frontiers | Frontiers in Water

TYPE Original Research
PUBLISHED 30 June 2023
DOI 10.3389/frwa.2023.1195029

OPEN ACCESS

EDITED BY
Thomas Graf,
Leibniz University Hannover, Germany

REVIEWED BY
Marwan Fahs,
National School for Water and Environmental Engineering, France
Salim Heddam,
University of Skikda, Algeria

*CORRESPONDENCE
George Kopsiaftis
✉ gkopsiaf@survey.ntua.gr

RECEIVED 27 March 2023
ACCEPTED 22 May 2023
PUBLISHED 30 June 2023

CITATION
Kopsiaftis G, Kaselimi M, Protopapadakis E, Voulodimos A, Doulamis A, Doulamis N and Mantoglou A (2023) Performance comparison of physics-based and machine learning assisted multi-fidelity methods for the management of coastal aquifer systems.
Front. Water 5:1195029.
doi: 10.3389/frwa.2023.1195029

Performance comparison of physics-based and machine learning assisted multi-fidelity methods for the management of coastal aquifer systems

George Kopsiaftis[1]*, Maria Kaselimi[2], Eftychios Protopapadakis[3], Athanasios Voulodimos[4], Anastasios Doulamis[2], Nikolaos Doulamis[2] and Aristotelis Mantoglou[1]

[1]Laboratory of Reclamation Works and Water Resources Management, School of Rural, Surveying and Geoinformatics Engineering, National Technical University of Athens, Athens, Greece, [2]Laboratory of Photogrammetry, School of Rural, Surveying and Geoinformatics Engineering, National Technical University of Athens, Athens, Greece, [3]School of Information Sciences, Department of Applied Informatics, University of Macedonia, Thessaloniki, Greece, [4]Artificial Intelligence and Learning Systems Laboratory, School of Electrical and Computer Engineering, National Technical University of Athens, Athens, Greece

In this work we investigate the performance of various lower-fidelity models of seawater intrusion in coastal aquifer management problems. The variable density model is considered as the high-fidelity model and a pumping optimization framework is applied on a hypothetical coastal aquifer system in order to calculate the optimal pumping rates which are used as a benchmark for the lower-fidelity approaches. The examined lower-fidelity models could be classified in two categories: (1) physics-based models, which include several widely used variations of the sharp-interface approximation and (2) machine learning assisted models, which aim to improve the efficiency of the SI approach. The Random Forest method was utilized to create a spatially adaptive correction factor for the original sharp-interface model, which improves its accuracy without compromising its efficiency as a lower-fidelity model. Both the original sharp-interface and Machine Learning assisted model are then tested in a single-fidelity optimization method. The optimal pumping rated which were calculated using the Machine Learning based SI model sufficiently approximate the solution from the variable density model. The Machine Learning assisted approximation seems to be a promising surrogate for the high-fidelity, variable density model and could be utilized in multi-fidelity groundwater management frameworks.

KEYWORDS

variable density, sharp interface, machine learning, coastal aquifer, pumping optimization, random forests

1. Introduction

Seawater intrusion (SWI) models based on density-dependent approach constitute an accurate and realistic emulation of the saltwater/freshwater interaction and movement in coastal aquifers, since they incorporate several components of the physical processes, such as the dispersion mechanism (Simmons, 2005; Younes et al., 2009, 2022; Dokou and Karatzas, 2012; Werner et al., 2013; Karatzas and Dokou, 2015; Kourgialas et al., 2015; Fahs et al., 2022). However, variable density (VD) models are computationally

demanding and entail considerably long runtimes, as the mathematical description of the model consists of a coupled partial differential equation system, which require the implementation of complex and time-consuming numerical methods (e.g., finite differences or finite element methods). The long simulation runtimes, usually lead to an unmanageable overall computational burden for several real-life applications that require a considerable number of iterations, such as simulation-optimization problems and inverse modeling in coastal aquifers, sensitivity and uncertainty analysis or the development of decision support systems in coastal water resources systems (Asher et al., 2015). A common approach to mitigate the excessive computational cost is the use of approximation models, which aim to simulate the same physical process considering only few components of the real physical system (Simpson et al., 2001; Forrester et al., 2008; Razavi et al., 2012). These low-complexity models—which are often called surrogate models, metamodels, lower-fidelity (LF) models, response surfaces etc.—are much quicker compared to the high-fidelity (HF) models, but they have lower fidelity, i.e., the degree of the realism of a simulation model (Razavi et al., 2012).

Approximation models are gradually gaining ground in many aspects of water resources, such as groundwater modeling. A thorough survey of the surrogate models in water resources is presented in the work of Razavi et al. (2012) and Asher et al. (2015). According to Razavi et al. (2012) there are two main categories of the surrogate models: (i) the statistical or empirical data-driven models, which emulate the HF model responses, and (ii) the physically-based surrogates, which are usually a simplified version of the original HF model.

On the broad category of the data-driven surrogate models, Machine Learning (ML) algorithms demonstrated good performance in reproducing the outcomes of HF models. One of the first to use an ML assisted metamodeling strategy for SWI was Bhattacharjya et al. (2007), who employed an Artificial Neural Network (ANN) to predict the concentration at specified observation locations at different times in a coastal aquifer. Trichakis et al. (2011) also used ANNs to model the groundwater levels in a karstified groundwater system. Roy and Datta (2017a) utilized the fuzzy C-means clustering algorithm to predict the extent of SWI, while Lal and Datta (2018) employed Support Vector Machines (SVM) as a ML regressor to predict the salinity concentration at specified monitoring wells and compared the results with a genetic programming (GP) based surrogate model. Kopsiaftis et al. (2019b) investigated several ML algorithms, such as Gaussian Process Regression (GPR), SVMs, RFs, and ANNs, regarding their ability to estimate the location of a critical isohaline.

In a number of studies, the data-driven surrogate models were exploited in groundwater management strategies. For example, Kourakos and Mantoglou (2009) coupled a simulation-optimization framework with an ANN model to reduce the computational burden arising from the VD simulations. In the same context, Kourakos and Mantoglou (2013) integrated modular neural networks into a multi-objective management scheme, whereas Ataie-Ashtiani et al. (2014) developed an efficient, ANN-based multi-objective system for groundwater management in freshwater lenses of small islands. Christelis and Mantoglou (2016b) used cubic radial basis functions (RBFs) in an adaptive-recursive metamodeling framework which provided optimal pumping rates which approximated the global optimum. Christelis et al. (2018) emphasized on the joint use of HF and LF models to improve the efficiency of a surrogate-based optimization framework under limited computational budgets and for problems of different dimensionality.

In some recent studies, researchers investigate the efficiency of the weighted combination of several surrogate models in the form of ensemble models. Sreekanth and Datta (2011) proposed an ensemble model based on the genetic programming algorithm. Roy and Datta (2017b,c) examined ensemble versions of multivariate adaptive regression spline and adaptive neuro-fuzzy inference systems respectively, in coastal aquifer pumping optimization problems to control the encroachment of SWI. Kopsiaftis et al. (2019a) utilized three well established physically-based surrogate models in groundwater management problems following a tow-step approach—first they solved a pumping optimization problem with each model separately, and then with all the possible combinations. The results indicated that in most cases the ensemble model predictions outperformed the individual surrogate models. Christelis and Mantoglou (2016a, 2019) introduced a variable-fidelity approach in pumping optimization of coastal aquifers to leverage the LF model features and specifically the reduced simulation runtimes.

In this paper we propose a method that combines a physical-based surrogate model and a data driven model, which—to the best of our knowledge—has not been used in coastal aquifer management problems. The core idea of the method lies on the correction of existing approaches, which, as mentioned before, are widely-used and of significant usefulness in seawater intrusion problems. In the following section of the paper, we initially investigate well-established, physical-based surrogate models of SWI, regarding their efficiency in coastal aquifer management problems. The preliminary results indicate that the use of single-fidelity models may have several limitations in capturing the extent of SWI for a wide range of coastal aquifer parameter values (e.g., pumping rates and recharge). To address this problem, the combined use of models of various fidelities is investigated in this paper, which could enhance their adaptability in different aquifer geometries and flow conditions. To this end, a data-driven model—specifically the Random Forest (RF) method—is utilized to increase the fidelity of a widely used physically-based SWI model. It is worth mentioning that the RF algorithms, as implemented in this study, is univariate and that Bayesian optimization is employed in the surrogate model, to improve its efficiency. The proposed correction method is applied on a hypothetical unconfined coastal aquifer and the calculated SWI extent is compared to the results of three additional physically-based surrogate models, that is the original model before the correction and its two main modifications under moderate and intensive pumping.The examined aquifer is a typical coastal aquifer in arid and semi-arid Mediterranean islands, which spans an approximate area of 20 km^2. The ML assisted LF model outperformed the examined surrogate models in capturing the VD SWI extent. To test its usefulness in applications of practical interest, the proposed LF model is subsequently incorporated in a pumping optimization framework demonstrating a good performance, comparable to VD model.

The remainder of this work is structured as follows: Section 2 analyzes the mathematical formulation of all the examined SWI models, including the HF model, the previous well-established LF approximations, and finally the proposed RF assisted model, whereas Section 3 presents the optimization framework for a coastal aquifer management problem. Section 4 provides the experimental results, including the comparison of the proposed method with previous approached, and the implementation of the pumping optimization framework. Finally, Section 5 concludes the paper.

2. Seawater intrusion models and application aquifer

Several mathematical models of different complexity have been proposed to describe SWI in coastal aquifers. As already mentioned in Section 1, in the present study we employ models of two fidelity options to simulate SWI: (i) a 3D VD model representing the HF case, and (ii) a 2D SI model as the LF case. The LF model does not consider the dispersion mechanism—is the driving force of seawater/freshwater mixing—leading to a model of reduced complexity. In the following sections we briefly present the mathematical formulation of all the SWI models.

2.1. Variable density model

VD models are based on the spatial variability of groundwater density, which ranges from saline water density to freshwater density. The density changes because of the mixing of freshwater and seawater, which results in the existence of a transition zone across the entire coastline. The width and exact position of the zone depends on the aquifer parameters and the pumping regime. In the current paper, thermal and viscosity effects are neglected and the density changes are attributed only to concentration effect. The mathematical formulation of the VD model has been thoroughly analyzed in several works (e.g., Frind, 1982; Kolditz et al., 1998; Younes et al., 2009; Pool and Carrera, 2011; Kourgialas et al., 2015; Fahs et al., 2022). In this study thesis the notation proposed by (Guo and Langevin, 2002; Langevin et al., 2008) is followed in order to be consistent with the SEAWAT numerical code, which is utilized for the VD simulations.

In general, the flow and solute-transport equations are used to describe mathematically the VD model. The two equations form a coupled differential equation system, which could be expressed as follows

$$-\nabla\cdot(\rho\mathbf{q})+\rho_s q_s=\rho S_f\frac{\partial h_f}{\partial t}+n\frac{\partial\rho}{\partial C}\frac{\partial C}{\partial t} \tag{1}$$

$$-\nabla\cdot(\mathbf{D}\cdot\nabla C)-\nabla\cdot(\mathbf{v}C)-\frac{q_s}{n}C_s=\frac{\partial C}{\partial t} \tag{2}$$

where ρ is fluid density, $\mathbf{q}$ is the Darcy velocity vector, ρ_s is the density of water entering from a source of leaving through a sink, q_s is the volumetric flow rate per unit volume of porous medium representing sources and sinks, S_f is the specific storage, h_f is the freshwater head, n is the porosity, C is the solute concentration, $\mathbf{D}$ is the hydrodynamic dispersion tensor, $\mathbf{v}$ is the fluid velocity vector, and C_s is the solute concentration of water entering or leaving through sources and sinks respectively. Since solute reaction is not considered, fluid density is only a function of the solute concentration C, according to the following equation

$$\rho=\rho_o\left(1+\frac{\epsilon}{C_s-C_o}(C-C_o)\right) \tag{3}$$

in which ρ_o is the freshwater density, ϵ is the density difference ratio (Equation 3), C_o is reference concentration, and C_s is the maximum concentration. In this paper, the following values are used for the parameters of Equation (3): $\rho_o = 1.000\,\mathrm{kg/m^3}$, $C_o = 0\,\mathrm{kg/m^3}$, and $C_s = 35\,\mathrm{kg/m^3}$.

The density difference ratio is expressed as

$$\epsilon=\frac{\rho_s-\rho_f}{\rho_f} \tag{4}$$

where ρ_s stands for the maximum seawater density. In this study, we consider $\rho_s = 1025\,\mathrm{kg/m^3}$.

The Darcy flux term $\mathbf{q}$ of Equation (1) for constant viscosity and freshwater properties could be expressed as

$$\begin{aligned} q_x&=-K_{fx}\left(\frac{\partial h_f}{\partial x}\right)\\ q_y&=-K_{fy}\left(\frac{\partial h_f}{\partial y}\right)\\ q_z&=-K_{fz}\left(\frac{\partial h_f}{\partial z}+\frac{\rho-\rho_f}{\rho}\right) \end{aligned} \tag{5}$$

where q_x, q_y, and q_z are the components of the specific discharge in the principal directions, K_{fx}, K_{fy}, and K_{fz} are the components of the freshwater hydraulic conductivity in the same directions and ρ_f is the freshwater density.

Equations (1)–(5) are the mathematical representation of VD approach of seawater intrusion. The well-established SEAWAT code is used to solve numerically the above equation set. SEAWAT is a modular finite-difference computer code created by USGS, which couples MODFLOW and MT3DMS, in order to solve iteratively the fluid flow and solute transport equations (Guo and Langevin, 2002).

2.2. Sharp interface model

The SI model employed in the present work is based on the single flow potential formulation proposed by Strack (1976). The Strack's approach considers three basic assumptions: (1) the Dupuit-Forchheimer approximation, (2) Ghyben-Herzberg approximation, and (3) Hydrostatic conditions in the saltwater zone. For steady state conditions, the incorporation of the three assumptions result in a simplified one-fluid, 2D SWI model, which, for the general case of irregular-shaped aquifers with spatially variable parameters (e.g., hydraulic conductivity and recharge), could be solved numerically using a groundwater flow code

(Mantoglou et al., 2004). The differential equations used to calculate the flow potential in an unconfined coastal aquifer has the following form

$$\frac{\partial}{\partial x}\left(K\frac{\partial \phi}{\partial x}\right)+\frac{\partial}{\partial y}\left(K\frac{\partial \phi}{\partial y}\right)+N-Q(x,y)=0 \tag{6}$$

where $K = K(x,y)$ is the hydraulic conductivity of the aquifer, which is usually considered as a function of the spatial coordinates (x,y), N is the aquifer recharge, and $Q(x,y)$ represents the distributed well pumping rates. The flow potential ϕ_τ at the location of the toe in an unconfined aquifer is calculated from the following equations (Mantoglou, 2003).

$$\phi_\tau = \frac{(1+\epsilon)\epsilon}{2}d^2 \tag{7}$$

where d denotes the thickness of the aquifer from the horizontal bottom to the sea level. From the flow potential value field we are able to calculate the exact location of the sharp interface toe, which is of considerable interest in coastal aquifer management problems.

2.3. Sharp interface model modifications

The SI model proposed by Strack (1976) and adopted in several studies (e.g., Werner et al., 2012; Ketabchi and Ataie-Ashtiani, 2015) have been widely used to simulate SWI in coastal aquifers or pumping optimization problems in coastal aquifers (e.g., Mantoglou, 2003; Mantoglou et al., 2004). However, it has been observed in several previous studies (e.g., Dausman et al., 2010; Koussis et al., 2015; Kopsiaftis et al., 2019a) that for a specific set of pumping rates the SI model tends to overestimate the extent of the seawater wedge compared to the variable density approach, which is considered as a reference result for the estimation of seawater intrusion. Llopis-Albert and Pulido-Velazquez (2014) further examined the efficiency of the Ghyben-Herzberg based SI model in producing comparable results with the VD model, within a range in the values of critical hydrogeological parameters and defined several limitations in the validity of the SI model.

In order to eliminate the discrepancy between the two models, several correction methods have been recently proposed (e.g., Pool and Carrera, 2011; Koussis et al., 2012, 2015; Lu et al., 2016; Werner, 2017; Koussis and Mazi, 2018; Christelis et al., 2019), which aim to incorporate the effect of the dispersion mechanism and improve the accuracy of the SI approach. Pool and Carrera (2011) were the first to suggest a novel formula to extend Strack's equations to the case of the mixing zone. In particular, they proposed an multiplication factor for the saltwater-freshwater density ratio, which is expressed as an empirical equation of specific physical parameters of the aquifer. The multiplication factor has the following form

$$\epsilon^* = \epsilon\left[1-\left(\frac{a_T}{d}\right)^{1/6}\right] \tag{8}$$

where ϵ^* is the modified buoyancy factor or saltwater-freshwater density ratio and a_T is transverse dispersivity. This specific correction factor has been widely used in a number of studies (e.g., Lu et al., 2012; Christelis and Mantoglou, 2013; Lu and Werner, 2013; Lu and Luo, 2014; Koussis et al., 2015). Lu and Werner (2013) proposed a modification of the (Pool and Carrera, 2011) correction factor. Specifically, they performed a significant number of variable density simulations and suggested that the 1/6 exponent of Equation (8) should be replaced with 1/4. In a more recent study, Christelis and Mantoglou (2016a) observed that in highly pumped aquifers, the value of the density ratio that allows the SI to resemble the salinity spatial distribution depends not only on the aquifer physical parameters included in Equation (8), but also on the well pumping rates. In order to incorporate this effect, they suggested a dynamic adjustment of the density ratio during the pumping optimization process. In this case, the correction is based on the modification of the density ratio in an intermediate optimization step, which minimizes the discrepancy between the VD and SI models. Kopsiaftis et al. (2019a) compared the SWI extent based on five models: (i) the VD model, (ii) the Strack SI model, (iii) the Pool and Carrera (2011) correction, (iv) the Lu and Werner (2013), and (v) several combinations of the above SI models, in the form of ensemble SI model prediction. In the same study, the authors used all the examined models in a pumping optimization model and concluded that the ensemble solutions provide promising results. The ensemble SI prediction could be expressed as follows

$$\hat{Y}_{ens} = \sum_{i=1}^{P} w_i Y_i \tag{9}$$

where $\hat{Y}_{ens}$ is the ensemble prediction of the toe, P is the number of the individual SI models, w_i is the weight corresponding to Y_i toe prediction of the i-th SI model. In the present work, three SI models are examined for the ensemble prediction of the toe: (1) the original Strack SI model Strack (1976), (2) the modified Strack model proposed by Pool and Carrera (2011), and (3) the modified Strack model proposed by Lu and Werner (2013). The contribution of the three SI models is considered equal, thus the weights are $w_1 = w_2 = w_3 = 1/3$ and the $\hat{Y}_{ens}$ is calculated as the average value of the individual SI model predictions.

2.4. Multi-fidelity approach of seawater intrusion

In this work an implicit correction of the Strack (1976) SI model is investigated using the position of a specific isohaline, which represents the extent of SWI based on the variable density approach. In the multi-fidelity (MF) related literature, the proposed correction is denoted as additive correction (e.g., Alexandrov and Lewis, 2001; Zhou et al., 2016; Giselle Fernandez-Godino et al., 2019) and, in general, it could be described as an additive shift of the low-fidelity response $y_{LF}(\mathbf{x})$ to yield the high-fidelity response $y_{HF}(\mathbf{x})$. The correction has the following general mathematical form Giselle Fernandez-Godino et al. (2019).

$$\hat{y}_{MF}(\mathbf{x}) = \hat{y}_{LF}(\mathbf{x}) + \hat{\delta}(\mathbf{x}) \tag{10}$$

where $\hat{y}_{MF}(\mathbf{x})$ is the MF surrogate model, $\hat{\delta}(\mathbf{x})$ is the surrogate constructed to model the discrepancy between the $y_{HF}(\mathbf{x})$ and $y_{LF}(\mathbf{x})$ at data point where we have model

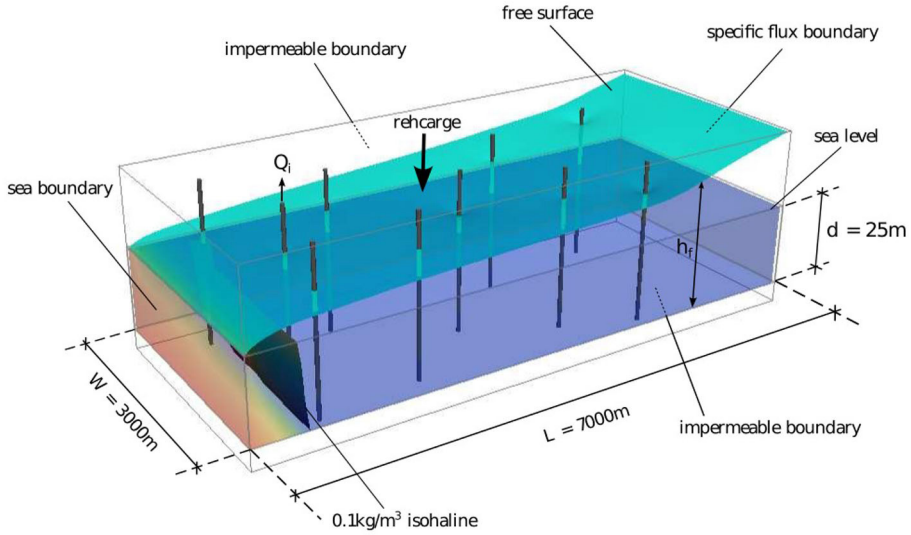

FIGURE 1
Conceptual model of the examined unconfined coastal aquifer, presenting the size of the aquifer, the boundary conditions, the location of the wells and an indicative location of the critical 0.1 kg/m^3 isohaline.

computations for both the HF and LF models. For the groundwater application examined in the present work, the input vector $\mathbf{x}$ corresponds to the vector $\mathbf{Q}$ of the pumping rates.

In this study, RF method was utilized as the discrepancy function $\hat{\delta}(\mathbf{x})$. RF algorithm is an ensemble learning technique for solving both classification and regression problems. RFs can be used to predict continuous target variable based on a set of input features. Each decision tree in the forest is built using a random subset of the available features, and the final output is obtained by averaging the predictions of all the trees. This results in a more stable and accurate model than using a single decision tree, as it reduces the risk of overfitting and the impact of noisy or irrelevant features. Ho (1998) was the first to embed a random subspace method in decision forest. Breiman (2001) further extended the algorithm and introduced the term "Random Forests". RFs have been widely used in several components of the water resources field under different context. For example, Papacharalampous et al. (2018, 2019) aimed at forecasting several hydrological processes and parameters (e.g., temperature and precipitation time-sires) and they compared the outcomes with stochastic methods. A popular implementation of RF refers to the predictive modeling of several groundwater pollution sources, which is crucial in areas with significant agricultural or urban growth, especially in arid and semi-arid climates (e.g., Wheeler et al., 2015; Canion et al., 2019; Knoll et al., 2019; Messier et al., 2019; Ouedraogo et al., 2019; Lahjouj et al., 2020; Pham et al., 2021; He et al., 2022). Other noteworthy applications of RF are the prediction of groundwater levels (e.g., Wang et al., 2018; Saha et al., 2022), assessment of drought impact on groundwater potential (Masroor et al., 2021), classification of surface-groundwater interaction (Yang et al., 2019), identification of potential sites for groundwater artificial recharge (Norouzi and Shahmohammadi-Kalalagh, 2019; Naghibi et al., 2020), investigation of the spatial and temporal variations of soil salinity (Fathizad et al., 2020) and modeling of the shallow water table at high spatial resolution to calculate the risk for groundwater-induced flood events (Koch et al., 2019).

It should be noted that in the present work, in order to capture the complex relation between the well pumping rates and the proposed additive correction factor, the 0.1 kg /m^3 isohaline and the corresponding SI toe are divided in several segments and a RF model is trained for the edge points of each segment. Thus, the proposed correction demonstrated good adaptive capabilities, which are further analyzed in Section 4.3.

3. Pumping optimization framework

In general, the objective of the groundwater management problem in the areas adjacent to the sea is the maximization of freshwater extraction while meeting several environmental criteria, which usually include the control of hydraulic head or the extent of SWI. In the present paper, we adopt a simulation-optimization approach to solve the groundwater management problem in coastal aquifers. The decision variables are the well pumping rates $Q_i, i = 1, 2, \ldots, n$, where n is the total number of wells. This approach is applied on both the VD and the SI models. However, the formulation slightly differs in the two SWI models, since each one of them has different output.

In the case of the VD model, the optimization problem has the following mathematical form (Kourakos and Mantoglou, 2013; Christelis and Mantoglou, 2016a; Christelis et al., 2019).

TABLE 1 Parameter values of the variable density model.

Parameter	Value
K_{fx}	15 m /d
K_{fy}	15 m /d
K_{fz}	1.5 m /d
Longitudinal dispersivity (a_L)	25 m
Transverse dispersivity (a_T)	2.5 m
Vertical dispersivity (a_V)	0.25 m
Recharge	5.479×10^{-5} m/d
Lateral inflow	600 m^3 /d
Density ratio	0.025
Seawater density	1025 kg /m^3
Freshwater density	1000 kg /m^3
Number of rows	60
Number of columns	140
Number of layers	5
Finite difference cell size (dx × dy)	50 m × 50 m

$$\begin{cases} \min & -\sum_{i=1}^{n} Q_i \\ \text{s.t.} & x_i^{C_i}(Q_1,\ldots,Q_n) \le xw_i, \quad \forall i = 1,\ldots,n \\ & h_i(Q_1,\ldots,Q_n) \ge 0, \quad \forall i = 1,\ldots,n \\ & Q_{\min} \le Q_i \le Q_{\max}, \quad i = 1,\ldots,n \end{cases} \tag{11}$$

where $x_i^{C_i}(Q_1,\ldots,Q_n)$ represents the distance of a critical isohaline—which is a function of the pumping rates and corresponds to a concentration value equal to C_i—from the coastline at the bottom of the aquifer. Here, the 0.1 kg/m^3 isohaline was selected as representative for SWI extent. According to the first constraint, SWI should not exceed the location of the i-th well denoted by xw_i. Likewise, the hydraulic head in the i-th well should not drop below the sea level. Finally, $Q_{\min}$ and $Q_{\max}$ defines the minimum and maximum allowed pumping rates for each well, respectively.

In the case of SI model, the extent of SWI is defined by the location of the equipotential line ϕ_τ which corresponds to the intersection of the SI with the bottom of the aquifer (toe of the SWI wedge). The value ϕ_τ of the toe is calculated by Equation (7) and its location is specified as a contour line of the potential flow field, which results from the solution of Equation (6). The mathematical expression of the optimization framework is similar to the one in Equation (11), with their main difference being that $x_i^{C_i}(Q_1,\ldots,Q_n)$ is substituted with $x_i^{\phi_\tau}(Q_1,\ldots,Q_n)$, which represents the distance of the toe from the coastline and it is a function of the well pumping rates and the hydraulic head $h_i(Q_1,\ldots,Q_n)$ with $\phi_i(Q_1,\ldots,Q_n)$, which denotes potential flow at the well locations. Similarly to the VD case, the constraints in the SI optimization framework prevent the toe from reaching the well locations and ensure that the potential flow remains above the sea level.

To solve the above optimization problems, we employed the evolutionary annealing-simplex (EAS) algorithm Efstratiadis and Koutsoyiannis (2002), a heuristic optimization algorithm that proved to be quite efficient in coastal aquifer management (Kourakos and Mantoglou, 2009; Christelis and Mantoglou, 2016b; Christelis and Hughes, 2018; Kopsiaftis et al., 2019a). To apply the EAS algorithm, the constraints in Equation (11) were incorporated in the objective function in the form of penalty terms (Christelis et al., 2018).

$$\min f(\mathbf{Q}) = \begin{cases} -\sum_{i=1}^{n} Q_i, \text{ if } \forall j;\ \left(g^j(\mathbf{Q}) \le 0\right),\ j = 1,\ldots,2n \\ M_g \sum_{i=1}^{2n} \left[\max\left(g^j, 0\right)\right]^2, \text{ if } \exists j;\ g^j(\mathbf{Q}) > 0,\ j = 1,\ldots,2n \end{cases} \tag{12}$$

where M_g is the number of constraint function violations, $\mathbf{Q}$ is the vector of the decision variables and $g^j(\mathbf{Q})$ is a term that contains all the constraints of Equation (11). It should be noted that Equation (12) applies for both the VD and the SI model.

4. Experimental results

4.1. Experiment setup

The SWI models described in Sections 2.1–2.3, as well as the optimization frameworks of Section 2.1 are applied on a hypothetical unconfined aquifer of a simplified rectangular shape. The horizontal dimensions of the examined aquifer are $L \times W = 7.000\,\text{m} \times 3.000\,\text{m}$ and the base of the aquifer is considered horizontal and is located at a depth of −25 m below sea level. Along the coastline, a hydrostatic boundary condition is set with a constant salinity concentration of 35 kg /m^3. Also, a specified flux is applied along the entire inland boundary, while the remaining boundaries—lateral and aquifer bottom—are considered impermeable. The aquifer is considered to be homogeneous and anisotropic with regard to the hydraulic conductivity and is pumped through ten fully penetrating wells. Finally, a uniform recharge replenishes the aquifer along its upper boundary. It should be noted that for both the recharge and the well pumping rates we used constant average values throughout the entire 50-year simulation period. Figure 1 is a conceptual representation of the coastal aquifer model, while Table 1 includes all the model's parameters. Each VD simulation requires approximately 75 s in a i7-4770 CPU of 3.7 GHz, 16 GB RAM to complete. The duration of the corresponding SI simulation is in the order of 0.3 s. The significant difference in computational time between the two SWI models highlights the benefit of the SI approach and its usefulness in applications which require a large number of simulations, such as a simulation-optimization method in groundwater management problems.

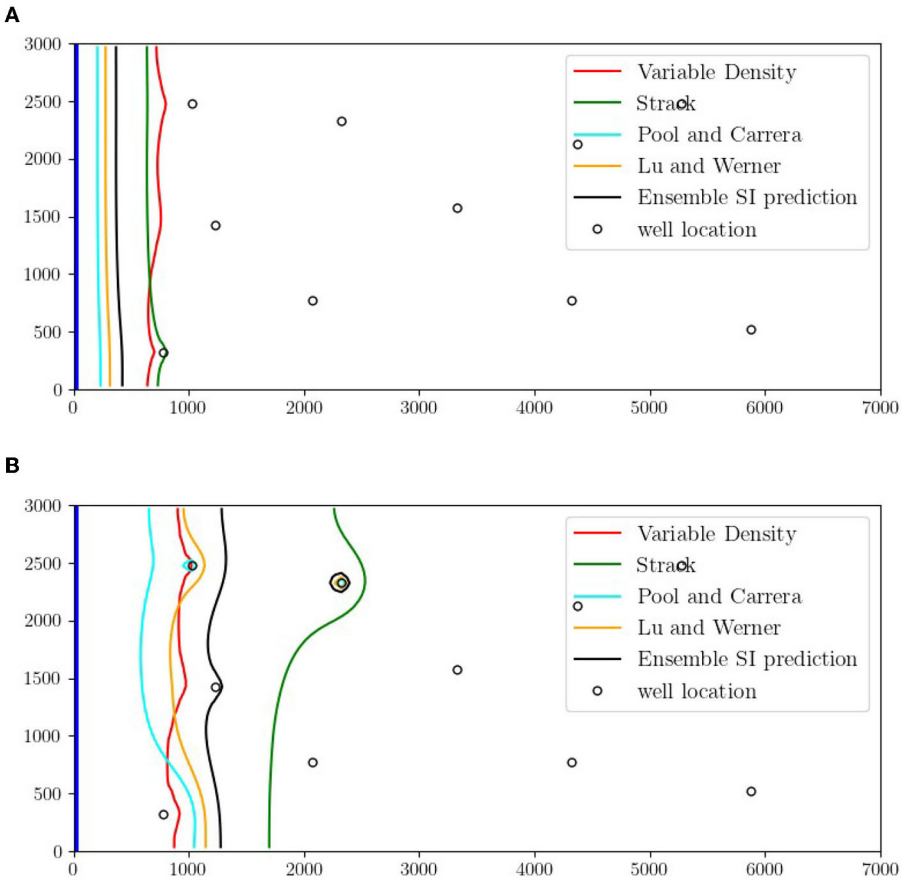

FIGURE 2
Top view of the bottom layer of the aquifer depicting the advancement of SWI for the VD model and the four variations of the SI model based on the same sets of pumping rates. **(A)** Results comparison based on a low total pumping scenario. **(B)** Results comparison based on a high total pumping scenario.

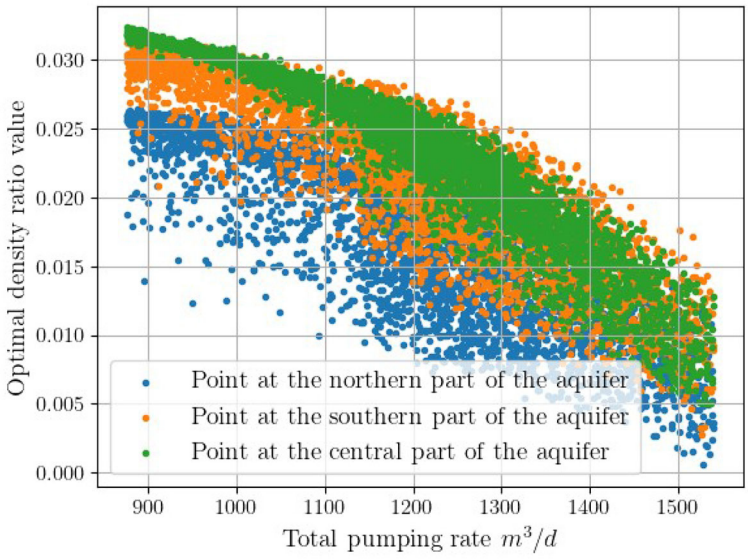

FIGURE 3
Optimal density ratio in relation to the total pumping rates for the 5,000 pumping rate scenarios. Each color corresponds to a point located at different part of the aquifer.

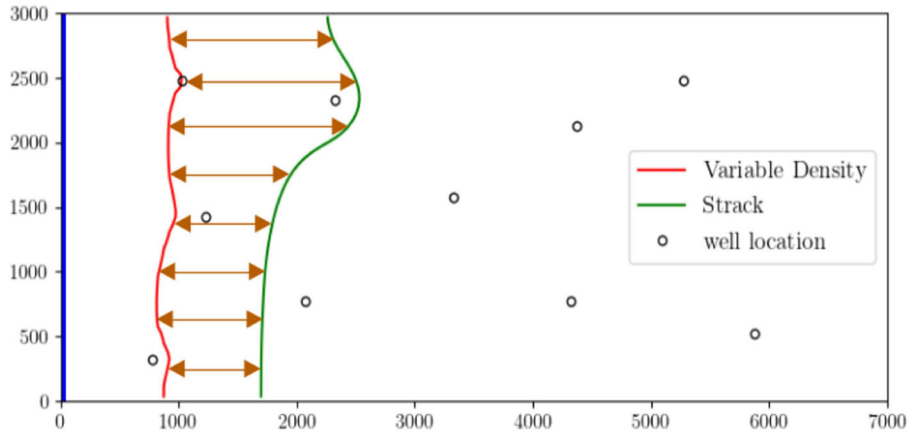

FIGURE 4
Relative position of the 0.1 kg/m^3 isohaline and the SI toe based on the Strack model. The arrows indicate the correction for each one of the points of the SI toe. A RF model is developed and trained for each one of the SI toe points.

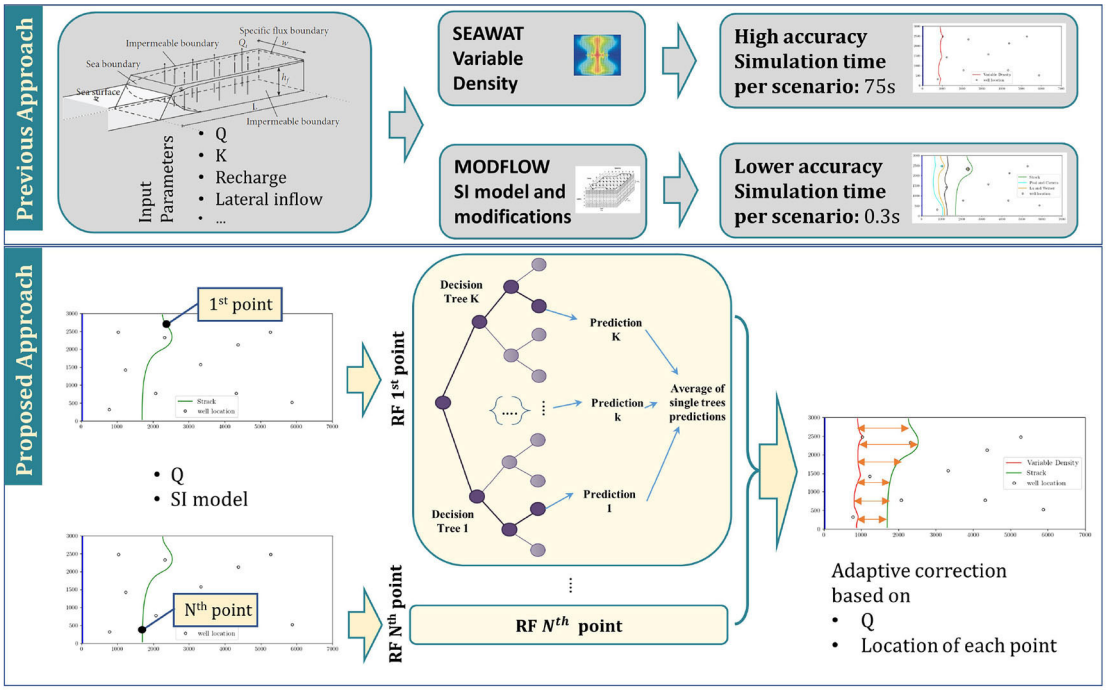

FIGURE 5
Overall concept of the proposed RF method for the correction of the Strack's based SI toe.

4.2. Comparison of seawater intrusion models

As already mentioned in Section 4.1, there is a clear advantage to utilizing the SI model rather than the VD model for groundwater management applications, due to its computational efficiency. In the present study, we further examine the correction of the SI model discussed in Section 2.3 by searching for any potential spatial dependencies of the density ratio, which would lead not to a unique correction factor for the entire flow field—as in previous studies—but to a locally adapted modification of the SI toe. To this end, we create a large dataset of 5,000 pumping scenarios using the Latin Hypercube Sampling (LHS) statistical method. The pumping scenarios are applied on both the VD and SI models and subsequently the locations of the 0.1 kg /m^3 isohaline and the SI toe are extracted from the concentration results and the flow potential field, respectively. Figures 2A, B depict the advancement of SWI for the VD model and the four variations of the SI model which include: (i) the original Strack model, (ii) the Pool and Carrera (2011) correction, (iii) the Lu and Werner (2013) correction, and (iv) the ensemble SI model described in Section 2.3. The results indicates that for low pumping rates the model proposed by Strack (Strack, 1976) outperforms the other SI models. However, in high pumping rate scenarios—which is of practical

TABLE 2 Value ranges of optimized hyperparameters of RF and indicative optimized values for one of the 60 points of the SI toe.

Parameter name	Value range or parameter options	Optimized values
Method	Bootstrap aggregation (bagging) or least-squares boosting	Least-squares boosting
Number of learning cycles	$10 - 500$	62
Learning rate	$10^{-3} - 1$	0.2281
Minimum leaf size	$1 - 1,998$	1,891

interest for groundwater management problems—the Strack model overestimates the extend of SWI, while the Lu and Luo (2014) models seems to capture more accurately the benchmark VD solution. The ensemble model toe closer resembles the shape of the 0.1 kg /m^3 isohaline, though shifted several hundred meters inland. Overall, the results demonstrate a considerable dependency of the SI models' efficiency on the total pumping rates, confirming the findings of previous studies (e.g., Christelis and Mantoglou, 2016a).

In order to define possible spatial dependencies of the correction factor both the 0.1 kg /m^3 isohaline and the SI toe based on Strack's model were divided in an equal number of segments and the correction of the latter was examined at the endpoints of each segment. In the hypothetical aquifer of the present study we selected a dense segmentation of the lines to achieve a relatively comprehensive and complete analysis. In particular, following the finite different discretization of the numerical models, we divided both lines into 59 segments. For each one of the segments' endpoints and for each pumping scenario we calculated an optimal density ratio that minimizes the distance between the two points. This problem is equivalent to calculating the potential flow value at each point of the 0.1 kg /m^3 isohaline through an interpolation method, and then solve Equation (7) for ϵ. Figure 3 presents the optimal values for the 5, 000 pumping scenarios, for three segments of the a the critical isohaline, representing both the northern and the southern part of the aquifer, as well as the middle area. The figure leads to the following two key conclusions

- The optimal values of the correction factor has a strong dependence on the pumping rates. In particular, the optimal values tend to reduce as the total pumping rates increase.
- The optimal correction factor follows the same pattern in all the segments of the compared lines. However, its mean value and range differs indicating a spatial variability, which should be further investigated.

4.3. Sharp interface correction based on random forests

The findings and observations of Section 4.2 suggest that the correction factor proposed in earlier studies for the Strack SI model (e.g., Pool and Carrera, 2011; Lu and Werner, 2013), enhance the determination of the toe position considerably, aligning with the benchmark VD approach. However, it has certain limitations and its overall precision is contingent on the pumping regime. Furthermore, applying a uniform correction factor across the entire flow field does not have the capability to adapt to the specific conditions of distinct areas within the aquifer, which—for the case of homogeneous aquifer, like the hypothetical aquifer examined in the present study—are mostly related to the relative position of the wells and the applied pumping rates.

Following the line segmentation approach of Section 4.2, a total number of 60 points is selected from both the 0.1 kg /m^3 and the SI toe based on Strack model, as depicted in Figure 4. The arrows in the figure denote the intended correction for indicative points of the line. In order to investigate a spatially variable correction factor, we utilize a ML method, and particularly the RF algorithm described in Section 2.3. The training sample consists of 4, 000 variable sets, representing the 80% of the total sample. Each variable set has 11 parameters, namely the pumping rates of the 10 extraction wells and the initial position of the examined point relative to the coastline. The output set consists of a single variable, that is the distance of the examined SI toe point from its relative point lying on the 0.1 kg /m^3 isohaline. Figure 5 schematically presents the overall concept of the proposed method in contrast to a typical approach in coastal aquifer managements problems. Regarding the RF based correction of the SI, the following points should be mentioned

- A multiple input/single output version of the RF algorithm was implemented in the present study. Thus, one RF model was developed and trained for each one of the points of the SI toe.
- Contrary to previous studies, we adopted a direct correction of the location of the SI toe and not an indirect correction through the modification of the density ratio. Modeling the distance of the examined line points seems to provide a relatively smooth corrected line, while the density ratio solution tends to overestimate or underestimate the displacement of the toe points, as small changes in the parameter value lead to a non-proportional correction.

In our experiments, we used a k-fold cross-validation technique to mitigate possible overfitting and achieve a more robust evaluation of the RF models' generalization ability. The train dataset is divided into $k = 5$ subsets (folds) of equal size. The RF hyperparameters were optimized using Bayesian optimization over these 5k-fold cross-validation sets. Table 2 summarizes the optimizable variables and the corresponding value ranges. The number of splits falls within the range of 1 to 3995.

For the evaluation of the RF models performance we use as metrics the mean absolute error (MAE), the root mean squared error (RMSE), and the R-squared (R^2), which are defined as follows:

$$\mathrm{MAE} = \frac{\sum_{i=1}^{N} |y_i - \hat{y}_i|}{N} \tag{13a}$$

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{N} (y_i - \hat{y}_i)^2}{N}} \tag{13b}$$

TABLE 3 Performance evaluation of the proposed method for half of the examined points.

Point ID	MAE train set	RMSE train set	R^2 train set	MAE test set	RMSE test set	R^2 test set
1	3.20 ± 2.46	4.53 ± 3.87	0.9997	7.24 ± 1.37	15.78 ± 2.34	0.9976
2	3.35 ± 1.18	4.99 ± 2.04	0.9997	5.27 ± 1.20	11.20 ± 2.77	0.9988
3	4.54 ± 3.04	6.27 ± 4.62	0.9994	7.20 ± 2.95	13.69 ± 5.26	0.9980
4	2.94 ± 1.16	3.95 ± 1.75	0.9998	5.81 ± 0.53	12.54 ± 1.29	0.9985
5	5.95 ± 3.49	8.30 ± 5.09	0.9991	8.23 ± 4.00	14.67 ± 7.25	0.9976
6	3.55 ± 2.09	5.06 ± 3.59	0.9996	6.88 ± 1.44	12.57 ± 2.40	0.9985
7	4.65 ± 2.80	6.29 ± 3.98	0.9995	7.09 ± 2.84	12.36 ± 4.51	0.9985
8	4.85 ± 2.23	6.85 ± 3.28	0.9995	7.31 ± 1.89	12.15 ± 3.06	0.9986
9	3.45 ± 1.49	4.81 ± 2.25	0.9998	5.94 ± 0.90	10.10 ± 1.36	0.9992
10	4.76 ± 1.97	6.45 ± 2.82	0.9996	8.66 ± 4.67	18.57 ± 15.95	0.9955
11	3.07 ± 1.10	5.34 ± 3.66	0.9968	5.41 ± 0.67	11.68 ± 1.46	0.9990
12	6.67 ± 2.52	10.72 ± 3.48	0.9977	8.41 ± 2.88	14.67 ± 5.19	0.9984
13	3.58 ± 1.16	5.88 ± 2.89	0.9997	5.50 ± 1.07	10.65 ± 1.74	0.9993
14	3.14 ± 1.59	4.40 ± 2.50	0.9998	5.83 ± 0.98	13.51 ± 3.05	0.9988
15	4.55 ± 1.65	6.22 ± 2.17	0.9997	7.07 ± 1.54	13.81 ± 4.28	0.9988
16	3.19 ± 1.00	8.86 ± 9.71	0.9990	4.64 ± 0.82	10.65 ± 3.47	0.9993
17	2.36 ± 1.11	3.28 ± 1.69	0.9999	4.86 ± 0.69	10.06 ± 2.76	0.9994
18	2.79 ± 1.59	3.74 ± 2.11	0.9999	4.59 ± 1.63	10.05 ± 2.21	0.9994
19	4.63 ± 2.81	6.54 ± 4.14	0.9996	6.84 ± 2.84	12.86 ± 4.19	0.9989
20	4.62 ± 2.27	6.53 ± 3.70	0.9996	6.91 ± 2.03	12.41 ± 3.85	0.9990
21	3.52 ± 1.57	5.24 ± 2.87	0.9998	5.11 ± 1.45	11.46 ± 4.99	0.9991
22	1.81 ± 0.51	2.51 ± 0.99	1.0000	4.22 ± 0.78	8.93 ± 1.98	0.9995
23	3.16 ± 1.69	4.31 ± 2.34	0.9998	5.07 ± 2.11	9.66 ± 3.66	0.9993
24	3.29 ± 1.58	4.52 ± 2.16	0.9998	5.62 ± 1.60	10.50 ± 2.40	0.9992
25	3.44 ± 1.02	6.96 ± 5.52	0.9994	6.77 ± 1.61	12.89 ± 3.30	0.9988
26	3.23 ± 1.10	4.25 ± 1.47	0.9999	6.40 ± 1.71	12.62 ± 3.48	0.9988
27	4.75 ± 1.86	6.76 ± 2.80	0.9996	7.55 ± 1.66	13.30 ± 2.69	0.9987
28	4.51 ± 2.37	6.37 ± 3.70	0.9996	8.26 ± 1.78	14.57 ± 3.01	0.9984
29	4.93 ± 1.43	8.07 ± 3.76	0.9994	8.85 ± 1.37	15.40 ± 2.73	0.9982
30	5.90 ± 1.63	9.34 ± 3.00	0.9993	9.45 ± 1.55	15.10 ± 2.06	0.9983
Average	3.78 ± 1.83	6.46 ± 3.92	0.9994	6.25 ± 1.78	11.99 ± 3.58	0.9988

The units of MAE and RMSE are m. The mean values at the end of the table refer to all 60 points.

$$R^2 = 1 - \frac{\sum_{i=1}^{N} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{N} (y_i - \bar{y}_i)^2} \tag{13c}$$

where N is the number of variable sets in the train or test dataset, y_i represents the true values of the modeled parameter, $\hat{y}_i$ stands for the estimated values and $\bar{y}_i$ is the mean value of the parameter. These metrics are applied in both training and test sets.

In order to ensure the accuracy and robustness of the final RF models, the entire training and validation process was repeated five times for all the examined points of the SI toe. The performance metrics were computed for each one of the five output values datasets. The repetition of the process allowed the approximate quantification of the uncertainty of the proposed method. The uncertainty is provided as the standard deviation of the metrics (e.g., MAE) over the five repetitions. In particular, the calculations for the MAE are based on the following formula:

$$SD = \sqrt{\frac{\sum_{i=1}^{r} \left(MAE_i - \overline{MAE}\right)^2}{r}} \tag{14}$$

Table 3 contains the values of the statistical measurements used to evaluate the performance of the proposed method, namely the mean absolute error (MAE), the root mean square error (RMSE) and the R^2. For the MAE we also provide the corresponding standard deviation. It should be noted that the values listed in the table are the averages over the five repetitions of the calculation process for half of the 60 points selected from the SI toe. Given the magnitude of the examined parameter and the size of the finite difference cell, the MAE and RMSE values are considered satisfactory. At the bottom of the table, we also provide the average values of all the metrics over the entire set of points. Regarding the high values of the R^2, the use of the cross-validation approach and the repetition of the process minimize the possibility of overfitting and ensure the model's generalization ability. The proposed methodology significantly improves the location of the SI toe, with respect to the 0.1 kg /m^3 isohaline.

4.4. Pumping optimization results

In order to test the usefulness of the proposed LF model in pumping optimization problems, we apply the optimization framework described in Section 3 on the unconfined hypothetical aquifer. For consistency, we solved the optimization problem not only for the HF model, but also for the original Strack model. The parameters of the EAS optimization algorithm are listed in Table 4 and their values are set based on the work of Efstratiadis and Koutsoyiannis (2002), Tsoukalas et al. (2016) and Christelis and Mantoglou (2019). The optimization is terminated if at least one of the following criteria is met: (i) the convergence criterion ε reaches its preset value, and (ii) the number of objective function evaluations exceeds the maximum number n_{max}.

Table 5 contains the optimal pumping rates for the examined HF and LF models. As expected, the Strack's SI model provides a relatively low estimation of the total pumping rates, which does not exceed 75% of the VD optimal results. Using the proposed correction of the SI significantly improved the results, allowing the extraction of a freshwater amount that corresponds to 95% of the VD estimation. Furthermore, it is worth mentioning that the evaluation of the optimal pumping rates with the VD optimization framework confirms that it is a feasible solution, since none of the constraint functions was violated. Figure 6 shows the 0.1 kg /m^3 isohaline corresponding to the optimal pumping rates for the two examined SI models. The RF-based correction of the Strack SI model provides a solution which approaches the benchmark VD solution, without violating the constraint set.

TABLE 4 Parameter values of the EAS optimization algorithm.

Parameter name		Parameter value
Initial population	n_{pop}	$8n$
Annealing schedule control parameters	λ_p	0.95
	ψ	2
Mutation probability	m_p	0.1
Convergence criterion	ε	0.1
Maximum objective function evaluations	n_{max}	$100n_{pop}$

Note that n is the number of pumping wells.

5. Conclusions

Due to the low computational budget and the relatively easy implementation the SI models are considered a promising surrogate model for the more accurate, but computationally expensive VD model. For this reason, the SI model have been extensively investigated in a number of previous studies. Most of them focus on improving the Strack SI models by modifying the density ratio. The present work aims to provide a ML based correction of the original Strack SI model. In a first step, the Strack model and three modifications were examined in several pumping levels. From the corresponding SWI extent it was evident that the behavior is significantly influenced by the total amount of the freshwater extraction, thus imposing limitations in their ability to function as effective substitutes for the VD model. This observation required further investigation, and to this end we calculated an optimal density ratio that minimizes the discrepancy between the SI toe and the 0.1 kg /m^3 isohaline. Despite the fact that the optimized density ratio corrected the location of the SI toe, it failed to encapsulate possible local conditions in the flow field (e.g., interaction between neighboring pumping wells).

The above findings highlighted the need to search for a regionally adapted correction factor, and to achieve this goal, the SI toe was divided in several segments, which were examined individually with respect to the density ratio. A considerable number of simulations were performed using both the VD and the SI approach and based on the results a secondary inverse problem setup was implemented to calculate the optimal values

TABLE 5 Optimal pumping rates for the HF (VD) model and the two LF models: (i) the original Strack SI model, and (ii) the corrected SI model using the RF (RF-SI).

SWI model	Well pumping rates (m /d)										Total pumping
	1	2	3	4	5	6	7	8	9	10	
VD	0.58	0	181.13	58.55	53.37	72.75	122.86	207.61	295.13	272.82	1564.78
Strack SI	0.06	15.41	81.59	1.60	259.17	73.48	1.52	462.80	270.63	5.17	1171.44
RF-SI	16.24	79.89	130.84	178.58	246.24	251.40	59.95	104.90	209.57	193.42	1471.03

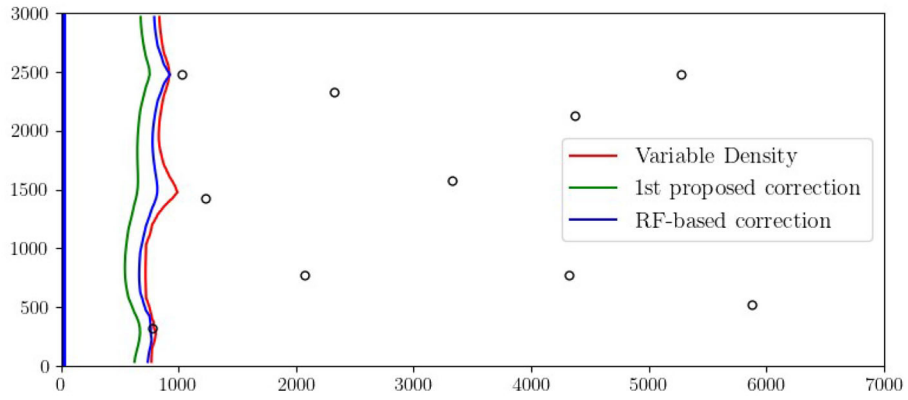

FIGURE 6
Projection of the 0.1 kg/m^3 isohaline at the bottom of the aquifer, which correspond to the VD evaluation of the optimal pumping rates for two SI models: (i) the original Strack model and (ii) the corrected SI model based on RF. The VD optimal solution is also included for comparison purposes.

of the correction factor. The outcome confirmed the spatial variability of the examined parameter, in terms of the magnitude and value range. In order to model the regional dependency of the correction factor, the RF algorithm was utilized for each one of the individual points of the Strack SI toe. In particular, instead of using the density ratio as a correction factor, the point to point distance between the SI toe and the 0.1 kg /m^3 isohaline was selected as the output of the ML method. Based on the RF evaluation metrics—specifically, the MAE and RMSE results for both the train and test data sets—, the proposed correction demonstrates a good performance, with a mean accuracy in the order of $5 - 10$ m.

Subsequently, the proposed correction factor was incorporated in a coastal aquifer optimization framework. The estimated optimal pumping rates provided feasible solutions, which approximated the benchmark VD optimal results within a 5% margin. To the best of our knowledge, this is the first study that examines in detail the spatial variability of the SI correction and implements the correction in a coastal aquifer pumping optimization problem. Overall, the proposed method is yielding promising results and could be used as an accurate and robust LF model in coastal aquifer management problems. Future research will focus on further scrutinizing the effectiveness of the proposed method in different aquifer setups and real applications.

Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

Author contributions

GK and MK: conceptualization and visualization. GK, MK, EP, and AV: methodology and formal analysis. GK, MK, EP, AV, AD, ND, and AM: validation and writing—review and editing. GK: investigation and data curation. GK, MK, and EP: resources and writing—original draft preparation. MK, EP, AV, AD, ND, and AM: supervision. All authors have read and agreed to the published version of the manuscript.

Funding

This paper was supported by the European Union Funded project euPOLIS Integrated NBS-based Urban Planning Methodology for Enhancing the Health and Well-being of Citizens: the euPOLIS Approach under the Horizon 2020 program H2020-EU.3.5.2., grant agreement no 869448.

Acknowledgments

This work was supported by computational time granted from the National Infrastructures for Research and Technology S.A. (GRNET) in the National HPC facility ARIS-under project ID pa211104 (MLGM).

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

References

Alexandrov, N. M., and Lewis, R. M. (2001). An overview of first-order model management for engineering optimization. *Opt. Eng.* 2, 413–430. doi: 10.1023/A:1016042505922

Asher, M. J., Croke, B. F., Jakeman, A. J., and Peeters, L. J. (2015). A review of surrogate models and their application to groundwater modeling. *Water Resour. Res.* 51, 5957–5973. doi: 10.1002/2015WR016967

Ataie-Ashtiani, B., Ketabchi, H., and Rajabi, M. M. (2014). Optimal management of a freshwater lens in a small island using surrogate models and evolutionary algorithms. *J. Hydrol. Eng.* 19, 339–354. doi: 10.1061/(ASCE)HE.1943-5584.0000809

Bhattacharjya, R. K., Datta, B., and Satish, M. G. (2007). Artificial neural networks approximation of density dependent saltwater intrusion process in coastal aquifers. *J. Hydrol. Eng.* 12, 273–282. doi: 10.1061/(ASCE)1084-0699(2007)12:3(273)

Breiman, L. (2001). Random forests. *Mach. Learn.* 45, 5–32. doi: 10.1023/A:1010933404324

Canion, A., McCloud, L., and Dobberfuhl, D. (2019). Predictive modeling of elevated groundwater nitrate in a karstic spring-contributing area using random forests and regression-kriging. *Environ. Earth Sci.* 78, 271. doi: 10.1007/s12665-019-8277-1

Christelis, V., and Hughes, A. G. (2018). Metamodel-assisted analysis of an integrated model composition: an example using linked surface water–groundwater models. *Environ. Modell. Softw.* 107, 298–306. doi: 10.1016/j.envsoft.2018.05.004

Christelis, V., Kopsiaftis, G., and Mantoglou, A. (2019). Performance comparison of multiple and single surrogate models for pumping optimization of coastal aquifers. *Hydrol. Sci. J.* 64, 336–349. doi: 10.1080/02626667.2019.1584400

Christelis, V., and Mantoglou, A. (2013). "Improved sharp interface models in coastal aquifers of finite dimensions," in *EGU General Assembly Conference Abstracts* (Vienna).

Christelis, V., and Mantoglou, A. (2016a). Coastal aquifer management based on the joint use of density-dependent and sharp interface models. *Water Resour. Manage.* 30, 861–876. doi: 10.1007/s11269-015-1195-4

Christelis, V., and Mantoglou, A. (2016b). Pumping optimization of coastal aquifers assisted by adaptive metamodelling methods and radial basis functions. *Water Resour. Manage.* 30, 5845–5859. doi: 10.1007/s11269-016-1337-3

Christelis, V., and Mantoglou, A. (2019). Pumping optimization of coastal aquifers using seawater intrusion models of variable-fidelity and evolutionary algorithms. *Water Resour. Manage.* 33, 555–568. doi: 10.1007/s11269-018-2116-0

Christelis, V., Regis, R. G., and Mantoglou, A. (2018). Surrogate-based pumping optimization of coastal aquifers under limited computational budgets. *J. Hydroinform.* 20, 164–176. doi: 10.2166/hydro.2017.063

Dausman, A. M., Langevin, C. D., Bakker, M., and Schaars, F. (2010). "A comparison between SWI and SEAWAT–the importance of dispersion, inversion and vertical anisotropy," in *Proceedings of SWIM* (Azores), 271–274.

Dokou, Z., and Karatzas, G. P. (2012). Saltwater intrusion estimation in a karstified coastal system using density-dependent modelling and comparison with the sharp-interface approach. *Hydrol. Sci. J.* 57, 985–999. doi: 10.1080/02626667.2012.690070

Efstratiadis, A., and Koutsoyiannis, D. (2002). "An evolutionary annealing-simplex algorithm for global optimisation of water resource systems," in *Hydroinformatics 2002: Proceedings of the Fifth International Conference on Hydroinformatics* (Cardiff), 1423–1428.

Fahs, M., Koohbor, B., Shao, Q., Doummar, J., Baalousha, H. M., and Voss, C. I. (2022). Effect of flow-direction-dependent dispersivity on seawater intrusion in coastal aquifers. *Water Resour. Res.* 58. doi: 10.1029/2022WR032315

Fathizad, H., Ali Hakimzadeh Ardakani, M., Sodaiezadeh, H., Kerry, R., and Taghizadeh-Mehrjardi, R. (2020). Investigation of the spatial and temporal variation of soil salinity using random forests in the central desert of Iran. *Geoderma* 365, 114233. doi: 10.1016/j.geoderma.2020.114233

Forrester, A. I. J., Sobester, A., and Keane, A. J. (2008). *Engineering Design via Surrogate Modelling: A Practical Guide*. Chichester; Hoboken, NJ: J. Wiley.

Frind, E. O. (1982). Simulation of long-term transient density-dependent transport in groundwater. *Adv. Water Resour.* 5, 73–88.

Giselle Fernandez-Godino, M., Park, C., Kim, N. H., and Haftka, R. T. (2019). Issues in deciding whether to use multifidelity surrogates. *AIAA J.* 57, 2039–2054. doi: 10.2514/1.J057750

Guo, W., and Langevin, C. D. (2002). *User's Guide to SEAWAT; A Computer Program for Simulation of Three-Dimensional Variable-Density Ground-Water Flow*. Technical report.

He, S., Wu, J., Wang, D., and He, X. (2022). Predictive modeling of groundwater nitrate pollution and evaluating its main impact factors using random forest. *Chemosphere* 290, 133388. doi: 10.1016/j.chemosphere.2021.133388

Ho, T. K. (1998). The random subspace method for constructing decision forests. *IEEE Trans. Pattern Anal. Mach. Intell.* 20, 832–844.

Karatzas, G. P., and 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. doi: 10.1007/s10040-015-1286-6

Ketabchi, H., and Ataie-Ashtiani, B. (2015). Evolutionary algorithms for the optimal management of coastal groundwater: a comparative study toward future challenges. *J. Hydrol.* 520, 193–213. doi: 10.1016/j.jhydrol.2014.11.043

Knoll, L., Breuer, L., and Bach, M. (2019). Large scale prediction of groundwater nitrate concentrations from spatial data using machine learning. *Sci. Tot. Environ.* 668, 1317–1327. doi: 10.1016/j.scitotenv.2019.03.045

Koch, J., Berger, H., Henriksen, H. J., and Sonnenborg, T. O. (2019). Modelling of the shallow water table at high spatial resolution using random forests. *Hydrol. Earth Syst. Sci.* 23, 4603–4619. doi: 10.5194/hess-23-4603-2019

Kolditz, O., Ratke, R., Diersch, H.-J. G., and 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., and Mantoglou, A. (2019a). Comparison of sharp interface to variable density models in pumping optimisation of coastal aquifers. *Water Resour. Manage.* 33, 1397–1409. doi: 10.1007/s11269-019-2194-7

Kopsiaftis, G., Protopapadakis, E., Voulodimos, A., Doulamis, N., and Mantoglou, A. (2019b). Gaussian process regression tuned by Bayesian optimization for seawater intrusion prediction. *Comput. Intell. Neurosci.* 2019, e2859429. doi: 10.1155/2019/2859429

Kourakos, G., and Mantoglou, A. (2009). Pumping optimization of coastal aquifers based on evolutionary algorithms and surrogate modular neural network models. *Adv. Water Resour.* 32, 507–521. doi: 10.1016/j.advwatres.2009.01.001

Kourakos, G., and Mantoglou, A. (2013). Development of a multi-objective optimization algorithm using surrogate models for coastal aquifer management. *J. Hydrol.* 479, 13–23. doi: 10.1016/j.jhydrol.2012.10.050

Kourgialas, N. N., Dokou, Z., Karatzas, G. P., Panagopoulos, G., Soupios, P., Vafidis, A., et al. (2015). Saltwater intrusion in an irrigated agricultural area: combining density-dependent modeling and geophysical methods. *Environ. Earth Sci.* 75, 15. doi: 10.1007/s12665-015-4856-y

Koussis, A. D., and 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. doi: 10.1007/s10040-018-1817-z

Koussis, A. D., Mazi, K., and 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. doi: 10.1016/j.jhydrol.2011.11.012

Koussis, A. D., Mazi, K., Riou, F., and Destouni, G. (2015). A correction for Dupuit–Forchheimer interface flow models of seawater intrusion in unconfined coastal aquifers. *J. Hydrol.* 525, 277–285. doi: 10.1016/j.jhydrol.2015.03.047

Lahjouj, A., El Hmaidi, A., Bouhafa, K., and Boufala, M. (2020). Mapping specific groundwater vulnerability to nitrate using random forest: case of Sais basin, Morocco. *Model. Earth Syst. Environ.* 6, 1451–1466. doi: 10.1007/s40808-020-00761-6

Lal, A., and Datta, B. (2018). Development and implementation of support vector machine regression surrogate models for predicting groundwater pumping-induced saltwater intrusion into coastal aquifers. *Water Resour. Manage.* 32, 2405–2419. doi: 10.1007/s11269-018-1936-2

Langevin, C. D., Thorne, D. T. Jr., Dausman, A. M., Sukop, M. C., and Guo, W. (2008). *SEAWAT Version 4: A Computer Program for Simulation of Multi-Species Solute and Heat Transport: U.S. Geological Survey Techniques and Methods*. USGS (Florida Integrated Science Center), Technical report.

Llopis-Albert, C., and 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. doi: 10.1002/hyp.9908

Lu, C., Chen, Y., and Luo, J. (2012). Boundary condition effects on maximum groundwater withdrawal in coastal aquifers. *Groundwater* 50, 386–393. doi: 10.1111/j.1745-6584.2011.00880.x

Lu, C., and Luo, J. (2014). Groundwater pumping in head-controlled coastal systems: the role of lateral boundaries in quantifying the interface toe location and maximum pumping rate. *J. Hydrol.* 512, 147–156. doi: 10.1016/j.jhydrol.2014.02.034

Lu, C., and Werner, A. D. (2013). Timescales of seawater intrusion and retreat. *Adv. Water Resour.* 59, 39–51. doi: 10.1016/j.advwatres.2013.05.005

Lu, C., Xin, P., Kong, J., Li, L., and Luo, J. (2016). Analytical solutions of seawater intrusion in sloping confined and unconfined coastal aquifers. *Water Resour. Res.* 52, 6989–7004. doi: 10.1002/2016WR019101

Mantoglou, A. (2003). Pumping management of coastal aquifers using analytical models of saltwater intrusion. *Water Resour. Res.* 39. doi: 10.1029/2002WR001891

Mantoglou, A., Papantoniou, M., and Giannoulopoulos, P. (2004). Management of coastal aquifers based on nonlinear optimization and evolutionary algorithms. *J. Hydrol.* 297, 209–228. doi: 10.1016/j.jhydrol.2004.04.011

Masroor, M., Rehman, S., Sajjad, H., Rahaman, M. H., Sahana, M., Ahmed, R., et al. (2021). Assessing the impact of drought conditions on groundwater potential in Godavari Middle Sub-Basin, India using analytical hierarchy process and random forest machine learning algorithm. *Groundwater Sustain. Dev.* 13, 100554. doi: 10.1016/j.gsd.2021.100554

Messier, K. P., Wheeler, D. C., Flory, A. R., Jones, R. R., Patel, D., Nolan, B. T., et al. (2019). Modeling groundwater nitrate exposure in private wells of North Carolina for the Agricultural Health Study. *Sci. Tot. Environ.* 655, 512–519. doi: 10.1016/j.scitotenv.2018.11.022

Naghibi, S. A., Vafakhah, M., Hashemi, H., Pradhan, B., and Alavi, S. J. (2020). Water resources management through flood spreading project suitability mapping using frequency ratio, k-nearest neighbors, and random forest algorithms. *Nat. Resour. Res.* 29, 1915–1933. doi: 10.1007/s11053-019-09530-4

Norouzi, H., and Shahmohammadi-Kalalagh, S. (2019). Locating groundwater artificial recharge sites using random forest: a case study of Shabestar region, Iran. *Environ. Earth Sci.* 78, 380. doi: 10.1007/s12665-019-8381-2

Ouedraogo, I., Defourny, P., and Vanclooster, M. (2019). Application of random forest regression and comparison of its performance to multiple linear regression in modeling groundwater nitrate concentration at the African continent scale. *Hydrogeol. J.* 27, 1081–1098. doi: 10.1007/s10040-018-1900-5

Papacharalampous, G., Tyralis, H., and Koutsoyiannis, D. (2018). Univariate time series forecasting of temperature and precipitation with a focus on machine learning algorithms: a multiple-case study from Greece. *Water Resour. Manage.* 32, 5207–5239. doi: 10.1007/s11269-018-2155-6

Papacharalampous, G., Tyralis, H., and Koutsoyiannis, D. (2019). Comparison of stochastic and machine learning methods for multi-step ahead forecasting of hydrological processes. *Stochast. Environ. Res. Risk Assess.* 33, 481–514. doi: 10.1007/s00477-018-1638-6

Pham, L. T., Luo, L., and Finley, A. (2021). Evaluation of random forests for short-term daily streamflow forecasting in rainfall- and snowmelt-driven watersheds. *Hydrol. Earth Syst. Sci.* 25, 2997–3015. doi: 10.5194/hess-25-2997-2021

Pool, M., and 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. doi: 10.1029/2010WR010256

Razavi, S., Tolson, B. A., and Burn, D. H. (2012). Review of surrogate modeling in water resources: REVIEW. *Water Resour. Res.* 48. doi: 10.1029/2011WR011527

Roy, D. K., and Datta, B. (2017a). Fuzzy C-mean clustering based inference system for saltwater intrusion processes prediction in coastal aquifers. *Water Resour. Manage.* 31, 355–376. doi: 10.1007/s11269-016-1531-3

Roy, D. K., and Datta, B. (2017b). Multivariate adaptive regression spline ensembles for management of multilayered coastal aquifers. *J. Hydrol. Eng.* 22, 04017031. doi: 10.1061/(ASCE)HE.1943-5584.0001550

Roy, D. K., and Datta, B. (2017c). "Optimal management of groundwater extraction to control saltwater intrusion in multi-layered coastal aquifers using ensembles of adaptive neuro-fuzzy inference system," in *World Environmental and Water Resources Congress 2017* (Sacramento, CA: American Society of Civil Engineers), 139–150.

Saha, S., Mallik, S., and Mishra, U. (2022). "Groundwater depth forecasting using machine learning and artificial intelligence techniques: a survey of the literature," in *Recent Developments in Sustainable Infrastructure (ICRDSI-2020)* eds B. B. Das, H. Hettiarachchi, P. K. Sahu, and S. Nanda (Singapore: Springer), 153–167.

Simmons, C. T. (2005). Variable density groundwater flow: from current challenges to future possibilities. *Hydrogeol. J.* 13, 116–119. doi: 10.1007/s10040-004-0408-3

Simpson, T., Poplinski, J., Koch, P. N., and Allen, J. (2001). Metamodels for Computer-based Engineering Design: Survey and recommendations. *Eng. Comput.* 17, 129–150. doi: 10.1007/PL00007198

Sreekanth, J., and Datta, B. (2011). Coupled simulation-optimization model for coastal aquifer management using genetic programming–based ensemble surrogate models and multiple-realization optimization. *Water Resour. Res.* 47. doi: 10.1029/2010WR009683

Strack, O. D. L. (1976). A single-potential solution for regional interface problems in coastal aquifers. *Water Resour. Res.* 12, 1165–1174.

Trichakis, I. C., Nikolos, I. K., and Karatzas, G. P. (2011). Artificial neural network (ANN) based modeling for karstic groundwater level simulation. *Water Resour. Manage.* 25, 1143–1152. doi: 10.1007/s11269-010-9628-6

Tsoukalas, I., Kossieris, P., Efstratiadis, A., and 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. doi: 10.1016/j.envsoft.2015.12.008

Wang, X., Liu, T., Zheng, X., Peng, H., Xin, J., and Zhang, B. (2018). Short-term prediction of groundwater level using improved random forest regression with a combination of random features. *Appl. Water Sci.* 8, 125. doi: 10.1007/s13201-018-0742-6

Werner, A. D. (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. doi: 10.1016/j.advwatres.2017.02.001

Werner, A. D., Bakker, M., Post, V. E. A., Vandenbohede, A., Lu, C., Ataie-Ashtiani, B., et al. (2013). Seawater intrusion processes, investigation and management: Recent advances and future challenges. *Adv. Water Resour.* 51, 3–26. doi: 10.1016/j.advwatres.2012.03.004

Werner, A. D., Ward, J. D., Morgan, L. K., Simmons, C. T., Robinson, N. I., and Teubner, M. D. (2012). Vulnerability indicators of sea water intrusion. *Groundwater* 50, 48–58. doi: 10.1111/j.1745-6584.2011.00817.x

Wheeler, D. C., Nolan, B. T., Flory, A. R., DellaValle, C. T., and Ward, M. H. (2015). Modeling groundwater nitrate concentrations in private wells in Iowa. *Sci. Tot. Environ.* 536, 481–488. doi: 10.1016/j.scitotenv.2015.07.080

Yang, J., Griffiths, J., and Zammit, C. (2019). National classification of surface-groundwater interaction using random forest machine learning technique. *River Res. Appl.* 35, 932–943. doi: 10.1002/rra.3449

Younes, A., Fahs, M., and Ahmed, S. (2009). Solving density driven flow problems with efficient spatial discretizations and higher-order time integration methods. *Adv. Water Resour.* 32, 340–352. doi: 10.1016/j.advwatres.2008.11.003

Younes, A., Koohbor, B., Belfort, B., Ackerer, P., Doummar, J., and Fahs, M. (2022). Modeling variable-density flow in saturated-unsaturated porous media: An advanced numerical model. *Adv. Water Resour.* 159. doi: 10.1016/j.advwatres.2021.104077

Zhou, Q., Shao, X., Jiang, P., Gao, Z., Wang, C., and Shu, L. (2016). An active learning metamodeling approach by sequentially exploiting difference information from variable-fidelity models. *Adv. Eng. Inform.* 30, 283–297. doi: 10.1016/j.aei.2016.04.004