NATIONAL TECHNICAL UNIVERSITY OF ATHENS

SCHOOL OF RURAL AND SURVEYING ENGINEERING

DEPARTMENT OF INFRASTRUCTURE AND RURAL DEVELOPMENT

SURROGATE-BASED OPTIMIZATION METHODS FOR COASTAL AQUIFER MANAGEMENT

Vasileios Christelis

SUBMITTED IN PARTIAL FULFILLMENT OF THE

REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

SUPERVISOR:

Professor Aristotelis Mantoglou

Athens, January 2021

NATIONAL TECHNICAL UNIVERSITY OF ATHENS

SCHOOL OF RURAL AND SURVEYING ENGINEERING

DEPARTMENT OF INFRASTRUCTURE AND RURAL DEVELOPMENT

SURROGATE-BASED OPTIMIZATION METHODS FOR COASTAL AQUIFER MANAGEMENT

Vasileios Christelis

SUBMITTED IN PARTIAL FULFILLMENT OF THE

REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

SUPERVISOR:

Professor Aristotelis Mantoglou

Athens, January 2021

THESIS COMMITTEE

SUPERVISOR

Professor Aristotelis Mantoglou, National Technical University of Athens

ADVISORY COMMITTEE

1. Professor Aristotelis Mantoglou, National Technical University of Athens
2. Professor George P. Karatzas, Technical University of Crete
3. Assistant Professor Aikaterini Nanou-Giannarou, National Technical University of Athens

EVALUATION COMMITTEE

1. Professor Aristotelis Mantoglou, National Technical University of Athens
2. Professor George P. Karatzas, Technical University of Crete
3. Assistant Professor Aikaterini Nanou-Giannarou, National Technical University of Athens
4. Professor Evangelos Baltas, National Technical University of Athens
5. Professor Konstantinos Katsifarakis, Aristotle University of Thessaloniki
6. Associate Professor Ioannis Nalbantis, National Technical University of Athens
7. Professor Vassilios Tsihrintzis, National Technical University of Athens

CONTENTS

«Η έγκριση της διδακτορικής διατριβής από την Ανώτατη Σχολή Αγρονόμων και Τοπογράφων Μηχανικών του Ε.Μ.Π. δεν υποδηλώνει αποδοχή των γνωμών του συγγραφέα (Ν. 5343/1932, Άρθρο 202)»

Acknowledgments

When I was younger, I was mostly good at running sprints. Some people say that the pursuit of a PhD resembles running a marathon and I knew it would be hard. However, life sometimes just like that, throws you extreme marathons and unpleasant surprises. So be it then!

First, I would like to express my sincere thanks to Professor Aristotelis Mantoglou, my thesis supervisor, who introduced me to the fascinating topic of surrogate-based optimization and provided to me constructive suggestions and support during this research. I would also like to thank him for encouraging me not to quit when I had to run on really “long and tough tracks”. I would also like to thank Professor George Karatzas and Assistant Professor Aikaterini Nanou-Giannarou for agreeing to serve on the supervising committee. I am grateful for their support and advice throughout the completion of this thesis.

Being a member of the Laboratory of Reclamation Works and Water Resources Management during those years was a delight! I am thankful for the warm, kind and stimulating environment that was zealously pursued by Professor George Tsakiris, Associate Professor Ioannis Nalbantis and Professor Vassilios Tsihrintzis. I would also like to thank them for their continuous support and friendly attitude.

I was lucky to meet and make new great friends during my PhD years, namely, George Kopsiaftis, Dr. Harris Vangelis, Dr. Dimitris Tigkas and Dr. Vasilis Bellos. We shared research ideas, enjoyed conferences, and had fun in our vivid coffee breaks! I would also like to thank them for the things that I learned via our collaboration on scientific papers and our discussions on hydrology and other engineering subjects. Special thanks go to George Kopsiaftis who, as a researcher in the same field of coastal aquifer management, selflessly offered his programming and hydrogeology knowledge during our fruitful discussions. Hopefully, his PhD thesis will soon be on its way to completion.

Herein, I wish to thank Professor Rommel Regis for generously offering his insights on the topic of optimization and surrogate modelling and for our ongoing research collaboration which randomly started three years ago. Also, I would like to thank Rob McLaren for his help and guidance regarding groundwater model development with the HydroGeoSphere code (then an academic research code) during my first PhD years. Also, I would like to thank Dr. Andreas Efstratiadis and Dr. Ioannis Tsoukalas for sharing the code of EAS and SEEAS algorithms.

Furthermore, I would like to thank my current colleagues in British Geological Survey and the organization itself for encouraging me to complete the PhD thesis from the very first moment I started working there as well as for their constant support during tough times. Particularly, I would like to thank my colleagues Dr. Andrew Barkwith, Dr. Andrew Hughes, Dr. Christopher Jackson, Dr. Majdi Mansoor, Dr. Marco Bianchi, Dr. Johana Scheidegger, Fiona McEvoy, Dr. Barbara Palumbo-Roe, Dr. Corinna Abesser, Sarah Collins, Dr. Elena Tamayo-Mas, Dr. Azucena Rodriguez Yebra, Dr. Jon Mackay, Dr. Lei Wang, Dr. Rob Ward and Dr. Evi Petravatzi.

I would like to thank Professor Vassilios Tsihrintzis, Associate Professor Ioannis Nalbantis, Professor Konstantinos Katsifarakis and Professor Evangelos Baltas for agreeing to be members of the evaluation committee and for their comments on improving the writing of this thesis. Furthermore, I am grateful of the financial support that was provided to me through the Graduate Scholarship awarded by the NTUA Office of the Special Account for Research (E.Λ.K.E.).

I cannot think of how much space I would ideally need to write down my feelings and my thankfulness to my beloved parents and sister, Christos, Angeliki and Leeda, to my wonderful wife

Zoi and our adorable little angel, little Angeliki. Each one of them, in their own unique way, helped me to keep things moving forward and lift my spirits up.

Physically writing this at Nottingham (mind travelling to Athens and Veroia)

December 2020

« *No alarms and no surprises, please…* »

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

Names	**Descriptions**
AR	Adaptive-recursive surrogate-assisted optimization framework
AR-coKRG^{CONS}	Adaptive-recursive surrogate-assisted optimization framework using individual co-Kriging surrogate models for the constraint functions
AR-coKRG^{OBJ}	Adaptive-recursive surrogate-assisted optimization framework using a single co-Kriging surrogate model for the penalized objective function
coKRG	co-Kriging surrogate model
ConstrLMSRBF	Local metric stochastic RBF algorithm for large-scale optimization involving expensive black-box objective and constraint functions
CSEEAS	Constrained surrogate-enhanced evolutionary annealing **s**implex algorithm
EAS	Evolutionary annealing simplex algorithm
EAS-VDST	Optimization framework based on evolutionary annealing simplex algorithm and variable density and solute transport numerical models

EAS-PB	Prediction-based infill strategy embedded in the operations of the evolutionary annealing **s**implex algorithm
GOSAC	Global optimization algorithm with **s**urrogate approximation of con**s**traints
HF	High-fidelity
KRG	Kriging
LF	Low-fidelity
LR-RSRBF	Random search surrogate-assisted algorithm with local refinement and exploration steps using RBF models
MHFr	Maximum available number of runs with the high-fidelity variable density and solute transport numerical model
MLMSRBF	Multi-start local metric stochastic RBF algorithm
RBF	Radial basis functions
SBO	Surrogate-based optimization
SEEAS	Surrogate enhanced evolutionary annealing simplex algorithm
SWI	Seawater intrusion
VDST	Variable density flow and solute transport

ABSTRACT

In pumping optimization of coastal aquifers, often the objective is to mitigate the phenomenon of saltwater intrusion while satisfying the demands for freshwater extraction. This management task is typically formulated as a nonlinear constrained optimization problem where the individual pumping rates are the decision variables. Evolutionary algorithms are considered highly competent to find a near global optimum to this difficult optimization problem, at the expense of thousands of function evaluations with the physics-based seawater intrusion model. To solve a pumping optimization problem of that type using variable-density flow and solute transport numerical models, leads to a computational cost of many hours, days or even months in the case of an extremely time-consuming simulation, depending on the number of abstraction wells and the length of the simulation period considered.

To that end, the present thesis focused on the use of surrogate modelling techniques as a realistic approach to computationally expensive problems of pumping optimization of coastal aquifers. The aim was to develop surrogate-based optimization methods that can realistically be applied in real-world coastal aquifer management problems. In surrogate-based optimization, fast approximation models, commonly called surrogate models or metamodels, are built using input-output data from the computationally intensive physics-based model that simulates the system under study. Then, the surrogate models are used in lieu of the physics-based model to search for the optimum in the decision variable space. Here, radial basis functions and Kriging models were selected to develop surrogate-based optimization frameworks. While these surrogate models have been fairly utilized in other fields of engineering optimization, their use is scarce in aquifer management studies. As interpolating surrogate models, they can exactly predict the response of the physics-based model to points that exist in the training dataset. This was considered as a

preferable feature for emulating the deterministic response of the seawater intrusion models utilized in this work. Individual surrogate models were constructed to approximate each nonlinear constraint function included in the pumping optimization problems defined in this thesis.

Emphasis was given on the development of online surrogate-based optimization methods where the accuracy of the surrogate models was further enhanced during the operations of the optimization process. That is, additional points were selectively evaluated with the physics-based model to update the surrogate models after the initial training. This approach, typically called infill strategy, allowed for a more efficient and effective search of feasible optimal solutions, as opposed to offline methods that typically require large training datasets to develop accurate surrogate models in the whole decision variable space.

Different infill strategies were investigated within the proposed surrogate-based optimization methods. One approach utilized a pure exploitation infill strategy that was embedded in the operations of an evolutionary algorithm (surrogate-assisted evolutionary strategy). This optimization framework adopts an aggressive strategy where the physics-based model evaluates a candidate solution only at the current optimum located by the surrogate models. Here, it was implemented by either using single predefined surrogate models for the constraint functions or via a multiple surrogate approach. In the latter case, a cross-validation strategy was employed to select the best surrogate model for each constraint function or to construct an ensemble surrogate model of weighted prediction. To the best of author's knowledge, the use of an ensemble surrogate with optimal weights in a surrogate-assisted evolutionary strategy was applied for the first time in pumping optimization of coastal aquifers. Another popular infill strategy in surrogate-based optimization, is to employ various criteria for balancing exploration and exploitation using the metamodels. That is, to evaluate certain points with the physics-based model away from the current optimum (exploration step) that can improve the global accuracy of the metamodel and increase its efficacy to locate near-optimal solutions. Such methods have received little attention in pumping optimization of coastal aquifers despite their proven success in the broader engineering optimization literature. In this thesis, a new surrogate-based optimization algorithm was developed which aims to balance exploration with exploitation within an adaptive-recursive optimization framework.

The surrogate-based optimization methods developed here, were applied in pumping optimization problems for a hypothetical and a real-world coastal aquifer case. By using several independent optimization trials, extensive comparisons were conducted to assess the effectiveness and the efficiency of the proposed surrogate-based methods in coastal aquifer management. In overall, results showed that the optimization methods based on surrogate models drastically reduced the computational cost of the corresponding optimization that was based solely on the variable-density flow and solute transport model. For pumping optimization problems of moderate dimensionality (i.e., 10 decision variables), all methods approached the region of the global optimum while those balancing exploration with exploitation provided near-optimal solutions within just 100 evaluations with the variable-density flow and solute transport model.

For pumping optimization problems of a larger dimensionality (i.e., 20 decision variables), the performance of the surrogate-assisted evolution framework, which utilized a pure exploitation infill strategy, was negatively affected. The use of the multiple-surrogate approach within this evolutionary strategy did not outperform the corresponding implementation with the single surrogate models. However, the optimization schemes which balance exploration and exploitation had a consistent performance independent of the initial training sample, particularly when the number of evaluations with the physics-based model was increased to 300. The new surrogate-based optimization method that balances exploration with exploitation and was developed in this thesis, performed better or comparable to other published surrogate-based optimization algorithms that have been successfully applied in the water resources optimization literature.

The present thesis also addressed those cases where the variable-density flow and solute transport models are extremely time-consuming and it is unrealistic to run more than a limited number of simulations. In such cases, the development of conventional surrogate-based optimization methods that utilize training points solely obtained from the high-fidelity simulation model are either impractical or of limited efficacy. Multi-fidelity optimization methods are considered a useful alternative for these problems, yet this approach is largely unexplored in coastal aquifer management. In practice, multi-fidelity optimization utilizes a large number of low-fidelity but computationally cheap simulations while only a limited number of runs with the high-fidelity model is used. A correction process is established via the development of a surrogate model which combines the abundant low-fidelity data and the limited high-fidelity data. This correction aims to smooth out the inaccuracies from the low-fidelity model and to enable the search for optimal solutions in a computationally affordable manner.

In this thesis, a novel multi-fidelity optimization method was developed based on an adaptive-recursive framework and co-Kriging surrogate models. Two levels of model fidelity were considered, the high-fidelity variable-density flow and solute transport model and a low-fidelity and computationally inexpensive sharp interface model. This was the first time that a multi-fidelity optimization approach based on co-Kriging surrogate models was developed for coastal aquifer management. The proposed multi-fidelity optimization framework delivered good local solutions by using as few as 21 simulations with the variable-density flow and solute transport model and 200 simulations with the sharp interface model for a pumping optimization problem of 10 decision variables. Additionally, the multi-fidelity optimization method outperformed conventional surrogate-based optimization for such a limited number of simulations with the high-fidelity seawater intrusion model.

EXTENDED ABSTRACT -ΕΚΤΕΤΑΜΕΝΗ ΠΕΡΙΛΗΨΗ

«Μέθοδοι βελτιστοποίησης βασισμένες σε μετα-μοντέλα στην διαχείριση παράκτιων υδροφορέων»

Η ραγδαία αύξηση της υπολογιστικής ισχύος διεύρυνε την δυνατότητα ανάπτυξης αριθμητικών μοντέλων υψηλής πιστότητας στην προσομοίωση των φυσικών συστημάτων. Η ενσωμάτωση πολύπλοκων διεργασιών στην ανάπτυξη ενός μαθηματικού μοντέλου οδηγεί αναπόφευκτα σε χρονοβόρες αριθμητικές επιλύσεις. Ωστόσο, η λήψη αποφάσεων, η ανάλυση αβεβαιότητας, η βελτιστοποίηση και άλλες επαναληπτικές μεθοδολογίες βασίζονται στην χρήση πολλαπλών προσομοιώσεων υψηλής πιστότητας για μια αποτελεσματική προσέγγιση στην επίλυση προβλημάτων σε πολύπλοκα φυσικά συστήματα. Η σύζευξη των εν λόγω μεθοδολογιών με λεπτομερή αριθμητικά μοντέλα συχνά κρίνεται ως μη πρακτική λόγω του μεγάλου υπολογιστικού φόρτου. Ένα αντίστοιχο πρόβλημα πραγματεύεται και η παρούσα διατριβή στην οποία η χρήση μοντέλων υψηλής πιστότητας για την προσομοίωση της παράκτιας υπόγειας ροής σε προβλήματα βελτιστοποίησης αντλήσεων, εμποδίζει την εφαρμογή καθιερωμένων σχημάτων προσομοίωσης-βελτιστοποίησης λόγω αυξημένου υπολογιστικού φόρτου.

Μια διαθέσιμη μέθοδος, αναγνωρισμένη στην βιβλιογραφία για την επιτυχή μείωση του υπολογιστικού κόστους κατά την βελτιστοποίηση, είναι η χρήση μετα-μοντέλων στην οποία εστιάζει η παρούσα διδακτορική διατριβή. Τα μετα-μοντέλα είναι μαθηματικά μοντέλα ικανά να περιγράφουν την σχέση μεταξύ των δεδομένων εισόδου-εξόδου ενός υπολογιστικού μοντέλου φυσικής βάσης. Βασικό χαρακτηριστικό τους είναι ότι εφόσον εκπαιδευτούν στην απόκριση του υπολογιστικού μοντέλου για διάφορους συνδυασμούς των μεταβλητών εισόδου, μπορούν να γενικεύσουν την ικανότητα πρόβλεψης μιας αντίστοιχης απόκρισης για νέα δεδομένα εισόδου. Η εκτέλεση ενός υπολογισμού με τα μετα-μοντέλα γίνεται σε πολύ μικρό υπολογιστικό χρόνο, ενώ η ακρίβεια της πρόβλεψης ενισχύεται με την συγκέντρωση μεγάλου αριθμού δειγμάτων εισόδου-εξόδου από το ακριβές υπολογιστικό μοντέλο.

Ωστόσο, σε προβλήματα βελτιστοποίησης η δημιουργία ενός μετα-μοντέλου υψηλής ακρίβειας σε ολόκληρο τον χώρο των μεταβλητών απόφασης συνεπάγεται υψηλό υπολογιστικό κόστος και καθιστά μη πρακτική την χρήση των μετα-μοντέλων. Η προσέγγιση που συνήθως προκρίνεται είναι η αρχική επίτευξη μιας μερικής ακρίβειας των μετα-μοντέλων, η οποία βελτιώνεται στην συνέχεια ανάλογα με την μεθοδολογία ανανέωσης του δείγματος εκπαίδευσης (δειγματοληψία) του μετα-μοντέλου. Μια αποτελεσματική στρατηγική δειγματοληψίας στοχεύει στην αναζήτηση ισορροπίας μεταξύ ευρείας εξερεύνησης (exploration) και στοχευμένης εκμετάλλευσης (exploitation) στον χώρο των μεταβλητών απόφασης, στηριζόμενη στην πληροφορία εισόδου-εξόδου που παρέχει η προσομοίωση με το υπολογιστικό μοντέλο φυσικής βάσης. Μια ικανοποιητική ακρίβεια του μετα-μοντέλου σε συνδυασμό με μια αποτελεσματική στρατηγική δειγματοληψίας, αναμένεται να οδηγήσει σε λύσεις πολύ κοντά στο ολικό βέλτιστο μειώνοντας δραστικά τον απαιτούμενο υπολογιστικό χρόνο.

Η παρούσα διατριβή εστιάζει στην ανάπτυξη μεθόδων βασισμένων σε μετα-μοντέλα για την επίλυση προβλημάτων βελτιστοποίησης αντλήσεων σε παράκτιους υδροφορείς. Δίνεται έμφαση στην ανάπτυξη και ταυτόχρονα διεξοδική σύγκριση ανάλογων μεθόδων σε προβλήματα βελτιστοποίησης με διαφορετική διάσταση αλλά και διαφορετικό αριθμό προϋπολογισμένων διαθέσιμων προσομοιώσεων με το υπολογιστικά απαιτητικό αριθμητικό μοντέλο. Το περιεχόμενο των κεφαλαίων της διατριβής περιγράφεται παρακάτω:

Στο κεφάλαιο 2 περιγράφεται το φυσικό φαινόμενο της υφαλμύρωσης που αφορά την προσομοίωση της παράκτιας υπόγειας ροής και δίνεται μια σύντομη περιγραφή των μαθηματικών μοντέλων που χρησιμοποιήθηκαν στην παρούσα έρευνα. Επίσης περιγράφεται η δυνατότητα να οριστούν μοντέλα διαφορετικής πιστότητας στην προσομοίωση της υφαλμύρωσης το οποίο αξιοποιείται στα επόμενα κεφάλαια κατά την ανάπτυξη των μεθόδων βελτιστοποίησης με μετα-μοντέλα. Ως υψηλής πιστότητας μοντέλο υφαλμύρωσης θεωρήθηκε το μοντέλο μεταβλητής πυκνότητας και μεταφοράς ρύπου ενώ ως χαμηλότερης πιστότητας αλλά και χαμηλότερου υπολογιστικού κόστους θεωρήθηκε το μοντέλο απότομης διεπιφάνειας κατά Strack (1976).

Το κεφάλαιο 3 περιλαμβάνει αρχικά μια σύντομη βιβλιογραφική ανασκόπηση του προβλήματος της βελτιστοποίησης αντλήσεων στους παράκτιους υδροφορείς. Στην συνέχεια, ορίζεται μαθηματικά το πρόβλημα βελτιστοποίησης με μη-γραμμικούς περιορισμούς που εξετάζεται στην παρούσα διατριβή για την εύρεση των βέλτιστων παροχών άντλησης με την χρήση εξελικτικών αλγορίθμων.

Το κεφάλαιο 4 περιλαμβάνει μια εκτενή αναφορά στις υπάρχουσες πρακτικές βελτιστοποίησης με μετα-μοντέλα σε άλλους κλάδους της μηχανικής αλλά και στην διαχείριση παράκτιων υδροφορέων. Γίνεται λεπτομερής συζήτηση για την διαφορά στην χρήση των μετα-μοντέλων όταν ενυπάρχουν μη-γραμμικοί περιορισμοί στην διατύπωση του προβλήματος βελτιστοποίησης και περιγράφονται οι δυνατότητες διαφορετικών μεθοδολογιών για την στρατηγική δειγματοληψίας με το υπολογιστικά απαιτητικό μοντέλο. Στην συνέχεια περιγράφονται τα μετα-μοντέλα που επιλέχθηκαν για την παρούσα έρευνα, τα οποία είναι οι συναρτήσεις ακτινικής βάσης (radial basis functions) και η μεθοδολογία Kriging. Παρουσιάζονται επίσης τα σχήματα που αναπτύχθηκαν για την αναζήτηση βέλτιστης λύσης με χρήση μετα-μοντέλων τα οποία συνοψίζονται στα ακόλουθα:

1. Σχήμα βελτιστοποίησης βασισμένο σε μετα-μοντέλα ενσωματωμένο στις διαδικασίες εξελικτικού αλγόριθμου και εφαρμογή στρατηγικής στοχευμένης εκμετάλλευσης.

2. Σχήμα βελτιστοποίησης βασισμένο σε πολλαπλά μετα-μοντέλα (επιλογή καλύτερου μετα-μοντέλου ή δημιουργία συνόλου μετα-μοντέλων με υπολογισμό βέλτιστων βαρών) ενσωματωμένο στις διαδικασίες εξελικτικού αλγόριθμου και εφαρμογή στρατηγικής στοχευμένης εκμετάλλευσης.

3. Αλγόριθμοι βελτιστοποίησης βασισμένοι σε μετα-μοντέλα και στρατηγική αναζήτησης ισορροπίας μεταξύ ευρείας εξερεύνησης και στοχευμένης εκμετάλλευσης.

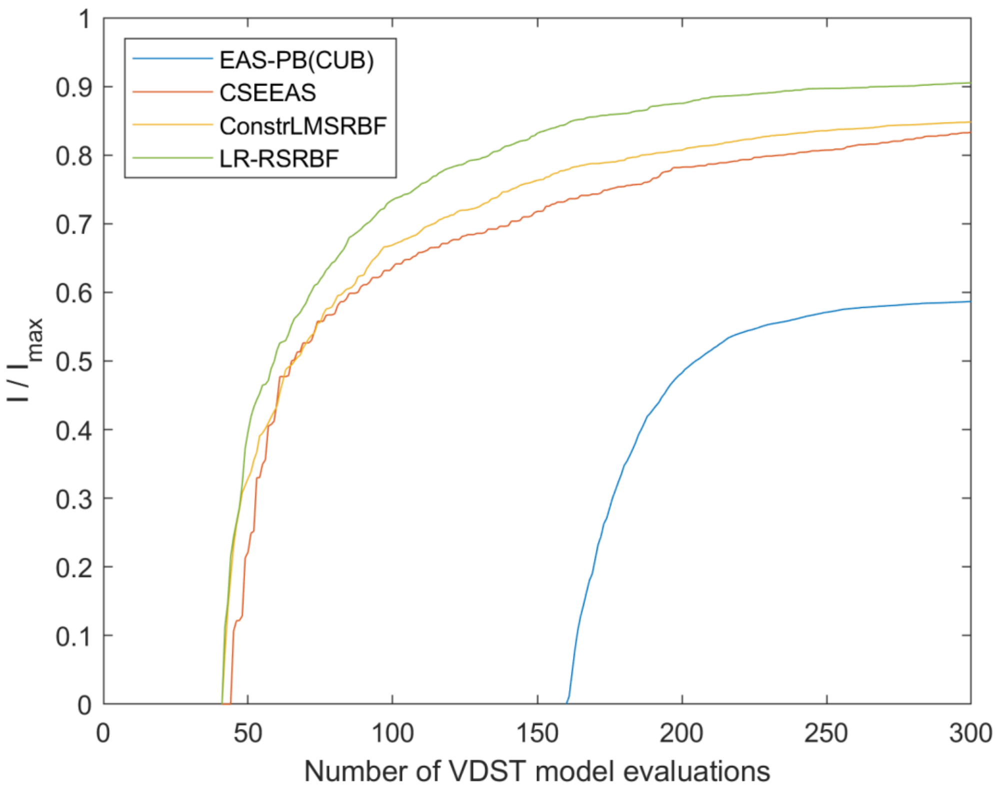
4. Βελτιστοποίηση βασισμένη σε μοντέλα υφαλμύρωσης πολλαπλής πιστότητας με χρήση της μεθοδολογίας co-Kriging.

Στο κεφάλαιο 5 πραγματοποιούνται εκτεταμένες συγκρίσεις των μεθοδολογιών που αναπτυχθήκαν, σε προβλήματα βελτιστοποίησης διαφορετικής διάστασης με χρήση υποθετικών μοντέλων παράκτιων υδροφορέων. Όλες οι συγκρίσεις βασίστηκαν σε πολλαπλές ανεξάρτητες δοκιμές επίλυσης της βελτιστοποίησης προκειμένου να αποδοθεί μια στατιστική σημαντικότητα στα αποτελέσματα. Οι αδυναμίες και οι δυνατότητες των μεθοδολογιών αναλύονται τόσο μέσω των στατιστικών χαρακτηριστικών του δείγματος όσο και με χρήση κριτηρίων που αξιολογούν την πορεία της επίδοσης των αλγορίθμων καθώς αυξάνονται τα δείγματα με το υπολογιστικά απαιτητικό μοντέλο. Τέλος, η βελτιστοποίηση βασισμένη σε μοντέλα πολλαπλής πιστότητας με χρήση της μεθοδολογίας co-Kriging, εφαρμόζεται ξεχωριστά υποθέτοντας σενάρια όπου η δυνατότητα προσθήκης νέων δειγμάτων είναι εξαιρετικά περιορισμένη. Ο συνδυασμός των μοντέλων πολλαπλής πιστότητας αφορά το μοντέλο μεταβλητής πυκνότητας και μεταφοράς ρύπου και το μοντέλο απότομης διεπιφάνειας κατά Strack (1976).

Βάσει των αποτελεσμάτων που παρουσιάστηκαν στο κεφάλαιο 5, ένα υποσύνολο των μεθοδολογιών με τις καλύτερες επιδόσεις αξιοποιήθηκε στο κεφάλαιο 6 για την εύρεση βέλτιστων παροχών άντλησης σε μοντέλο πραγματικού παράκτιου υδροφορέα, συγκεκριμένα στο Βαθύ Καλύμνου. Οι προτεινόμενες μεθοδολογίες παρείχαν λύσεις στην περιοχή του ολικού βέλτιστου όταν αυτές συγκρίθηκαν με την βέλτιστη λύση που υπολογίσθηκε με το αριθμητικό μοντέλο μεταβλητής πυκνότητας και μεταφοράς ρύπου.

Το κεφάλαιο 7 παρουσιάζει μια σύνοψη των αποτελεσμάτων από την εφαρμογή των προτεινόμενων μεθοδολογιών σε υποθετικούς και πραγματικούς υδροφορείς. Αναφέρονται οι προτάσεις για την εφαρμογή των μετα-μοντέλων στην διαχείριση παράκτιων υδροφορέων όπως προέκυψαν από την παρούσα έρευνα και συζητιούνται δυνατότητες περαιτέρω εμβάθυνσης στην εν λόγω ερευνητική περιοχή.

Όσον αφορά την συμβολή της παρούσας εργασίας στην επιστημονική περιοχή της διαχείρισης των υδατικών πόρων, αναφέρονται παρακάτω συνοπτικά τα κύρια σημεία πρωτοτυπίας αλλά και προτάσεις ή πρακτικές που προτείνονται ως αποτέλεσμα της συγκεκριμένης έρευνας. Ως κύριος στόχος της εργασίας είναι η ανάπτυξη νέων αποτελεσματικών σχημάτων προσομοίωσης-βελτιστοποίησης με χρήση μετα-μοντέλων, στην διαχείριση παράκτιων υδροφορέων. Σημαντικό βαθμό πρωτοτυπίας παρουσιάζει η μεθοδολογία LR-RSRBF (Local Refinement Random Search with RBF models) που αναπτύχθηκε στα πλαίσια της διατριβής. Η LR-RSRBF υλοποιεί μια αλγοριθμική διαδικασία προσαρμοστικής αναζήτησης ισορροπίας μεταξύ ευρείας εξερεύνησης και στοχευμένης εκμετάλλευσης, συνδυάζοντας αποτελεσματικά κριτήρια και βήματα προηγούμενων μεθοδολογιών βελτιστοποίησης με μετα-μοντέλα. Η ύπαρξη μη-γραμμικών περιορισμών στο πρόβλημα βελτιστοποίησης αντλήσεων που εξετάζεται, αντιμετωπίστηκε επιτυχώς με το εν λόγω σχήμα και την χρήση συναρτήσεων ακτινικής βάσης (RBF) ως μετα-μοντέλων. Προέκυψαν αξιόπιστες λύσεις στην περιοχή του ολικού βέλτιστου, ικανοποιώντας ταυτόχρονα την ανάγκη για μικρό αριθμό προσομοιώσεων με το υπολογιστικά απαιτητικό μοντέλο μεταβλητής πυκνότητας και μεταφοράς ρύπου. Επιπλέον η προτεινόμενη μεθοδολογία συγκρίνεται επαρκώς με δημοσιευμένες τεχνικές βελτιστοποίησης με χρήση μετα-μοντέλων που χαίρουν ευρείας αποδοχής στην αντίστοιχη επιστημονική περιοχή για προβλήματα που περιλαμβάνουν μη-γραμμικούς περιορισμούς (Γράφημα 1).



Γράφημα 1: Δειγματοληπτικός μέσος όρος (από 30 ανεξάρτητες δοκιμές βελτιστοποίησης) της σχετικής βελτίωσης της απόδοσης του αλγορίθμου έναντι του αριθμού προσομοιώσεων με το μοντέλο μεταβλητής πυκνότητας και μεταφοράς ρύπου. Όταν ο λόγος $r_I = I/I_{\max}$ πλησιάζει την τιμή 1 για όσο το δυνατόν μικρότερο αριθμό προσομοιώσεων, είναι ένδειξη ότι ο αλγόριθμος είναι αποτελεσματικός στην προσέγγιση λύσεων στην περιοχή του ολικού βέλτιστου και αποδοτικός όσον αφορά την ταχύτητα εντοπισμού τους. Το παραπάνω γράφημα αφορά πρόβλημα βελτιστοποίησης παροχών αντλήσεων 20 πηγαδιών, θεωρώντας την τιμή 300 ως τον μέγιστο επιτρεπτό αριθμό προσομοιώσεων με το μοντέλο μεταβλητής πυκνότητας και

μεταφοράς ρύπου. Η προτεινόμενη μεθοδολογία LR-RSRBF παρουσιάζει καλύτερη απόδοση σε σχέση με τους συγκρινόμενους αλγορίθμους στο συγκεκριμένο πρόβλημα που εξετάστηκε.

Σημαντικό βαθμό πρωτοτυπίας παρουσιάζουν επίσης τα σχήματα προσομοίωσης-βελτιστοποίησης που αναπτύχθηκαν στην παρούσα διατριβή για μαθηματικά μοντέλα υφαλμύρωσης πολλαπλής πιστότητας. Στην διεθνή βιβλιογραφία σε διάφορους κλάδους της μηχανικής, η σύνδεση δύο ή και περισσότερων μοντέλων διαφορετικής πιστότητας που προσομοιώνουν ένα σύστημα, παρουσιάζει αυξανόμενο ενδιαφέρον ιδιαίτερα για τις περιπτώσεις που η προσομοίωση με ένα αριθμητικό μοντέλο υψηλής πιστότητας είναι πολύ υψηλού υπολογιστικού κόστους. Η βελτιστοποίηση με μοντέλα πολλαπλής πιστότητας στηρίζεται στην θεώρηση πως μοντέλα φυσικής βάσης ή ακόμα και εμπειρικές σχέσεις χαμηλότερης πιστότητας αλλά και χαμηλότερου υπολογιστικού κόστους, μπορούν να παρέχουν χρήσιμη πληροφορία για την γενική απόκριση του συστήματος, παρά την έλλειψη ακρίβειας σε σχέση με το λεπτομερές αλλά και υπολογιστικά ακριβό μοντέλο φυσικής βάσης. Κατά αυτόν τον τρόπο, τα χαμηλότερης πιστότητας μοντέλα χρησιμοποιούνται για ένα μεγάλο αριθμό προσομοιώσεων ενώ ένας σημαντικά μικρότερος αριθμός προσομοιώσεων με το μοντέλο υψηλής πιστότητας διορθώνει τις αποκρίσεις των μοντέλων χαμηλής πιστότητας. Αυτή η προσέγγιση, που συνήθως εφαρμόζεται ως μια επαναληπτική διαδικασία, δημιουργεί ένα μετα-μοντέλο το οποίο ενσωματώνει γνώση από μοντέλα φυσικής βάσης διαφόρων επίπεδων πιστότητας. Θεωρείται ότι ως μέθοδος προσφέρει πλεονεκτήματα στις περιπτώσεις που οι διαθέσιμες προσομοιώσεις με το μοντέλο υψηλής πιστότητας είναι μη πρακτικό να υπερβούν έναν ορισμένο αριθμό ή δεν επαρκούν για την ανάπτυξη συμβατικών μεθόδων βελτιστοποίησης με μετα-μοντέλα.

Όσον αφορά στην βελτιστοποίηση πολλαπλής πιστότητας, υπάρχει ένας μικρός αριθμός δημοσιεύσεων στην επιστημονική περιοχή της διαχείρισης των υδατικών πόρων ενώ ειδικότερα στην διαχείριση παράκτιων υδροφορέων μια πρώτη παρουσίαση ανάλογων μεθόδων έγινε στις εργασίες Christelis and Mantoglou (2016) και Christelis and Mantoglou (2019). Στην παρούσα διατριβή παρουσιάζεται ένα νέο σχήμα βελτιστοποίησης με μοντέλα πολλαπλής πιστότητας που συνδυάζει το υψηλής πιστότητας μοντέλο μεταβλητής πυκνότητας και μεταφοράς ρύπου με το χαμηλότερης πιστότητας αλλά και υπολογιστικά μη απαιτητικό μοντέλο απότομης διεπιφάνειας κατά Strack (1976). Η προτεινόμενη μεθοδολογία αναπτύχθηκε σε ένα σχήμα προσαρμοστικής-αναδρομικής δειγματοληψίας αξιοποιώντας τις δυνατότητες της μεθοδολογίας co-Kriging για την ανάπτυξη μετα-μοντέλων. Τα αποτελέσματα που προέκυψαν είναι ιδιαίτερα ενθαρρυντικά για την χρήση της προτεινόμενης μεθοδολογίας σε προβλήματα βελτιστοποίησης αντλήσεων παράκτιων υδροφορέων με πολύ υψηλό υπολογιστικό κόστος.

Όπως φαίνεται στον Πίνακα 1, η μεθοδολογία AR-coKRG που προτείνεται στην παρούσα διατριβή παρείχε ικανοποιητικές λύσεις με εύρεση τοπικών ακρότατων για υπόθεση ακραίου σεναρίου όπου μόνο 21 προσομοιώσεις ήταν διαθέσιμες με το μοντέλο μεταβλητής πυκνότητας και μεταφοράς ρύπου. Τα δεδομένα υψηλής πιστότητας από το μοντέλο μεταβλητής πυκνότητας και μεταφοράς ρύπου συνδυάστηκαν με τα δεδομένα χαμηλότερης πιστότητας που προέκυψαν από έναν αριθμό 200 προσομοιώσεων με το μοντέλο απότομης διεπιφάνειας. Η μέθοδος συγκρίθηκε έναντι του αλγορίθμου ConstrLMSRBF που αναπτύσσει μετα-μοντέλα για ένα μόνο επίπεδο πιστότητας και θεωρείται ιδιαίτερα αποτελεσματικός στο πρόβλημα βελτιστοποίησης αντλήσεων παράκτιων υδροφορέων (Christelis and Mantoglou 2018). Τα αποτελέσματα δείχνουν πως για ένα σύνολο 30 ανεξάρτητων δοκιμών βελτιστοποίησης, ο δειγματοληπτικός μέσος όρος των λύσεων με την μεθοδολογία AR-coKRG υπερτερεί έναντι των άλλων προσεγγίσεων για περιορισμένο αριθμό προσομοιώσεων με το μοντέλο μεταβλητής πυκνότητας και μεταφοράς ρύπου.

Πίνακας 1: Συγκριτικά αποτελέσματα των βέλτιστων λύσεων μεταξύ της μεθοδολογίας AR-coKRG με αυτές που προέκυψαν από το χαμηλότερης πιστότητας μοντέλο απότομης διεπιφάνειας (EAS-SH) και τον αλγόριθμο ConstrLMSRBF που χρησιμοποιεί μετα-μοντέλα για ένα μόνο επίπεδο πιστότητας. Τα αποτελέσματα αφορούν πρόβλημα βελτιστοποίησης παροχών αντλήσεων 10 πηγαδιών. Οι καλύτερες λύσεις απεικονίζονται με μαύρο φόντο ενώ το «πραγματικό» βέλτιστο με το υψηλής πιστότητας μοντέλο μεταβλητής πυκνότητας και μεταφοράς ρύπου (EAS-VDST) αφορά 4967 προσομοιώσεις.

Optimization method	Worst	Best	Mean	StDev	VDST runs	Sharp runs	Time (hr)
EAS-VDST		<u>4857.5</u>			4967	NA*	14.45
EAS-SH		4049.2			NA*	4164	1.6
AR-coKRG^{CONS}	3966.3	**4815.1**	**4605.0**	187.38	21	200	0.56
ConstrLMSRBF	4004.8	4740.6	4397.3	205.94	21	NA*	0.095

* NA: Not Applicable

Ενδιαφέρον συμπέρασμα αυτής της εργασίας αποτελεί το ότι για έναν μέτριο αριθμό μεταβλητών απόφασης (π.χ. 10 πηγάδια άντλησης), η χρήση απλών μετα-μοντέλων, όπως οι συναρτήσεις ακτινικής βάσης κυβικού τύπου, με στρατηγική στοχευμένης εκμετάλλευσης ενσωματωμένη στις διαδικασίες εξελικτικού αλγόριθμου, αποδίδει λύσεις στην περιοχή του ολικού βέλτιστου μειώνοντας δραστικά τον υπολογιστικό χρόνο κατά 90-95%. Η ίδια στρατηγική αναπτύχθηκε και για την περίπτωση συνδυασμού πολλαπλών μετα-μοντέλων (συναρτήσεις ακτινικής βάσης και Kriging) που, κατά την γνώση των συγγραφέων, αποτελεί και την πρώτη εφαρμογή αυτής της μεθοδολογίας στην διαχείριση παράκτιων υδροφορέων (Christelis et al. 2019). Πρέπει επίσης να σημειωθεί ότι η παρούσα διατριβή παρουσίασε την αποτελεσματικότητα και τις συγκρίσεις των διαφόρων μεθοδολογιών βάσει πολλαπλών ανεξάρτητων δοκιμών βελτιστοποίησης η σημασία των οποίων έχει γενικότερα παραλειφθεί σε προηγούμενες μελέτες στο συγκεκριμένο αντικείμενο, παρά το ότι κρίνεται αναγκαία για την διερεύνηση της αμεροληψίας των αλγορίθμων ως προς την περιοχή εκκίνησης της διαδικασίας αναζήτησης λύσεων. Επιχειρήθηκε επίσης μια εκτενής ανασκόπηση των μεθοδολογιών βελτιστοποίησης με χρήση μετα-μοντέλων τόσο στους διάφορους κλάδους της μηχανικής όσο και από τον χώρο της διαχείρισης των υδατικών πόρων. Ως αποτέλεσμα αυτής της ανάλυσης, επιμέρους τεχνικές που υιοθετούνται ως state-of-the-art μελετήθηκαν και όσες από αυτές κρίθηκαν κατάλληλες για τα προβλήματα βελτιστοποίησης αντλήσεων εντάχθηκαν στην παρούσα έρευνα. Σημειώνεται ότι στην διατριβή δεν επιχειρείται κάποια εμβάθυνση από την σκοπιά της μαθηματικής περιγραφής των μετα-μοντέλων και της θεωρίας βελτιστοποίησης. Ο κύριος σκοπός της εργασίας είναι να διερευνηθούν οι δυνατότητες της βελτιστοποίησης με χρήση μετα-μοντέλων στα προβλήματα διαχείρισης παράκτιων υδροφορέων και στο πως εξειδικεύονται λαμβάνοντας υπόψη τις ιδιαιτερότητες της προσομοίωσης του φυσικού συστήματος της παράκτιας ροής.

Chapter 1

Introduction

1.1 Background

Groundwater is often the main resource of freshwater in many coastal regions around the world. This is particularly the case for arid and semi-arid areas due to low rainfall and limited availability of surface water bodies. Approximately 60% of the world's population resides within coastal zones and exploits aquifers for domestic water supply, agricultural and industrial purposes (Essink 2001). As coastal groundwater is in hydraulic contact with seawater, uncontrolled groundwater abstraction is considered as the main driver for triggering the landward movement of seawater (Ferguson and Gleeson 2012). This phenomenon, known as seawater intrusion (SWI), results in the reduction of freshwater volumes and in the contamination of pumping wells. To prevent and manage SWI in coastal aquifers, several mitigation actions have been proposed such as, freshwater injection barriers, artificial recharge mechanisms or groundwater abstraction control (Georgopoulou et al. 2001; Schwartz and Zhang 2003; Mantoglou et al. 2004; Pool and Carrera 2010). The present thesis focuses on the development of optimal groundwater abstraction strategies to control SWI by means of simulation-optimization methods.

Simulation-optimization techniques have long been used in coastal aquifer management to deliver sustainable groundwater extraction plans. A standard class of such applications is the pumping optimization of coastal aquifers under environmental constraints which protect the coastal groundwater resources. The main objective is to control the landward movement of seawater due to pumping while satisfying the demands for freshwater. The implementation of a simulation-optimization routine to address this management problem is based on the coupling of a coastal aquifer flow model with an optimization algorithm. There is a wide body of literature in coastal aquifer management which combines seawater intrusion models with a variety of optimization algorithms and different formulations of the objective functions and constraints. Many of those methods are summarized in the review papers of Singh (2014) and Sreekanth and Datta (2015). The formulation of such management problems involves two primary decisions:

1. To define the level of model fidelity which would be considered adequate to simulate seawater intrusion for the coastal aquifer under study.

2. To select an appropriate optimization algorithm which is capable to find a near global optimum, given the mathematical formulation of the pumping optimization problem.

The next two sections shortly discuss these decisions whereas more details are given in the following chapters.

1.2 The fidelity concept in seawater intrusion modelling

As is the case with many other complex physical systems, a variety of mathematical models is also available to simulate coastal aquifer flow. Not all of them describe the coastal aquifer processes in the same detail and a hierarchy can be defined based on their complexity and accuracy. Variable density flow and solute transport (VDST) numerical models simulate the density variability in space and the saltwater movement which constitute key features for the realistic prediction of SWI (Simmons 2005; Abarca et al. 2007; Dausman et al. 2010; Pool and Carrera 2011; Dokou and Karatzas 2012). Thus, VDST models represent a high-fidelity choice for simulating SWI. A short classification may set as low-fidelity SWI models those that neglect dispersion mechanisms but simulate saltwater movement (Essaid 1986) or incorporate density effects in coastal aquifer flow

(Bakker 2003). An additional simplification and thus an even low-fidelity level, can be introduced by those sharp interface models which assume static seawater (e.g., Strack 1976; Mantoglou et al 2004; Koussis et al. 2012). Obviously, depending on the specifications of each SWI modelling study, further refinements within the same fidelity level can be defined. For example, VDST models with a coarser discretization, two-dimensional instead of three-dimensional VDST models, steady-state flow modelling instead of transient flow, etc.

1.3 Problem statement and optimization algorithm selection

In the present thesis, focus is given on single-objective pumping optimization problems of coastal aquifers. Here, it is described as a nonlinear constrained optimization problem of the following form:

$$\begin{aligned}
&\min f(\mathbf{x}) \\
&s.t.\ g_i(\mathbf{x}) \le 0,\ i = 1,\ldots,k \\
&l_b \le \mathbf{x} \le u_b,\ \mathbf{x} \in \mathbb{R}^k
\end{aligned} \tag{1.1}$$

where f, g_i represent the deterministic objective function and the ith nonlinear inequality constraint function, respectively. The decision variable vector $\mathbf{x}$ takes values in the $k-$dimensional continuous space $[l_b, u_b] \subset \mathbb{R}^k$. A real vector $\mathbf{x}^*$ is sought so that $f(\mathbf{x}^*) = \min f(\mathbf{x})$ subject to ($s.t.$) the constraints defined in problem (1.1). It is assumed that the derivatives of f and g_i are not available, the bound constraints l_b and u_b define the search space of the optimization problem while the set of the inequality constraints $g_i,\ i = 1,\ldots,k$ define the feasible solution space.

Nonlinear programming methods, such as sequential quadratic programming (SQP), that directly handle the constraint functions, have been applied in past coastal aquifer management studies. However, it has been observed that they might easily get trapped in local optima (Mantoglou et al. 2004). Pumping optimization problems of coastal aquifers, have been generally identified as a non-convex optimization problem with multiple local optima and the relevant literature suggests the use of global optimizers based on evolutionary principles (Ketabchi and Ataie-Ashtiani 2015). Thus, evolutionary algorithms have been increasingly used for coastal aquifer management problems although they require much larger number of objective function evaluations compared to the gradient-based optimizers (Mantoglou et al. 2004; Sreekanth and Datta 2015). Nevertheless, and depending on the optimization problem complexity and dimensionality, some evolutionary algorithms may increase the computational cost whilst failing to locate near global optimal solutions for SWI management problems (Karpouzos and Katsifarakis 2013; Ketabchi and Ataie-Ashtiani 2015).

1.4 Motivation

It is acknowledged that the combination of VDST models with evolutionary algorithms can potentially deliver a high-fidelity optimization outcome (Sreekanth and Datta 2015; Ketabchi and Ataie-Ashtiani 2015). However, the computationally burdensome runtimes of VDST simulations hinder the application of standard simulation-optimization routines for coastal aquifer management. The optimization task can be further challenged by the presence of many decision variables and nonlinear constraints which in turn increase the function evaluations required by the evolutionary algorithm to converge. As a result, solving a pumping optimization problem of coastal aquifers may lead to a computational cost of many hours, days or even months in the case of extremely time-consuming simulations.

A pragmatic approach to confront the computational cost in such cases, is to employ fast approximation models, commonly called surrogate models or metamodels, to search for promising solutions in the decision variable search space. Surrogate models are initially trained with input-output data obtained from simulations with the physics-based model. This is an essential step where the metamodels acquire a certain accuracy to predict outputs of the original model to unseen input data. Then, the metamodels may be used in lieu of the computationally expensive physics-based model to explore the search space. However, due to limitations on the computational cost, it is unlikely that a single initial training set will provide a globally accurate surrogate model. Thus, a framework should be designed to efficiently sample the physics-based model at additional points that are considered informative for updating the surrogate models while keeping the number of model simulations at a minimum. This framework is commonly implemented in an iterative fashion and constitutes the basic structure of an online surrogate-based optimization (SBO) method. The advantages of SBO have been well-documented and acknowledged in engineering optimization literature (Forrester and Keane 2009; Asher et al. 2015; Bhosekar and Ierapetritou 2018; Yondo et al. 2018). In general, a fair metamodel accuracy combined with an effective sampling strategy can possibly steer the SBO algorithm to locate near optimal solutions in a fraction of the time required with the costly physics-based model. Nevertheless, every engineering optimization problem involves unique features which may favor the use of specific surrogate modelling approaches or their use might not even recommended at all (Razavi et al. 2012a).

Although SBO has gained increased interest in coastal aquifer management literature during the last 15 years, there are still techniques and approaches which are largely unexplored. The present thesis addresses specific issues in the implementation of SBO methods for the problem of pumping optimization of coastal aquifers. These are mainly focused on the following two aspects of SBO methods and set the motivation for this work.

First, the application and development of SBO frameworks which balance exploration and exploitation has not received the required attention in coastal aquifer management. A more crude and aggressive approach evaluates a candidate solution with the physics-based model only at the current optimum located by the surrogate models. This strategy aims to quickly improve the accuracy of the metamodel in the optimum region (exploitation step). However, the global improvement of the metamodel is neglected and depending on the problem at hand, the search might get stuck in local optima (Forrester et al., 2008). Evaluating certain points with the physics-based model away from the current optimum (exploration step) can improve the global accuracy of the metamodel and it can potentially locate better solutions as new samples are added. Based on findings in the engineering optimization literature, such methods can converge to high quality solutions using a relatively small number of evaluations with the physics-based model. This is particularly desirable due to the generally time-consuming VDST simulations. Therefore, the present thesis focused on the development and application of SBO methods that balance exploration with exploitation and investigated their performance against less comprehensive infill strategies.

Second, there is a gap in coastal aquifer management research regarding multi-fidelity optimization methods. There is often a variety of mathematical models that describe a physical process at different levels of fidelity. The idea pursued by multi-fidelity methods in general, is to exploit information from available low-fidelity models which usually are much faster than the computationally intensive high-fidelity model. Multi-fidelity optimization is considered advantageous over the conventional SBO methods when only a few high-fidelity simulations can be conducted in a manageable computational time. In those cases, either the number of simulations might not be sufficient to train a surrogate model or the available high-fidelity simulations after the initial surrogate model training, are too few to successfully apply a standard SBO framework. Several models of lower fidelity have been presented in the literature to simulate coastal aquifer flow in a computationally affordable manner (e.g. Strack 1976; Essaid 1986; Bakker 2003; Mantoglou 2003; Bakker 2006; Pool and Carrera 2011; Koussis et al. 2012; Koussis et al. 2015; Lu et al. 2015; Lu et al. 2016; Werner 2017). Yet, only a few papers have combined low-fidelity (LF) and high-fidelity (HF) data to solve pumping optimization problems of coastal aquifers (e.g. Christelis and Mantoglou 2016; Christelis and Mantoglou 2017; Christelis and Mantoglou 2019; Dey and Prakash 2020). In the present thesis, a new multi-fidelity optimization framework is developed for coastal aquifer management to tackle hypothetical cases where only limited high-fidelity data can be obtained (i.e., less than 30 simulations available with the VDST model).

1.5 Key objectives, assumptions, and thesis overview

The main aim of the present thesis is to develop new SBO methods that provide a computationally affordable route to high-fidelity solutions for pumping optimization problems of coastal aquifers. Furthermore, it is anticipated that some conclusions and methodologies from the present analysis can be generalized and be applied for similar computationally expensive optimization problems in groundwater management. The accomplishment of the above objectives is pursued through:

- The development of efficient and effective SBO methodologies using appropriate surrogate models and sampling strategies.

- The development of SBO methods which provide near optimal solutions using a relatively small number of VDST simulations which is convenient for real-world pumping optimization problems.

- The exhaustive comparison of the various SBO methods to identify limitations and advantages depending on the dimensionality and the characteristics of the present optimization problem.

The analysis conducted in this thesis involves hypothetical as well as real-world coastal aquifer case studies. The seawater intrusion modelling and the optimization methods were based on the following assumptions:

- Seasonal variations of recharge are not considered in the coastal aquifer simulations.
- The parameter uncertainty in coastal aquifer simulations is not considered.
- Constant pumping rates over the specified management plan.
- Two levels of model fidelity are employed to simulate seawater intrusion.
- The seawater intrusion simulation models will not crash for any input $\mathbf{x} \in [l_b, u_b]$.

The thesis is organized as follows. Chapter 2 presents the mathematical modelling of seawater intrusion for VDST and sharp interface models. In chapter 3, previous works in coastal aquifer management are discussed and the mathematical formulation of the present optimization problem is presented, along with the specifications related to each simulation model. Chapter 4 includes a detailed discussion of metamodelling in the broader engineering optimization literature as well as in the field of coastal aquifer management. Furthermore, the chosen surrogate models are presented along with the details of the SBO frameworks implemented in this thesis. To the best of our knowledge, the proposed SBO algorithms are presented for the first time in coastal aquifer management. In chapter 5, we demonstrate and discuss the results based on a thorough comparison of the various SBO methods under alternative computational budgets (the total number of runs with the VDST model) and different dimensionalities of the pumping optimization problem (the number of decision variables). The main issues which are discussed include:

1) The importance of choosing appropriate sampling strategies as the dimensionality of the optimization problem increases.

2) To what extent the sophistication of the developed surrogate models affects the optimization outcome.

3) The choice of approximating each nonlinear constraint function using an individual surrogate model over that of a single surrogate model of the penalized objective function and what is the impact on the efficiency and effectiveness of the SBO methods.

4) How multi-fidelity optimization is compared to conventional SBO methods when only a few VDST model runs are available.

In chapter 6, the best SBO methods, as identified from the comparisons conducted in chapter 5, are employed to solve a real-world pumping optimization problem for a coastal aquifer in the Greek island of Kalymnos. Finally, chapter 7 discusses the main conclusions of this thesis and presents some thoughts for further research and future applications of the developed SBO methods.

Chapter 2

Seawater intrusion modelling

2.1 Conceptual models

The mathematical formulation of SWI models is mainly based on two different conceptualizations of the freshwater/saltwater interactions. As discussed previously, the model fidelity is also related to these conceptual models. Accordingly, the high-fidelity (HF) modelling approach aims to a more realistic representation of coastal aquifer flow where between the two fluids a transition zone is formed and is controlled by hydrodynamic dispersion mechanisms. Within this zone the fluid density and salt concentration gradually vary from that of freshwater to seawater. The lower-fidelity (LF) modelling approach is a simplification of the physical system where the dispersion zone is idealized as a sharp interface. It is a reasonable approximation in regional coastal aquifers where the dispersive zone is narrow compared to the scale of the problem (Mantoglou et al. 2004). The sharp interface approximation can be either formulated on the assumption that only flow in the freshwater zone is modelled (one-fluid approach) or on the two-fluid approach where a coupled system of flow equations in the salt and fresh water zones is solved (Essaid 1986). They are both simplifications of the dispersive flow, as simulated by the VDST models, however, the one-fluid approach is mostly limited to reproduce long term responses in coastal aquifer systems. Figure 2-1 demonstrates the two conceptual approaches for modelling seawater intrusion in coastal aquifers.

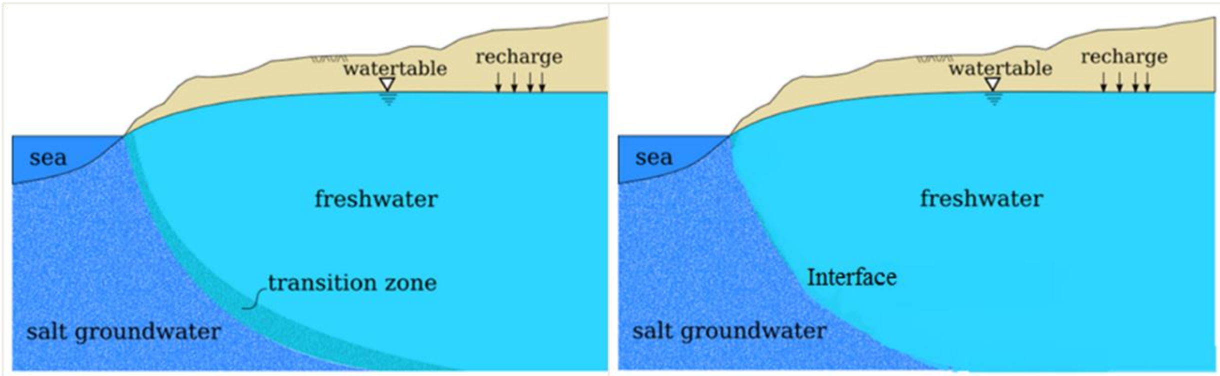


Figure 2-1 A typical unconfined coastal aquifer under natural undisturbed conditions considering a dispersive zone (left) and a sharp interface (right).

2.2 Mathematical modelling of seawater intrusion

In this thesis, the VDST models and the one-fluid sharp interface models based on the flow potential formulation of Strack (1976) are utilized. The two modelling approaches have been widely used to study several aspects of coastal aquifer processes. Some examples are:

- The influence of hydraulic and transport properties on SWI (e.g. Dagan and Zeitoun 1998; Simmons et al. 2001; Al-Bitar and Ababou 2005; Abarca et al. 2007; Kerrou and Renard 2010; Chang and Yeh 2010; Walther et al. 2017).

- The impacts of climate change on coastal aquifer systems for various sea-level rise scenarios (Bobba 1993; Sherif and Singh 1999; Werner and Simmons 2009; Watson et al. 2010; Chang et al. 2011; Webb and Howard 2011; Mazi et al. 2013; Yang et al. 2015).

- The response of coastal aquifers to seasonal variations of groundwater recharge and pumping (e.g. Mahesha and Nagaraja 1996; Paniconi et al. 2001; Michael et al. 2005; Gingerich and Voss 2005; Prieto et al. 2006; Kopsiaftis et al. 2009; Mazi et al. 2014; Kopsiaftis et al. 2017).

The following sections present in more detail the mathematical formulation of these SWI models.

2.2.1 *Variable density flow and solute transport models*

A coupled system of nonlinear partial differential equations governs the mathematical description of the transition/mixing zone between freshwater and saltwater (Pool and Carrera 2011). The solution of this coupled system of equations is obtained by applying numerical models and a detailed analysis of the mathematical formulation can be found in Kolditz et al. (1998). Here, a brief description of the governing equations is presented as per Frind (1982):

$$\frac{\partial}{\partial x_i}\left[K_{ij}\left(\frac{\partial h_f}{\partial x_j}+\rho_r n_j\right)\right]+Q_\rho = S_s\frac{\partial h_f}{\partial t} \tag{2.1}$$

$$\frac{\partial}{\partial x_i}\left(D_{ij}\frac{\partial c}{\partial x_j}\right)-\frac{\partial}{\partial x_i}\left(v_i c\right)+Q_c = \frac{\partial\left(c\right)}{\partial t} \tag{2.2}$$

In flow Equation (2.1), the variable $h_f\ [L]$ is the equivalent freshwater head defined as:

$$h_f = \left(p/\rho_f g\right)+z \tag{2.3}$$

where $p\left[ML^{-1}T^{-2}\right]$ is the fluid pressure, $\rho_f\left[ML^{-3}\right]$ is the reference freshwater fluid density while ρ_r is the relative density defined as:

$$\rho_r = \frac{\rho}{\rho_f}-1 \tag{2.4}$$

Also, $g\left[LT^{-2}\right]$ is the gravity acceleration constant and $z\ [L]$ is the elevation above horizontal datum. $K_{ij}\left[LT^{-1}\right]$ are the coefficients of the freshwater hydraulic conductivity tensor, Q_ρ

$\left[(L^3T^{-1})\,L^{-3}\right]$ is a volumetric fluid source/sink term per unit aquifer volume, $t\left[T\right]$ is time, $n_j=1$ indicates the vertical direction, $n_j=0$ indicates the horizontal directions and $S_s\left[L^{-1}\right]$ is the specific storage. Under isothermal conditions and neglecting viscosity effects, fluid density depends only on concentration and this relation can be defined as:

$$\rho=\rho_f\left(1+\varepsilon c\right) \tag{2.5}$$

where $\rho\left[ML^{-3}\right]$ is the fluid density, $\varepsilon=\left(\rho_{\max}/\rho_f\right)-1$ is a constant derived from the maximum fluid density $\rho_{\max}$ (seawater density) and c is the relative concentration (dimensionless) which varies between 0 and 1 for fluid densities varying from ρ_f to $\rho_{\max}$.

Similarly, the transport equation (2.2), is conveniently written in terms of the dimensionless relative concentration $c\left[-\right]$ where D_{ij} $\left[L^2T^{-1}\right]$ are the coefficients of the dispersion tensor and $Q_c\left[(L^3T^{-1})\,L^{-3}\right]$ is a solute source/sink term per aquifer volume. The fluid velocity v_i $\left[LT^{-1}\right]$ is defined as:

$$v_i=\frac{q_i}{\nu} \tag{2.6}$$

where ν $\left[-\right]$ is the porosity while the Darcy flux $q_i\left[LT^{-1}\right]$ is expressed in the case of variable-density flow as:

$$q_i=-K_{ij}\left(\frac{\partial h_f}{\partial x_j}+\rho_r n_j\right) \tag{2.7}$$

Equation (2.7) describes the relation of fluid flow and solute transport based on the Boussinesq approximation where the only changes in density appear in the buoyancy term ρg of Darcy's flow in the z-direction.

Some examples of numerical codes that can simulate variable-density flow and solute transport are SUTRA (Voss, 1984), CODESA-3D (Gambolati et al. 1999), SEAWAT (Guo and Langevin 2002), FEFLOW (Diersch and Kolditz 2002), HydroGeoSphere (Therrien, et al., 2006), etc. In the present thesis the academic version of HydroGeoSphere code (HGS) was used to simulate seawater intrusion. The HGS code uses the control volume finite element method to solve the above system of partial differential equations and employs a Picard iteration scheme with adaptive time-stepping to cycle between the solutions of the flow equation and the transport equation (Thompson et al. 2007).

It is noted that VDST modelling is a computationally expensive task, as fine spatial and time discretizations are required for accurate representation of the flow and transport processes. To take full advantage of the simulation capabilities of VDST models, plenty of field data are required and together with the associated computational cost, their application in regional-scale coastal aquifer models is not straightforward (Sanford and Pope 2010). Nevertheless, VDST models have been successfully used in the past for gaining insights in real-world coastal aquifers (e.g. Gingerich and Voss 2005; Kopsiaftis et al. 2009; Kerrou et al. 2013; Giambastiani et al. 2017).

2.2.2 Sharp interface models

In the present thesis, the sharp interface model based on the single-potential formulation of Strack (1976) is also used to simulate SWI as a low-fidelity model. It is based on the Ghyben-Herzberg relation and Dupuit approximation and neglects density variability in space as well as mixing between freshwater and saltwater. The saltwater is assumed static and aquifer flow is assumed horizontal and steady-state. The depth of the interface is estimated using the Ghyben-Herzberg approximation which assumes that horizontally flowing freshwater floats above static saltwater (Essaid 1986b).

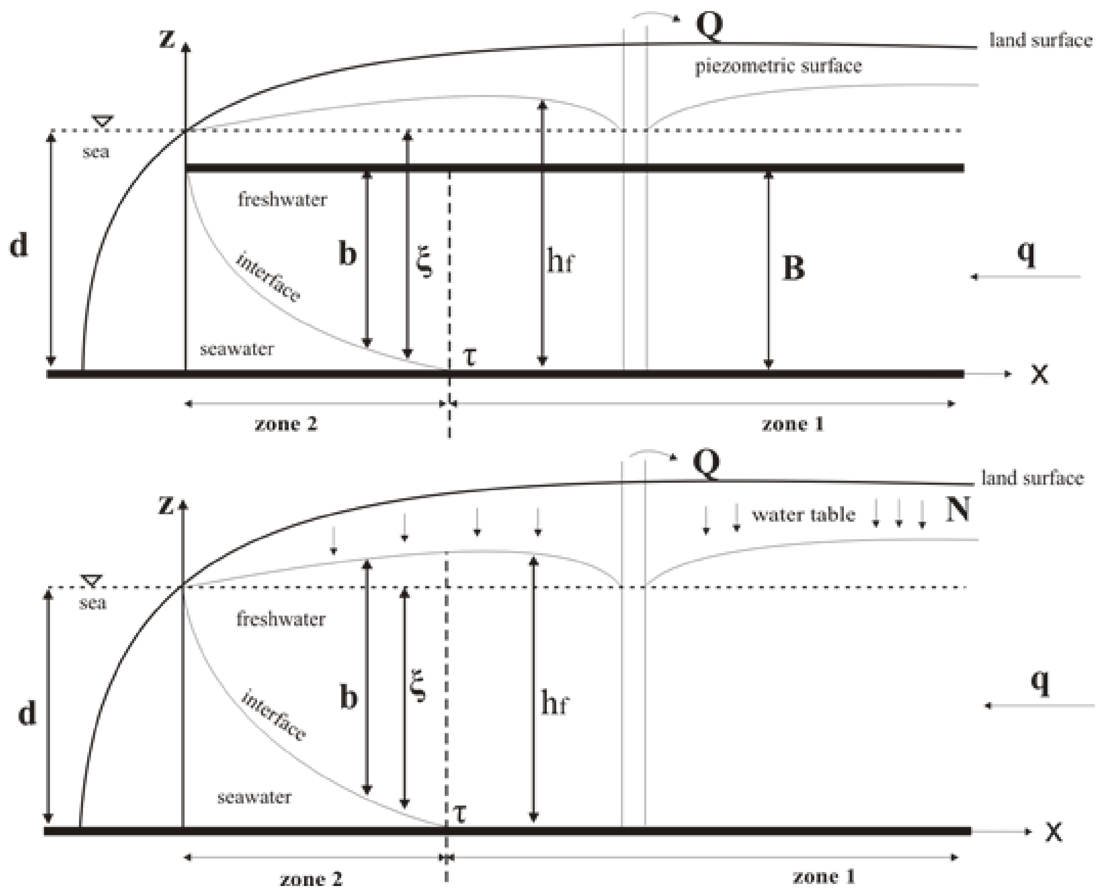


Figure 2-2 Schematic vertical cross-sections of a confined (upper figure) and an unconfined (lower figure) coastal aquifer, based on the sharp interface approximation.

As shown in Figure 2-2, two vertical cross-sections for both confined and unconfined coastal aquifers are presented, and a sharp interface is assumed to separate freshwater from saltwater. Two distinct zones are developed in both coastal aquifer types. In zone 1, the aquifer behaves as a confined (upper view) or as an unconfined (lower view) while in zone 2 a freshwater lens floats above the static saltwater layer. In zone 1, fresh groundwater is pumped by fully penetrating pumping wells. Variable $d\ [L]$ represents the depth from sea level to the aquifer base and variable $\xi(x,y)\ [L]$ is the freshwater depth from sea level to the interface. Point τ indicates the point where the interface intersects the base of the coastal aquifers. It is usually called “toe” of seawater wedge and comprises a typical measure of the extent of seawater intrusion.

A freshwater discharge volume rate per unit width of aquifer $q\ \left[L^3/LT\right]$, recharges the aquifer from the east inland boundary. Variable $b(x,y)\ [L]$ is the thickness of the confined flow region. In zone 1, $b = B$ where $B\ [L]$ is aquifer thickness defined by the two confining boundaries, while In zone 2, $b(x,y) = \xi(x,y) - (d - B)$. Variable $h_f(x,y)\ [L]$ is the freshwater piezometric head with reference to the impermeable aquifer base. In the case of the unconfined aquifer, a groundwater recharge rate $N\ \left[LT^{-1}\right]$ replenishes the aquifer. Variable $b(x,y)\ [L]$ is the total freshwater depth where in zone 1, $b = h_f$ and in zone 2, $b(x,y) = h_f - d + \xi(x,y)$, where $h_f(x,y)$ $[L]$ is the freshwater head with reference to the impermeable aquifer base. The Ghyben-Herzberg relation links the hydraulic head $h_f(x,y)$ and depth $\xi(x,y)$ via the saltwater/freshwater density ratio $\varepsilon = (\rho_s - \rho_f)/\rho_f$ according to $(1/\varepsilon)(h_f - d) = \xi(x,y)$.

The following differential equations govern the steady-state flow applicable for both zones of the aquifer (Strack 1976; Mantoglou et al. 2004):

$$\left.\begin{aligned}&\frac{\partial}{\partial x}\left(K\frac{\partial \phi}{\partial x}\right)+\frac{\partial}{\partial y}\left(K\frac{\partial \phi}{\partial y}\right)-Q(x,y)=0, && \text{confined interface flow}\\ &\frac{\partial}{\partial x}\left(K\frac{\partial \phi}{\partial x}\right)+\frac{\partial}{\partial y}\left(K\frac{\partial \phi}{\partial y}\right)+N-Q(x,y)=0, && \text{unconfined interface flow}\end{aligned}\right\} \quad (2.8)$$

where $\phi\left[L^2\right]$ is the flow potential and $K\left[LT^{-1}\right]$ is the aquifer's hydraulic conductivity. The distributed pumping rate $Q(x,y)\left[(L^3T^{-1})L^{-2}\right]$ is $Q(x,y)=\sum_{j=1}^{M} Q_j\,\delta\left(x-x_{wj},y-y_{wj}\right)$ where (x_{wj},y_{wj}) are the coordinates of pumping wells $j=1,\ldots M$ with rates Q_j and $\delta\left(x-x_{wj},y-y_{wj}\right)$ is the Dirac delta function. In the case of confined aquifers, the flow potential is defined as (Bear et al. 1999):

$$\left.\begin{aligned}&\phi=Bh_f+\frac{\varepsilon B^2}{2}-(1+\varepsilon)Bd && \text{zone 1}\\ &\phi=\frac{1}{2\varepsilon}\left[h_f+\varepsilon B-(1+\varepsilon)d\right]^2 && \text{zone 2}\end{aligned}\right\} \quad (2.9)$$

while in the case of unconfined aquifers the flow potential is expressed as (Bear et al. 1999),

$$\left.\begin{aligned}&\phi=\frac{1}{2}\left[h_f^2-(1+\varepsilon)d^2\right], && \text{zone 1}\\ &\phi=\frac{(1+\varepsilon)}{2\varepsilon}\left(h_f-d\right)^2, && \text{zone 2}\end{aligned}\right\} \quad (2.10)$$

At the location of the toe, the flow potential is calculated based on the following equations for each coastal aquifer type (Mantoglou 2003):

$$\left.\begin{aligned}\phi_{toe} &= \frac{\varepsilon}{2}B^2, && \text{confined aquifer}\\ \phi_{toe} &= \left[\frac{\varepsilon(\varepsilon+1)}{2}\right]d^2, && \text{unconfined aquifer}\end{aligned}\right\} \tag{2.11}$$

The sharp interface model of Strack (1976) has been widely used to simulate coastal aquifer flow due to its simplicity and low computational cost. The flow equations presented above can be easily solved numerically using a groundwater flow code with a rather coarse spatial discretization. However, the sharp interface assumption might introduce significant errors in the estimation of the extent of seawater intrusion under pumping conditions (Dausman et al. 2010; Christelis and Mantoglou 2013; Llopis-Albert and Pulido-Velazquez 2014; Koussis et al. 2015). On the other hand, VDST models usually provide a benchmark estimation of the SWI extent when pumping is present since they are considered more accurate (high-fidelity) models.

To improve the accuracy of sharp interface models, recent studies have proposed correction formulas to mitigate the overestimation of SWI by implicitly incorporating dispersion effects (e.g. Pool and Carrera 2011; Koussis et al. 2015; Lu and Werner 2013; Werner 2017; Koussis and Mazi 2018). A particular correction is the one proposed by Pool and Carrera (2011) who developed an empirical equation for the sharp interface model of Strack (1976). Practically, a modified density ratio for the sharp interface model is calculated, based on the aquifer thickness B and the transverse dispersivity value a_T as follows:

$$\varepsilon^* = \varepsilon\left[1-\left(\frac{a_T}{B}\right)^{1/6}\right] \tag{2.12}$$

where ε^* denotes a modified saltwater-freshwater density ratio. Lu and Werner (2013) conducted a series of numerical experiments with VDST models and they suggested that the exponent in

Equation (2.12) should be replaced with $1/4$. The above correction of the density ratio has motivated many studies to search for improved versions of correcting the sharp interface model and match the salinity profiles of VDST models. Procedurally speaking, the density ratio is reduced in Equation (2.12) and thereby the toe location is moved seawards allowing for larger extraction of groundwater volumes. As it will be discussed later, this concept can be effectively implemented through a dynamic adjustment of the density ratio which is more suitable for problems of pumping optimization (Christelis and Mantoglou 2016a).

Chapter 3

Simulation-optimization in coastal aquifer management

3.1 Previous pumping optimization studies

As mentioned before, simulation optimization methods in coastal aquifer management require the coupling of a SWI model with an optimization algorithm (Figure 3-1).

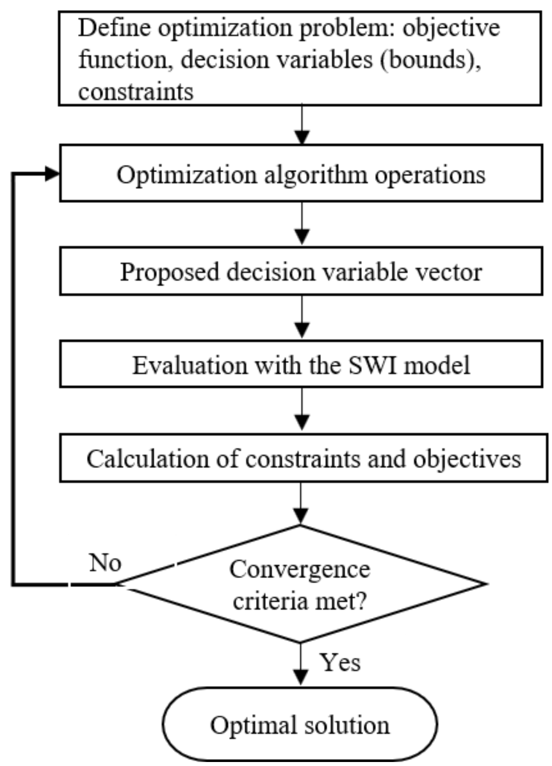


Figure 3-1 Typical workflow for coupling an optimization algorithm with a SWI model.

In brief, the optimization algorithm calls the SWI model to evaluate the decision variable vector (e.g., pumping rates) and return the simulated state variables (e.g., hydraulic heads, salinity concentrations). Then, based on the values of the state variables at certain points of interest, the constraint and the objective functions are calculated. This procedure is repeated many times until specific convergence criteria are met, as defined by the operation rules of the optimization algorithm.

Several formulations of optimization problems have been adopted by using SWI models of various fidelities (e.g. Katsifarakis and Petala 2006; Ferreira da Silva and Haie 2007; Karterakis et al. 2007; Uddameri and Kuchanur 2007; Kacimov et al. 2009; Sedki and Ouazar 2011; Doulgeris and Zissis 2014; Karatzas and Dokou 2015). The sharp interface model based on the Strack's (1976) single-potential formulation, has been coupled with nonlinear programming methods as well as evolutionary algorithms to calculate optimal pumping rates. Cheng et al. (2000), extended the single-well analysis of Strack (1976) for the case of multiple wells and they calculated optimal pumping rates using a genetic algorithm. Mantoglou (2003) used the method of images and the single-potential formulation to develop analytical solutions for coastal aquifers of finite size. The SQP algorithm was employed to solve a nonlinear pumping optimization problem with inequality constraints to control SWI. Mantoglou et al. (2004) applied a numerical solution for the sharp interface model of Strack (1976) to account for spatial variations of hydraulic conductivity and recharge and solved a pumping optimization problem. They also compared the performance of SQP and genetic algorithms. They found that using a genetic algorithm, provided better optimal solutions than the SQP method, at the expense of increased computational cost.

In a follow-up paper, Mantoglou and Papantoniou (2008) used the same numerical solution of the sharp interface model to find optimal pumping rates and optimal well locations based on a hybrid scheme with evolutionary algorithms and SQP. Ataie-Ashtiani and Ketabchi (2011) utilized a numerical solution of the sharp interface approximation and investigated the applicability of ant-colony optimization algorithms to solve coastal aquifer management problems. Christelis et al. (2012) solved a pumping optimization problem of coastal aquifers assuming different statistical properties of hydraulic conductivity random fields. They used SQP and the model of Strack (1976) to reduce the computational cost within a Monte-Carlo based optimization framework. Kourakos and Mantoglou (2015) developed an efficient coupling scheme between the same sharp interface model and an evolutionary algorithm to enable fast calculation of optimal pumping rates for a real-world coastal aquifer management problem.

Although the sharp interface models have been a popular choice in coastal aquifer management, the corresponding studies based on VDST models are limited mostly due to the increased requirements in computational resources. Das and Datta (1999) explored different SWI management models, based on VDST simulations, and demonstrated the complexity of attaining optimal solutions under various objective function formulations. Qahman et al. (2005) coupled VDST models with genetic algorithms to solve hypothetical coastal aquifer management problems considering different formulations of the objective functions. A strategy to control SWI while reducing the cost of operating pumping wells was proposed in Abd-Elhamid and Javadi (2011). They coupled a transient VDST model with a genetic algorithm. Javadi et al. (2015) developed a multi-objective management model based on VDST models to investigate the effectiveness of combined strategies for SWI mitigation.

Regardless of the management model or the selection of the optimization algorithm, the simulation-optimization routines developed in the relevant literature demonstrate their efficacy to cope with the complex decision-making problems in coastal aquifer management. The unavoidable uncertainties and limitations related to the application of SWI models for regional coastal aquifer systems may hinder the accuracy of the optimal designs obtained through simulation-optimization.

Nevertheless, it is rather unrealistic to explore systematically the best actions to control SWI without the use of simulation-optimization methods.

3.2 Definition of the pumping optimization problem

VDST and sharp interface models do not share the same physics and thus they differ in terms of input parameters and output variables. VDST simulations provide a salinity concentration field for the calculation of SWI. On the contrary, the output from the sharp interface model of Strack (1976) is a single-potential flow field which is used to define the sharp interface location and thereby to calculate SWI (Mantoglou 2003; Mantoglou et al. 2004). Hence, the optimization problem is presented separately for each SWI model. The VDST-based optimization is mathematically defined as follows (Kourakos and Mantoglou 2009):

$$\begin{aligned} & \min -\sum_{i=1}^{k} Q_i \\ & s.t.\ C_i\left(Q_1,\dots,Q_k\right) \le C_t, \forall\, i=1,\dots,k \\ & Q_{\min} \le Q_i \le Q_{\max} \end{aligned} \tag{3.1}$$

or by expressing the constraint functions in terms of iso-salinity contours,

$$\begin{aligned} & \min -\sum_{i=1}^{k} Q_i \\ & s.t.\ x_i^{C_t}\left(Q_1,\dots Q_k\right) \le xw_i, \forall\, i=1,\dots,k \\ & Q_{\min} \le Q_i \le Q_{\max}, i=1,\dots,k \end{aligned} \tag{3.2}$$

The overall goal is to maximize (the reason for the negative sign in the objective function) the total groundwater extraction, subject to constraints that do not allow the salinity levels in pumped groundwater to exceed a specified salinity concentration threshold C_t. Based on that, the salinity concentration C_i in (3.1) for each pumping well is not allowed to exceed a maximum concentration of $C_t = 0.1 kg/m^3$ which was chosen as an acceptable limit for drinking water according to World Health Organization guidelines (1996). The objective of pumping maximization is similarly achieved with the alternative expression of constraints in (3.2), where the variable xw_i is the pumping well location and $x_i^{C_t}$ represents the horizontal distance of the iso-salinity C_t from the coast, as a function of the pumping rates.

With $\mathbf{Q}$ representing the decision vector of pumping rates $\mathbf{Q} = (Q_1, ..., Q_k)$, the objective function $f(\mathbf{Q}) = -\sum_{i=1}^{k} Q_i$ is linear in respect to the decision variables Q_i, $i = 1, ..., k$, that is, the individual pumping rates. $Q_{\min}$ and $Q_{\max}$ define the lower and upper bounds of pumping rates, respectively. In the following sections and depending on the settings of the developed simulation-optimization methods, either (3.1) or (3.2) can be used to solve the pumping optimization problems. The corresponding optimization formulation for the sharp interface models is:

$$\begin{aligned} &\min -\sum_{i=1}^{k} Q_i \\ &s.t.\ \ x_i^{toe}\left(Q_1, ..., Q_k\right) < xw_i ,\forall\, i = 1, ..., k \\ &\qquad Q_{\min} \le Q_i \le Q_{\max} , i = 1, ..., k \end{aligned} \tag{3.3}$$

where the set of the constraint functions here do not allow the “toe” of the interface x^{toe} to reach the pumping wells (Mantoglou et al. 2004). The variable x_i^{toe} is the horizontal distance of the toe from the coast, as a function of the pumping rates. The inequality constraints defined in problems (3.1) and (3.2) are nonlinear due to the inherent nonlinear equations involved in the VDST model formulation (Dhar and Datta 2009). The nonlinearity of the optimization problem (3.3) is due to the nonlinear relationship between the pumping rates and the variable x^{toe} (Mantoglou 2003; Mantoglou et al. 2004).

3.3 Constraints handling with evolutionary algorithms

As discussed in the introduction part of this chapter, evolutionary algorithms are preferred over the gradient-based optimizers in coastal aquifer management studies, due to their ability to handle the presence of multiple local optima (Ketabchi and Ataie-Ashtiani 2015). Typically, in evolutionary optimization, the nonlinear constraints are embedded into the objective function using penalty terms. Therefore, the optimization problems defined in (3.1), (3.2) and (3.3) are translated to bound-constrained optimization problems. Here, the objective function corresponding to (3.1) is penalized according to the following formulation for the VDST model (Christelis et al. 2018):

$$\min f\left(\mathbf{Q}\right)=\begin{cases}-\sum_{i=1}^{k}Q_i, if\ \forall\, i=1,...,k\,; C_i\left(Q_1,...,Q_k\right)\leq C_t \\ M_v\sum_{i=1}^{k}\left[\max\left(\left(C_i-C_t\right),0\right)\right]^2, if\ \exists\ i=1,...,k\,; C_i\left(Q_1,...,Q_k\right)>C_t\end{cases} \tag{3.4}$$

while in the case of (3.2) the objective function is similarly defined as:

$$\min f\left(\mathbf{Q}\right)=\begin{cases}-\sum_{i=1}^{k}Q_i, if\ \forall\, i=1,...,k\,; x_i^{C_t}\left(Q_1,...,Q_k\right)\leq xw_i \\ M_v\sum_{i=1}^{k}\left[\left(\left(x_i^{C_t}-xw_i\right),0\right)\right]^2 if\ \exists\ i=1,...,k\,; x_i^{C_t}\left(Q_1,...,Q_k\right)>xw_i\end{cases} \tag{3.5}$$

where M_v represents the number of pumping wells that the constraint is violated. The above formulation attributes a separate score for each violated constraint. Furthermore, the penalized objective function is multiplied by M_v to include the number of constraint violations for the case of a non-feasible vector $\mathbf{Q}$ (Forrester et al. 2008). It is noted that the number of the constraint functions equals the number of pumping wells. Therefore, each constraint function is associated with an individual pumping well and they all contribute to the objective function score. A similar formulation is defined for the sharp interface model for problem (3.3):

$$\min f(\mathbf{Q}) = \begin{cases} -\sum_{i=1}^{k} Q_i, if\ \forall\, i = 1,...,k\,; x_i^{toe}(Q_1,...,Q_k) < xw_i \\ M_v \sum_{i=1}^{k} \left[\max\left(\left(x_i^{toe} - xw_i\right), 0\right)\right]^2, if\ \exists\, i = 1,...,k\,; x_i^{toe}(Q_1,...,Q_k) \geq xw_i \end{cases} \tag{3.6}$$

The pumping optimization problems defined in this section can be directly solved by combining the VDST or the sharp interface model with an evolutionary algorithm. Although the development of efficient and robust optimization algorithms is a very active research field, it is rather unrealistic to expect that a specific evolutionary algorithm will always provide the best optimal solutions for optimization problems with different characteristics (Behrangi et al. 2008). In this work, a probabilistic heuristic global optimization algorithm, namely the evolutionary annealing-simplex (EAS) algorithm (Efstratiadis and Koutsoyiannis 2002) is employed to solve the optimization problems defined above. EAS combines the concepts of the downhill simplex method and simulated annealing to improve efficiency and effectiveness in smooth and rugged search spaces, respectively (Tsoukalas et al. 2016). Details on the principles and steps for the implementation of EAS can be found in Efstratiadis and Koutsoyiannis (2002), Rozos et al. (2004), Kourakos and Mantoglou (2009) and Tsoukalas et al. (2016). A few parameters need to be initialized before applying EAS, that is, the initial population size $mpop$, the annealing schedule parameter ξ_{an}, a mutation probability criterion p_m and a convergence criterion $\varepsilon_{\mathrm{cov}}$. The initial population size is one of the EAS parameters that has a considerable impact on the global exploration capabilities of the algorithm (Kourakos and Mantoglou 2009). The EAS parameters were set to $mpop = 8 \times k$, where k is the number of the decision variables, $\xi_{an} = 2$, $p_m = 0.1$ and

$\varepsilon_{\text{cov}} = 10^{-4}$, according to the suggestions in Rozos et al. (2004). The EAS algorithm has demonstrated comparable performance against well-established optimization algorithms often used in water resources and has been successfully applied in problems of automated calibration (e.g. Rozos et al. 2004; Efstratiadis et al. 2015; Tigkas et al. 2016; Christelis et al. 2016c). Recently, EAS has been also applied in coastal aquifer management problems demonstrating robust performance (Kourakos and Mantoglou 2009; Christelis and Mantoglou 2016b; Christelis et al. 2018; Kopsiaftis et al. 2019a). The solution of problems (3.4) and (3.5) using the VDST model and an evolutionary algorithm is computationally expensive and hinders the implementation of simulation-optimization routines for coastal aquifer management. The following chapter presents various optimization frameworks based on the use of surrogate models which enable a high-fidelity solution of the present problem in a computationally manageable manner.

Chapter 4

Surrogate modelling and optimization

4.1 Literature review on surrogate-based optimization

4.1.1 Metamodels and optimization

The reliable solution of engineering optimization problems typically requires the use of high-fidelity, yet computationally expensive numerical models (Koziel and Leifsson 2016). Despite the advances in computer power, a robust decision-making process requires hundreds to thousands of numerical simulations to gain an understanding of the system's response to various inputs and stresses. One option to cope with these computationally demanding tasks, is the development of fast mathematical approximations of the original physics-based simulation models. These approximation models are typically called surrogate models or metamodels and their function is to mimic the responses of the time-intensive original model in a variety of input variables of interest. Henceforth, the term surrogate models or metamodels will be used interchangeably.

Popular examples of surrogate models presented in the engineering optimization literature are Kriging, radial basis functions, support vector machines, artificial neural networks and Gaussian processes (Forrester and Keane 2009; Asher et al. 2015). While it is generally accepted that surrogate modelling can alleviate the computational burden associated with the physics-based numerical simulations, its success and effectiveness depends on the problem at hand (Razavi et al. 2010). The use of surrogate models for the various engineering optimization problems is commonly reported as surrogate-based optimization (SBO).

There are mainly two ways to implement SBO methods which share a common step at the beginning. First, an initial set of input-output data from the physics-based models is obtained to train the surrogate models and attain a certain level of accuracy for predicting responses to unseen data (Solomatine and Ostfeld 2008). These initial training points are usually derived from space-filling designs, such as Latin Hypercube Sampling, to achieve a better understanding of the variability of the unknown original model response surface (Razavi et al. 2012b). However, after the initial training procedure either an offline or an online framework might follow (Figure 4-1).

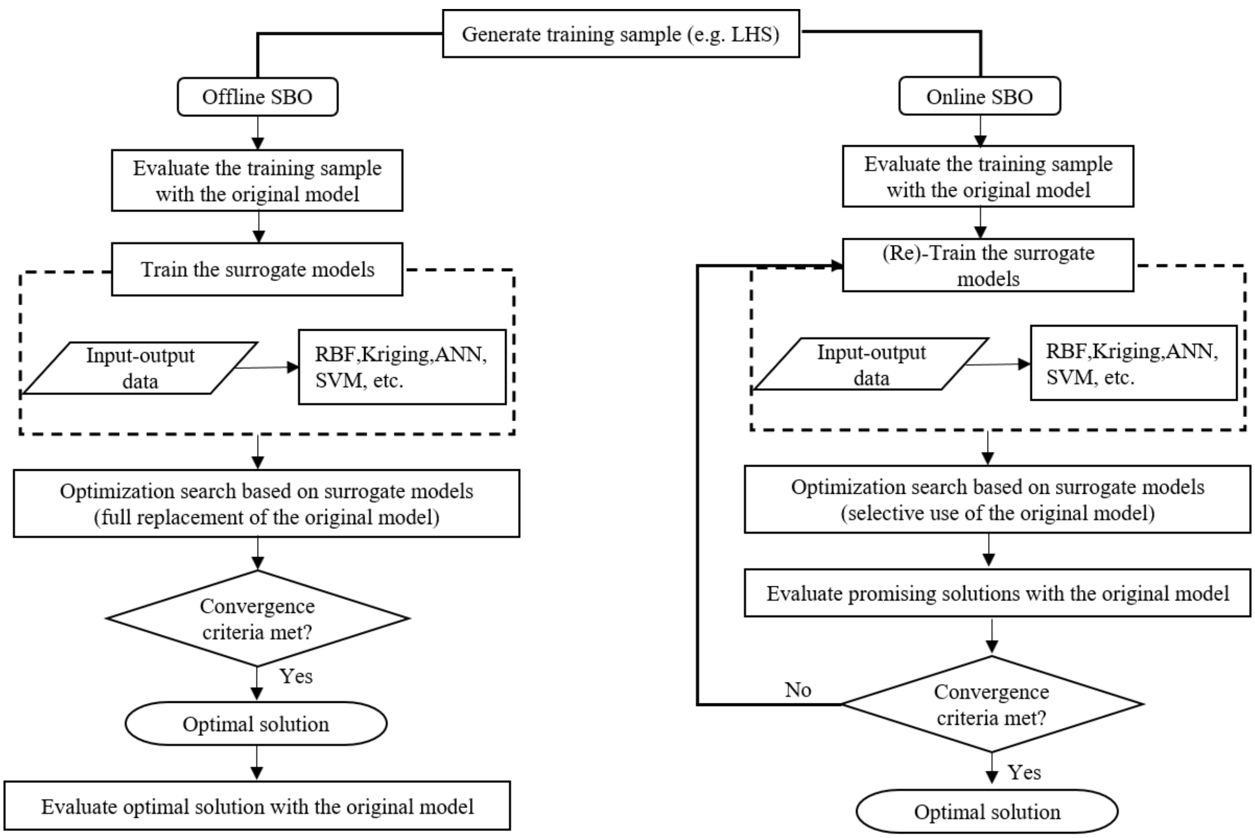


Figure 4-1 Generic workflow examples of an offline and an online SBO method.

As shown in Figure 4-1, the offline or basic sequential approach is the simplest one and assumes that the surrogate model can be used alone to search for the optimal solution after having been trained once with the available input-output data from the original model (Razavi et al. 2012b). It is noted though that the accuracy of the surrogate models generally depends on the available training data, both in terms of size and quality. An empirical rule suggests that the initial training sample could be in the order of $m_{init} = 10k$ although this might lead to unaffordable computational cost as the number of design/decision variables k increases (Jones et al. 1998). In real-world applications, there is often a practical limit on the number of runs with the expensive physics-based model. Therefore, it is unlikely that a single training set will provide a globally accurate surrogate model to explore the search space (Forrester et al. 2008). In that case, the optimization process will probably suffer from false optima, introduced due to inaccurate surrogate model predictions of the original model responses (Jin 2011). Therefore, the optimal solution found with the metamodels should be evaluated using the original model at the end of this framework.

On the contrary, the online SBO method allows for a selective communication between the surrogate model and the original model during the optimization steps. If the online framework is embedded within the operations of an evolutionary algorithm, is usually called surrogate-assisted evolutionary strategy (Jin 2011). Certain criteria are employed in a surrogate-assisted evolutionary algorithm to define whether any promising solutions obtained with the metamodels, during the operations of the evolutionary algorithm, should also be evaluated with the original model (Jin 2011; Karakasis and Giannakoglou 2006). Another common online SBO method is the so-called adaptive-recursive framework which starts with an initial fit of the metamodels to a set of points generated from a space-filling design. Then, the metamodels are used to search for new promising points based on an optimization step or a random sampling procedure combined with distance metrics, probability criteria, etc., (e.g. Regis and Shoemaker 2007; Müller and Woodbury 2017).

The best points found with the metamodels are evaluated using the original model, the new training data are added to the initial design and the metamodel is updated. Stopping criteria, such as maximum number of evaluations with the original model, are usually applied to mark the termination of the adaptive-recursive framework. Both the surrogate-assisted evolutionary strategy and the adaptive-recursive framework ensure that the original model evaluates the feasibility of candidate solutions and the SBO method delivers reliable objective function values.

4.1.2 Exploration, exploitation, and infill criteria

With online SBO methods, sampling strategies are developed where the available runs with the expensive numerical model are split to evaluate a set of initial training points and sequentially another set of additional points (also called infill or update points). The additional points are chosen by using infill strategies based on various criteria. One type of infill criteria is the so-called prediction-based exploitation which adds points at the current optimum found by the metamodel. It is a greedy approach that aims to quickly improve the accuracy of the surrogate models in the optimum region. However, it neglects the global improvement of the metamodel and can potentially weaken the effectiveness of the SBO methods and get stuck in local optima (Forrester et al. 2008).

Other more comprehensive infill criteria seek to find a balance between exploration and exploitation using the metamodels. An exploration step calls the original model to evaluate points away from the current optimum which are then added to the training dataset and further improve the global predictions skills of the metamodel. The probability of improvement and the expected improvement, are popular criteria in SBO that guide the search with a metamodel to new points that show increased probability to improve the current known best solution as well as to regions where the metamodel’s prediction uncertainty is significant (Jones et al. 1998; Forrester and Jones

2008; Couckuyt et al. 2010). This is a beneficial characteristic of specific metamodels, such as Kriging, where their prediction is accompanied with an estimated error and thus allows for locating sampling points where uncertainty is higher. Another approach is to utilize weighted distance metrics from previously evaluated points and probabilistic criteria for generating new sampling points to be evaluated with the original model (e.g. Regis and Shoemaker 2007; Regis and Shoemaker 2013).

There is a wide body of literature that focuses on the development of infill strategies that effectively utilize the physics-based model simulations to update the surrogate model. Typically, the aim is to start with a small initial training sample and then sequentially increase the accuracy of the surrogate model within regions of interest as well as globally in the decision variable space (e.g. Jones et al. 1998; Leary et al. 2004; Queipo et al. 2005; Mugunthan et al. 2005; Regis and Shoemaker 2007; Forrester and Keane 2009; Villemonteix et al. 2009; Jin 2011; Kleijnen et al. 2012; Yao et al. 2014; Tsoukalas et al. 2016; Yang et al. 2020). The efficient global optimization algorithm (EGO) (Jones et al. 1998), is one of the most popular SBO algorithms that deals with unconstrained optimization problems for computationally expensive numerical models. EGO, in its original version, utilizes Kriging metamodels to emulate the response of the objective function to the input variables and it is usually implemented with the expected improvement criterion to maintain a balance between exploration and exploitation. It is noted that since the introduction of EGO, the development of SBO algorithms on bound constrained problems continues to flourish and numerous papers have been published proposing improved versions of EGO or EGO-like algorithms (e.g. Huang et al. 2006; Knowles 2006; Forrester and Jones 2008; Villemonteix et al. 2009; Kleijnen et al. 2012; Viana et al. 2013; Couckuyt et al. 2014; Sun et al. 2020). Recently, the EAS algorithm, which is utilized in this thesis, was also modified to develop a surrogate-assisted version, namely, SEEAS (Tsoukalas et al. 2016). SEEAS is a surrogate-assisted evolutionary algorithm which employs a cubic RBF metamodel to emulate the response of the objective function to decision variables and utilizes a weighted acquisition function to generate new sampling points within the operations of the original EAS algorithm.

4.1.3 The presence of nonlinear inequality constraints

The majority of the SBO algorithms found in the literature, deal with computationally expensive problems where the only constraints are bound constraints (Müller and Woodbury 2017). An example is the calibration of a simulation model where its decision variables are restricted to certain lower and upper limits. However, most engineering optimization problems involve constraints which may or may not be computationally expensive (Forrester et al. 2008). In the presence of expensive-to-evaluate constraint functions, surrogate models are required to emulate their response. However, one could argue that if the number of constraints is large then building separate surrogate models for each one of them can be complicated or costly. Surely, the use of cheap-to-train surrogate models is essential to alleviate the overall computational burden, assuming that their prediction capability is acceptable for a specific optimization problem.

Another way to deal with the case of multiple constraints is to build a metamodel only for the penalized objective function. However, it has been realized in the literature that building metamodels for the penalized objective function only, limits the capabilities of SBO to approach the region of global optimum (Regis 2011; Dong et al. 2018). This is even more evident when a prediction-based exploitation infill strategy is selected for the SBO algorithm (Forrester et al. 2008). Ignoring the information provided by the constraints, hinders the effectiveness of SBO for optimization problems with inequality constraints (Jiao et al. 2019).

Several studies have suggested modifications to the expected improvement criterion to accommodate a faster and more accurate solution of nonlinear constrained optimization problems (e.g. Sasena et al. 2002; Basudhar et al. 2012; Li et al. 2017; Bouhlel et al. 2018; Dong et al. 2018; Jiao et al. 2019). The development of generic convergent schemes for nonlinear constrained SBO, is an active research area where studies aim to develop efficient algorithms for cases where the constraints are correlated, their number is large or an initial feasible solution is not known (Regis 2011; Basudhar et al. 2012; Regis 2014; Datta and Regis 2016). Sometimes it is possible for an analyst/engineer to eliminate a few constraints prior to any optimization runs, assuming that some of those are not active or are not essential for deriving an optimal result (Forrester et al. 2008). Generally, the presence of nonlinear inequality constraints might further complicate the development of SBO methods and many studies discuss how to specifically handle these difficulties (e.g. Forrester et al. 2008; Regis 2011; Parr et al. 2012; Boukouvala and Ierapetritou 2014; Datta and Regis 2016; Li et al. 2017; Müller and Woodbury 2017; Regis and Wild 2017; Bouhlel et al. 2018; Dong et al. 2018; Nuñez et al. 2018; Wu et al. 2018; Li et al. 2019).

4.1.4 Multiple surrogates

Instead of employing a single type of surrogate models to approximate the responses of an expensive physic-based model, various researchers in engineering optimization have investigated the use of multiple surrogates to improve accuracy and effectiveness (e.g. Viana et al. 2009; Acar 2010; Müller and Piché 2011; Viana et al. 2013; Nikolos 2013; Müller and Shoemaker 2014; Jiang et al. 2015; Shi et al. 2016; Hou et al. 2017; Bhosekar and Ierapetritou 2018). The generic approach in a multiple surrogate framework is to identify a suite of reliable surrogates for the problem at hand. This is usually achieved through a cross-validation process and then either the best or an ensemble of surrogates (e.g. weighted average surrogate) may be utilized (Viana et al. 2010).

The superiority of employing multiple against single type surrogates is somewhat debatable, since the nature of each optimization problem may favor the one approach or the other (Viana et al. 2009; Babaei and Pan 2016). It should be also considered that the use of multiple surrogates might increase the effort and computational time in SBO. For example, the analyst has to decide whether a cross-validation strategy will identify the best surrogate models or it is known a priori which surrogates should work best based on previous experience with the optimization problem (Viana et al. 2009). In the first case, it remains to be defined how many surrogates should be explored and what computational budget is available to apply an efficient and informative cross-validation framework. Furthermore, the training time of surrogate models varies and exploring more sophisticated surrogate models will eventually add considerable computational time.

4.1.5 Multi-fidelity optimization

Thus far, the discussion considered only a single, high-fidelity (HF) level for the physics-based model. Often, the exploration of the search space may be facilitated by using simpler, computationally cheap models which simulate the physical system at low-fidelity (LF) levels (Forrester et al. 2008). This possibility has motivated the development of the so-called multi-fidelity or variable-fidelity optimization (Robinson et al. 2006). Within this context, surrogate models are built upon faster LF models which may be simplifications of the physical system or might share the same physics with the HF models, but, are less accurate in terms of grid resolution, convergence criteria, dimensionality, etc., (Razavi et al. 2012a). LF models utilize their embedded knowledge of the physical system to produce an output and potentially can offer additional benefits for the implementation of SBO methods, particularly when the available HF model runs are limited (Koziel and Leifsson 2016). The main hypothesis is that an analyst can only afford to run the HF model a few times but it is affordable to run LF models many times and gather enough samples to acquire useful information for a computationally demanding optimization problem (Fernández-Godino et al. 2019).

There is variety of multi-fidelity modelling approaches presented in the literature. Many examples can be found in electromagnetic simulations through the application of the space mapping technique (Bandler et al. 1994; Bakr et al. 2000; Bandler et al. 2004; Koziel et al. 2008; Koziel et al. 2009; Cervantes-González et al. 2016; Feng et al. 2019), as well as, in aerospace engineering (e.g. Alexandrov et al. 2001; Gano et al. 2004; Marduel et al. 2006; Forrester et al. 2007; Karakasis et al. 2007; Han et al. 2013; Leifsson and Koziel 2015; Tyan et al. 2015; Zhou et al. 2016; Cheng et al. 2019; Shu et al. 2019). The usefulness of multi-fidelity surrogates in cases where only a few HF data can be obtained, is rather unquestionable (Koziel et al. 2011). Various multi-fidelity methodologies have been successfully developed in the literature to integrate the information from both HF and LF models (e.g. Gano et al. 2004; Forrester et al. 2007; Koziel et al. 2008; Leifsson and Koziel 2015; Zaefferer et al. 2016; Liu et al. 2016; Zhou et al. 2017). However, there is some ambiguity regarding their success and the computational gains in cases where it is affordable to gather enough HF data and thus a conventional surrogate model can be constructed instead (Fernández-Godino et al. 2019).

4.2 SBO in coastal aquifer management

Metamodels have been used in coastal aquifer management to approximate the responses of VDST models and alleviate the computational burden which results from coupled simulation-optimization routines (Singh 2014). Artificial Neural Networks is a popular example of surrogate model selection in pumping optimization problems of coastal aquifers (e.g. Rao et al. 2004; Dhar and Datta 2009; Kourakos and Mantoglou 2009; Sreekanth and Datta 2010; Sreekanth and Datta 2011; Grundmann et al. 2012; Kourakos and Mantoglou 2013). Recent studies have also proposed other surrogate models such as, evolutionary polynomial regression (Hussain et al. 2015), polynomial chaos expansions (Rajabi et al. 2015), Gaussian process models (Rajabi and Ketabchi 2017; Kopsiaftis et al. 2019), fuzzy inference systems (Roy and Datta 2017a), multivariate adaptive regression splines (Roy and Datta 2017b), extreme learning machine (Yadav et al. 2018) and support vector machine regression (Lal and Datta 2018).

Former SBO applications in coastal aquifer management have mostly implemented the offline approach. Rao et al. (2004), first replaced a VDST numerical model with Artificial Neural Network (ANN) models for coastal aquifer management. They solved the optimization problem with a simulated annealing algorithm. Also, they discussed the limitations of fully replacing the VDST model with the surrogate model for the accurate exploration of the search space. Bhattacharjya and Datta (2005) approximated a three-dimensional VDST model using a trained ANN model which provided considerable computational savings in the search for optimal solutions. The impact of the large training sample size on the performance of a global ANN model was also addressed.

Kourakos and Mantoglou (2006) substituted a VDST model with a trained ANN model in a pumping optimization problem of coastal aquifers. Their optimization approach was based on a nonlinear programming algorithm and results were in good agreement with the VDST-based optimization. Furthermore, they mentioned the necessity to explore the structure of the ANN model to improve its generalization capabilities. It is noted that with the above offline SBO methods large training data sets were generated to construct accurate surrogate models. Nevertheless, large training sets have been utilized in coastal aquifer modelling to exhaustively compare the performance and prediction skills of different metamodeling techniques (e.g. Kopsiaftis et al. 2019b).

Kourakos and Mantoglou (2009) developed an online approach that aimed to bypass the large training time required for a global ANN model and the results showed a significant computational gain. Modular neural sub-networks were selectively trained in a surrogate-assisted evolutionary strategy and the VDST model evaluated only the current best solution during the operations of an evolutionary algorithm. Sreekanth and Datta (2010) and Sreekanth and Datta (2011), used Genetic Programming as a metamodelling method in coastal aquifer management and compared it with the widely used ANN. Their results showed advantages of Genetic Programming over ANN metamodels, in terms of number of model parameters, parameter estimation and training data requirements, as well as in finding the global optimal solution. Christelis and Mantoglou (2016) compared the performance of the surrogate-assisted evolutionary framework against the adaptive-recursive approach for pumping optimization of coastal aquifers. Both frameworks were based on the prediction-based exploitation infill method. Their results showed that the surrogate-assisted evolutionary framework outperformed the adaptive-recursive approach in computational efficiency while it successfully located feasible optimal solutions. In general, the use of online SBO frameworks for coastal aquifer management, significantly improved the exploration of the search space within reasonable computational times (e.g. Kourakos and Mantoglou 2009; Papadopoulou et al. 2010; Song et al. 2018). Nevertheless, the application of SBO strategies which utilize the information obtained from the surrogate models to balance exploration and exploitation is rather limited in pumping optimization of coastal aquifers.

Despite the recognized success of surrogate modelling in computationally demanding tasks, the benefit from its use in optimization problems with increased dimensionality and under limited computational budgets is controversial according to Razavi et al. (2012b). Such limitations are also investigated in the present thesis while in the recent work of Christelis et al. (2018) it was found that SBO methods outperform the direct optimization with the VDST model, particularly as dimensionality and number of constraints are increased.

Regarding the multiple surrogate approach in coastal aquifer management, there are fewer applications in the literature compared to the use of single surrogate models. Sreekanth and Datta (2011) first utilized an ensemble of genetic programming surrogate models in lieu of a VDST model, to solve a multi-objective pumping optimization problem. Their study showed that the ensemble performed better than using a single genetic programming surrogate model. Recently, Roy and Datta (2017a) and Roy and Datta (2017b), solved multi-objective pumping optimization problems by utilizing ensembles of fuzzy inference systems and of multivariate adaptive regression splines, respectively. Their approach reduced the uncertainty in the prediction of the surrogate models while the multivariate adaptive regression splines provided a more efficient ensemble surrogate model. Recently, Roy and Datta (2020) utilized an illustrative coastal aquifer model and developed ensembles of metamodels based on the Dempster-Shafer theory to predict SWI under pumping conditions. Their results showed some advantage of the ensemble over the standalone metamodels, yet, in overall the performance of the ensemble was comparable to the metamodel identified as best. On a similar study, Lal and Datta (2020) concluded that, for their case study, a homogeneous ensemble of Gaussian process regression models performed better than standalone models of artificial neural networks, support vector regression, genetic programming and Gaussian process regression, or any heterogeneous combination of the above in the form of an ensemble. In the present thesis, the performance of multiple and single surrogates in pumping optimization is also compared based on previous results published in Christelis et al. (2019b). A surrogate-assisted evolutionary framework is implemented using heterogeneous and homogeneous ensembles of surrogate models. Their performance is compared against the corresponding single-type surrogate approach.

To the best of our knowledge, the combined use of multi-fidelity models for SWI management has received little attention so far. Also, in their review paper, Sreekanth and Datta (2015) do not report any multi-fidelity optimization methods developed for coastal aquifer management. Although a considerable number of papers compares or develops faster low-fidelity models to better approximate the VDST model responses, only few have proposed a combined use for pumping optimization problems.

In the broader context of multi-fidelity optimization for coastal aquifer management, Christelis and Mantoglou (2016a) and Christelis and Mantoglou (2017) developed methods which adaptively correct the parameters of a LF SWI model, to mimic the response of a HF SWI model. Specifically, Christelis and Mantoglou (2016a) have developed an optimization strategy to improve the calculated maximum pumping rates within the operations of an evolutionary algorithm. Their approach was based on the concept proposed by Pool and Carrera (2011) to correct the density ratio in the sharp interface model of Strack (1976) (practically reducing it) and allow for larger maximum pumping rates. Using a one-off correction of the density ratio cannot ensure that the resulting optimal solution would be comparable with that obtained from the VDST model, given that a specific iso-salinity contour is usually set as a threshold in the constraint functions (Christelis and Mantoglou 2016a). On the contrary, it is possible that such a correction would result in non-feasible solutions when the optimal pumping rates will be evaluated using the VDST model (Kopsiaftis et al. 2019a). Therefore, instead of using the corrected LF sharp interface model alone, it was proposed to selectively call the VDST model during optimization and adaptively correct the density ratio. The main purpose was to locate feasible solutions for the

VDST model, based on the computationally efficient sharp interface model and only on a small number of the expensive VDST simulations. In other words, a limited number of VDST simulations were utilized to evaluate the sharp interface predictions, adjust the density ratio, and control the optimization search towards feasible solutions. During optimization, the SWI models simulate the impact of numerous combinations of groundwater extraction rates on the development of the seawater intrusion front. Several density ratio corrections of the sharp interface model might be required to approximate the response of the VDST model. To that end, it was suggested that the density ratio can be treated as a black-box parameter, with no physical meaning, subjected to a form of calibration process where an optimal value of this ratio is sought.

This adaptive adjustment of the density ratio proposed by Christelis and Mantoglou (2016a), was probably the first attempt to combine SWI models of different fidelity for solving pumping optimization problems in coastal aquifers. It involves two fidelity levels, the sharp interface model and the VDST model. The advantage of the method is that it provides a steep improvement of the objective function within a few iterations of the optimization algorithm. It has similarities to the implicit space mapping techniques (Bandler et al. 2004) where fixed parameters of the HF model are iteratively adjusted in the LF model to minimize the differences between their responses. However, it should be noted that the sequential adjustment of the density ratio, during the operations of the evolutionary algorithm, does not imply a continuous improvement of the sharp interface model. The evaluation with the VDST model and the updated density ratio only allows the algorithm to search for better solutions with the computationally cheap sharp interface model. At a certain number of function evaluations with the VDST model, the updated density ratio will oscillate around a set of values where the sharp interface model matches the VDST output. Thus, it will eventually update the density ratio only in the search region that the evolutionary algorithm has located as more promising.

In general, the method proposed by Christelis and Mantoglou (2016a) can be useful in the case of limited computational budgets but it is not expected to outperform conventional surrogate methods when it is affordable to get more samples from the VDST model. Furthermore, if the two SWI models are in good agreement for a certain range of pumping rates, the algorithm might unnecessarily call the expensive VDST model for evaluation. It should be noted that the success of this multi-fidelity optimization strategy is strongly depended on the parameterization of the SWI models as well as, the type of constraints utilized in the pumping optimization problem. Kopsiaftis et al. (2019a), examined the concept of density ratio modification by forming ensembles of corrected sharp interface models to approximate the optimal solution of the VDST model. Their results showed that a combination of three density modifications could mitigate the errors produced by sharp interface models for pumping optimization.

Recently, Christelis and Mantoglou (2019a) used simple response correction techniques between a sharp interface model and a VDST model in a multi-fidelity optimization framework. Their work showed that multi-fidelity SBO methods could be, under conditions, a promising choice for coastal aquifer management. Nevertheless, in a series of independent runs presented in that work, there was no evidence that this multi-fidelity method can outperform the conventional data-driven approach if an adequate number of HF samples can be obtained. Dey and Prakash (2020) applied a similar methodology to Christelis and Mantoglou (2016a) by developing an iterative process where the density ratio is optimized to update the sharp interface predictions and calculate optimal solutions, using a particle swarm optimization algorithm. Their results showed computational gains within a few iterations and their algorithm quickly located local optima of the

VDST model. In general, as multi-fidelity optimization has received little attention in coastal aquifer management, further research should follow to identify cases or problems which could be benefit from this approach.

4.3 Types of surrogate models utilized in this thesis

The nonlinear constraints described in chapter 2 are computationally expensive to evaluate. The VDST numerical simulations compute the distribution of the salinity concentration for a specified management period and then the constraints are calculated. Even for an optimization problem of moderate dimensionality, the task is considered impractical given the computational cost of VDST simulations and the thousands of runs required by an evolutionary algorithm to converge. To reduce the computational burden, surrogate models are employed to approximate the VDST model response to pumping and enable an efficient simulation-optimization routine.

Two types of surrogate models were chosen to develop SBO methods for coastal aquifers, namely, radial basis functions (RBF) and Kriging (KRG). While these surrogate models have been fairly utilized in other fields of engineering optimization, their use is scarce in aquifer management studies. Their interpolating capabilities allow for an exact estimation of the previously evaluated sampling points with the physics-based model. This is considered beneficial in terms of surrogate model accuracy for deterministic computer simulations (Kleijnen 2009) which is the case for the present thesis. Both RBF and KRG are very popular choices in SBO (Jin et al. 2001). An increasing interest in these metamodels is also observed in the broader field of water resources optimization (e.g. Baú and Mayer 2006; Shoemaker et al. 2007; Razavi et al. 2012b; Tsoukalas et al. 2016).

4.3.1 *Radial basis functions (RBF)*

The radial basis functions were originally developed for the purpose of scattered multivariate data interpolation (Hardy 1971). One attractive characteristic is that RBF models can express highly nonlinear responses, a common case in engineering optimization, while they still preserve a simple mathematical formulation (Forrester et al. 2008). The construction of RBF surrogate models in this thesis is based on the MATLAB codes developed for MATSuMoTo toolbox by Müller (2014). In the pumping optimization problem described in chapter 2, the objective function is a linear function of the k decision variables $Q_1,...,Q_k$ that is, the pumping rates. A unique RBF model is constructed for each one of the k inequality constraints associated with each one of the k pumping wells. Let the decision vector of pumping rates denoted by $\mathbf{Q}=(Q_1,...,Q_k)$. Then, given a set of m training points $\mathbf{Q}^{(1)},\mathbf{Q}^{(2)},...,\mathbf{Q}^{(m)}\in\mathbb{R}^k$, the corresponding responses obtained from the evaluation with the VDST model are either $G_i=\left[C_i\left(\mathbf{Q}^{(1)}\right),...C_i\left(\mathbf{Q}^{(m)}\right)\right]^T, i=1,...,k$ or $G_i=\left[x_i^{C_T}\left(\mathbf{Q}^{(1)}\right),...,x_i^{C_T}\left(\mathbf{Q}^{(m)}\right)\right]^T, i=1,...,k$ (the reader is referred to chapter 3 for the definition of these quantities). The following RBF model is employed (Powell, 1992):

$$s\left(\mathbf{Q}\right)=\sum_{j=1}^{m}\lambda_j\varphi\left(\|\mathbf{Q}-\mathbf{Q}^{(j)}\|\right)+p\left(\mathbf{Q}\right) \tag{4.1}$$

where $s\left(\mathbf{Q}\right)$ in our case denotes the prediction of the RBF model for an input vector of pumping rates $\mathbf{Q}$, $\boldsymbol{\lambda}=\left[\lambda_1,...,\lambda_m\right]^T\in R$ are coefficients to be determined, $\left(\cdot\right)^T$ represents the transpose of a vector or a matrix , $\varphi\left(\cdot\right)$ denotes the form of the radial basis function, $\|\cdot\|$ is the Euclidean distance and $p\left(\mathbf{Q}\right)=\boldsymbol{\beta}^T\mathbf{Q}+\beta_0$ is a linear polynomial tail with $\boldsymbol{\beta}=\left(\beta_1,...,\beta_k\right)^T$ whose coefficients also need to be determined such that the resulting RBF model passes through all the m design points. The coefficients $\boldsymbol{\lambda}$, $\boldsymbol{\beta}$ and β_0 are obtained by solving a linear system of equations:

$$\begin{bmatrix} \mathbf{\Phi} & \mathbf{P} \\ \mathbf{P}^T & \mathbf{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{\lambda} \\ \mathbf{w} \end{bmatrix} = \begin{bmatrix} G_i \\ \mathbf{0} \end{bmatrix} \tag{4.2}$$

where $\mathbf{\Phi}$ is a $m \times m$ matrix with entries $\mathbf{\Phi}_{j,l} = \varphi\left(\| \mathbf{Q}^{(j)} - \mathbf{Q}^{(l)} \| \right)$, $\mathbf{P}$ is a $m \times (k+1)$ matrix whose ith row is $\left[1, \left(\mathbf{Q}^{(i)}\right)^T\right]$, $\mathbf{w} = [\beta_1, ..., \beta_k, \beta_0]^T$ and $\mathbf{0}$ denotes a matrix with all entries equal to zero. The interpolation matrix is invertible if and only if $rank(\mathbf{P}) = k+1$ which in turn implies that at least $k+1$ points are required to train the RBF model (Müller and Woodbury 2017). Here, two types of basis functions are considered, the cubic form where $\varphi(r) = r^3$ and a thin plate spline where $\varphi(r) = r^2 \ln r$. The construction time for the RBF models is negligible in comparison to the VDST simulation time even for large number of training points. To facilitate the presentations of the results in the following chapters, the cubic and the thin plate spline RBF models will be abbreviated as CUB and TPS, respectively.

4.3.2 *Kriging (KRG)*

Kriging was mainly introduced as an approximation model for multidimensional input-output data from a simulation model, in the original paper of Sacks et al. (1989). However, the origins of the Kriging method are generally attributed to the work of the South African mining engineer Krige (Kleijnen 2009). The concept utilized by the KRG surrogate models is that when the distance between two input vectors $\mathbf{Q}$ and $\mathbf{Q'}$ is small, then the resulting scalar responses $C_i(\mathbf{Q})$ and $C_i(\mathbf{Q'})$ (or $x_i^{C_t}(\mathbf{Q})$ and $x_i^{C_t}(\mathbf{Q}'))$ should be closely correlated (Forrester et al. 2008). Details about the mathematical background of Kriging in metamodeling can be found elsewhere in the literature (e.g. Jones et al. 1998; Lophaven et al. 2002; Forrester et al. 2008). Here, we only briefly present the general concept of a KRG surrogate model as it applied in our problem. Kriging treats the responses G_i from the VDST model to the input sample points $\mathbf{Q}^{(1)},\mathbf{Q}^{(2)},...,\mathbf{Q}^{(m)}\in\mathbb{R}^k$, as if they were generated from a stochastic process $Y(\mathbf{Q})$ defined as (Sacks et al. 1989):

$$Y(\mathbf{Q})=\sum_{i=1}^{p}\alpha_i F_i(\mathbf{Q})+Z(\mathbf{Q}) \tag{4.3}$$

where $\sum_{i=1}^{p}a_i F_i(\mathbf{Q})$ is a regression model and $\boldsymbol{\alpha}=(\alpha_1,...,\alpha_p)$ are the regression coefficients while $F_i(\mathbf{Q}),\ i=1...p$ are known basis functions. The second term $Z(\mathbf{Q})$ is a zero mean Gaussian process, with variance σ^2 and a $m\times m$ correlation matrix $\mathbf{\Psi}$ with entries given by the correlation function $corr\left[Z(\mathbf{Q}),Z(\mathbf{Q}^{'})\right]=\exp\left[-\sum_{j=1}^{k}\theta_j\left|Q_j-Q_j^{'}\right|^{\pi}\right]$. A Gaussian correlation function is selected where $\pi=2$. Although π is considered fixed here it can also be a parameter to be estimated. The set of parameters $\theta_1,...,\theta_k$ is identified using the Maximum Likelihood Estimation method and numerical optimization techniques (Viana et al. 2010). In brief, a prediction $y(\mathbf{Q}^{new})$ with the KRG surrogate model at a new point $\mathbf{Q}^{new}$ is given by (Palar and Shimoyama 2018):

$$y(\mathbf{Q}^{new})=\Lambda\boldsymbol{\alpha}+r(\mathbf{Q}^{new})\mathbf{\Psi}^{-1}(G_i-\mathbf{F}\boldsymbol{\alpha}) \tag{4.4}$$

where $\Lambda=\left(F_1(\mathbf{Q}^{new})...F_p(\mathbf{Q}^{new})\right)$, $\boldsymbol{\alpha}=\left(\mathbf{F}^T\mathbf{\Psi}^{-1}\mathbf{F}\right)^{-1}\mathbf{F}^T\mathbf{\Psi}^{-1}G_i$ denotes a $p\times 1$ vector of the coefficients of the regression function as obtained by the Generalized Least Squares procedure. It is reminded that G_i is a $m\times 1$ vector of the observed VDST model responses associated with the ith constraint function, given the m training points $\mathbf{Q}^{(1)},\mathbf{Q}^{(2)},...,\mathbf{Q}^{(m)}\in\mathbb{R}^k$. Also, a $1\times m$ vector of correlations between the observed data and the new prediction is defined as

$r\left(\mathbf{Q}^{new}\right)=\left(cor\left[Y\left(\mathbf{Q}^{(1)}\right),Y\left(\mathbf{Q}^{new}\right)\right]\cdots cor\left[Y\left(\mathbf{Q}^{(m)}\right),Y\left(\mathbf{Q}^{new}\right)\right]\right)$. A $m\times p$ matrix $\mathbf{F}$ of the regression functions of the form is also defined as:

$$\mathbf{F}=\begin{pmatrix} F_1\left(\mathbf{Q}^1\right) \dots F_p\left(\mathbf{Q}^1\right) \\ \vdots \quad \vdots \quad \vdots \\ F_1\left(\mathbf{Q}^m\right) \cdots F_p\left(\mathbf{Q}^m\right) \end{pmatrix} \tag{4.5}$$

The SURROGATES MATLAB toolbox (Viana 2011) was used to facilitate the construction of KRG surrogate models which are based on the DACE toolbox (Design and Analysis of Computer Experiments) developed by Lophaven et al. (2002). The toolbox allows for regression models of different order polynomials and correlation models of various structures. After performing several numerical tests, two KRG surrogate models were employed in this work, namely, Gaussian correlation models combined with zero-order (denoted henceforth as G0) and first-order regression models (denoted henceforth as G1).

4.4 Surrogate-based optimization algorithms

The SBO algorithms developed in the present thesis employ individual surrogate models for each constraint function instead of a single surrogate model that emulates the response of a penalized objective function. This approach is expected to increase the accuracy of the SBO methods for nonlinear constrained optimization problems. The SBO methods are developed either based on the surrogate-assisted evolutionary strategy or the adaptive-recursive framework. The surrogate-assisted evolutionary algorithms mainly utilize the prediction-based exploitation approach which adds points at the current optimum found by the metamodel. On the contrary, the adaptive-recursive SBO methods apply a balanced exploitation and exploration strategy for sampling the VDST model.

As it has been discussed before, despite the level of sophistication in VDST models, there are still many limitations in their prediction skills particularly for regional coastal aquifers (Sanford and Pope 2010). In that sense, the methodologies proposed in this thesis aim to efficiently solve pumping optimization problems of coastal aquifers, assuming that the VDST models are as accurate as possible. It is also noted that no attempt is being made in this thesis to delve into the mathematical details of surrogate modelling theory and optimization. The aim is to understand how surrogate modelling fits within the practical requirements of coastal aquifer management plans and how the specifications of SWI modelling favor the use of one or the other SBO method.

4.4.1 Surrogate-assisted evolutionary strategy using single surrogate models

The method will be denoted throughout the thesis as EAS-PB followed by the type of the metamodel used in the algorithm. The EAS component of EAS-PB means that the method is embedded in the operations of the EAS algorithm and PB stands for prediction-based exploitation. The steps of this SBO approach can be summarized as follows:

1. Create an initial experimental design (LHS in our case) to provide the set S_m of training points (e.g. $\mathbf{Q}^{(1)},...,\mathbf{Q}^{(m)}$). Evaluate these initial points with the HF model (here, the VDST model) and store the outputs related to the constraint functions (for example, $C_i\left(\mathbf{Q}^{(1)}\right),...C_i\left(\mathbf{Q}^{(m)}\right), i=1,...,k$) in the set C_m.

2. Fit k surrogate models (e.g. k cubic RBF models), corresponding to the k pumping wells, using the available training sets S_m and C_m.

3. Run EAS algorithm based on the surrogate models and if during optimization a better optimum is found do the following:

   a) Re-evaluate the current decision vector of pumping rates with the VDST model.

b) Replace the objective function value with that obtained from the VDST model.

c) Compute the minimum Euclidean distance of the current candidate point $\mathbf{Q}^*$ from previously evaluated points in the training sample S_m. If the distance criterion $\Delta_{\min}$ is satisfied, add the new input-output data into the training sets S_m and C_m, and retrain the surrogate models.

4. Are stopping criteria of EAS algorithm met? If yes, return final solution; otherwise return to step 3 and continue running EAS algorithm using the surrogate models.

At step 3c, a distance criterion $\Delta_{\min}$ from previously evaluated points with the VDST model is calculated, before deciding to add the new sample point in the existing training dataset S_m. This is imposed because the accuracy of RBF and KRG models can be adversely affected by closely sampled points. If Δ_D denotes the minimum distance between the low and upper limits of the decision variables and k represents the number of the decision variables, we set $\Delta_{\min} = 0.0005 \times \Delta_D \times \sqrt{k}$, as suggested in the implementation of the ConstrLMSRBF algorithm (Regis 2011). The rest of the required parameters of EAS-PB were selected similarly to the original EAS algorithm (see chapter 3). That is, the initial population $mpop = 8 \times k$, the annealing schedule parameter $\xi_{an} = 2$ and the mutation probability criterion $p_m = 0.1$. In the present implementation of EAS-PB, the initial m training points also serve as the initial population m_{pop}. The EAS-PB algorithm will terminate either if $N_{eval} = m_{pop} \times 100$, which is the maximum number of objective function evaluations using the metamodels or if $\varepsilon_{\text{cov}} < 10^{-4}$ which is a convergence parameter defined in the original EAS algorithm. A schematic representation of EAS-PB is presented in Figure 4-2 below.

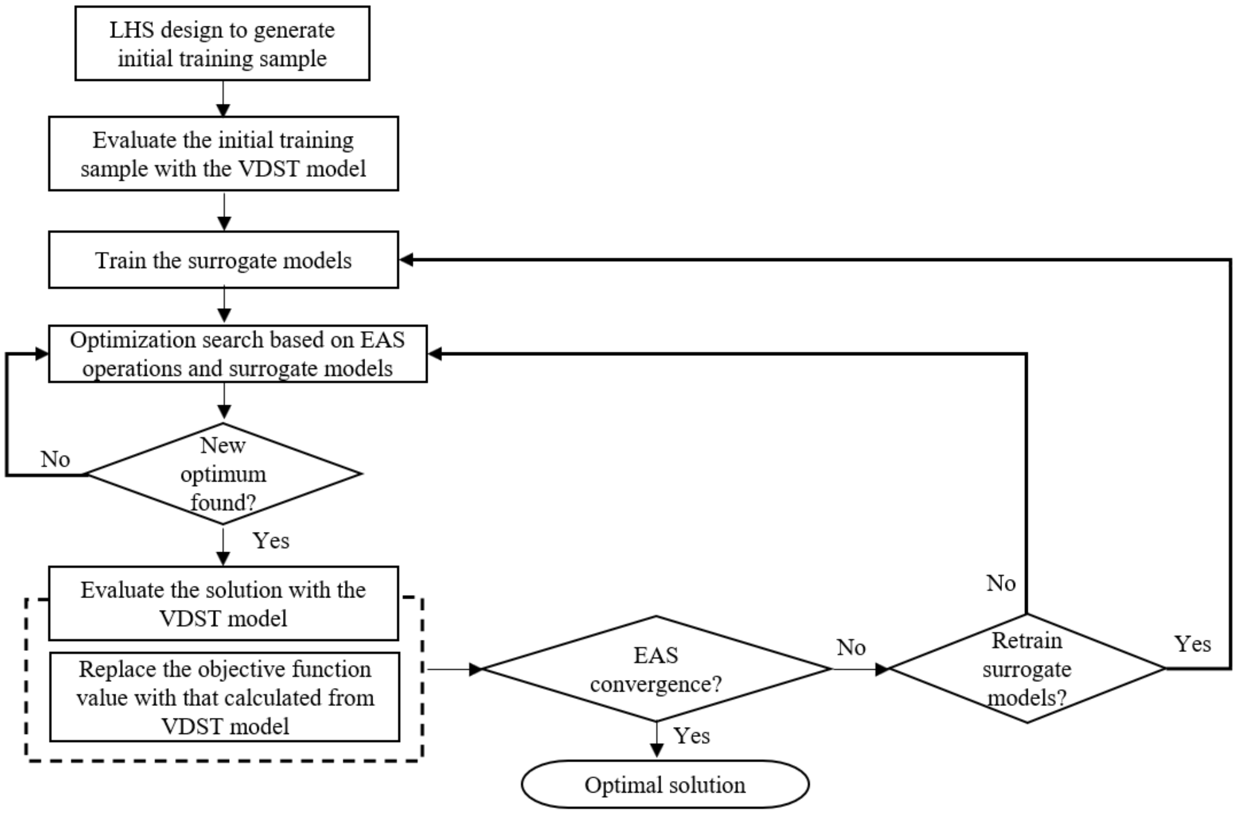


Figure 4-2 Optimization workflow for the EAS-PB method

4.4.2 Surrogate-assisted evolutionary strategy using multiple surrogates

A similar optimization method to EAS-PB is presented here but it is based on the multiple surrogate approach. Whether SBO methods using multiple surrogates derive better results than using single type surrogate models is inconclusive in the existing literature (Viana et al. 2010). To that end, two multiple surrogate approaches were developed here to examine their performance in pumping optimization of coastal aquifers. One of them, selects the best surrogate model for each constraint function and updates this knowledge as more evaluations with the VDST model are added during optimization. The other, selects the two best surrogate models for each constraint function and constructs an ensemble with weighted prediction. The weights are calculated through an internal optimization scheme within the operations of the EAS-PB algorithm. The ensemble means that the prediction of the approximation model is constructed from surrogates of different type, for example, a KRG model and a RBF model or a cubic RBF and a TPS RBF model.

To assess the prediction skills of the surrogate models and identify the best for each constraint function, a cross-validation strategy was employed. The RBF and KRG surrogates emulate the scalar response of the VDST model to pumping rates. Thus, a total of k surrogate models corresponding to the k pumping wells, need to be constructed. First, l roughly equal subsets including m_{sub} training points are generated from the whole of the training sample points. Then, iteratively each surrogate model is fitted to $l-1$ subsets and provides predictions for the points that were left out from the fitting process. This process is repeated l times to gather all the predicted errors based on different validation datasets. The root mean squared error (RMSE) was used as a cross-validation score between the known VDST responses and the predictions from the surrogate models. It was defined as follows for the ith pumping well:

$$RMSE^{(i)} = \sqrt{\frac{1}{m_{sub}}\sum_{j=1}^{m_{sub}}\left(y_i^j - \hat{y}_i^j\right)^2},\ i = 1,\dots,k \tag{4.6}$$

In the above RMSE definition, y_i^j denotes the known value from the VDST model for the j training point (the vector of pumping rates in our case) and $\hat{y}_i^j$ is the surrogate model prediction (either C_i or $x_i^{C_t}$) for the same point. For the best surrogate model approach, only the surrogates with the lowest RMSE values for each pumping well are selected. Within the EAS-PB optimization framework the surrogate models are re-constructed as new training points become available through the algorithm's operations. Every m_{new} training points, the cross-validation score is re-calculated and the best surrogates are identified again.

Apart from selecting the best surrogate model an ensemble formulation was also implemented. The two best surrogate models identified for each constrained function from the cross-validation strategy, were used to form an ensemble of surrogates by constructing a weighted average surrogate. The prediction of the ensemble surrogate model is formulated as:

$$\hat{y} = \sum_{q=1}^{N_{SM}} w_q \hat{y}_q \tag{4.7}$$

where $\hat{y}_q$ is the prediction produced by the qth surrogate model, N_{SM} is the number of the surrogate models which form the ensemble and in our case $N_{SM} = 2$. Variable w_q denotes the optimal weights attributed to the qth surrogate model prediction subject to the equality constraint $\sum_{q=1}^{N_{SM}} w_q = 1$. The construction of the weighted average surrogate model was based on optimal weights obtained through minimization of the cross-validation score. It is noted that the optimization task for the optimal weights was performed for each one of the constraint functions to form the ensembles of surrogates. The relevant optimization problem is defined as (Zhou et al. 2013; Jiang et al. 2015):

$$\begin{aligned} &\min f\left(w\right) = \sqrt{\frac{1}{m_{sub}} \sum_{j=1}^{m_{sub}} \left(w_1 \hat{y}_1 + w_2 \hat{y}_2 - y\right)^2} \\ &s.t. \sum_{q=1}^{N_{SM}} w_q = 1 \end{aligned} \tag{4.8}$$

where $\hat{y}_1$ and $\hat{y}_2$ are the individual predictions of the best two surrogate models for each constraint function value and y is the corresponding response of the VDST model. The above optimization problem was solved using the SQP method provided in MATLAB. Since the number of constraint functions might be large, the construction of the weighted average surrogate may add considerable computational time. The optimal weights are also updated every m_{new} training points within the operations of the EAS-PB method. Here, rather arbitrarily m_{new} was selected based on the dimensionality D of the optimization problem. For example, if $k = 10$ then the weights were re-calculated every $m_{new} = 10$ evaluations with the VDST model. Since the optimization framework is the same with the EAS-PB method, a flow diagram of the multiple surrogate approach is presented in Figure 4-3.

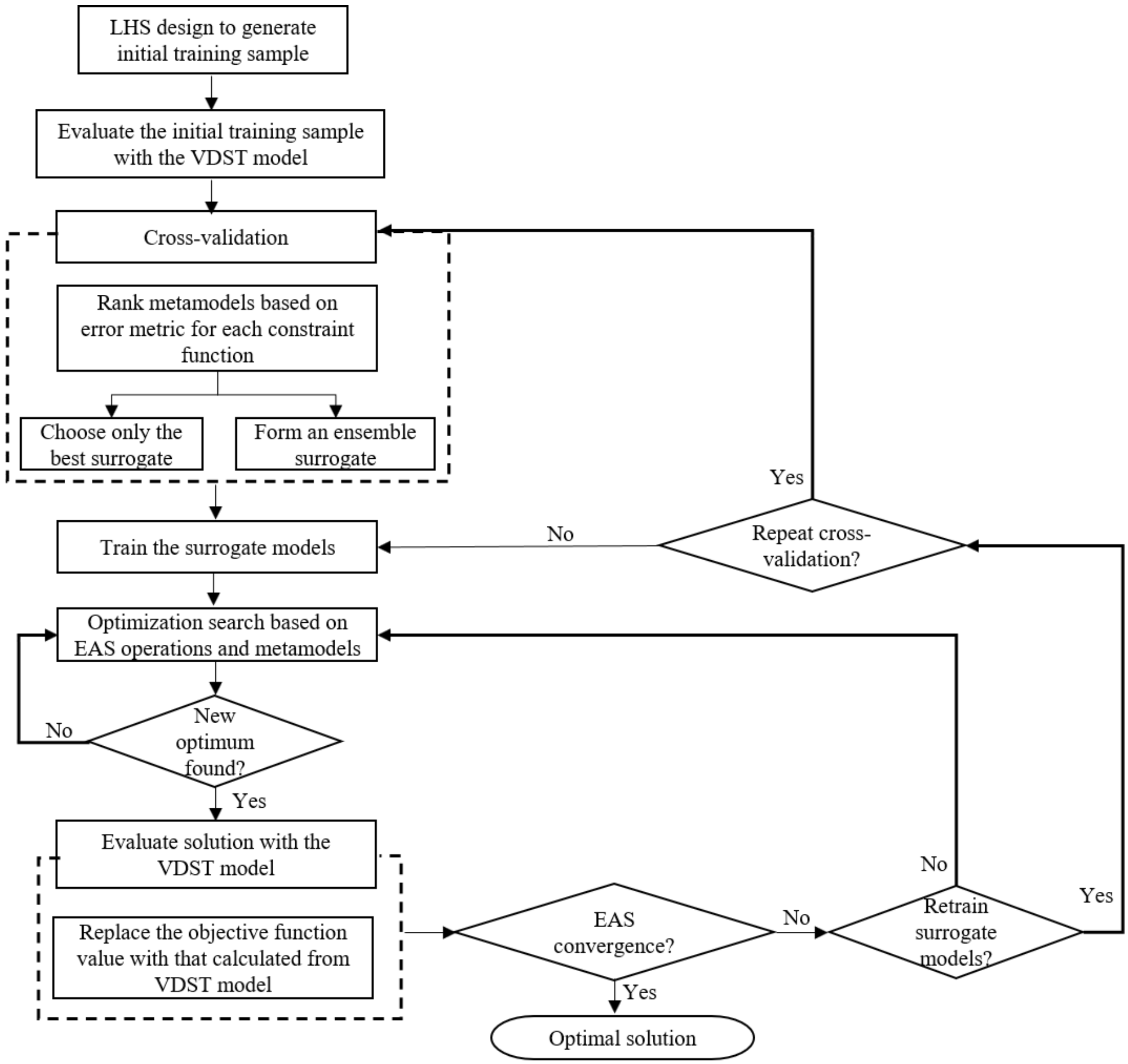


Figure 4-3 Workflow diagram of the surrogate-assisted evolutionary framework using multiple surrogates (using the best surrogate or an ensemble).

It is noted that the multiple surrogate approach is somewhat sample dependent which means that using a different training sample may result in a different selection/combination of the best surrogate models. If an ensemble is constructed by different instances of the same surrogate model, resampling techniques may be beneficial to increase the diversity of the surrogate model response (Roy and Datta 2017b). To distinguish the implementation of EAS-PB using multiple surrogates, the abbreviation EAS-PB(BE) is used when only the best model for each constraint function is selected, and EAS-PB(OW) denotes the case where optimal weights are computed to obtain the prediction of the ensemble surrogate model.

4.4.3 Modification to the SEEAS algorithm for nonlinear constrained optimization

Recently, EAS algorithm was enhanced with surrogate models to tackle computationally expensive optimization problems with bound constraints (Tsoukalas et al. 2016). The resulting SEEAS (Surrogate-enhanced evolutionary annealing simplex) algorithm, showed considerable advantages over other widely used SBO algorithms for water resources optimization problems. A main difference from EAS is that SEEAS employs an internal optimization task that minimizes an acquisition function and generates candidate points, based on an efficient search with surrogate models. The scope of the acquisition function is to balance exploration and exploitation using the metamodel and locate promising areas of the original objective function landscape (Tsoukalas et al. 2016). For the development of SEEAS, some modifications were also implemented to the operations of the original version of EAS. Detailed steps and description of SEEAS can be found in Tsoukalas (et al. 2016) and therefore we only refer here to the specific modification that was done for the purposes of the present thesis.

Originally, in SEEAS algorithm, the acquisition function calculates a weighted score of candidate points based on the predicted objective function value from the surrogate model and the distance from previously evaluated points with the computationally expensive model. Let $y_{st}\left(\mathbf{x}_c\right)$ be the standardized response value of a surrogate model to a set of candidate points $\mathbf{x}_c$ where $y_{st}\left(\mathbf{x}_c\right)=\left[y\left(\mathbf{x}_c\right)-y^{\min}\right]\big/\left[y^{\max}-y^{\min}\right]$. The values $y^{\min}$ and $y^{\max}$ represent the minimum and maximum values of the non-standardized response values $y\left(\mathbf{x}_c\right)$, respectively. In addition, the Euclidean distance d_E is calculated, for each point in the set $\mathbf{x}_c$, from all previously evaluated points $\mathbf{x}_p$ with the original physics-based model. The set of points $\mathbf{x}_p$ also serve as the existing training points for the surrogate model. The standardized distance $d_E^{st} = \left[d^{\max} - d_E\right] / \left[d^{\max} - d^{\min}\right]$ is also calculated and the weighted score of the acquisition function is obtained as $Sc = w y_{st}\left(\mathbf{x}_c\right) + (1-w) d_E^{st}$, where w is a dimensionless weighting coefficient. The latter is dynamically adjusted based on the empirical formula $w = \max\left[0.75, \min\left(PI, 0.95\right)\right]$ (Tsoukalas et al. 2016).

The parameter PI represents a dimensionless progress index calculated as $PI = \log\left(HFr\right) / \log\left(MHFr\right)$, where HFr is the current number of runs with the HF model and $MHFr$ is the maximum allowed number of runs with the HF model or in other words the assigned computational budget. This practically means that at the beginning, more weight (up to 0.25) is assigned to the acquisition function to promote exploration using the metamodel and improve global accuracy (for details, see Tsoukalas et al. 2016). The best point found from the minimization of the acquisition function replaces the worst point of the population, if it is better than the existing worst solution, and it is evaluated by the HF model. It also enters the archive of the training points and the surrogate model is updated. As more iterations are carried out by the algorithm and thus more HF model runs are added, the value of PI increases and the minimization of the acquisition function focuses on exploitation. The algorithm stops iterating when the predefined computational budget, i.e., the maximum number of HF model runs, is exhausted.

In our case, the pumping optimization problem includes nonlinear constraints and thus a slight modification was done to SEEAS to accommodate surrogate models for each constraint function. Thus, separate surrogate models associated with each pumping well were employed to approximate the constraint functions of the VDST model. Accordingly, the calculation of the acquisition function value was modified to include the information from all the surrogate models of the constraint functions and compute the weighted score. In SEEAS, the generated candidate points that violate a distance criterion from previously evaluated points with the original model are attributed a high score to drive the internal optimization away from these points. An additional penalty was also attributed in the modified constrained version for those points that are infeasible based on the predictions from the surrogate models of the constraint functions. In the next chapter, we investigate the performance of the original SEEAS in coastal aquifer management, assuming initially only a single surrogate model for the penalized objective function. Then, we test the performance of the modified version of SEEAS using surrogate models for the constraints and for convenience this approach will be denoted henceforth as CSEEAS (Constrained-Surrogate-enhanced evolutionary annealing simplex). As in SEEAS, we also utilize RBF surrogate models to run CSEEAS.

4.4.4 The ConstrLMSRBF algorithm

To further assess the SBO methods developed in this thesis, a comprehensive and effective SBO method developed by Regis (2011) was also included in the comparison. ConstrLMSRBF is a SBO method which balances global and local improvement of the surrogate model and has been applied in other engineering optimization problems successfully. Recently, it was applied in coastal aquifer management problems showing very promising results (Christelis et al. 2018). The algorithm simultaneously deals with the objective function and the constraints of the optimization problem, by constructing RBF surrogate models for each one of them.

The version of ConstrLMSRBF applied in this thesis, requires that at least one feasible point exists among the initial training points of the surrogate models. After evaluating the initial points with the computationally expensive model (here, the VDST model) RBF models are fit for the objective and constraint functions using all available data points. Then, the algorithm goes through a loop that involves generating a large number of random candidate points obtained by perturbing some or all of the coordinates of the current best feasible point, using Gaussian distributions with zero mean and with known standard deviations. The standard deviations can vary adaptively, depending on the performance of the algorithm and based on heuristic rules, to facilitate either local or global search. The algorithm gathers the candidate points that are predicted to be feasible or that have the minimum number of predicted constraint violations. The next point where the HF simulation will run is chosen to be the best point among all the valid candidate points according to two criteria: predicted objective function value of the candidate point according to the RBF model of the objective, and its minimum distance from previously evaluated points. Once the HF model has evaluated the selected valid candidate point, the algorithm re-trains the RBF surrogate model with the new data point included. Then it goes back to generating a new set of random candidate points and continues the loop until the computational budget is exhausted, that is, the maximum number of HF simulations has been reached. More details on the theoretical aspects of ConstrLMSRBF can be found in Regis (2011).

4.4.5 LR-RSRBF (Local Refinement Random Search with RBF models)

In this section, a new optimization framework is presented which adopts and combines concepts and steps from three different existing SBO algorithms, namely GOSAC, ConstrLMSRBF and SEEAS. The general structure is based on the concept proposed by Müller and Woodbury (2017) who developed GOSAC algorithm. GOSAC is a surrogate-assisted global optimization algorithm for problems with computationally expensive constraint functions and computationally inexpensive objective functions. This concept meets the specifications of the present pumping optimization problem where the objective function is calculated as a simple summation of the decision variables and thus it has negligible computational cost.

In its original formulation, GOSAC algorithm includes two main optimization phases. The first, seeks for a starting feasible point for those problems where an initial sampling design cannot guarantee the existence of feasible points. To that end, a multi-objective optimization problem is solved based on trained surrogate models with the initial sampling design. When a feasible point is found, the second phase of the algorithm aims at improving the current best feasible point. This is achieved by minimizing the objective function and evaluating the constraints using surrogate models, for example, RBF models.

Each time a new infill point is proposed by the surrogates the original model evaluates the solution and the point is added to the existing training sample. Since the second optimization phase aims at exploitation using the surrogate models, if the proposed solution is too close to previously evaluated points with the expensive computer model, another sample point is preferred which maximizes the minimum distance from the existing sample. This step enables the sampling of the original model in unexplored regions of the decision variable space and thus improves the global accuracy of the surrogates while it may locate promising areas with local or global optima (Müller and Woodbury 2017). GOSAC's stopping criterion is based on the depletion of the available computational budget.

Here, we implement another version of a GOSAC-like SBO method where the first phase is omitted since it is generally easy to locate a feasible starting point. However, the structure of our proposed method incorporates additional features from the SEEAS algorithm and the ConstrLMSRBF algorithm presented previously. Specifically, we make use of a progress index as in SEEAS while we also incorporate a specific part of the ConstrLMSRBF algorithm which appears to be highly effective for the present optimization problem. The latter involves generating many random candidate points obtained by perturbing some (or all) of the coordinates of the current best feasible point, using normal distributions with zero mean and with specified standard deviations. As it will be shown later in the numerical application, this refinement of the current best solution using perturbed points, leads to a steep improvement of the current best solution within only few evaluations with the VDST model and is considered beneficial for the optimization problem at hand. This exploitation step is further enhanced in the proposed algorithm by selectively activate an optimization step using surrogate models and an evolutionary algorithm. Apart from exploitation steps, the present methodology uses a criterion based on random number generation and the number of HF model evaluations, to decide if a new point should be added that maximizes the minimum distance from the existing sample, as in GOSAC algorithm. The method is implemented using cubic RBF models, but other surrogates could be employed as well. The steps of the proposed methodology are presented below:

1. Create an initial experimental design set S_m of training points (e.g. $\mathbf{Q}^{(1)},...,\mathbf{Q}^{(m)}$). Evaluate the initial points with the HF model. Store the outputs related to the constraint functions (for example, $C_i\left(\mathbf{Q}^{(1)}\right),...C_i\left(\mathbf{Q}^{(m)}\right), i=1,...,k$) into C_m.

2. Denote by $\mathbf{x}_{best}$ that point in S_m which corresponds to the best feasible objective function value f_{best}.

3. Initialize the progress index $PI=\log\left(HFr\right)/\log\left(MHFr\right)$ which lies between zero and one. Set a distance criterion $\Delta_{\min}$ (see section 3.4.1) for accepting a new entry in S_m and a

tolerance criterion F_{tol} for the number of subsequent failures F_{count} with the surrogate models to improve the current best solution .

4. Fit the k cubic RBF models on the evaluated points S_m .
5. **while** $HF_r < MHF_r$
6. Generate a random number r_{nd} in $[0,1]$,
7. **if** $r_{nd} > PI$, find point $\mathbf{x}_{new}$ that maximises the minimum distance of $\mathbf{x}_{best}$ from previously evaluated points in S_m .
8. **else** generate N_{cand} candidate points by perturbing the coordinates of $\mathbf{x}_{best}$. Evaluate the N_{cand} points using the RBF models and denote the best point found by $\mathbf{x}_{temp}$ and the best objective function value f_{temp} .
    a. **if** $f_{temp} < f_{best}$ and $F_{count} < F_{tol}$, set $\mathbf{x}_{new} = \mathbf{x}_{temp}$
    b. **else** minimize the objective function $f(\mathbf{x})$ subject to the RBF surrogate models of the constraint functions and denote the optimum point by $\mathbf{x}_{temp}$ and the objective function value f_{temp} .
    c. **end if**
9. **if** the distance criterion $\Delta_{\min}$ is violated find another point $\mathbf{x}_{new}$ that maximises the minimum distance of $\mathbf{x}_{temp}$ from previously evaluated points in S_m .

10. **else** set $\mathbf{x}_{new} = \mathbf{x}_{temp}$

11. **end if**

12. **end if**

13. Evaluate the point $\mathbf{x}_{new}$ with the HF model, return the objective function value f_{HF} and update HF_r.

14. Add $\mathbf{x}_{new}$ in S_m and the corresponding constraint function outputs in C_m.

15. **if** $f_{HF} < f_{best}$, update the values of $\mathbf{x}_{best}$ and f_{best}

16. **end if**

17. **end while**

18. Return the best solution $\mathbf{x}_{best}$, f_{best}.

As described in the above optimization framework, there are two cases where an exploration task takes place. First, it is decided based on the random number generation at step 7, if a new point will be added far from the current best solution $\mathbf{x}_{best}$. However, the chance of implementing this step is reduced as HF_r increases and thus, PI increases as well. This is a common practice in SBO methods to allow more exploration at early stages and focus on exploitation as the size of the training sample increases. The second time that exploration is selected regards the case where the proposed point $\mathbf{x}_{temp}$ is close to the existing training points. In such cases, it is preferable to add another point $\mathbf{x}_{new}$ to improve the global accuracy of the surrogate models as it is done in step 9.

The exploitation part of the algorithm is implemented at steps 8 and 8b either with the perturbation of the current best solution or by searching for a promising point using a global optimization algorithm and the RBF models. The latter step is critical for the performance of the proposed method. First, it adds considerable computational effort despite the exclusive use of surrogate models. An evolutionary algorithm will evaluate the objective function a few thousand times. Even using surrogate models, this implies an additional computational cost which depends on the number of the decision variables and the type of surrogate models used. For example, KRG models will lead to a slower optimization run than RBF models. Considering that in our case separate surrogate models are constructed for each constraint function, step 8b will eventually add some computational load, particularly if this step is repeated many times until the stopping criterion is met. On the other hand, selecting a fast-convergent optimization algorithm but with low robustness might adversely affect the progress of the proposed optimization framework. Therefore, it is recommended that one should implement this SBO method using an evolutionary algorithm with known global search capabilities for the problem at hand.

The increased computational load of step 8b can be treated effectively using vectorized versions of evolutionary algorithms. Due to the large number of independent trials that we employ in this thesis, the vectorized form of MATLAB’s particle swarm algorithm was found appropriate to employ for the present optimization problem. Preliminary runs showed that the original EAS algorithm can improve further the performance of this SBO method at the expense of increased computational time. As mentioned before, EAS is a particularly robust choice for pumping optimization in coastal aquifers. Obviously, if the single runtime of the VDST model ranges from several minutes to hours, the additional computational effort contributed by step 8b is not as noticeable as for a VDST model which runs in seconds or in few minutes. Furthermore, in practice, an analyst expects to obtain a good optimum with one or two runs of the SBO methods. Otherwise, it might be worthless to use SBO algorithms and train surrogate models instead of directly using the original HF model. For convenience, the proposed method will be denoted as LR-RSRBF (Local Refinement Random Search with RBF models) and a schematic workflow is presented in

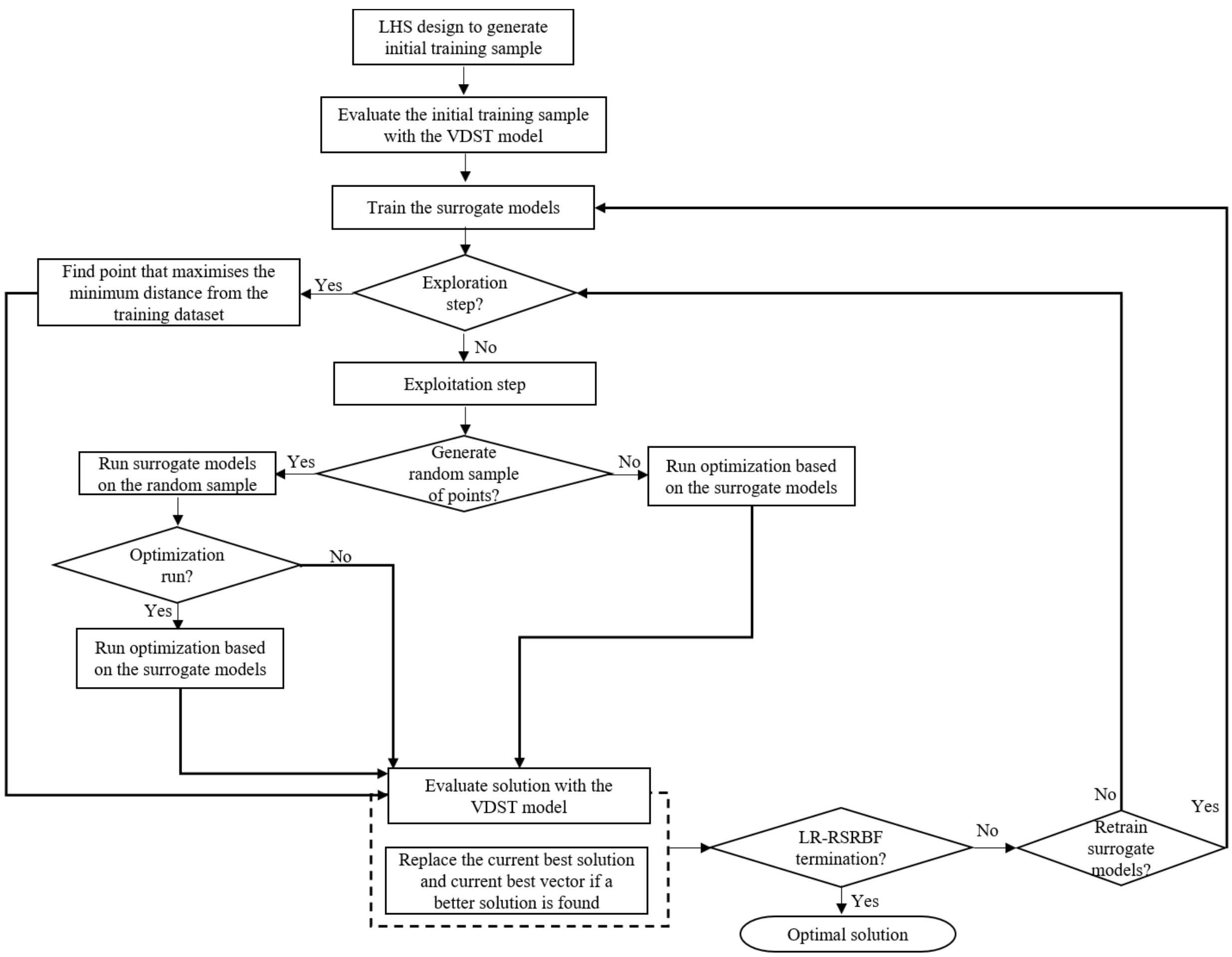


Figure 4-4 Workflow diagram of the LR-RSRBF optimization method.

4.4.6 Multi-fidelity optimization using co-Kriging metamodels

As discussed previously, using LF models in optimization could be beneficial for those cases where only limited data from the HF model can be obtained, due to extremely high computational cost (Fernández-Godino et al. 2019). There are different ways to define model fidelity levels, depending on the physical problem at hand. Based on preliminary runs with the available SWI models and the lack of studies that use multi-fidelity methods for pumping optimization in coastal aquifers, some fidelity models for SWI were not considered in this thesis.

For example, reducing the computational cost of the VDST model using coarser discretization was considered as a problematic option for the HydroGeoSphere code. Although different codes have different restrictions regarding the mesh Peclet number for transport solution (Essink, 2001), here, the coarser model was prone to numerical errors while the saving in runtimes might not be adequate. Other simpler SWI models that ignore salt transport mechanisms but consider saltwater movement have not been widely tested in pumping optimization studies while their computational cost, although much lower than VDST models, is still considerable. On the contrary, the one-fluid approach using Strack's potential (1976) is a very fast model of SWI that has been widely applied in coastal aquifer management (e.g. Mantoglou 2003; Mantoglou et al. 2004; Ferreira da Silva and Haie 2007; Mantoglou and Papantoniou 2008; Ataie-Ashtiani and Ketabchi 2011; Christelis et al., 2012; Karatzas and Dokou 2015). Presently, this sharp interface model was selected as an appropriate choice to explore multi-fidelity optimization in coastal aquifer management. As such, two levels of fidelity for simulating SWI were considered in this thesis, the VDST model and the sharp interface model of Strack (1976).

It should be noted that previous SWI studies have shown that this sharp interface model leads to conservative estimations of the optimal pumping rates compared to the VDST models (Pool and

Carrera 2011; Christelis and Mantoglou 2013; Christelis and Mantoglou 2016a; Kopsiaftis et al. 2019a). Also, Llopis-Albert and Pulido-Velazquez (2014) have demonstrated that depending on the hydraulic properties of the coastal aquifer, the sharp interface model may deviate significantly from the SWI predictions provided by the VDST model. Finding those regions where the responses between the LF and the HF model are significantly different and improving methods for multi-fidelity modelling, is a research area of growing interest in the relevant literature (e.g. Han et al. 2013; Zhou et al. 2016).

For the numerical experiments that follow in the next chapter we focus on a new approach to combine the outputs from the sharp interface and the VDST model. A special formulation of KRG models is the method of co-Kriging (coKRG) which exploits the combination of large amounts of LF data with few HF data to develop fast approximation models (Forrester et al. 2008). The constructed coKRG model is expected to be more accurate than the LF model alone, in approximating the HF responses. The success of a coKRG model also depends on the training samples and the differences between the HF and the LF model. The method of co-Kriging can be useful in cases where the HF model is very time-consuming and there is one or more LF models available that can enhance the analysis on a much lower computational effort. On the other hand, if the computational cost of running simulations with LF physics-based models is high or if the dimensionality of the optimization problem is large, the development of co-Kriging models may be an impractical choice (Forrester et al. 2008). The theory of co-Kriging dates back to 40 years ago with pioneering published works in geostatistics (Matheron 1973; Matheron 1979; Journel and Huijbregts 1978; Francois-Bongarcon 1981; Myers 1981). Here, the recent formulation presented in Forrester et al. (2007; 2008) is followed. The concept of co-Kriging is briefly presented below with emphasis given to the next chapter where the practical aspects of its implementation for coastal aquifer management is further discussed.

Consider a set of m_{HF} HF training points $\mathbf{Q}^{(1)},...,\mathbf{Q}^{(m_{HF})} \in R^k$ denoted as $\mathbf{Q}_{HF}$, which is a subset of m_{LF} LF training points $\mathbf{Q}^{(1)},...,\mathbf{Q}^{(m_{LF})} \in R^k$ denoted as $\mathbf{Q}_{LF}$. The associated HF and LF responses of the models can be denoted by $\mathbf{Y}_{HF} = \{Y(\mathbf{Q}^{(1)}),...,Y(\mathbf{Q}^{(m_{HF})})$ and $\mathbf{Y}_{LF} = \{Y(\mathbf{Q}^{(1)}),...,Y(\mathbf{Q}^{(m_{LF})})$, respectively. Then, let $Z_{LF}(\cdot)$ to represent a Gaussian process of the LF model data $(\mathbf{Q}_{LF}, \mathbf{Y}_{LF})$, $Z_{HF}(\cdot)$ a Gaussian process of the HF model data $(\mathbf{Q}_{HF}, \mathbf{Y}_{HF})$ and $Z_d(\cdot)$ a Gaussian process that represents the differences between $\rho Z_{LF}(\cdot)$ and $Z_{HF}(\cdot)$. In essence, the HF model is approximated as (Kennedy and O'Hagan 2000, Forrester et al. 2007):

$$Z_{HF}(\mathbf{Q}) = \rho_s Z_{LF}(\mathbf{Q}) + Z_d(\mathbf{Q}) \tag{4.9}$$

where ρ_s is a scaling factor estimated as part of the Maximum Likelihood Estimation method through optimization when the KRG model of the residuals between the LF and HF data is constructed. The co-Kriging prediction at a new point has a similar definition to (4.4) but the interested reader is referred to Forrester et al. (2007) and Forrester et al. (2008) for a proper mathematical presentation of the method, as it involves a considerable amount of matrix algebra for its derivation which is out of the scope of this work. The MATLAB implementation of coKRG provided in Forrester et al. (2008) was used to develop the multi-fidelity optimization method for this thesis. Also, the original version of the EAS algorithm was used to search for the optimal parameter sets of the coKRG model.

The multi-fidelity optimization framework using coKRG models, was developed here based on a specific loop included in the operations of the ConstrLMSRBF algorithm. As explained previously, ConstrLMSRBF goes through a loop that involves generating a large number of random candidate points obtained by perturbing some (or all) of the coordinates of the current best feasible point, using Gaussian distributions with zero mean and with standard deviations. We are implementing the multi-fidelity method assuming a restrictive computational budget of HF runs, and we utilize this concept aiming at a local search where all coordinates of the current best feasible point are perturbed. The standard deviation remains constant throughout the iterations of the method as opposed to the dynamic adjustment which is applied in the ConstrLMSRBF algorithm.

Initially, a space-filling design using the LHS method is created and the objective function values are evaluated using the LF sharp interface model. The LHS points along with the responses of the sharp interface model for each constraint function are stored in an archive which represents the LF data. To build the coKRG model a set of HF data is then required. It is noted that selecting those points for HF evaluation is not always straightforward (Forrester et al. 2008). In our case, the HF VDST model first evaluates the best point identified from the LF model based on the LHS design. Then the VDST model evaluates the worst point to gather the initial HF data. If, however, there is no feasible point then an internal loop searches for the next best point to ensure that the initial HF data include at least one feasible solution, before proceeding to the next steps of the method. This initial search may lead to spending some HF runs but on the other hand it is essential for the method to verify the feasibility of the current best point. A more conservative approach is to add an evaluation of a vector of low pumping rates that can ensure feasibility after evaluation with the VDST model. After gathering the HF and LF data, the coKRG models of the constraint functions are constructed. Then, the current best solution is perturbed, and a random sample is generated while the next point where the VDST simulation will run is the best point predicted from the coKRG models among all the generated candidate points. This new training pattern is added to the HF archive, the coKRG models are re-trained and the current best solution is updated if the new point is better.

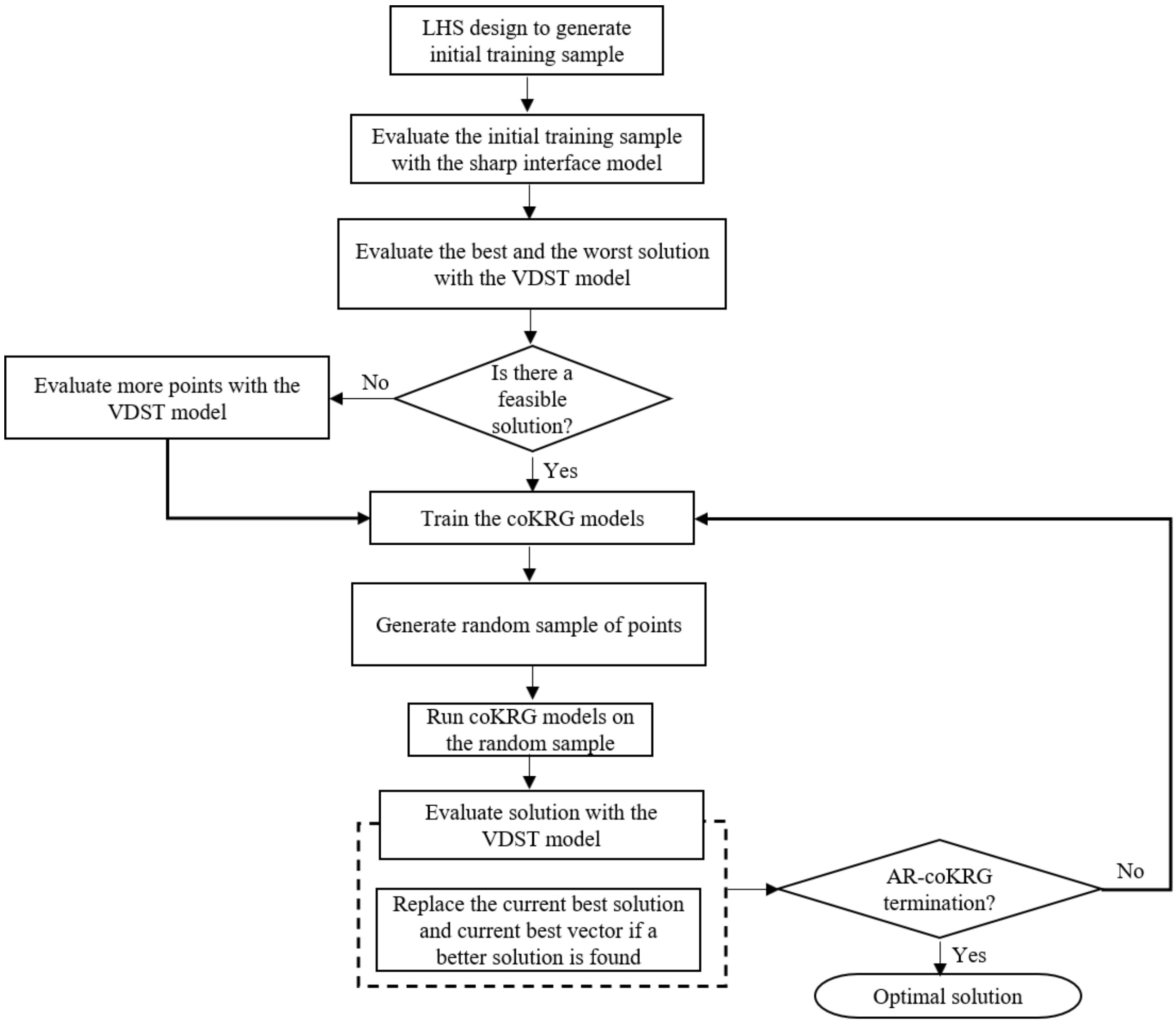


Figure 4-5 Workflow diagram of the multi-fidelity optimization framework using co-Kriging surrogate models and an adaptive-recursive sampling strategy.

Note that if the new point violates the tolerance for the minimum required distance from previously evaluated points with the HF model, then it does not enter the archive. However, with such a limited computational budget this is unlikely to happen. The above iterative framework stops when the maximum number of HF runs has been reached. As this framework follows an adaptive-recursive scheme, for convenience it is named as AR-coKRG (Adaptive-Recursive with coKRG models).

Chapter 5

SBO for hypothetical coastal aquifers

5.1 Settings and rationale of the numerical experiments

This chapter presents the findings from the several SBO frameworks that were developed for the pumping optimization problem described in chapter 3. We investigate the usefulness of surrogate modelling for this engineering optimization problem by conducting exhaustive comparisons among the proposed SBO algorithms. As previously discussed, VDST numerical simulations are associated with increased computational effort mainly originating from the spatial and time discretization requirements of the solute transport step (Werner et al. 2013). To facilitate the analysis, VDST models of simple geometry have been employed here and seawater intrusion is simulated based on hypothetical coastal aquifer models.

To enable the computationally expensive comparisons among the SBO algorithms, the settings of the numerical VDST model are chosen in such a way that a single simulation requires on average 11 seconds. Therefore, a relatively fast VDST model is utilized to perform such a demanding computational task for the generic comparison purposes. The optimization runs were performed on a 2.7 GHz Intel i5 processor with 8 GB of RAM in a 64-bit Windows 10 system. Furthermore, to assess the impact of problem dimensionality on the performance of the SBO methods, two scenarios are assumed, one with 10 operating pumping wells and the other with 20 operating pumping wells. The comparisons include the types of SBO algorithms described in chapter 4 and a summary is given also here for convenience.

One category utilizes the concept of prediction-based exploitation infill strategy and is implemented within the operations of the EAS algorithm. It is reminded that this approach aims at a fast convergence to an optimal solution, by adding infill points at the current optimum predicted by the surrogate models. This SBO method, which belongs to the broad group of surrogate-assisted evolutionary strategies, utilizes the inherent convergence criteria of the EAS algorithm. The surrogate-assisted evolutionary strategy is developed for both the single and the multiple surrogate approach. The multiple surrogate approach is based on the hypothesis that the combined use of surrogate models will smooth out the prediction errors from a single surrogate and thus will enable a better exploration of the search space. However, this approach implies additional computational effort.

The second category involves SBO algorithms that have been designed to balance local exploitation and global exploration using the metamodels and thus, they can potentially explore the search space with more accuracy. For the implementation of this type of SBO algorithms, we rely on the concept of the available computational budget. That is, a certain number of VDST simulations is available for evaluating the initial training points and the rest of VDST model runs are utilized to update the surrogate models with additional sampling points through iterative optimization procedures. Therefore, the depletion of the computational budget serves as a stopping criterion for these SBO algorithms.

To ensure that the comparison is as fair as possible, and that it is not affected by the different optimal choices for each algorithm, the following settings were employed. First, two alternative budgets of 100 and 300 VDST model runs were tested for those SBO methods that utilize the available computational budget as a stopping criterion. The computational budget of 100 simulations represents a hypothetical case where the VDST model is time-consuming and a restricted number of simulations is realistic to run. Furthermore, it serves the purpose of challenging the capability of the algorithms to near-optimal solutions despite the small computational budget. Based on preliminary runs, the computational budget of 300 simulations with the VDST model, was considered sufficient for all SBO methods to search for optimal solutions considering the dimensionality of the present optimization problems.

The size of the initial training points for fitting the KRG and RBF models was decided according to the type of the SBO algorithm. The surrogate-assisted evolutionary strategy depends, to some extent, on the evolutionary algorithm that is utilized (Razavi et al. 2012a). For example, the size of the initial population of EAS algorithm is considered critical to ensure global search capabilities and therefore the recommended value of $m = 8 \times k$ was selected (Kourakos and Mantoglou 2009). It is reminded that k is the number of the decision variables (pumping rates) and m denotes the number of the training points. Thus, the initial training points were associated with the initial population of the EAS algorithm for the surrogate-assisted evolutionary strategy. Obviously, this choice requires a substantial number of runs with the HF (high-fidelity) model. When the dimensionality of the optimization problem is large, this might be a non-affordable option in terms of computational cost. In those cases, one should reduce the initial population size at the possible cost of affecting the algorithm's capabilities for a global search.

The SBO algorithms that balance exploration and exploitation have been designed to perform successfully based on initial training designs of smaller sizes, such as $2 \times k + 1$ points. In fact, they may even perform worse in some cases, if a considerable number of HF runs is consumed on the initial design instead of exploiting the computational budget within their iterations. As more evaluations with the HF model are added, they have the potential to further improve the current optimum and thus, there is a clear reasoning on testing their performance on the available computational budget. The empirical rule $m = \max\left[2\times k+1, 0.1MHFr\right]$ as proposed by Razavi et al. (2012b), was used to set the initial training points for this type of SBO algorithms. The variable $MHFr$ represents the maximum available runs with the HF model as defined in each computational budget. Therefore, more weight is given on using the majority of the VDST simulations for the sampling of the subsequent infill points instead of evaluating a large initial training sample.

It is noted that for all the SBO methods, multiple independent optimization runs are performed to produce a statistically meaningful comparison, given the stochastic nature of the proposed algorithms and the inherent randomness in sampling designs. To ensure a fair comparison, the different SBO algorithms share the same initial training points for each optimization run whenever this is possible. Their performance is evaluated based on the sample statistics calculated from these independent optimization runs. Of course, in real world problems this luxury is not available but, here, it serves the purpose of comparing the sample statistics of the surrogate-based optimal solutions and assess the robustness of the SBO methods.

Finally, as a special case, we also examine the effectiveness of multi-fidelity optimization in the hypothetical scenario where only a few VDST simulations are affordable, that is 11 and 21 VDST model runs. To construct the RBF and KRG surrogate models, a minimum of $k+1$ points is required. Thus, the 11 sampling points with the VDST model are marginally sufficient to train the metamodels for the case of 10 pumping wells. The case of 21 sampling points is utilized to enable the performance comparison of multi-fidelity optimization against a conventional SBO algorithm. Based on these extremely limited computational budgets, we develop an iterative multi-fidelity optimization method utilizing co-Kriging models to search for an optimal solution. The comparison of the multi-fidelity method against the ConstrLMSRBF algorithm is conducted to assess the practical importance of multi-fidelity optimization in the case of computationally heavy VDST simulations.

5.2 Conceptualization of flow conditions and model description

A brief description of the hypothetical coastal aquifer geometry is presented herein. The conceptual model is based on a real-world coastal aquifer in the Greek island of Kalymnos, according to a previous work from Mantoglou et al. (2004). An orthogonal shape approximation of the real aquifer was utilized to enable the setup of a convenient VDST numerical model to support the exploratory nature of the optimization runs. Furthermore, the model assumes a homogeneous, anisotropic coastal aquifer under unconfined saturated flow conditions. The aquifer is replenished by both surface recharge and inland fluxes in the presence of multiple fully penetrating pumping wells (Figure 5-1).

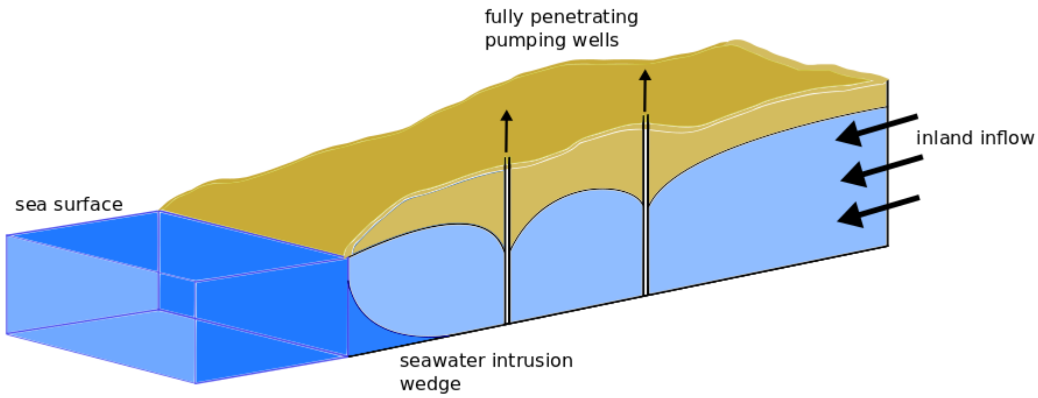


Figure 5-1 Conceptual flow model for the numerical experiments.

The horizontal dimensions of the coastal aquifer model are $x = 7km$, $y = 3km$ and the aquifer base is at $z = 25m$ below sea-level. The hydraulic conductivities are set to $K_x = 50\,m/day$,

$K_y = 50\,m/day$ and $K_z = 5\,m/day$ while dispersivity values are set to $\alpha_L = 100m$, $\alpha_T = 10m$ and $\alpha_{TV} = 1m$. On the left side of the aquifer model a hydrostatic boundary condition is applied to represent the sea-boundary with a constant specified salinity concentration of $35\,Kg/m^3$ for a saltwater density of $1025\,Kg/m^3$. The two lateral model boundaries are no-flow boundaries while the aquifer receives a total recharge (surface + inland) of $5410\,m^3/day$ (Figure 5-2).

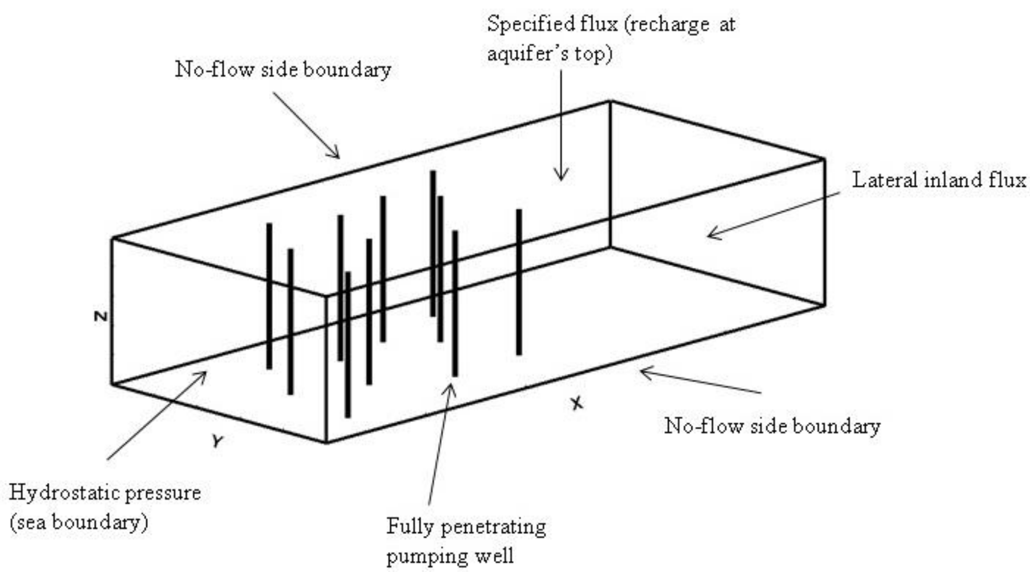


Figure 5-2 Representation of the applied boundary conditions on the numerical model and the distribution of the pumping wells (the case of 10 pumping wells is presented here).

5.3 Optimal results from the direct optimization with the VDST model

Initially, coastal aquifer flow is simulated without pumping until hydraulic head and salinity distribution reach steady-state conditions. The time horizon of the pumping management plan is set to 30 years (10950 days). However, running the optimization using as initial conditions those obtained without pumping, will result in a solution which is beyond the replenishment of the aquifer. Therefore, to represent a more realistic management situation, an initial optimization run is performed with the VDST model, to set a scenario where the aquifer is already being heavily pumped. The resulting head and salinity distributions from this first optimization run, are then used as initial conditions for all the subsequent optimization runs.

For the following numerical experiments, the direct optimization with the VDST model is solved once to provide benchmark solution and computational time. Henceforth, for convenience, this approach is denoted as EAS-VDST. As discussed previously, two pumping optimization problems were considered. The first scenario includes 10 pumping wells with lower and upper bounds of pumping rates set at $Q_{\min} = 0\, m^3/day$ and $Q_{\max} = 1000\, m^3/day$, respectively. The second scenario includes 20 pumping wells with lower and upper bounds of pumping rates set at $Q_{\min} = 0\, m^3/day$ and $Q_{\max} = 500\, m^3/day$, respectively. For this case, the upper bound of pumping rates was reduced by half to allow for more pumping wells to remain active since the total recharge remained the same. The salinity threshold was set to $C_t = 0.1 kg/m^3$ and this value was used to calculate the nonlinear constraint functions.

Figure 5-3 and Figure 5-4 present the results from the two optimization scenarios with the VDST model. In both cases, a similar pattern is developed where the three pumping wells closer to the sea boundary are at more risk of contamination if the calculated optimal pumping scheme operates for longer than the management period of 30 years. For the case of 10 pumping wells, the total pumping rate evaluated by EAS algorithm is $Q_{tot} = 4857.5\, m^3/day$ which corresponds to 89.8% of the total aquifer recharge whereas for the case of 20 pumping wells, the total pumping rate evaluated by EAS algorithm is $Q_{tot} = 4832.2\, m^3/day$ which corresponds to 89.4% of the total aquifer recharge.

Obviously, these calculated high percentages of total pumping are not sustainable and eventually after longer management periods more wells will be contaminated. However, for the numerical experiments designed in this section the scope is to investigate the performance of the SBO methods and not to focus on the impact of pumping on aquifer's sustainability. A longer design period would ensure a better protection of the groundwater resources for this hypothetical model, at the expense of more time-consuming simulations, which in turn, could hinder the implementation of the present exhaustive comparisons. The EAS-VDST optimization framework for the case of 10 pumping wells, converged after 4697 evaluations of the objective function through the VDST model, resulting in a total computational time of 14.35 hours. For the case of 20 pumping wells, 11020 evaluations of the objective function through the VDST model were required which resulted in a total computational time of 33.68 hours.

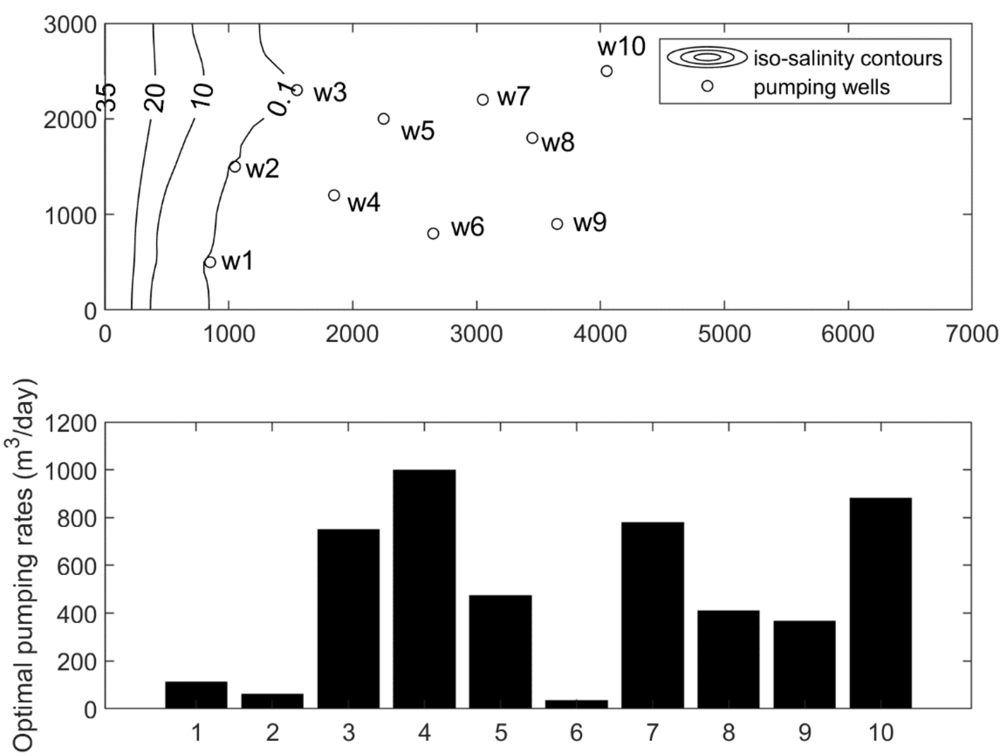


Figure 5-3 Plan view of the simulated salinity distribution at the aquifer base, for the optimal vector of pumping rates shown in the bar graph below (pumping wells are shown with numbers, i.e., w1). The results

are from the direct optimization with the VDST model for 10 pumping wells. It is noted that the iso-salinity representing $C_t = 0.1 kg/m^3$ marginally reaches but does not intersect the pumping wells.

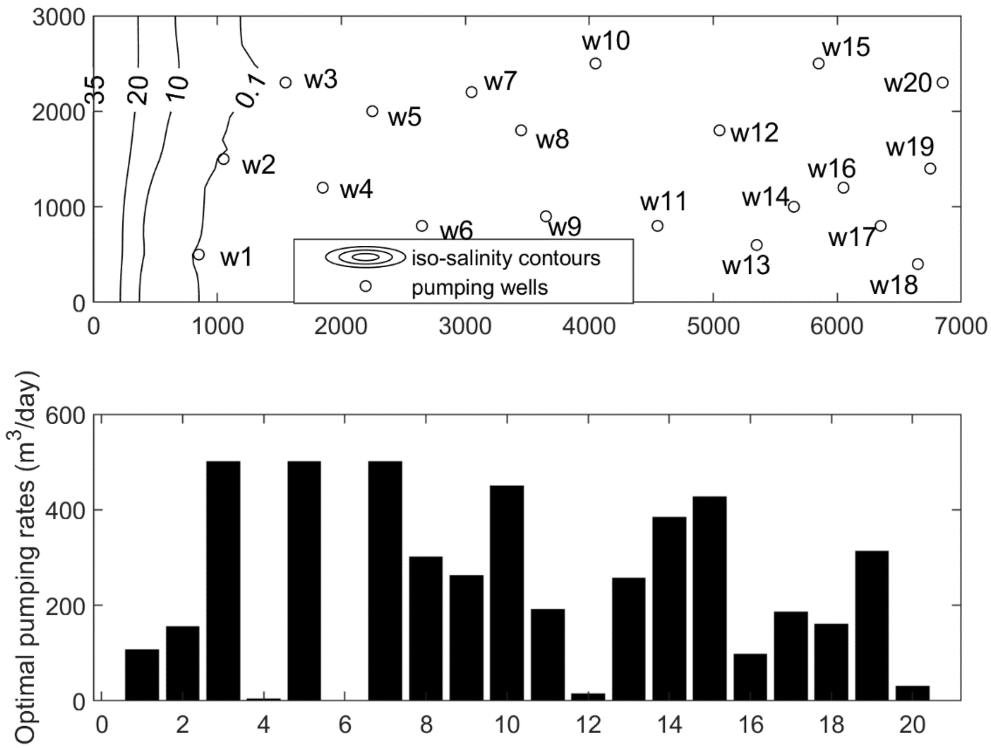


Figure 5-4 Plan view of the simulated salinity distribution at the aquifer base, for the optimal vector of pumping rates shown in the bar graph below (pumping wells are shown with numbers, i.e., w1). The results are from the direct optimization with the VDST model for 20 pumping wells. It is noted that the iso-salinity representing $C_t = 0.1 kg/m^3$ marginally reaches but does not intersect the pumping wells.

5.4 Optimal results with surrogate models of the penalized function only

Before proceeding to the main results from the SBO methods, it is presented how a single metamodel of the penalized objective function would perform in pumping optimization of coastal aquifers. Note that the scope is not to restate any difficulties that these SBO strategies encounter when dealing with nonlinear constrained optimization. Instead, the following numerical experiments are utilized to demonstrate their performance for the present pumping optimization problem and to highlight the differences with the explicit handling of nonlinear constraints using separate metamodels. In addition, it has not been shown before if any of these algorithms would perform satisfactorily in a pumping optimization problem of coastal aquifers and if these methods can successfully locate a good optimal solution. To that end, the pumping optimization problem was solved using three well-documented SBO algorithms:

i. The classic EGO algorithm based on the expected improvement criterion as implemented by Viana (2011) in SURROGATES MATLAB toolbox.

ii. The Multistart Local Metric Stochastic RBF (MLMSRBF) method (Regis and Shoemaker 2007).

iii. The SEEAS algorithm (Tsoukalas et al. 2016a), the surrogate-enhanced version of EAS for bound constrained problems.

All three algorithms above, have been applied successfully in a variety of optimization problems where the only constraints are bound constraints on all the decision variables. The details of each method can be found in the relative published papers and it is out of the scope of this work to present the theory. The classic EGO algorithm was implemented using KRG surrogate models.

Based on preliminary runs and literature examples of EGO algorithm in water resources management (Tsoukalas and Makropoulos 2015), a KRG model of zero-order polynomial and Gauss correlation function was selected to emulate the response of the objective function to pumping rates. MLMSRBF and SEEAS are implemented using a cubic RBF metamodel.

A set of 30 independent optimization runs were performed for each algorithm to avoid dependencies of the results on the random features of each method. The optimal results are compared against the benchmark solution obtained from the single run based on EAS-VDST optimization. The pumping optimization problem refers to the case of 10 pumping wells. The above SBO methods utilize as a stopping criterion the maximum number of function evaluations with the original computationally expensive model. Here, the limit is set to $MHFr = 300$, that is, 300 VDST model runs. The initial training points were set to $m = 30$ according to the rule $m = \max\left[2 \times k + 1, 0.1MHFr\right]$. The results are presented first in Figure 5-5 using a boxplot visualization of the optimal solutions. The comparison demonstrates that MLMSRBF and SEEAS had a more consistent performance than EGO as indicated by the boxplots.

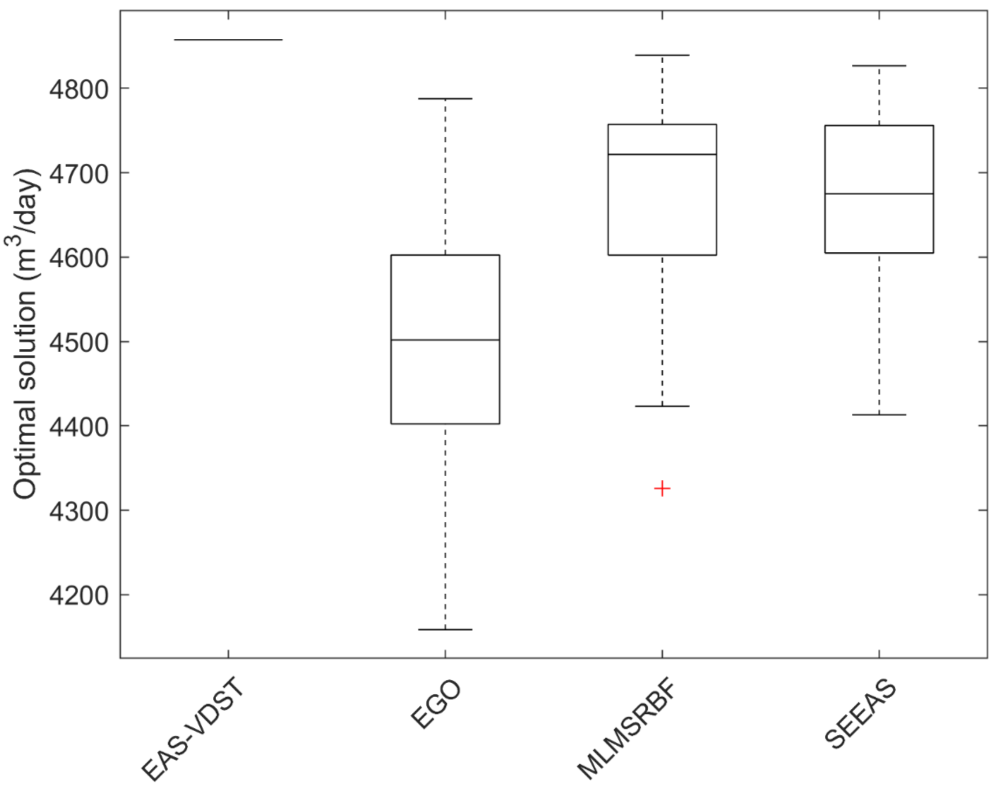


Figure 5-5 Comparison of direct optimization with the VDST model against the EGO, the MLMSRBF and the SEEAS algorithms. (Surrogate models have been constructed only for the penalized objective function).

The more reliable performance of MLMSRBF and SEEAS, is further demonstrated in Table 5-1 when comparing their mean and standard deviation values. The non-parametric Wilcoxon rank sum test, provided in the Statistics and Machine Learning MATLAB Toolbox, (2019b), was also employed to test the statistical significance of the results. The null hypothesis for the test, is that the data of two groups are independent samples from continuous distributions with equal medians. The test returns the p-value and a logical value h which indicates a rejection of the null hypothesis when $h = 1$ and a failure to reject the null hypothesis when $h = 0$ at the 5% significance level. The calculated p-value of 0.652 and $h = 0$ indicates that there is no statistically significant difference between the medians of MLMSRBF and SEEAS algorithms. On the contrary, their p-values against EGO were in the range of 10^{-5} with $h = 1$ indicating that the difference in their sample medians is statistically significant. Interestingly, all algorithms were able to locate at least one good optimal solution not far from the benchmark optimal solution using the VDST model alone.

Table 5-1 Sample statistics of the feasible optimal solutions obtained from the 30 independent optimization runs (best results are in bold and the benchmark solution is underlined).

Optimization frameworks	Worst	Best	Mean	Median	SD*
EAS-VDST**		<u>4857.5</u>			
EGO	4158.5	4787.7	4488.4	4501.8	174.93
MLMSRBF	4326.1	**4839.1**	**4670.3**	**4721.6**	126.49
SEEAS	**4413.1**	4826.7	4663.2	4675.0	**111.02**

* SD stands for standard deviation, ** one run with EAS-VDST is available

MLMSRBF provided the best feasible optimal solution in one of the optimization trials, which is remarkably close to the VDST-based solution. Figure 5-6 demonstrates the steep improvement of the average best feasible objective function value for MLMSRBF and SEEAS, as the number of VDST model runs increases. It appears that the infill strategy followed by these two algorithms shows promise in improving the objective function value and probably more runs with the VDST model could further improve their average performance, at the expense of additional computational effort. Notwithstanding the results are problem dependent, MLMSRBF and SEEAS algorithms showed that they have the potential to provide solutions of good quality. However, it should be noted that the performance of these algorithms is hindered, mainly due to the limited knowledge that the single surrogate model of the penalized function carries while searching the objective function landscape. In real-world applications, more effective SBO frameworks should be chosen for coastal aquifer management and this is pursued in the following numerical experiments where the surrogate models deal directly with the nonlinear constraints of the pumping optimization problem.

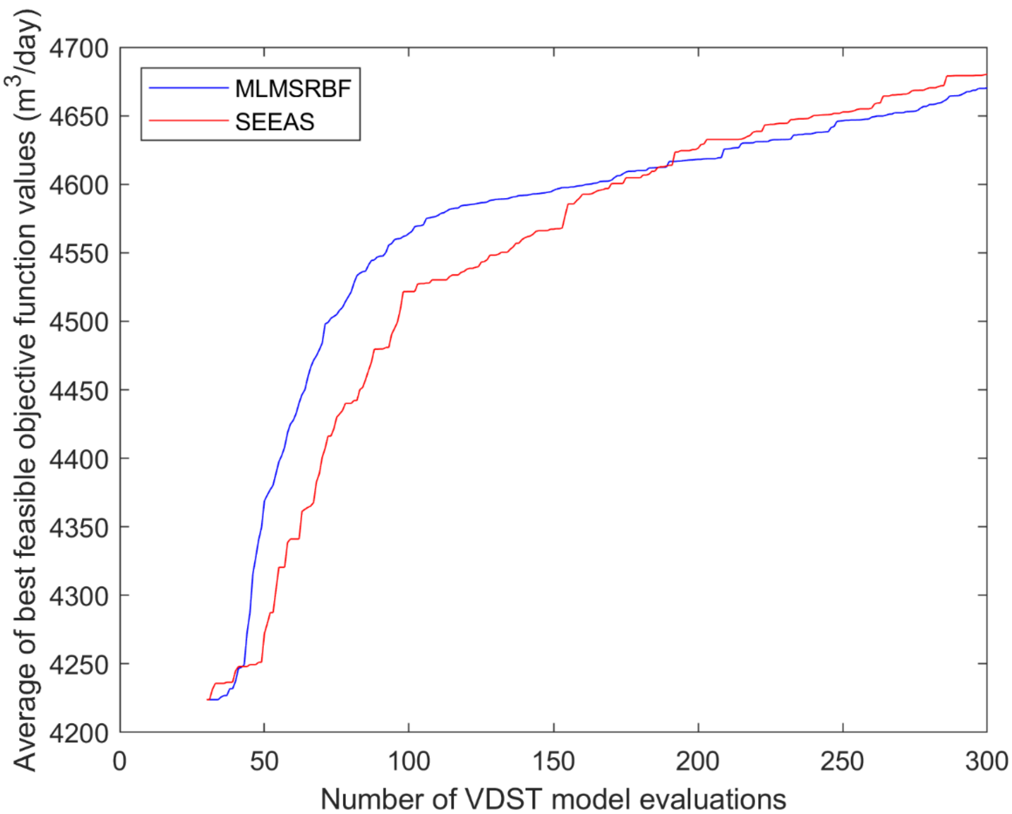


Figure 5-6 Convergence curves of the SBO methods, based on their average best feasible objective function value from the 30 independent runs and for $MHFr = 300$ (plotting starts at $m = 30$ initial design points).

5.5 Comparison of SBO methods using surrogate models for the constraints

It is worth reiterating that pumping optimization problems of coastal aquifers typically involve nonlinear inequality constraints (Mantoglou et al. 2004b). There are examples in coastal aquifer management that by using clustering/zonation methodologies the number of decision variables and constraints were reduced to simplify the optimization problem and thereby facilitating the training of surrogate models (e.g. Ataie-Ashtiani et al. 2014). In the present work, it is assumed that it is of practical importance to include the operation of each pumping well in the management plan and that all pumping wells are associated with nonlinear constraints (see Chapter 3).

In this section, the SBO frameworks focus on exploiting the information from each constraint function by constructing individual surrogate models to predict the quantity of interest. In such cases, one should consider the increase in computational effort to build separate surrogate models for each constraint and the implications when the number of associated constraint functions is large. However, the RBF models utilized in this thesis have small training time and there are examples in the literature of their efficiency and effectiveness in high-dimensional nonlinear constrained problems (e.g. Regis 2014). On the other hand, the KRG models are computationally more intensive, especially in high dimensions, but possess other advantages such as the error estimates calculated with each prediction. The latter feature of KRG models is utilized in the implementation of EGO algorithm presented in the previous section.

It is reminded that the KRG models are utilized here within the surrogate-assisted evolutionary strategy only. As this strategy focuses on local exploitation with the metamodels it is of interest to examine if the predictions skills of the KRG models, which have a more sophisticated structure, outperform the simpler RBF models. Clearly, using the KRG metamodels in a pure exploitation infill strategy does not exploit their full potential in SBO, since the estimated error produced in each prediction is not considered. Nevertheless, keeping the computational effort to a minimum is important and if simpler surrogate models can provide comparable optimal solutions, this is preferable. Table 5-2 summarizes the abbreviated names for the SBO methods and for the surrogate models that were described in chapter 4.

Table 5-2 List of abbreviations and short descriptions of the SBO methods and the surrogate models.

Abbreviation	Description
EAS-PB	Surrogate-assisted evolutionary strategy using the prediction-based exploitation infill method
CSEEAS	Constrained surrogate-enhanced evolutionary annealing simplex algorithm, based on SEEAS (Tsoukalas et al. 2016)
ConstrLMSRBF	Stochastic surrogate-assisted algorithm based on RBF models (Regis, 2011)
LR-RSRBF	Random search surrogate-assisted algorithm based on local refinement and exploration using RBF models
CUB	Cubic RBF model with polynomial tail
TPS	Thin plate spline RBF model with polynomial tail
G0	KRG model with Gauss correlation function and zero-order polynomial
G1	KRG model with Gauss correlation function and first-order polynomial
BE	Only the best surrogate for each constraint function is used
OW	An ensemble surrogate model with optimal weights is constructed for each constraint function

5.5.1 Optimal solutions for the case of 10 pumping wells

The results obtained from the SBO methods are discussed next for the case of 10 pumping wells which sets the dimensionality of the optimization problem to $k=10$. A set of 30 LHS designs of size $m \times k$, were evaluated with the VDST model to provide the initial training points for all algorithms and for all 30 individual optimization runs. The EAS-PB algorithms used an initial training sample of $m=80$ points which also serves as an initial population for EAS and satisfies the $8 \times k$ recommended population size (Kourakos and Mantoglou 2009). For ConstrLMSRBF, CSEEAS and LR-RSRBF two alternative computational budgets of $MHFr = 100$ and $MHFr = 300$ were employed. That is, 100 and 300 VDST model runs were available for the implementation of these methods. Based on the empirical rule $m = \max\left[2 \times k + 1, 0.1MHFr\right]$, the initial training sample of $m = 21$ was used in the case of $MHFr = 100$ and $m = 30$ for the case of $MHFr = 300$.

Table 5-3 Sample statistics of the feasible optimal solutions obtained from the 30 independent optimization runs for 10 pumping wells (best results are in bold and the benchmark solution is underlined).

Optimization method	Worst	Best	Mean	StDev	VDST runs*	Time (hr) **
EAS-VDST		<u>4857.5</u>			4967	14.45
EAS-PB(TPS)	4545.4	4852.6	4810.9	69.12	193	0.84
EAS-PB(CUB)	4700.2	4853.9	4823.4	36.38	181	0.81
EAS-PB(G0)	4517.2	4854.2	4802.1	81.57	175	0.79
EAS-PB(G1)	4491.7	4852.2	4797.5	102.18	160	0.76
EAS-PB(BE)	4517.2	4851.9	4808.3	81.81	210	1.11
EAS-PB(OW)	4545.4	**4858.5**	4810.5	75.42	186	1.35
CSEEAS^{100}	4757.7	4842.2	4803.8	22.66	100	0.35
$\text{ConstrLMSRBF}^{100}$	4723.6	4850.6	4813.8	25.72	100	0.31
LR-RSRBF^{100}	4802.0	4854.2	4839.2	12.32	100	0.33
CSEEAS^{300}	4783.9	4852.1	4828.1	19.57	300	0.95
$\text{ConstrLMSRBF}^{300}$	**4834.1**	4856.5	**4847.3**	**5.17**	300	0.93
LR-RSRBF^{300}	4821.6	4856.3	4846.1	7.29	300	1.10

* For EAS-PB methods this is an average value from the 30 optimization runs, ** average computational time

Regarding computational savings, it is evident that SBO reduced the computational time significantly compared to the direct optimization with the VDST model. A 90-98 percent reduction in time was achieved, depending on the method used and the computational budget that was specified to terminate the SBO algorithm. Obviously, forcing the algorithm to stop at 100 VDST model runs, produced the larger reduction in computational time. On the other hand, the EAS-PB algorithms which employ the multiple surrogate approach required more computational time. This is attributed to the extra steps included in the multiple surrogate approach, such as, the internal optimization task for calculating optimal weights of the ensemble and the cross-validation method to select the best surrogates for each constraint function. For LR-RSRBF, it has been already discussed that the step which involves the search with a global optimization algorithm increases the overall computational time. It is reminded that the number of evaluations with the VDST model varies for the EAS-PB methods, since the stopping criteria of the algorithm are not based on predefined computational budgets. Practically, this means that as the surrogate models predict better optimum values than the current optimum, the VDST model is called by the EAS-PB algorithm to evaluate the feasibility of the proposed solution regardless of the number of VDST model runs.

What is important though, is to assess if the computational gains were associated with a successful exploration of the search space by the SBO algorithms. A first thing to note, is that the performance of the SBO algorithms was markedly improved by using separate surrogate models for the constraint functions compared to just using a single surrogate for the penalized function (see Table 5-1). This was also the case for the lower computational budget where $MHFr = 100$, as demonstrated with the results obtained from CSEEAS, ConstrLMSRBF and LR-RSRBF algorithms. All SBO methods found an optimal solution which is notably close to the benchmark optimum in at least one of the 30 trials, as indicated by the best value in Table 5-1. Given the calculated sample statistics, the use of more sophisticated surrogate models, such as KRG or the multiple surrogate approach, did not provide any noticeable advantages over the simpler single surrogate approach using RBF models.

However, if one should pick the most promising method from the multiple surrogate frameworks, the case of the ensemble with optimal weights EAS-PB(OW) marginally outperforms the EAS-PB(BE). Interestingly, EAS-PB(OW) in one of the optimization trials found an optimal solution which is slightly better than that obtained from EAS-VDST. This is possible since the direct optimization with the VDST model was performed only once to get the benchmark solution and it cannot be concluded as the “true” global optimum. It is noted though, that the extra optimization step for the weight calculation of the ensemble, complicates the algorithm and increases the computational cost considerably. However, the solution found by EAS-PB(OW) is an indication that the algorithm successfully located the region of the global optimum.

Among the EAS-PB algorithms, EAS-PB(CUB) had the best performance, as indicated by its mean and standard deviation values. In fact, it performed well against CSEEAS, ConstrLMSRBF and LR-RSRBF although it follows a less comprehensive infill strategy. It is reminded that EAS-PB follows an aggressive sampling strategy focused on local exploitation and thus, there is a point where a further increase in the VDST model runs will not improve the algorithm’s capability to escape from a possible local optimum. Practically, the average value of 181 VDST model runs for the EAS-PB(CUB) algorithm means that convergence is achieved around that number of VDST evaluations for this specific problem. Adding more training points from VDST simulations will only provide minor improvements on the current optimum and eventually will deteriorate the prediction skills of the RBF models due to closely sampled points. That is also the reason for setting a distance criterion from previously evaluated points as described in chapter 4. The fact that EAS-PB(CUB) had a sample mean close to the benchmark solution and a relatively low standard deviation, shows that for a moderate dimensionality of 10 pumping wells, this approach will likely provide optimal solutions of good quality in a fraction of the time required by the EAS-VDST optimization.

The sample mean along with the standard deviation values are indicative of how consistent the performance of each SBO algorithm is. A simple metric for the sample means can be defined as $F_{mean} = \left| Y_{HF} - Y_{mean} \right|$, where Y_{HF} is the benchmark solution and Y_{mean} is the sample mean of the SBO method. The lower the value the closer is the sample mean to the solution with the VDST model. In quantitative terms, F_{mean} also shows the difference of total pumping in cubic meters per day, between the benchmark solution and the sample mean.

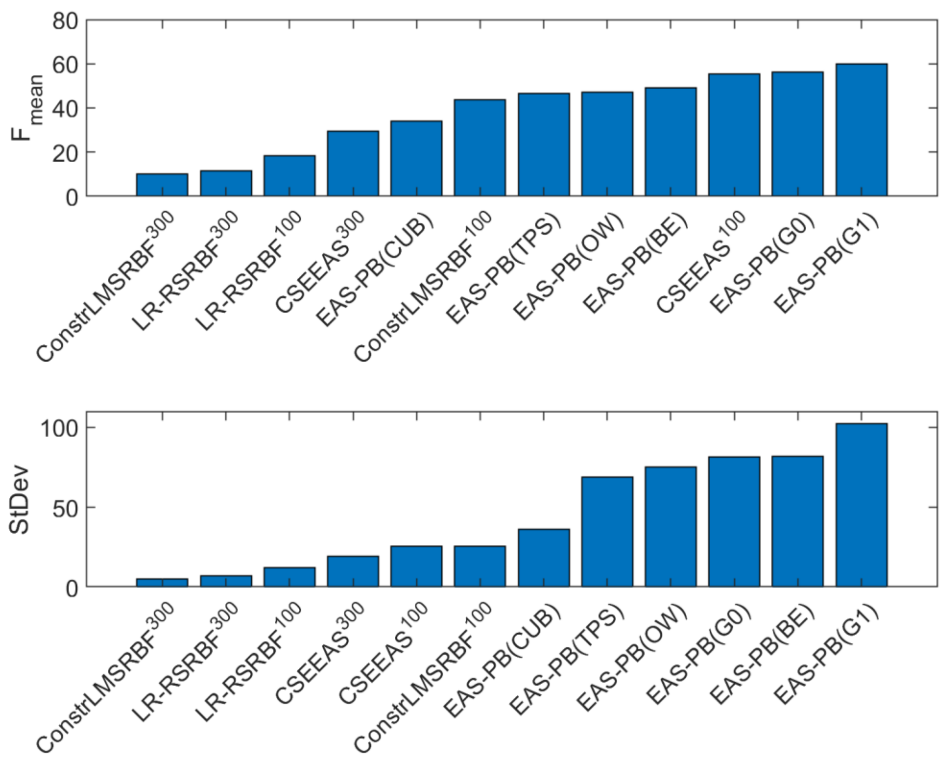


Figure 5-7 The performance of all SBO algorithms based on the metric F_{mean} (upper plot) and the calculated standard deviation (lower plot) from the 30 independent optimization trials.

Based on the ranking shown in Figure 5-7, the SBO algorithms which involve both exploration and exploitation steps appear more reliable than the pure exploitation strategy which is pursued with the EAS-PB methods. The most reliable performance is observed with the ConstrLMSRBF algorithm when the computational budget increases to $MHFr = 300$ (i.e., ConstrLMSRBF^{300}). Interestingly, LR-RSRBF performed equally well with ConstrLMSRBF for $MHFr = 300$ while it performed better than ConstrLMSRBF in the case of $MHFr = 100$. CSEEAS also showed a reliable performance especially for the computational budget of $MHFr = 300$. In general, CSEEAS, ConstrLMSRBF and LR-RSRBF algorithms had amongst the best mean values along with the lowest standard deviation values, which indicates a reliable performance independent of initial training designs and random operations of each algorithm. EAS-PB(CUB) was ranked 5th and 7th for F_{mean} and standard deviation values, respectively, which is the best rank among the EAS-PB methods.

Another useful metrics to compare the SBO algorithms are the actual relative improvement I, the maximum possible relative improvement $I_{\max}$ and the relative improvement ratio r_I (Viana et al. 2010a). These quantities are particularly useful in our case where the algorithms in comparison have different convergence criteria. They are defined as follows:

$$I = \frac{y_{inB} - y^*}{|y_{inB}|},\ I_{\max} = \frac{y_{inB} - y^*_{HF}}{|y_{inB}|},\ r_I = \frac{I}{I_{\max}} \tag{5.1}$$

where y_{inB} is the initial best solution (could be the one from the initial design points), y^* is the actual solution found from the SBO algorithm and y^*_{HF} is the global optimum known from the optimization with the HF model. If $I = 0$ means that the SBO algorithm did not improve further from the known starting best feasible solution. When $I > 0$ means that there is improvement over the starting feasible solution which could be significant depending on the progress of the algorithm. Accordingly, $I_{\max}$ sets a standard of how far the initial best feasible point from the "true" global optimum is. Therefore, a ratio r_I closer to 1 suggests an algorithm that produced significant progress during optimization and found a solution in the region of the global optimum. These measures can be used in a progress plot showing the improvement over the number of HF model runs.

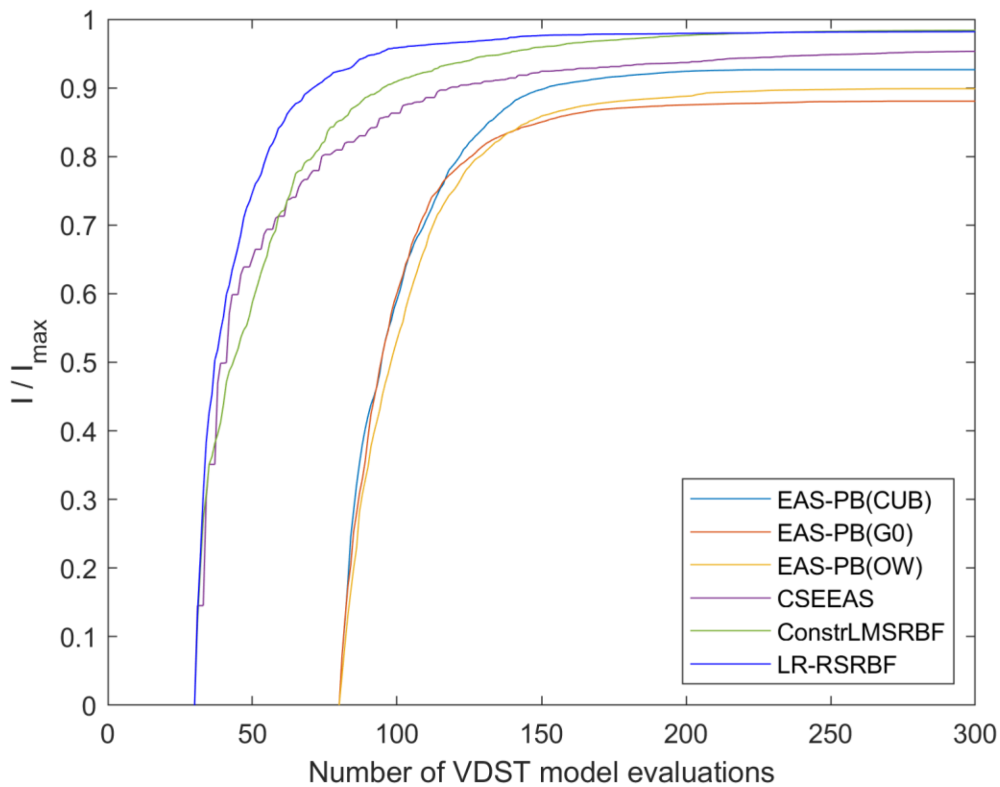


Figure 5-8 Mean of the relative improvement ratio $r_I = I/I_{\max}$ (from the 30 optimization trials) for $MHFr = 300$. Plotting starts at $m = 21$ initial design points for CSEEAS, ConstrLMSRBF and LR-RSRBF while it starts at $m = 80$ initial design points for the EAS-PB methods.

As shown in Figure 5-8 the mean relative improvement ratio for six SBO algorithms for the case of $MHFr = 300$ is demonstrated. Three of them belong to the EAS-PB method which performed best in their category, given their sample statistics. Thus, we select the EAS-PB(CUB) from the implementation with the RBF models, the EAS-PB(G0) from the implementation with the KRG models and the EAS-PB(OW) from the implementation with the multiple surrogate models. It is reminded that these SBO algorithms do not terminate on a specific computational budget and thus each trial may or may not have used 300 VDST model runs. However, as it was explained earlier, by adding more HF runs only minor improvements are expected, if any at all. As the plot indicates, at a certain number of VDST evaluations their progress has reached a plateau. The other three methods in the plot are CSEEAS, LR-RSRBF and ConstrLMSRBF which utilize the number of HF model runs as a stopping criterion.

Using 300 runs with the VDST model represents a moderate computational budget which emulates the scenario of a relatively time-consuming VDST model. In such cases, finding good solutions can be challenging. It appears that all six SBO methods performed well, exhibiting steep improvements on their mean r_I values within the first 100 function evaluations of the VDST model. LR-RSRBF, climbs faster than other algorithms to $r_I = 0.9$ in fewer than 100 evaluations with the VDST model, demonstrating its capability to guide the search to promising regions fast. CSEEAS has the third best progress while the performance of ConstrLMSRBF, after 80 VDST model runs, is markedly improved and is similar to LR-RSRBF. Among the EAS-PB methods the EAS-PB(CUB) has the best progress and after 150 VDST model runs exhibits a good performance closer to the CSEEAS algorithm.

Based on the present results, there is a strong indication that some of the proposed methods deliver good and reliable solutions within much less computational effort than the EAS-VDST approach. There is always room for an improved performance for each of these SBO algorithms by tuning specific parameters. This is especially the case for the more comprehensive approaches such as ConstrLMSRBF, CSEEAS and LR-RSRBF. However, this would have made the intercomparison even more complicated. As discussed in Razavi et al. (2012), this is one of the puzzling things to consider when applying SBO methods and obviously the user's experience with each individual algorithm as well as with the design of the optimization problem might have a positive or negative effect on the success of each method. Some aspects of the effect of parameter tuning on the SBO algorithms performance are discussed below.

As mentioned before in chapter 4, the original SEEAS utilizes the dimensionless progress index $PI = \log\left(HFr\right)/\log\left(MHFr\right)$, where HFr is the current number of runs with the HF model while $MHFr$ defines the available computational budget. The PI index, sets the weight for the acquisition function based on the empirical formula $w = \max\left[0.75, \min\left(PI, 0.95\right)\right]$ in the original paper (Tsoukalas et al. 2016). As more iterations are carried out by the algorithm and thus more HF model runs are added, the value of PI increases and the minimization of the acquisition function focuses on exploitation. For the present implementation of CSEEAS algorithm, it was observed that the performance of the algorithm was sensitive to the contribution of exploration through the weighting formula $w = \max\left[0.75, \min\left(PI, 0.95\right)\right]$. In Tsoukalas et al. (2016), SEEAS, initially designed for optimization problems with bound constraints only, was tested for computational budgets of $MHFr = 500$ and of $MHFr = 1000$. As this is a first attempt to modify SEEAS for nonlinear constrained optimization problems and its performance was tested for lower computational budgets of $MHFr = 100$ and $MHFr = 300$, we examined different values of weights given to exploration from the beginning of the optimization. For this specific optimization problem and for the lower computational budget of $MHFr = 100$, it was observed that CSEEAS performed better and more reliably when the minimization of the acquisition function was oriented to exploitation from earlier stages of the optimization. Therefore, exploration was set to a minimum by setting $w = 0.95$. For the case of $MHFr = 300$, we set $w = \max\left[0.85, \min\left(PI, 0.95\right)\right]$ as optimal choice. This selection for the weighted acquisition function could be problem dependent, and thus, it cannot be concluded that the above settings for CSEEAS would also be optimal in similar cases with small computational budgets.

ConstrLMSRBF algorithm is considered as a promising method for pumping optimization problems in coastal aquifers as it was demonstrated in Christelis et al. (2018). It is expected that the present optimization problem has multiple local optima and the stochastic features of ConstrLMSRBF show promise in quickly improving the objective function value within a small number of iterations. ConstrLMSRBF generates candidate points for evaluation with the RBF metamodels, by applying normal random perturbations on all or on a subset of the coordinates of the best feasible solution found so far (Regis 2011). By varying the algorithm settings for the present optimization problem, it was observed that in our case the performance of ConstrLMSRBF was better if all coordinates were perturbed to generate new candidate points. Since LR-RSRBF adopts the perturbation scheme of ConstrLMSRBF is expected to be also sensitive to the same parameters although in ConstrLMSRBF the standard deviation varies dynamically while in LR-RSRBF remains constant during the optimization steps.

5.5.2 Increasing the dimensionality to 20 pumping wells

The increase in dimensionality of the optimization problem challenges the effectiveness of the SBO algorithms in locating the region of the global optimum. Therefore, the comparison of algorithms' performance is also conducted for the case of 20 pumping wells. The computational budgets are again $MHFr = 100$ and $MHFr = 300$ for CSEEAS, ConstrLMSRBF and LR-RSRBF.

The initial training points this time are $m = 41$ based on $m = \max\left[2\times k + 1, 0.1MHFr\right]$ with $k = 20$. An initial training sample of $m = 160$ points was created for the EAS-PB algorithms that satisfies the $8\times k$ criterion for the initial population size. It should be noted that although the $8\times k$ value is selected as a preferable choice for EAS-PB, it does not mean that its performance will necessarily degrade if smaller initial population sizes are selected. It is obvious though, that as the dimensionality of the optimization problem increases, spending many HF model runs on the initial population evaluation is not ideal and certainly puts a lot of effort on a random design instead of a more guided search towards promising solutions.

Table 5-4 Sample statistics of the feasible optimal solutions obtained from the 30 independent optimization runs for 20 pumping wells (best results are in bold and the benchmark solution is underlined).

Optimization method	Worst	Best	Mean	StDev	VDST runs*	Time (hr) **
EAS-VDST		<u>4832.2</u>			11020	33.68
EAS-PB(TPS)	4540.7	4824.3	4698.3	73.90	244	1.23
EAS-PB(CUB)	4545.7	4815.0	4718.5	68.82	243	1.23
EAS-PB(G0)	4419.4	4762.7	4628.2	71.46	231	1.19
EAS-PB(G1)	4465.0	4813.0	4662.4	99.39	215	1.14
EAS-PB(BE)	4465.0	4806.6	4659.2	91.38	218	1.41
EAS-PB(OW)	4465.0	4804.6	4672.5	92.28	186	1.52
CSEEAS^{100}	4596.7	4775.2	4684.3	44.33	100	0.35
$\text{ConstrLMSRBF}^{100}$	4546.4	4768.5	4690.4	60.72	100	0.32
LR-RSRBF^{100}	4661.6	4793.0	4729.4	40.70	100	0.34
CSEEAS^{300}	4697.1	4822.1	4763.0	33.67	300	0.97
$\text{ConstrLMSRBF}^{300}$	4656.8	4814.5	4768.6	31.19	300	0.94
LR-RSRBF^{300}	**4729.0**	**4829.5**	**4792.5**	**23.19**	300	1.47

* For EAS-PB methods this is an average value from the 30 optimization runs, ** average computational time

Table 5-4 presents the sample statistics for the case of 20 pumping wells and for the same SBO algorithms that were tested before. Notable computational savings are achieved despite that for each algorithm 20 separate surrogate models were constructed and updated during optimization. However, the increased dimensionality affected the performance of the SBO algorithms. In the previous case of 10 pumping wells, the maximum difference between the sample mean of EAS-PB(G1) and the HF global optimum was $61.5\, m^3/day$. Now, the maximum difference is $204\, m^3/day$ between the sample mean of EAS-PB(G0) and the HF solution from the VDST model. The best solution with the SBO algorithms was found in one of the optimization trials with LR-RSRBF for $MHFr = 300$. For the same computational budget, LR-RSRBF had the best lowest feasible solution, the best mean value, and the lowest standard deviation. However, it was also the second most expensive SBO algorithm due to the internal global optimization step which this time involved 20 pumping wells (see 4.4.5 for details).

Based on the mean values, CSEEAS, LR-RSRBF and ConstrLMSRBF struggled to approximate the region of global optimum for $MHFr = 100$ as successful as in the case of 10 pumping wells. It is noted that for all SBO algorithms, we kept the same settings as in the optimization problem with $k = 10$. Apparently, tuning those settings may improve the performance of the algorithms, however, this would further complicate the intercomparison. The SBO methods based on the EAS-PB approach utilized generally a larger number of VDST model evaluations than before until they converged to an optimum. Again, given the sample statistics, EAS-PB(CUB) appears to perform better than the other EAS-PB algorithms.

As with the case of 10 pumping wells, Figure 5-9 illustrates the mean relative improvement ratio. This time, only four SBO algorithms are compared for the case of $MHFr = 300$. Only one belongs to the EAS-PB method, namely, EAS-PB(CUB). The other three methods in the plot are CSEEAS, LR-RSRBF and ConstrLMSRBF which utilize the number of HF model runs as a stopping criterion and balance exploration and exploitation using the surrogate models.

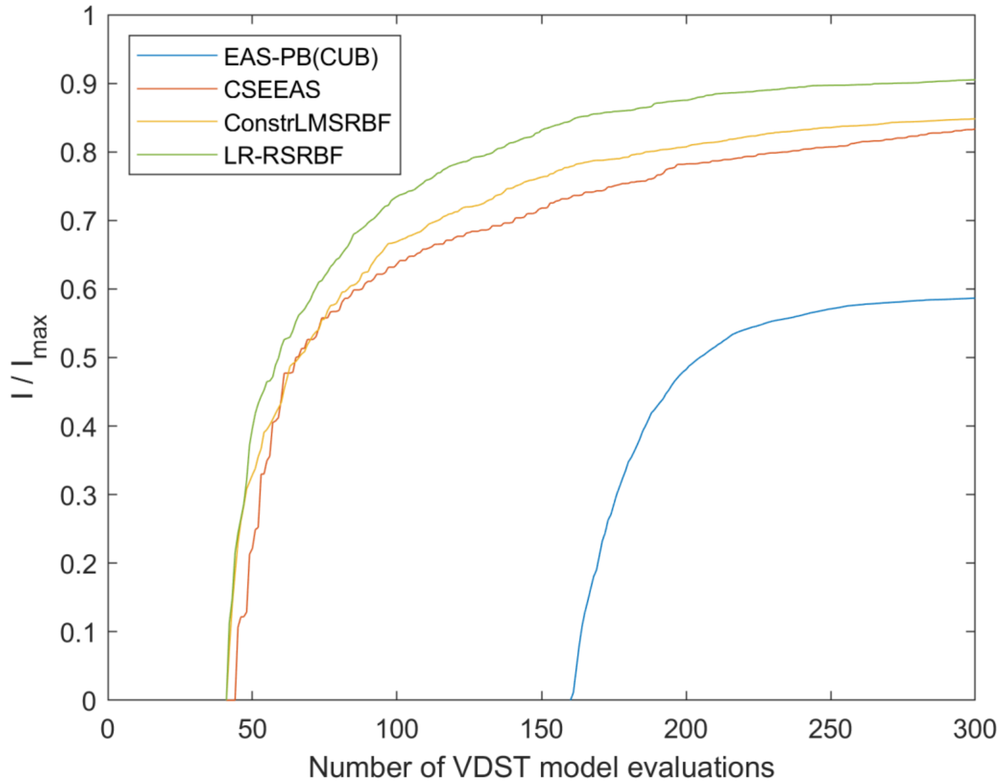


Figure 5-9 Mean of the relative improvement ratio $r_I = I/I_{\max}$ (from the 30 optimization trials) for $MHFr = 300$. Plotting starts at $m = 41$ initial design points for CEEAS, LR-RSRBF and ConstrLMSRBF while it starts at $m = 160$ initial design points for the EAS-PB methods.

While in the case where $D = 10$ EAS-PB(CUB) approached a value of $r_I = 0.9$, here, it appears that the algorithm hardly reaches a mean relative improvement ratio of $r_I = 0.6$. On the other hand, CSEEAS, LR-RSRBF and ConstrLMSRBF performed much better by having an improvement of $r_I = 0.6$ in fewer than 100 VDST model runs while LR-RSRBF exhibits the most promising progress among the SBO algorithms reaching a mean improvement ratio of $r_I = 0.9$ at the end of the 300 evaluations with the VDST model. A comparison between LR-RSRBF and ConstrLMSRBF as well as between LR-RSRBF and CSEEAS, using the non-parametric Wilcoxon rank sum test, returned in both cases $h = 1$ and p-values in the range of 10^{-4}. Thus, at least for the present problem and for $MHFr = 300$, LR-RSRBF is preferable over the other two SBO algorithms.

5.6 Multi-fidelity optimization for limited computational budgets

In the previous numerical experiments, the available computational budgets were adequate to build surrogate models and use a considerable number of HF VDST model evaluations to search for an optimal solution within each SBO framework. Here, it is hypothesized that the VDST model is extremely time-consuming and therefore we can only afford small computational budgets of $MHFr = 11$ and of $MHFr = 21$. The $MHFr = 11$ budget, marginally satisfies the requirement of $k + 1$ training points for the case of 10 pumping wells. Practically that means we certainly cannot use the previous conventional SBO methods and thus we can either only evaluate a sampling design of 11 points to look for a feasible solution or use a low-fidelity but computationally less expensive model to run the optimization.

An alternative approach in such cases, is to combine the available HF model evaluations with a substantially larger set of input-output data from low-fidelity (LF) models. As mentioned before, we utilize here the method of co-Kriging to achieve this combination of training points from the LF sharp interface model and the HF VDST model (see section 4.4.6). As it was shown previously, in the presence of nonlinear constraints the SBO methods perform better when separate surrogate models are built for each constraint function. However, training coKRG models for each constraint function requires larger computational effort than the RBF and the KRG models. It involves running first a considerable number of simulations with the LF model and then fit each coKRG model on selected training points. It also includes a separate optimization task, like KRG models, to estimate the required coKRG parameter set. Due to this increased computational cost, the same problem is also solved by constructing only a single coKRG model for the penalized objective function of the VDST and the sharp interface models. This approach aims to investigate if the errors from ignoring the constraint responses are smoothed out due to the knowledge-based approach of the LF model despite its known inaccuracies.

In general, the multi-fidelity methodology developed here assumes that the sharp interface model is capable to explain part of the VDST model behavior. Our effort was focused on using the two models in an iterative multi-fidelity optimization framework for the case of limited VDST samples. In the first scenario where $MHFr = 11$, it was assumed that the available VDST model runs are not adequate to construct conventional surrogate models and thus we need to resort to multi-fidelity modelling. In the second scenario, where $MHFr = 21$, it is possible to acquire samples from the VDST model and construct surrogate models but the overall computational budget is very low, meaning that only a few runs with the VDST model are available after the initial training. For the second case, the multi-fidelity optimization approach is compared against the robust ConstrLMSRBF algorithm. Therefore, a total of four optimization methods are compared in this section:

i. Optimization using the LF sharp interface model of Strack (1976) as an alternative to the HF EAS-VDST approach. It will be denoted as EAS-SH.

ii. An adaptive-recursive (AR) optimization framework presented in chapter 4, using coKRG models for each of the constraint functions. It is implemented for two computational budgets, i.e., $MHFr = 11$ and $MHFr = 21$. For convenience, the method will be denoted as AR-coKRG$^{\mathrm{CONS}}$.

iii. An adaptive-recursive (AR) optimization framework presented in chapter 4, using coKRG models only for the penalized objective function output. It is implemented only for $MHFr = 21$. For convenience, the method will be denoted as AR-coKRG$^{\mathrm{OBJ}}$.

iv. The ConstrLMSRBF algorithm as a conventional SBO method to compare against multi-fidelity optimization, for the computational budget of $MHFr = 21$. Practically, 11 training points are used to build the cubic RBF models and the rest 10 VDST simulations are utilized to update the surrogates and evaluate promising points.

It is noted that apart from the known optimum obtained with the VDST model for the case of 10 pumping wells, the EAS-SH optimization was also run once to get the corresponding optimum. The SBO methods run for a set of 30 independent optimization trials to obtain a statistical meaning of algorithms' performance.

Table 5-5 Comparison of optimal solutions found with the AR-coKRG SBO method, the sharp interface model, the VDST model and the ConstrLMSRBF method for $MHFr = 21$ are also presented. (The optimal solutions with the VDST model and the sharp interface model are underlined).

Optimization method	Worst	Best	Mean	StDev	VDST runs	Sharp runs	Time (hr)
EAS-VDST		<u>4857.5</u>			4967	NA*	14.45
EAS-SH		<u>4049.2</u>			NA*	4164	1.6
AR-coKRG^{CONS}	3866.0	4740.0	4363.2	241.96	11	100	0.12
AR-coKRG^{CONS}	4155.7	4730.1	4480.9	150.82	11	200	0.18
AR-coKRG^{CONS}	**4201.9**	4773.5	4552.3	**144.51**	21	100	0.31
AR-coKRG^{CONS}	3966.3	**4815.1**	**4605.0**	187.38	21	200	0.56
AR-coKRG^{OBJ}	4106.1	4763.6	4415.8	167.39	21	200	0.45
ConstrLMSRBF	4004.8	4740.6	4397.3	205.94	21	NA*	0.095

* NA: Not Applicable

Based on the above results from Table 5-5, it is evident that on average the multi-fidelity optimization provided a better solution than using the sharp interface model alone. The mean values obtained from the AR-coKRG^{CONS} method are considerably higher than the EAS-SH approach, even for the case of 11 model evaluations with the VDST model. This is also true for

AR-coKRG^{OBJ}, where only one surrogate model is constructed to combine the output of the penalized objective function from the sharp interface model and from the penalized objective function of the VDST model. Interestingly, in one of the optimization trials AR-coKRG^{CONS} found a solution equal to $4815.1 m^3/day$ which is close to the known optimum with the VDST model. These generally promising results from the multi-fidelity optimization are attributed to the capabilities of coKRG models as well as to the effectiveness of the local refinement sampling strategy included in ConstrLMSRBF algorithm which is utilized in our multi-fidelity method (see section 4.4.4 and 4.4.6).

We also tested the impact of using more LF simulations on the accuracy and performance of the AR-coKRG^{CONS} method. Thus, we set two additional scenarios where the LF model runs are set to 100 and 200, respectively. Notably, the increase in the LF simulations improved the performance of the AR-coKRG^{CONS} method. However, the non-parametric Wilcoxon rank sum test returned $h=0$ with a p-value of 0.0679 for $MHFr=11$, and $h=0$ with a p-value of 0.0575 for $MHFr=21$, which means that the two groups are independent samples from continuous distributions with equal medians, at the 5% significance level. Nevertheless, the comparison between AR-coKRG^{OBJ} and AR-coKRG^{CONS} for $MHFr=21$, returned $h=1$ with a p-value of 9.21×10^{-5} indicating again the better performance of AR-coKRG method when using separate coKRG models for the constraint functions. This improvement comes at the cost of increased computational effort. This would have been more evident if more LF samples were used for an optimization problem or in the case of a larger number of constraint functions. Finally, AR-coKRG^{CONS} outperformed the data-driven ConstrLMSRBF algorithm for a computational budget as small as $MHFr=21$. The non-parametric Wilcoxon rank sum test returned $h=1$ with a p-value of 2.68×10^{-4}, meaning that there is statistically significant difference between the medians of AR-coKRG^{CONS} and ConstrLMSRBF for the given budget.

Chapter 6

Application of SBO to a real-world coastal aquifer model

6.1 Overview of the study area

In this chapter, a VDST numerical model is developed for the case of a real-world coastal aquifer. It is an unconfined, elongated aquifer along the Vathi valley, located at the central part of the Greek island of Kalymnos (Figure 6-1). The aquifer consists mainly of highly permeable limestone which outcrops at the valley margins. The available borehole lithological data indicate that the aquifer's bottom is bounded by an impermeable schist formation which underlies the limestones. The geological formations in the valley plain consist of high permeable scree, alluvium deposits of medium permeability and almost impermeable volcanic formations (tuff) (Mantoglou et al. 2004).

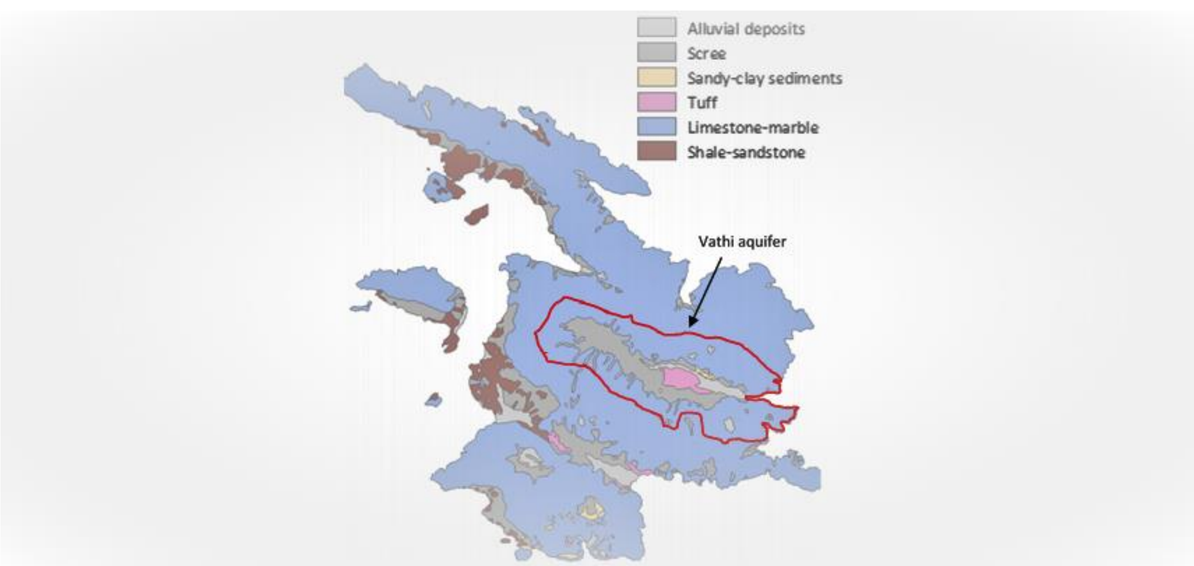


Figure 6-1 Hydro-lithological map of Kalymnos island (Hellenic Ministry of development, 2005).

For the purposes of the present analysis, it is assumed that for the aquifer flow, which partially takes place in the limestone fissures, the concept of equivalent porous medium is valid. Therefore, although the carbonate rocks are characterized by secondary porosity, a uniform hydraulic conductivity is considered to simplify the numerical model development. The aquifer is broadly divided in four uniform hydraulic conductivity zones, as shown in Figure 6-2.

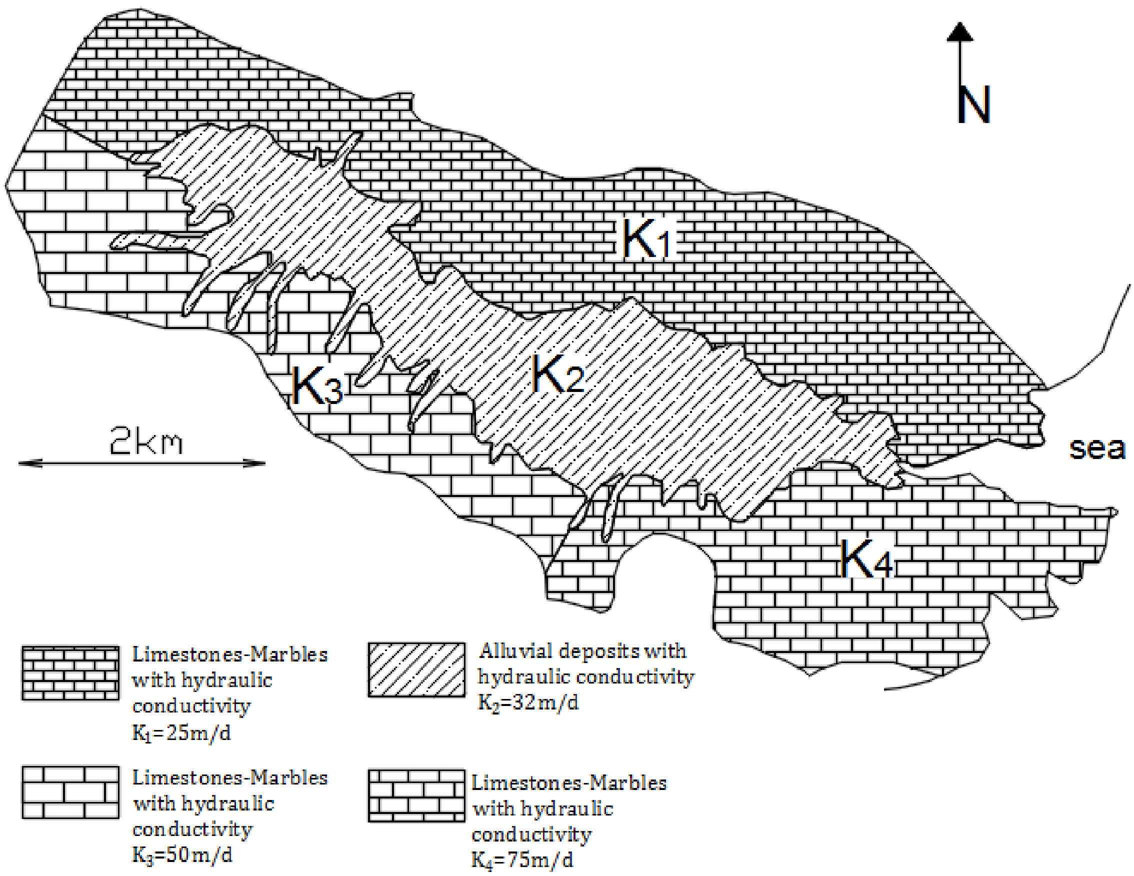


Figure 6-2 Hydraulic conductivity zones of the Vathi aquifer

Figure 6-3 depicts the four different recharge zones of the aquifer, with the limestones receiving the largest amount of recharge due to the presence of large crevices which facilitate the surface water percolation (Mantoglou et al. 2004). On the contrary, the alluvium parts mixed with clay receive reduced recharge rates of $R = 30\,mm/yr$ while the tuffs are considered impermeable with no recharge. Finally, the recharge rate at the scree area is estimated at $R = 70\,mm/yr$ (Mantoglou et al. 2004).

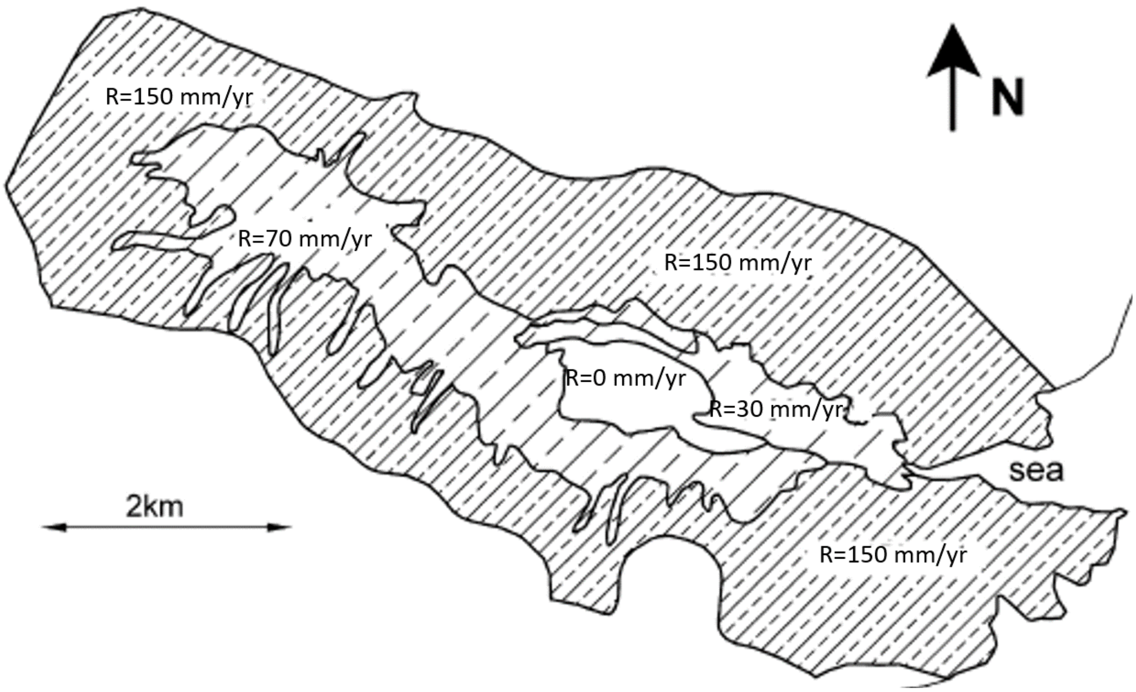


Figure 6-3 The spatial distribution of surface recharge rates across the Vathi aquifer into four different zones (source: Mantoglou et al. 2004).

6.2 Numerical model development

A numerical coastal aquifer model was developed to perform the three-dimensional VDST simulations. Initially, a two-dimensional irregular mesh composed of triangular elements was generated to cover the area of interest and then it was further discretized on the vertical direction to create a three-dimensional layered mesh. That resulted in a total of 44864 prism elements. As illustrated in Figure 6-4, a denser discretization was applied to the east side of the model to simulate more accurately the evolvement of the salinity plume close to the sea boundary. The aquifer's bottom was approximated as horizontal, at a depth of -25 m below sea level. The north, south and west boundaries are considered impermeable. At the east sea boundary, first-type (Dirichlet) boundary conditions are applied. That is, a specified hydrostatic equivalent freshwater head and a constant relative concentration corresponding to the maximum fluid density of seawater are applied to represent sea-boundaries. The aquifer is pumped by 11 pumping wells of known locations.

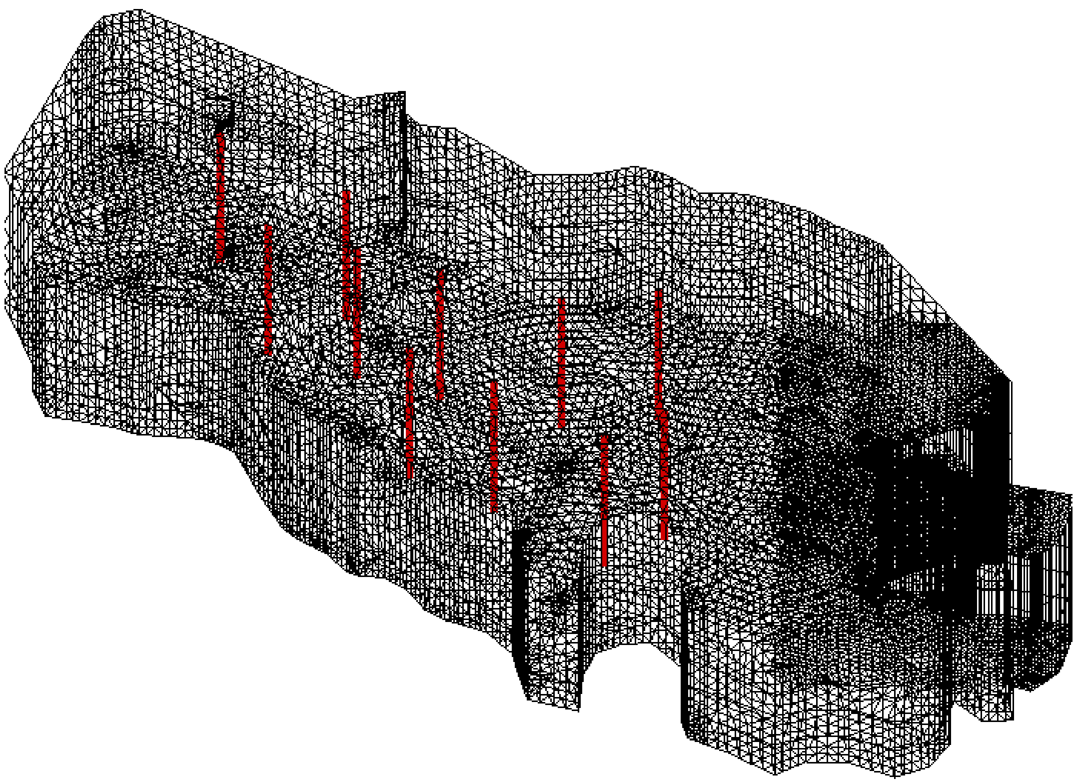


Figure 6-4 Three-dimensional view of the discretized model domain along with the location of the pumping wells shown in red colour.

Due to the lack of hydrogeological data and hydraulic head measurements, it was not possible to perform a reliable and accurate calibration of the aquifer model before optimization. Therefore, the parameter values are rough estimates, based on a trial-and-error procedure (Mantoglou et al., 2004). The aquifer is replenished by a total uniform surface recharge of approximately

$7849\, m^3/day$. A single run of the VDST simulation requires approximately 80.67 seconds (1.34 minutes) running on a 2.7 GHz Intel i5 processor with 8 GB of RAM in a 64-bit Windows 10 system.

6.3 Formulation of the pumping optimization problem

In Mantoglou et al. (2004), a pumping optimization problem was solved for the Vathi aquifer based on the sharp interface model and a hybrid optimization scheme using SQP and genetic algorithms. The authors observed that for the specific well coordinates in Vathi aquifer, the constraints associated with the piezometric head at the wells are those which are active rather than the constraints related to the toe location. Previous runs with the VDST model in our case, showed a similar situation where a narrow salinity front is developed, if the imposed constraints are related only to the salinity levels at the wells (Christelis et al. 2019b).

In the long run, a total pumping rate, based on the toe formulation of constraints, results in a significant lowering of the piezometric head at the wells which is not a sustainable approach. On the other hand, it is desirable to eliminate any inactive constraints from an optimization problem at the outset, to ease the search with the optimization algorithm (Forrester et al. 2008). To that end, instead of using the constraints which are related to salinity levels we reformulate here the optimization problem and impose constraints which maintain the piezometric head at a level greater than zero with respect to the sea level. Therefore, the objective function is penalized again in a way that can be handled by the EAS algorithm as follows:

$$\min\, f(\mathbf{Q}) = \begin{cases} -\sum_{i=1}^{k} Q_i,\ if\ \forall\, i = 1,\dots,k;\ h_i(Q_1,\dots,Q_k) > 0 \\ M_v \sum_{i=1}^{k} \left[\max(h_i, 0)\right]^2,\ if\ \exists\ i = 1,\dots,k;\ h_i(Q_1,\dots Q_k) \le 0 \end{cases} \tag{6.1}$$

where M_v represents the number of pumping wells that the constraint is violated while h_i denotes the piezometric head at the ith pumping well. The above formulation is similar to that presented in chapter 3 (see equations 3.5 and 3.6) but this time, the nonlinear constraints require that the piezometric head is maintained above zero with respect to sea level. However, it is noted that this is a simpler optimization problem since the constraints based solely on the piezometric heads formulate a less strong nonlinearity compared to the examples presented in chapter 5. Although this is in favor of constructing surrogate models with good prediction skills, it can still provide insights on the capabilities of the SBO algorithms.

The present optimization problem includes 11 pumping wells with minimum and maximum pumping flow rates at $Q_{\min} = 0\, m^3/day$ and $Q_{\max} = 1000\, m^3/day$, respectively. Based on the direct optimization with the VDST model, the total pumping rate for the Vathi aquifer is $Q_{tot} = 6184.8\, m^3/day$, given a management plan of 30 years. The optimal solution obtained from the EAS-VDST algorithm, converged after 5189 evaluations of the objective function with the numerical VDST model. The overall computational cost was estimated at 116.28 hours (4.84 days).

The optimal solution from EAS-VDST corresponds to a 78.8% of the total recharge volume of the aquifer which is $7849\, m^3/day$. The optimal solution found presently with the VDST model is stressing the aquifer’s sustainability. Kourakos and Mantoglou (2013) discuss a framework which is closer to sustainable groundwater extraction by performing a second optimization run by using the optimal results from the first optimization run as initial conditions. This reduces significantly the dependance on the initial aquifer storage and was also employed in the previous chapter with the hypothetical aquifer models. It should be further noted that such an approach is more to the safe side and it might provide a low total pumping rate which does not meet the freshwater demands (Kourakos and Mantoglou 2013). Here, it is assumed that the constraints related to the piezometric heads in combination with the locations of the pumping wells maintain the salinity front seaward while providing a reasonable rate of total groundwater extraction. Thus, only the first optimization run is conducted.

Nevertheless, it is possible that in the event of a dramatic decrease in aquifer's recharge, this pumping scheme will result in a significant encroachment of seawater inland. For illustrative purposes, Figure 6-5 compares two recharge scenarios for a 30-year management plan, one with the present recharge and the other for the extreme case of a 35% reduction in total recharge.

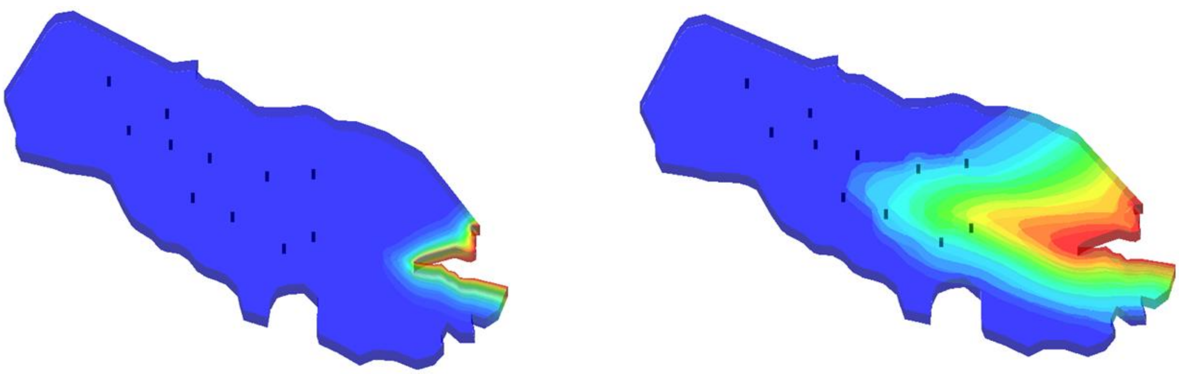


Figure 6-5 Salinity distribution after 30 years of simulation with the optimal pumping rates for the present recharge scenario (left view) and the corresponding output using the same optimal pumping rates for the 35% total recharge reduction (right view).

The above plots demonstrate that the piezometric head constraints apply a strict control on the salinity front by maintaining it seaward. However, if recharge rates decrease significantly, the present management plan is no longer sustainable as shown on the right view of Figure 6-5. The saltwater front moves further inland while increased salinity levels are observed in a large part of the aquifer.

6.4 Optimal results from the SBO algorithms

Given the results from the numerical experiments presented in chapter 5, we employed four of the most promising SBO algorithms, namely, EAS-PB(CUB), CSEEAS, LR-RSRBF and ConstrLMSRBF. Therefore, one algorithm was selected to represent the prediction-based exploitation infill strategy, i.e., EAS-PB(CUB), while the other three belong to methods that balance exploration and exploitation. For those three algorithms a computational budget of $MHFr = 300$ was assumed. This time only one optimization run was conducted for each SBO algorithm.

CSEEAS, LR-RSRBF and ConstrLMSRBF share the same 23 initial training points while the rest 277 evaluations with the VDST model were utilized by the framework of each algorithm to search for promising points and update the surrogate models. The EAS-PB(CUB) method used an initial population of 88 points (vectors of pumping rates) which also served as initial training points for the surrogate models. It is worth to mention that all SBO algorithms were based on cubic RBF models. For this problem, the surrogate models were fitted on pumping rates as input data and the corresponding piezometric heads at each one of the pumping wells as output data. In total, 11 surrogate models were employed to predict the hydraulic head values to unseen vectors of pumping rates during the operations of the optimization frameworks.

Table 6-1 Optimal results from the optimization with the VDST model alone as well as with the SBO methods. (Best result is in bold and the benchmark solution is underlined)

Optimization method	Optimal solution (m^3/day)	VDST runs	Time (hr)
EAS-VDST	<u>6184.8</u>	5189	116.28
EAS-PB(CUB)	6184.4	125	2.59
CSEEAS	6176.3	300	6.28
ConstrLMSRBF	6174.3	300	5.57
LR-RSRBF	**6185.4**	300	6.29

The results presented in Table 6-1 demonstrate that all SBO algorithms found optimal solutions which are remarkably close to that obtained from the EAS-VDST method. EAS-PB(CUB) converged in only 125 evaluations with the VDST model and provided the largest reduction in computational time which is almost 98% compared to EAS-VDST. It is reminded that EAS-PB(CUB) relies on the convergence criteria of the original EAS algorithm and it does not use a specific computational budget as a stopping criterion, unlike to the other three SBO methods. These methods, due to the predefined computational budget of 300 VDST model runs, required more time to converge providing a smaller reduction of 94% of the VDST-based optimization.

Given this single run for each SBO algorithm, the best optimal solution was found by LR-RSRBF which is slightly better than the one obtained from EAS-VDST. Again, this is possible since all algorithms involve random operations, and a single run cannot ensure finding the “true” global optimum. Interestingly, EAS-PB(CUB) performed equally well with the other three more comprehensive SBO algorithms and found a solution in the region of global optimum in only 125 VDST runs. To our understanding, this is attributed to the present formulation of the constraints which are not as highly nonlinear as in the case with the salinity-based constraints used in chapter 5 while the number of decision variables is moderate. Here, the optimization problem has probably fewer local optima and a less rugged objective function landscape which implies that a simpler SBO algorithm like EAS-PB(CUB) might be quite adequate.

Figure 6-6 presents the optimal pumping rates from each optimization framework. The distribution of the pumping rates from EAS-PB(CUB) and LR-RSRBF exhibit a notable similarity with that obtained from EAS-VDST. The fact that LR-RSRBF found a slightly better solution with similar distribution of pumping rates strengthens the possibility that this solution is near the global optimum.

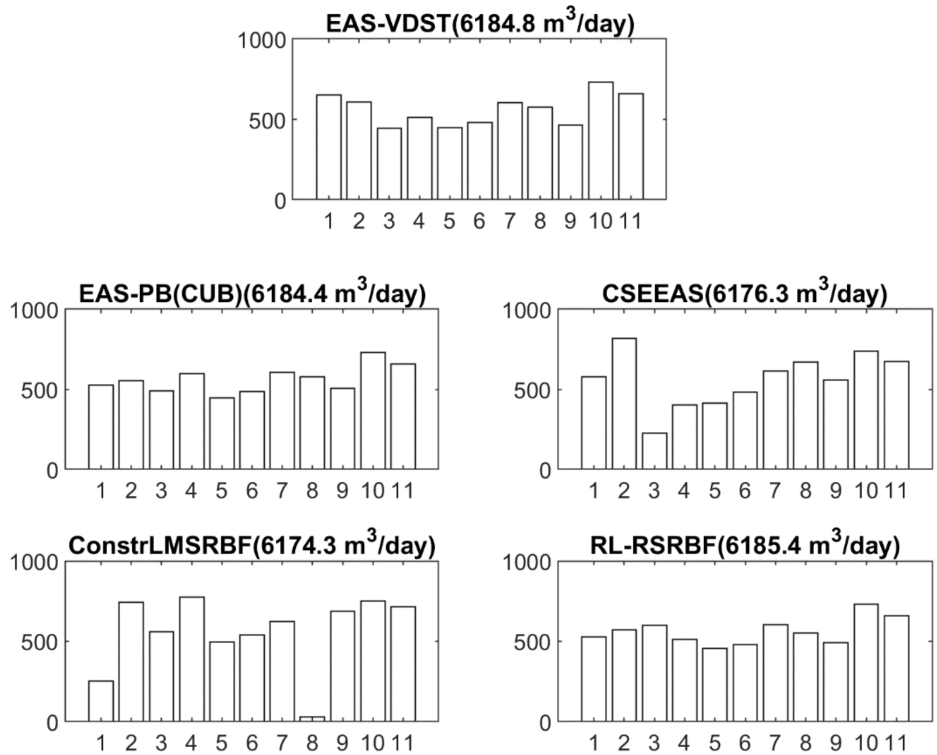


Figure 6-6 Distribution of the optimal pumping rates. The number in the parenthesis corresponds to the total pumping rate obtained from each optimization framework.

6.5 Testing the multi-fidelity approach

The multi-fidelity optimization framework that was developed for cases of extremely time-consuming VDST models, is also tested here for the real-world coastal aquifer model. Although in the previous numerical experiments AR-coKRG^{CONS} performed better than AR-coKRG^{OBJ}, it also increases the computational cost due to the separate construction of coKRG models for each constraint function. Here, given that the nonlinearity in the constraints is less strong, the AR-coKRG^{OBJ} approach was chosen as a simpler and faster implementation of the proposed multi-fidelity optimization method. It is noted that if the implementation cost of a multi-fidelity method is higher than a few runs of the VDST model, then it might worth spending this extra time to obtain more HF samples and rely on a conventional SBO approach.

Initially, 200 runs with the LF sharp model were conducted on a set of training points generated by a LHS design. Then, a coKRG model was constructed based on the output values of the penalized objective functions from the VDST and the sharp interface models. Similar to the hypothetical examples presented in chapter 5, a computational budget of $MHFr = 21$ was set for the AR-coKRG^{OBJ} algorithm, as a stopping criterion. Only one run of the AR-coKRG^{OBJ} algorithm was performed. To compare the effectiveness of AR-coKRG^{OBJ}, the optimization problem was also solved using the sharp interface model alone. The sharp interface simulations were based on the 2-D mesh of the VDST model and required 2.7 seconds for a single run.

Table 6-2 Comparison of optimal solutions found with the AR-coKRG^{OBJ} method, the sharp interface model and the VDST model for $MHFr = 21$. The optimal solutions with the VDST model and the sharp interface model are underlined).

Optimization method	Optimal solution (m^3/day)	VDST runs	Sharp runs	Time (hr)
EAS-VDST	<u>6184.8</u>	5189	NA*	116.28
EAS-SH	<u>5880.1</u>	NA*	3800	2.81
AR-coKRG^{OBJ}	**5993.1**	21	200	1.05

* NA: Not Applicable

As expected, the optimization with the sharp interface model provided a lower optimum than the VDST model, due to an overestimation of the seawater intrusion. The difference here is less than $300\, m^3/day$ which from a practical perspective represents a safer extraction rate, given the uncertainties related to the coastal aquifer model predictions. However, in terms of algorithmic performance, AR-coKRG^{OBJ} provided an even better outcome using only 21 VDST model evaluations and 200 runs with the sharp interface model. It is mentioned herein that the AR-coKRG^{OBJ} may depend on the initial training design and another run of the algorithm might not necessarily provide a better optimal solution than EAS-SH. However, and based on the present results, the proposed AR-coKRG optimization method appears as an efficient and effective approach, particularly when the discrepancies between the VDST and the sharp interface model are not large. As also mentioned previously, the fact that the present optimization formulation has a smoother objective function landscape facilitates the search with the multi-fidelity approach leading to a good local solution within a few HF model runs.

Chapter 7

Conclusions and suggestions for further research

A summary of the present work as well as the main conclusions and contributions are discussed next. Also, a section regarding suggestions for future research is included.

7.1 Summary

Pumping optimization based on the combination of variable density flow and solute transport (VDST) numerical models with evolutionary algorithms, is considered an impractical task due to the resulting computational burden. In this thesis, the development of surrogate-based optimization (SBO) methods was proposed to alleviate the computational cost. Various types of surrogate models, sampling strategies and adaptive optimization frameworks were developed and extensively compared for the practical applications of coastal aquifer management.

SBO methods were tested either by using a single surrogate model of the penalized function or by constructing individual surrogate models associated with each nonlinear constraint function. A total of nine implementations of SBO algorithms were employed. Radial basis functions and Kriging models were used to emulate the response of the nonlinear constraint functions to pumping rates. They were chosen as appropriate surrogate model types for the current deterministic numerical simulations due to their interpolating capabilities. Six out of these nine methods were developed based on the surrogate-assisted evolutionary framework and the prediction-based exploitation infill strategy. An evolutionary algorithm, namely EAS, was used as the optimization platform where this online surrogate-assisted approach, denoted as EAS-PB, was developed. For this framework, radial basis functions (RBF), Kriging models (KRG) or a combination of those via a multiple surrogate approach were utilized.

The other three SBO algorithms were based on infill strategies that balance exploration and exploitation using the metamodels. One of them was the robust and well-documented ConstrLMSRBF algorithm (Regis 2011). It was selected as a comprehensive method to compare against those developed in this thesis, due to its promising performance for pumping optimization problems of coastal aquifers with large dimensionalities and under limited computational budgets (Christelis et al. 2018). A constrained version of the SEEAS algorithm (denoted as CSEEAS) was also presented. CSEEAS was based on a short modification of the original SEEAS algorithm to pass the information from the separate surrogate models of the nonlinear constraint functions within the operations of SEEAS. Finally, a new adaptive-recursive surrogate-assisted framework, denoted as LR-RSRBF, was developed combining features of exploitation and exploration steps from three existing algorithms, namely, GOSAC, ConstrLMSRBF and SEEAS.

Also, a new adaptive-recursive optimization scheme was developed using co-Kriging models for hypothetical cases where the VDST simulation is extremely time-consuming. A few high-fidelity data from the VDST model were combined with a significantly larger number of lower-fidelity data with the sharp interface model of Strack (1976), to investigate the effectiveness of multi-fidelity optimization in coastal aquifer management.

7.2 Conclusions

The performance of all SBO algorithms was compared for multiple independent optimization trials to attribute a statistically meaningful assessment. A first thing to note is that the proposed surrogate-optimization algorithms reduced the overall computational cost by 90-98% of the direct optimization with the VDST model. Furthermore, the comparison confirmed that individual surrogate models of the constraint functions outperform the choice of a single surrogate model for the penalized objective function. However, it should be mentioned that MLMSRBF (Regis and Shoemaker 2007a) and SEEAS (Tsoukalas et al. 2016b), designed originally for bound constrained optimization problems, can potentially find good solutions for the present optimization problem. The analysis of their average converge progress showed that for a moderate dimensionality of 10 pumping wells, there is a continuous improvement of the objective function value, provided that the available computational budget is large enough.

Another finding is that as the dimensionality increases, the performance of the EAS-PB type algorithms is adversely affected. For a moderate dimensionality of 10 pumping wells, the average performance of EAS-PB algorithms was competent, delivered near-optimal solutions close to the benchmark global optimum obtained from the VDST model. Thus, despite the greedy approach of adding infill points only at the current optimum, EAS-PB algorithms are considered an efficient and effective choice for pumping optimization problems with similar dimensionalities. On the other hand, for the case of 20 pumping wells, all EAS-PB algorithms struggled to approach the region of the global optimum and mainly returned local solutions of moderate quality.

In general, the inclusion of more sophisticated surrogate models in the EAS-PB method, either KRG models or via the multiple surrogate approach, did not show any advantage in a consistent manner over the simpler RBF models. Although it is difficult to generalize, due to the specifications of each optimization problem, the use of cubic RBF models provided fair to notably good solutions especially when they were included within the operations of the other type of algorithms, namely, ConstrLMSRBF, SEEAS and LR-RSRBF. It is believed that cubic RBF models are a fast and fairly accurate surrogate model type that fits conveniently within the computational requirements of pumping optimization in coastal aquifers. The cubic RBF model is a favorable choice for problems with large number of constraint functions since their fitting time to the input-output data is small compared to the single runtime of a VDST numerical model. However, their use was investigated for adaptive online SBO frameworks in coastal aquifer management and no recommendations can be made on their prediction skills for offline SBO methods. The offline methods typically utilize large training datasets, and in such cases other surrogate models may perform more accurately.

The performance of ConstrLMSRBF, SEEAS and LR-RSRBF algorithms was tested for two alternative computational budgets. For the case of 10 pumping wells, these algorithms were able to locate near-optimal solutions for a computational budget that included as few as 100 simulations with the VDST model. This is a notable performance in both efficiency and effectiveness suggesting a significant computational tool for pumping optimization problems with a similar dimension. When the dimensionality was increased to 20 pumping wells, they were mainly trapped to local solutions for the budget of 100 VDST simulations. However, for a larger budget of 300 VDST simulations, their performance was considerably improved and approached the region of the global optimum. That was prominently the case for LR-RSRBF which is a newly introduced optimization framework developed in the present thesis. For the cases examined here, LR-RSRBF showed great promise for the efficient solution of coastal aquifer management problems.

The application of the above surrogate-assisted optimization schemes on a real-world coastal aquifer case, successfully returned near global solutions for a problem of 11 pumping wells. The EAS-PB and LR-RSRBF algorithms, both using cubic RBF models, provided the best optimal solutions. Indicative of their successful application was that LR-RSRBF found an even better solution than the one provided with the VDST model while both SBO methods returned optimal solutions with a similar distribution of pumping rates against the benchmark solution. The computational gains were in the order of 95% reduction of the original optimization task with the VDST model.

The multi-fidelity optimization framework developed in this thesis returned satisfactory solutions with as few as 21 evaluations with the computationally expensive VDST model. This is a very promising performance which could be potentially utilized in real-word coastal aquifer models with remarkably high computational requirements. The use of multi-fidelity optimization methods is largely unexplored in coastal aquifer management and, to the best of our knowledge, this is one of the few attempts existing in the literature to combine the VDST model with low-fidelity models for pumping optimization problems. This new multi-fidelity optimization method outperformed conventional SBO algorithms which are designed for a single fidelity model in the case of limited computational budgets.

7.3 Contributions

In short and to the best of the author's knowledge, the contributions of this thesis in coastal aquifer management research can be summarized as follows:

- The use of RBF and KRG surrogate models is scarce in coastal aquifer management despite their successful application in many other engineering fields. In this thesis, they were utilized for the first time in surrogate-assisted evolutionary frameworks for pumping optimization of coastal aquifers.

- Heterogeneous ensembles of KRG and RBF models using optimal weights and online updating within the operations of an evolutionary algorithm, is a new approach presented for pumping optimization problems of coastal aquifers.

- Pumping optimization problems of coastal aquifers involve nonlinear constraints. The standard choice is to develop individual surrogate models for each constraint function. In this thesis, a formal comparison was presented for the first time of this standard approach against comprehensive SBO methods where a single surrogate model is developed only for the penalized objective function.

- Exhaustive comparisons were performed for SBO methods which follow different sampling strategies, and this is a first presentation of such outcomes in coastal aquifer management.
- A constrained version of the SEEAS algorithm (Tsoukalas et al. 2016) was implemented for the requirements of the pumping optimization problems of coastal aquifers.
- A new adaptive-recursive SBO framework, denoted as LR-RSRBF, was developed to balance exploration and exploitation using surrogate models for coastal aquifer management. The new SBO method combines features from three existing algorithms, namely, GOSAC (Müller and Woodbury 2017), ConstrLMSRBF (Regis, 2011) and SEEAS (Tsoukalas et al. 2016).
- A new adaptive-recursive optimization scheme was developed using co-Kriging models for multi-fidelity optimization in coastal aquifer management. This is the first time such an approach is developed for the purposes of pumping optimization of coastal aquifers.

7.4 Some thoughts for future applications and further research

It is indisputable that for the present optimization problem, SBO methods offered a computationally affordable route to locate near-optimal solutions depending on the method and the available computational budget. As with all algorithms, tuning some parameters that are expected to affect their performance is a practical thing to do although a second or even more runs with the SBO methods is not always affordable.

Given the inherent random processes usually encountered in SBO frameworks, a series of independent optimization trials is generally recommended. So far, only few works present such an analysis in the relevant literature of coastal aquifer management for SBO methods. To our understanding, if the task is affordable, research on SBO should include multiple independent runs to avoid dependencies on initial training designs and to recognize the challenges that these optimization problems might present for SBO.

The above optimization problems were set on the assumption that seasonal recharge variations are not considered, and that constant pumping is assumed during the pumping management period. This is justified in terms of long-term management plans and based on the assumption that coastal aquifer response to normal fluctuations of recharge is relatively slow. However, it remains a challenge to implement SBO methods considering time-variable flow conditions which represent a more detailed management option of coastal aquifers. This might be particularly necessary given the anticipated impacts of climate change on coastal aquifer systems, particularly during dry months.

Regarding the implementation of SBO methods in coastal aquifer management, it is of practical interest to test more algorithms that balance exploitation with exploration. These optimization frameworks are expected to perform more reliably than offline or pure exploitation methods. While there is an increasing interest on these methods, their application for the nonlinear constrained pumping optimization problems of coastal aquifers is still limited, compared to other engineering fields.

Finally, multi-fidelity optimization methods deserve more attention for the efficient solution of real-world coastal aquifer management problems. There are already various low-fidelity models proposed in the literature but their inclusion in multi-fidelity optimization frameworks is mainly unexplored. Although, the preparation of a multi-fidelity optimization framework requires additional effort, at the same time some aspects might be beneficial. For example, in the case of transient simulations or in the presence of spatial heterogeneity of aquifer parameters where the use of conventional surrogate models might be inefficient or complicated. Two or more models of different fidelity might return output variables with different physical meaning. In addition, the HF model might need to run on a fine discretized grid compared to the grid requirements of the LF model. Such dependencies may complicate the implementation and increase the analyst time on developing a SBO method based on models of different fidelity. However, all these specifications remain to be confronted with the expectation that multi-fidelity modelling could be proved beneficial, especially for real-world optimization coastal aquifer management problems.

REFERENCES

Abarca E, Carrera J, Sánchez-Vila X, Dentz M (2007) Anisotropic dispersive Henry problem. Adv Water Resour 30:913–926.

Abd-Elhamid HF, Javadi AA (2011) A Cost-Effective Method to Control Seawater Intrusion in Coastal Aquifers. Water Resour Manag 25:2755–2780.

Acar E (2010) Various approaches for constructing an ensemble of metamodels using local measures. Struct Multidiscip Optim 42:879–896.

Al-Bitar A, Ababou R (2005) Random field approach to seawater intrusion in heterogeneous coastal aquifers: unconditional simulations and statistical analysis. In: Geostatistics for Environmental Applications. Springer, Berlin, Heidelberg, pp 233–248

Alexandrov NM, Lewis RM, Gumbert CR, et al (2001) Approximation and Model Management in Aerodynamic Optimization with Variable-Fidelity Models. J Aircr 38:1093–1101.

Asher MJ, Croke BFW, Jakeman AJ, Peeters LJM (2015) A review of surrogate models and their application to groundwater modeling. Water Resour Res 51:5957–5973.

Ataie-Ashtiani B, Ketabchi H (2011) Elitist Continuous Ant Colony Optimization Algorithm for Optimal Management of Coastal Aquifers. Water Resour Manag 25:165–190.

Ataie-Ashtiani Behzad, Ketabchi Hamed, Rajabi Mohammad Mahdi (2014) Optimal Management of a Freshwater Lens in a Small Island Using Surrogate Models and Evolutionary Algorithms. J Hydrol Eng 19:339–354.

Babaei M, Pan I (2016) Performance comparison of several response surface surrogate models and ensemble methods for water injection optimization under uncertainty. Comput Geosci 91:19–32.

Bakker M (2003) A Dupuit formulation for modeling seawater intrusion in regional aquifer systems. Water Resour Res 39:. https://doi.org/10.1029/2002WR001710

Bakker M (2006) Analytic solutions for interface flow in combined confined and semi-confined, coastal aquifers. Adv Water Resour 29:417–425.

Bandler JW, Biernacki RM, Shao Hua Chen, et al (1994) Space mapping technique for electromagnetic optimization. IEEE Trans Microw Theory Tech 42:2536–2544.

Bandler JW, Cheng QS, Nikolova NK, Ismail MA (2004) Implicit space mapping optimization exploiting preassigned parameters. IEEE Trans Microw Theory Tech 52:378–385.

Basudhar A, Dribusch C, Lacaze S, Missoum S (2012) Constrained efficient global optimization with support vector machines. Struct Multidiscip Optim 46:201–221.

Baú DA, Mayer AS (2006) Stochastic management of pump-and-treat strategies using surrogate functions. Adv Water Resour 29:1901–1917.

Bear J, Cheng AH-D, Sorek S, et al (eds) (1999) Seawater Intrusion in Coastal Aquifers — Concepts, Methods and Practices. Springer Netherlands, Dordrecht

Behrangi A, Khakbaz B, Vrugt JA, et al (2008) Comment on "Dynamically dimensioned search algorithm for computationally efficient watershed model calibration" by Bryan A. Tolson and Christine A. Shoemaker: Water Resour Res 44.

Bhattacharjya RK, Datta B (2005) Optimal Management of Coastal Aquifers Using Linked Simulation Optimization Approach. Water Resour Manag 19:295–320.

Bhosekar A, Ierapetritou M (2018) Advances in surrogate based modeling, feasibility analysis, and optimization: A review. Comput Chem Eng 108:250–267.

Bobba AG (1993) Mathematical models for saltwater intrusion in coastal aquifers. Water Resour Manag 7:3–37.

Bouhlel MA, Bartoli N, Regis RG, et al (2018) Efficient global optimization for high-dimensional constrained problems by using the Kriging models combined with the partial least squares method. Eng Optim 50:2038–2053.

Boukouvala F, Ierapetritou MG (2014) Derivative-free optimization for expensive constrained problems using a novel expected improvement objective function. AIChE J 60:2462–2474.

Cervantes-González JC, Rayas-Sánchez JE, López CA, et al (2016) Space mapping optimization of handset antennas considering EM effects of mobile phone components and human body. Int J RF Microw Comput-Aided Eng 26:121–128.

Chang C-M, Yeh H-D (2010) Spectral approach to seawater intrusion in heterogeneous coastal aquifers. Hydrol Earth Syst Sci 14:719–727.

Chang SW, Clement TP, Simpson MJ, Lee K-K (2011) Does sea-level rise have an impact on saltwater intrusion? Adv Water Resour 34:1283–1291.

Cheng AH-D, Halhal D, Naji A, Ouazar D (2000) Pumping optimization in saltwater-intruded coastal aquifers. Water Resour Res 36:2155–2165.

Cheng J, Jiang P, Zhou Q, et al (2019) A lower confidence bounding approach based on the coefficient of variation for expensive global design optimization. Eng Comput 36:830–849.

Christelis V (2017) Physics-based and data-driven surrogate models for pumping optimization of coastal aquifers. European Water 57:481-488

Christelis V, Bellos V, Tsakiris G (2016) Employing surrogate modelling for the calibration of a 2D flood simulation model. In Sustainable Hydraulics in the Era of Global Change, Proceedings of 4th IAHR Congress, Liege, Belgium, 27–29 July 2016; Erpicum, S., Dewals, B., Archambeau, P., Pirotton, M., Eds.; CRC Press: Boca Raton, FL, USA, 2016; pp. 727–732.

Christelis V, Kopsiaftis G, Mantoglou A (2019b) Performance comparison of multiple and single surrogate models for pumping optimization of coastal aquifers. Hydrol Sci J 64:336–349.

Christelis V, Mantoglou A (2016a) Coastal Aquifer Management Based on the Joint use of Density-Dependent and Sharp Interface Models. Water Resour Manag 30:861–876.

Christelis V, Mantoglou A (2019a) Pumping Optimization of Coastal Aquifers Using Seawater Intrusion Models of Variable-Fidelity and Evolutionary Algorithms. Water Resour Manag 33:555–568.

Christelis V, Mantoglou A (2013) Improved Sharp Interface Models in Coastal Aquifers of Finite Dimensions. 15: EGU 2013-12228

Christelis V, Mantoglou A (2016b) Pumping Optimization of Coastal Aquifers Assisted by Adaptive Metamodelling Methods and Radial Basis Functions. Water Resour Manag 30:5845–5859.

Christelis V, Regis RG, Mantoglou A (2018) Surrogate-based pumping optimization of coastal aquifers under limited computational budgets. J Hydroinformatics 20:164–176.

Couckuyt I, Declercq F, Dhaene T, et al (2010) Surrogate-based infill optimization applied to electromagnetic problems. Int J RF Microw Comput-Aided Eng 20:492–501. https://doi.org/10.1002/mmce.20455

Couckuyt I, Deschrijver D, Dhaene T (2014) Fast calculation of multiobjective probability of improvement and expected improvement criteria for Pareto optimization. J Glob Optim 60:575–594.

Dagan G, Zeitoun DG (1998) Seawater-freshwater interface in a stratified aquifer of random permeability distribution. J Contam Hydrol 29:185–203.
Das A, Datta B (1999) Development of Management Models for Sustainable Use of Coastal Aquifers. J Irrig Drain Eng 125:112–121.
Datta R, Regis RG (2016) A surrogate-assisted evolution strategy for constrained multi-objective optimization. Expert Syst Appl 57:270–284.
Dausman AM, Langevin C, Bakker M, Schaars F (2010) A comparison between SWI and SEAWAT – the importance of dispersion, inversion and vertical anisotropy. In: de Melo MTC, Lebbe L, Cruz JV, Coutinho R, Langevin C, Buxo A, editors. Proceedings of SWIM 21 – 21st salt water intrusion meeting, Azores, Portugal; 2010. p. 271–4.
Dey S, Prakash O (2020) Managing saltwater intrusion using conjugate sharp interface and density dependent models linked with pumping optimization. Groundw Sustain Dev 11:100446.
Dhar A, Datta B (2009) Saltwater Intrusion Management of Coastal Aquifers. I: Linked Simulation-Optimization. J Hydrol Eng 14:1263–1272.
Diersch H-JG, Kolditz O (2002) Variable-density flow and transport in porous media: approaches and challenges. Adv Water Resour 25:899–944.
Dokou Z, Karatzas GP (2012) Saltwater intrusion estimation in a karstified coastal system using density-dependent modelling and comparison with the sharp-interface approach. Hydrol Sci J 57:985–999.
Dong H, Song B, Dong Z, Wang P (2018) SCGOSR: Surrogate-based constrained global optimization using space reduction. Appl Soft Comput 65:462–477.
Doulgeris C, Zissis T (2014) 3D Variable Density Flow Simulation to Evaluate Pumping Schemes in Coastal Aquifers. Water Resour Manag 28:4943–4956.
Efstratiadis A, Koutsoyiannis D An evolutionary annealing-simplex algorithm for global optimisation of water resource systems. Proc., 5th Int. Conf. on Hydroinformatics (Hydroinformatics 2002), IWA Publishing, Colchester, U.K.
Efstratiadis A, Nalbantis I, Koutsoyiannis D (2015) Hydrological modelling of temporally-varying catchments: facets of change and the value of information. Hydrol Sci J 60:1438–1461.
Essaid HI (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. J Hydrol 86:169–193.
Feng F, Zhang J, Zhang W, et al (2019) Coarse- and Fine-Mesh Space Mapping for EM Optimization Incorporating Mesh Deformation. IEEE Microw Wirel Compon Lett 29:510–512.
Ferguson G, Gleeson T (2012) Vulnerability of coastal aquifers to groundwater use and climate change. Nat Clim Change 2:342–345.
Ferreira da Silva JF, Haie N (2007) Optimal Locations of Groundwater Extractions in Coastal Aquifers. Water Resour Manag 21:1299–1311.
Forrester A, Jones D (2008) Global Optimization of Deceptive Functions with Sparse Sampling. In: 12th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference. American Institute of Aeronautics and Astronautics, Victoria, British Columbia, Canada
Forrester AIJ, Keane AJ (2009) Recent advances in surrogate-based optimization. Prog Aerosp Sci 45:50–79.
Forrester AIJ, Sóbester A, Keane AJ (2007) Multi-fidelity optimization via surrogate modelling. Proc R Soc Math Phys Eng Sci 463:3251–3269.

Francois-Bongarcon, D., 1981, Les coregionalisations, le cokrigeage;Centre de Géostatistique, Fontainebleau, C-86.
Frind EO (1982) Simulation of long-term transient density-dependent transport in groundwater. Adv Water Resour 5:73–88.
Gambolati G, Putti M, Paniconi C (1999) Three-Dimensional Model of Coupled Density-Dependent Flow and Miscible Salt Transport. In: Seawater Intrusion in Coastal Aquifers — Concepts, Methods and Practices. Springer, Dordrecht, pp 315–362
Gano S, Sanders B, Renaud J (2004) Variable Fidelity Optimization Using a Kriging Based Scaling Function. In: 10th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference. American Institute of Aeronautics and Astronautics, Albany, New York
Georgopoulou E, Kotronarou A, Koussis A, et al (2001) A methodology to investigate brackish groundwater desalination coupled with aquifer recharge by treated wastewater as an alternative strategy for water supply in Mediterranean areas. Desalination 136:307–315.
Giambastiani BMS, Colombani N, Greggio N, et al (2017) Coastal aquifer response to extreme storm events in Emilia-Romagna, Italy. Hydrol Process 31:1613–1621.
Gingerich SB, Voss CI (2005) Three-dimensional variable-density flow simulation of a coastal aquifer in southern Oahu, Hawaii, USA. Hydrogeol J 13:436–450.
Fernández-Godino GM, Park C, Kim NH, Haftka RT (2019) Issues in Deciding Whether to Use Multifidelity Surrogates. AIAA J 57:2039–2054.
Grundmann J, Schütze N, Schmitz GH, Al-Shaqsi S (2012) Towards an integrated arid zone water management using simulation-based optimisation. Environ Earth Sci 65:1381–1394.
Guo W, Langevin CD (2002) User's guide to SEAWAT; a computer program for simulation of three-dimensional variable-density ground-water flow
Han Z-H, Görtz S, Zimmermann R (2013) Improving variable-fidelity surrogate modeling via gradient-enhanced kriging and a generalized hybrid bridge function. Aerosp Sci Technol 25:177–189.
Hardy RL (1971) Multiquadric equations of topography and other irregular surfaces. J Geophys Res 1896-1977 76:1905–1915.
Hou Z, Lu W, Xue H, Lin J (2017) A comparative research of different ensemble surrogate models based on set pair analysis for the DNAPL-contaminated aquifer remediation strategy optimization. J Contam Hydrol 203:28–37.
Huang D, Allen TT, Notz WI, Zeng N (2006) Global Optimization of Stochastic Black-Box Systems via Sequential Kriging Meta-Models. J Glob Optim 34:441–466.
Hussain MS, Javadi AA, Ahangar-Asr A, Farmani R (2015) A surrogate model for simulation–optimization of aquifer systems subjected to seawater intrusion. J Hydrol 523:542–554.
Javadi A, Hussain M, Sherif M, Farmani R (2015) Multi-objective Optimization of Different Management Scenarios to Control Seawater Intrusion in Coastal Aquifers. Water Resour Manag 29:1843–1857.
Jiang X, Lu W, Hou Z, et al (2015) Ensemble of surrogates-based optimization for identifying an optimal surfactant-enhanced aquifer remediation strategy at heterogeneous DNAPL-contaminated sites. Comput Geosci 84:37–45.
Jiao R, Zeng S, Li C, et al (2019) A complete expected improvement criterion for Gaussian process assisted highly constrained expensive optimization. Inf Sci 471:80–96.
Jin R, Chen W, Simpson TW (2001) Comparative studies of metamodelling techniques under multiple modelling criteria. Struct Multidiscip Optim 23:1–13.

Jin Y (2011) Surrogate-assisted evolutionary computation: Recent advances and future challenges. Swarm Evol Comput 1:61–70.

Jones DR, Schonlau M, Welch WJ (1998) Efficient Global Optimization of Expensive Black-Box Functions. J Glob Optim 13:455–492.

Journel, A. and Huijbrechts, Ch. J., 1978, Mining Geostatistics: Academic Press.

Kacimov AR, Sherif MM, Perret JS, Al-Mushikhi A (2009) Control of sea-water intrusion by salt-water pumping: Coast of Oman. Hydrogeol J 17:541–558.

Karakasis MK, Giannakoglou KC (2006) On the use of metamodel-assisted, multi-objective evolutionary algorithms. Eng Optim 38:941–957.

Karakasis MK, Koubogiannis DG, Giannakoglou KC (2007) Hierarchical distributed metamodel-assisted evolutionary algorithms in shape optimization. Int J Numer Methods Fluids 53:455–469.

Karatzas GP, Dokou Z (2015) Optimal management of saltwater intrusion in the coastal aquifer of Malia, Crete (Greece), using particle swarm optimization. Hydrogeol J 23:1181–1194.

Karpouzos DK, Katsifarakis KL (2013) A Set of New Benchmark Optimization Problems for Water Resources Management. Water Resour Manag 27:3333–3348.

Karterakis SM, Karatzas GP, Nikolos IK, Papadopoulou MP (2007) Application of linear programming and differential evolutionary optimization methodologies for the solution of coastal subsurface water management problems subject to environmental criteria. J Hydrol 342:270–282.

Katsifarakis KL, Petala Z (2006) Combining genetic algorithms and boundary elements to optimize coastal aquifers' management. J Hydrol 327:200–207.

Kerrou J, Renard P (2010) A numerical analysis of dimensionality and heterogeneity effects on advective dispersive seawater intrusion processes. Hydrogeol J 18:55–72.

Ketabchi H, Ataie-Ashtiani B (2015) Evolutionary algorithms for the optimal management of coastal groundwater: A comparative study toward future challenges. J Hydrol 520:193–213.

Kleijnen JPC (2009) Kriging metamodeling in simulation: A review. Eur J Oper Res 192:707–716.

Kleijnen JPC, van Beers W, van Nieuwenhuyse I (2012) Expected improvement in efficient global optimization through bootstrapped kriging. J Glob Optim 54:59–73.

Knowles J (2006) ParEGO: a hybrid algorithm with on-line landscape approximation for expensive multiobjective optimization problems. IEEE Trans Evol Comput 10:50–66.

Kolditz O, Ratke R, Diersch H-JG, Zielke W (1998) Coupled groundwater flow and transport: 1. Verification of variable density flow and transport models. Adv Water Resour 21:27–46.

Kopsiaftis G, Christelis V, Mantoglou A (2019a) Comparison of Sharp Interface to Variable Density Models in Pumping Optimisation of Coastal Aquifers. Water Resour Manag 33:1397–1409.

Kopsiaftis G, Mantoglou A, Giannoulopoulos P (2009) Variable density coastal aquifer models with application to an aquifer on Thira Island. Desalination 237:65–80.

Kopsiaftis G, Protopapadakis E, Voulodimos A, et al (2019b) Gaussian Process Regression Tuned by Bayesian Optimization for Seawater Intrusion Prediction. In: Comput. Intell. Neurosci. https://www.hindawi.com/journals/cin/2019/2859429/.

Kopsiaftis G, Tigkas D, Christelis V, Vangelis H (2017) Assessment of drought impacts on semi-arid coastal aquifers of the Mediterranean. J Arid Environ 137:7–15.

Kourakos G, Mantoglou A (2006) pumping optimization of coastal aquifers using 3-d density models and approximations with neural networks. In: XVI International Conference on Computational Methods in Water Resources, Copenhagen.

Kourakos G, Mantoglou A (2015) An efficient simulation-optimization coupling for management of coastal aquifers. Hydrogeol J 23:1167–1179. 7

Kourakos G, Mantoglou A (2009) Pumping optimization of coastal aquifers based on evolutionary algorithms and surrogate modular neural network models. Adv Water Resour 32:507–521.

Kourakos G, Mantoglou A (2013) Development of a multi-objective optimization algorithm using surrogate models for coastal aquifer management. J Hydrol 479:13–23.

Koussis AD, Mazi K (2018) Corrected interface flow model for seawater intrusion in confined aquifers: relations to the dimensionless parameters of variable-density flow. Hydrogeol J 26:2547–2559.

Koussis AD, Mazi K, Destouni G (2012) Analytical single-potential, sharp-interface solutions for regional seawater intrusion in sloping unconfined coastal aquifers, with pumping and recharge. J Hydrol 416–417:1–11.

Koussis AD, Mazi K, Riou F, Destouni G (2015) A correction for Dupuit–Forchheimer interface flow models of seawater intrusion in unconfined coastal aquifers. J Hydrol 525:277–285.

Koziel S, Bandler JW, Madsen K (2008) Quality assessment of coarse models and surrogates for space mapping optimization. Optim Eng 9:375–391.

Koziel S, Bandler JW, Madsen K (2009) Space Mapping With Adaptive Response Correction for Microwave Design Optimization. IEEE Trans Microw Theory Tech 57:478–486.

Koziel S, Ciaurri DE, Leifsson L (2011) Surrogate-Based Methods. In: Koziel S, Yang X-S (eds) Computational Optimization, Methods and Algorithms. Springer Berlin Heidelberg, Berlin, Heidelberg, pp 33–59

Lal A, Datta B (2018) Development and Implementation of Support Vector Machine Regression Surrogate Models for Predicting Groundwater Pumping-Induced Saltwater Intrusion into Coastal Aquifers. Water Resour Manag 32:2405–2419.

Lal A, Datta B (2020) Performance Evaluation of Homogeneous and Heterogeneous Ensemble Models for Groundwater Salinity Predictions: a Regional-Scale Comparison Study. Water Air Soil Pollut 231:320.

Leary SJ, Bhaskar A, Keane AJ (2004) A Derivative Based Surrogate Model for Approximating and Optimizing the Output of an Expensive Computer Simulation. J Glob Optim 30:39–58.

Leifsson L, Koziel S (2015) Aerodynamic shape optimization by variable-fidelity computational fluid dynamics models: A review of recent progress. J Comput Sci 10:45–54.

Li C, Fang H, Gong C (2019) Expensive Inequality Constraints Handling Methods Suitable for Dynamic Surrogate-based Optimization. In: 2019 IEEE Congress on Evolutionary Computation (CEC). pp 2010–2017

Li Y, Wu Y, Zhao J, Chen L (2017) A Kriging-based constrained global optimization algorithm for expensive black-box functions with infeasible initial points. J Glob Optim 67:343–366.

Liu B, Koziel S, Zhang Q (2016) A multi-fidelity surrogate-model-assisted evolutionary algorithm for computationally expensive optimization problems. J Comput Sci 12:28–37.

Llopis-Albert C, Pulido-Velazquez D (2014) Discussion about the validity of sharp-interface models to deal with seawater intrusion in coastal aquifers Hydrol Process 28:3642–3654.

Lu C, Werner AD (2013) Timescales of seawater intrusion and retreat. Adv Water Resour 59:39–51.

Lu C, Werner AD, Simmons CT, Luo J (2015) A Correction on Coastal Heads for Groundwater Flow Models. Groundwater 53:164–170.
Lu C, Xin P, Kong J, et al (2016) Analytical solutions of seawater intrusion in sloping confined and unconfined coastal aquifers. Water Resour Res 52:6989–7004.
Mahesha A, Nagaraja SH (1996) Effect of natural recharge on sea water intrusion in coastal aquifers. J Hydrol 174:211–220.
Mantoglou A (2003) Pumping management of coastal aquifers using analytical models of saltwater intrusion. Water Resour Res 39:
Mantoglou A, Papantoniou M (2008) Optimal design of pumping networks in coastal aquifers using sharp interface models. J Hydrol 361:52–63.
Mantoglou A, Papantoniou M, Giannoulopoulos P (2004) Management of coastal aquifers based on nonlinear optimization and evolutionary algorithms. J Hydrol 297:209–228.
Marduel X, Tribes C, Trépanier J-Y (2006) Variable-fidelity optimization: Efficiency and robustness. Optim Eng 7:479–500.
Matheron, G., (1973) The intrinsic random functions and their applications: Adv. Appl. Prob. 5:437–468.
Matheron G., 1979 Recherche de simplification dans un probleme de cokrigeage: Centre de Géostatistique, Fountainebleau, N-698.
Mazi K, Koussis AD, Destouni G (2014) Intensively exploited Mediterranean aquifers: resilience to seawater intrusion and proximity to critical thresholds. Hydrol Earth Syst Sci 18:1663–1677.
Michael HA, Mulligan AE, Harvey CF (2005) Seasonal oscillations in water exchange between aquifers and the coastal ocean. Nature 436:1145–1148.
Mugunthan P, Shoemaker CA, Regis RG (2005) Comparison of function approximation, heuristic, and derivative-based methods for automatic calibration of computationally expensive groundwater bioremediation models. Water Resour Res 41:
Müller J, Piché R (2011) Mixture surrogate models based on Dempster-Shafer theory for global optimization problems. J Glob Optim 51:79–104.
Müller J, Shoemaker CA (2014) Influence of ensemble surrogate models and sampling strategy on the solution quality of algorithms for computationally expensive black-box global optimization problems. J Glob Optim 60:123–144.
Müller J, Woodbury JD (2017) GOSAC: global optimization with surrogate approximation of constraints. J Glob Optim 69:117–136.
Myers, D. E., 1981, Joint estimation of random functions: the matrix form: Research report, Department of Mathematics, University of Arizona.
Nikolos IK (2013) On the use of multiple surrogates within a differential evolution procedure for high-lift airfoil design. Int J Adv Intell Paradig 5:319.
Nuñez L, Regis RG, Varela K (2018) Accelerated Random Search for constrained global optimization assisted by Radial Basis Function surrogates. J Comput Appl Math 340:276–295.
Oude Essink GHP (2001) Improving fresh groundwater supply—problems and solutions. Ocean Coast Manag 44:429–449.
Palar PS, Shimoyama K (2018) On efficient global optimization via universal Kriging surrogate models. Struct Multidiscip Optim 57:2377–2397.
Paniconi C, Khlaifi I, Lecca G, et al (2001) Modeling and Analysis of Seawater Intrusion in the Coastal Aquifer of Eastern Cap-Bon, Tunisia. Transp Porous Media 43:3–28.

Papadopoulou MP, Nikolos IK, Karatzas GP (2010) Computational benefits using artificial intelligent methodologies for the solution of an environmental design problem: saltwater intrusion. Water Sci Technol 62:1479–1490.

Parr JM, Keane AJ, Forrester AIJ, Holden CME (2012) Infill sampling criteria for surrogate-based optimization with constraint handling. Eng Optim 44:1147–1166.

Pool M, Carrera J (2010) Dynamics of negative hydraulic barriers to prevent seawater intrusion. Hydrogeol J 18:95–105.

Pool M, Carrera J (2011) A correction factor to account for mixing in Ghyben-Herzberg and critical pumping rate approximations of seawater intrusion in coastal aquifers. Water Resour Res 47:

Prieto C, Kotronarou A, Destouni G (2006) The influence of temporal hydrological randomness on seawater intrusion in coastal aquifers. J Hydrol 330:285–300.

Qahman K, Larabi A, Ouazar D, et al (2005) Optimal and sustainable extraction of groundwater in coastal aquifers. Stoch Environ Res Risk Assess 19:99–110.

Queipo NV, Haftka RT, Shyy W, et al (2005) Surrogate-based analysis and optimization. Prog Aerosp Sci 41:1–28.

Rajabi MM, Ataie-Ashtiani B, Simmons CT (2015) Polynomial chaos expansions for uncertainty propagation and moment independent sensitivity analysis of seawater intrusion simulations. J Hydrol 520:101–122.

Rajabi MM, Ketabchi H (2017) Uncertainty-based simulation-optimization using Gaussian process emulation: Application to coastal groundwater management. J Hydrol 555:518–534.

Rao SVN, Bhallamudi SM, Thandaveswara BS, Mishra GC (2004) Conjunctive Use of Surface and Groundwater for Coastal and Deltaic Systems. J Water Resour Plan Manag 130:255–267.

Razavi S, Tolson BA, Burn DH (2012a) Numerical assessment of metamodelling strategies in computationally intensive optimization. Environ Model Softw 34:67–86.

Razavi S, Tolson BA, Burn DH (2012b) Review of surrogate modeling in water resources: REVIEW. Water Resour Res 48:.

Razavi S, Tolson BA, Matott LS, et al (2010) Reducing the computational cost of automatic calibration through model preemption. Water Resour Res 46:.

Regis RG (2011) Stochastic radial basis function algorithms for large-scale optimization involving expensive black-box objective and constraint functions. Comput Oper Res 38:837–853.

Regis RG (2014) Constrained optimization by radial basis function interpolation for high-dimensional expensive black-box problems with infeasible initial points. Eng Optim 46:218–243.

Regis RG, Shoemaker CA (2007a) A Stochastic Radial Basis Function Method for the Global Optimization of Expensive Functions. Inf J Comput 19:497–509.

Regis RG, Shoemaker CA (2013) Combining radial basis function surrogates and dynamic coordinate search in high-dimensional expensive black-box optimization. Eng Optim 45:529–555.

Regis RG, Shoemaker CA (2007b) Improved Strategies for Radial basis Function Methods for Global Optimization. J Glob Optim 37:113–135.

Regis RG, Wild SM (2017) CONORBIT: constrained optimization by radial basis function interpolation in trust regions. Optim Methods Softw 32:552–580.

Robinson T, Willcox K, Eldred M, Haimes R (2006) Multifidelity Optimization for Variable-Complexity Design. In: 11th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference. American Institute of Aeronautics and Astronautics, Portsmouth, Virginia

Roy DK, Datta B (2020) Saltwater intrusion prediction in coastal aquifers utilizing a weighted-average heterogeneous ensemble of prediction models based on Dempster-Shafer theory of evidence. Hydrol Sci J 65:1555–1567.

Roy DK, Datta B (2017a) Fuzzy C-Mean Clustering Based Inference System for Saltwater Intrusion Processes Prediction in Coastal Aquifers. Water Resour Manag 31:355–376.

Roy DK, Datta B (2017b) Multivariate Adaptive Regression Spline Ensembles for Management of Multilayered Coastal Aquifers. J Hydrol Eng 22:04017031.

Rozos E, Efstratiadis A, Nalbantis I, Koutsoyiannis D (2004) Calibration of a semi-distributed model for conjunctive simulation of surface and groundwater flows / Calage d'un modèle semi-distribué pour la simulation conjointe d'écoulements superficiels et souterrains. Hydrol Sci J 49:.

Sacks J, Welch WJ, Mitchell TJ, Wynn HP (1989) Design and Analysis of Computer Experiments. Stat Sci 4:409–423

Sanford WE, Pope JP (2010) Current challenges using models to forecast seawater intrusion: lessons from the Eastern Shore of Virginia, USA. Hydrogeol J 18:73–93.

Sasena MJ, Papalambros P, Goovaerts P (2002) Exploration of Metamodeling Sampling Criteria for Constrained Global Optimization. Eng Optim 34:263–278.

Schwartz FW, Zhang H (2003) Groundwater. New York: John Wiley & Sons Inc

Sedki A, Ouazar D (2011) Simulation-Optimization Modeling for Sustainable Groundwater Development: A Moroccan Coastal Aquifer Case Study. Water Resour Manag 25:2855–2875.

Sherif MM, Singh VP (1999) Effect of climate change on sea water intrusion in coastal aquifers. Hydrol Process 13:1277–1287.

Shi R, Liu L, Long T, Liu J (2016) An efficient ensemble of radial basis functions method based on quadratic programming. Eng Optim 48:1202–1225.

Shoemaker CA, Regis RG, Fleming RC (2007) Watershed calibration using multistart local optimization and evolutionary optimization with radial basis function approximation. Hydrol Sci J 52:450–465. https://doi.org/10.1623/hysj.52.3.450

Shu L, Jiang P, Song X, Zhou Q (2019) Novel Approach for Selecting Low-Fidelity Scale Factor in Multifidelity Metamodeling. AIAA J 57:5320–5330. https://doi.org/10.2514/1.J057989

Simmons CT (2005) Variable density groundwater flow: From current challenges to future possibilities. Hydrogeol J 13:116–119.

Simmons CT, Fenstemaker TR, Sharp JM (2001) Variable-density groundwater flow and solute transport in heterogeneous porous media: approaches, resolutions and future challenges. J Contam Hydrol 52:245–275.

Singh A (2014) Optimization modelling for seawater intrusion management. J Hydrol 508:43–52.

Solomatine DP, Ostfeld A (2008) Data-driven modelling: some past experiences and new approaches. J Hydroinformatics 10:3–22.

Song J, Yang Y, Wu J, et al (2018) Adaptive surrogate model based multiobjective optimization for coastal aquifer management. J Hydrol 561:98–111.

Sreekanth J, Datta B (2015) Review: Simulation-optimization models for the management and monitoring of coastal aquifers. Hydrogeol J 23:1155–1166.

Sreekanth J, Datta B (2010) Multi-objective management of saltwater intrusion in coastal aquifers using genetic programming and modular neural network based surrogate models. J Hydrol 393:245–256.
Sreekanth J, Datta B (2011a) Comparative Evaluation of Genetic Programming and Neural Network as Potential Surrogate Models for Coastal Aquifer Management. Water Resour Manag 25:3201–3218.
Sreekanth J, Datta B (2011b) Coupled simulation-optimization model for coastal aquifer management using genetic programming-based ensemble surrogate models and multiple-realization optimization. Water Resour Res 47:.
Strack ODL (1976) A single-potential solution for regional interface problems in coastal aquifers. Water Resour Res 12:1165–1174.
Sun G, Tian Y, Wang R, et al (2020) Parallelized multiobjective efficient global optimization algorithm and its applications. Struct Multidiscip Optim 61:763–786.
Thompson C, Smith L, Maji R Hydrogeological modeling of submarine groundwater discharge on the continental shelf of Louisiana. J Geophys Res Oceans 112:.
Tigkas D, Christelis V, Tsakiris G (2016) Comparative Study of Evolutionary Algorithms for the Automatic Calibration of the Medbasin-D Conceptual Hydrological Model. Environ Process 3:629–644.
Tsoukalas I, Kossieris P, Efstratiadis A, Makropoulos C (2016) Surrogate-enhanced evolutionary annealing simplex algorithm for effective and efficient optimization of water resources problems on a budget. Environ Model Softw 77:122–142.
Tsoukalas I, Makropoulos C (2015) Multiobjective optimisation on a budget: Exploring surrogate modelling for robust multi-reservoir rules generation under hydrological uncertainty. Environ Model Softw 69:396–413.
Tyan M, Nguyen NV, Lee J-W (2015) Improving variable-fidelity modelling by exploring global design space and radial basis function networks for aerofoil design. Eng Optim 47:885–908.
Uddameri V, Kuchanur M (2007) Simulation-optimization approach to assess groundwater availability in Refugio County, TX. Environ Geol 51:921–929.
Viana F, Haftka R, Watson L (2010a) Why Not Run the Efficient Global Optimization Algorithm with Multiple Surrogates? In: 51st AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference. 18th AIAA/ASME/AHS Adaptive Structures Conference 12th. American Institute of Aeronautics and Astronautics, Orlando, Florida
Viana FAC, Gogu C, Haftka RT (2010b) Making the Most Out of Surrogate Models: Tricks of the Trade. In: Volume 1: 36th Design Automation Conference, Parts A and B. ASMEDC, Montreal, Quebec, Canada, pp 587–598
Viana FAC, Haftka RT, Steffen V (2009) Multiple surrogates: how cross-validation errors can help us to obtain the best predictor. Struct Multidiscip Optim 39:439–457.
Viana FAC, Haftka RT, Watson LT (2013) Efficient global optimization algorithm assisted by multiple surrogate techniques. J Glob Optim 56:669–689.
Villemonteix J, Vazquez E, Sidorkiewicz M, Walter E (2009a) Global optimization of expensive-to-evaluate functions: an empirical comparison of two sampling criteria. J Glob Optim 43:373–389.
Villemonteix J, Vazquez E, Walter E (2009b) An informational approach to the global optimization of expensive-to-evaluate functions. J Glob Optim 44:509–534.

Walther M, Graf T, Kolditz O, et al (2017) How significant is the slope of the sea-side boundary for modelling seawater intrusion in coastal aquifers? J Hydrol 551:648–659.
Watson TA, Werner AD, Simmons CT (2010) Transience of seawater intrusion in response to sea level rise. Water Resour Res 46:.
Webb MD, Howard KWF (2011) Modeling the Transient Response of Saline Intrusion to Rising Sea-Levels. Ground Water 49:560–569.
Werner AD (2017) Correction factor to account for dispersion in sharp-interface models of terrestrial freshwater lenses and active seawater intrusion. Adv Water Resour 102:45–52.
Werner AD, Bakker M, Post VEA, et al (2013) Seawater intrusion processes, investigation and management: Recent advances and future challenges. Adv Water Resour 51:3–26.
Werner AD, Simmons CT (2009) Impact of Sea-Level Rise on Sea Water Intrusion in Coastal Aquifers. Ground Water 47:197–204.
Wu Y, Yin Q, Jie H, et al (2018) A RBF-based constrained global optimization algorithm for problems with computationally expensive objective and constraints. Struct Multidiscip Optim 58:1633–1655.
Yadav B, Mathur S, Ch S, Yadav BK (2018) Data-based modelling approach for variable density flow and solute transport simulation in a coastal aquifer. Hydrol Sci J 63:210–226.
Yang J, Graf T, Ptak T (2015) Impact of climate change on freshwater resources in a heterogeneous coastal aquifer of Bremerhaven, Germany: A three-dimensional modeling study. J Contam Hydrol 177–178:107–121.
Yang Z, Qiu H, Gao L, et al (2020) Surrogate-assisted classification-collaboration differential evolution for expensive constrained optimization problems. Inf Sci 508:50–63.
Yao W, Chen XQ, Huang YY, Tooren M van (2014) A surrogate-based optimization method with RBF neural network enhanced by linear interpolation and hybrid infill strategy. Optim Methods Softw 29:406–429.
Yondo R, Andrés E, Valero E (2018) A review on design of experiments and surrogate models in aircraft real-time and many-query aerodynamic analyses. Prog Aerosp Sci 96:23–61.
Zaefferer M, Gaida D, Bartz-Beielstein T (2016) Multi-fidelity modeling and optimization of biogas plants. Appl Soft Comput 48:13–28.
Zhou Q, Jiang P, Shao X, et al (2017) A variable fidelity information fusion method based on radial basis function. Adv Eng Inform 32:26–39.
Zhou Q, Shao X, Jiang P, et al (2016) An active learning metamodeling approach by sequentially exploiting difference information from variable-fidelity models. Adv Eng Inform 30:283–297.
Zhou X, Ma Y, Tu Y, Feng Y (2013) Ensemble of Surrogates for Dual Response Surface Modeling in Robust Parameter Design. Qual Reliab Eng Int 29:173–197.