: Flowcharts of the procedure for constructing homogeneous (a) and heterogeneous (b) ensemble models

7.3.2 Datasets and cross-validation

Transient groundwater pumping rates from all operating wells (inputs) and salinity concentrations (outputs) in the aquifer were required to train and test the standalone models. The input training pumping rate data from all operating pumping wells were obtained using the Latin hypercube sampling strategy. Each set of pumping patterns was fed into the numerical model as input and the corresponding salinities at different monitoring wells were obtained. This procedure was repeated *D* times to gather *D* input-output datasets. For cross-validation purposes, these input-output datasets were partitioned into construction and independent sets at the ratio of 0.9:0.1. The construction set was used to develop (train and test) the standalone models. The standalone models were constructed using the "hold-out strategy", in which 80 % of the construction datasets were used to train the models while the remaining 20 % were used to test the models. The independent set was used to determine the performances (validate) each ensemble model type. Once all the standalone models were trained and tested, the input data (pumping rates) of the independent set was used to establish the predictive capabilities of both the ensemble models. Finally, the predicted outputs from the ensemble models were compared with the output of the independent set.

7.3.3 Statistical performance evaluation criteria

Various statistical criteria were used to assess the performance of the developed standalone and ensemble models: root mean square error (RMSE), coefficient of determination (R^2), Nash-Sutcliffe coefficient (NSE) and Wilmott's index of agreement (WI). These evaluation criteria are defined by Eqs. 7.6 to 7.9.

$$RMSE = \sqrt{\frac{1}{d}\sum_{i=1}^{d}(t_i - p_i)^2} \tag{7.6}$$

$$R^2 = \frac{\sum_{i=1}^{d}(p_i - \bar{t})^2}{\sum_{i=1}^{d}(t_i - \bar{t})^2} \tag{7.7}$$

$$NSE = 1 - \frac{\sum_{i=1}^{d}(t_i - p_i)^2}{\sum_{i=1}^{d}(t_i - \bar{p})^2} \tag{7.8}$$

$$WI = \left|1 - \left(\frac{\sum_{i=1}^{d}(t_i - p_i)^2}{\sum_{i=1}^{d}|p_i - \bar{t}| + |t_i - \bar{t}|}\right)\right| \tag{7.9}$$

Here, d represents the total number of datasets, t_i is the true salinity from the numerical model, p_i represents the predictive model's salinity estimates, $\bar{t}$ is the mean true salinity from the numerical model and $\bar{p}$ denotes the mean predictive model's salinity estimates.

7.3.4 Evaluation of the developed method

For evaluation purposes, the developed methodology was applied to the Bonriki aquifer system located in Kiribati. Detailed descriptions of the study area and its hydrogeology and field data are provided in Chapter 4. The FEMWATER computer package was used to construct a 3D coupled flow and transport model of the Bonriki aquifer system. The modelling details are extensively presented in Chapter 4. Details on the calibration and validation of the numerical model and its results are also presented in Chapter 4.

After calibration and validation, the numerical model was simulated for the next four years to generate the input-output datasets required to develop the predictive models. Salinity datasets at six monitoring wells were collected by feeding 100 input pumping rates (yearly pumping rate from 25 pumping wells × 4 years) into the numerical model at once. A total of 700 datasets were generated by running the validated numerical model 700 times. Four different realizations of the datasets were generated using the random sampling without replacement technique (Friedman et al. 2001). These different realizations of the datasets were used to train, test and validate the standalone and ensemble models, as discussed in Section 7.3.2. Salinity was monitored at six monitoring wells. Therefore, six different standalone models were developed using each predictive modelling algorithm. Six ensemble models were also developed using both homogeneous and heterogeneous modelling techniques. These standalone and ensemble models were only responsible for predicting salinity concentrations at particular monitoring wells. In this study, all the predictive models were developed using the MATLAB 2017b platform. A feed-forward neural network with a back propagation training algorithm was used to develop the ANN-based predictive models. In developing the GP models, minimization of the sum of the squared errors was chosen as the fitness function. For the GPR models, a squared-exponential covariance function with automatic relevance determination was used as the covariance function. All the related hyper-parameters of this covariance function were set to the default MATLAB parameters. Lastly, in developing the SVMR-based models, a Gaussian kernel function was used with the values of parameters ε (insensitive tube), C (cost function) and γ (Gaussian kernel parameter) set to 0.0004, 10 and 0.05, respectively.

7.4 Results and discussion

7.4.1 Performance of the standalone models

The performance evaluation results for the ANN, GP, SVMR and GPR standalone groundwater salinity prediction models are presented in Tables 7.1, 7.2, 7.3 and 7.4, respectively.

Table 7.1: Performance evaluation results for the standalone ANN models

Monitoring well	Performance indicator	Model							
		ANN^1		ANN^2		ANN^3		ANN^4	
		Train	Test	Train	Test	Train	Test	Train	Test
MW_1	RMSE	6.33	6.27	6.12	6.53	6.68	6.9	6.15	6.62
	R^2	0.92	0.93	0.93	0.94	0.95	0.92	0.92	0.93
	WI	0.91	0.93	0.92	0.93	0.92	0.92	0.91	0.92
	NSE	0.97	0.98	0.99	0.98	0.98	0.97	0.98	0.98
MW_2	RMSE	6.15	6.09	6.28	6.37	6.61	5.99	6.24	6.15
	R^2	0.95	0.93	0.93	0.93	0.94	0.94	0.93	0.93
	WI	0.92	0.93	0.93	0.93	0.93	0.93	0.92	0.93
	NSE	0.99	0.98	0.98	0.98	0.99	0.98	0.99	0.98
MW_3	RMSE	5.33	6.12	5.88	6.46	6.37	5.94	5.68	5.88
	R^2	0.92	0.93	0.94	0.92	0.94	0.93	0.92	0.92
	WI	0.93	0.94	0.93	0.93	0.94	0.92	0.92	0.91
	NSE	0.98	0.99	0.99	0.99	0.98	0.98	0.99	0.99
MW_4	RMSE	6.33	6.21	6.18	6.03	6.17	6.28	5.99	5.87
	R^2	0.93	0.92	0.94	0.93	0.95	0.93	0.94	0.93
	WI	0.94	0.93	0.94	0.94	0.94	0.92	0.94	0.93
	NSE	0.99	0.99	0.98	0.99	0.99	0.98	0.99	0.99
MW_5	RMSE	4.85	4.67	4.92	4.13	4.88	4.41	4.06	4.18
	R^2	0.95	0.94	0.95	0.95	0.94	0.94	0.95	0.95
	WI	0.94	0.94	0.95	0.94	0.94	0.95	0.94	0.95
	NSE	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.99
MW_6	RMSE	4.99	4.87	4.67	4.54	4.83	4.99	4.01	4.88
	R^2	0.95	0.95	0.95	0.94	0.95	0.94	0.95	0.94
	WI	0.94	0.94	0.95	0.94	0.94	0.95	0.94	0.95
	NSE	0.99	0.99	0.99	0.99	0.99	0.99	0.99	0.98

ANN^k: denotes the ANN model developed using realisation k of the dataset, where k = 1, 2, 3, 4

Table 7.2: Performance evaluation results for the standalone GP models

Monitoring well	Performance indicator	Model							
		GP1		GP2		GP3		GP4	
		Train	Test	Train	Test	Train	Test	Train	Test
MW1	RMSE	5.21	5.68	5.98	5.72	5.63	5.84	5.39	5.67
	R^2	0.96	0.92	0.96	0.94	0.96	0.93	0.96	0.94
	WI	0.93	0.94	0.93	0.93	0.94	0.92	0.93	0.92
	NSE	0.99	0.99	0.99	0.99	0.98	0.98	0.98	0.99
MW2	RMSE	5.46	5.13	5.1	5.67	5.24	5.59	5.27	5.09
	R^2	0.96	0.94	0.97	0.95	0.95	0.95	0.96	0.94
	WI	0.93	0.93	0.94	0.93	0.93	0.93	0.94	0.93
	NSE	0.99	0.98	0.98	0.99	0.99	0.99	0.99	0.99
MW3	RMSE	4.65	5.84	5.13	5.9	5.68	5.53	5.08	5.37
	R^2	0.95	0.93	0.95	0.93	0.96	0.94	0.95	0.93
	WI	0.94	0.95	0.95	0.93	0.95	0.92	0.93	0.94
	NSE	0.98	0.99	0.99	0.99	0.98	0.98	0.99	0.99
MW4	RMSE	5.96	5.79	5.19	5.68	5.39	5.56	5.83	5.44
	R^2	0.95	0.92	0.95	0.95	0.96	0.96	0.96	0.95
	WI	0.95	0.95	0.96	0.94	0.96	0.93	0.93	0.94
	NSE	1	0.99	0.98	1	0.99	0.97	0.99	0.99
MW5	RMSE	4.65	4.38	4.18	5.76	4.33	4.61	4.28	4.91
	R^2	0.97	0.92	0.97	0.96	0.96	0.96	0.97	0.95
	WI	0.95	0.95	0.96	0.96	0.97	0.96	0.96	0.96
	NSE	0.99	0.99	1	0.99	1	0.99	1	1
MW6	RMSE	4.68	4.37	4.3	4.04	4.68	4.62	4.82	4.31
	R^2	0.96	0.96	0.97	0.95	0.96	0.95	0.96	0.95
	WI	0.96	0.97	0.96	0.96	0.96	0.95	0.96	0.96
	NSE	1	0.99	1	1	0.99	1	1	1

GP^k denotes GP model developed using realisation k of the dataset, where k = 1, 2, 3, 4

Table 7.3: Performance evaluation results for the standalone SVMR models.

Monitoring well	Performance indicator	Model							
		SVMR1		SVMR2		SVMR3		SVMR4	
		Train	Test	Train	Test	Train	Test	Train	Test
MW1	RMSE	4.33	5.44	5.18	5.66	5.07	5.12	5.18	4.93
	R^2	0.98	0.93	0.97	0.94	0.97	0.95	0.98	0.95
	WI	0.94	0.95	0.95	0.94	0.96	0.94	0.94	0.93
	NSE	1	1	1	1	1	1	1	0.99
MW2	RMSE	4.79	4.61	4.59	4.48	4.06	4.82	4.62	4.02
	R^2	0.98	0.94	0.98	0.95	0.98	0.94	0.97	0.93
	WI	0.95	0.94	0.95	0.94	0.96	0.93	0.96	0.95
	NSE	1	1	1	0.99	1	0.99	0.99	0.99
MW3	RMSE	5.54	5.13	4.66	4.08	4.79	4.37	4.08	3.82
	R^2	0.98	0.95	0.97	0.93	0.98	0.95	0.97	0.94
	WI	0.95	0.96	0.96	0.95	0.97	0.94	0.96	0.95
	NSE	1	0.99	1	0.99	1	1	0.99	0.99
MW4	RMSE	5.58	5.46	4.56	4.12	5.18	5.47	5.65	4.89
	R^2	0.96	0.94	0.97	0.95	0.98	0.96	0.98	0.96
	WI	0.97	0.96	0.97	0.96	0.98	0.95	0.98	0.95
	NSE	1	0.99	1	1	0.99	1	0.99	0.99
MW5	RMSE	3.62	3.33	3.10	3.47	3.68	3.34	3.18	3.92
	R^2	0.98	0.96	0.98	0.95	0.98	0.95	0.98	0.94
	WI	0.96	0.97	0.98	0.98	0.98	0.97	0.98	0.97
	NSE	1	0.99	1	0.99	1	0.99	1	1
MW6	RMSE	3.71	3.79	3.22	3.87	3.25	2.89	3.46	3.88
	R^2	0.98	0.97	0.98	0.97	0.98	0.97	0.98	0.98
	WI	0.97	0.98	0.98	0.97	0.98	0.97	0.98	0.98
	NSE	1	0.99	1	1	1	1	1	1

$SVMR^k$ denotes SVMR model developed using realisation k of the dataset, where k = 1, 2, 3,

Table 7.4: Performance evaluation results for the standalone GPR models.

Monitoring well	Performance indicator	Model							
		GPR^1		GPR^2		GPR^3		GPR^4	
		Train	Test	Train	Test	Train	Test	Train	Test
MW_1	RMSE	3.21	4.15	3.98	4.58	4.37	4.28	4.26	4.1
	R^2	0.98	0.96	0.98	0.95	0.98	0.96	0.98	0.97
	WI	0.96	0.96	0.97	0.96	0.97	0.96	0.97	0.96
	NSE	1	1	1	1	1	1	1	0.99
MW_2	RMSE	2.72	3.68	3.33	4.18	4.07	3.98	4.11	3.42
	R^2	0.98	0.95	0.98	0.96	0.98	0.96	0.98	0.96
	WI	0.96	0.96	0.96	0.96	0.97	0.95	0.97	0.96
	NSE	1	1	1	1	1	0.99	1	0.99
MW_3	RMSE	3.11	3.18	2.88	3.52	3.16	3.45	3.2	3.28
	R^2	0.98	0.96	0.98	0.98	0.98	0.95	0.98	0.97
	WI	0.97	0.97	0.97	0.96	0.98	0.96	0.97	0.96
	NSE	1	0.99	1	1	1	1	1	1
MW_4	RMSE	4.11	4.2	3.79	3.73	4.15	4.83	4.11	3.85
	R^2	0.98	0.95	0.98	0.97	0.98	0.96	0.98	0.97
	WI	0.98	0.97	0.98	0.97	0.98	0.97	0.98	0.96
	NSE	1	1	1	1	1	1	1	1
MW_5	RMSE	3.25	3.1	2.81	2.57	2.31	2.46	2.18	2.46
	R^2	0.98	0.94	0.98	0.97	0.98	0.96	0.99	0.97
	WI	0.98	0.98	0.99	0.98	0.98	0.99	0.98	0.98
	NSE	1	1	1	1	1	1	1	1
MW_6	RMSE	2.22	2.61	2.37	2.16	3.04	3.12	2.58	2.34
	R^2	0.99	0.98	0.99	0.98	0.99	0.98	0.99	0.98
	WI	0.98	0.99	0.98	0.98	0.98	0.98	0.98	0.98
	NSE	1	1	1	1	1	1	1	1

GPR^k denotes GPR model developed using realisation k of the dataset, where k = 1, 2, 3, 4

7.4.2 Performance evaluation of the ensemble predictive models

A comparison of the performance evaluation of the homogeneous and heterogeneous ensemble models is presented in Table 7.5.

Table 7.5: Performance evaluation results for the homogeneous and heterogeneous ensemble models

Monit-oring well	Performance indicator	Ensemble model type				
		Homogeneous ensemble models				Heterogeneous ensemble model
		ANN_En	GP_En	SVMR_En	GPR_En	ANN/GP/SVMR/GPR_En
MW_1	RMSE	6.23	5.46	4.62	3.99	4.56
	R^2	0.92	0.96	0.96	0.98	0.97
	WI	0.92	0.94	0.95	0.97	0.96
	NSE	0.99	0.99	0.99	1	0.99
MW_2	RMSE	5.82	5.07	3.99	3.16	3.25
	R^2	0.94	0.96	0.97	0.98	0.97
	WI	0.94	0.94	0.96	0.97	0.96
	NSE	0.99	0.99	0.99	1	0.99
MW_3	RMSE	5.67	5.19	3.82	3.07	3.56
	R^2	0.94	0.95	0.96	0.97	0.97
	WI	0.95	0.96	0.97	0.98	0.96
	NE	0.99	0.99	0.99	1	1
MW_4	RMSE	5.48	5.23	4.11	2.98	3.13
	R^2	0.94	0.96	0.96	0.98	0.97
	WI	0.95	0.96	0.97	0.98	0.96
	NSE	0.99	1	0.99	1	1
MW_5	RMSE	4.66	4.28	3.17	2.19	2.79
	R^2	0.96	0.96	0.97	0.98	0.97
	WI	0.95	0.97	0.98	0.99	0.97
	NSE	1	1	1	1	1
MW_6	RMSE	4.42	4.11	2.88	2.32	2.95
	R^2	0.95	0.97	0.98	0.99	0.97
	WI	0.96	0.97	0.98	0.99	0.97
	NSE	1	1	1	1	1

The predictive models used in this study are common modelling tools that have been widely used in groundwater salinity prediction. All standalone models performed reasonably well, as demonstrated by the evaluation results presented in Tables 7.1, 7.2, 7.3 and 7.4. For a particular monitoring well, four standalone models (each using a different realization of the dataset) were developed for predicting salinity concentration at that specified location. When developing the standalone ANN predictive models (as per Table 7.1), the calculated RMSE values ranged from approximately 4.06 mg/L to 6.90 mg/L. The calculated R^2 values ranged from 0.92 to 0.95.

Similarly, the WI and NSE values ranged from 0.91 to 0.95 and 0.97 to 0.99, respectively. For the standalone GP models, the RMSE values ranged from 4.04 mg/L to 5.98 mg/L. The minimum and maximum R^2 values obtained when training and testing the standalone GP models were 0.92 and 0.97, respectively. Likewise, the values of WI and NSE ranged from 0.92 to 0.97 and 0.98 to 1, respectively. For the standalone SVMR models, the RMSE values ranged from 3.10 mg/L to 5.66 mg/L. Similarly, the R^2 and WI values ranged from 0.93 to 0.98 and 0.93 to 0.98, respectively. The calculated NSE values for all the standalone SVMR models were either 0.99 or 1. Lastly, for the standalone GPR models, the RMSE values ranged from 2.12 mg/L to 4.58 mg/L. Also, the values for R^2 and WI ranged from 0.95 to 0.99 and 0.96 to 0.99, respectively. The NSE values obtained for all standalone GPR models were either 0.99 or 1. Overall, the performance evaluation results presented in Tables 7.1, 7.2, 7.3 and 7.4 establish that all the standalone predictive models based on the four algorithms approximated the variable density flow and salt transport numerical model with reasonable accuracy. However, the standalone GPR models were found to be superior to the corresponding ANN, GP and SVMR standalone models. The standalone ANN models were the least effective and had lower prediction accuracy.

The developed homogeneous ANN_En, GP_En, SVMR_En, GPR_En and heterogeneous ANN_GP_SVMR_GPR_En models demonstrated reliable salinity prediction capabilities, as demonstrated by the performance evaluation results (Table 7.5). Of the four homogeneous ensemble models, GPR_En was found to be the most reliable and to have better prediction accuracy. The GPR_En model had the lowest RMSE and highest R^2, WI and NE values of the homogeneous ensemble models. In addition, the GPR_En model was found to perform better and be more accurate than the heterogeneous ANN_GP_SVMR_GPR_En model in terms of the four performance indicators. Overall, as per the findings of this study, the homogeneous GPR_En models were found to be better suited to salinity prediction at the six monitoring wells.

Along with accuracy, efficiency in computational time is another major factor influencing decision-makers' choice of which model to use for groundwater salinity prediction. The training and testing time for each model type is different, and the decision-maker needs to prioritize accuracy or efficiency and achieve a balance between the two. In this study, all four standalone model types required different training and testing times. The ANN was the least accurate model in terms of the four evaluation criteria and it also took a reasonable amount of time (~10 minutes) to train. Similarly, the standalone GP models took approximately 15 minutes to train.

On the other hand, the SVMR and GPR models took much less time to train (~0.03 minutes). The time required to train a standalone model is dependent on the size of the training dataset; a larger dataset requires significantly more training time. In addition, developing ensemble models requires extra computational effort. For example, the standalone models can be integrated into an ensemble using various techniques. In this study, a simple average methodology was applied to construct homogeneous and heterogeneous ensemble models.

Using other techniques, such as majority voting and weighted voting, can yield more accurate results but may require more computational effort and development time. All these factors need to be evaluated before deciding on a modelling algorithm to be used for a particular prediction activity. Out of all the algorithms used in this study, the training of standalone GPR models required the least computational time and they also had the best prediction accuracy. The accuracy and efficiency of the standalone GPR models were reflected in the accuracy and efficiency the of GPR_En models. Hence, GPR_En models were established as the most preferred model (both in terms of accuracy and efficiency) for predicting salinity at monitoring wells in the Bonriki aquifer. On the other hand, ANN_En models were the least accurate and, hence, the least preferred. A ranking of the ensemble models from most to least preferred is provided in Figure 7.2.

Finally, the main purpose of developing the ensemble models was to predict salinity conditions in the Bonriki aquifer with reasonable accuracy and efficiency. The performance of the ensemble models was entirely dependent on the variable density flow and salt transport numerical simulation models used to develop the predictive models (i.e., on the training/testing datasets). Determining the precision of a numerical simulation model in approximating actual groundwater salinity conditions is essential. The calibration and validation of numerical simulation models remain important issues in real-life applications. Despite having acceptable calibration and validation results, the variable density flow and salt transport numerical model may contain uncertainties resulting from inadequate calibration/validation datasets, uncertain aquifer parameters and erroneous boundary conditions. These aspects of groundwater modelling need to be thoroughly explored and correctly implemented into numerical simulation models.

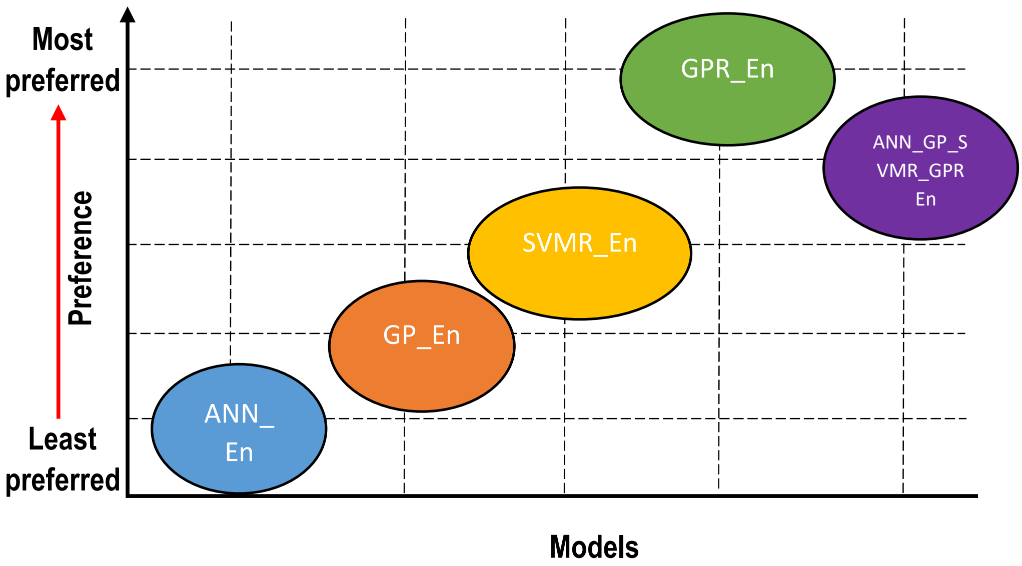


Figure 7.2: Preference ranking of the ensemble models.

7.5 Conclusions

This study compared the groundwater salinity prediction performance of standalone models and homogeneous and heterogeneous ensemble models. Specifically, ANN, GP, SVMR and GP standalone models were developed to construct homogeneous ensemble models (ANN_En, GP_En, SVMR_En and GPR_En) and a heterogeneous ensemble (ANN_GP_SVMR_GPR_En) model capable of predicting salinity concentration in the Bonriki aquifer. A summary of the major contributions of this study is as follows.

1. All the tested standalone models predicted the salinity concentration at respective monitoring wells with reasonable accuracy. However, in terms of the four performance indicators, standalone GPR models displayed better predictive accuracy than the corresponding ANN, GP and SVMR standalone models.
2. The standalone SVMR and GPR models required significantly less time to train compared to the ANN and GP standalone models.
3. The homogeneous GPR_En model was the best-performing model, even compared to the heterogeneous ANN_GP_SVMR_GPR_En model.
4. Overall, the GPR_En model performed the best of all the models evaluated. Hence, it is a potentially powerful tool for predicting salinity levels in the Bonriki aquifer. In addition, with its accurate and efficient prediction capabilities, the GPR_En model can be employed as an approximate simulator in the simulation-optimization models needed to develop regional-scale saltwater intrusion management strategies for coastal aquifers.

Comparisons of homogeneous and ensemble models of groundwater salinity have not been conducted in previous studies. The present comparison is the major contribution of this study. Also, the results presented in this study help in the establishment of better-performing ensemble model. This provides a valuable reference for decision-makers and engineers who may use these methods to predict groundwater salinity in coastal aquifers. Although the results presented are promising, the prediction capabilities of all the standalone models can be increased further. This can be achieved by using an optimal number of training and testing datasets. Also, recent developments and advancements in computational power will aid in developing hybrid models with increased predictive power. In addition, more comparative studies using other new predictive modelling algorithms are recommended in line with the objectives of the current work. It is hoped that future research will focus on these areas, which may eventually lead to the establishment of a more robust, accurate, efficient and versatile salinity prediction tool. In the next chapter, group method of data handling models are introduced into the field of saltwater intrusion modelling.

Chapter 8: Application of Group Method of Data Handling and variable importance analysis for prediction and modelling of saltwater intrusion processes in coastal aquifers

A manuscript based on the main contents of this chapter has been submitted to a journal and is currently under review, as outlined below:

Lal, A., and Datta, B. (2019). "Application of the Group Method of Data Handling and variable importance analysis for prediction and modelling of saltwater intrusion processes in coastal aquifers." *Neural Computing and Applications.*

8.1 Summary

Data-driven mathematical models are powerful predictive tools that can approximate the responses of saltwater intrusion numerical simulation models. Employing data-driven predictive models instead of complex groundwater flow and transport models enables the prediction of future scenarios. Most importantly, they help save computational time, effort and requirements when developing optimal coastal aquifer management strategies using complex and large-scale coupled simulation-optimization models. In this study, a new data-driven model, namely, the group method of data handling (GMDH) model, is developed and utilized to predict salinity concentrations in a coastal aquifer by mimicking the responses of a variable-density flow and solute transport numerical simulation model. In addition, an important characteristic of GMDH models is explored and evaluated; i.e., the ability to identify a set of input variables (pumping rates) that are most influential to the outcomes (salinity concentration at monitoring locations). To confirm variable importance, three tests are conducted in which new GMDH models are constructed using subsets of the original datasets. In TEST 1, new GMDH models are constructed using a set of the most influential variables (consisting of pumping rates at selected locations). In TEST 2, a subset of 20 variables (10 of the most and least influential variables) is used to develop new GMDH models. In TEST 3, a subset of the least influential variables is used to develop GMDH models. Performance evaluation demonstrates that the GMDH models developed using the entire dataset had reasonable prediction accuracy and efficiency. Comparative performance evaluation of the three test scenarios highlights the importance of the appropriate selection of relevant input pumping rates when developing accurate predictive models. The results suggest that incorporating the least influential variables decreases the accuracy of predictive models; thus, by considering the most influential pumping rates, it is possible to develop more accurate and efficient salinity prediction models. Overall, the evaluation results of this

study establish that GMDH models and the inherent input variable ranking capability can be utilized to construct accurate and efficient coastal saltwater intrusion prediction models. Hence, GMDH models are viable saltwater intrusion modelling tools, which can be employed in future regional-scale saltwater intrusion prediction and management investigations.

8.2. Background

Thus far, many data-driven mathematical models have delivered reasonably accurate and reliable saltwater intrusion prediction results. However, investigators are still challenged by two complications. First, not all predictive models explicitly describe the influence of individual input variables on the output. It is important for investigators to employ a predictive model that allows them to assess the impact of each input variable on the outcome. This is a significant component of predictive modelling that needs further attention. Second, the development of data-driven predictive models requires investigators to be experts in setting the optimal model parameters. In reality, investigators do not always have such in-depth knowledge in handling model parameters. Determining optimal model structures is always considered a burden; however, it is very important as it can affect the accuracy and efficiency of the model. These two critical issues can be resolved by using the GMDH model as a complex saltwater intrusion process prediction tool.

In the hydrogeological literature, GMDH models have not been used for the approximation of complex saltwater intrusion processes in coastal aquifer systems despite possessing significant advantages over other approaches. However, GMDH-based predictive models have been employed in various engineering and science applications due to their four advantages. First, GMDH is a self-organizing and inductive evolutionary algorithm, which has the potential to deal with multi-dimensional and complex variables (Liu et al. 2018). Second, GMDH models have an automatic modelling mechanism, which allows self-structuring of the models with optimal parameters without any user input (Xiao et al. 2017). Third, GMDH models have a strong anti-interference capability, which allows them to handle noisy datasets (Xiao et al. 2017). Lastly, GMDH models can display explicit expressions that relate input variables to the output. Lastly, GMDH models have the capability to select and present subsets of the most influential input predictor variables (Teng et al. 2017). This characteristic allows the user to understand the effect of each input predictor variable on the output.

In pumping-induced saltwater intrusion modelling investigations, determining the pumping rates that have the greatest influence on salinity at monitoring locations is of great significance. First, selection of the most influential variables for the prediction task can provide higher prediction accuracy (Mohammadi et al. 2016). Second, the inclusion of the least influential variables in predictive models will lead to unnecessary model complexity (Braun and Oswald 2011). Also, the need for a higher number of input datasets can significantly increase the cost of data monitoring and collection. To the best of our knowledge, there are few studies that have determined the pumping rates most influential to the salinity at monitoring locations in coastal aquifer systems. The important contribution of this study is its utility of GMDH to predict the pumping rates most influential to saltwater intrusion. The evaluations conducted in this study establish the influence of pumping rates on saltwater intrusion, as well as the importance of properly selecting input pumping rates when developing accurate and efficient groundwater salinity prediction models.

In the present work, GMDH models were trained and tested to predict groundwater salinity concentration in a coastal aquifer, and to determine the pumping rates most influential to groundwater salinity. Sets of the most- and least-influential pumping rates, and a combination of them, were identified and used to construct new GMDH-based predictive models. This process illustrates the importance of identifying influential variables when developing predictive models, and also assesses how the inclusion of less-influential variables degrades their performance. The outcomes of this study establish GMDH models as accurate and efficient saltwater intrusion modelling tools that can be applied to solve future regional-scale saltwater intrusion modelling problems. Most specifically, they can also be used in coupled simulation-optimization models needed for the sustainable and optimal management of coastal aquifer systems (Sreekanth and Datta 2010). Based on this study, it is anticipated that, in future, GMDH models may be considered as potential candidates for evaluating input variable importance, and for developing reasonably efficient and accurate predictive tools for various groundwater management-related applications.

8.3. Methods

8.3.1 Coastal groundwater mathematical simulation code

The mixing of freshwater and saltwater due to their density difference makes the saltwater intrusion process highly non-linear. For the accurate simulation of saltwater intrusion dynamics, flow and transport equations need to be solved simultaneously. In this study, the finite element code FEMWATER (Lin et al. 1997) was utilized to numerically simulate three-dimensional variable-density flow and solute transport processes in the study area. The governing equations used to solve coupled groundwater flow and mass transport problems in porous media can be found in Lin et al. (1997). Depending on the specified values of hydrological properties such as hydraulic conductivities, recharge, initial conditions and boundary conditions, the governing flow and transport equations can be solved. The FEMWATER code has been successfully applied in various regional-scale saltwater intrusion modelling investigations, such as Datta et al. (2009), Kim et al. (2012), Insigne and Kim (2010) and Lal and Datta (2019). A complete description of the FEMWATER model is given in Chapter 3.

8.3.2 Development of GMDH predictive models

8.3.2.1 The GMDH algorithm

The GMDH algorithm was first suggested by the former Soviet scientist Ivakhnenko. It is a method for identifying non-linear relationships between sets of input and output data (Fernández and Lozano 2010). In principle, the GMDH model functions by generating a high-order polynomial network, which is principally a feed-forward and multilayer neural network. The GMDH algorithm provides a self-organizing data mining platform, which automatically decides the variables to be used in the modelling framework, and the structure (neurons in hidden layers) and parameters of the model. The model itself provides an optimal structure, thus reducing the need for prior knowledge and assumptions. This feature of the GMDH algorithm reduces the potential for user bias and also minimises the complexity of the model (Xiao et al. 2017). Construction of a GMDH model requires division of the input dataset into two groups. The first group is used to approximate the parameters of each neuron to obtain a partial description of the process, while the second group is used to weigh the performance of the candidate models that describe the process most efficiently (Fernández and Lozano 2010). Specifically, the training dataset is used for the approximation of the coefficients of the Kolmogorov-Gabor polynomial, while the testing set is used in the GMDH network for error evaluation. GMDH works by constructing successive layers with connections that are the individual terms of a polynomial (Srinivasan 2008). The output of each neuron is assessed and evaluated by an external criterion. The model eliminates the neurons that provide the poorest predictions and preserves the neurons with excellent performance for use in the next layer. These steps are repeated to create new layers until the error criterion stops decreasing. The whole process of training and assortment is repeated on this new layer. Once neurons that best satisfy the pre-specified criterion are chosen, the model is verified using the testing dataset. More comprehensive descriptions of the GMDH modelling algorithm are available in the literature (Farlow 1984; Liu et al. 2018; Srinivasan 2008).

8.3.2.2 Datasets, model development and cross-validation

Development of the GMDH models required several input-output datasets. The pumping patterns from all operational pumping wells made up the input dataset. The outputs were the resulting salinity concentration values obtained after implementing each set of the input pumping patterns into the numerical simulation model. Each time a new pumping pattern was implemented into the numerical simulation model, a new output dataset was recorded (salinities at the respective monitoring wells). Thus, collecting N input-output datasets required running the variable-density flow and solute transport numerical model N times. The collected input-output datasets were used to develop and assess the performance of the GMDH predictive models. The N datasets were separated into a development set and an independent set. The development set was used to train and test the predictive models. In principle, the GMDH algorithm only approximates the relationship between multiple input variables and a single output variable. Hence, for m monitoring wells, m GMDH models need to be developed. Each model is only able to predict salinity at each monitoring location. During model development, the collected input-output datasets were divided into training and testing datasets. In this study, the "holdout" cross-validation procedure (Kohavi 1995) was used, in which a portion of the dataset is randomly held and not used in the training process. This holdout dataset was used to examine the performance of the trained models. Lastly, the independent set was used to assess the performance of the predictive models.

8.3.2.3 Performance assessment indices

The accuracy of the GMDH models was assessed using the statistical indices root mean square error (RMSE), mean absolute error (MAE), Nash-Sutcliffe efficiency (NSE), coefficient of determination (R^2) and correlation coefficient (r). These are standard accuracy evaluation measures that have been used in various other predictive modelling studies in the field of hydrology. Their mathematical formulations are in Eqs. 8.1 to 8.5.

$$RMSE = \sqrt{\frac{1}{d}\sum_{i=1}^{d}(C_i^p - C_i^o)^2} \tag{8.1}$$

$$MAE = \frac{1}{d}\sum_{i=1}^{d}\left|C_i^p - C_i^o\right| \tag{8.2}$$

$$NSE = 1 - \frac{\sum_{i=1}^{d}(C_i^o - C_i^p)^2}{\sum_{i=1}^{d}(C_i^o - \overline{C^o})^2} \tag{8.3}$$

$$R^2 = \left(\frac{\sum_{i=1}^{d}(C_i^p - \overline{C^p})(C_i^o - \overline{C^o})}{\sqrt{\sum_{i=1}^{d}(C_i^p - \overline{C^p})^2 \sum_{i=1}^{d}(C_i^o - \overline{C^o})^2}}\right)^2 \tag{8.4}$$

$$r = \frac{\sum_{i=1}^{d}(C_i^p - \overline{C^p})(C_i^o - \overline{C^o})}{\sqrt{\sum_{i=1}^{d}(C_i^p - \overline{C^p})^2 \sum_{i=1}^{d}(C_i^o - \overline{C^o})^2}} \tag{8.5}$$

Here, C_i^o is observed salinity concentration, C_i^p is predicted salinity concentration, $\overline{C^o}$ is the mean observed salinity concentration, $\overline{C^p}$ is the mean predicted salinity concentration, and d represents total number of data points.

8.3.3 Data dimensionality reduction

The originally developed (with input variables) GMDH models provided a subset of the most influential pumping rates (input variables), which made the greatest contribution to the resulting salinities at the monitoring locations. To establish the importance of identifying and utilizing influential pumping rates in the construction of accurate GMDH models, new GMDH models were developed using only a subset of the original dataset. In this study, three subsets were used to construct new GMDH models, as follows:

TEST 1: New GMDH models were constructed using the most influential variables only.

TEST 2: New GMDH models were constructed using a mixture of the ten most- and ten least-influential variables.

TEST 3: New GMDH models were constructed using only the least influential variables.

The performance of the GMDH models constructed for three scenarios consisting of three tests were evaluated and compared using the five statistical criteria (Eqs. 8.1 to 8.5).

8.3.4 Application

For performance evaluation purposes, GMDH models were used to predict salinity concentration in an illustrative coastal aquifer system. A portion of a coastal aquifer (2.5 km^2 in area) was simulated using the FEMWATER computer code. The aquifer was 60 m in depth and was divided into three equal layers. The seaside boundary (Boundary A) was about 2.1 km in length. Similarly, Boundaries B and C were 2.0 and 2.7 km in length, respectively. The aquifer comprised eight groundwater abstraction wells (A_1, A_2, A_3, A_4, A_5, A_6, A_7 and A_8), which were utilized to extract groundwater for beneficial local use. Additionally, a set of five barrier wells (B_1, B_2, B_3, B_4 and B_5) were installed in the aquifer. These barrier wells were installed closer to the seaside boundary and were used to extract saline water. The extraction of saline water creates a hydraulic pressure towards the sea, which helps control saltwater intrusion into the aquifer. Salinity was monitored at three monitoring wells (M_1, M_2 and M_3). The specific well locations and the study area model domain are illustrated in Fig. 8.1. The study area conceptual model was discretised into thin, triangular, finite elements containing a total of 4660 nodes. Element size was selected after conducting several numerical trials. Boundary A was the seaside boundary, which was constant head; i.e., 0 m at both ends, and constant concentration (35 kg/m^3) boundary. Boundaries B and C were treated as no-flow boundaries. Groundwater recharge was considered constant over the simulation period. A recharge rate of 0.00054 m/d was distributed uniformly over the entire model domain. The aquifer was considered homogeneous (all 3 layers had the same hydraulic properties) but anisotropic with different hydraulic conductivities in the *x*-, *y*- and *z*-directions. The hydraulic conductivity of the aquifer was set to 15 m/d, 7.5 m/d and 1.5 m/d in the *x*-, *y*- and *z*-directions, respectively. The porosity of the aquifer materials was set to 0.4. The bulk density of the aquifer materials was taken as 1600 kg/m^3. The longitudinal and lateral dispersivity were taken as 50 m and 25 m, respectively. The molecular dispersion coefficient was set to 0.69 m^2/d. The density reference ratio and compressibility of the aquifer were taken as 0.69 and 8.5×10^{-15} md^2/kg, respectively. The resulting 3D finite element mesh was used to generate the initial conditions for the simulation model. The initial head and concentration conditions were obtained by executing the simulation model for a period of 20 years. Only the abstraction wells were used during this simulation. After 20 years, the head and concentration values had very minimal deviations compared to the previous year's values. Using this as the initial conditions, the simulation was run for a period of four years in a transient state (different annual pumping rates) using all the operational wells (8 abstraction and 5 barrier wells).

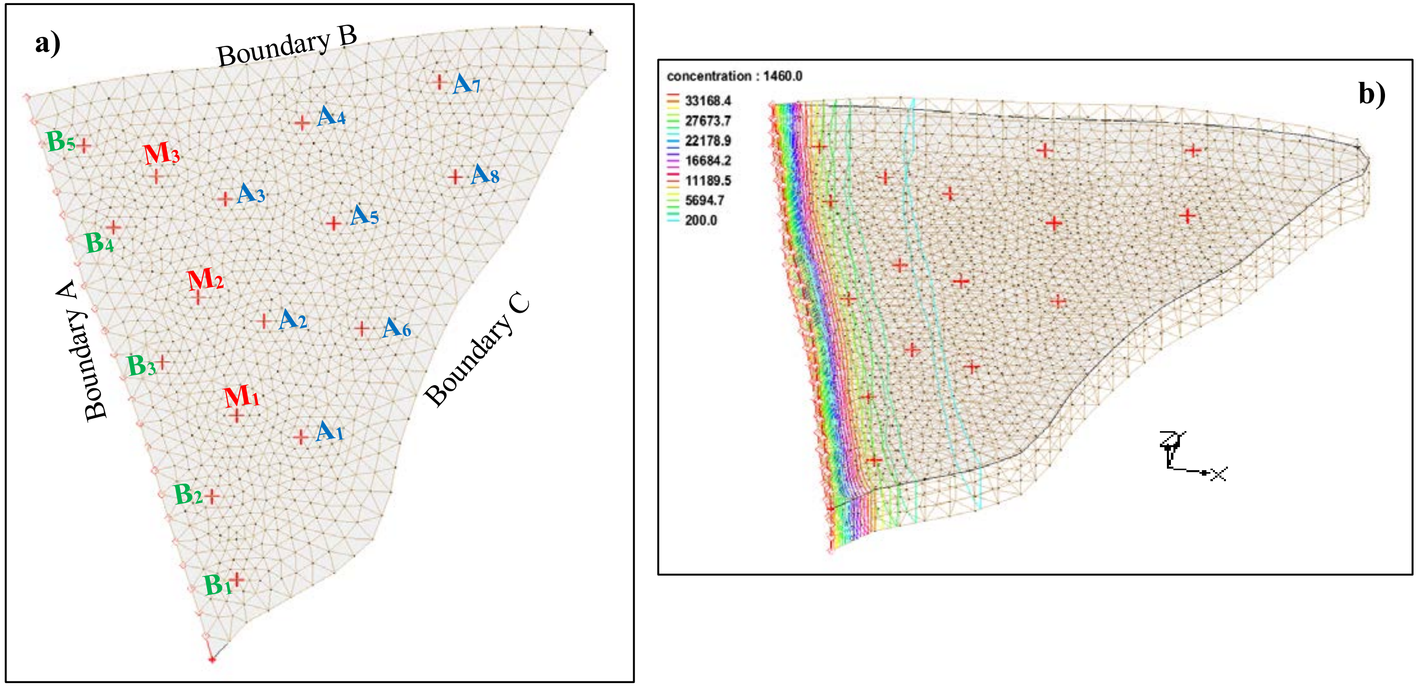


Figure 8.1: a) 2D triangular mesh representation of the study area with locations of barrier wells (B1 - B5), abstraction wells (A1 - A8) and monitoring wells (M1 - M3) and b) 3D view of the study area model domain.

The total of 13 pumping wells with different yearly pumping values resulted in 52 decision variables (13 wells × 4-year simulation period). These 52 variables (labelled *x1*-*x52*) represented yearly the annual pumping rates from each well over the four-year simulation period. A total of 800 sets of randomized input pumping patterns were produced using Latin hypercube sampling. Each time a pumping pattern was initiated in the numerical model, salinity outputs at the three monitoring wells were recorded. The numerical model took approximately four minutes to converge. The 800 input-output datasets were obtained by successfully executing the simulation model 800 times. Out of the 800 datasets, the development set consisted of 700 datasets and the independent set contained 100 datasets. GMDH shell software was used to construct the predictive models. Of the 700 development datasets, 560 were used for model training, while 140 were used for testing. Dividing the data into training and testing sets is user-dependent and can be easily implemented in the GMDH shell software. For model development, RMSE was used as the external criterion. The GMDH shell simultaneously provided three predictive models capable of estimating salinity at the three monitoring wells. For easy identification, the GMDH models used for predicting salinity at monitoring wells 1, 2 and 3 were labelled $GMDH_1$, $GMDH_2$ and $GMDH_3$, respectively. Once the models were developed, the independent set was imported into the shell and used to measure the true prediction competencies of the models.

8.4 Results and discussion

8.4.1 Accuracy of the predictive models

The performance assessment of the GMDH predictive models is summarized in Table 8.1. In addition, the comparison between observed and predicted salinities made during the testing phase is presented in Fig. 8.2. According to the performance evaluation results obtained for the training and testing phases, the three GMDH models predicted salinity concentration with reasonable accuracy. This was verified using the independent dataset. The performance of the developed models in predicting salinity during the training and testing stages are summarized in Table 8.1. A similar performance result, in terms of the five evaluation criteria, was observed for the independent set as well.

Table 8.1: GMDH model performance evaluation results

Model	Stage	RMSE	MAE	NSE	R^2 (%)	r
$GMDH_1$	Training	0.366	0.292	1	99.68	0.997
	Testing	0.358	0.289	1	99.72	0.997
	Prediction	0.379	0.299	1	99.61	0.996
$GMDH_2$	Training	0.221	0.175	1	99.67	0.998
	Testing	0.218	0.174	1	99.66	0.998
	Prediction	0.243	0.186	1	99.73	0.997
$GMDH_3$	Training	0.440	0.298	1	99.68	0.998
	Testing	0.368	0.293	1	99.69	0.998
	Prediction	0.458	0.306	1	99.66	0.996

For a similar aquifer system, Lal and Datta (2018) used genetic programming and support vector machine regression models to predict salinity concentration. It was observed that GMDH models delivered comparable or even better results than genetic programming and support vector machine regression models. This establishes the fact that GMDH models can accurately approximate the relationship between input transient pumping patterns and corresponding output salinity concentrations. The results also show that GMDH models provide a better approximation of the saltwater intrusion process than genetic programming and support vector machine regression models. Hence, in future, GMDH models can be used to predict salinity concentrations at monitoring wells in response to transient groundwater pumping from operational abstraction and barrier wells.

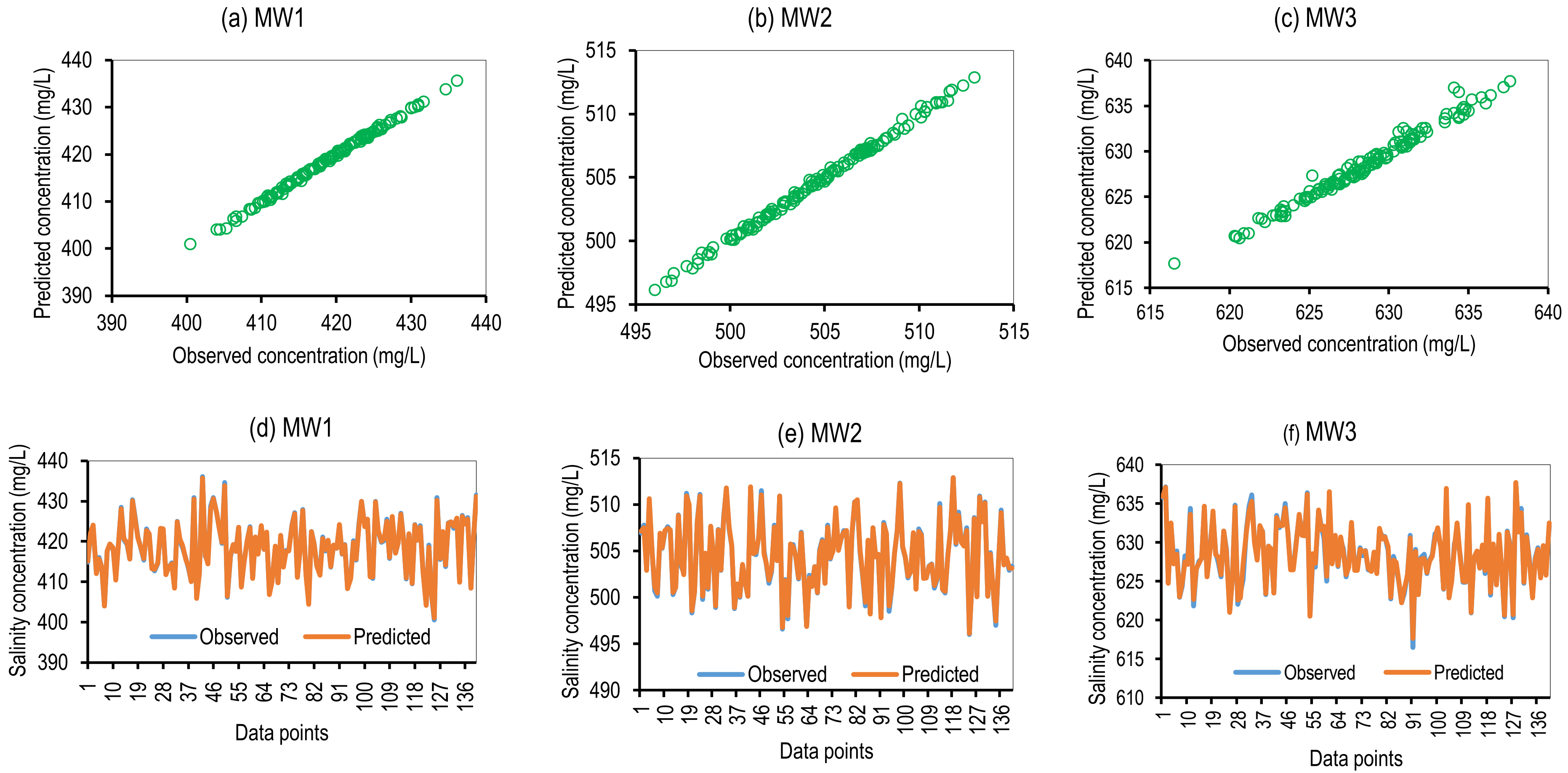


Figure 8.2: Comparisons of observed and predicted salinity concentration at the three monitoring wells using scatter plots (a, b and c) and line graphs (d, e and f)

8.4.2 Efficiency of the predictive models

Efficiency is another major factor that needs consideration before using predictive models for saltwater intrusion approximation. Salinity prediction using data-driven modelling techniques are not always efficient. The time taken for model development and prediction using independent datasets is presented in Table 8.2. It was observed that the development of all three GMDH models only needed a computation time of 0.45 minutes. This is significantly less than that of other predictive models, such as genetic programming-based models. In a similar aquifer system, development of a genetic programming predictive model required approximately 45 minutes of CPU runtime (Lal and Datta 2018). On the other hand, compared to support vector machine regression models, the development of GMDH models takes slightly more time. For a similar coastal aquifer problem, developing support vector machine regression models took only a few seconds (Lal and Datta 2018). These results establish that GMDH models are reasonably efficient, because they were developed in significantly less time than genetic programming-based models and in a similar timeframe to support vector machine regression models.

Once the GMDH models were developed, a new set of independent datasets was used to verify the prediction capability of the models. Prediction using the independent dataset comprising of 100 sets of pumping patterns took almost 0.40 mins of runtime. This is significantly less than the time required for the original variable density and solute transport numerical model. For each set of pumping patterns, the numerical model took about five minutes to converge to a solution. Running 100 sets of pumping patterns would take approximately 500 minutes. Therefore, it can be established that the ability to predict salinity for 100 input datasets at once within 0.40 minutes makes the GMDH model highly efficient.

Table 8.2: GMDH efficiency results

Stage	Time (minutes)
Model development (training and testing)	0.45
Prediction using an independent set	0.40

8.4.3 Variable importance ranking

One of the key features of the GMDH modelling algorithm is that it automatically recognizes and selects the dominant variables for a prediction model. In this study, the GMDH model only selected the pumping rates that had the greatest influence on the output salinity concentration. Out of the 52 input pumping rates (ranked variables *x1-x52*), those most influential to resulting salinity concentrations are presented in Fig. 8.3

It is observed that only 15 variables were most influential and, hence, were used for the development of model $GMDH_1$. Variables *x9*, *x12* and *x22* were found to be the most important and had the greatest impact on the salinity output. Likewise, 20 variables were used in developing model $GMDH_2$. Variables *x10*, *x23* and *x7* had the greatest impact on the output salinity. Lastly, only the 14 most important variable was used for developing model $GMDH_3$. Variables *x10*, *x8* and *x11* were the most influential. The variable ranking feature of the GMDH algorithm provides two major benefits to the decision-maker. First, it allows them to focus only on the most influential variables for the efficient and accurate prediction of saltwater intrusion processes. The most influential pumping rates can also be modified to control the concentration at respective monitoring wells. Also, the least influential pumping rates can be modified (possibly increased) to maximize the total feasible beneficial pumping in the study area, since they do not have much impact on the salinity at specified or critical locations. Second, fewer input variables means less computational burden in terms of model training and prediction time, as demonstrated in Lal and Datta (2018). Also, the cost in terms of data monitoring, collection and preparation will be significantly lower when fewer variables are used in the development of prediction models.

a)

#	If replaced with mean	Impact on RMSE	Bars	RMSE
1	x9	50.47%		3.42911
2	x12	43.27%		2.99222
3	x22	39.45%		2.76024
4	x25	33.49%		2.39884
5	x35	24.67%		1.86327
6	x38	17.34%		1.41899
7	x6	14.79%		1.26385
8	x19	12.08%		1.09927
9	x32	5.67%		0.71055
10	x48	2.83%		0.538432
11	x3	1.71%		0.470043
12	x51	1.15%		0.436332
13	x16	0.47%		0.394962
14	x29	0.08%		0.371036
15	x4	0.06%		0.370177

b)

#	If replaced with mean	Impact on RMSE	Bars	RMSE
1	x10	66.04%		2.64668
2	x23	47.67%		1.97198
3	x7	24.29%		1.11341
4	x36	23.10%		1.06974
5	x20	15.53%		0.791669
6	x9	14.71%		0.761751
7	x22	10.87%		0.620543
8	x35	4.90%		0.401266
9	x4	4.50%		0.386733
10	x33	2.30%		0.306043
11	x46	2.08%		0.297727
12	x17	1.34%		0.270635
13	x49	1.20%		0.265433
14	x5	1.08%		0.260946
15	x6	0.26%		0.230864
16	x8	0.22%		0.229584
17	x30	0.19%		0.22859
18	x18	0.19%		0.228532
19	x21	0.06%		0.223533
20	x29	0.01%		0.221851

c)

#	If replaced with mean	Impact on RMSE	Bars	RMSE
1	x10	41.54%		1.94042
2	x8	35.65%		1.72742
3	x11	35.24%		1.71257
4	x23	32.39%		1.60994
5	x24	27.87%		1.44631
6	x21	21.64%		1.22125
7	x36	16.38%		1.03121
8	x37	11.06%		0.839002
9	x34	5.76%		0.647692
10	x5	2.20%		0.519007
11	x47	1.27%		0.485321
12	x49	0.88%		0.471478
13	x18	0.52%		0.458154
14	x50	0.23%		0.447709

Figure 8.3: Selected pumping rates used for the development of models a) $GMDH_1$, b) $GMDH_2$ and c) $GMDH_3$

8.4.4 Analysis of variable importance

The performance evaluation results for TEST 1, TEST 2 and TEST 3 are presented in Tables 8.3, 8.4 and 8.5, respectively. In TEST 1, the GMDH models were constructed using a subset of the original dataset comprising the most influential variables only. The prediction performance in terms of the five evaluation criteria was similar if not even better than that of the models constructed during the original datasets. This result establishes that using less influential variables can help achieve satisfactory prediction results without any significant loss in accuracy.

In TEST 2, new GMDH models were constructed using 20 variables instead of the original dataset of 52 variables. These 20 variables comprised 10 of the most and least influential variables. The performance evaluation results presented in Table 8.4 show a reduction in the accuracy of the predictive models in terms of all five evaluation criteria when compared to the results of models constructed using the original dataset. This result shows that the inclusion of less influential variables in the model-construction phase can diminish the accuracy of the models.

Lastly, in TEST 3, a set of the least influential variables was used in the construction of GMDH predictive models. The developed models had worse prediction capability than the other models constructed using the original dataset. This is highlighted by the high values of RMSE, MAE and extremely small values of NSE, R^2 and r.

Table 8.3: Performance of the GMDH models constructed using the most influential variables (Test 1)

Model	No. of variables used	Stage	RMSE	MAE	NSE	$\boldsymbol{R}^2$ (%)	$\boldsymbol{r}$
$GMDH_1$	15	Training	0.356	0.284	1	99.67	0.998
		Testing	0.441	0.346	1	99.62	0.998
		Prediction	0.388	0.295	1	99.68	0.997
$GMDH_2$	20	Training	0.213	0.172	1	99.69	0.998
		Testing	0.261	0.185	1	99.62	0.998
		Prediction	0.262	0.173	1	99.63	0.998
$GMDH_3$	14	Training	0.448	0.340	1	98.70	0.993
		Testing	0.451	0.336	1	98.83	0.994
		Prediction	0.492	0.321	1	99.76	0.995

Table 8.4: Performance of the GMDH models constructed using the 10 most and 10 least influential variables (Test 2)

Model	No. of variables used	Stage	RMSE	MAE	NSE	R^2 (%)	r
		Training	0.560	0.454	0.98	99.18	0.996
$GMDH_1$	20	Testing	0.575	0.468	0.98	99.35	0.997
		Prediction	0.552	0.484	0.98	99.27	0.996
		Training	0.406	0.319	0.98	98.86	0.994
$GMDH_2$	20	Testing	0.409	0.329	0.98	99.06	0.995
		Prediction	0.411	0.365	0.98	98.83	0.995
		Training	0.528	0.403	0.98	98.29	0.991
$GMDH_3$	20	Testing	0.551	0.432	0.97	98.27	0.992
		Prediction	0.602	0.448	0.97	98.22	0.991

Table 8.5: Performance of the GMDH models constructed using the least influential variables (Test 3)

Model	No. of variables used	Stage	RMSE	MAE	NSE	R^2 (%)	r
		Training	6.078	4.787	0.13	3.57	0.189
$GMDH_1$	37	Testing	7.394	6.077	0.11	7.07	0.086
		Prediction	7.221	6.451	0.10	7.35	0.095
		Training	3.684	3.011	0.18	6.35	0.252
$GMDH_2$	32	Testing	4.298	3.398	0.15	3.64	0.016
		Prediction	4.228	3.295	0.16	3.28	0.224
		Training	3.928	3.194	0.17	5.33	0.231
$GMDH_3$	38	Testing	4.111	3.282	0.16	3.92	0.214
		Prediction	4.089	3.322	0.16	3.87	0.228

Overall, these performance evaluation results show that GMDH models can predict groundwater salinity with reasonable accuracy, and also identifying the most and least influential pumping rates. Identification of the variables most and least influential to output salinity concentrations offers practical and economic benefits to decision-makers and managers. First, it allows construction of predictive models containing fewer variables without any loss in prediction accuracy. Second, the cost related to the monitoring, collection and preparation of the dataset is lower as the decision-maker only needs focus on the most influential pumping rates.

8.5 Conclusions

In this study, a new application of GMDH models was employed to predict salinity concentration at specified monitoring wells in a coastal aquifer in response to variable transient groundwater pumping patterns. The performance assessment results suggest that the developed models can predict salinity concentrations with reasonable accuracy and efficiency. Apart from displaying accurate and efficient prediction capabilities, GMDH models provide a way to identify the most influential pumping rates. Identification of the most and least influential pumping rates offers substantial benefits to the decision-maker.

Specifically, it allows them to focus on the pumping rates most influential to the management outcomes and prediction accuracy. These characteristics offer both computational and economic benefits. Second, the results of this study establish that the identification of influential pumping rates should be given important consideration. Including the least influential variables in predictive models can diminish their accuracy. In general, the results of this study suggest that GMDH models can be considered accurate and efficient saltwater intrusion modelling tools that can be used for regional-scale saltwater intrusion prediction modelling purposes. GMDH models can also be used to comprehend the importance or influence of input pumping rates, which can be useful in developing efficient management plans for aquifer systems. In addition, the highly accurate and efficient GMDH predictive models can be used as approximate simulators in coupled simulation-optimization models to develop computationally-practicable regional-scale coastal aquifer management methodologies. However, it is advisable that researchers first compare the performance of GMDH models with other well-established predictive modelling techniques in a more rigorous manner. Such comparative study will be part of future investigations. In the next chapter, the benefit of using artificial freshwater recharge to control saltwater intrusion is qualified and a multi-objective management model incorporating artificial freshwater recharge is developed and evaluated.

Chapter 9: Development and evaluation of a multi-objective strategy for management of coastal aquifers utilizing planned artificial freshwater recharge

The major contents of this chapter have been published, as outlined below:

Lal, A., and Datta, B. (2017). "Modelling saltwater intrusion processes and development of a multi-objective strategy for management of coastal aquifers utilizing planned artificial freshwater recharge." *Modeling Earth Systems and Environment*, 1-16.

9.1 Summary

The need for freshwater is emerging as the most critical resource issue facing humanity. In several arid and semi-arid parts of the world, groundwater resources are being used as an alternative source of freshwater. Excessive and/or unplanned groundwater withdrawals have negative impacts on aquifers. Groundwater withdrawn from coastal aquifers is susceptible to contamination by saltwater intrusion. This study investigates the efficiency and viability of using artificial freshwater recharge (AFR) to increase fresh groundwater pumping from production wells. A 3D, transient, density-dependent, finite element-based flow and transport model of an illustrative coastal aquifer was implemented using the FEMWATER code. First, the effect of AFR on the inland encroachment of saline water is quantified for existing scenarios. Specifically, groundwater head and salinity differences at monitoring locations before and after AFR are presented. Second, a multi-objective management model incorporating groundwater pumping and AFR is implemented to control groundwater salinization in an illustrative coastal aquifer system. To avoid computational burden and ensure computational feasibility, support vector machine regression (SVMR) predictive models are used as an approximate simulator in the coupled simulation-optimization framework. Performance evaluation indicates that the SVMR models were adequately trained and capable of approximating saltwater intrusion processes in the aquifer. A multi-objective genetic algorithm (MOGA) was used to solve the multi-objective optimization problem. The Pareto-optimal front obtained from the SVMR-MOGA optimization model presented a set of optimal solutions for the sustainable management of the aquifer. The pumping strategies, obtained as Pareto optimal solutions with and without freshwater recharge wells, show that saltwater intrusion is sensitive to AFR. Also, the hydraulic head lenses created by AFR can be used to control saltwater intrusion in coastal aquifers. The developed 3D saltwater intrusion model, the predictive capability of the developed SVMR models, and the feasibility of using the proposed linked multi-objective SVMR-MOGA

optimization model, makes the proposed method potentially attractive in solving large-scale regional saltwater intrusion management problems.

9.2 Background

This chapter incorporates the benefits of AFR with the S/O methodology for the management of coastal aquifers and focuses on the real-world benefits of using the presented method to solve large regional-scale groundwater salinization management problems. Use of AFR can be an efficient way to solve water scarcity problems (Benseddik et al. 2017; Clark et al. 2015; Hashemi et al. 2013; Owusu et al. 2017) and control saltwater intrusion into coastal aquifers (Koussis et al. 2010; Roumasset and Wada 2010). Hashemi et al. (2013) stated that artificial recharge is a method for balancing and recovering groundwater resources and is a vital component of groundwater systems. In addition, Guttman et al. (2017) explained that artificial recharge is needed to increase the water budget of aquifers, improve water quality and prevent saltwater intrusion by creating freshwater barriers. In addition, Hussain et al. (2016) demonstrated the effects of artificial recharge on the general advancement of salt water into hypothetical and real-world coastal aquifers. However, the study did not develop optimal rates of groundwater pumping and freshwater recharge. Despite numerous benefits, only a limited number of studies have integrated the methods of artificial recharge into an S/O framework to control saltwater intrusion in coastal aquifers. Bhattacharjya and Datta (2009) presented a coastal aquifer management model which maximized groundwater pumping and minimized artificial recharge. However, the study utilised ANN models as surrogate models, which have some complexities and limitations (Tu 1996). More robust and efficient surrogate modelling techniques are needed to obtain improved and reliable optimal results. More recently, Abd-Elhamid and Javadi (2011) presented a methodology termed ADR (abstraction, desalination and recharge) to control saltwater intrusion into coastal aquifers. However, their proposed management model only considered a single objective function.

A linked S/O model used for developing a coastal aquifer management strategy requires linking of a coastal aquifer simulation model to an appropriate optimization algorithm. However, as discussed in the previous chapters, linking a simulation model to an optimization model is highly computationally demanding. An efficient solution to this problem is to use a trained and validated surrogate model instead of the complex numerical simulation model. Utilising a surrogate model significantly decreases the computational requirements, which subsequently reduces the total optimization time. In this study, a comparatively new predictive modelling technique; SVMR-based surrogate models instead of the complex the numerical simulation model are used in the S/O framework. A detailed description of SVMR models is given in Chapter 3.

Another key feature of this study is its incorporation of planned AFR as a management strategy within a multi-objective S/O framework to control saltwater intrusion in a coastal aquifer. As mentioned earlier, only a few studies have explored the use of multi-objective coastal aquifer management models incorporating artificial recharge as decision variables. In this study, groundwater pumping from production wells and artificial recharge injection at recharge wells is considered as a method for controlling saltwater intrusion in a simulated coastal aquifer. This study is divided into three components: 1) quantifying the effects of artificial recharge on the study area and demonstrating lens formation due to AFR; 2) development and evaluation of SVMR models for approximating salinity concentration at respective monitoring locations; and 3) using the developed SVMR models as surrogate models in an S/O framework to prescribe optimal groundwater pumping and artificial recharge rates.

9.3 Methods

9.3.1 Quantifying the effects of artificial freshwater recharge on the study area

The use of artificial recharge to control saltwater intrusion has been proposed as an effective method for controlling groundwater salinization. In a laboratory-scale study, Luyun et al. (2011) demonstrated that artificial recharge creates a hydraulic barrier by raising the piezometric head of the aquifer, thereby controlling saltwater intrusion. Similarly, Ros and Zuurbier (2017) suggested that water artificially recharged through wells remains at the top of the aquifer and forms local freshwater lenses that form barriers to saltwater intrusion. Figure 9.1 illustrates the impact of groundwater pumping and artificial recharge on saltwater intrusion processes. Before developing an optimal coastal aquifer management strategy based on using AFR in an S/O approach, assessing the effect of artificial recharge on salinity levels at monitoring locations was essential. Salinity concentrations and head differences over time at monitoring locations with and without recharge were evaluated. In addition, differences in head contours before and after AFR were appraised, and the possibility of lens formation due to AFR was considered. Random pumping and recharge rate datasets generated via Latin hypercube sampling (LHS) were implemented in the numerical simulation model, and a comparison between salt concentrations before and after artificial recharge initiation was carried out. LHS is a homogeneous stratified sampling technique established from the Monte-Carlo method (Loh 1996).

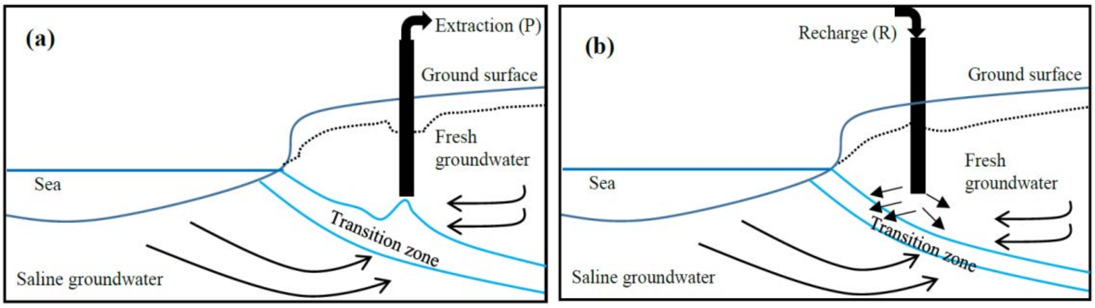


Figure 9.1: (a) Movement of saline water into fresh groundwater as a result of groundwater pumping (modified from Essink, 2001) and (b) controlling fresh groundwater salinization through artificial recharge.

9.3.2 Management model utilizing production wells and artificial recharge wells

The key components of the proposed coastal aquifer consisted of objective functions, variables and constraints. The main goal of the proposed coastal aquifer management model was to maximize the total volume of groundwater pumped from pumping wells and minimize injection at recharge wells. Hence, the management model consisted of two objective functions. The variables used are the transient pumping and recharge rates at pumping and recharge wells, respectively. The linked SVMR model within the SVMR-MOGA optimization framework acted as a binding constraint defining the response of the aquifer to various stresses; i.e., recharge and pumping. Also, maintaining the salinity levels at pre-specified values were treated as constraints i.e., defining the permissible salinity limits at different locations based on specified water use types. Mathematical expressions of the two conflicting objective functions, the constraints, and the pumping and recharge bounds; similar to Sreekanth and Datta (2011) are given by:

Objective Function I

Maximize,

$$F_1(P) = \sum_{l=1}^{L}\sum_{t=1}^{T} P_l^t \qquad (9.1)$$

Objective Function II

Minimize,

$$F_2(P) = \sum_{q=1}^{Q}\sum_{t=1}^{T} R_q^t \tag{9.2}$$

Constraints $c_i = \xi(P, R)$ (9.3)

$c_i \leq c_{max}\ \forall i, t$ (9.4)

Bounds $P_{min} \leq P_l^t \leq P_{max}$ (9.5)

$R_{min} \leq R_q^t \leq R_{max}$ (9.6)

P_l^t denotes pumping from the n^{th} PW at time t and R_q^t denotes artificial recharge (injection) from the m^{th} RW at time t. c_i represents the salinity concentration at the i^{th} monitoring well at the end of the management time period. $\xi(,)$ symbolizes the surrogate model replacing the numerical FEMWATER model and constraint (9.3) denotes the coupling of the surrogate model within the S/O framework. Variables L, Q and T denote the total number of pumping wells, recharge wells, and time steps in the management model, respectively. Inequality (9.4) represents the constraints imposed to keep salinities within specified limits at the respective monitoring locations. Inequalities (9.5) and (9.6) represent the upper and lower bounds of pumping and artificial recharge injection rates at pumping wells and recharge wells, respectively.

9.4 Application of the management model to an illustrative study area

9.4.1 Description of the study area

The developed management model was applied to an illustrative coastal aquifer. A detailed description of the aquifer system is presented in Chapter 3. The study area (Fig. 9.2) contained a portion of a multi-layered coastal aquifer system. The FEMWATER modelling paradigm was used to develop a saltwater intrusion numerical simulation model for the study area. Details of the FEMWATER modelling paradigm, well locations, and aquifer layers are presented in Chapter 3. The study area incorporated five recharge wells (RWs), eight production wells (PWs) and three monitoring locations (MLs). PWs were installed for withdrawing fresh groundwater for beneficial use, whereas RWs were installed to prevent saltwater intrusion by creating freshwater barriers (Guttman et al. 2017). Recharge wells were installed outside the intruding saltwater wedge to allow saltwater intrusion repulsion with increasing recharge rates, as demonstrated in Luyun et al. (2011). Salinity concentration was recorded at the respective MLs.

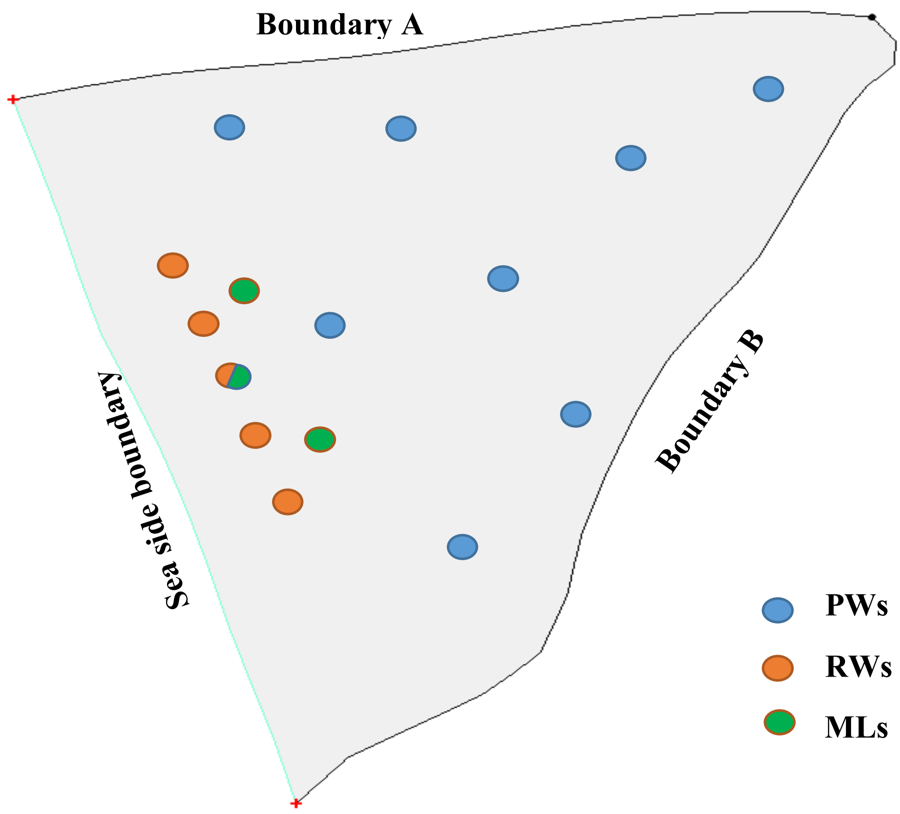


Figure 9.2: Study area and locations of PWs, RWs and MLs

9.4.2 Boundary conditions, model discretization and key aquifer properties

The seaside boundary had a constant head and constant concentration boundary with a concentration of 35 kg/m^3. Boundary A and Boundary B of the study area were treated as no-flow boundaries. The modelled aquifer was discretised into triangular finite elements with an average element size of 150 m. Constant groundwater recharge of 0.00054 m/d was specified over the entire study area. The compressibility and dynamic velocity of water were taken as 6.69796 × 10^{-20} md^2/kg and 131.328 kg.md, respectively. Other key parameters used for the aquifer simulation are listed in Table 9.1.

For the present case study, pumping and injection bounds for both the PWs and RWs were set between 0 – 1300 m^3/day. The constraints imposed as permissible limits on salinity concentration (assumed to be a conservative pollutant) were $c_i \le c_{max,i}$ of 1000 mg/L at ML1, $c_i \le c_{max,j}$ of 400 mg/L at ML2 and $c_i \le c_{max,k}$ of 400 mg/L at ML3. A management period of 4 years (1460 days) was considered for this study. A total of 52 variables (8 PW × 4-year management period and 5 RWs × 4-year management period) was considered.

Table 9.1: Aquifer properties used in the numerical simulation model

Property		Layer 1 value	Layer 2 value
Hydraulic conductivity	x-direction	15 m/d	20 m/d
	y-direction	7.5 m/d	10 m/d
	z-direction	1.5 m/d	2 m/d
Bulk density		1600 kg/m^3	1500 kg/m^3
Longitudinal dispersivity		50 m/d	50 m/d
Lateral dispersivity		25 m/d	25 m/d
Molecular dispersion coefficient		0.69 m^2/d	0.69 m^2/d
Density reference ratio		7.14 x 10^{-7}	7.14 x 10^{-7}
Soil porosity		0.43	0.46
Compressibility		8.5 x 10^{-15} md^2/kg	8.5 x 10^{-15} md^2/kg

9.4.3 Development of surrogate models and cross-validation

The coastal aquifer flow and transport models were used to generate sets of input-output patterns to be used during the surrogate model construction and prediction phases. Some 600 transient pumping and injection rates (inputs) were obtained from a uniform sampling distribution using LHS, with an upper bound of 1300 m^3/d and lower bound of 0 m^3/d. These 600 pumping/injection rate input sets were fed into the numerical simulation model and the resulting salinity corresponding to each input dataset was recorded. Each numerical simulation model took approximately 4-5 minutes to converge. The 600 input datasets and resulting output salinities were assembled by running the simulation 600 times. These input-output patterns were used in the surrogate model construction phase.

Three SVMR models were constructed (M I, M II and M III). These models are capable of approximating salt concentrations at the three corresponding MLs. The SVMR models were constructed by learning from training data with the intent of determining the functional relationship between the pumping/injection rate and salinity datasets. For cross-validation (CV) purposes, the 'holdout' option (Lin et al. 2008) was chosen, in which the 600 datasets to be used during the surrogate model construction phase were partitioned randomly into training (80%) and testing datasets (20%) without replacement. Once the models were trained and tested, a separate test case of transient pumping and recharge rates (100 datasets obtained via LHS sampling) were fed into the respective predictive models to obtain resulting salinity concentration datasets from the corresponding MLs. The salinity concentration data from the SVMR was compared with the data from the numerical simulation model to evaluate the predictive capabilities of the developed models. Selecting the best kernel function and its associated parameters was a key challenge of this study. However, after trial-and-error, a Gaussian kernel was selected with parameters $\varepsilon = 0.60$, $C = 6.49$ and $\gamma = 0.001$. The predictive performance of the developed SVMR models was tested using the statistical indices RMSE, RE, NSE and *r*.

9.4.4 Coupled simulation-optimization model

The developed SVMR surrogate models were linked to the MOGA after evaluation of their predictive capabilities. The MOGA model available in the R2017a MATLAB environment allowed computation of an optimal solution in the proposed multi-objective management model. Selecting the key parameters for the MOGA was a challenge and was accomplished after doing numerous runs and trials. The important MOGA parameters used during the implementation of the SVMR-MOGA optimization model are listed in Table 9.2.

Assessing the validity of the SVMR-MOGA model's optimal solutions was an integral part of the present work. The optimal solutions obtained from the SVMR-MOGA model utilized the developed surrogate models instead of the original complex numerical model. Therefore, the solutions had to be verified with the original numerical simulation model. To quantify the validity of the optimal solutions, random optimal solutions (comprising transient pumping and recharge rates) were chosen and implemented into the complex numerical simulation model. The resulting salinity values obtained from the numerical model were compared with the corresponding salinity predictions of the SVMR models.

Table 9.2: Key MOGA parameters used in SVMR-MOGA framework

Parameter	Population size	Generations	Function tolerance	Constraint tolerance	Crossover fraction	Mutation probability
Value	1500	5200	1×10^{-5}	1×10^{-4}	0.8	0.02

9.5 Results and discussion

9.5.1 Three-dimensional saltwater intrusion modelling results

The flow and transport process in an illustrative coastal aquifer was modelled using a finite element-based numerical flow and transport simulation model (FEMWATER). Figure 9.3 shows the modelled aquifer with the respective well locations and salinity concentration front after a period of four years. Each set of pumping and artificial recharge rates at the PWs and RWs were fed into the simulated model and the resulting concentrations at MLs were recorded. These input-output datasets values were used to construct SVMR predictive models capable of approximating salinity concentration in response to pumping and artificial recharge rates within a similar domain. Therefore, to generate datasets for surrogate model construction and to evaluation, the numerical simulation model was executed a number of times, each with different PW pumping and RW artificial recharge rates obtained by LHS. The training, testing and prediction results are discussed in the next section.

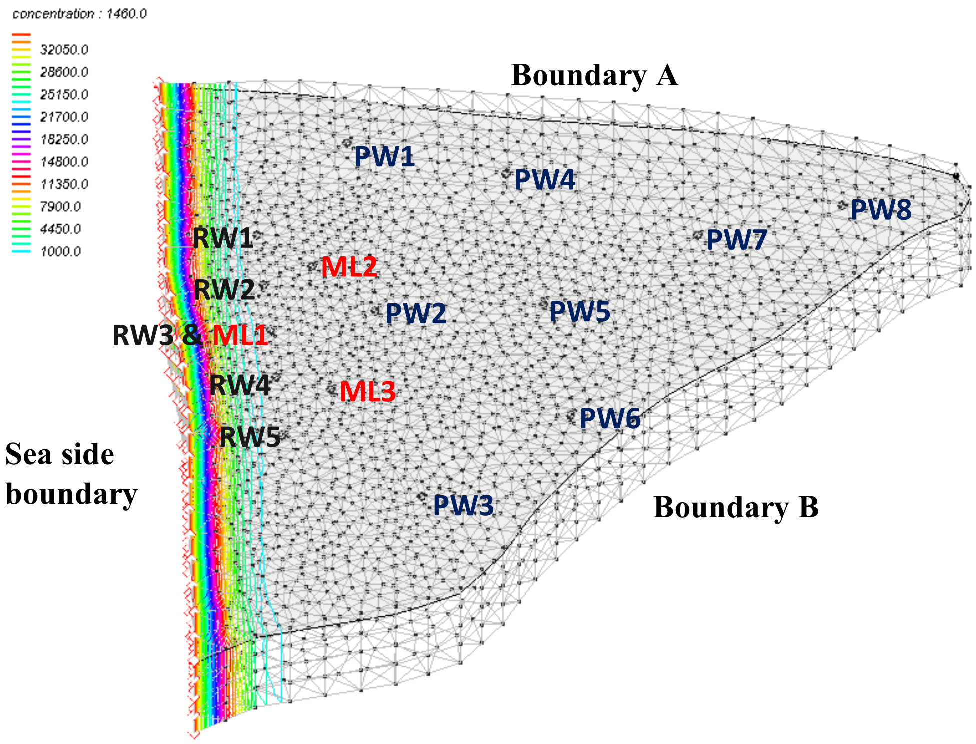


Figure 9.3: 3D view of the simulated coastal aquifer with concentration contours after 4 years of simulation

9.5.2 Effect of artificial recharge on saltwater intrusion

9.5.2.1 Salinity concentrations and head differences at monitoring wells over time

The potential effect of artificial recharge was evaluated by monitoring heads and salinities at the three MLs over time. Figures 9.4 and 9.5 illustrate the difference in these values with and without artificial recharge. It was observed that artificial recharge caused an increase in the head values. Head values at ML1 increased more than at ML2 and ML3 because it was also used as a RW for freshwater injection. In terms of the salinities at MLs, an inverse trend was observed: a decrease in salinity concentration was recorded when artificial recharge was initiated. Without any freshwater recharge, salinity increased over time (from the 1st to 4th time steps). However, initiation of artificial recharge decreased the salinity concentration compared to salinity concentration with no artificial recharge.

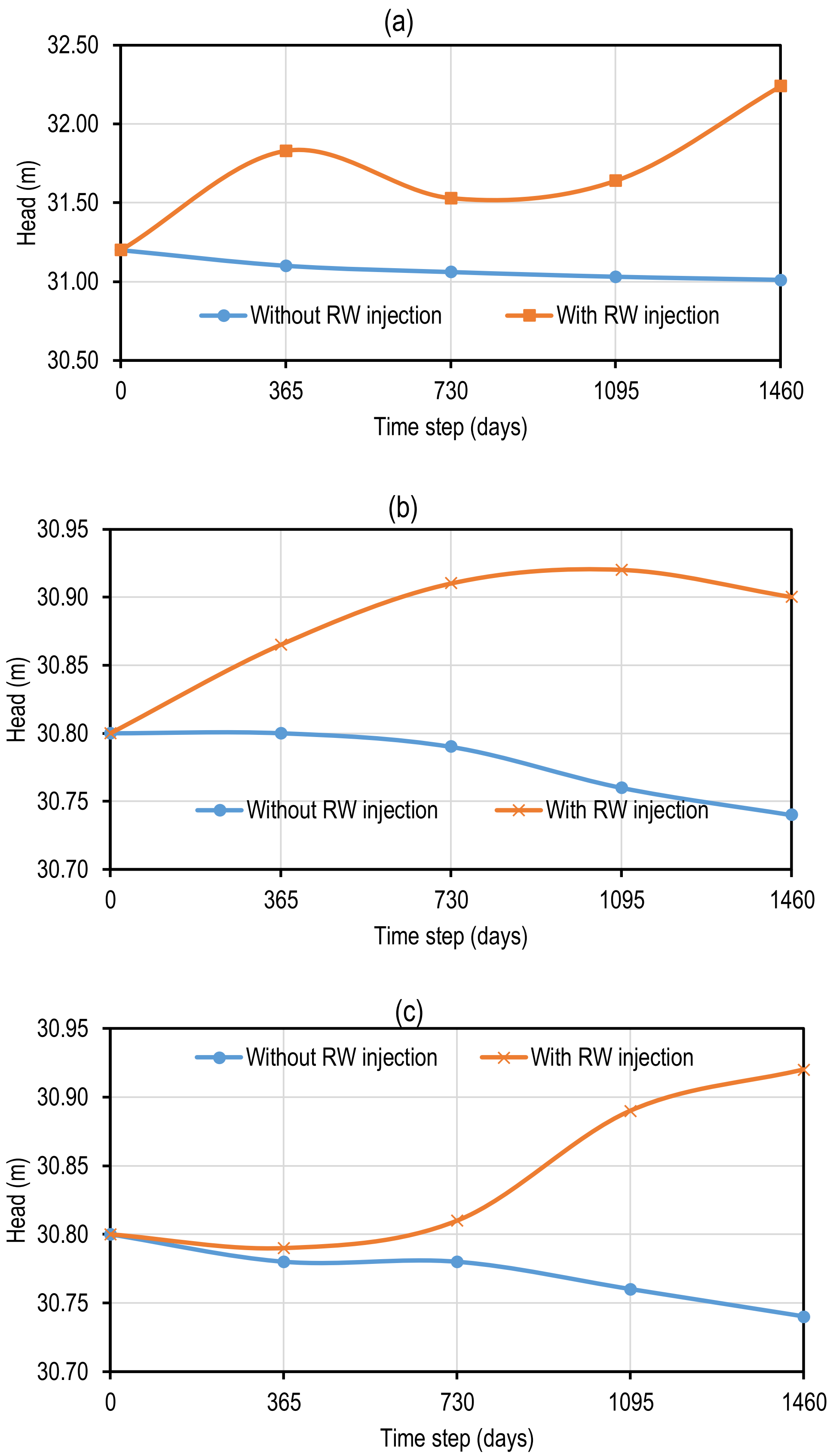


Figure 9.4: Changes in head over time at (a) ML1 (b) ML2 and (c) ML3

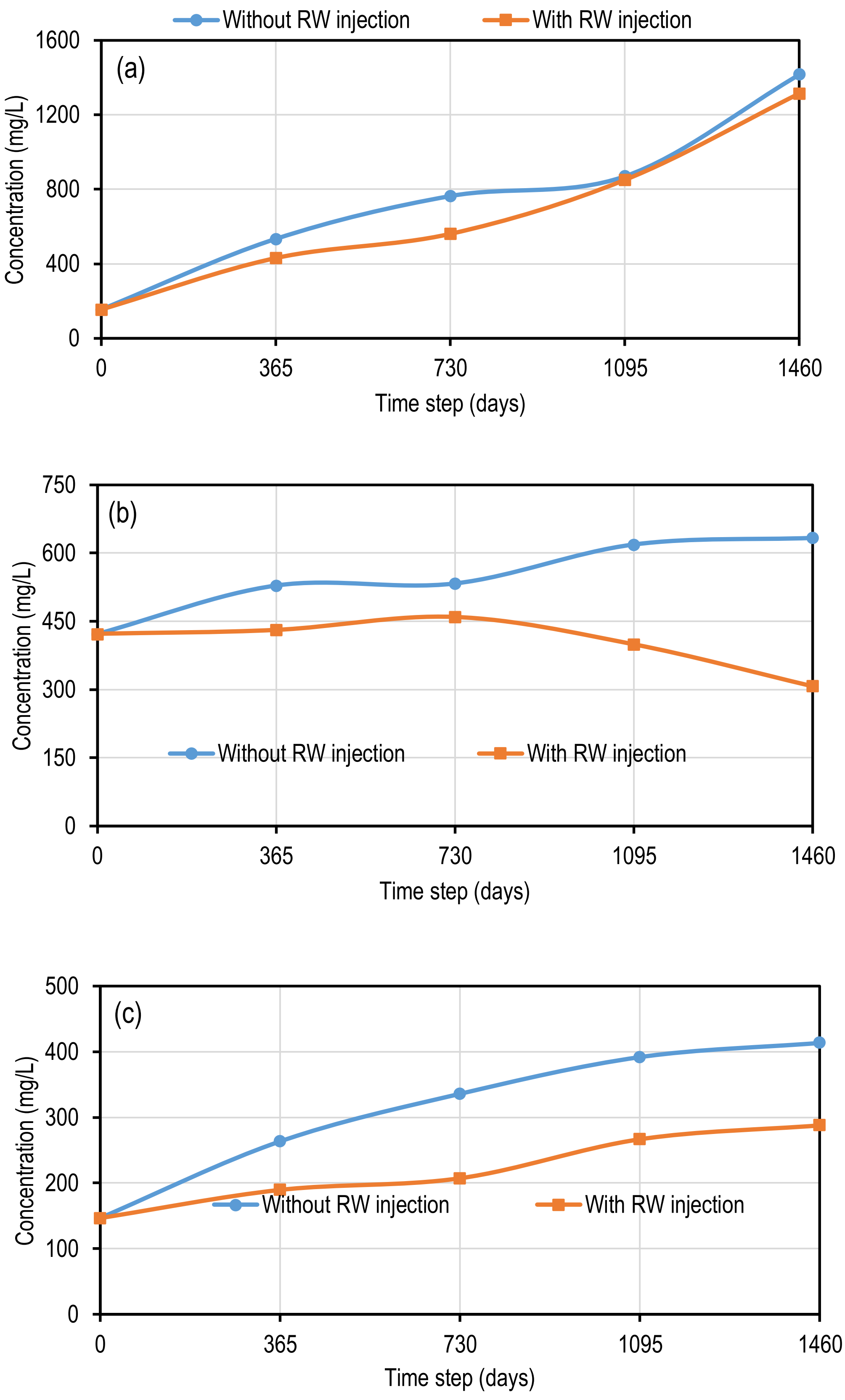


Figure 9.5: Change in salinity concentrations over time at (a) ML1 (b) ML2 and (c) ML3

9.5.2.2 Comparison of head and lens formation due to artificial freshwater recharge

The head contours before and after AFR using a set of random pumping and artificial recharge rates are illustrated in Fig. 9.6. An increase in head was observed after AFR initiation, as shown in Fig. 9.6 (b). This increase in head was responsible for reducing saltwater intrusion into the aquifer and controlling salinity levels at the MLs. To further demonstrate the formation of lenses due to AFR, two cross-sections (A-A' and B-B') were created. Figure 9.7 shows these two cross-sections and the observed lens formation near the RWs. At cross-section A-A', lens formation was detected, as seen in the obvious head increase near the RWs. However, no such increase in head was observed at cross-section B-B' where no RWs were installed.

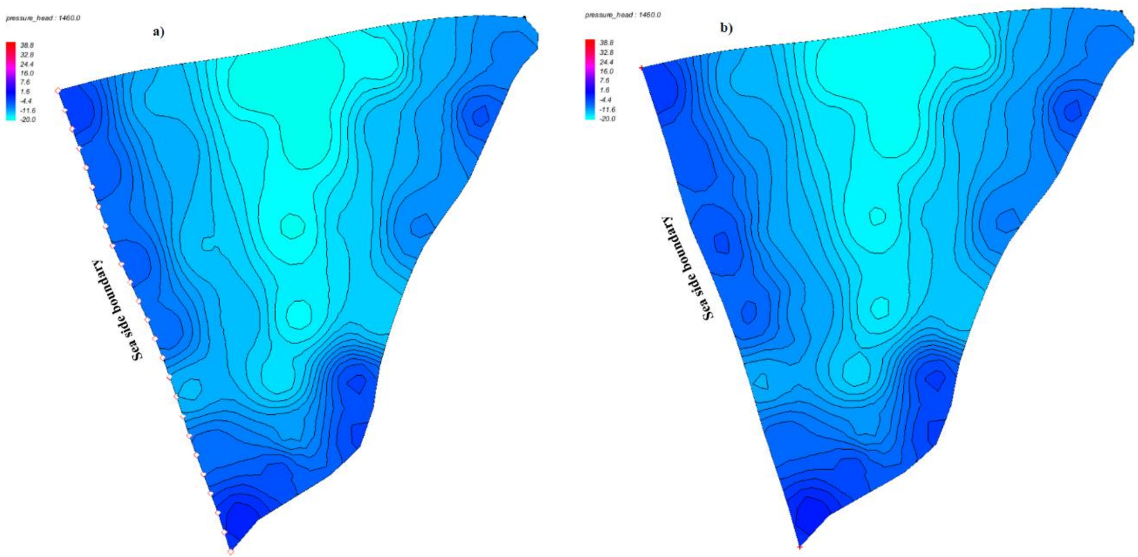


Figure 9.6: Head contours before (a) and after (b) artificial freshwater recharge initiation

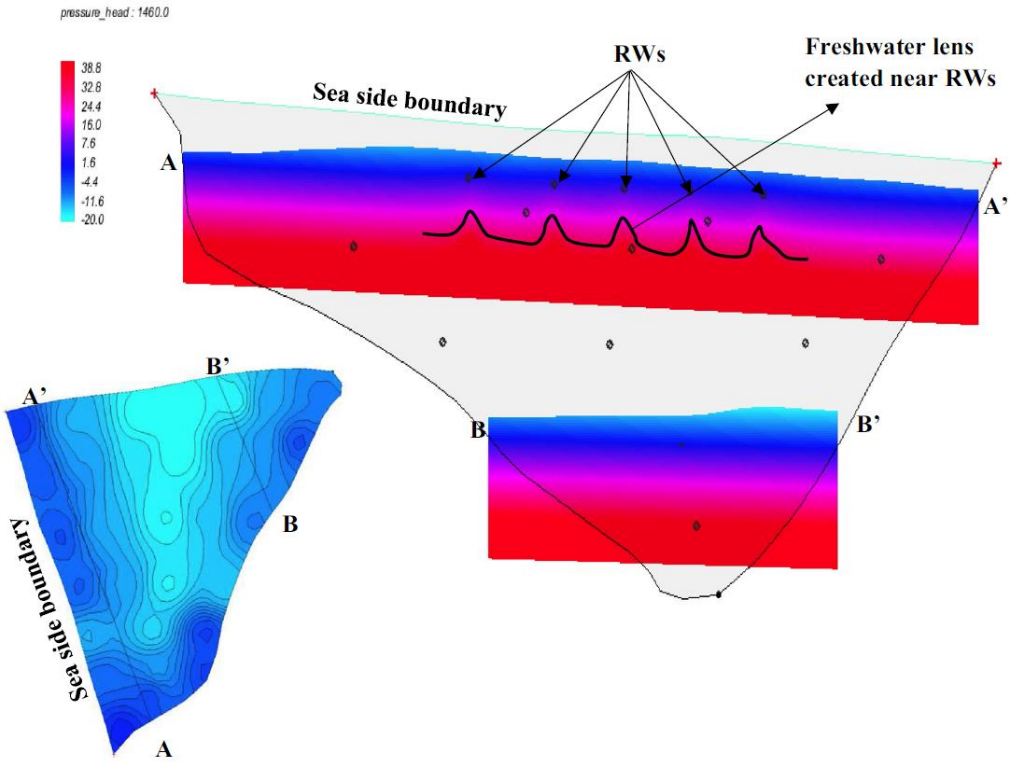


Figure 9.7: Lens formation due to AFR at cross-section A-A' after 1460 days

9.5.2.3 Salinity concentrations at monitoring wells after the 4th time step

Five randomly selected pumping and recharge sets were selected to evaluate the effect the artificial recharge on the study area. The comparison in Fig. 9.8 establishes the influence of artificial recharge on the simulated aquifer. For each of the five sets of pumping and recharge rates used, significant decreases in salinity were observed at all MLs after artificial recharge initiation. This evaluation suggests that artificial recharge has the potential to control salinity intrusion into the aquifer, resulting in a substantial increase in beneficial pumping. Hence, the proposed management model, incorporating groundwater pumping with an AFR option, was formulated and solved.

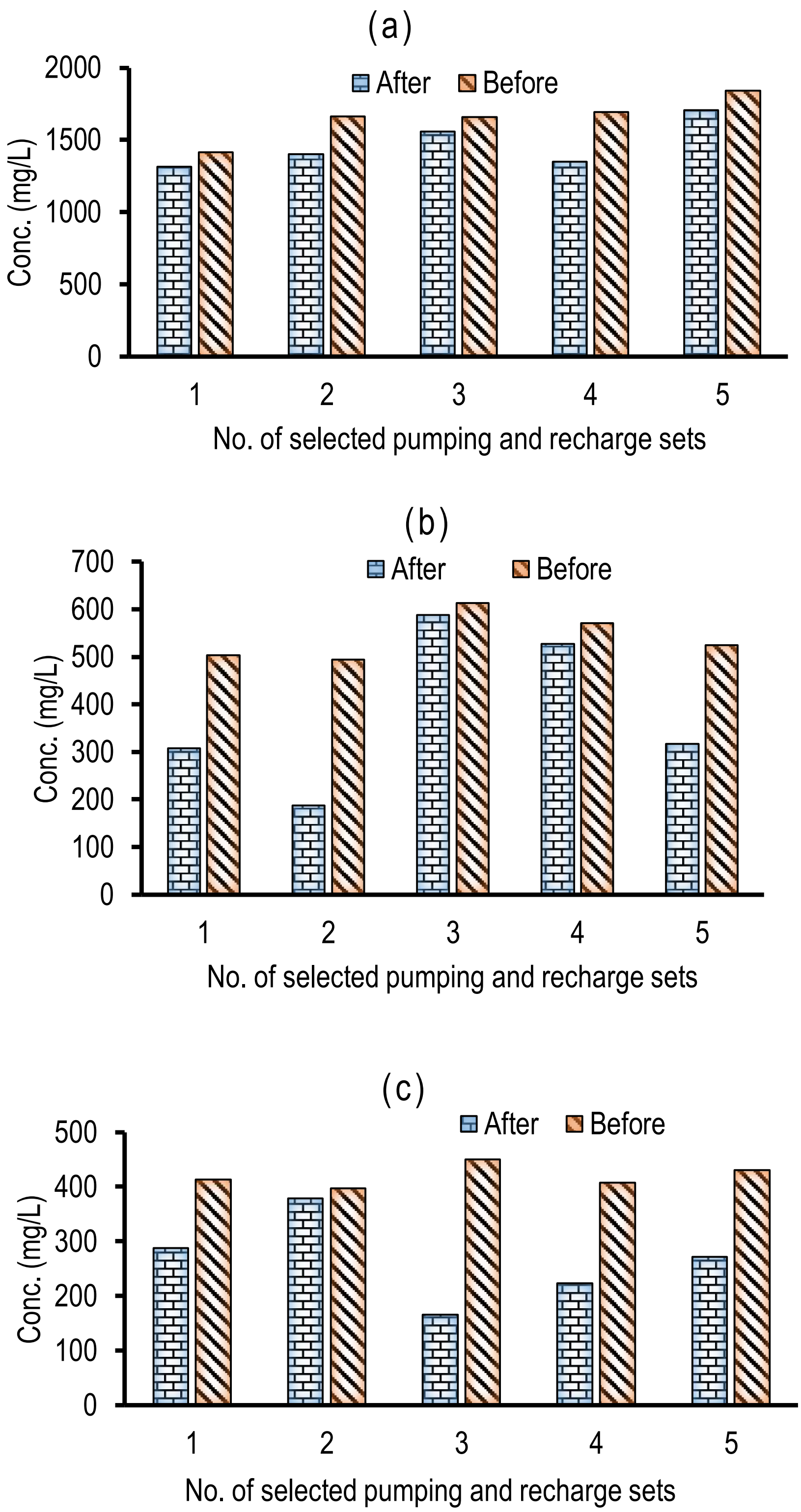


Figure 9.8: Salinity concentrations before and after artificial recharge initiation at (a) ML1 (b) ML2 and (c) ML3

9.5.3 Training- and testing-phase performance

Three SVMR surrogate models (M I, M II and M III) were developed for predicting salinity concentrations at the corresponding MLs. Performance measures of the developed SVMR models at the training and testing phases are given in Table 9.3. The mean RMSE, RE (%), NSE and *r* (%) at the training stage were 3.19, 0.03, 0.98 and 97.4, respectively. A similar trend in the results was found during the testing phase. The training and testing results indicate that properly trained and tested models are capable of emulating the complex numerical simulation model's responses when supplied with pumping and injection rate datasets within the same domain.

Table 9.3: Training and testing phase performance of the surrogate models (SM)

SM	Training phase				Testing phase			
	RMSE	RE (%)	NSE	*r* (%)	RMSE	RE (%)	NSE	*r* (%)
M I	3.08	0.01	0.98	96.6	4.38	0.03	0.98	96.9
M II	2.92	0.02	0.99	97.9	2.71	0.03	0.99	98.7
M III	3.56	0.06	0.98	97.7	2.42	0.09	0.99	97.6

9.5.4 SVMR model prediction capabilities

The performance of the developed SVMR models in terms of prediction accuracy was quantified using a totally different prediction dataset. The prediction results for M I, M II and M III are listed in Table 9.4. The prediction results agree with the training and testing results in terms of the calculated error estimates. On average, the values of RMSE, RE (%), NSE and *r* (%) were 5.19, 0.11, 0.98 and 97.7, respectively. A slight increase in the RMSE was found for M I due to a greater range in the salinity values obtained at ML1. However, the NSE and *r* (%) values of 0.97 and 96.1, respectively, for M I are within an acceptable range.

Table 9.4: SVMR surrogate models (SM) prediction errors

SM	Prediction performance			
	RMSE	RE (%)	NSE	*r* (%)
M I	8.70	0.06	0.97	96.1
M II	2.94	0.13	0.99	98.8
M III	3.93	0.14	0.98	98.3

An NSE value of 1 indicates a perfect model (He et al. 2014); i.e., the errors in estimation are virtually zero. A model can be considered accurate if the calculated NSE value is greater than 0.8 (Shu and Ouarda 2008). All the NSE values for M I, M II, and M III were near 1, indicating accurate salinity concentration prediction capabilities. Overall, the observed trends in the prediction results indicate that the three SVMR models were adequately trained and could be utilised for predicting salinity concentrations at the respective MLs.

9.5.5 Optimal threshold groundwater pumping and recharge rates

The executed SVMR-MOGA optimization model presented a set of optimal solutions from which one can be chosen and implemented, although the set may need additional preference ranking. The Pareto-front obtained from the multi-objective optimization methodology (Fig. 9.9) provides a trade-off between two objectives values so that decision-makers can select the best solution from multiple Pareto optimal solutions.

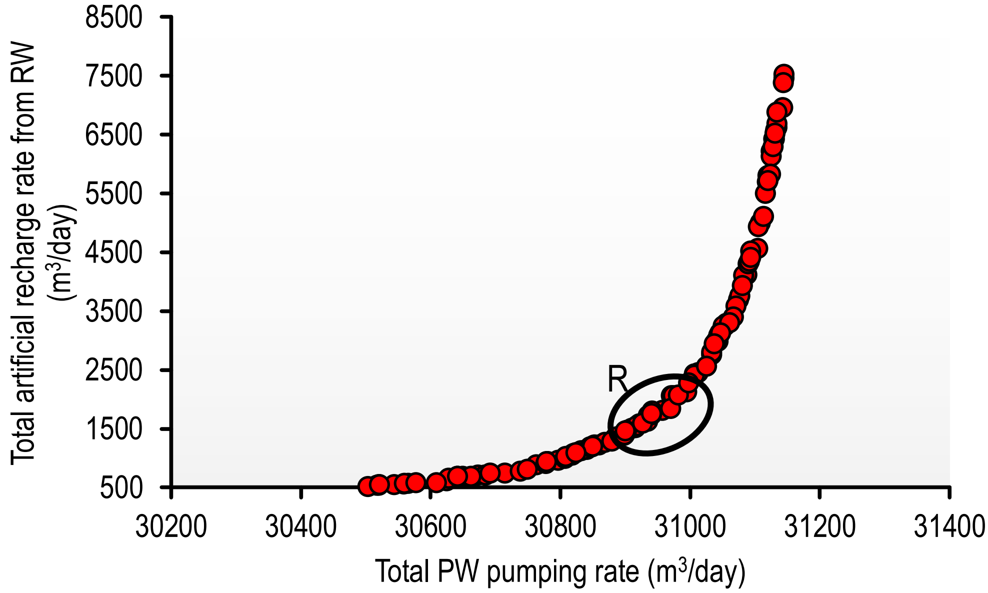


Figure 9.9: Optimal Pareto-front of the executed SVMR-MOGA model

The Pareto optimal solutions for groundwater pumping rates at PWs ranged between ~30,500 m^3/day and ~–31,200 m^3/day. The corresponding artificial recharge rates at the RWs ranged from ~500 m^3/day to ~8000 m^3/day. Also, a decent spread in the solutions was seen between the two extrema of the Pareto optimal solutions obtained. However, it is to be noted that the marginal gain in total pumping resulting from increased artificial recharge at the RWs was generally small, especially at smaller total pumping values, as evident in Fig. 9.9. This may not represent typical outcomes in general, as these trade-offs are dependent on the location of the RWs and their effectiveness in raising the groundwater table and head. Also, the study horizon was limited to four years and, therefore, the beneficial consequences of freshwater recharge were not evident, as it takes time to influence hydraulic heads at locations far from RWs. Therefore, the potential benefit of recharging the aquifer was further analysed in terms of increases to the water table/hydraulic heads. A random optimal solution from the optimal Pareto front was selected and implemented in the numerical model. The changes in heads at all three MLs before and after AFR are illustrated in Fig. 9.10. Increases in the heads at all three MLs were recorded. A significant increase in the head was recorded for ML1 because it was also used as an AFR well. Salinities obtained as an optimal solution were also recorded and found to be within specified limits as restricted by the constraints (refer Fig. 9.11).

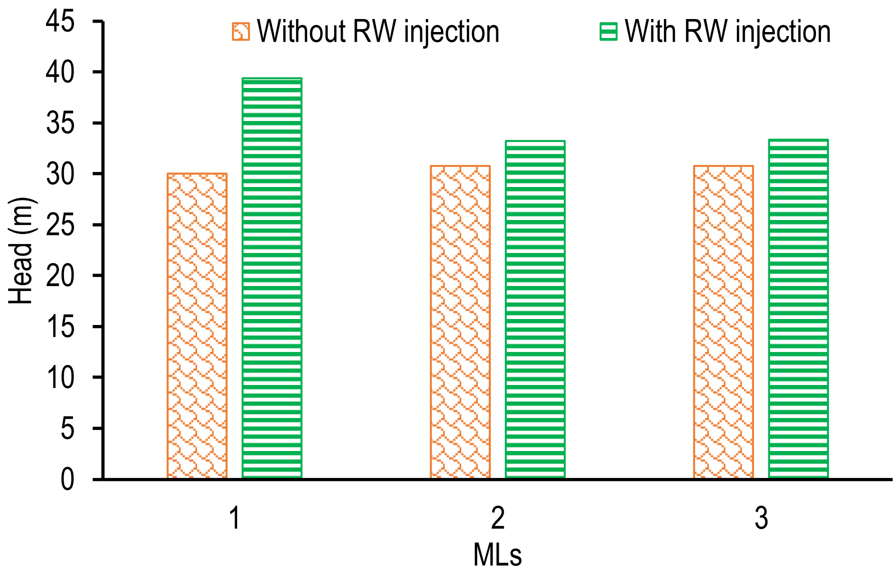


Figure 9.10: Changes in the head at MLs with and without AFR initiation

Selecting a solution from a Pareto-front is difficult when the preferences of the decision-maker are unknown. Recently, many studies have proposed methods that allows decision-makers to easily analyse huge sets of trade-off optimal solutions on the Pareto-front based on their preference ranking. The preference order is generally based on the implicit utility function of the decision-makers and the trade-offs required for changing the level of objectives (Datta and Peralta 1986). Another common approach is to detect and implement a solution from the knee region of the Pareto-front (Gong et al. 2016; Juang et al. 2014). Knee regions are the parts of Pareto-fronts that represent the maximum trade-off between objectives (Bechikh et al. 2011). Solutions from knee regions are preferred because of the fact that a small improvement in either objective will cause a large deterioration in at least one other objective, which makes moving in either direction unrealistic. Thus, in cases where the preferences of the decision-maker are unknown, a solution from the knee region is suggested. Similarly, for the present case, a solution from the knee region (region *R* in Fig. 9.9), which is the best compromise solution, is most likely to be chosen so that optimal pumping from PWs and artificial recharge injection from RWs can be implemented. This implementation will ensure the optimal management of the simulated coastal aquifer by maintaining salinity levels beyond specified limits.

9.5.6 Validation of the optimal solutions

The accuracy of the surrogate model in predicting salinity concentrations at the respective MLs in the proposed SVMR-MOGA optimization model was evaluated by comparing the surrogate model's solutions with those of the original numerical flow and transport model. Ten random solutions from the Pareto-front were chosen and implemented in the original numerical simulation model. The relative error (RE) between the predicted salinities and the numerical responses are illustrated in Fig. 9.11. It was observed that the salinity concentration values from both models were very close. The minimum RE (within 5%) signified that the concentration values from the SVMR and numerical simulation model were very close to each other. Also, it was observed that when the optimal transient pumping and recharge rates were implemented into the numerical simulation model, the concentrations values were below the maximum allowable salinity concentration limit as specified in the constraints at MLs. Also, the optimal solutions converged to the specified upper bound of the salinity constraint. Overall, the validation results demonstrate that the SVMR surrogate models accurately predicted salinity at the MLs and can be utilized as efficient tools replacing complex, nonlinear, coupled, density-dependent, numerical simulation models; for developing sustainable coastal aquifer management strategies.

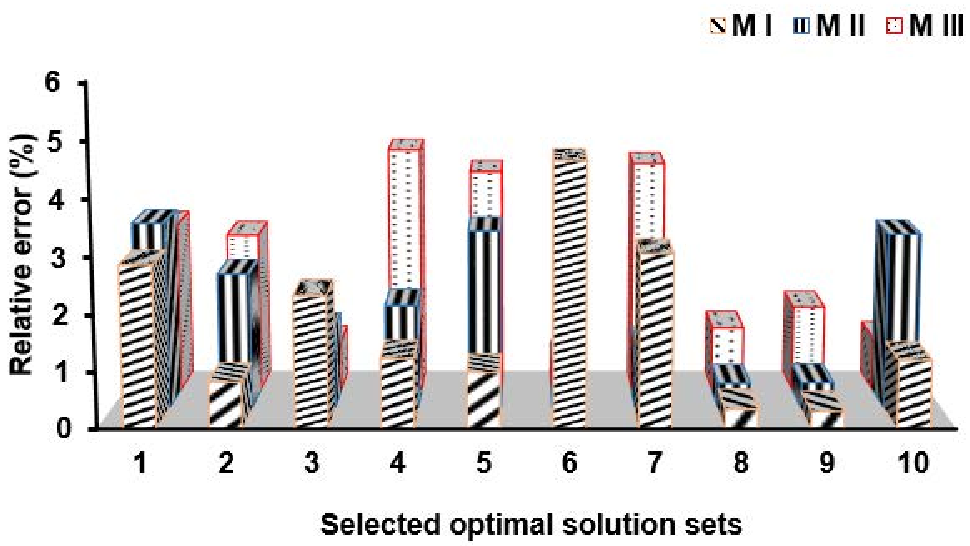


Figure 9.11: Comparison of salinity concentration obtained from the numerical model and the SVMR predictive models, generated utilizing ten random optimal solution sets

9.6 Conclusions

This study demonstrates the significant role of artificial recharge in controlling saltwater intrusion in coastal aquifers. The findings of this study demonstrate that integrated utilization of freshwater recharge and pumping strategies can form an efficient and practical solution to the current groundwater salinity crisis affecting people residing in coastal areas. The numerical modelling results quantify the effects of artificial recharge on heads and salinity levels at the three MLs. A considerable decrease in salinity levels occurred when groundwater pumping was combined with artificial recharge. Injected freshwater maintained a seaward gradient in the system by raising the inland piezometric head. Following the analysis of the developed 3D model results, and quantification of changes in head and salinity at monitoring wells over time, it could be inferred that saltwater repulsion increases with freshwater recharge. The simulation model results indicate that freshwater lenses created by AFR initiation could be used to efficiently manage coastal aquifers subject to saltwater intrusion. After this quantification of the practical benefits, recharge wells and production wells were considered in the development of a coastal aquifer management model capable of specifying regional-scale management strategies.

A key component of this study included the use of a comparatively new, trained, SVMR-based surrogate model to predict saltwater intrusion processes in coastal aquifers. In the pursuit of achieving a linked surrogate model-based S/O optimization methodology for the sustainable management of coastal aquifers, improvements in the surrogate model's prediction capabilities were critical. The SVMR approach has several benefits and enhanced prediction efficiency. SVMR models were used to emulate the flow and transport simulation model once trained using the numerical model's solutions for randomized inputs. The surrogate model's responses were then linked to an S/O framework to ensure computational feasibility of the optimization-based management model. Evaluation of the surrogate model's performance indicated that models M I, M II and M III were sufficiently trained and could accurately predict salinity at the respective MLs. After performance evaluation, the SVMR models were successfully linked to a MOGA optimization model to obtain optimal pumping and recharge rates while satisfying salinity limits specified as constraints. The multi-objective optimization model derived a set of Pareto-optimal spatio-temporal pumping and artificial recharge strategies, which can be implemented for regional-scale optimal management of the coastal aquifer.

The obtained Pareto-front offers a substantial benefit to decision-makers in terms of optimal management strategy selection. The Pareto-front obtained from the linked SVMR-MOGA model will enable decision-makers to understand the trade-offs between pumping from PWs and artificial recharge injection at RWs. In general, the proposed method can successfully deliver optimal, integrated, pumping and recharge solutions while maintaining saltwater intrusion levels within acceptable limits. The accurate predictive capability of the developed SVMR models and the feasibility of the proposed linked multi-objective SVMR-MOGA optimization model makes the proposed methodology potentially attractive in solving large-scale, regional, coastal aquifer management strategies. In future, assessing the practicality of the developed saltwater intrusion management strategy in a real case study area would be beneficial. Lastly, while this study has demonstrated the potential of utilizing artificial recharge to minimise seawater intrusion in coastal aquifers, it would be valuable to identify optimal locations for installing recharge wells. Optimal locations could enhance the effect of artificial recharge, potentially improve groundwater usability, and provide more efficient and cost-effective ways to manage saltwater intrusion in coastal aquifers. The next chapter summarises the major findings of this thesis and draws the overall conclusions.

Chapter 10: Conclusions and recommendations

10.1 Conclusions

The main aim of this thesis was to develop and evaluate the performance of computationally-efficient and feasible strategies for the management of coastal aquifers. First, SVMR models were developed and utilized to predict saltwater intrusion by approximating the responses of a complex variable-density flow and transport numerical model. For evaluation purposes, the performance of the SVMR models was compared with that of the well-established GP modelling algorithm. The performance evaluation results revealed that the SVMR model is superior to GP models and can be applied to obtain precise and dependable SWI predictions. The evaluation results suggest that the SVMR predictive models can be applied in groundwater salinity management as computationally-efficient substitutes for numerical simulation models. Another advantage of utilizing SVMR surrogates is that the time required to train and validate them is significantly less than that required by GP models. Also, a sensitivity analysis method was presented and evaluated for ranking input variables for surrogate models. The sensitivity analysis presented a set of input variables most influential to the corresponding outputs. Thus, retraining a surrogate model by refining the input dataset and using only the most influential variables can yield a superior predictive model that offers substantial benefits in saltwater intrusion prediction.

Secondly, after validating the use of the SVMR surrogate model, it was used in an S/O model to develop a computationally-efficient and feasible multi-objective management strategy for the Bonriki aquifer system. The management model incorporated pumping from FPWs and the option of pumping from BWs to maximize the supply of freshwater to the local South Tarawa community. The SVMR surrogate models were trained and tested using input (pumping rates at FPWs and BWs) and output (salinity concentration) datasets generated by the calibrated and validated Bonriki aquifer numerical simulation model. The developed SVMR surrogate models were externally linked to a MOGA optimization model. The executed, multi-objective, linked S/O-based management model generated an optimal Pareto-front presenting different trade-offs between the total FPW and BW pumping solutions. Selection of a single solution from this huge optimal solution set was deemed a difficult task, especially for a group of decision-makers. Therefore, the *k*-means clustering technique was used to group solutions with similar features. The *k*-means clustering technique pruned the Pareto front and presented a more workable reduced set of optimal solutions for decision-makers. Overall, our limited performance evaluations show that the suggested methodology for solving multi-objective aquifer management problems has the potential for application to other islands facing similar saltwater intrusion problems.

Thirdly, to ensure the robustness of the linked S/O model, a method was presented that incorporates uncertainty in aquifer parameters (hydraulic conductivity and porosity). Aquifer parameter uncertainty was accounted for by utilizing an ensemble surrogate model-based coupled multi-objective S/O model to develop reliable, optimal pumping strategies for the unconfined Bonriki aquifer system. Uncertainties in the numerical saltwater intrusion simulation model were characterised by developing several standalone SVMR surrogate models based on different combinations of uncertain hydraulic conductivity and porosity values. Therefore, instead of linking a standalone SVMR surrogate model, an ensemble of ten standalone SVMR models for each monitoring location were coupled to the MOGA optimisation model. The accuracy and reliability of the standalone SVMR models ensured that the ensemble models also retained reliable predictive capabilities. Execution of the ensemble surrogate-based multi-objective S/O model produced 600 optimal solutions in the form of a Pareto front. The 600 optimal solutions represented 600 pumping strategies that could be evaluated and implemented to ensure the sustainable management of the Bonriki aquifer system.

This thesis also demonstrated the application of an adaptive management framework, which utilizes sequential feedback information from a designed monitoring network, to the management of the Bonriki aquifer. In achieving the targeted adaptive management goal, an optimal strategy of pumping from production and barrier wells was determined by an S/O model. This is an option for the sustainable control of saltwater intrusion into the Bonriki aquifer. Using this prescribed optimal strategy, optimal monitoring wells were identified. A new monitoring objective function was developed to determine the optimal locations of monitoring wells in high-salinity areas. The resulting optimal monitoring wells were then used to monitor compliance with the prescribed management strategy (recommended by the S/O model) when compared to those actually implemented in the field and also considering uncertainty of aquifer parameters (hydraulic conductivity and porosity). Based on the field-level deviations between actual and planned salinity levels, the pumping rates for future time periods in the management horizon were modified according to a sequentially-updated coupled S/O model. It was noted that field-level deviations in implementing accurate pumping rates and uncertain aquifer parameters could lead to significant differences in the salinities measured at optimal monitoring wells. Hence, updating the management model using feedback information from earlier time periods is crucial to the management of the Bonriki aquifer.

A key feature of this study was its comparison of the performance of homogeneous and heterogeneous ensemble models. Specifically, ANN, GP, SVMR and GP standalone models were developed to construct homogeneous ensemble models (ANN_En, GP_En, SVMR_En and GPR_En) and a heterogeneous ensemble model (ANN_GP_SVMR_GPR_En) capable of predicting salinity concentration in the Bonriki aquifer. The results of this investigation demonstrate that all standalone models predicted salinity concentration at respective monitoring wells with reasonable accuracy. However, in terms of the four performance indicators considered, standalone GPR models displayed better prediction accuracy than the corresponding ANN, GP and SVMR standalone models. Also, the standalone SVMR and GPR models required significantly less time to train than the corresponding ANN and GP standalone models. The homogeneous GPR_En model was the best-performing ensemble model, even compared to the heterogeneous ANN_GP_SVMR_GPR_En model. Overall, the GPR_En model performed better than all four standalone models and the heterogeneous ensemble model. Hence, it can be used as a potentially powerful tool for predicting salinity in the Bonriki aquifer. In addition, with their accurate and efficient prediction capabilities, GPR_En models can also be employed as approximate simulators in simulation-optimization models used for developing regional-scale saltwater intrusion management strategies for coastal aquifers. The comparative investigation presented in this thesis provides a valuable reference for decision-makers and engineers who may choose to apply these methods for predicting groundwater salinity concentrations in coastal aquifers.

In this thesis, a new application of group method of data handling (GMDH) models was employed to predict salinity concentration at specified monitoring wells in a coastal aquifer in response to variable transient groundwater pumping patterns. The performance assessment results suggest that the developed models can predict salinity concentrations with reasonable accuracy and efficiency. Apart from exhibiting accurate and efficient prediction capabilities, the GMDH models provide a method for identifying the most influential pumping rates. The identification of the most and least influential pumping rates offers substantial benefits to the decision-maker. Specifically, it allows the decision-maker to determine the variables most influential to the management outcomes and prediction accuracy. These characteristics offer both computational and economic benefits. Second, the results of this study establish that the determination of influential variables when developing groundwater salinity predictive models should be given important consideration. Including the least influential variables in predictive models can diminish their accuracy.

Another important aspect of this thesis was its utilization of artificial freshwater recharge (AFR) to control saltwater intrusion into coastal aquifers. The findings of this study demonstrate that the integrated utilization of freshwater recharge and groundwater pumping can provide an efficient, practical solution to the current groundwater scarcity problems affecting people in coastal areas. The numerical modelling results quantified the effects of artificial recharge on heads and salinity levels at monitoring locations. A considerable decrease in salinity occurs when groundwater pumping is combined with artificial recharge. Injected freshwater maintains a seaward gradient in the system by raising the inland piezometric head. Following analysis of the developed 3D models' results, and quantification of changes in head and salinity at monitoring wells over time, it can be inferred that saltwater repulsion is increased by freshwater recharge. The simulation models' results indicate that freshwater lenses created by AFR can be a practical alternative considered for the efficient management of coastal aquifers subject to saltwater intrusion. After this quantification and detection of practical benefits, recharge wells and production wells were utilized to develop an SVMR surrogate model-assisted linked S/O model capable of specifying regional-scale coastal aquifer management strategies. The multi-objective optimization model derived a set of Pareto-optimal spatio-temporal pumping and artificial recharge rates, which can be implemented in the regional-scale optimal management of coastal aquifers. In general, the proposed methods of incorporating planned AFR in multi-objective groundwater management models can provide solutions to saltwater intrusion problems in coastal aquifers.

10.2 Recommendations

This study can make several recommendations based on some of its limitations. A few of these aspects are as follows. First, in the present field-scale application study, a two-layer aquifer system was modelled due to limited data availability and to ensure convergence of the 3D FEMWATER-based numerical model. The layers (constructed using limited borehole data) were considered heterogeneous vertically, based on the geological stratification of the layers. However, the materials in each layer were the same, albeit anisotropic. The proposed method can be applied to completely heterogeneous coastal aquifer systems in other geological settings. Secondly, the multi-objective management model developed in this study only considered two management goals denoted by the two different objective functions. Other management objectives, such as assigning pumping well locations, prescribing optimal operating costs, and incorporating recharge wells as management options can also be considered and applied to the Bonriki aquifer system for evaluation purposes. Thirdly, the implications of tidal fluctuations and seasonal variations on the movement of saltwater fronts could be investigated. These factors were not considered in the present study but could be investigated further and incorporated into the proposed management model. Implementing these factors presents significant challenges, as higher requirements for 3D modelling convergence tolerance, mesh tolerance, computing power, and hydrological data would be needed for numerical model calibration and validation.

The comparison of the homogeneous and heterogeneous models presented in this study showed promising results. However, the predictive capabilities of all standalone models could be further increased. This could be accomplished by using an optimal number of training and testing datasets. Also, recent developments in computational power can aid in the development of hybrid models that have increased predictive capability. In addition, more comparative studies using new modelling algorithms are recommended in line with the objectives of the current work. It is hoped that future research will focus on these directions, which may eventually lead to the establishment of a more robust, accurate, efficient and versatile salinity concentration prediction tool.

The results of this study suggest that adequately trained and tested GMDH models can be considered accurate and efficient saltwater intrusion prediction tools, which can be used for developing saltwater intrusion management models. GMDH models also allows decision-makers to comprehend the importance of input variables (e.g. pumping rates), which can be useful in developing efficient management plans for aquifer systems. In addition, the highly accurate and efficient GMDH predictive models can be used as approximate simulators in coupled simulation-optimization models to develop computationally practicable regional-scale coastal aquifer management methodologies. However, it is advisable that researchers first compare the performance of GMDH models with that of other well-established modelling techniques in a more rigorous manner.

This study has demonstrated that AFR can be used as a strategy to control saltwater intrusion in coastal aquifers. However, in future, assessing the practicality of this approach in a real case study would be beneficial. On the other hand, while this study has demonstrated the potential for utilizing artificial recharge to minimize seawater intrusion in coastal aquifers, it would be valuable to identify defined zones for installing recharge wells. Optimal zones would enhance the effects of artificial recharge and potentially provide a more efficient and possibly more cost-effective way to manage saltwater intrusion.

Lastly, the results presented in this study demonstrate pathways for future studies on other small island countries, where saltwater intrusion due to excessive groundwater withdrawal is a threat to sustainability of freshwater resources. The developed and evaluated predictive modelling tools, multi-objective management models, ensemble models incorporating aquifer parameter estimation uncertainty, and adaptive management methods can potentially be applied to other regional-scale coastal aquifers subject to saltwater intrusion. However, such applications require the development of a variable density groundwater flow and transport numerical model of the study area. Development of such a model necessitates numerical modelling skills, software, high computational power, and groundwater head and salinity data. Such datasets are not always readily available and may require field investigations, which can be costly and time-consuming.

References

Abarca, E., Carrera, J., Sánchez-Vila, X., and Voss, C. I. (2007). "Quasi-horizontal circulation cells in 3D seawater intrusion." *Journal of Hydrology*, 339(3), 118-129.

Abarca, E., Vázquez-Suñé, E., Carrera, J., Capino, B., Gámez, D., and Batlle, F. (2006). "Optimal design of measures to correct seawater intrusion." *Water resources research*, 42(9).

Abd-Elhamid, H., and Javadi, A. (2008). "An investigation into control of saltwater intrusion considering the effects of climate change and sea level rise." *Proceeding of 20th SWIM*, 2327.

Abd-Elhamid, H. F., and Javadi, A. A. (2011). "A cost-effective method to control seawater intrusion in coastal aquifers." *Water resources management*, 25(11), 2755-2780.

Aguirre, O., and Taboada, H. (2011). "A clustering method based on dynamic self organizing trees for post-pareto optimality analysis." *Procedia Computer Science*, 6, 195-200.

Akhtar, T., and Shoemaker, C. A. (2016). "Multi objective optimization of computationally expensive multi-modal functions with RBF surrogates and multi-rule selection." *Journal of Global Optimization*, 64(1), 17-32.

Aler, R., Borrajo, D., and Isasi, P. (2002). "Using genetic programming to learn and improve control knowledge." *Artificial Intelligence*, 141(1), 29-56.

Allen, D. M., Klassen, J., Van Pelt, S., Gurdak, J., and Taniguchi, M. (2014). "ASSESSMENT OF SIDS GROUNDWATER SYSTEMS."

Alley, W. M., Reilly, T. E., and Franke, O. L. (1999). "Sustainability of ground-water resources." US Department of the Interior, US Geological Survey, Denver, Colorado.

Allow, K. A. (2012). "The use of injection wells and a subsurface barrier in the prevention of seawater intrusion: a modelling approach." *Arabian Journal of Geosciences*, 5(5), 1151-1161.

Ammar, K., Khalil, A., McKee, M., and Kaluarachchi, J. (2008). "Bayesian deduction for redundancy detection in groundwater quality monitoring networks." *Water Resources Research*, 44(8).

Andrews, R. W. (1981). "Salt-Water Intrusion in the Costa de Hermosillo, Mexico: A Numerical Analysis of Water Management Proposals." *Ground Water*, 19(6), 635-647.

Angelaki, A., Singh Nain, S., Singh, V., and Sihag, P. (2018). "Estimation of models for cumulative infiltration of soil using machine learning methods." *ISH Journal of Hydraulic Engineering*, 1-8.

Anifowose, F. A., Labadin, J., and Abdulraheem, A. (2017). "Ensemble machine learning: An untapped modeling paradigm for petroleum reservoir characterization." *Journal of Petroleum Science and Engineering*, 151, 480-487.

Antonellini, M., Mollema, P., Giambastiani, B., Bishop, K., Caruso, L., Minchio, A., Pellegrini, L., Sabia, M., Ulazzi, E., and Gabbianelli, G. (2008). "Salt water intrusion in the coastal aquifer of the southern Po Plain, Italy." *Hydrogeology Journal*, 16(8), 1541-1556.

Arnell, N. W. (1999). "Climate change and global water resources." *Global environmental change*, 9, S31-S49.

Asefa, T., Kemblowski, M., Urroz, G., and McKee, M. (2005). "Support vector machines (SVMs) for monitoring network design." *Ground Water*, 43(3), 413-422.

Ataie-Ashtiani, B., and Ketabchi, H. (2011). "Elitist continuous ant colony optimization algorithm for optimal management of coastal aquifers." *Water resources management*, 25(1), 165-190.

Ataie-Ashtiani, B., Ketabchi, H., and Rajabi, M. M. (2013). "Optimal management of a freshwater lens in a small island using surrogate models and evolutionary algorithms." *Journal of Hydrologic Engineering*, 19(2), 339-354.
Aye, S., and Heyns, P. (2017). "An integrated Gaussian process regression for prediction of remaining useful life of slow speed bearings based on acoustic emission." *Mechanical Systems and Signal Processing*, 84, 485-498.
Ayers, J. F., and Vacher, H. (1986). "Hydrogeology of an atoll island: a conceptual model from detailed study of a Micronesian example." *Ground Water*, 24(2), 185-198.
Baalousha, H. (2010). "Assessment of a groundwater quality monitoring network using vulnerability mapping and geostatistics: a case study from Heretaunga Plains, New Zealand." *Agricultural Water Management*, 97(2), 240-246.
Bailey, R. T., Jenson, J., and Olsen, A. (2009). "Numerical modeling of atoll island hydrogeology." *Groundwater*, 47(2), 184-196.
Bailey, R. T., Jenson, J. W., and Olsen, A. E. (2010). "Estimating the ground water resources of atoll islands." *Water*, 2(1), 1-27.
Bandyopadhyay, S., and Maulik, U. (2002). "An evolutionary technique based on K-means algorithm for optimal clustering in RN." *Information Sciences*, 146(1), 221-237.
Banerjee, P., Singh, V., Chatttopadhyay, K., Chandra, P., and Singh, B. (2011). "Artificial neural network model as a potential alternative for groundwater salinity forecasting." *Journal of Hydrology*, 398(3), 212-220.
Barlow, P. M. (2003). *Ground Water in Fresh Water-salt Water Environments of the Atlantic*, Geological Survey (USGS).
Barlow, P. M., and Reichard, E. G. (2010). "Saltwater intrusion in coastal regions of North America." *Hydrogeology Journal*, 18(1), 247-260.
Barrett, B., Heinson, G., Hatch, M., and Telfer, A. (2002). "Geophysical methods in saline groundwater studies: locating perched water tables and fresh-water lenses." *Exploration Geophysics*, 33(2), 115-121.
Bartram, J., and Ballance, R. (1996). *Water quality monitoring: a practical guide to the design and implementation of freshwater quality studies and monitoring programmes*, CRC Press.
Bashi-Azghadi, S. N., and Kerachian, R. (2010). "Locating monitoring wells in groundwater systems using embedded optimization and simulation models." *Science of the Total Environment*, 408(10), 2189-2198.
Bear, J., Cheng, A. H.-D., Sorek, S., Ouazar, D., and Herrera, I. (1999). *Seawater intrusion in coastal aquifers: concepts, methods and practices*, Springer Science & Business Media.
Bear, J., and Verruijt, A. (2012). *Modeling groundwater flow and pollution*, Springer Science & Business Media.
Bechikh, S., Said, L. B., and Ghédira, K. (2011). "Searching for knee regions of the Pareto front using mobile reference points." *Soft Computing*, 15(9), 1807-1823.
Ben-Jemaa, F., Mariño, M. A., and Loaiciga, H. A. (1994). "Multivariate geostatistical design of ground-water monitoring networks." *Journal of Water Resources Planning and Management*, 120(4), 505-522.
Bendixson, V. M., Jenson, J. W., and Habana, N. C. (2014). "THE NORTHERN GUAM LENS AQUIFER DATABASE." University of Guam.
Benhachmi, M., Ouazar, D., Naji, A., Cheng, A., and El Harrouni, K. "Chance Constrained Optimal Management in Saltwater Intruded Coastal Aquifers using Genetic Algorithms." *Proc., Proc. of the 1st Inter. Conf. on Saltwater Intrusion and Coastal Aquifers-Monitoring, Modeling, and Management, Essaouira, Morocco*.
Benseddik, B., El Mrabet, E., El Mansouri, B., Chao, J., and Kili, M. (2017). "Delineation of artificial recharge zones in Mnasra Aquifer (NW, Morocco)." *Modeling Earth Systems and Environment*, 3(1), 10.
Berthe, L., Seng, D. C., and Asora, L. (2014). "Multiple stresses, veiled threat: Saltwater intrusion in Samoa."

Bhattacharjya, R. K., and Datta, B. (2005). "Optimal Management of Coastal Aquifers Using Linked Simulation Optimization Approach." *Water Resources Management*, 19(3), 295-320.

Bhattacharjya, R. K., and Datta, B. (2009). "ANN-GA-based model for multiple objective management of coastal aquifers." *Journal of Water Resources Planning and Management*, 135(5), 314-322.

Bhattacharjya, R. K., Datta, B., and Satish, M. G. (2007). "Artificial neural networks approximation of density dependent saltwater intrusion process in coastal aquifers." *Journal of Hydrologic Engineering*, 12(3), 273-282.

Binley, A., Buckley, K., Calore, C., Parodi, U., and BARBERA, P. L. (1997). "Modelling uncertainty in estimates of recharge to a shallow coastal aquifer." *Hydrological sciences journal*, 42(2), 155-168.

Bishop, C. M. (1995). *Neural networks for pattern recognition*, Oxford university press.

Bobba, A. (1993). "Mathematical models for saltwater intrusion in coastal aquifers." *Water Resources Management*, 7(1), 3-37.

Bosserelle, A., Jakovovic, D., Post, V., Rodriguez, S. G., Werner, A., and Sinclair, P. (2015). "Bonriki Inundation Vulnerability Assessment (BIVA): Assessment of sea-level rise and inundation effects on Bonriki Freshwater Lens, Tarawa Kiribati - Groundwater Modelling report." Secretariat of the Pacific Community (SPC)

Bouwer, H. (2002). "Artificial recharge of groundwater: hydrogeology and engineering." *Hydrogeology Journal*, 10(1), 121-142.

Branke, J., and Deb, K. (2005). "Integrating user preferences into evolutionary multi-objective optimization." *Knowledge incorporation in evolutionary computation*, Springer, 461-477.

Braun, M. T., and Oswald, F. L. (2011). "Exploratory regression analysis: A tool for selecting models and determining predictor importance." *Behavior research methods*, 43(2), 331-339.

Brunner, P., and Simmons, C. T. (2012). "HydroGeoSphere: a fully integrated, physically based hydrological model." *Ground Water*, 50(2), 170-176.

Bui, L. T., and Alam, S. (2008). "An introduction to multi-objective optimization." *Multi-Objective Optimization in Computational Intelligence: Theory and Practice*, 1-19.

Burns, W. C. (2002). "Pacific island developing country water resources and climate change." *The World's Water*, 2003, 113-132.

Cai, J., Taute, T., and Schneider, M. (2014). "Saltwater upconing below a pumping well in an inland aquifer: a theoretical modeling study on testing different scenarios of deep saline-groundwater pathways." *Water, Air, & Soil Pollution*, 225(11), 1-11.

Calvache, M., and Pulido-Bosch, A. (1997). "Effects of geology and human activity on the dynamics of salt-water intrusion in three coastal aquifers in southern Spain." *Environmental Geology*, 30(3-4), 215-223.

Carbajal, S. G. (2007). "Time Series Prediction Using Grammar-directed Genetic Programming Methods."

Carrera, J., Usunoff, E., and Szidarovszky, F. (1984). "A method for optimal observation network design for groundwater management." *Journal of Hydrology*, 73(1-2), 147-163.

Carruthers, P. (2009). "COOK ISLANDS COASTAL VULNERABILITY ASSESSMENTS: A SMALL PACIFIC ISLAND NATION'S EXPERIENCE."

Chadalavada, S., and Datta, B. (2008). "Dynamic optimal monitoring network design for transient transport of pollutants in groundwater aquifers." *Water Resources Management*, 22(6), 651-670.

Chan Hilton, A. B., and Culver, T. B. (2005). "Groundwater remediation design under uncertainty using genetic algorithms." *Journal of water resources planning and management*, 131(1), 25-34.

Chaudhari, P., Dharaskar, R., and Thakare, V. (2010). "Computing the most significant solution from Pareto front obtained in multi-objective evolutionary." *International Journal of Advanced Computer Science and Applications*, 1(4), 63-68.
Chebrolu, S., Abraham, A., and Thomas, J. P. (2005). "Feature deduction and ensemble design of intrusion detection systems." *Computers & security*, 24(4), 295-307.
Cheikh, M., Jarboui, B., Loukil, T., and Siarry, P. (2010). "A method for selecting Pareto optimal solutions in multiobjective optimization." *Journal of Informatics and Mathematical Sciences*, 2(1), 51-62.
Chen, Z., Huang, G., and Chakma, A. (2003). "Hybrid fuzzy-stochastic modeling approach for assessing environmental risks at contaminated groundwater systems." *Journal of Environmental Engineering*, 129(1), 79-88.
Cheng, A., Benhachmi, M., Halhal, D., Ouazar, D., Naji, A., and El Harrouni, K. (2004). *Pumping optimization in saltwater-intruded aquifers*, Lewis Publishers.
Cheng, A. H., and Ouazar, D. (2016). *Coastal aquifer management-monitoring, modeling, and case studies*, CRC Press.
Cheng, J. M., and Chen, C. X. (2001). "Three-Dimensional Modeling of Density-Dependent Salt Water Intrusion in Multilayered Coastal Aquifers in Jahe River Baisn, Shandong Province, China." *Ground water*, 39(1), 137-143.
Chevalier, R. F., Hoogenboom, G., McClendon, R. W., and Paz, J. A. (2011). "Support vector regression with reduced training sets for air temperature prediction: a comparison with artificial neural networks." *Neural Computing and Applications*, 20(1), 151-159.
Choquet, C., Diédhiou, M. M., and Rosier, C. (2016). "Derivation of a Sharp-Diffuse Interfaces Model for Seawater Intrusion in a Free Aquifer. Numerical Simulations." *SIAM Journal on Applied Mathematics*, 76(1), 138-158.
Christelis, V., and Mantoglou, A. (2016). "Pumping Optimization of Coastal Aquifers Assisted by Adaptive Metamodelling Methods and Radial Basis Functions." *Water Resources Management*, 1-15.
Clark, R., Gonzalez, D., Dillon, P., Charles, S., Cresswell, D., and Naumann, B. (2015). "Reliability of water supply from stormwater harvesting and managed aquifer recharge with a brackish aquifer in an urbanising catchment and changing climate." *Environmental Modelling & Software*, 72, 117-125.
Clauser, C. (2012). *Numerical simulation of reactive flow in hot aquifers: SHEMAT and processing SHEMAT*, Springer Science & Business Media.
Coello, C. A. C., Lamont, G. B., and Van Veldhuizen, D. A. (2007). *Evolutionary algorithms for solving multi-objective problems*, Springer.
Connor, R. (2015). *The United Nations world water development report 2015: water for a sustainable world*, UNESCO Publishing.
Corzo, G., Solomatine, D., Wit, M. d., Werner, M., Uhlenbrook, S., and Price, R. (2009). "Combining semi-distributed process-based and data-driven models in flow simulation: a case study of the Meuse river basin." *Hydrology and Earth System Sciences*, 13(9), 1619-1634.
Crevillén-García, D., Wilkinson, R., Shah, A., and Power, H. (2017). "Gaussian process modelling for uncertainty quantification in convectively-enhanced dissolution processes in porous media." *Advances in Water Resources*, 99, 1-14.
Cui, X., Zhu, P., Yang, X., Li, K., and Ji, C. (2014). "Optimized big data K-means clustering using MapReduce." *The Journal of Supercomputing*, 70(3), 1249-1259.
Custodio, E. (1997). "Studying, monitoring and controlling seawater intrusion in coastal aquifers." *Guidelines for Study, Monitoring and Control, FAO Water Reports*, 11, 7-23.
Custodio, E. (2010). "Coastal aquifers of Europe: an overview." *Hydrogeology Journal*, 18(1), 269-280.
Dagan, G., and Bear, J. (1968). "Solving the problem of local interface upconing in a coastal aquifer by the method of small perturbations." *Journal of Hydraulic Research*, 6(1), 15-44.

Das, A., and Datta, B. (1999). "Development of multiobjective management models for coastal aquifers." *Journal of Water Resources Planning and Management*, 125(2), 76-87.

Datta, B., Chakrabarty, D., and Dhar, A. (2009). "Optimal dynamic monitoring network design and identification of unknown groundwater pollution sources." *Water Resources Management*, 23(10), 2031-2049.

Datta, B., and Peralta, R. C. (1986). "Interactive computer graphics-based multiobjective decision-making for regional groundwater management." *Agricultural Water Management*, 11(2), 91-116.

Datta, B., Prakash, O., and Sreekanth, J. (2014). "Application of genetic programming models incorporated in optimization models for contaminated groundwater systems management." *EVOLVE-A Bridge between Probability, Set Oriented Numerics, and Evolutionary Computation V*, Springer, 183-199.

Datta, B., Vennalakanti, H., and Dhar, A. (2009). "Modeling and control of saltwater intrusion in a coastal aquifer of Andhra Pradesh, India." *Journal of hydro-environment Research*, 3(3), 148-159.

Dawe, P. (2001). *Review of the Rotuma Water Supply and Distribution System, Fiji Islands*, SOPAC Secretariat.

Dawoud, M. A. (2004). "Design of national groundwater quality monitoring network in Egypt." *Environmental monitoring and assessment*, 96(1-3), 99-118.

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

Deb, K., Pratap, A., Agarwal, S., and Meyarivan, T. (2002). "A fast and elitist multiobjective genetic algorithm: NSGA-II." *IEEE transactions on evolutionary computation*, 6(2), 182-197.

Destandau, F., and Zaiter, Y. "Optimal spatio-temporal design for water quality monitoring networks in maximizing economic value of information." *Proc., MATEC Web of Conferences*, EDP Sciences, 03004.

Dhar, A., and Datta, B. (2007). "Multiobjective design of dynamic monitoring networks for detection of groundwater pollution." *Journal of water resources planning and management*, 133(4), 329-338.

Dhar, A., and Datta, B. (2009). "Logic-based design of groundwater monitoring network for redundancy reduction." *Journal of water resources planning and management*, 136(1), 88-94.

Dhar, A., and Datta, B. (2009). "Saltwater intrusion management of coastal aquifers. I: Linked simulation-optimization." *Journal of Hydrologic Engineering*, 14(12), 1263-1272.

Dhar, A., and Datta, B. (2009). "Saltwater intrusion management of coastal aquifers. II: Operation uncertainty and monitoring." *Journal of Hydrologic Engineering*, 14(12), 1273-1282.

Diersch, H.-J., and Kolditz, O. (2002). "Variable-density flow and transport in porous media: approaches and challenges." *Advances in Water Resources*, 25(8), 899-944.

Dokou, Z., and Karatzas, G. P. (2012). "Saltwater intrusion estimation in a karstified coastal system using density-dependent modelling and comparison with the sharp-interface approach." *Hydrological Sciences Journal*, 57(5), 985-999.

Don, N. C., Araki, H., Yamanishi, H., and Koga, K. (2005). "Simulation of groundwater flow and environmental effects resulting from pumping." *Environmental Geology*, 47(3), 361-374.

Duan, Q., Ajami, N. K., Gao, X., and Sorooshian, S. (2007). "Multi-model ensemble hydrologic prediction using Bayesian model averaging." *Advances in Water Resources*, 30(5), 1371-1386.

Eason, J., and Cremaschi, S. (2014). "Adaptive sequential sampling for surrogate model generation with artificial neural networks." *Computers & Chemical Engineering*, 68, 220-232.

Ellison, J., and Fiu, M. (2010). "Vulnerability of Fiji's mangroves and associated coral reefs to climate change." *World Wildlife Fund*.
Emch, P. G., and Yeh, W. W. (1998). "Management model for conjunctive use of coastal surface water and ground water." *Journal of water resources planning and management*, 124(3), 129-139.
Esfahani, H. K., and Datta, B. (2015). "Use of Genetic Programming Based Surrogate Models to Simulate Complex Geochemical Transport Processes in Contaminated Mine Sites." *Handbook of Genetic Programming Applications*, Springer, 359-379.
Essaid, H. (1986). "A comparison of the coupled fresh water-salt water flow and the Ghyben-Herzberg sharp interface approaches to modeling of transient behavior in coastal aquifer systems." *Journal of Hydrology*, 86(1-2), 169-193.
Essaid, H. I. (1990). *The computer model SHARP, a quasi-three-dimensional finite-difference model to simulate freshwater and saltwater flow in layered coastal aquifer systems*, Department of the Interior, US Geological Survey.
Essawy, B. T. A. (2013). "Effect of Climate Change on Sea Water Intrusion in the Nile Delta Coastal Aquifer." Faculty of Engineering at Cairo University In Partial Fulfillment of the Requirements for the Degree of MASTER OF SCIENCE In Irrigation and Hydraulics Faculty of Engineering, Cairo University Giza, Egypt.
Essink, G. (2003). "Mathematical models and their application to salt water intrusion problems." *TIAC*, 3, 57-77.
Falkland, A. (1992). "Review of Tarawa freshwater lenses, Republic of Kiribati." *Australian International Development Assistance Bureau. Report No. HWR*, 92, 682.
Farag, A., and Mohamed, R. M. (2004). "Regression using support vector machines: basic foundations."
Farlow, S. J. (1984). *Self-organizing methods in modeling: GMDH type algorithms*, CrC Press.
Ferguson, G., and Gleeson, T. (2012). "Vulnerability of coastal aquifers to groundwater use and climate change." *Nature Climate Change*, 2(5), 342-345.
Fernández, F. H., and Lozano, F. H. "GMDH algorithm implemented in the intelligent identification of a bioprocess." *Proc., ABCM Symposium Series in Mechatronics.–2010.–4.–P*, 278-287.
Fletcher, C. H., and Richmond, B. M. (2010). "Climate change in the Federated States of Micronesia: Food and water security, climate risk management, and adaptive strategies." University of Hawai 'i Sea Grant College Program.
Forrester, A., Sobester, A., and Keane, A. (2008). *Engineering design via surrogate modelling: a practical guide*, John Wiley & Sons. USA
Foster, J. A. (2001). "Review: Discipulus: A commercial genetic programming system." *Genetic Programming and Evolvable Machines*, 2(2), 201-203.
Foster, S., and Chilton, P. (2003). "Groundwater: the processes and global significance of aquifer degradation." *Philosophical Transactions of the Royal Society B: Biological Sciences*, 358(1440), 1957-1972.
Freeze, R. A., and Cherry, J. A. (1979). "Groundwater."
Friedman, J., Hastie, T., and Tibshirani, R. (2001). *The elements of statistical learning*, Springer series in statistics New York.
Friedman, J. H. (1991). "Multivariate adaptive regression splines." *The annals of statistics*, 1-67.
Füssel, H.-M. (2012). "Vulnerability of Coastal Populations." *Climate Change, Justice and Sustainability*, Springer, 45-57.
Gad, M., and Khalaf, S. (2013). "Application of sharing genetic algorithm for optimization of groundwater management problems in Wadi El-Farigh, Egypt." *Applied Water Science*, 3(4), 701-716.
Geiger, M. J. (2006). "Solving multi-objective scheduling problems-An integrated systems approach." *INTERNATIONAL FEDERATION FOR INFORMATION PROCESSING-PUBLICATIONS-IFIP*, 217, 493.

Gelhar, L. W. (1993). *Stochastic subsurface hydrology*, Prentice-Hall.
Ghassemi, F., Jakeman, A., and Jacobson, G. (1990). "Mathematical modelling of sea water intrusion, Nauru Island." *Hydrological Processes*, 4(3), 269-281.
Ghassemi, F., Jakeman, A., Jacobson, G., and Howard, K. (1996). "Simulation of seawater intrusion with 2D and 3D models: Nauru Island case study." *Hydrogeology Journal*, 4(3), 4-22.
Godinez, I. G., and Darnault, C. J. (2008). "Coastal aquifers and saltwater intrusion." *Overexploitation and Contamination of Shared Groundwater Resources*, Springer, 185-201.
Goel, T., Haftka, R. T., Shyy, W., and Queipo, N. V. (2007). "Ensemble of surrogates." *Structural and Multidisciplinary Optimization*, 33(3), 199-216.
Gong, M., Li, H., and Jiang, X. (2016). "A multi-objective optimization framework for ill-posed inverse problems." *CAAI Transactions on Intelligence Technology*, 1(3), 225-240.
Gorelick, S. M., Voss, C. I., Gill, P. E., Murray, W., Saunders, M. A., and Wright, M. H. (1984). "Aquifer reclamation design: The use of contaminant transport simulation combined with nonlinear programing." *Water Resources Research*, 20(4), 415-427.
Grbić, R., Kurtagić, D., and Slišković, D. (2013). "Stream water temperature prediction based on Gaussian process regression." *Expert systems with applications*, 40(18), 7407-7414.
Gretton, A., Doucet, A., Herbrich, R., Rayner, P. J., and Scholkopf, B. "Support vector regression for black-box system identification." *Proc., Statistical Signal Processing, 2001. Proceedings of the 11th IEEE Signal Processing Workshop on*, IEEE, 341-344.
Grundmann, J., Schütze, N., and Lennartz, F. (2012). "Sustainable management of a coupled groundwater–agriculture hydrosystem using multi-criteria simulation based optimisation." *Water Science and Technology*, 67(3), 689-698.
Grundmann, J., Schütze, N., Schmitz, G. H., and Al-Shaqsi, S. (2012). "Towards an integrated arid zone water management using simulation-based optimisation." *Environmental Earth Sciences*, 65(5), 1381-1394.
Guo, W., and Langevin, C. D. (2002). "User's guide to SEAWAT; a computer program for simulation of three-dimensional variable-density ground-water flow."
Guttman, J., Negev, I., and Rubin, G. (2017). "Design and Testing of Recharge Wells in a Coastal Aquifer: Summary of Field Scale Pilot Tests." *Water*, 9(1), 53.
Hanjra, M. A., and Qureshi, M. E. (2010). "Global water crisis and future food security in an era of climate change." *Food Policy*, 35(5), 365-377.
Hashemi, H., Berndtsson, R., Kompani-Zare, M., and Persson, M. (2013). "Natural vs. artificial groundwater recharge, quantification through inverse modeling." *Hydrology and Earth System Sciences*, 17(2), 637-650.
Hay, J., and Kaluwln, C. "ASSESSMENT OF THE VULNERABILITY TO SEA LEVEL RISE FOR THE KINGDOM OF TONGA: SUMMARY OF THE TONGAN CASE STUDY." *Proc., Second SPREP Meeting on Climate Change and Sea Level Rise in the South Pacific Region: Noumea, New Caledonia, 6-10 April 1992: proceedings*, South Pacific Regional Environment Programme, 186.
He, Z., Wen, X., Liu, H., and Du, J. (2014). "A comparative study of artificial neural network, adaptive neuro fuzzy inference system and support vector machine for forecasting river flow in the semiarid mountain region." *Journal of Hydrology*, 509, 379-386.
Hillstrom, K., and Hillstrom, L. C. (2003). "Australia, Oceania, and Antarctica."
Hiscock, K. M. (2011). "Groundwater in the 21st century–meeting the challenges." *Sustaining Groundwater Resources*, Springer, 207-225.
Hong, S.-H., Song, S.-H., Bae, S. K., and Park, N. (2004). "Verification and validation of an optimization model for groundwater development in coastal areas." *Proceeding of 18th SWIM*, 77-90.

Hsu, S.-K. (1998). "Plan for a groundwater monitoring network in Taiwan." *Hydrogeology Journal*, 6(3), 405-415.
Huang, C., and Mayer, A. S. (1997). "Pump-and-treat optimization using well locations and pumping rates as decision variables." *Water Resources Research*, 33(5), 1001-1012.
Hudak, P. F., and Loaiciga, H. A. (1992). "A location modeling approach for groundwater monitoring network augmentation." *Water Resources Research*, 28(3), 643-649.
Hudak, P. F., and Loaiciga, H. A. (1993). "An optimization method for monitoring network design in multilayered groundwater flow systems." *water resources research*, 29(8), 2835-2845.
Hussain, M., Javadi, A., and Sherif, M. (2016). "Artificial recharge of coastal aquifers using treated wastewater to control saltwater intrusion." *Proc., 24th UK Conference of the Association for Computational Mechanics in Engineering*, 4.
Hussain, M. S., Javadi, A. A., Ahangar-Asr, A., and Farmani, R. (2015). "A surrogate model for simulation–optimization of aquifer systems subjected to seawater intrusion." *Journal of Hydrology*, 523, 542-554.
Huyakorn, P. S., Andersen, P. F., Mercer, J. W., and White, H. O. (1987). "Saltwater intrusion in aquifers: Development and testing of a three-dimensional finite element model." *Water Resources Research*, 23(2), 293-312.
Insigne, M. S. L., and Kim, G.-S. (2010). "Saltwater intrusion modeling in the aquifer bounded by Manila Bay and Parañaque River, Philippines." *Environmental Engineering Research*, 15(2), 117-121.
Jafari, S., Khalaf, P., and Montazeri-Gh, M. (2014). "Multi-Objective Meta Heuristic Optimization Algorithm with Multi Criteria Decision Making Strategy for Aero-Engine Controller Design." *International Journal of Aerospace Sciences*, 3(1), 6-17.
Jakovovic, D., Werner, A. D., de Louw, P. G., Post, V. E., and Morgan, L. K. (2016). "Saltwater upconing zone of influence." *Advances in Water Resources*, 94, 75-86.
Jakovovic, D., Werner, A. D., and Simmons, C. T. (2011). "Numerical modelling of saltwater up-coning: comparison with experimental laboratory observations." *Journal of hydrology*, 402(3), 261-273.
Javadi, A., Hussain, M., Sherif, M., and Farmani, R. (2015). "Multi-objective optimization of different management scenarios to control seawater intrusion in coastal aquifers." *Water Resources Management*, 29(6), 1843-1857.
Jha, M. K., and Datta, B. (2014). "Linked simulation-optimization based dedicated monitoring network design for unknown pollutant source identification using dynamic time warping distance." *Water Resources Management*, 28(12), 4161-4182.
Jin, Y., Okabe, T., and Sendhoff, B. "Neural network regularization and ensembling using multi-objective evolutionary algorithms." *Proc., Evolutionary Computation, 2004. CEC2004. Congress on*, IEEE, 1-8.
Johnson, V. M., and Rogers, L. L. (2000). "Accuracy of neural network approximators in simulation-optimization." *Journal of Water Resources Planning and Management*, 126(2), 48-56.
Juang, C. H., Wang, L., Hsieh, H.-S., and Atamturktur, S. (2014). "Robust geotechnical design of braced excavations in clays." *Structural Safety*, 49, 37-44.
Kacimov, A., and Obnosov, Y. V. "Analytical solution for a sharp interface problem in sea water intrusion into a coastal aquifer." *Proc., Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, The Royal Society, 3023-3038.
Kacimov, A., and Sherif, M. (2006). "Sharp interface, one-dimensional seawater intrusion into a confined aquifer with controlled pumping: Analytical solution." *Water resources research*, 42(6).

Kallioras, A., Pliakas, F.-K., Schuth, C., and Rausch, R. (2013). "Methods to countermeasure the intrusion of seawater into coastal aquifer systems." *Wastewater Reuse and Management*, Springer, 479-490.

Kao, C. (2010). "Weight determination for consistently ranking alternatives in multiple criteria decision analysis." *Applied Mathematical Modelling*, 34(7), 1779-1787.

Karatzas, G. P., and Dokou, Z. (2015). "Optimal management of saltwater intrusion in the coastal aquifer of Malia, Crete (Greece), using particle swarm optimization." *Hydrogeology Journal*, 23(6), 1181-1194.

Karbasi, M. (2017). "Forecasting of Multi-Step Ahead Reference Evapotranspiration Using Wavelet-Gaussian Process Regression Model." *Water Resources Management*, 1-18.

Kattan, A., and Ong, Y.-S. (2015). "Surrogate Genetic Programming: A semantic aware evolutionary search." *Information Sciences*, 296, 345-359.

Kelman, I., and West, J. J. (2009). "Climate change and small island developing states: a critical review." *Ecological and Environmental Anthropology*, 5(1), 1-16.

Ketabchi, H., and Ataie-Ashtiani, B. (2015). "Evolutionary algorithms for the optimal management of coastal groundwater: a comparative study toward future challenges." *Journal of Hydrology*, 520, 193-213.

Ketabchi, H., and Ataie-Ashtiani, B. (2015). "Review: Coastal groundwater optimization—advances, challenges, and practical solutions." *Hydrogeology Journal*, 23(6), 1129-1154.

Khadri, S., and Pande, C. (2016). "Ground water flow modeling for calibrating steady state using MODFLOW software: a case study of Mahesh River basin, India." *Modeling Earth Systems and Environment*, 2(1), 39.

Kim, S.-D., Lee, H.-J., and Park, J.-S. (2012). "Simulation of seawater intrusion range in coastal aquifer using the FEMWATER model for disaster information." *Marine Georesources & Geotechnology*, 30(3), 210-221.

Kipp, K. L. (1987). "HST3D; a computer code for simulation of heat and solute transport in three-dimensional ground-water flow systems."

Kodinariya, T. M., and Makwana, P. R. (2013). "Review on determining number of Cluster in K-Means Clustering." *International Journal*, 1(6), 90-95.

Koh, B. R., Oh, Y. H., and Park, D. I. (2016). "A study on formation of saltwater-freshwater interface and upconing in coastal aquifers based on experiments with hydraulic models."

Kohavi, R. "A study of cross-validation and bootstrap for accuracy estimation and model selection." *Proc., Ijcai*, Montreal, Canada, 1137-1145.

Kollat, J. B., and Reed, P. M. (2006). "Comparing state-of-the-art evolutionary multi-objective algorithms for long-term groundwater monitoring design." *Advances in Water Resources*, 29(6), 792-807.

Kong, D., Chen, Y., and Li, N. (2018). "Gaussian process regression for tool wear prediction." *Mechanical Systems and Signal Processing*, 104, 556-574.

Konikow, L. F. (2011). "The Secret to Successful Solute-Transport Modeling." *Groundwater*, 49(2), 144-159.

Kordjazi, A., Pooya Nejad, F., and Jaksa, M. (2015). "Prediction of load-carrying capacity of piles using a support vector machine and improved data collection." *Proc., 12th Australia New Zealand Conference on Geomechanics*, The New Zealand Geotechnical Society and the Australian Geomechanics Society, New Zealand, 1-8.

Kourakos, G., and Mantoglou, A. (2009). "Pumping optimization of coastal aquifers based on evolutionary algorithms and surrogate modular neural network models." *Advances in water resources*, 32(4), 507-521.

Kourakos, G., and Mantoglou, A. (2011). "Simulation and multi-objective management of coastal aquifers in semi-arid regions." *Water resources management*, 25(4), 1063-1074.

Kourakos, G., and Mantoglou, A. (2013). "Development of a multi-objective optimization algorithm using surrogate models for coastal aquifer management." *Journal of Hydrology*, 479, 13-23.

Koussis, A., Georgopoulou, E., Kotronarou, A., Lalas, D., Restrepo, P., Destouni, G., Prieto, C., Rodriguez, J., Rodriguez-Mirasol, J., and Cordero, T. (2010). "Cost-efficient management of coastal aquifers via recharge with treated wastewater and desalination of brackish groundwater: general framework." *Hydrological Sciences Journal–Journal des Sciences Hydrologiques*, 55(7), 1217-1233.

Koussis, A., Kotronarou, A., Destouni, G., and Prieto, C. (2003). "Intensive groundwater development in coastal zones and small islands." *Intensive Use of Groundwater: Challenges and Opportunities*, 90(5809), 390.

Koza, J. R., Mydlowec, W., Lanza, G., Yu, J., and Keane, M. A. (2000). "Reverse engineering and automatic synthesis of metabolic pathways from observed data using genetic programming." *Biomedical Informatics*, 1-12.

Kumar, S., Sondhi, S., and Phogat, V. (2005). "Network design for groundwater level monitoring in upper Bari Doab canal tract, Punjab, India." *Irrigation and Drainage: The journal of the International Commission on Irrigation and Drainage*, 54(4), 431-442.

Kura, N. U., Ramli, M. F., Ibrahim, S., Sulaiman, W. N. A., Zaudi, M. A., and Aris, A. Z. (2014). "A preliminary appraisal of the effect of pumping on seawater intrusion and upconing in a small tropical island using 2D resistivity technique." *The Scientific World Journal*, 2014.

Lal, A., and Datta, B. (2018). "Development and implementation of support vector machine regression surrogate models for predicting groundwater pumping-induced saltwater intrusion into coastal aquifers." *Water resources management*, 32(7), 2405-2419.

Lal, A., and Datta, B. (2019). "Multi-objective groundwater management strategy under uncertainties for sustainable control of saltwater intrusion: Solution for an island country in the South Pacific." *Journal of environmental management*, 234, 115-130.

Lee, C. H., and Cheng, R. T. S. (1974). "On seawater encroachment in coastal aquifers." *Water Resources Research*, 10(5), 1039-1043.

Lee, J.-Y., and Song, S.-H. (2007). "Evaluation of groundwater quality in coastal areas: implications for sustainable agriculture." *Environmental Geology*, 52(7), 1231-1242.

Lee, Y.-M., and Ellis, J. H. (1996). "Comparison of algorithms for nonlinear integer optimization: application to monitoring network design." *Journal of Environmental Engineering*, 122(6), 524-531.

Lew, T., Spencer, A., Scarpa, F., Worden, K., Rutherford, A., and Hemez, F. (2006). "Identification of response surface models using genetic programming." *Mechanical Systems and Signal Processing*, 20(8), 1819-1831.

Li, Y., and Hilton, A. B. C. (2005). "Reducing spatial sampling in long-term groundwater monitoring networks using ant colony optimization." *International Journal of Computational Intelligence Research*, 1(1), 9-28.

Li, Y., and Hilton, A. B. C. (2007). "Optimal groundwater monitoring design using an ant colony optimization paradigm." *Environmental Modelling & Software*, 22(1), 110-116.

Lim, D., Ong, Y.-S., Jin, Y., and Sendhoff, B. "A study on metamodeling techniques, ensembles, and multi-surrogates in evolutionary computation." *Proc., Proceedings of the 9th annual conference on Genetic and evolutionary computation*, ACM, 1288-1295.

Lin, H.-C. J., Richards, D. R., Yeh, G.-T., Cheng, J.-R., and Cheng, H.-P. (1997). "FEMWATER: A Three-Dimensional Finite Element Computer Model for Simulating Density-Dependent Flow and Transport in Variably Saturated Media." DTIC Document, Vicksburg, MS 39180-6199.

Lin, R. S., Horn, S. D., Hurdle, J. F., and Goldfarb-Rumyantzev, A. S. (2008). "Single and multiple time-point prediction models in kidney transplant outcomes." *Journal of biomedical informatics*, 41(6), 944-952.

Liong, S.-Y., Lim, W.-H., and Paudyal, G. N. (2000). "River stage forecasting in Bangladesh: neural network approach." *Journal of computing in civil engineering*, 14(1), 1-8.

Liong, S., Tariq, A., and Lee, K. (2004). "Application of evolutionary algorithm in reservoir operations." *Journal of the Institution of Engineers, Singapore*, 44(1), 39-53.

Liu, J., and Zio, E. (2016). "An adaptive online learning approach for Support Vector Regression: Online-SVR-FID." *Mechanical Systems and Signal Processing*, 76, 796-809.

Liu, L., Gao, C., Xuan, W., and Xu, Y.-P. (2017). "Evaluation of medium-range ensemble flood forecasting based on calibration strategies and ensemble methods in Lanjiang Basin, Southeast China." *Journal of Hydrology*, 554, 233-250.

Liu, Q., Chen, C., Zhang, Y., and Hu, Z. (2011). "Feature selection for support vector machines with RBF kernel." *Artificial Intelligence Review*, 36(2), 99-115.

Liu, W., Dou, Z., Wang, W., Liu, Y., Zou, H., Zhang, B., and Hou, S. (2018). "Short-term load forecasting based on elastic net improved GMDH and difference degree weighting optimization." *Applied Sciences*, 8(9), 1603.

Llopis-Albert, C., and Pulido-Velazquez, D. (2014). "Discussion about the validity of sharp-interface models to deal with seawater intrusion in coastal aquifers." *Hydrological Processes*, 28(10), 3642-3654.

Loaiciga, H. A. (1989). "An optimization approach for groundwater quality monitoring network design." *Water Resources Research*, 25(8), 1771-1782.

Loaiciga, H. A., Charbeneau, R. J., Everett, L. G., Fogg, G. E., Hobbs, B. F., and Rouhani, S. (1992). "Review of ground-water quality monitoring network design." *Journal of Hydraulic Engineering*, 118(1), 11-37.

Loh, W.-L. (1996). "On Latin hypercube sampling." *The annals of statistics*, 24(5), 2058-2080.

Luo, J., and Lu, W. (2014). "Comparison of surrogate models with different methods in groundwater remediation process." *Journal of Earth System Science*, 123(7), 1579-1589.

Luo, Q., Wu, J., Yang, Y., Qian, J., and Wu, J. (2014). "Optimal design of groundwater remediation system using a probabilistic multi-objective fast harmony search algorithm under uncertainty." *Journal of hydrology*, 519, 3305-3315.

Luyun, R., Momii, K., and Nakagawa, K. (2011). "Effects of recharge wells and flow barriers on seawater intrusion." *Ground Water*, 49(2), 239-249.

Luyun, R., Momii, K., and Nakagawa, K. (2011). "Effects of recharge wells and flow barriers on seawater intrusion." *Groundwater*, 49(2), 239-249.

MacQueen, J. "Some methods for classification and analysis of multivariate observations." *Proc., Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, Oakland, CA, USA, 281-297.

Mahar, P. S., and Datta, B. (1997). "Optimal monitoring network and ground-water–pollution source identification." *Journal of water resources planning and management*, 123(4), 199-207.

Mantoglou, A. (2003). "Pumping management of coastal aquifers using analytical models of saltwater intrusion." *Water resources research*, 39(12).

Martin, J. D., and Simpson, T. W. (2005). "Use of kriging models to approximate deterministic computer models." *AIAA journal*, 43(4), 853-863.

Masoumi, F., and Kerachian, R. (2010). "Optimal redesign of groundwater quality monitoring networks: a case study." *Environmental monitoring and assessment*, 161(1-4), 247-257.

Mattson, C. A., Mullur, A. A., and Messac, A. (2004). "Smart Pareto filter: Obtaining a minimal representation of multiobjective design space." *Engineering Optimization*, 36(6), 721-740.

Mehr, A. D., and Kahya, E. (2017). "A Pareto-optimal moving average multigene genetic programming model for daily streamflow prediction." *Journal of hydrology*, 549, 603-615.

Metai, E. "Vulnerability of freshwater lens on Tarawa–the role of hydrological monitoring in determining sustainable yield." *Proc., Case study presented as part of Theme*.

Metutera, T. (2002). "Water management in Kiribati with special emphasis on groundwater development using infiltration galleries." *Proceedings of the Pacific Regional Consultation on Water in Small Island Countries, Sigatoka, Fiji Islands*, 29.

Meyer, P. D., and Brill, E. D. (1988). "A method for locating wells in a groundwater monitoring network under conditions of uncertainty." *Water Resources Research*, 24(8), 1277-1282.

Meyer, P. D., Valocchi, A. J., and Eheart, J. W. (1994). "Monitoring network design to provide initial detection of groundwater contamination." *Water Resources Research*, 30(9), 2647-2659.

Mogheir, Y., De Lima, J., and Singh, V. (2009). "Entropy and multi-objective based approach for groundwater quality monitoring network assessment and redesign." *Water resources management*, 23(8), 1603-1620.

Mogheir, Y., and Singh, V. (2002). "Application of information theory to groundwater quality monitoring networks." *Water Resources Management*, 16(1), 37-49.

Mohammadi, K., Shamshirband, S., Petković, D., Yee, L., and Mansor, Z. (2016). "Using ANFIS for selection of more relevant parameters to predict dew point temperature." *Applied Thermal Engineering*, 96, 311-319.

Mohsen, M. S., Singh, V. P., and Amer, A. M. (1990). "A note on saltwater intrusion in coastal aquifers." *Water Resources Management*, 4(2), 123-134.

Moriasi, D. N., Arnold, J. G., Van Liew, M. W., Bingner, R. L., Harmel, R. D., and Veith, T. L. (2007). "Model evaluation guidelines for systematic quantification of accuracy in watershed simulations." *Transactions of the ASABE*, 50(3), 885-900.

Morris, B. L., Lawrence, A. R., Chilton, P., Adams, B., Calow, R. C., and Klinck, B. A. (2003). "Groundwater and its susceptibility to degradation: a global assessment of the problem and options for management."

Mosley, L., and Carpenter, C. (2005). *Niue coastal water quality and groundwater resources assessment*, SOPAC, Fiji.

Mourits, L. J. (1996). *An assessment of saltwater intrusion in babai pits and some water supply projects on Makin, Butaritari and Abaiang, Republic of Kiribati*, Citeseer.

Mugunthan, P., and Shoemaker, C. A. (2004). "Time varying optimization for monitoring multiple contaminants under uncertain hydrogeology." *Bioremediation Journal*, 8(3-4), 129-146.

Muttil, N., and Lee, J. H. (2005). "Genetic programming for analysis and real-time prediction of coastal algal blooms." *Ecological modelling*, 189(3), 363-376.

Narayan, K., Schleeberger, C., Charlesworth, P., and Bistrow, K. "Effects of groundwater pumping on saltwater intrusion in the lower Burdekin Delta, North Queensland." *Proc., MODSIM 2003 International Congress on modelling and simulation*, 212-217.

Nash, J. E., and Sutcliffe, J. V. (1970). "River flow forecasting through conceptual models part I—A discussion of principles." *Journal of hydrology*, 10(3), 282-290.

Nazeer, K. A., and Sebastian, M. "Improving the Accuracy and Efficiency of the k-means Clustering Algorithm." *Proc., Proceedings of the world congress on engineering*, 1-3.

Newport, B. D. (1977). *Salt water intrusion in the United States*, Environmental Protection Agency, Office of Research and Development, Robert S. Kerr Environmental Research Laboratory.

Nguyen-Tuong, D., Seeger, M., and Peters, J. (2009). "Model learning with local gaussian process regression." *Advanced Robotics*, 23(15), 2015-2034.
Nguyen, B. T. (2016). "Coupling geophysical and isotopic approaches to better simulate saltwater intrusion into coastal aquifers: a case study in the Crau aquifer." Université d'Avignon.
Nikolos, I. K., Stergiadi, M., Papadopoulou, M. P., and Karatzas, G. P. (2008). "Artificial neural networks as an alternative approach to groundwater numerical modelling and environmental design." *Hydrological Processes*, 22(17), 3337-3348.
Nishikawa, T. (1998). "Water-resources optimization model for Santa Barbara, California." *Journal of water resources planning and management*, 124(5), 252-263.
Nouiri, I., Yitayew, M., Maßmann, J., and Tarhouni, J. (2015). "Multi-objective optimization tool for integrated groundwater management." *Water Resources Management*, 29(14), 5353-5375.
Nunes, L., Cunha, M., and Ribeiro, L. (2004). "Groundwater monitoring network optimization with redundancy reduction." *Journal of Water Resources Planning and Management*, 130(1), 33-43.
Oberdorfer, J. A., Hogan, P. J., and Buddemeier, R. W. (1990). "Atoll island hydrogeology: flow and freshwater occurrence in a tidally dominated system." *Journal of Hydrology*, 120(1-4), 327-340.
Owusu, S., Mul, M. L., Ghansah, B., Osei-Owusu, P. K., Awotwe-Pratt, V., and Kadyampakeni, D. (2017). "Assessing land suitability for aquifer storage and recharge in northern Ghana using remote sensing and GIS multi-criteria decision analysis technique." *Modeling Earth Systems and Environment*, 1-11.
Padilla, F., and Cruz-Sanjulián, J. (1997). "Modeling Sea-Water Intrusion with Open Boundary Conditions." *Ground Water*, 35(4), 704-712.
Papadopoulou, M. P., Nikolos, I. K., and Karatzas, G. P. (2010). "Computational benefits using artificial intelligent methodologies for the solution of an environmental design problem: saltwater intrusion." *Water Science and Technology*, 62(7), 1479-1490.
Park, N., and Shi, L. (2015). "A comprehensive sharp-interface simulation-optimization model for fresh and saline groundwater management in coastal areas." *Hydrogeology Journal*, 23(6), 1195-1204.
Parveen, N., Zaidi, S., and Danish, M. (2016). "Support vector regression model for predicting the sorption capacity of lead (II)." *Perspectives in Science*, 8, 629-631.
Pawar, P. J., Vidhate, U. S., and Khalkar, M. Y. (2017). "Improving the quality characteristics of abrasive water jet machining of marble material using multi-objective artificial bee colony algorithm." *Journal of Computational Design and Engineering*.
Perrone, M. P., and Cooper, L. N. (1995). "When networks disagree: Ensemble methods for hybrid neural networks." *How We Learn; How We Remember: Toward An Understanding Of Brain And Neural Systems: Selected Papers of Leon N Cooper*, World Scientific, 342-358.
Person, M., Taylor, J. Z., and Dingman, S. L. (1998). "Sharp interface models of salt water intrusion and wellhead delineation on Nantucket Island, Massachusetts." *Ground Water*, 36(5), 731-742.
Petrakova, A., Affenzeller, M., and Merkurjeva, G. (2015). "Heterogeneous versus homogeneous machine learning ensembles." *Information Technology and Management Science*, 18(1), 135-140.
Polemio, M., Dragone, V., and Limoni, P. (2009). "Monitoring and methods to analyse the groundwater quality degradation risk in coastal karstic aquifers (Apulia, Southern Italy)." *Environmental Geology*, 58(2), 299-312.
Pousa, J., Tosi, L., Kruse, E., Guaraglia, D., Bonardi, M., Mazzoldi, A., Rizzetto, F., and Schnack, E. (2007). "Coastal processes and environmental hazards: the Buenos Aires (Argentina) and Venetian (Italy) littorals." *Environmental Geology*, 51(8), 1307-1316.

Prakash, M., and Singh, V. (2000). "Network design for groundwater monitoring–a case study." *Environmental Geology*, 39(6), 628-632.

Prakash, O., and Datta, B. (2013). "Sequential optimal monitoring network design and iterative spatial estimation of pollutant concentration for identification of unknown groundwater pollution source locations." *Environmental monitoring and assessment*, 185(7), 5611-5626.

Presley, T. K. (2005). "Effects of the 1998 drought on the freshwater lens in the Laura area, Majuro Atoll, Republic of the Marshall Islands." Geological Survey (US).

Preziosi, E., Petrangeli, A., and Giuliano, G. (2013). "Tailoring groundwater quality monitoring to vulnerability: a GIS procedure for network design." *Environmental monitoring and assessment*, 185(5), 3759-3781.

Prinos, S. T., Lietz, A., and Irvin, R. (2002). "Design of a real-time ground-water level monitoring network and portrayal of hydrologic data in southern Florida." Geological Survey (US).

Pulukuri, S., Keesara, V. R., and Deva, P. (2018). "Flow Forecasting in a Watershed using Autoregressive Updating Model." *Water Resources Management*, 1-16.

Putti, M., and Paniconi, C. (1995). "Picard and Newton linearization for the coupled model for saltwater intrusion in aquifers." *Advances in water resources*, 18(3), 159-170.

Qu, B., Zhang, X., Pappenberger, F., Zhang, T., and Fang, Y. (2017). "Multi-model grand ensemble hydrologic forecasting in the Fu river basin using Bayesian model averaging." *Water*, 9(2), 74.

Ranjithan, S., Eheart, J., and Garrett, J. (1993). "Neural network-based screening for groundwater reclamation under uncertainty." *Water Resources Research*, 29(3), 563-574.

Rao, S., Bhallamudi, S. M., Thandaveswara, B., and Mishra, G. (2004). "Conjunctive use of surface and groundwater for coastal and deltaic systems." *Journal of Water Resources Planning and Management*, 130(3), 255-267.

Rao, S., and Manju, S. (2007). "Optimal pumping locations of skimming wells." *Hydrological sciences journal*, 52(2), 352-361.

Rasmussen, K., May, W., Birk, T., Mataki, M., Mertz, O., and Yee, D. (2009). "Climate change on three Polynesian outliers in the Solomon Islands: impacts, vulnerability and adaptation." *Geografisk Tidsskrift-Danish Journal of Geography*, 109(1), 1-13.

Rastogi, A., Choi, G. W., and Ukarande, S. (2004). "Diffused interface model to prevent ingress of sea water in multi-layer coastal aquifers." *Journal of Spatial Hydrology*, 4(2).

Razavi, S., Tolson, B. A., and Burn, D. H. (2012). "Review of surrogate modeling in water resources." *Water Resources Research*, 48(7).

Razi, M. A., and Athappilly, K. (2005). "A comparative predictive analysis of neural networks (NNs), nonlinear regression and classification and regression tree (CART) models." *Expert Systems with Applications*, 29(1), 65-74.

Reed, P., Minsker, B., and Valocchi, A. J. (2000). "Cost-effective long-term groundwater monitoring design using a genetic algorithm and global mass interpolation." *Water Resources Research*, 36(12), 3731-3741.

Reed, P., Minsker, B. S., and Goldberg, D. E. (2001). "A multiobjective approach to cost effective long-term groundwater monitoring using an elitist nondominated sorted genetic algorithm with historical data." *Journal of Hydroinformatics*, 3(2), 71-89.

Reed, P. M., and Minsker, B. S. (2004). "Striking the balance: long-term groundwater monitoring design for conflicting objectives." *Journal of Water Resources Planning and Management*, 130(2), 140-149.

Reilly, T., and Goodman, A. (1987). "Analysis of saltwater upconing beneath a pumping well." *Journal of Hydrology*, 89(3), 169-204.

Reilly, T. E., and Goodman, A. S. (1985). "Quantitative analysis of saltwater-freshwater relationships in groundwater systems—A historical perspective." *Journal of Hydrology*, 80(1-2), 125-160.

Rey, J., Martínez, J., Barberá, G., García-Aróstegui, J., García-Pintado, J., and Martínez-Vicente, D. (2013). "Geophysical characterization of the complex dynamics of groundwater and seawater exchange in a highly stressed aquifer system linked to a coastal lagoon (SE Spain)." *Environmental earth sciences*, 70(5), 2271-2282.

Richardson, R. R., Osborne, M. A., and Howey, D. A. (2017). "Gaussian process regression for forecasting battery state of health." *Journal of Power Sources*, 357, 209-219.

Rijsberman, F. R. (2006). "Water scarcity: fact or fiction?" *Agricultural water management*, 80(1), 5-22.

Rokach, L. (2010). *Pattern classification using ensemble methods*, World Scientific.

Ros, S. E. M., and Zuurbier, K. G. (2017). "The Impact of Integrated Aquifer Storage and Recovery and Brackish Water Reverse Osmosis (ASRRO) on a Coastal Groundwater System." *Water*, 9(4), 273.

Rosen, M. R. (2009). "Groundwater monitoring networks." *Groundwater-Volume II*, 441.

Roumasset, J. A., and Wada, C. A. (2010). "Optimal and sustainable groundwater extraction." *Sustainability*, 2(8), 2676-2685.

Roy, D. K., and Datta, B. (2016). "Fuzzy C-Mean Clustering Based Inference System for Saltwater Intrusion Processes Prediction in Coastal Aquifers." *Water Resources Management*, 1-22.

Roy, D. K., and Datta, B. (2017). "Genetic algorithm tuned fuzzy inference system to evolve optimal groundwater extraction strategies to control saltwater intrusion in multi-layered coastal aquifers under parameter uncertainty." *Modeling Earth Systems and Environment*, 1-19.

Roy, D. K., and Datta, B. (2017). "Multivariate Adaptive Regression Spline Ensembles for Management of Multilayered Coastal Aquifers." *Journal of Hydrologic Engineering*, 22(9), 04017031.

Roy, T., Schütze, N., Grundmann, J., Brettschneider, M., and Jain, A. (2016). "Optimal groundwater management using state-space surrogate models: a case study for an arid coastal region." *Journal of Hydroinformatics*, jh2016086.

Rubin, H., and Pinder, G. F. (1977). "Approximate analysis of upconing." *Advances in Water Resources*, 1(2), 97-101.

Saeed, M. M., Bruen, M., and Asghar, M. (2002). "A review of modeling approaches to simulate saline-upconing under skimming wells." *Hydrology Research*, 33(2-3), 165-188.

Sakr, S. A. (1999). "Validity of a sharp-interface model in a confined coastal aquifer." *Hydrogeology Journal*, 7(2), 155-160.

Sanchez, E., Pintos, S., and Queipo, N. V. (2008). "Toward an optimal ensemble of kernel-based approximations with engineering applications." *Structural and multidisciplinary optimization*, 36(3), 247-261.

Sanders, T. G. (1983). *Design of networks for monitoring water quality*, Water Resources Publication.

Scharage, L. (1999). "Optimization Modeling with LINGO. LINDO Systems." *Inc, Chicago, Il*.

Scheidleger, A., Grath, J., and Lindinger, H. "Saltwater intrusion due to groundwater over-exploitation EEA inventory throughout Europe." *Proc., 18th Saltwater Intrusion Meeting, Cartagena, Spain*.

Seiller, G., Anctil, F., and Perrin, C. (2012). "Multimodel evaluation of twenty lumped hydrological models under contrasted climate conditions." *Hydrology and Earth System Sciences*, 16(4), p. 1171-p. 1189.

Servan-Camas, B., and Tsai, F. T.-C. (2009). "Saltwater intrusion modeling in heterogeneous confined aquifers using two-relaxation-time lattice Boltzmann method." *Advances in water resources*, 32(4), 620-631.

Shah, T., Molden, D., Sakthivadivel, R., and Seckler, D. (2000). *Groundwater: Overview of opportunities and challenges*, IWMI.

Shah, T., Roy, A. D., Qureshi, A. S., and Wang, J. "Sustaining Asia's groundwater boom: an overview of issues and evidence." *Proc., Natural Resources Forum*, Wiley Online Library, 130-141.

Shammas, M. I. (2008). "The effectiveness of artificial recharge in combating seawater intrusion in Salalah coastal aquifer, Oman." *Environmental geology*, 55(1), 191-204.

Sherif, M., Kacimov, A., Javadi, A., and Ebraheem, A. A. (2012). "Modeling groundwater flow and seawater intrusion in the coastal aquifer of Wadi Ham, UAE." *Water resources management*, 26(3), 751-774.

Sherif, M., and Singh, V. (2002). "Effect of groundwater pumping on seawater intrusion in coastal aquifers." *Agricultural Sciences*, 7(2), 61-67.

Sherif, M. M., and Hamza, K. I. (2001). "Mitigation of seawater intrusion by pumping brackish water." *Transport in porous media*, 43(1), 29-44.

Sheta, A. F., and Mahmoud, A. "Forecasting using genetic programming." *Proc., Southeastern Symposium on System Theory*, 343-348.

Shu, C., and Burn, D. H. (2004). "Artificial neural network ensembles and their application in pooled flood frequency analysis." *Water Resources Research*, 40(9).

Shu, C., and Ouarda, T. (2008). "Regional flood frequency analysis at ungauged sites using the adaptive neuro-fuzzy inference system." *Journal of Hydrology*, 349(1), 31-43.

Sinclair, P. (2011). "Groundwater in Paradise: Challenges faced in developing and sustaining groundwater resources in the Pacific Island Countries."Japan.

Singh, A. (2014). "Optimization modelling for seawater intrusion management." *Journal of Hydrology*, 508, 43-52.

Singh, R. B., Hales, S., de Wet, N., Raj, R., Hearnden, M., and Weinstein, P. (2001). "The influence of climate variation and change on diarrheal disease in the Pacific Islands." *Environmental health perspectives*, 109(2), 155.

Smola, A., and Vapnik, V. (1997). "Support vector regression machines." *Advances in neural information processing systems*, 9, 155-161.

Soldal, O., Mauring, E., Halvorsen, E., and Rye, N. (1994). "Seawater intrusion and fresh groundwater hydraulics in fjord delta aquifers inferred from ground penetrating radar and resistivity profiles—Sunndalsøra and Esebotn, western Norway." *Journal of applied geophysics*, 32(4), 305-319.

Sorek, S., and Pinder, G. (1999). "Survey of computer codes and case histories." *Seawater Intrusion in Coastal Aquifers—Concepts, Methods and Practices*, Springer, 399-461.

Sreekanth, J., and Datta, B. (2010). "Multi-objective management of saltwater intrusion in coastal aquifers using genetic programming and modular neural network based surrogate models." *Journal of Hydrology*, 393(3-4), 245-256.

Sreekanth, J., and Datta, B. (2011). "Comparative evaluation of genetic programming and neural network as potential surrogate models for coastal aquifer management." *Water resources management*, 25(13), 3201-3218.

Sreekanth, J., and Datta, B. (2011). "Coupled simulation-optimization model for coastal aquifer management using genetic programming-based ensemble surrogate models and multiple-realization optimization." *Water Resources Research*, 47(4).

Sreekanth, J., and Datta, B. (2012). "Comment on "Artificial neural network model as a potential alternative for groundwater salinity forecasting" by Pallavi Banerjee et al.[J. Hydrol. 398 (2011) 212–220]." *Journal of Hydrology*, 420, 419-420.

Sreekanth, J., and Datta, B. (2013). "Design of an optimal compliance monitoring network and feedback information for adaptive management of saltwater intrusion in coastal aquifers." *Journal of Water Resources Planning and Management*, 140(10), 04014026.

Sreekanth, J., and Datta, B. (2014). "Stochastic and robust multi-objective optimal management of pumping from coastal aquifers under parameter uncertainty." *Water resources management*, 28(7), 2005-2019.

Sreekanth, J., and Datta, B. (2015). "Review: simulation-optimization models for the management and monitoring of coastal aquifers." *Hydrogeology Journal*, 23(6), 1155-1166.

Srinivasan, D. (2008). "Energy demand prediction using GMDH networks." *Neurocomputing*, 72(1-3), 625-629.

Storck, P., Eheart, J. W., and Valocchi, A. J. (1997). "A method for the optimal location of monitoring wells for detection of groundwater contamination in three-dimensional heterogenous aquifers." *Water Resources Research*, 33(9), 2081-2088.

Storey, D., and Hunter, S. (2010). "Kiribati: an environmental 'perfect storm'." *Australian Geographer*, 41(2), 167-181.

Strobl, R., Robillard, P., Shannon, R., Day, R., and McDonnell, A. (2006). "A water quality monitoring network design methodology for the selection of critical sampling points: Part I." *Environmental Monitoring and Assessment*, 112(1-3), 137-158.

Suganyadevi, M., and Babulal, C. (2014). "Support vector regression model for the prediction of loadability margin of a power system." *Applied Soft Computing*, 24, 304-315.

Sun, A. Y., Wang, D., and Xu, X. (2014). "Monthly streamflow forecasting using Gaussian process regression." *Journal of Hydrology*, 511, 72-81.

Taboada, H. A., and Coit, D. W. (2007). "Data clustering of solutions for multiple objective system reliability optimization problems." *Quality Technology & Quantitative Management*, 4(2), 191-210.

Taboada, H. A., and Coit, D. W. (2008). "Multi-objective scheduling problems: Determination of pruned Pareto sets." *IIE Transactions*, 40(5), 552-564.

Taylor, R. (1990). "Interpretation of the correlation coefficient: a basic review." *Journal of diagnostic medical sonography*, 6(1), 35-39.

Teng, G., Xiao, J., He, Y., Zheng, T., and He, C. (2017). "Use of group method of data handling for transport energy demand modeling." *Energy Science & Engineering*, 5(5), 302-317.

Terry, J. P., Chui, T. F. M., and Falkland, A. (2013). "Atoll Groundwater Resources at Risk: Combining Field Observations and Model Simulations of Saline Intrusion Following Storm-Generated Sea Flooding." *Groundwater in the Coastal Zones of Asia-Pacific*, Springer, 247-270.

Todd, D. K. (1974). "Salt-water intrusion and its control." *Journal (American Water Works Association)*, 180-187.

Toth, E. (2009). "Classification of hydro-meteorological conditions and multiple artificial neural networks for streamflow forecasting." *Hydrology and Earth System Sciences*, 13(9), 1555.

Trefry, M. G., and Muffels, C. (2007). "FEFLOW: A Finite-Element Ground Water Flow and Transport Modeling Tool." *Ground Water*, 45(5), 525-528.

Tu, J. V. (1996). "Advantages and disadvantages of using artificial neural networks versus logistic regression for predicting medical outcomes." *Journal of clinical epidemiology*, 49(11), 1225-1231.

Tularam, G., and Krishna, M. (2009). "Long term consequences of groundwater pumping in Australia: a review of impacts around the globe." *J Appl Sci Environ Sanit*, 4(2), 151-166.

Van Camp, M., Mtoni, Y., Mjemah, I. C., Bakundukize, C., and Walraevens, K. (2014). "Investigating seawater intrusion due to groundwater pumping with schematic model simulations: The example of the Dar es Salaam coastal aquifer in Tanzania." *Journal of African Earth Sciences*, 96, 71-78.

Van Der Velde, M., Green, S., Vanclooster, M., and Clothier, B. (2007). "Sustainable development in small island developing states: Agricultural intensification, economic development, and freshwater resources management on the coral atoll of Tongatapu." *Ecological Economics*, 61(2), 456-468.

Vandenbohede, A., Van Houtte, E., and Lebbe, L. (2009). "Sustainable groundwater extraction in coastal areas: a Belgian example." *Environmental geology*, 57(4), 735-747.
Vapnik, V., Golowich, S. E., and Smola, A. J. (1997). "Support vector method for function approximation, regression estimation and signal processing." *Proc., Neural Information Processing Systems 1996*, MIT Press, Cambridge, 281-287.
Velázquez, J., Anctil, F., Ramos, M., and Perrin, C. (2011). "Can a multi-model approach improve hydrological ensemble forecasting? A study on 29 French catchments using 16 hydrological model structures." *Advances in Geosciences*, 29, 33-42.
Voss, C. I. (1984). "SUTRA (Saturated-Unsaturated Transport). A Finite-Element Simulation Model for Saturated-Unsaturated, Fluid-Density-Dependent Ground-Water Flow with Energy Transport or Chemically-Reactive Single-Species Solute Transport." DTIC Document.
Voss, C. I., and Souza, W. R. (1987). "Variable density flow and solute transport simulation of regional aquifers containing a narrow freshwater-saltwater transition zone." *Water Resources Research*, 23(10), 1851-1866.
Wada, Y., van Beek, L. P., van Kempen, C. M., Reckman, J. W., Vasak, S., and Bierkens, M. F. (2010). "Global depletion of groundwater resources." *Geophysical research letters*, 37(20).
Wang, C., Duan, Q., Gong, W., Ye, A., Di, Z., and Miao, C. (2014). "An evaluation of adaptive surrogate modeling based optimization with two benchmark problems." *Environmental Modelling & Software*, 60, 167-179.
Wang, Z., and Rangaiah, G. P. (2017). "Application and Analysis of Methods for Selecting an Optimal Solution from the Pareto-Optimal Front obtained by Multiobjective Optimization." *Industrial & Engineering Chemistry Research*, 56(2), 560-574.
Ward, D., Reeves, M., and Duda, L. E. (1984). "Verification and field comparison of the sandia waste-isolation flow and transport model (SWIFT)." Sandia National Labs., Albuquerque, NM (USA).
Webb, A. (2007). "Assessment of salinity of groundwater in swamp taro (Cyrtosperma chamissonis)"puluka pits in Tuvalu." *EU EDF8-SOPAC Project report Suva, Fiji, SOPAC*, 75, 37.
Wen, Y., Cai, C., Liu, X., Pei, J., Zhu, X., and Xiao, T. (2009). "Corrosion rate prediction of 3C steel under different seawater environment by using support vector regression." *Corrosion Science*, 51(2), 349-355.
Werner, A. D. (2010). "A review of seawater intrusion and its management in Australia." *Hydrogeology journal*, 18(1), 281-285.
Werner, A. D., Bakker, M., Post, V. E., Vandenbohede, A., Lu, C., Ataie-Ashtiani, B., Simmons, C. T., and Barry, D. A. (2013). "Seawater intrusion processes, investigation and management: recent advances and future challenges." *Advances in Water Resources*, 51, 3-26.
Werner, A. D., Jakovovic, D., and Simmons, C. T. (2009). "Experimental observations of saltwater up-coning." *Journal of Hydrology*, 373(1), 230-241.
Westerhuis, J. A., Hoefsloot, H. C., Smit, S., Vis, D. J., Smilde, A. K., van Velzen, E. J., van Duijnhoven, J. P., and van Dorsten, F. A. (2008). "Assessment of PLSDA cross validation." *Metabolomics*, 4(1), 81-89.
White, I., and Falkland, A. "Practical responses to climate change: Developing national water policy and implementation plans for pacific small island countries." *Proc., Water and Climate: Policy Implementation Challenges; Proceedings of the 2nd Practical Responses to Climate Change Conference*, Engineers Australia, 439.
White, I., and Falkland, T. (2010). "Management of freshwater lenses on small Pacific islands." *Hydrogeology Journal*, 18(1), 227-246.
White, I., Falkland, T., Crennan, L., Jones, P., Metutera, T., Etuati, B., and Metai, E. (1999). "Groundwater recharge in low coral island Bonriki, South Tarawa, Republic of Kiribati: isues, traditions and conflicts in groundwater use and management." *Technical documents in hydrology*, UNESCO.

White, I., Falkland, T., and Rebgetz, M. (2008). "Report on the protection and management of water reserves, South Tarawa." Kiribati Adaptation Project, Phase.

Wichard, J., and Ogorzalek, M. "Time series prediction with ensemble models." *Proc., Proceedings of International Joint Conference on Neural Networks (IJCNN 2004)*, Citeseer, 1625-1629.

Willmott, C. J. (1981). "On the validation of models." *Physical geography*, 2(2), 184-194.

Wu, C.-H., Ho, J.-M., and Lee, D.-T. (2004). "Travel-time prediction with support vector regression." *IEEE transactions on intelligent transportation systems*, 5(4), 276-281.

Wu, J., Zheng, C., and Chien, C. C. (2005). "Cost-effective sampling network design for contaminant plume monitoring under general hydrogeological conditions." *Journal of Contaminant Hydrology*, 77(1), 41-65.

Wu, X., Kumar, V., Quinlan, J. R., Ghosh, J., Yang, Q., Motoda, H., McLachlan, G. J., Ng, A., Liu, B., and Philip, S. Y. (2008). "Top 10 algorithms in data mining." *Knowledge and information systems*, 14(1), 1-37.

Xia, Z., Mao, K., Wei, S., Wang, X., Fang, Y., and Yang, S. (2017). "Application of genetic algorithm-support vector regression model to predict damping of cantilever beam with particle damper." *Journal of Low Frequency Noise, Vibration and Active Control*, 36(2), 138-147.

Xiao, J., Cao, H., Jiang, X., Gu, X., and Xie, L. (2017). "GMDH-based semi-supervised feature selection for customer classification." *Knowledge-Based Systems*, 132, 236-248.

Xue, Y., Xie, C., Wu, J., Liu, P., Wang, J., and Jiang, Q. (1995). "A Three-Dimensional Miscible Transport Model For Seawater Intrusion in China." *Water Resources Research*, 31(4), 903-912.

Yadollahi, M., Abd Majid, M. Z., and Mohamad Zin, R. (2015). "Post-Pareto optimality approach to enhance budget allocation process for bridge rehabilitation management." *Structure and Infrastructure Engineering*, 11(12), 1565-1582.

Yang, F.-g., Cao, S.-y., Liu, X.-n., and Yang, K.-j. (2008). "Design of groundwater level monitoring network with ordinary kriging." *Journal of Hydrodynamics*, 20(3), 339-346.

Yang, H., Chan, L., and King, I. "Support vector machine regression for volatile stock market prediction." *Proc., International Conference on Intelligent Data Engineering and Automated Learning*, Springer, 391-396.

Yangxiao, Z. (1994). "Objectives, criteria and methodologies for the design of primary groundwater monitoring networks." *IAHS Publications-Series of Proceedings and Reports-Intern Assoc Hydrological Sciences*, 222, 285-296.

Yeh, M.-S., Lin, Y.-P., and Chang, L.-C. (2006). "Designing an optimal multivariate geostatistical groundwater quality monitoring network using factorial kriging and genetic algorithms." *Environmental Geology*, 50(1), 101-121.

Yin, H., Wen, G., Fang, H., Qing, Q., Kong, X., Xiao, J., and Liu, Z. (2014). "Multiobjective crashworthiness optimization design of functionally graded foam-filled tapered tube based on dynamic ensemble metamodel." *Materials & Design*, 55, 747-757.

Yoon, H., Jun, S.-C., Hyun, Y., Bae, G.-O., and Lee, K.-K. (2011). "A comparative study of artificial neural networks and support vector machines for predicting groundwater levels in a coastal aquifer." *Journal of Hydrology*, 396(1), 128-138.

Yu, P.-S., Chen, S.-T., and Chang, I.-F. (2006). "Support vector regression for real-time flood stage forecasting." *Journal of Hydrology*, 328(3), 704-716.

Yu, S., Zheng, S., Gao, S., and Yang, J. (2017). "A multi-objective decision model for investment in energy savings and emission reductions in coal mining." *European Journal of Operational Research*, 260(1), 335-347.

Žalik, K. R. (2008). "An efficient k′-means clustering algorithm." *Pattern Recognition Letters*, 29(9), 1385-1391.

Zektser, S., Loáiciga, H. A., and Wolf, J. (2005). "Environmental impacts of groundwater overdraft: selected case studies in the southwestern United States." *Environmental Geology*, 47(3), 396-404.
Zhang, C., Wei, H., Zhao, X., Liu, T., and Zhang, K. (2016). "A Gaussian process regression based hybrid approach for short-term wind speed prediction." *Energy Conversion and Management*, 126, 1084-1092.
Zhang, J., Li, W., Zeng, L., and Wu, L. (2016). "An adaptive Gaussian process-based method for efficient Bayesian experimental design in groundwater contaminant source identification problems." *Water Resources Research*, 52(8), 5971-5984.
Zhang, Y., Pinder, G. F., and Herrera, G. S. (2005). "Least cost design of groundwater quality monitoring networks." *Water Resources Research*, 41(8).
Zhao, Z.-w., Zhao, J., and Fu, C.-s. (2013). "Uncertainty analysis of seawater intrusion forecasting." *Water Science and Engineering*, 6(4), 380-391.
Zhou, Q., Bear, J., and Bensabat, J. (2005). "Saltwater upconing and decay beneath a well pumping above an interface zone." *Transport in Porous Media*, 61(3), 337-363.
Zhou, X., Ma, Y., Tu, Y., and Feng, Y. (2013). "Ensemble of surrogates for dual response surface modeling in robust parameter design." *Quality and Reliability Engineering International*, 29(2), 173-197.
Zhou, X. J., Ma, Y. Z., and Li, X. F. (2011). "Ensemble of surrogates with recursive arithmetic average." *Structural and Multidisciplinary Optimization*, 44(5), 651-671.
Zhou, Y., Dong, D., Liu, J., and Li, W. (2013). "Upgrading a regional groundwater level monitoring network for Beijing Plain, China." *Geoscience Frontiers*, 4(1), 127-138.
Zhou, Z.-H. (2012). *Ensemble methods: foundations and algorithms*, CRC press.
Zhu, X., Yue, Y., Wong, P. W., Zhang, Y., and Ding, H. (2019). "Designing an Optimized Water Quality Monitoring Network with Reserved Monitoring Locations." *Water*, 11(4), 713.
Zio, E., and Bazzo, R. (2010). "Multiobjective optimization of the inspection intervals of a nuclear safety system: A clustering-based framework for reducing the Pareto Front." *Annals of Nuclear Energy*, 37(6), 798-812.