Surrogate-based pumping optimization of coastal aquifers under limited computational budgets

Vasileios Christelis[1,*], Rommel G. Regis[2], 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] Department of Mathematics, Saint Joseph's University Philadelphia, USA

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

* British Geological Survey, Environmental Science Centre, Keyworth, Nottingham, UK

Abstract

The long runtimes of variable density and salt transport numerical models hinder the implementation of simulation-optimization routines for coastal aquifer management. To reduce this excessive computational cost, surrogate models have been successfully applied in several studies. However, it has not been previously addressed how effective is surrogate modelling in pumping optimization of coastal aquifers, given a limited number of available runs with the seawater intrusion model. To that end, two surrogate-based optimization frameworks are employed and compared against the direct optimization approach under restricted computational budgets. The first surrogate-assisted algorithm, utilizes an infill strategy aiming at a fast local improvement of the surrogate model around optimal values. The other, balances global and local improvement of the surrogate model while it is applied for the first time in coastal aquifer management. The performance of the algorithms is investigated for optimization problems of moderate and large dimensionality. Results indicate that for all problems, the surrogate-based optimization methods provide higher objective function values than the direct optimization. Additionally, the selection of cubic radial basis function surrogate models, enables the construction of very fast approximations for problems with up to 40 decision variables and 40 constraint functions.

INTRODUCTION

Variable density and salt transport (VDST) numerical models are indispensable tools for simulating seawater intrusion (SWI) in coastal aquifers (Werner *et al*. 2013). They have been effectively employed to improve understanding in real-world SWI problems (e.g. Gingerich & Voss 2005; Giambastiani *et al*. 2007; Kopsiaftis *et al*. 2009; Kerrou *et al*. 2013). Additionally, the simulation of dispersive flow between seawater and freshwater by using VDST models, enables a more accurate management of groundwater abstraction in coastal aquifers (Pool and Carrera , 2011).

However, VDST models are computationally expensive, as is the case with most of the high-fidelity computer simulations. Hence, their use in iterative numerical tasks, such as sensitivity analysis or optimization, is hindered by the increased computational cost. To address this issue, several studies have employed data-driven surrogate modelling techniques either to partly or fully replace the computationally expensive VDST simulations (Sreekanth & Datta 2015). Examples of surrogate models in coastal aquifer management comprise artificial neural networks (e.g. Rao *et al.* 2004; Bhattacharjya & Datta 2005; Kourakos & Mantoglou 2009; Ataie-Ashtiani *et al.* 2013; Kourakos & Mantoglou 2013; Roy *et al.* 2016), genetic programming (Sreekanth & Datta 2011), evolutionary polynomial regression (Hussain *et al*. 2015), polynomial chaos expansions (Rajabi *et al*. 2015), radial basis functions (Christelis & Mantoglou 2016a) or fuzzy inference systems (Roy & Datta 2016).

Typically, an initial set of input-output data from the physics-based models is used to train the surrogate models in order to attain a certain level of accuracy for predicting responses to unseen data (Solomatine & Ostfeld 2008). It is unlikely though that a global accurate surrogate model can be constructed, given that the number of available runs with the original model is usually limited due to computational restrictions (Forrester *et al*. 2008). In certain coastal aquifer management studies, hundreds to thousands input-output patterns were used to construct an accurate surrogate model (Sreekanth & Datta 2015). The use of large training patterns may lead to impractical computational cost even for a VDST model with simulation runtimes of few minutes.

Most coastal aquifer management studies, have applied surrogate-based optimization (SBO) methods without pre-specified restrictions on the overall computational budget. The use of adaptive surrogate training frameworks has significantly reduced the associated computational burden (e.g. Kourakos & Mantoglou 2009; Papadopoulou *et al*. 2010; Christelis & Mantoglou 2016a). Alternatively, Ataie-Ashtiani *et al*. (2014) proposed a zonation methodology as a practical approach to reduce the dimensionality of the optimization problem and therefore the required training data for building the surrogate models. It is also worth noting that pumping optimization problems of coastal aquifers usually involve non-linear constraints (Mantoglou *et al*. 2004). The presence of non-linear constraints further complicates the development of SBO methods (Forrester et al. 2008).

Nevertheless, many engineering optimization studies have focused on approximating the global optimum based on a specified number of runs with the original expensive computer model. There is a wide body of SBO literature which develops adaptive sampling strategies that effectively utilize the expensive original model runs, to update the surrogate and increase its accuracy within regions of interest (e.g. Jones 1998; Mugunthan *et al.* 2005; Regis & Shoemaker 2007; Forrester & Keane 2009; Parr *et al*. 2012; Regis & Shoemaker 2013; Regis 2014; Tsoukalas *et al*. 2016). However, the application of comprehensive SBO strategies which exploit information from the surrogate models in order to sample the expensive original model is rather limited in groundwater modelling and optimization (Asher *et al*. 2015). Furthermore, it is debatable if there is a benefit from the use of surrogate models in optimization problems with increased dimensionality and under limited computational budgets (Razavi *et al*. 2012a).

In the present paper, we address the effectiveness of surrogate modelling in pumping optimization of coastal aquifers, given a limited number of available runs with the expensive SWI model. Two SBO frameworks are employed in order to solve single-objective pumping optimization problems. The first SBO algorithm utilizes a metamodel-embedded evolution framework which constructs radial basis function (RBF) surrogate models for the constraints functions only. RBF surrogate models have been successfully applied in several SBO optimization problems (Razavi *et al.* 2012b). The other is an advanced SBO algorithm, namely, ConstrLMSRBF (Regis, 2011), which simultaneously deals with the objective function and the constraints of the optimization problem, by constructing RBF surrogate models for each one of them. ConstrLMSRBF algorithm is applied for the first time in water resources optimization and for problems of pumping optimization of coastal aquifers. The goal of this study is to investigate the performance of these SBO algorithms on different dimensionalities of the decision variable space while imposing strong restrictions on the number of available runs with the VDST model. The latter assumption is closer to real-world cases where coastal aquifer management problems involve computationally heavy numerical models of SWI. The SBO algorithms are compared against direct optimization with the VDST model in order to evaluate the usefulness of constructing surrogate models in the case of limited computational budgets.

The rest of the paper includes 4 sections. Section 2 presents the SWI numerical simulation model, the coastal aquifer model and the formulation of the pumping optimization problem. In section 3 the surrogate models along with their implementation in SBO strategies are described. Section 4 presents the optimization results and finally section 5 concludes on the findings of the present study.

METHODS

SWI modelling

VDST models utilize numerical codes which solve a coupled system of partial differential equations of flow and transport in order to simulate SWI (Voss & Souza 1987). It is considered a complicated and computationally expensive numerical task mostly due to the spatial and time discretization requirements of the solute transport step (Werner *et al*. 2013). In the present paper, the HydroGeoSphere code (HGS) (Therrien & Sudicky 1996; Graf & Therrien 2005; Therrien *et al*. 2006) was used to simulate SWI. The HGS code applies the control volume finite element method with adaptive time-stepping while a Picard iteration scheme is utilized to iteratively solve the system of flow and transport equations for VDST simulations (Thompson *et al.* 2007). The mathematical formulation of VDST modelling is briefly described below whereas comprehensive presentations can be found elsewhere (e.g. Kolditz *et al.* 1998).

$$\frac{\partial}{\partial x_i}\left[K_{ij}\left(\frac{\partial h_f}{\partial x_j}+\rho_r n_j\right)\right]+Q_\rho = S_s\frac{\partial h_f}{\partial t} \tag{1}$$

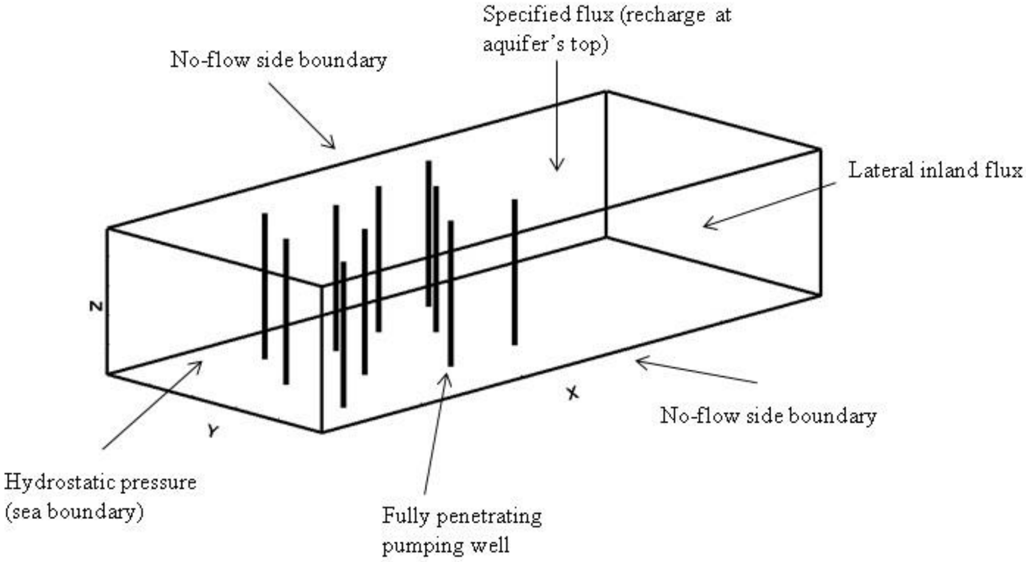
$$\frac{\partial}{\partial x_i}\left(D_{ij}\frac{\partial c}{\partial x_j}-q_i c\right)+Q_c = \frac{\partial(\phi c)}{\partial t} \tag{2}$$

In the flow equation (1) the equivalent freshwater head $hf\ [L]$ is the flow variable given by $h_f = (p/\rho_f g)+z$, where $p\ [ML^{-1}T^{-2}]$ is the fluid pressure, $\rho_f\ [ML^{-3}]$ is the reference fluid density, $g\ [ML^{-2}]$ is the gravity acceleration constant and $z\ [L]$ is the elevation above horizontal datum. The indices i, j represent the unit vectors in *x* and *y* directions respectively, while n_j represents the direction of flow and it equals 1 in the vertical direction and 0 for the horizontal directions. In transport equation (2) the dimensionless relative concentration $c\ [-]$ is the transport variable which varies between 0 and 1. It is linearly related to fluid density ρ $[ML^{-3}]$ through $(\rho-\rho_f)/\rho_f = [(\rho_{\max}-\rho_f)/\rho_f]c$, under the assumption that the solute concentration of a fluid is $c_{\max}=1$ when $\rho=\rho_{\max}$. The term $(\rho-\rho_f)/\rho_f$ represents the dimensionless relative density ρ_r. $K_{ij}\ [LT^{-1}]$ are the coefficients of freshwater hydraulic conductivity tensor, $D_{ij}\ [L^2T^{-1}]$ are the coefficients of the dispersion tensor, $\phi\ [-]$ is porosity, $t\ [T]$ is time, $Q_\rho\ [L^3L^{-3}T^{-1}]$ is a volumetric fluid source/sink term per unit aquifer volume, Q_c $[ML^{-3}T^{-1}]$ is a solute mass source/sink term and $S_s\ [L^{-1}]$ is the specific storage. The Darcy flux term q_i is expressed for freshwater properties as:

$$q_i = -K_{ij}\left(\frac{\partial h_f}{\partial x_j}+\rho_r n_j\right) \tag{3}$$

Coastal aquifer application model

The numerical SWI simulations are based on a coastal aquifer model of rectangular shape (figure 1) which is an approximation of a real aquifer at the Greek Island of Kalymnos (Mantoglou *et al.* 2004).



[FIGURE 1]

The horizontal dimensions of the coastal aquifer model are $x = 7000m$, $y = 3000m$ and the aquifer base is at $z = -25m$ below sea-level. On the west side of the aquifer model a hydrostatic specified head boundary condition is applied along with a specified salinity concentration of $35\,Kg/m^3$ for a saltwater density of approximately $1025\,Kg/m^3$. The aquifer is replenished by both recharge and inland fluxes. The two lateral boundaries are no-flow boundaries while fully penetrating pumping wells extract groundwater from the coastal aquifer. A homogeneous and anisotropic coastal aquifer is assumed where the values of hydraulic conductivity are $K_x = K_y = 100\,m/day$ and $K_z = 10\,m/day$. The longitudinal dispersivity value was set to $100m$ and the transverse dispersivity value to $10m$. In the absence of field data and due to the exploratory features of this study, relatively large dispersivity values were selected to facilitate the setup of a faster VDST model since spatial discretization is related to dispersivity values (Werner *et al*. 2013). Note that for all the optimization problems described in the following sections, multiple independent optimization runs are performed in order to produce a statistical output due to the stochastic nature of the algorithms. In that sense, a relatively fast VDST model is required to realize such a demanding computational task for generic comparison purposes. 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.

Formulation of the pumping optimization problem

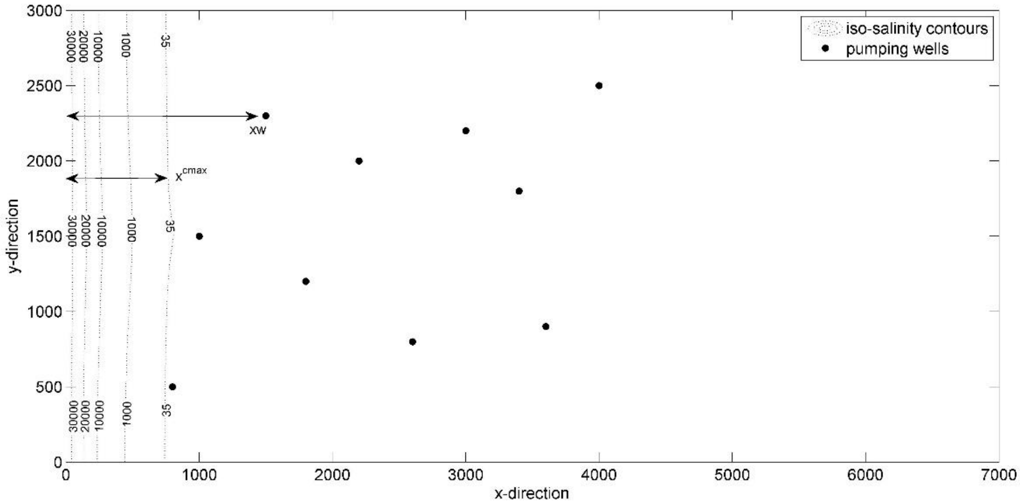
The pumping optimization problem of the present work lies in the category of non-linear constrained optimization problems described as follows:

$$\begin{aligned} &\min f(\xi) \\ &s.t. \quad g_i(\xi) \le 0,\ i = 1,2 \ldots M,\ l \le \xi \le u \end{aligned} \tag{4}$$

where f, g_i represent the objective function and inequality constraint functions respectively. The vector ξ takes values in the $N-$dimensional continuous space $[l,u] \subset R^N$. A real vector ξ^* is sought so that $f(\xi^*) = \min f(\xi)$, subject to the constraints defined in equation (4). It is assumed that the derivatives of f, g are not available while the bound constraints define the search space of the optimization problem. The corresponding single-objective pumping optimization problem can be mathematically described as (Mantoglou 2003; Mantoglou *et al.* 2004):

$$\begin{aligned} &\min -\sum_{i=1}^{M} Q_i \\ &s.t. \quad x_i^{c\max}(Q_1, Q_2, \ldots Q_M) \le xw_i,\ \forall i = 1,2,\ldots M \\ &\qquad Q_{\min} \le Q_i \le Q_{\max},\ i = 1,2,\ldots M \end{aligned} \tag{5}$$

where Q_i is the individual pumping rate of each pumping well and $x_i^{c\max}$ is the horizontal distance of the iso-salinity $c_{\max}$ from the coast, as a function of pumping rates from each pumping well. The variable xw_i refers to the pumping well location, while $Q_{\min}$ and $Q_{\max}$ define the lower and upper limits which pumping rates can take. The goal is to maximize (the reason for the negative sign in the objective function) the total groundwater extraction, subject to constraints which maintain the salinity levels in pumped groundwater at the specified limit of $c_{\max} = 35\, mg/lt$. Figure 2 illustrates a plan view of the simulated iso-salinity contours at the aquifer base, for a feasible vector Q of pumping rates.



[FIGURE 2]

The optimization problem in (5) can be translated to a bound-constrained optimization problem using penalty terms in the objective function. Thus, the objective function value is penalized every time that a constraint of the problem is violated. In this study we have applied the following objective function penalty formulation:

$$\min f\left(Q\right)=\begin{cases}-\sum_{i=1}^{M}Q_i,\ if\ \forall\, i=1,2...M;\, x_i^{c\max}\left(Q_1,Q_2,..Q_M\right)\leq xw_i \\ M_v\sum_{i=1}^{M}\left[\max\left(\left(x_i^{c\max}-xw_i\right),0\right)\right]^2, if\ \exists\, i=1,2...M;\, x_i^{c\max}\left(Q_1,Q_2,..Q_M\right)>xw_i\end{cases} \tag{6}$$

where M_v represents the number of pumping wells that the constraint is violated. The above formulation aims to attribute a separate score for each violated constraint while it involves the magnitude of violation through the squared difference between $x_i^{c\max}$ and xw_i. The penalized objective function is also multiplied by M_v to incorporate the number of constraint violations for a non-feasible vector Q. The pumping optimization problem defined above can be directly solved using the VDST model combined with a proper optimization algorithm. In pumping optimization of coastal aquifers, evolutionary algorithms tend to perform better than conventional gradient-based algorithms which might get trapped in local minima (Ketabchi & Ataie-Ashtiani 2015). However, evolutionary algorithms require a large number of function evaluations to converge and their performance may vary depending on the application

(Mantoglou & Papantoniou 2008; Karpouzos & Katsifarakis 2013; Ketabchi & Ataie-Ashtiani, 2015). Therefore, the direct solution of pumping optimization problems using VDST models and evolutionary algorithms may result in excessive computational burden.

In this study, a heuristic optimization method, namely, the evolutionary annealing-simplex (EAS) algorithm (Efstratiadis & Koutsoyiannis, 2002), is utilized to solve the penalized formulation of the optimization problem defined in (6). EAS algorithm employs the concepts of evolutionary search, the downhill simplex scheme and simulated annealing (Rozos *et al.* 2004). It has shown a robust performance for various pumping optimization problems of coastal aquifers (Kourakos & Mantoglou, 2009, Christelis & Mantoglou 2016a, Christelis & Mantoglou 2016b). Thereinafter, the direct optimization approach with the SWI model will be referred as VDST-EAS.

The surrogate model

The VDST-EAS approach may considerably increase the required computational effort to get an optimal solution. In some cases, the VDST simulations can be very expensive so that only a small number of them can be utilized to estimate a feasible solution in reasonable computational times (e.g. Christelis & Mantoglou 2016b). In this section, surrogate models are proposed as an alternative method for attaining an improved optimal solution based on a specified number of runs with the VDST model.

In the pumping optimization problem described in (5), the objective function is just a linear function of the decision variables $Q_1, Q_2, ..., Q_M$ which are the pumping rates, while the constraint functions are computationally expensive to evaluate. There are a variety of surrogate modelling techniques that can be used to approximate the constraints, including Kriging, RBF and Support Vector Machines (SVM). This paper employs a cubic RBF model augmented with a linear polynomial tail, in order to build a surrogate model for each of the M inequality constraint functions $x_i^{c\max}\left(Q_1, Q_2, ..., Q_M\right) \leq xw_i, i = 1, ..., M$. This type of surrogate was chosen because of its prior success when used with some SBO algorithms for constrained black-box optimization (e.g. Regis 2011; Regis 2014).

For convenience, denote the decision vector of pumping rates by $Q = (Q_1, Q_2, ..., Q_M)$ and the objective function by $f\left(Q\right) = -\sum_{i=1}^{M} Q_i$, and rewrite each inequality constraint function in the form $g_i(Q) \leq 0$, where $g_i\left(Q\right) = x_i^{c\max}\left(Q\right) - xw_i$. Now, given the vectors

$Q^{(1)}, Q^{(2)}, ..., Q^{(m)} \in R^M$ (which are simply referred to as points) where the constraint functions have been evaluated (i.e. so that the values $g_i\left(Q^{(1)}\right), g_i\left(Q^{(2)}\right), ...g_i\left(Q^{(m)}\right)$ are known for all $i = 1, ..., M$), this paper uses an RBF model of the form (Powell, 1992):

$$S_m(Q) = \sum_{k=1}^{m} \lambda_k \phi\left(\left\|\left(Q - Q^{(k)}\right\|\right)\right) + p(Q) \tag{7}$$

for each of the M inequality constraints. Here, $\phi(r) = r^3$ (the cubic form), $\lambda_1, ..., \lambda_m \in R$ are coefficients to be determined, and $p(Q)$ is a linear polynomial whose coefficients also need to be determined. Training the above RBF surrogate model for a constraint function means obtaining suitable values for the coefficients of the RBF part and the polynomial part so that the error between the constraint function and the RBF model at the training points $Q^{(1)}, Q^{(2)}, ..., Q^{(m)}$, is minimized. For the particular RBF model and training method used in this paper, the training error will always be zero, which means that the resulting RBF model passes through all the data points, that is, the surrogate model is an exact emulator. To obtain the coefficients in the above cubic RBF model for the *ith* constraint function g_i, define the matrix $\Phi \in R^{M x M}$ where $\Phi_{k,l} = \phi\left(\left\| Q^{(k)} - Q^{(l)} \right\|\right)$ and the matrix $P \in R^{m x (M+1)}$ whose *ith* row is $\left[1, \left(Q^{(i)}\right)^T\right]$. Moreover, define the vector $G_i = \left[g_i\left(Q^{(1)}\right), g_i\left(Q^{(2)}\right), ...g_i\left(Q^{(m)}\right)\right]^T$. Now, the vector of coefficients $\lambda = [\lambda_1, ..., \lambda_m]^T$ for the RBF part and the coefficients $c = [c_0, c_1, ...c_M]^T$ for the polynomial part are obtained by solving the following system of linear equations:

$$\begin{pmatrix} \Phi & P \\ P^T & 0 \end{pmatrix} \begin{pmatrix} \lambda \\ c \end{pmatrix} = \begin{pmatrix} G_i \\ 0 \end{pmatrix} \tag{8}$$

Under some simple conditions on the training points, namely that the matrix P has full column rank, the interpolation matrix in the above system is guaranteed to be invertible (Powell 1992). Since the above system can be solved quickly and efficiently, even when M is large, the training time for the cubic RBF model is negligible in comparison to the simulation time needed to generate the constraint function values. In all, the computational benefits from the negligible training time of cubic RBF models and their exact interpolation characteristics appear attractive for deterministic pumping optimization problems of large dimensionalities.

SBO optimization using a prediction-based infill strategy

Adaptive SBO methods have successfully applied in problems of pumping optimization of coastal aquifers. Those approaches managed to reduce the number of input-output patterns required from the surrogate model to provide reasonable approximations of the VDST model during optimization (Sreekanth & Datta 2015). Recently, Christelis and Mantoglou (2016a) applied cubic RBF surrogate models for a pumping optimization problem of coastal aquifers, which involved ten pumping wells and ten corresponding constraint functions for each pumping well. In their work, an online training scheme of the RBF models was embedded within the EAS algorithm. Their approach was to add infill points to the initial sampling plan by using the current best solutions found by the RBF model during the optimization operations. This infill strategy favours a fast improvement of the RBF model at the region of the current optimum (local exploitation). However, it neglects the global improvement of the surrogate model and might fail to identify the region of the global optimum (Forrester *et al.* 2008). In that study, the above approach reduced by 96% the corresponding computational time with the VDST-EAS approach while it successfully located the region of the global optimum. We apply the same method here, in order to test its performance as a basic SBO strategy and evaluate its performance for problems of larger dimensions and under limited computational budgets. The steps of the method, denoted hereinafter as RBF-EAS, are briefly presented below since the details have been presented in Christelis & Mantoglou (2016a):

1. Use a Latin Hypercube Sampling method to produce the initial population for the EAS algorithm and evaluate the VDST model at these points.
2. Store the initial sampling plan of the evaluation points $Q^{(1)}, Q^{(2)}, ..., Q^{(m)}$, along with the responses of the VDST model for the constraint functions $g_i(Q), i = 1, ..., M$ and train the RBF surrogate models.
3. Run EAS algorithm based on the RBF models and if a new optimum is found, use the VDST model to evaluate the current best solution Q. Add the new input-output data to the initial sampling plan, and re-train the RBF models.
4. Is the computational budget exhausted? If yes, return final solution, else go to step 3.

The ConstrLMSRBF algorithm

The ConstrLMSRBF algorithm (Regis 2011) is an SBO algorithm for constrained black-box optimization that uses the RBF interpolation model described previously, to approximate the black-box objective and inequality constraint functions. In the case of the pumping optimization problem in (5), only the constraint functions are computationally expensive to evaluate. However, in the standard implementation of ConstrLMSRBF, the algorithm also maintains an RBF surrogate model for the objective function. In this case though, since the objective function is linear in the decision variables (the pumping rates), one can mathematically prove that the resulting surrogate will also be linear and will be identical to the objective function, provided there are at least $M+1$ training points.

ConstrLMSRBF begins by evaluating the objective and constraint functions at a feasible starting point and at the points of a space-filling design, specifically a Latin hypercube design (LHD) with $2M+1$ points, over the region defined by the bound constraints of the problem $\left[Q_{\min}, Q_{\max}\right]$. Together, the feasible starting point and the LHD points constitute the initial training points. The space-filling design points possibly include infeasible points, and for the version of ConstrLMSRBF used in this paper, the first initial point must be feasible. The requirement of having a feasible point is not unreasonable since in many applications a feasible solution is often available or easy to obtain, as is the case for the above pumping optimization problem, and the practitioner is simply looking to improve this feasible solution. However, an extension of ConstrLMSRBF in Regis (2014) allows all initial points to be infeasible.

After evaluating the objective and inequality constraint functions at the initial points, 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 standard deviations that are allowed to vary adaptively depending on performance, to facilitate either local search or global search. When generating a candidate point, the choice of which coordinates of the current best point are perturbed is random, and is controlled by a parameter p_{select} which is the probability that a given coordinate is perturbed. In the numerical experiments, p_{select} equals 0.5 or 1. Next, the algorithm gathers the candidate points that are predicted to be feasible or that have the minimum number of predicted constraint violations. These points will be referred to as the valid candidate points. The next point where the simulation will be run (or where objective and constraint functions will be evaluated) 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. More precisely, for each valid candidate point Q, the algorithm calculates a score for the RBF criterion, $V_{RBF}\left(Q\right)$, and a score for the distance criterion, $V_{DIST}(Q)$. These scores vary from 0 to 1, with the preferred candidate points having scores closer to zero. Then, the next point where the simulation will take place is the valid candidate point Q that minimizes the value of:

$$V\left(Q\right) = w_{RBF}V_{RBF}\left(Q\right) + w_{DIST}V_{DIST}\left(Q\right) \tag{9}$$

where w_{RBF} and w_{DIST} are the weights for the two criteria and they satisfy $w_{RBF} + w_{DIST} = 1$. In the numerical experiments, these weights were fixed to $w_{RBF} = 0.95$ and $w_{DIST} = 0.05$ to put more emphasis on the RBF criterion.

Once the VDST simulation has taken place at the selected valid candidate point, the algorithm re-trains the RBF surrogate model with the new data point. Then it goes back to generating a new set of random candidate points and continues in the same manner as before until the computational budget is exhausted (e.g. the maximum number of VDST simulations has been reached). More details on ConstrLMSRBF can be found in Regis (2011).

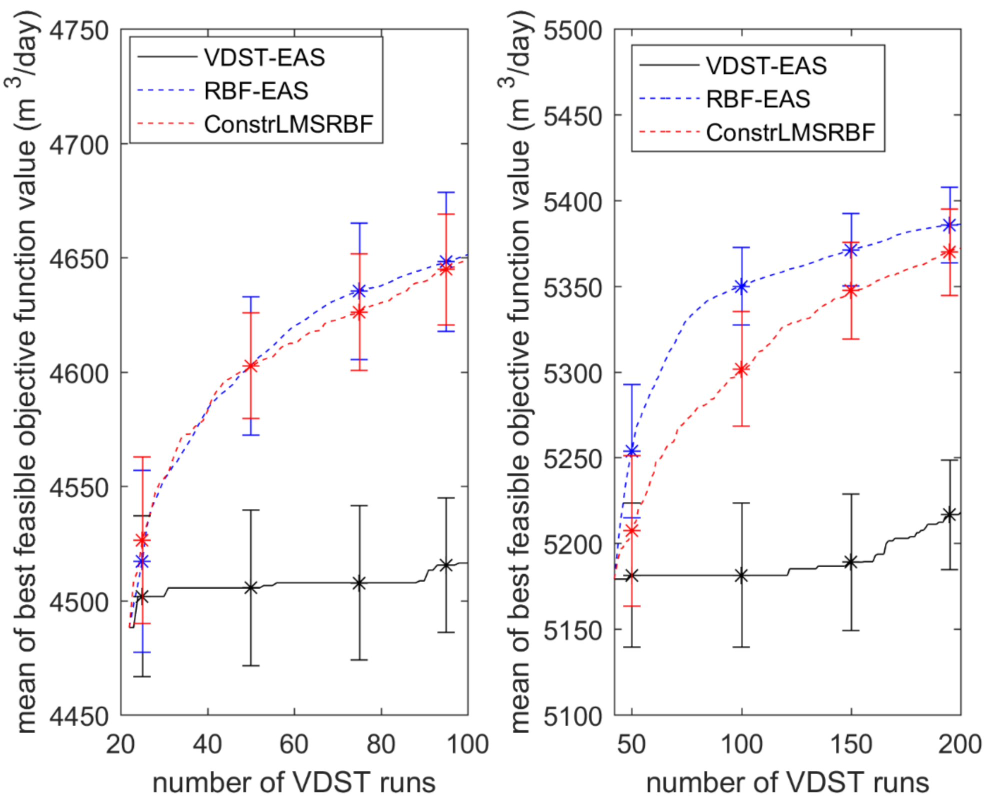
Problem settings

Four pumping optimization problems of different dimensionality were solved to test the performance of the algorithms described above. That is, $M = 10$, $M = 20$, $M = 30$ and $M = 40$. For each increase in the number of pumping wells the total recharge of the coastal aquifer model was also modified accordingly. This facilitated the comparison on the performance of the algorithms by moving the region of the global optimum in a different location. Therefore, for $M = 10$ the total recharge was set to $5409.86\,m^3/day$, for $M = 20$ the total recharge was set to $6159.8\,m^3/day$, for $M = 30$ the total recharge was set to $6909.8\,m^3/day$ and for $M = 40$ the total recharge was set to $7659.8\,m^3/day$. For each optimization problem (due to the different total recharge rates) an initial VDST model run was performed with no pumping present, until the head and salinity concentration fields reached steady-state. These were used as the initial conditions for the subsequent VDST simulations during the optimization task.

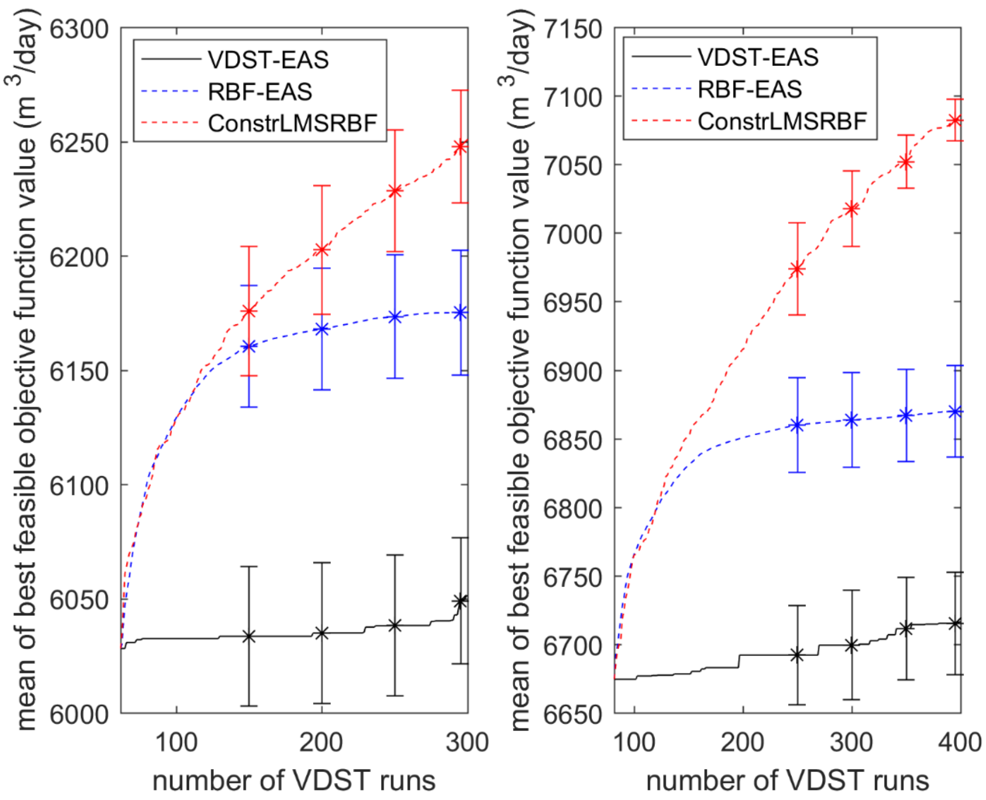
Each optimization problem was solved based on a specified budget of VDST simulations. The maximum allowed number of VDST model runs was set to $100 \times M$. Since the optimization methods of this study are based on stochastic operators, a set of 30 independent optimization runs is used for each approach in order to perform an adequate statistical comparison. In addition, for each independent optimization run a new initial population is generated which is applied to all the optimization methods to ensure same starting conditions.

RESULTS AND DISCUSSION

Figures 3 and 4 present the performance of the optimization methods based on their best average feasible objective function value among the 30 independent runs. The problems considering 10 and 20 pumping wells are considered of moderate dimensionality and are grouped together. The problems with 30 and 40 pumping wells are considered as of larger dimensionality and are also grouped together.



[FIGURE 3]



[FIGURE 4]

Results demonstrate that the SBO methods outperform the direct VDST-EAS optimization for all test problems. The SBO methods were able to improve the objective function value given the available number of runs with the VDST model. The more global search capabilities of ConstrLMSRBF against the predictive-based infill strategy of EAS-RBF algorithm are also demonstrated, particularly for the higher dimensional problems (figure 4). In both SBO frameworks, the objective function value exhibits a rapid improvement after the initial population evaluation comparative to VDST-EAS. However, in problems where $M = 30$ and $M = 40$, RBF-EAS appears to stall as the computational budget is exhausted. On the other hand, ConstrLMSRBF displays a continuous improvement of the average objective function value as the number of VDST runs are increased for all problems. A one-way analysis of variance (ANOVA) test was also performed on the above samples using the built-in MATLAB functions *anova1* and *multcompare* (Statistics and Machine Learning Toolbox, 2016b). The results are shown in the following table.

[TABLE 1]

Optimization frameworks		p-value			
		M=10	**M=20**	**M=30**	**M=40**
VDST-EAS	RBF-EAS	3.365-07	1.053-09	4.088-08	6.424-08
VDST-EAS	ConstrLMSRBF	5.158-07	2.812-09	9.561-10	9.560-10
RBF-EAS	ConstrLMSRBF	0.994	0.798	0.0021	9.569-10

It is demonstrated that the p-values between VDST-EAS and the two SBO strategies are close to zero for all optimization problems which confirms that the difference in their sample mean values is statistically significant. Furthermore, the comparison between RBF-EAS and ConstrLMSRBF shows that the sample means of the two methods for the 30 and the 40 decision variable problems are also significantly different.

CONCLUSIONS

A single-objective pumping optimization problem of coastal aquifer was solved using both direct and surrogate-based optimization methods. The direct optimization (VDST-EAS) involved the combination of a variable density and salt transport numerical model with an evolutionary algorithm. The two SBO methods were applied by utilizing the same surrogate models, namely, cubic RBF models. However, they were based on different update strategies for the surrogate model. The first (RBF-EAS) employed a classic prediction-based infill strategy (local exploitation) embedded in the same evolutionary algorithm with the direct optimization framework. The second (ConstrLMSRBF) was based on a comprehensive infill strategy which aims at both local exploitation and global exploration of the decision variable space.

To the best of our knowledge, this is the first time in coastal aquifer management that optimization problems of moderate and large dimensionalities are employed and compared for both direct and SBO methods. It is also the first time that a comprehensive generic SBO method (ConstrLMSRBF algorithm) is tested for single-objective pumping optimization problems of coastal aquifers. Results demonstrated an outperformance of the SBO methods against the direct optimization for the case of four different optimization problems with increased dimensionality (from 10 to 40 pumping wells). In particular, ConstrLMSRBF algorithm is considered a promising SBO method for coastal aquifer management since located the best solutions under limited computational budgets and demonstrated a robust performance for all optimization problems. The ANOVA tests confirmed the statistical significance of the differences in the sample means between the direct optimization and the SBO methods. Furthermore, the simple and fast implementation of cubic RBF surrogate models, in both SBO approaches, facilitated the individual treatment of a large number of constraint functions (up to 40) in negligible computational cost.

REFERRENCES

Asher, M. J., Croke, B. F. W., Jakeman, A.J. & Peeters, L. J. M. 2015 A review of surrogate models and their application to groundwater modelling. *Water Resour. Res.* **51**(8), 5957-5973.

Ataie-Ashtiani, B., Ketabchi, H. & Rajabi, M. M. 2014 Optimal management of a freshwater lens in a small island using surrogate models and evolutionary algorithms. *J. Hydrolog. Eng.* **19**, 339-354.

Bhattacharjya, R. K. & Datta, B. 2005 Optimal management of coastal aquifers using linked simulation optimization approach. *Water Resour. Manage.* **19**(3), 295-320.

Christelis, V. & Mantoglou, A. 2016a Pumping optimization of coastal aquifers assisted by adaptive metamodelling methods and radial basis functions. *Water Resour. Manage.* Doi:10.1007/s11269-016-1337-3.

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.

Dhar, A. & Datta, B. 2009 Saltwater intrusion management of coastal aquifers, I: linked simulation-optimization. *J. Hydrolog. Eng.* **14**, 1263–1272.

Efstratiadis, A. & Koutsoyiannis, D. 2002 An evolutionary annealing-simplex algorithm for global optimization of water resource systems. *Proceedings of the Fifth International Conference on Hydroinformatics*, Cardiff, UK, International Water Association Publishing **2**, 1423-1428.

Forrester, A. I. J., Sóbester, A. & Keane, A. J. 2008 *Engineering design via surrogate modelling-A practical guide*. Wiley, New York.

Forrester, A. I. J. & Keane, A. J. 2009 Recent advances in surrogate-based optimization. *Prog. Aerosp. Sci.* **49**, 50-79.

Giambastiani, B. M., Antonellini, M., Essink, G. H. O. & Stuurman, R. J. 2007 Saltwater intrusion in the unconfined coastal aquifer of Ravenna (Italy): a numerical model. *J. Hydrol*. **340**(1), 91-104.

Gingerich, S. B. & Voss, C. I. 2005 Three-dimensional variable-density flow simulation of a coastal aquifer in southern Oahu, Hawaii, USA. *Hydrogeol. J.* **13**(2), 436-450.

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.

Hussain, M. S., Javadi, A. A., Ahangar-Asr, A. & Farmani, R. 2015 A surrogate model for simulation–optimization of aquifer systems subjected to seawater intrusion. *J. Hydrol*. **523**, 542-554.

Jones, D., Schonlau, M. & Welch, W. 1998 Efficient global optimization of expensive black-box functions. *J. Global Optim.* **13**, 455-492.

Karpouzos, D. K. & Katsifarakis, K. L. 2013 A set of new benchmark optimization problems for water resources management. *Water Resour. Manage.* **27**(9), 3333-3348.

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.

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.

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

Kopsiaftis, G., Mantoglou, A. & Giannoulopoulos, P. 2009 Variable density coastal aquifer models with application to an aquifer on Thira Island. *Desalination* **237**(1), 65-80.

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.

MATLAB, Statistics and Machine Learning Toolbox, Release 2016b. The MathWorks, Inc., Natick, Massachusetts, United States.

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.

Mantoglou, A. & Papantoniou, M. 2008 Optimal design of pumping networks in coastal aquifers using sharp interface models. *J. Hydrol*. **361**, 52–63.

Mugunthan, P., Shoemaker, C. A., & Regis, R. G. 2005. Comparison of function approximation, heuristic and derivative-based methods for automatic calibration of computationally expensive groundwater bioremediation models. *Water Resour. Res.* Doi:10.1029/2005WR004134.

Papadopoulou, M. P., Nikolos, I. K., & Karatzas, G. P. 2010. Computational benefits using artificial intelligent methodologies for the solution of an environmental design problem: saltwater intrusion. *Water Sci. Technol.* **62**(7), 1479-1490.

Parr, J. M., Keane, A. J., Forrester, A. I., & Holden, C. M. E. 2012. Infill sampling criteria for surrogate-based optimization with constraint handling. Eng. Optimiz. **44**(10), 1147-1166.

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**(5).

Powell, M. J. D. 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.

Rao, S. V. N., Sreenivasulu, V., Bhallamudi, S. M., Thandaveswara, B. S. & Sudheer, K.P. 2004 Planning groundwater development in coastal aquifers. *Hydrol. Sci. J.* **49**(1).

Razavi, S., Tolson, B. A. & Burn, D. H. 2012a Review of surrogate modelling in water resources. *Water Resour. Res*. **48**(7).

Razavi, S., Tolson, B. A. & Burn, D. H. 2012b Numerical assessment of metamodelling strategies in computationally intensive optimization. Environ. Modell. Softw., **34**, 67-86.

Regis, R. G. & Shoemaker, C. A. 2005 Constrained global optimization of expensive black box functions using radial basis functions. *J. Global Optim.* **31**, 153-171.

Regis, R. G. & Shoemaker, C. A. 2007 Improved strategies for radial basis function methods for global optimization. *J. Global Optim.* **37**, 113-135.

Regis, R. G. 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, R. G. & Shoemaker, C. A. 2013 A quasi-multistart framework for global optimization of expensive functions using response surface methods. *J. Global Optim.* **56**, 1719-1753.

Regis, R.G. 2014 Evolutionary programming for high-dimensional constrained expensive black-box optimization using radial basis functions. *IEEE T. Evolut. Comput.* **18**(3).

Regis, R. G. & Shoemaker, C.A. 2013 Combining radial basis function surrogates and dynamic coordinate search in high-dimensional expensive black-box optimization. *Eng. Optimiz.* **45**(5), 529-555.

Rozos, E., Efstratiadis, A., Nalbantis, I. & Koutsoyiannis, D. 2004 Calibration of a semi-distributed model for conjunctive simulation of surface and groundwater flows. *Hydrol. Sci. J.* **49**(5), 819-842.

Roy, T., Schütze, N., Grundmann, J., Brettschneider, M. & Jain, A. 2016 Optimal groundwater management using state-space models: a case study for an arid coastal region. *J. Hydroinf.* **18**(4), 666-686.

Solomatine, D. P. & Ostfeld, A. 2008 Data-driven modelling: some past experiences and new approaches. *J. Hydroinf.* **10** (1), 3-22.

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.* **23**(6), 1155-1166.

Therrien, R. & Sudicky, E. A. 1996 Three-dimensional analysis of variably-saturated flow and solute transport in discretely-fractured porous media. *J. Contam. Hydrol.* **23**, 1-44.

Therrien, R., McLaren, R. G., Sudicky, E. A. & Panday, S. M. 2006 *HydroGeoSphere-A three-dimensional numerical model describing fully-integrated subsurface and surface flow and solute transport. Groundwater Simulations Group*, University of Waterloo, Canada, 2006.

Thompson, C., Smith, L. & Maji, R, 2007 Hydrogeological modelling of submarine groundwater discharge on the continental shelf of Louisiana. *J. Geophys. Res.* Doi: 10.1029/2006JC003557.

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.

Voss, C. I. & Souza, W. R. 1987 Variable density flow and solute transport simulation of regional aquifers containing a narrow freshwater- saltwater transition zone. *Water Resour. Res.* **23**(10), 1851-1866.

Werner, A. D., Bakker, M., Post, V. E. A., Vandenbohede, A, Lu, C., Ataie-Ashtiani, B., Simmons, C. T. & Barry D. A. 2013 Seawater intrusion processes, investigation and management: Recent advances and future challenges. *Adv. Water Resour*. **51**, 3-26.