Pumping optimization of coastal aquifers using seawater intrusion models of variable-fidelity and evolutionary algorithms

Vasileios Christelis[1,2] and Aristotelis Mantoglou[1]

[1] Laboratory of Reclamation Works and Water Resources Management, School of Rural and Surveying Engineering, National Technical University of Athens, Athens, Greece

[2] British Geological Survey, Environmental Science Centre, Keyworth, Nottingham, UK

Corresponding author: Vasileios Christelis (vchrist@survey.ntua.gr)

Abstract:

Variable-fidelity modelling has been utilized in several engineering optimization studies to construct surrogate models. However, similar approaches have received much less attention in coastal aquifer management problems. A variable-fidelity optimization framework was developed utilizing a lower-fidelity and computationally cheap model of seawater intrusion, based on the sharp interface assumption, and a simple correction process. The variable-fidelity method was compared to the direct optimization with the high-fidelity variable density and salt transport model and to conventional surrogate-based optimization. The surrogate-based approaches were embedded into the operations of an evolutionary algorithm to implement an efficient online update of the surrogate models and control the feasibility of the optimal solutions. Multiple independent optimization runs were performed to provide more insightful comparison outcomes. Although the variable-fidelity method found a better optimum than the conventional approach, the overall sample statistics showed that the surrogate-based optimization frameworks performed equally well and provided good approximations to the high-fidelity solution. Despite the potential for an improved exploration of the optimal search space by using the variable-fidelity method, the conventional approach had a 30% faster average convergence time.

1. Introduction

The seawater intrusion phenomenon is a common problem of coastal aquifers, particularly for semi-arid areas where low recharge and increased groundwater extraction threatens the sustainability of freshwater resources (Petalas et al. 2009). Typically, simulation-optimization routines based on seawater intrusion models and optimization algorithms, are utilized to calculate maximum groundwater abstraction subject to constraints that control the landward advancement of seawater (Singh 2014). Variable density and salt transport (VDST) numerical models have been effectively employed in several seawater intrusion studies (e.g. Kerrou et al. 2013; Mahmoodzadeh et al. 2014), however, the resulting computational cost hinders their use in simulation-optimization frameworks for coastal aquifer management.

To reduce the computational burden, recent coastal aquifer management studies have used surrogate modelling techniques to emulate the response of VDST models and enable a computationally tractable optimization model (Sreekanth and Datta 2015). The use of surrogate modelling in the various engineering optimization problems is often reported as surrogate-based optimization (SBO) (Forrester and Keane 2009). Common practice is to create first a space-filling design for the input data and then run the expensive computer models to obtain the required quantity of interest. The surrogate models are then constructed based on these input-output data (Razavi et al. 2012a). Given that the trained surrogate models have attained a reasonable level of accuracy, they are used to provide fast approximations of the original numerical model response to unseen data (Forrester and Keane 2008). The successful construction of a surrogate model largely depends on the size and the spread of the sampling design. However, it is often impractical to use a large sample of input-output data from the original computer models due to computational restrictions. In those cases, it is more effective and efficient to update the initial surrogate model with new training points by utilizing an iterative process (Regis 2011; Zhou et al. 2017).

Significant improvements on the computational requirements for coastal aquifer management have been achieved through the application of the so-called adaptive SBO frameworks (e.g. Kourakos and Mantoglou 2009; Papadopoulou et al. 2010; Christelis et al. 2017; Song et al. 2018). Several surrogate modelling techniques have been proposed in the coastal aquifer management literature including artificial neural networks (Ataie-Ashtiani et al. 2013; Huang and Chiu 2018), genetic programming (Sreekanth and Datta 2011), evolutionary polynomial regression (Hussain et al. 2015), Gaussian process models (Rajabi and Ketabchi

2017), radial basis functions (Christelis and Mantoglou 2016a), fuzzy inference systems (Roy and Datta 2017a), multivariate adaptive regression splines (Roy and Datta 2017b) and support vector machine regression (Lal and Data 2018).

Often, the exploration of the optimal search space may be informed by simpler, computationally efficient models which simulate the physical system at a lower-fidelity level (Forrester et al. 2008). This possibility has motivated the development of variable-fidelity or multi-fidelity optimization (Robinson et al. 2006). Under this framework, surrogate models are constructed using faster lower fidelity models which may share the same physics with the computationally expensive high-fidelity models but are less accurate in terms of grid resolution, convergence criteria, dimensionality, or can be conceptual simplifications of the physical system (Razavi et al., 2012b; Asher et al. 2015). Numerous variable-fidelity frameworks have been developed in the field of electromagnetic simulations through the application of the space mapping technique (Bandler et al. 1994; Koziel et al. 2006), as well as, in aerospace engineering utilizing response correction techniques (e.g. Alexandrov et al. 2001; Gano et al. 2004). Most of the times, the variable-fidelity methods involve the construction of a surrogate model which combines the available fidelity levels and corrects the less accurate but fast lower-fidelity models towards the response of the high-fidelity model (Park et al. 2016).

Since lower-fidelity models utilize their embedded knowledge of the physical system to produce an output, they may offer some benefits for the implementation of SBO (Kennedy and O' Hagan 2000; Koziel and Leifsson 2016). That is, the lower-fidelity model is capable to explain part of the high-fidelity model behavior which in turn may effectively direct the SBO algorithm to promising regions. It is possible though, that the differences between the lower-fidelity and the high-fidelity model to be significant. In that case, the surrogate model may have to approximate abrupt changes in the variable-fidelity data which can be addressed by applying an informative sampling strategy (Zhou et al. 2016). It should be noted that a variable-fidelity SBO may add computational effort and this must always be considered as a possible drawback. Therefore, it is convenient to employ lower-fidelity models which retain a certain degree of accuracy while being much faster that the high-fidelity model.

In the case of seawater intrusion simulation, a coarse taxonomy of the fidelity levels may consider VDST models as high-fidelity models. Then, seawater intrusion models that neglect dispersion mechanisms but simulate saltwater movement (Essaid 1986) or coastal aquifer flow with density variations (Bakker 2003), may represent lower levels of fidelity. An additional simplification and thus an even lower-fidelity level could be defined by sharp interface models which assume static seawater (Strack 1976; Mantoglou et al 2004; Koussis et al. 2012). Obviously, there are other fidelity levels that could be identified in a seawater intrusion simulation framework (VDST models of coarser resolution, steady-state instead of transient coastal aquifer models, etc). In their recent review paper, Sreekanth and Datta (2015) do not report any variable-fidelity applications developed for coastal aquifer management. In the context of variable-fidelity optimization for coastal aquifer management, Christelis and Mantoglou (2016b) recently proposed a method which adaptively corrects the density ratio of a lower-fidelity sharp interface model to adjust its response towards that of the high-fidelity VDST model. Although not explicitly formulated in their work, that approach has some similarities to the implicit space mapping method, but it lacks proper convergence and it is mostly capable of quickly locating promising solutions.

Given the little investigation of variable-fidelity methods in coastal aquifer management, the present work, implements such an approach for a single-objective pumping optimization problem. Here, VDST numerical simulations represent the high-fidelity data whereas a sharp interface model, based on the single-potential formulation of Strack (1976), is employed as a lower-fidelity model. The latter is a simple and computationally efficient model of seawater intrusion which has been employed in several studies to develop simulation-optimization routines (e.g. Mantoglou et al. 2004; Karatzas and Dokou 2015). Therefore, it is worthwhile to investigate its applicability in variable-fidelity optimization frameworks for coastal aquifer management. For comparison purposes, the performance of the variable-fidelity optimization method is evaluated against direct optimization with the VDST model and against conventional SBO with radial basis functions as surrogate models. To enable a more comprehensive comparison among the proposed methods, multiple independent runs of the SBO frameworks are performed. To the best of authors' knowledge, this is the first study in coastal aquifer management which employs a variable-fidelity optimization strategy for reducing the computational cost of the VDST-based optimization and compares its applicability against conventional SBO approaches.

2. Coastal aquifer simulation and pumping optimization

A brief discussion follows regarding the coastal aquifer simulation models that were used in this study. The mathematical formulation of these models has been extensively presented in the relevant literature and it is omitted here for brevity.

2.1 The SI models

VDST models emulate dispersion mechanisms and density variability in space and are considered high-fidelity approximations of coastal aquifer processes (Dokou and Karatzas 2012). The 3D simulations of variable density and salt transport dynamics are based on numerical codes which solve a coupled system of partial differential equations (Werner et al. 2013). VDST modelling is considered a computationally expensive task due to the spatial and time discretization requirements of the solute transport step (Werner et al. 2013). In the present paper, the HydroGeoSphere numerical code (Graf and Therrien 2005) was used to simulate seawater intrusion. HydroGeoSphere applies the control volume finite element method with adaptive time-stepping to solve the coupled system of flow and transport equations and utilizes a Picard iteration scheme to cycle between them (Therrien et al. 2006). Thereinafter, the VDST model will be interchangeably called high-fidelity model.

Sharp interface models based on Ghyben-Herzberg approximation and the single-potential formulation of Strack (1976), are considered as lower-fidelity models since they neglect density variability in space and mixing between freshwater and saltwater. Seawater is assumed static and aquifer flow is horizontal and steady-state. Thus, the dimensionality of the seawater intrusion simulation is reduced to a simple 2D flow equation problem (Mantoglou et al. 2004). Several comprehensive theoretical presentations of the sharp interface approximation can be found elsewhere in the literature (e.g. Strack 1976; Cheng and Quazar 1999). Here, we apply the numerical solution of the flow problem for coastal aquifers of finite size as described in Mantoglou et al. (2004). Thereinafter, the sharp interface model will be interchangeably called lower-fidelity model.

2.2 Coastal aquifer model settings

The numerical simulations are based on an illustrative coastal aquifer model of rectangular shape, which approximates a real aquifer at the Greek Island of Kalymnos (Mantoglou et al. 2004). The dimensions of the coastal aquifer model are $x = 7000m$, $y = 3000m$ while the aquifer thickness is $z = 25m$. Unconfined, steady-state and saturated flow conditions are assumed and the aquifer is replenished by both recharge and lateral fluxes. Table 1 summarizes the basic input parameters for both the VDST and the sharp interface numerical models.

Table 1. Parameters for the numerical SI simulations

Model parameters	VDST (3D)	Sharp interface (2D)
$K_x, K_y, K_z (m/day)$	100,100,10	100, 100,*
$R_{gw}\left(m^3/day\right)$	5422	5422
$\alpha_L, \alpha_T (m)$	100,10	*
$\Delta_x, \Delta_y, \Delta_z \left(m\right)$	100,100,5	100,100,*

K_x, K_y, K_z: hydraulic conductivities, R_{gw}: total aquifer recharge, α_L, α_T: longitudinal and transverse dispersivity values, $\Delta_x, \Delta_y, \Delta_z$: grid discretization settings, *: not applicable

In the absence of real-world data, the dispersivity values for the VDST model were selected to facilitate the setup of a faster VDST model since spatial discretization is related to the dispersivity values via the mesh Peclet number. Note that due to the exploratory nature of this work, multiple runs are performed for the SBO frameworks to take into account the stochastic nature of the evolutionary algorithm and produce an insightful comparison output. In that sense, a relatively fast VDST model is required to realize such a demanding computational task. A single run of the VDST model required an approximate CPU time of 30 seconds, running on a 2.53 GHz Intel i5 processor with 6 GB of RAM in a 64-bit Windows 7 system. On the contrary, a single run of the sharp interface model based on the same computer settings is 0.52 seconds. An initial simulation run of the VDST model was performed with no pumping present, until the head and salinity concentration fields reached steady-state. This step provided the initial conditions for the VDST simulations related to the optimization part of this study.

2.3 Pumping optimization based on the SI models

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 seawater intrusion. On the contrary, the output from the sharp interface model is a single-potential flow field which is used to calculate the "toe" of interface (Mantoglou et al. 2004). The formulation of the optimization problem is presented for nonlinear constraint functions considering fully penetrating pumping wells. The VDST-based optimization is mathematically defined as follows:

$$\begin{aligned} &\min -\sum_{i=1}^{M} Q_i \\ &s.t.\ \ x_i^{c\max}\left(Q_1, Q_2, \ldots Q_M\right) \le xw_i\ , \forall i = 1, 2, \ldots M \\ &\qquad Q_{\min} \le Q_i \le Q_{\max}\ , i = 1, 2, \ldots M \end{aligned} \tag{1}$$

where M is the number of pumping wells, Q_i represents the individual pumping rate, xw_i is the horizontal distance of the well from the coastline, $x_i^{c\max}$ is the horizontal distance of the iso-salinity line $c\max$ from the coast as a function of the pumping rates, and $Q_{\min}$ and $Q_{\max}$ define the lower and upper limits of pumping rates, respectively. Therefore, the problem is set as the maximization of the extracted groundwater amount by M pumping wells, subject to constraints that maintain the levels of salinity concentration in pumped groundwater up to a potable limit $c\max$. The maximum salinity level of $c\max = 35\,mg/lt$, was selected to formulate the high-fidelity constraint functions.

The corresponding optimization formulation for the sharp interface models is:

$$\begin{aligned} &\min -\sum_{i=1}^{M} Q_i \\ &s.t.\ \ x_i^{toe}\left(Q_1, Q_2, \ldots Q_M\right) < xw_i\ , \forall i = 1, 2, \ldots M \\ &\qquad Q_{\min} \le Q_i \le Q_{\max}\ , i = 1, 2, \ldots M \end{aligned} \tag{2}$$

where the set of the constraint functions do not allow the "toe" of the interface x^{toe} to reach the location of 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. In this study, the evolutionary annealing-simplex (EAS) algorithm (Efstratiadis and Koutsoyiannis 2002) was used to solve the optimization problems defined in (1) and (2). To apply heuristic optimization, the nonlinear constraints are embedded in the objective function using penalty terms. Here, the objective function is penalized according to the following formulation for the VDST model (Christelis et al. 2017):

$$\min f\left(Q\right)=\begin{cases}-\sum_{i=1}^{M}Q_i,\ if\ \forall\, i=1,2...M;\, x_i^{c\max}\left(Q_1,Q_2,..Q_M\right)\le xw_i\\ M_v\sum_{i=1}^{M}\left[\max\left(\left(x_i^{c\max}-xw_i\right),0\right)\right]^2,if\ \exists\, i=1,2...M;\, x_i^{c\max}\left(Q_1,Q_2,..Q_M\right)>xw_i\end{cases}$$

(3)

where M_v represents the number of pumping wells that the constraint is violated. A similar formulation is defined for the sharp interface model:

$$\min f\left(Q\right)=\begin{cases}-\sum_{i=1}^{M}Q_i,\ if\ \forall\, i=1,2...M;\, x_i^{toe}\left(Q_1,Q_2,..Q_M\right)<xw_i\\ M_v\sum_{i=1}^{Mv}\left[\max\left(\left(x_i^{toe}-xw_i\right),0\right)\right]^2,if\ \exists\, i=1,2...M;\ x_i^{toe}\left(Q_1,Q_2,..Q_M\right)\ge xw_i\end{cases}$$

(4)

The few parameters that must be initially set for the EAS algorithm were defined according to Efstratiadis and Koutsoyiannis (2002) and Tsoukalas et al. (2016). Thus, the initial population size is set to $n_{pop}=8M$, the two annealing schedule control parameters to $\lambda p = 0.95$ and $\psi = 2$, the mutation probability is set to $mp = 0.1$ and the convergence criterion to $\varepsilon = 10^{-4}$. For all optimization frameworks the termination criteria are met if the convergence criterion ε equals its pre-set value or the number of maximum objective function evaluations equal $n_{\max} = 100 n_{pop}$.

3. Development of the SBO frameworks

3.1 The conventional approach

As mentioned in the introduction section there are various surrogate models that can be employed for SBO. Here, cubic radial basis function (RBF) models are utilized. RBF models have been successfully applied in several SBO studies (e.g. Mugunthan et al. 2005; Sun et al. 2011; Tsoukalas et al. 2016; Christelis et al. 2017). The training time of cubic RBF surrogate models has a low computational cost which is desirable for the development of a SBO framework. In addition, RBF models are interpolating surrogates which means that they pass through all training data, a favourable feature for approximating the deterministic computer outputs produced in the present work.

For convenience, let Q be the decision vector of pumping rates $Q = (Q_1, Q_2, ..., Q_M)$ and $x_i^{c\max}(Q)$ the scalar response for the ith pumping well as calculated by the VDST simulation. It is noted that the number of the constraint functions equals the number of pumping wells. Therefore, a unique RBF model is constructed for each one of the pumping wells. A Latin Hypercube Sampling (LHS) design was utilized to uniformly sample the decision vector space, evaluate the VDST model and create an initial set of m training patterns for each RBF model. The set of the decision vectors $Q^{(1)}, Q^{(2)}, ..., Q^{(m)} \in R^M$ obtained from the LHS design and the values $x_i^{c\max}\left(Q^{(1)}\right), x_i^{c\max}\left(Q^{(2)}\right), ... x_i^{c\max}\left(Q^{(m)}\right), i = 1, ..., M$ obtained from the VDST model, define a cubic RBF model, augmented with a linear polynomial tail, of the following form (Powell, 1992):

$$S_m\left(Q\right) = \sum_{k=1}^{m} \lambda_k \phi\left(\|\left(Q - Q^{(k)}\|\right)\right) + p\left(Q\right) \qquad (5)$$

Here, the cubic form is applied, where $\phi\left(r\right) = r^3$, $\lambda_1, ..., \lambda_m \in R$ are coefficients to be determined, and $p\left(Q\right)$ is a linear polynomial whose coefficients also need to be determined. To obtain the coefficients in the above cubic RBF model the matrix $\Phi \in R^{M x M}$ is defined where $\Phi_{k,l} = \phi\left(\| Q^{(k)} - Q^{(l)} \|\right)$ and the matrix $P \in R^{m x (M+1)}$ whose *ith* row is $\left[1, \left(Q^{(i)}\right)^T\right]$. The vector $x_i^{VDST} = \left[x_i^{c\max}\left(Q^{(1)}\right), x_i^{c\max}\left(Q^{(2)}\right), ... x_i^{c\max}\left(Q^{(m)}\right)\right]^T$ is also defined. The vector of coefficients $\lambda = \left[\lambda_1, ..., \lambda_m\right]^T$ for the RBF part and the coefficients $c = \left[c_0, c_1, ... c_M\right]^T$ for the polynomial part are obtained by solving the following system of linear equations:

$$\begin{pmatrix} \Phi & P \\ P^T & 0 \end{pmatrix} \begin{pmatrix} \lambda \\ c \end{pmatrix} = \begin{pmatrix} x_i^{VDST} \\ 0 \end{pmatrix} \qquad (6)$$

In the present study, new training points are added to the initial sampling plan by evaluating the VDST model whenever the surrogate models find a new better solution during the optimization algorithm operations. Whereas this is not considered as a global SBO approach, it can efficiently locate near global optimal solutions and it has been successfully applied in previous single-objective pumping optimization studies of coastal aquifers (Kourakos and Mantoglou 2009). It is possible, particularly during the first few iterations of the EAS algorithm, that the surrogate models will incorrectly predict that $x_i^{c\max}(Q) < x_{init}^{c\max}$, where $x_{init}^{c\max}$ is obtained from the initial simulation with the VDST model without pumping. This could be avoided in a certain degree if we add more training points in the initial sampling design. However here, it was used as an additional model-based criterion to retrain the RBF models during optimization. A summary of the adaptive SBO framework is given below:

1. Use LHS design to provide the initial m training points $Q^{(1)}, Q^{(2)}, ..., Q^{(m)}$ and get $x_i^{c\max}\left(Q^{(1)}\right), x_i^{c\max}\left(Q^{(2)}\right), ... x_i^{c\max}\left(Q^{(m)}\right), i = 1, ..., M$ through m VDST simulations.
2. Construct M RBF models and create an external archive of training patterns.
3. Run EAS algorithm based on the RBF models and if during optimization a better optimum is found or any *ith* RBF model prediction is $x_i^{c\max}(Q) < x_{init}^{c\max}$ do the following:
   a) Re-evaluate the current best decision vector with the VDST model.
   b) Replace the objective function value with the VDST solution.

c) Store the new input-output data to the archive and re-train the surrogate models.

4. Are stopping criteria for EAS algorithm met? If yes, return final solution, else go to step 3.

3.2 The variable-fidelity approach

Previous seawater intrusion studies have demonstrated that the sharp interface model utilized here, provides conservative estimations of the optimal pumping rates compared to the VDST model (Pool and Carrera 2011; Kopsiaftis et al. 2017). An example is shown in Figure 1 for a specific input Q where the sharp interface model output shows a more severe landward advancement of the seawater wedge.

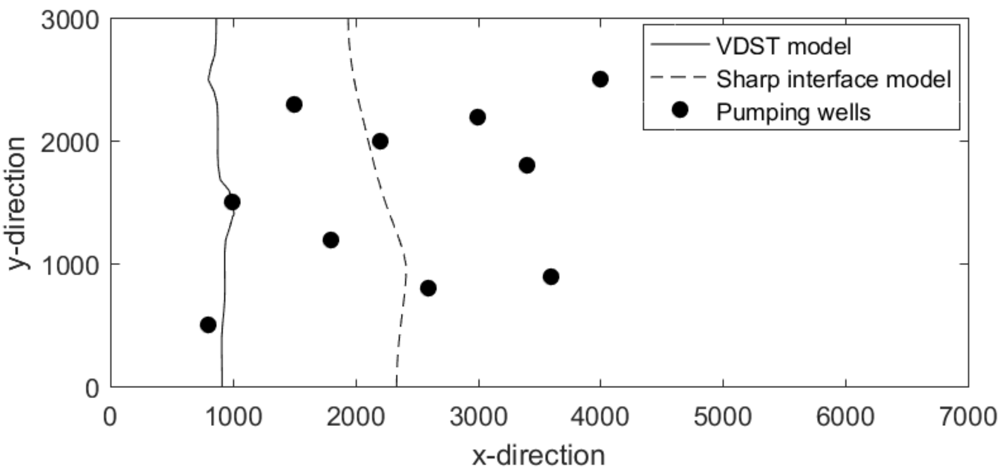

Figure 1. Advancement of seawater intrusion as simulated by the VDST and the sharp interface model for the same set of pumping rates.

In a previous study, Pool and Carrera (2011) developed an empirical correction for the sharp interface model to better match the maximum pumping rates calculated from the VDST models. Practically, a modified density ratio for the sharp interface model of Strack (1976) is calculated based on the aquifer depth

B and the transverse dispersivity value α_T defined in the VDST model as follows (see Pool and Carrera, 2011):

$$\varepsilon^* = \varepsilon\left[1-\left(\frac{a_T}{B}\right)^{1/6}\right] \tag{7}$$

where ε is the saltwater-freshwater density ratio defined as $\varepsilon = \left(\rho_s - \rho_f\right)/\rho_f$, with ρ_s being the saltwater density and ρ_f the freshwater density. Lu and Werner (2013) proposed that the exponent in equation (7) could be changed to $1/4$. However, as it is discussed in Christelis and Mantoglou (2016b) different pumping stresses may require different modifications of the density ratio to achieve a better match for the two models in a pumping optimization problem. Here, for comparison purposes, the variable-fidelity optimization framework is also implemented using the sharp interface model corrected by the different values of the density ratio proposed in Pool and Carrera (2011) and Lu and Werner (2013). Furthermore, all 3 sharp interface models considered here, are also utilized to formulate an ensemble surrogate where each of the sharp interface model response is given an equal weight. This is to investigate if the combination of the variable-fidelity data with a simple averaging approach can improve the exploration of the optimal search space.

Several response correction techniques have been suggested in the literature such as multiplicative or additive formulations (Leary et al. 2003). A simple correction process can be modeled by any approximation model (e.g. radial basis functions, kriging, etc.) which is fitted, for example, to the ratio R_{HF}/R_{LF} or to the difference $R_{HF} - R_{LF}$, with R_{HF} and R_{LF} being the high-fidelity and lower-fidelity model responses, respectively (Forrester et al. 2008). Let $x_i^{toe}(Q)$ be the scalar response for the *ith* pumping well as calculated by the sharp interface simulation. The initial training points $Q^{(1)}, Q^{(2)}, \ldots, Q^{(m)} \in R^M$ produced by the LHS design are also used here to obtain $x_i^{toe}\left(Q^{(1)}\right), x_i^{toe}\left(Q^{(2)}\right), \ldots x_i^{toe}\left(Q^{(m)}\right), i = 1, \ldots, M$ through simulation with the sharp interface model. The corresponding values from the VDST model are $x_i^{c\max}\left(Q^{(1)}\right), x_i^{c\max}\left(Q^{(2)}\right), \ldots x_i^{c\max}\left(Q^{(m)}\right), i = 1, \ldots, M$. A cubic RBF model is then trained to approximate a simple multiplicative formulation of the form $x_i^{c\max}(Q) / x_i^{toe}(Q)$ adopted for each one of the pumping wells. Thus, the surrogate model predicts the $x_i^{c\max}(Q)$ value for the *ith* pumping well based on the following correction:

$$\hat{x}_i^{c\max}(Q) = x_i^{toe}(Q) S(Q) \tag{8}$$

Therefore, the sharp interface model corrected by the RBF model is the surrogate model for the VDST model. For each pumping well, a unique RBF model S is constructed. This scaling model between the lower-fidelity and the high-fidelity data, may be reconstructed during optimization, if the evaluation with the high-fidelity model indicates a poor performance of the corrected LF model (Thokala and Martins, 2007). A similar adaptive SBO framework with the conventional approach is also used for the variable-fidelity optimization strategy:

1. Use LHS design to provide the initial m training points $Q^{(1)}, Q^{(2)}, ..., Q^{(m)}$ and get $x_i^{c\max}\left(Q^{(1)}\right), x_i^{c\max}\left(Q^{(2)}\right), ... x_i^{c\max}\left(Q^{(m)}\right), i = 1, ..., M$ through m high-fidelity simulations and $x_i^{toe}\left(Q^{(1)}\right), x_i^{toe}\left(Q^{(2)}\right), ... x_i^{toe}\left(Q^{(m)}\right), i = 1, ..., M$ through m lower-fidelity simulations.
2. Create an external archive of training patterns of the above high-fidelity and lower-fidelity data and train M RBF models to learn the ratio $x_i^{c\max}(Q)/x_i^{toe}(Q)$.
3. Run EAS algorithm based on the surrogate model described in equation (8).
4. If during optimization a better optimum is found or the *ith* surrogate model prediction is $x_i^{c\max}(Q) < x_{init}^{c\max}$ do the following:
    a. Re-evaluate the current decision vector with the VDST model.
    b. Replace the objective function value with the VDST solution.
    c. Store the new input-output data to the archive and re-train the surrogate models.
5. Are stopping criteria for EAS algorithm met? If yes, return final solution, else go to step 3.

4. Results and discussion

Since EAS is a probabilistic optimization method, a set of 30 independent optimization runs was used for each SBO approach to build more confidence among the optimal results. Furthermore, for each optimization run a different initial training set was produced with the VDST model and it was applied as initial population for the EAS algorithm. Thus, for each run, all surrogate models were trained under the same sampling design and the optimal solutions were also based on the same initial population to ensure a fair comparison among them. To handle the computational burden, a single run was executed for each of the direct optimization with the VDST and the sharp interface models. This is also justified by the robust performance of EAS in previous coastal aquifer management studies with these models (Christelis and Mantoglou 2016a). For convenience, the optimization frameworks are henceforth referred with their abbreviated names as follows:

1. EAS-HF: direct optimization using the high-fidelity VDST model.
2. EAS-LF: direct optimization using the sharp interface model
3. EAS-LF-PC: direct optimization using the sharp interface model with density ratio corrected as in Pool and Carrera (2011).
4. EAS-LF-LW: direct optimization using the sharp interface model with density ratio corrected as in Lu and Werner (2013).
5. EAS-RBF: Conventional SBO using the RBF models.
6. EAS-VF: Variable-fidelity SBO using the sharp interface model.
7. EAS-VF-PC: Variable-fidelity SBO using the sharp interface model with density ratio corrected as in Pool and Carrera (2011).
8. EAS-VF-LW: Variable-fidelity SBO using the sharp interface model with density ratio corrected as in Lu and Werner (2013).
9. EAS-VF-ENS: Variable-fidelity SBO with all 3 sharp interface models forming an ensemble in a simple averaging approach.

Table 2 presents the summary statistics of the SBO runs along with the results from direct optimization with the seawater intrusion models.

Table 2. Summary statistics comparisons among the optimization frameworks. The best SBO performance is highlighted while the high-fidelity solution from the VDST model is underlined.

	Best (m^3/day)	Worst (m^3/day)	Median (m^3/day)	CPU time (hr) (average*)
EAS-HF	4871.6	n/a	n/a	44.8
EAS-LF	2683.9	n/a	n/a	0.8
EAS-LF-PC	4810.3**	n/a	n/a	0.8
EAS-LF-LW	4724.4**	n/a	n/a	0.8
EAS-RBF	4856.8	4378.1	4703.6	**1.64**
EAS-VF	**4868.3**	4372.3	4697.4	2.36
EAS-VF-PC	4768.2	**4585.9**	4720	2.36
EAS-VF-LW	4780.1	4538.4	4697.6	2.36
EAS-VF-ENS	4796.9	4577.1	**4731.2**	2.66

n/a: not applicable

*: only for the SBO frameworks

**: not a feasible solution after evaluation with the VDST model

As expected, EAS-LF produced the lowest best objective function value, since the sharp interface model tends to overestimate seawater intrusion and therefore lower maximum pumping rates are calculated (Pool and Carrera, 2011). On the contrary, EAS-LF-PC and EAS-LF-LW provided higher optimal solutions than EAS-LF since the modification of the density ratio allows for larger groundwater extraction. Note that the optimal solutions obtained from the sharp interface models were evaluated using the VDST model. This is because EAS-LF, EAS-LF-PC and EAS-LF-LW frameworks do not involve any high-fidelity VDST runs. Thus, their optimal solutions were evaluated with the VDST model to check if they are feasible based on the set of constraints which are defined on the VDST model. The EAS-LF solutions were feasible but the EAS-LF-PC and EAS-LF-LW violated the constraints for the VDST model. Nevertheless, the above results do not conclude that EAS-LF-PC or EAS-LF-LW optimization would generally fail to provide feasible optimal solutions. The corrections proposed in Pool and Carrera (2011) and Lu and Werner (2013) constitute one-off corrections of the sharp interface model and the latter may approximate different parts of the dispersive zone of the VDST model depending also on the hydraulic parameter sets in both models. Therefore, other salinity levels defined in the VDST-based optimization may be satisfied by the solutions produced from EAS-LF-PC or EAS-LF-LW. It cannot be ignored that these approaches can potentially provide information for regions with good local optima for the VDST model.

As also shown in Table 2, EAS-RBF was the most computationally efficient method among the SBO frameworks, requiring less than an hour to converge. It also came second on providing the best objective function value after EAS-VF. The variable-fidelity methods in general, required approximately 30% more computational time to converge than the conventional EAS-RBF approach. The ensemble approach (EAS-VF-ENS) provided the highest median but also had the largest computational cost among the SBO methods. Figure 2 presents the distribution of the optimal solutions via box plot visualization for a more detailed comparison of the SBO frameworks.

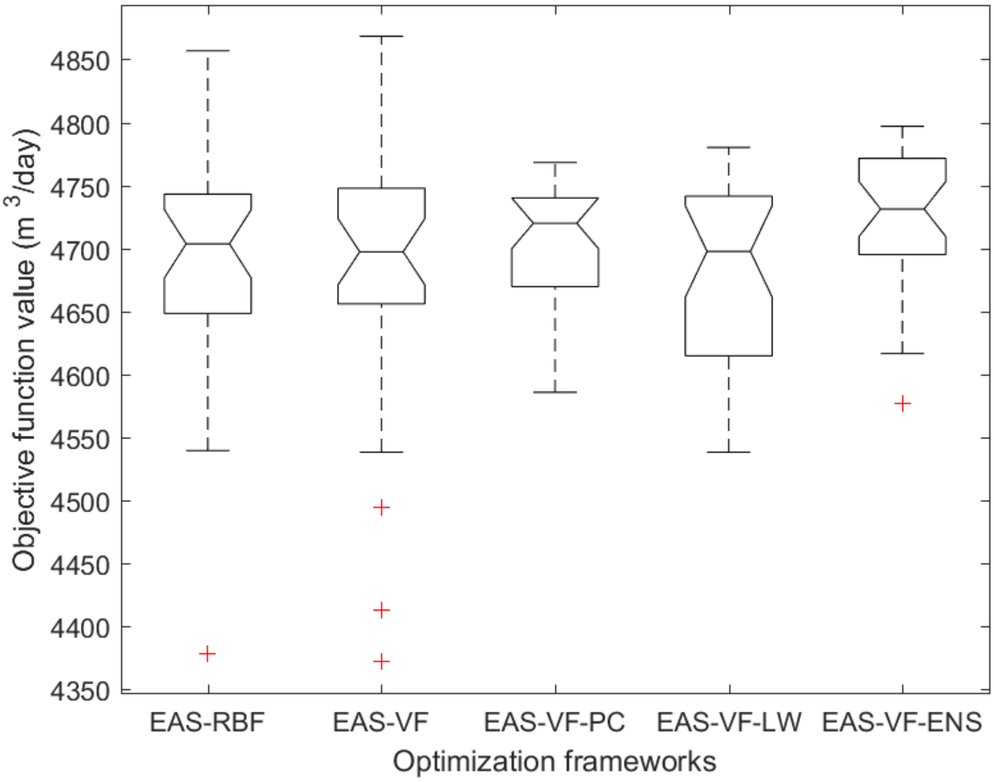

Figure 2. Performance of the SBO frameworks based on boxplots

As demonstrated the interval endpoints of the boxplot notches appear to overlap which implies that the medians of the SBO frameworks probably do not exhibit any statistically significant difference. Although the SBO sample runs are limited due to computational restrictions, a one-way analysis of variance (ANOVA) test also provided a p-value of 0.18 which increases the belief that for the case studied here the medians are not significantly different at a 5% significance level. It is interesting to see the average percentage of the successful predictions of feasible solutions with the surrogate models as these were evaluated by the VDST model during the operations of the optimization algorithm. That is, how many of the constraint function predictions from the surrogates were feasible whenever the VDST was called by the SBO algorithm to evaluate the current solution. The EAS-VF-ENS has a higher percentage than the other SBO methods while the other variable-fidelity methods (EAS-VF, EAS-VF-PC and EAS-VF-LW) demonstrate similar performance but not as good as the conventional EAS-RBF approach. Given, the similar statistical performance of all the SBO frameworks it appears that even the less accurate constructed surrogate models can still drive the optimization algorithm to good optimal solutions.

Table 3. Successful surrogate model predictions of feasible solutions based on the VDST model evaluation during the optimization operations.

	EAS-RBF	EAS-VF	EAS-VF-PC	EAS-VF-LW	EAS-VF-ENS
Percentage of feasible surrogate model predictions (average)	60.19	57.26	57.89	57.30	61.71

5. Conclusions

The problem of pumping optimization of coastal aquifers was solved considering direct optimization, as well as, SBO methods to reduce the computational cost derived from variable density and salt transport simulations. The SBO methods were developed using an adaptive framework by embedding the surrogate model update process in the operations of an evolutionary optimization algorithm. A conventional and a variable-fidelity surrogate model approach were employed. The objective was to identify whether a variable-fidelity approach which utilizes a simple scaling function that corrects the lower-fidelity models can outperform the conventional SBO and provide good approximations to the direct high-fidelity optimization. The variable-fidelity method was developed using the sharp interface assumption and the single-potential formulation of Strack (1976), as well as variations of this model based on recent proposed correction factors. In addition, an ensemble surrogate model, based on a simple averaging approach, was constructed by using all the variations of the sharp interface model.

Results demonstrated that the SBO approaches performed equally well and found optimal solutions close to those obtained from the direct optimization with the high-fidelity VDST model. The computational gains by applying the SBO methods were above 90% of the computational time from the VDST-based optimization. Although the variable-fidelity approach provided the best optimal solution and the highest median, it also added computational cost. Furthermore, the overall sample statistics implied that there wasn't a statistically significant difference among the medians of the SBO methods. The combination of all sharp interface models to form an ensemble surrogate model reduced the spread of the solutions and provided the highest median but failed to find the best objective function value among the SBO frameworks. The overall understanding is that given the lower computational cost of the conventional SBO approach, the variable-fidelity method followed in this study couldn't provide strong evidence than can perform better based on the derived sample statistics.

However, the results here are limited to the coastal aquifer model settings that were adopted in this study. It is of practical interest to investigate if variable-fidelity methods can provide an effective alternative to SBO for coastal aquifer management studies by testing other model settings and different correction and enhancement techniques of lower-fidelity seawater intrusion models. Future work will focus on utilizing other levels of fidelity apart from the sharp interface assumption, as well as investigating the construction of efficient surrogate ensembles based on models of variable-fidelity.

References

Alexandrov NM, Lewis RM, Gumbert CR, Green LL, Newman PA (2001) Approximation and model management in aerodynamic optimization with variable-fidelity models. J Aircraft 38(6):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(8): 5957-5973

Ataie-Ashtiani B, Ketabchi H, Rajabi MM (2013) Optimal management of a freshwater lens in a small island using surrogate models and evolutionary algorithms. J Hydrol Eng 19(2):339-354

Bakker M (2003) A Dupuit formulation for modeling seawater intrusion in regional aquifer systems. Water Resour Res 39(5)

Bandler JW, Biernacki RM, Chen SH, Grobelny PA, Hemmers RH (1994) Space mapping technique for electromagnetic optimization. IEEE T Microw Theory 42(12):2536-2544

Bandler JW, Cheng QS, Nikolova NK, Ismail MA (2004) Implicit space mapping optimization exploiting preassigned parameters. IEEE T Microw Theory 52(1): 378-385.

Christelis V, Mantoglou A (2016a) Pumping optimization of coastal aquifers assisted by adaptive metamodelling methods and radial basis functions. Water Resour Manage 30(15):5845-5859

Christelis V, Mantoglou A (2016b) Coastal aquifer management based on the joint use of density-dependent and sharp interface models. Water Resour Manage 30(2):861-876

Christelis V, Regis RG, Mantoglou A (2017) Surrogate-based pumping optimization of coastal aquifers under limited computational budgets. J Hydroinform 20(1):164-176

Dokou Z, Karatzas GP (2012) Saltwater intrusion estimation in a karstified coastal system using density-dependent modeling and comparison with the sharp-interface approach. Hydrol Sci J 57(5):985–999

Essaid HI (1986) A multilayered sharp interface model of coupled freshwater and saltwater flow in coastal systems: Model development and application. Water Resour Res 26(7):1431-1454

Park C, Haftka RT, Kim NH (2017) Remarks on multi-fidelity surrogates. Struct Multidiscip Optim 55(3):1029-1050.

Forrester AIJ, Sóbester A, Keane AJ (2008) Engineering design via surrogate modelling-A practical guide. Wiley, New York

Forrester AIJ, Keane AJ (2009) Recent advances in surrogate-based optimization. Prog Aerosp Sci 49:50-79

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 (p. 4460).

Graf T, Therrien R (2005) Variable-density groundwater flow and solute transport in porous media containing nonuniform discrete fractures. Adv Water Resour 28:1351-1367

Huang PS, Chiu YC (2018) A Simulation-Optimization Model for Seawater Intrusion Management at Pingtung Coastal Area, Taiwan. Water 10(3):251.

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

Karatzas GP, Dokou Z (2015) Optimal management of saltwater intrusion in the coastal aquifer of Malia, Crete (Greece) using particle swarm optimization. Hydrogeol J doi: 10.1007/s10040-015-1286-6

Kennedy MC, O'Hagan A (2000) Predicting the output from a complex computer code when fast approximations are available. Biometrika 87(1):1-13

Kerrou J, Renard P, Cornaton F, Perrochet P (2013) Stochastic forecasts of seawater intrusion towards sustainable groundwater management: application to the Korba aquifer (Tunisia). Hydrogeol J 21(2):425-440

Kopsiaftis G, Christelis V, Mantoglou A (2017) Pumping optimization in coastal aquifers: Comparison of sharp interface and density dependent models. European Water, 57: 443-449

Kourakos G, Mantoglou A (2009) Pumping optimization of coastal aquifers based on evolutionary algorithms and surrogate modular neural network models. Adv Water Resour 32(4): 507-521

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 416:1-11

Koziel S, Bandler JW, Madsen K (2006) A space-mapping framework for engineering optimization—Theory and implementation. IEEE T Micro Theory Techniques, 54(10):3721-3730

Koziel S, Leifsson L (2016) Simulation-Driven Design by Knowledge-Based Response Correction Techniques. New York, Springer

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 Manage 32(7):2405-2419

Leary SJ, Bhaskar A, Keane AJ (2003) A knowledge-based approach to response surface modelling in multifidelity optimization. J Global Optim 26(3):297-319

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

Mahmoodzadeh D, Ketabchi H, Ataie-Ashtiani B, Simmons CT (2014) Conceptualization of a fresh groundwater lens influenced by climate change: A modeling study of an arid-region island in the Persian Gulf. Iran. J Hydrol 519: 399-413

Mantoglou A, Papantoniou M, Giannoulopoulos P (2004) Management of coastal aquifers based on nonlinear optimization and evolutionary algorithms. J Hydrol 297(1–4):209–228

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(11)

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(7):1479-1490

Petalas C, Pisinaras V, Gemitzi A, Tsihrintzis VA, Ouzounis K (2007) Current conditions of saltwater intrusion in the coastal Rhodope aquifer system, northeastern Greece. Desalination 237:22-41

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 doi:10.1029/2010WR010256

Powell MJD (1992) The theory of Radial Basis Function Approximation in 1990, in: Light, W. (ed.), Advances in Numerical Analysis, Volume 2: Wavelets, Subdivision Algorithms and Radial Basis Functions. pp. 105–210, Oxford University Press

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

Razavi S, Tolson BA, Burn DH (2012a) Review of surrogate modeling in water resources. Water Resour Res 48(7)

Regis RG, Shoemaker CA (2005) Constrained global optimization of expensive black box functions using radial basis functions. J Global Optim 31(1):153-171.

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

Robinson T, Willcox K, Eldred M, Haimes R (2006) Multifidelity optimization for variable-complexity design. In 11th AIAA/ISSMO multidisciplinary analysis and optimization conference (p. 7114)

Roy DK, Datta B (2017a) Fuzzy C-Mean Clustering Based Inference System for Saltwater Intrusion Processes Prediction in Coastal Aquifers. Water Resour Manage 31(1):355-376

Roy DK, Datta B (2017b) Multivariate Adaptive Regression Spline Ensembles for Management of Multilayered Coastal Aquifers. J Hydrol Eng 22(9):04017031

Song J, Yang Y, Wu J, Wu J, Sun X, Lin J (2018) Adaptive surrogate model based multiobjective optimization for coastal aquifer management. J Hydrol 561:98-111

Sreekanth J, Datta B (2011) Comparative evaluation of Genetic Programming and Neural Network as potential surrogate models for coastal aquifer management. Water Resour Manage 25:3201-3218

Sreekanth J, Datta B (2015) Review: Simulation-optimization models for the management and monitoring of coastal aquifers. Hydrogeol J doi:10.1007/s10040-015-1272-z

Strack ODL (1976) A single-potential solution for regional interface problems in coastal aquifers. Water Resour Res 12(6):1165–1174

Sun G, Li G, Gong Z, He G, Li Q (2011) Radial basis functional model for multi-objective sheet metal forming optimization. Eng Optim 43(12):1351-1366

Therrien R, McLaren RG, Sudicky EA, Panday SM (2006) HydroGeoSphere-A three-dimensional numerical model describing fully-integrated subsurface and surface flow and solute transport. Groundwater Simulations Group, University of Waterloo, Canada, draft ed. 2006

Thokala P, Martins JR (2007) Variable-complexity optimization applied to airfoil design. Eng Optimiz 39(3):271-286

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 Modell Softw 77: 122-142

Werner AD, Bakker M, Post VEA, Vandenbohede A, Lu C, Ataie-Ashtiani B, Simmons CT, Barry DA (2013) Seawater intrusion processes, investigation and management: Recent advances and future challenges. Adv Water Resour 51:3-26

Zhou Q, Shao X, Jiang P, Gao Z, Wang C, Shu L (2016) An active learning metamodeling approach by sequentially exploiting difference information from variable-fidelity models. Adv Eng Inform, 30(3):283-297

Zhou Q, Wang Y, Jiang P, Shao X, Choi SK, Hu J, Cao L, Meng X (2017) An active learning radial basis function modeling method based on self-organization maps for simulation-based design problems. Knowl-Bas Syst, 131:10-27