Journal of Environmental Management 370 (2024) 122535

Contents lists available at ScienceDirect

Journal of Environmental Management

journal homepage: www.elsevier.com/locate/jenvman

Research article

Beach nourishment for coastal aquifers impacted by climate change and population growth using machine learning approaches

N.L. Kushwaha [a], Kallem Sushanth [b], Abhishek Patel [c], Ozgur Kisi [d,e], Ashraf Ahmed [f,*], Ismail Abd-Elaty [g,**]

[a] Division of Agricultural Engineering, ICAR-Indian Agricultural Research Institute, New Delhi, 110012, India
[b] Department of Agricultural and Food Engineering, IIT Kharagpur, Kharagpur, West Bengal, 721302, India
[c] ICAR-Central Arid Zone Research Institute, Regional Research Station, Kukma, Bhuj, Gujrat, 370105, India
[d] Department of Civil Engineering, Technical University of Lübeck, 23562, Lübeck, Germany
[e] Department of Civil Engineering, School of Technology, Ilia State University, 0162, Tbilisi, Georgia
[f] Department of Civil and Environmental Engineering, Brunel University London, Kingston Lane, Uxbridge, UB8 3PH, UK
[g] Department of Water and Water Structures Engineering, Faculty of Engineering, Zagazig University, Zagazig, 4451, Egypt

ARTICLE INFO

Keywords:
Sea level rise
Pumping
Saltwater intrusion
Beach nourishment
Biscayne
Random forest

ABSTRACT

Groundwater in coastal regions is threatened by saltwater intrusion (SWI). Beach nourishment is used in this study to manage SWI in the Biscayne aquifer, Florida, USA, using a 3D SEAWAT model nourishment considering the future sea level rise and freshwater over-pumping. The present study focused on the development and comparative evaluation of seven machine learning (ML) models, i.e., additive regression (AR), support vector machine (SVM), reduced error pruning tree (REPTree), Bagging, random subspace (RSS), random forest (RF), artificial neural network (ANN) to predict the SWI using beach nourishment. The performance of ML models was assessed using statistical indicators such as coefficient of determination (R^2), Nash–Sutcliffe efficiency (NSE), means absolute error (MAE), root mean square error (RMSE), and root relative squared error (RRSE) along with the graphical inspection (i.e., Radar and Taylor diagram). The findings indicate that applying SVM, Bagging, RSS, and RF models has great potential in predicting the SWI values with limited data in the study area. The RF model emerged as the best fit and closely matched observed values; it obtained R^2 (0.999), NSE (0.999), MAE (0.324), RRSE (0.209), and RMSE (0.416) during the testing process. The present study concludes that the RF model could be a valuable tool for accurate predictions of SWI and effective water management in coastal areas.

1. Introduction

Natural changes include phenomena such as sea level rise (SLR), which can result from climate change and the melting of polar ice caps. Man-made changes encompass activities like excessive groundwater pumping and the management of water resources. Low precipitation levels further compound the problem by reducing the natural recharge of aquifers, leading to a depletion of freshwater resources. These forces collectively exacerbate the issue of saltwater intrusion (SWI), where seawater encroaches into freshwater aquifers (Abd-Elaty et al., 2023a; Abdelgawad et al., 2018; Abdoulhalik et al., 2020, 2024). Saltwater intrusion (SWI) influences the water demands in coastal areas and changes the biogeochemistry and dynamic systems of coastal aquifers (Moore and Joye, 2021). The Intergovernmental Panel on Climate Change (IPCC, 2007) expected the SLR would range from 58 cm to 88 cm by the year 2100, whereas the IPCC et al. (2014) predicted that SLR rates would vary from 8 to 16 mm year^{-1} by the year 2100 (Pearce et al., 2014). The SLR increases the hydrostatic pressure of the saline water and thus increases the SWI in the land direction, leading to more freshwater contamination. Abd-Elaty et al. (2023a) demonstrated that SLR impacted groundwater salinity and damaged large quantities of freshwater bodies in these regions due to SWI.

Coastal areas often provide excellent soil and climatic conditions for agriculture, which has been practised for thousands of years and plays a

* Corresponding author.
** Corresponding author.
E-mail addresses: nand.kushwaha@icar.gov.in (N.L. Kushwaha), kallemsushi@gmail.com (K. Sushanth), abhi.patel.ape121@gmail.com (A. Patel), ozgur.kisi@th-luebeck.de, ozgur.kisi@iliauni.edu.ge (O. Kisi), ashraf.ahmed@brunel.ac.uk (A. Ahmed), eng_abdelaty2006@yahoo.com, eng_abdelaty@zu.edu.eg (I. Abd-Elaty).

https://doi.org/10.1016/j.jenvman.2024.122535
Received 8 November 2023; Received in revised form 27 June 2024; Accepted 15 September 2024
Available online 26 September 2024

vital role in the economy of the coastal regions by providing food for communities and raw materials for industry (Singh, 2020). Physical surface barriers, in the context of controlling Saltwater Intrusion (SWI), represent an engineered approach involving the expansion of coastal shorelines into the sea or ocean through the deposition of artificial coastal earth-fill materials (Guo and Jiao, 2007).

Beach nourishment is a widespread coastal management technique that can not only counteract coastal erosion but also change coastal groundwater dynamics. Understanding the mechanisms and drivers of this subsurface flow and transport processes is key to predicting future groundwater resources for beach nourishment (Yu et al., 2024). Land reclamation, as a strategy commonly employed in coastal regions, serves the dual purpose of mitigating the challenges posed by increasing urbanisation and population growth while also establishing new areas for freshwater resource utilisation (Oude Essink, 2001) (see Fig. 1).

Hu and Jiao (2010) researched the physical barriers, demonstrating that implementing land reclamation in coastal areas leads to notable increases in aquifer discharge and groundwater levels. Abd-Elaty et al. (2023a) developed a numerical study for cost-effective management measures of coastal aquifers impacted by SWI and climate change; the study showed that placing aquifer fill along the shallow shoreline increases net revenues increased agricultural production, delaying aquifer salinity, protecting fresh groundwater bodies, increasing agricultural lands, supporting surface water supplies by harvesting rainfall and flash flooding, and desalinating saline water using wave energy. Amrouni et al. (2024) showed that Southern California's shoreline retreat rates for sandy beaches will increase from ~−1.45 to −2.12 m/year in 2050 and to −3.18 m/year in 2100. Moreover, the annual volume of sand required for beach nourishment could triple by 2050, increasing from the present-day amount of ~1223–~3669 m^3/year per kilometre.

The machine learning algorithms (MLA) is a good tool for solving this problem owing to the high cost of field investigation of coastal aquifers salinity. Machine learning techniques hold significant promise in addressing intricate and non-linear engineering challenges (Zanoni et al., 2022). Particularly in developing countries, ML is being harnessed to alleviate the financial and labour-intensive burden associated with investigating irrigation water quality indices (IWQI) by farmers. Based on physical parameters, it is applied to forecast and assess IWQI in groundwater systems (El Bilali et al., 2021). For better management of the salinity problem, it is essential to be able to predict the SWI using MLA (Hoaiet et al., 2022). Deep learning models promise supplements to the existing process-based mode in estuarine salinity modelling when pre-trained using augmented data (Qi et al., 2023; Ahmed et al., 2024). Also, MLA-assisted approximation seems to be a promising surrogate for the high-fidelity, variable-density model and could be utilised in multi-fidelity water resources management (Kopsiaftis et al., 2023).

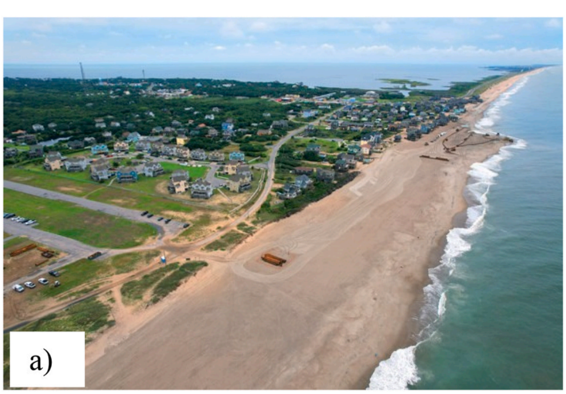
In recent years, limited studies explored machine learning (ML) techniques to assess and predict saltwater intrusion (SWI) (Nosair et al., 2022; Tran et al., 2022; Taşan et al., 2023). These investigations have employed diverse ML algorithms, including decision trees, Artificial Neural Networks (ANN), and Support Vector Machine Regression Surrogates (SVMr), to forecast the extent of SWI. Furthermore, these studies have convincingly demonstrated the potential of ML in the accurate prediction of SWI and concluded the most suitable algorithms and input variables for SWI predictive modelling (El Bilali et al., 2021; Wagh et al., 2016; Wang et al., 2020a,b).

Wagh et al. (2016) successfully applied ANN to predict groundwater quality in agricultural water processes using 13 physicochemical parameters, showcasing the efficacy and precision of ANN in IWQI prediction. Singh et al. (2021) used random forest (RF), k-nearest neighbour (k-NN), gradient boosting machine (GBM), decision tree (DT), and extra tree (ET) regression algorithms for short-term wind power forecasting and assessed them. The results show the GBM-based ensemble algorithm consistently demonstrated superior accuracy compared to the RF, k-NN, DT, and ET algorithms.

The effectiveness of ML techniques hinges on the nature and quantity of predictors used in the forecasting process. Lal and Datta (2018) developed an SVMr model to predict SWI resulting from over-pumping, revealing that enhancing the training dataset can significantly improve the SVMr models' capacity to inform the design and management of strategies in coastal regions. Furthermore, Shamshirb et al. (2019) leveraged a composite model, integrating multiple wavelet-ANNs, to enhance the performance of individual models for addressing chlorophyll and salinity concerns in coastal groundwater. Hoai et al. (2022) forecasted the SWI using ML in Vietnam. The study indicated that ANN outperformed the Multiple and Random Forest Regression for SWI prediction. A comprehensive review of the applications of machine learning applications to water resources management can be found in Ahmed et al. (2024).

Drawing upon the comprehensive literature survey presented above, it is evident that there is currently a dearth of studies employing machine learning (ML) techniques to forecast the management of Saltwater Intrusion (SWI) within coastal regions, particularly in the context of Sea Level Rise (SLR). Furthermore, there exists a notable research gap concerning the comparative evaluation of prominent ML models, such as additive regression (AR), Support Vector Machines (SVM), reduced error pruning tree (REPTree), bagging, random subspace (RSS), random forest (RF), and Artificial neural network (ANN) models, for the prediction of SWI dynamics.

The present investigation has been undertaken with the objectives of achieving precise predictions of SWI dynamics within the Biscayne aquifer located in Florida, USA, while accounting for the influences of Sea Level Rise (SLR) and groundwater decline due to over-pumping and reduction in rainfall. This research endeavour holds significance in the context of assessing groundwater salinity in coastal aquifers that are constrained by limited field data availability. The principal objectives of



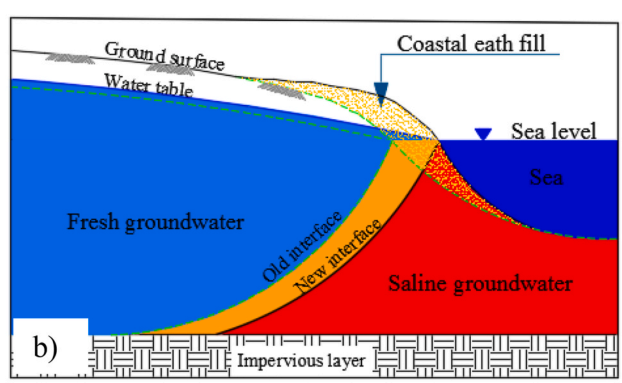


Fig. 1. a) Buxton Beach nourishment project, Carolina, USA (https://islandfreepress.org/outer-banks-news/25-of-buxton-beach-nourishment-project-complete-as-of-july-15/) and **b)** Schematic sketch shows SWI and land reclamation.

this study encompass i) The utilisation of the SEAWAT model to simulate SWI dynamics, ii) The application of ML models to forecast SWI, leveraging the outcomes derived from the SEAWAT numerical simulation, and iii) a comparative performance evaluation of diverse ML models in the prediction of SWI dynamics.

2. Material and methods

2.1. Real-world case study: Biscayne Aquifer, Florida, USA

Fig. 2a presents the main Floridan aquifer system, including the upper Floridan aquifer, a semi-confining unit of the middle confining unit located at the base of the upper aquifer, and the lower Floridan aquifer which lies below the middle confining unit (Upchurch et al., 2019). The location of the Biscayne aquifer was conducted in the Cutler Ridge area of south-eastern Florida, USA (Fig. 2a), where the current research was simulated. The water depth ranges between 1.8 m and 4 0m in the bay, except in dredged areas with depths exceeding 12 m (Caccia and Boyer, 2005). The aquifer was investigated for the SLR influence and over-pumping on SWI investigation and mitigation of aquifer salinity using land reclamation along the shoreline. The Biscayne aquifer is located in this region and is used in the current study. The aquifer has an extension of 10,000 km^2; the average length is 300 m, and the width is 615 m from the shoreline and 33 m below mean sea level (MSL) (Kohout, 1960; Kohout and Kolipinski, 1964). The saline groundwater circulates about 12.5 % to Biscayne Bay (Langevin, 2001).

The Biscayne aquifer underlies near the shore of the Atlantic Ocean and the area surrounding Biscayne Bay and extends beneath. The aquifer consists of highly permeable interbedded limestone and sandstone and covers most places only with a thin layer of porous soil (Miller, 1990). The climate in Southeast Florida is characterised by hot and humid rainy summers and mild winters; the total annual precipitation average is 1507 mm, while the average potential evapotranspiration (ETP) loss ranges from 1220 mm to 1320 mm per year (Alarcon et al., 2022).

2.2. Numerical model

In the Biscayne aquifer, a 3-D SEAWAT model was used to analyse the SWI for the baseline case, investigate the impact of SLR and over-pumping, and management scenarios; the mitigation was applied using the coastal beach nourishment (BN) effect on SWI considering fill precipitation, permeability, and width. The variable-density flow process (VDF) solves the governing equation of groundwater flow that was published by Guo and Langevin (2002). Moreover, the integrated MT3DMS transport (IMT) process was developed by Zheng and Wang (1999) to solve the advection-dispersion equation. The miscible variable-density process is governed by the following coupled system of flow and transport equations, Eqs. (1) and (2).

The Variable-Density Flow (VDF) Process solves the following variable density groundwater flow equation (Guo and Langevin, 2002):

$$\nabla\left[\rho^*\frac{\mu_o}{\mu}{}^*K_o\left(\nabla^*h_0+\frac{\rho-\rho_f}{\rho_f}{}^*\nabla Z\right)\right]=\rho^*S_{S,0}\left(\frac{\partial h_O}{\partial t}\right)+\theta^*\left(\frac{\partial\rho}{\partial C}\right)\left(\frac{\partial C}{\partial t}\right)-\rho_S{}^*q_S^{\backslash} \quad (1)$$

The Integrated MT3DMS Transport (IMT) process solves the following solute transport equation (Zheng and Wang, 1999):

$$\left(1+\frac{\rho_b{}^*K_d{}^k}{\theta}\right)\frac{\partial(\theta^*C)}{\partial t}=\nabla\left(\theta D^*\nabla C^k\right)-\nabla\left(q^*C^k\right)-\left(q_s^{\backslash}{}^*C_s{}^k\right) \quad (2)$$

where ρ_0: is the fluid density [ML^{-3}], ρ: is density of saline ground water [ML^{-3}], μ_0: is dynamic viscosity of the fresh groundwater [$ML^{-1}T^{-1}$], μ: is dynamic viscosity of saline ground water [ML^{-1} T^{-1}], K_0: is the hydraulic conductivity [LT^{-1}], h_0: is the hydraulic head [L], $S_{s,\,0}$: is the

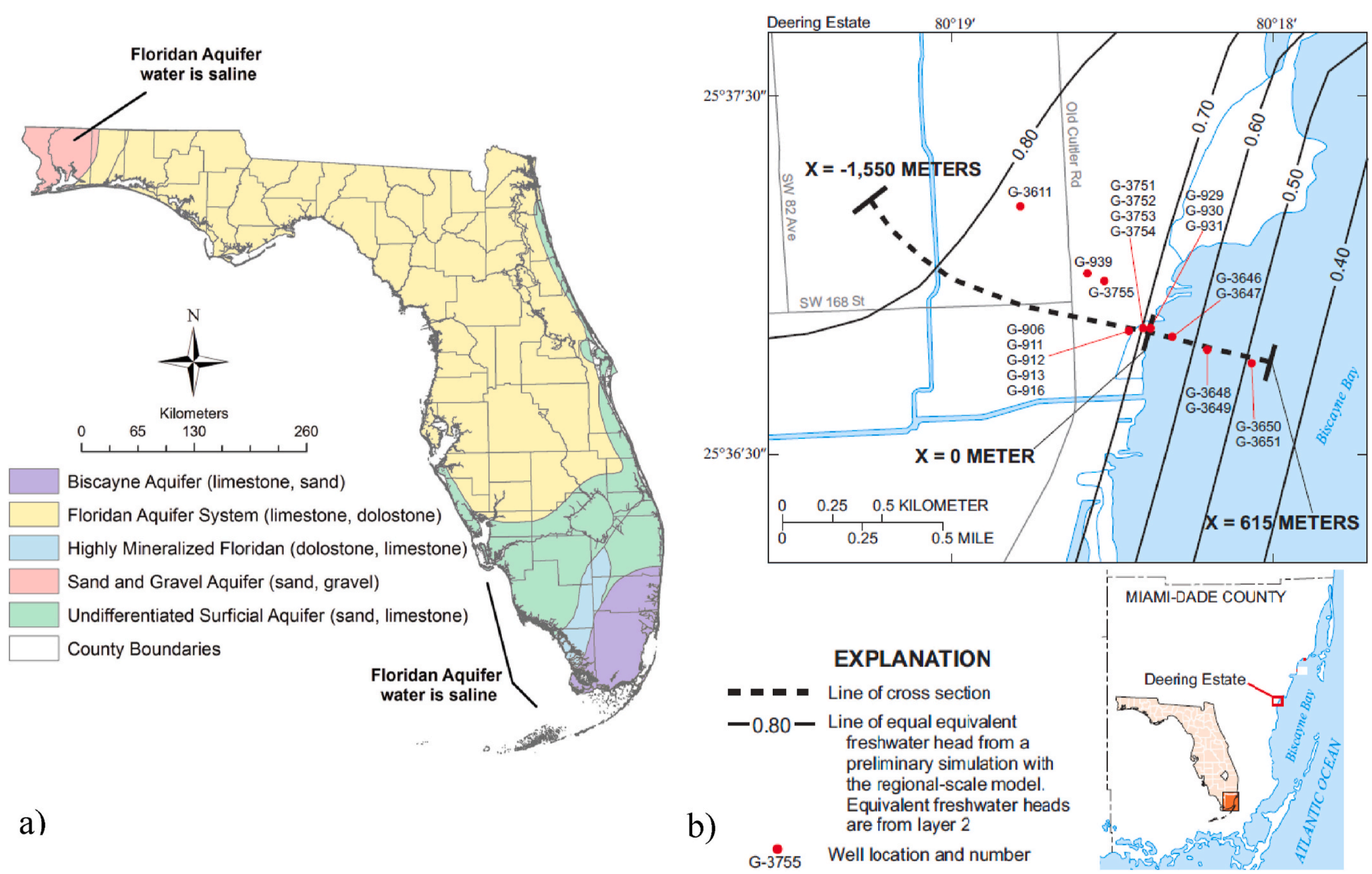


Fig. 2. Study area for (a) lithology of the principal aquifers of Florida (Upchurch et al., 2019) and (b) location at Biscayne Bay, Florida, USA (Langevin, 2001).

specific storage [L^{-1}], t:is time [T]; θ: is porosity [-]; C: is salt concentration [ML^{-3}]; and q'_s:is a source or sink [T^{-1}] of fluid with density ρ_s, ρ_b: is the bulk density [ML^{-3}], K_d^k: is the distribution coefficient of species k [$L^3 M^{-1}$], C_k: is the concentration of species k [ML^{-3}], D: is the hydrodynamic dispersion coefficient [L^2T^{-1}], q: is specific discharge [LT^{-1}], and C_s^k: is the source or sink concentration [ML^{-3}] of species k.

2.2.1. Aquifer geometry and boundary conditions

The aquifer domain was simulated using a 2D model consisting of 43 columns and 34 layers and with 1 row. The dimensions of vertical cells are 50 m and 1 m for the x and z-direction, while the cells' dimensions are 1 m for the y-direction, respectively. The model domains for the boundary conditions were set on the saline water side (right side) with a constant head boundary of 0.22 m above MSL and a salinity of 35,000 mg/l. Also, at the land boundary (left side), a specified freshwater flux was applied using 15 m^3 day^{-1} m^{-1} and a constant salinity of 1000 mg/l (Fig. 3a), the land side salinity for fresh groundwater of 0.0 mg/l an initial salinity of 1000 mg/l (Langevin, 2001).

2.2.2. Hydraulic parameters

Langevin (2001) published input model data for the hydrological and dispersion coefficients of the aquifer. Moreover, the hydraulic conductivities in vertical and horizontal directions were assigned values of 100 and 1000 m day^{-1}, respectively. The porosity of porous media is 0.20, and the freshwater and saline water densities are 1 and 1.025 g cm^{-3}, respectively. Furthermore, precipitation flow values in the study area is about 380 mm.$year^{-1}$, the longitudinal (α_L) and transverse (α_T) dispersions are 10 m and 1 m, while the aquifer diffusion coefficient (D*) is about 0 m^2 day^{-1}.

2.2.3. SWI in Biscayne aquifer

The Biscayne aquifer was calibrated using the SEAWAT code through field observation data developed by monitoring groundwater wells as presented and published by Langevin (2001) where the transition zone between saline water and the fresh groundwater was characterized using the monitoring wells which were installed both inland and offshore. Florida Geological Survey was installed the inland monitoring wells while the offshore monitoring wells were installed by U.S. Geological Survey (USGS) while the offshore wells were installed from a floating barge using the methods presented in Shinn and others (1994). The ground-water monitoring wells were located on or near the SWI line Langevin (2001). According to the solute transport modelling results, the groundwater salinity for the minimum and maximum residual reached −5552 mg/l and 388.30 mg/l, the residual mean is 347.05 mg/l, the absolute residual mean is 1738.50 mg/l, the standard error of estimation is 459.88 mg/l, root mean square (RMS) is 2085.70 mg/l, normalized RMSE is 6.13%. Also, SEAWAT results showed that the biscayne aquifer salinity reached 473 m for the isochore 17,500 mg/l, measured at the aquifer's base, as presented in Fig. 3b.

The outcomes were derived from the SEAWAT numerical simulation, followed by a comparative performance evaluation of diverse ML models in predicting SWI dynamics based on beach nourishment techniques.

2.3. ML models

In this study, we employ a diverse range of ML models to address the prediction of SWI. Each model offers unique strengths and capabilities. The goal of the present study is to rigorously evaluate their performance and determine which model is best suited for SWI prediction. The machine learning models include additive regression (AR), support vector machine (SVM), reduced error pruning tree (REPTree), Bagging, random subspace (RSS), random forest (RF), and artificial neural network (ANN). By thoroughly examining these models, the study aims to make an informed decision about the most effective ML model for our prediction endeavour.

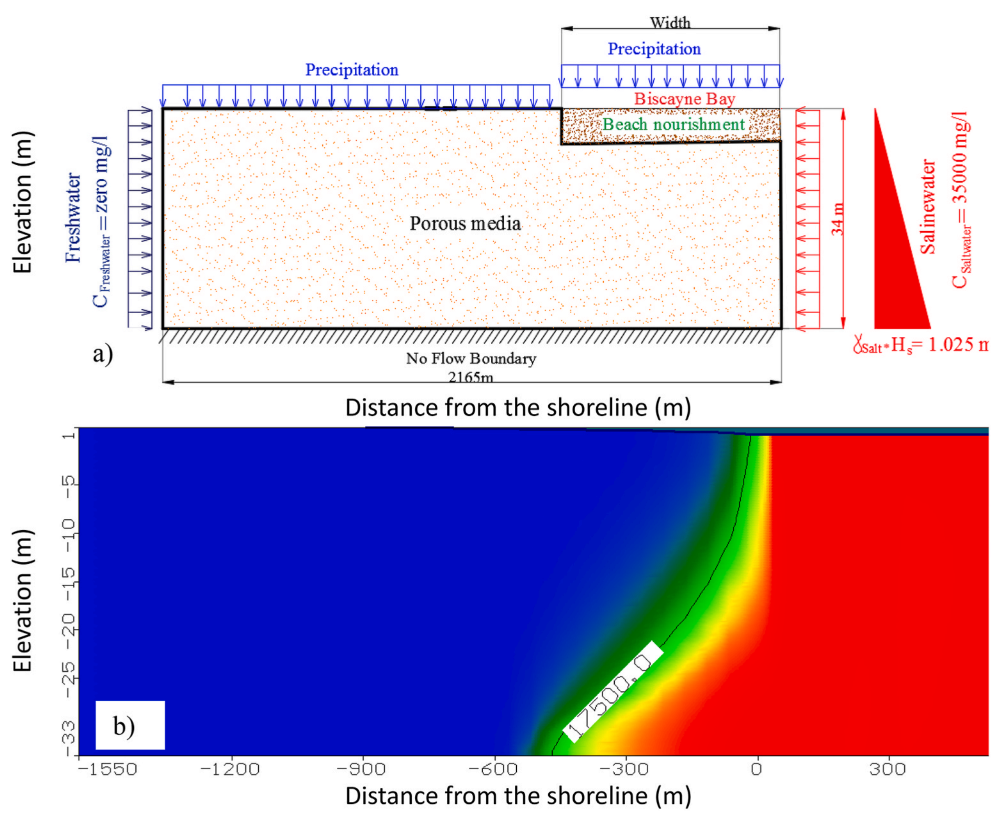


Fig. 3. Biscayne aquifer, Florida, USA for (a) Vertical section for model boundary conditions and (b) baseline SWI.

2.3.1. AR model

Friedman and Stuetzle (1981) proposed AR as a non-parametric regression approach. Unlike traditional regression, AR employs a single smoother function to explain the predictors and predictands relationship. As a result, it overcomes the curse dimensionality problem that other p-dimensional smoothers. Table 1 documented the parameters that selected during the model development. The form of the AR model is as follows:

$$E\left[y_i \middle| x_{i1}, \ldots, x_{ip}\right] = \beta_o + \sum_{j=1}^{p} f_i\left(x_{ij}\right) \tag{3}$$

Where $\sum_{j=1}^{p} f_i\left(x_{ij}\right)$ are smooth functions fitted from data and β_o is the regression coefficient. **2.3.2 REPTree model**

REPTree is a fast learning and decision-making approach introduced by Breiman *et al.* (1984). This learning happens in two different phases, i.e., constructing the tree as correctly as possible from the training dataset and pruning the tree to minimise the dependency of the learning tree on the provided dataset (Breiman et al., 1984; Salzberg, 1994). Moreover, by splitting instances into parts, it handles the missing values. The number of instances per leaf, tree depth, and training set variation can be set to minimum, maximum, and minimum, respectively, with an optimal number of pruning folds (Witten et al., 2011). It generates and prunes regression trees through backfitting by reducing/increasing the variance or error (Joseph K and Ravichandran, 2012). It is one of the easiest yet most popular and effective ML techniques for categorizing issues (Bharti et al., 2017). Table 1 presents the parameters of REPTree model.

2.3.2. SVM model

The SVM (support vector machine) is a kernel-based method that follows the Vapnik–Chervonenkis (VC) principle (Vapnik, 1998). It was originally introduced to classify binary problems using the concept of a hyperplane that divides/separates the data into suitable classes (Cristianini and Shawe-Taylor, 2000). However, it has evolved significantly and is used for function approximation and different pattern classification (Pan et al., 2009; Singh et al., 2011). Cortes and Vapnik (1995) introduced standard structure and its linear SVP and hyperplane can be given as follows:

$$x_1, y_1 \ldots\ldots x_n, y_n \tag{4}$$

$$wTx - b = 0 \tag{5}$$

where y_n will be -1 or 1, depending on which class x_n fall; x is the normal vector to the hyperplane.

Table 1
The ML models' parameters used for the prediction of SWI.

Model name	Description of parameters
RSS	Random seed = 1, Classifier = REPTree, batch size = 100, executions slots numbers = 1, subspace size numbers = 0. 5, and iteration = 10
AR	Number of iteration = 30, batch size = 100, Classifier = Bagging, and shrinkage = 1,
Bagging	Batch size-100, bag Size percent = 100, Classifier = REPTree, max depth = 0, executions slots numbers = 1, iterations numbers = 10, random seed = 1
RF	Batch size-100, bag Size percent = 100, max depth = 0, executions slots numbers = 1, iterations number = 100, random seed = 1
REPTree	Batch size = 100, random seed = 1, Initial count = 0, number of folds = 3, minimum proportion of the variance = 0.001, minimum number = 2, and max depth = 1
SVM	Kernel = Normalized Poly, Regression Optimizer = SMO Improved, Filter type = Normalize training dat, batch size = 100, C = 1, and cache size = 250,000
ANN	Batch size-100; Learning rate = 0.3, Hidden Layer = 5,6; Momentum = 0.2; Nominal to filter = True; Normalize attributes = True; Normalized Numeric Class = True

SVMs can be mapped in feature space (Fig. 4a). Using the hyperplane concept, SVM can minimise the estimation error and simultaneously generalise (avoiding overfitting) the model with reduced dimensions. Table 1 presents the parameters of the SVM model development.

2.3.3. Bagging

The Bagging, an acronym for Bootstrap Aggregating, is frequently used as an integrated (ensemble) decision tree classifier (Halmy and Gessler, 2015; Miao et al., 2012) by deploying on training samples to reduce variance (Briem et al., 2002). This integrated technique operates several independent predicting parameters and combines them with weighting or averaging (Breiman, 1996). Bagging is similar to the boosting technique in terms of ensemble decision tree classifiers; however, these two differ in essential technicality (Jafarzadeh et al., 2021). The Bagging algorithm layout is presented in Fig. 4b. Bagging creates a model based on bootstrap replicates of the given learning dataset; each is fed into a classification algorithm. The outcomes of each iteration are averaged with corresponding weights (generally with equal weights) to get an ensembled output, and classes' labels are assigned (Halmy and Gessler, 2015). Table 1 presents the parameters of the Bagging model used for prediction.

2.3.4. RSS model

The RSS ML method is an ensemble learning method that minimises correlation among predictors in an ensemble through their training via random sampling rather than a complete feature set (Fig. 4c). These representative samples are used to generate a set of decision agents (Li et al., 2011; Pham et al., 2018) and further aggregate the outcomes on a voting basis (Ho, 1998). RSS is beneficial in case of a less/limited number of learning points. It also generates quality decision agents (classifiers) when the primary dataset contains redundant points (Skurichina and Duin, 2002). RSS technique can be represented as follows:

Let $X = [x_1, x_2, \ldots, x_n]$ be a set of n numbers of independent parameters, and $Y = [Y_1, Y_2, \ldots, Y_n]$ is a set of corresponding dependent parameters in the feature dataset. For sub-setting, the feature dataset, N samples, each of a size of Z, are randomly selected with uniform distribution to ensure no replacement is needed. Each random sample expresses a subspace of X. Each subspace is deployed in the algorithm to generate a decision classifier. These classifiers are tested against the testing dataset and further aggregated/ensembled to produce most decisions. The parameters of RSS model used in the prediction are listed in Table 1.

2.3.5. RF model

Breiman proposed a tree-based ensemble learning model called RF, a supervised classification model (Breiman, 2001). RF is an improved version of the bagging algorithm to predict regression problems (Fig. 4d). To minimise the variance with maximised outcome accuracy, RF parallelly trains several decision trees over different subsets of the original learning dataset with viewpoints of sample dimension and feature dimension (Dong et al., 2020). For better generalisation and to avoid overfitting, the final decision is made via integration/aggregation of all individual trees' outcomes (Misra and Li, 2020). Further details about RF may be found in (Feng et al., 2017; Rahman and Islam, 2019). The parameters of the RF model used in the prediction are listed in Table 1.

2.3.6. Artificial neural network (ANN)

Since the early 19th century, artificial neural network models (ANN) have been utilised extensively as a "black box" model for water in the forecasting of stream flow, groundwater, water quality, water management strategy, rainfall forecasting, and reservoir management (Bagherzadeh et al., 2021b; Elbeltagi et al., 2022a; Sayadi Shahraki et al., 2021). Bahrami et al. (2016) and (Li et al. (2012) recently demonstrated that this model does not require sophisticated

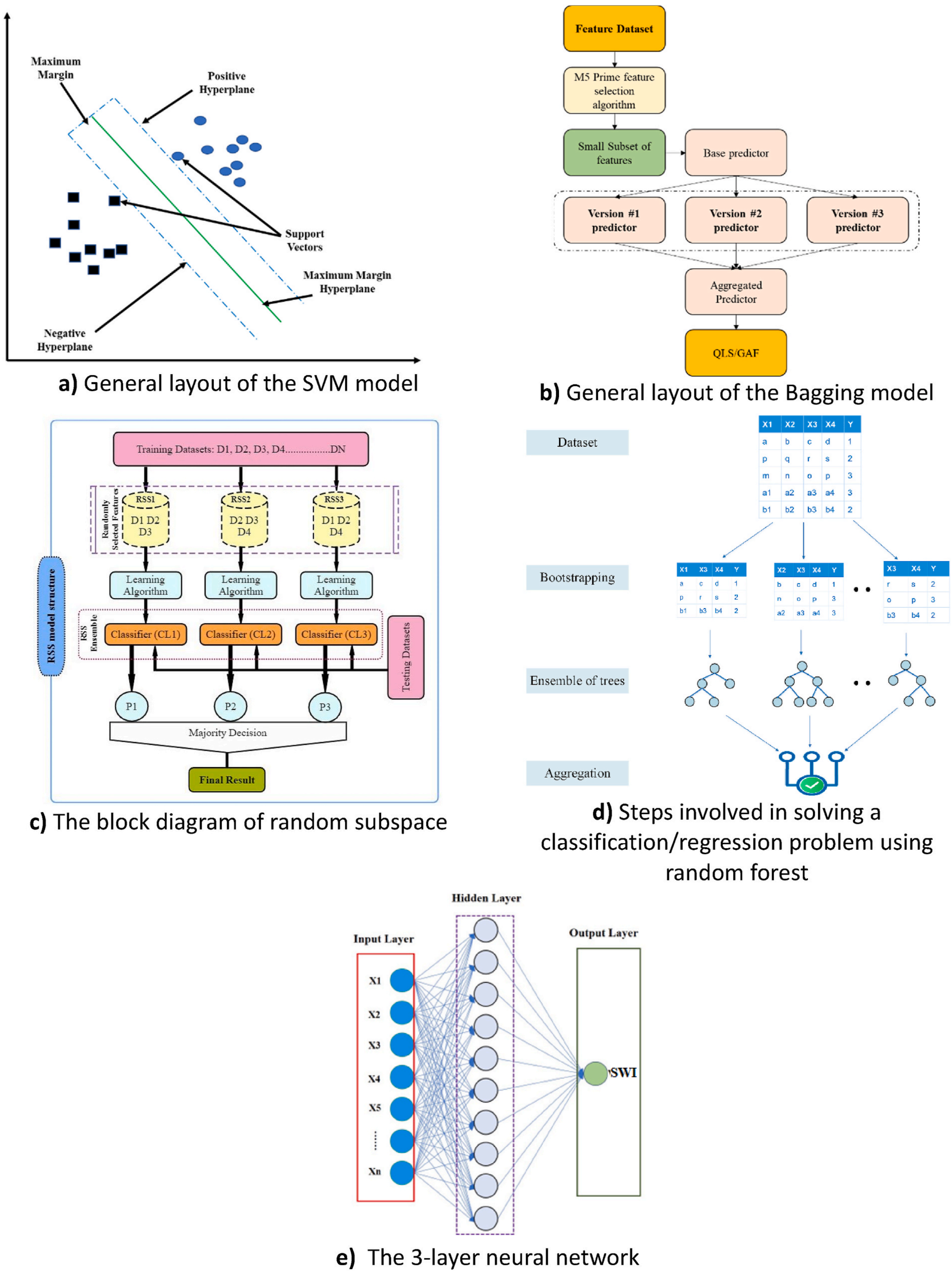


Fig. 4. The methods under consideration include AR, REPTree, SVM, Bagging, RSS, RF, and ANN.

environmental processes. Using available historical data, Ubah et al. (2021) examined the impact of physical characteristics on water distribution. Dutta et al. (2010) employed the ANN model to simulate reactive dye adsorption and photocatalysis on a TiO2 surface system. A simple 3-layer neural network is shown in Fig. 4e. Table 1 presents the parameters of developed models used in prediction.

2.4. Statistical performance criteria

The comparison between the computed datasets of SWI and the foreseen values was carried out using several statistical indicators, including the root mean square error (RMSE), coefficient of determination (R^2), mean absolute error (MAE), root relative squared error (RRSE), and Nash–Sutcliffe efficiency (NSE) (Elbeltagi et al., 2022a,b; Kushwaha et al., 2021, 2022; Pande et al., 2022). All these indicators defined as follows: N is total number of data points, SWI_A^i is real value, SWI_P^i is predicted value, and $SWI_{A\ or\ P}^{-}$ is the average value of reference or predicted samples.

The RMSE denotes the sample standard deviation of the variations between actual and predicted SWI values. The performance improves as the RMSE value decreases. A value of 0 signifies accurate prediction (Kushwaha et al., 2022). It is computed by:

$$\mathrm{RMSE} = \sqrt{\frac{1}{\mathrm{N}}\sum_{i=1}^{N}\left(SWI_A^i - SWI_P^i\right)^2} \tag{6}$$

The MAE is a commonly measures the difference between the actual and predicted values in a regression problem. The formula for MAE is calculated as the average of the absolute differences between the predicted and actual values of SWI, over all instances in the dataset. The value range of MAE is from 0 to infinity, where lower values indicate better predictive performance. A MAE value of 0 would mean that the model's predictions are perfect and exactly match the actual values (Kushwaha et al., 2021). It is defined as follows:

$$MAE = \frac{1}{\mathrm{N}}\sum_{i=1}^{N}\left|SWI_P^i - SWI_A^i\right| \tag{7}$$

The RRSE is calculated by standardising the overall squared error (SR) between the observed and predicted SWI values, and then dividing it by the total SR of predicted SWI. The RRSE index ranges from 0 to infinity, with 0 corresponding to the ideal (Kushwaha et al., 2022). The

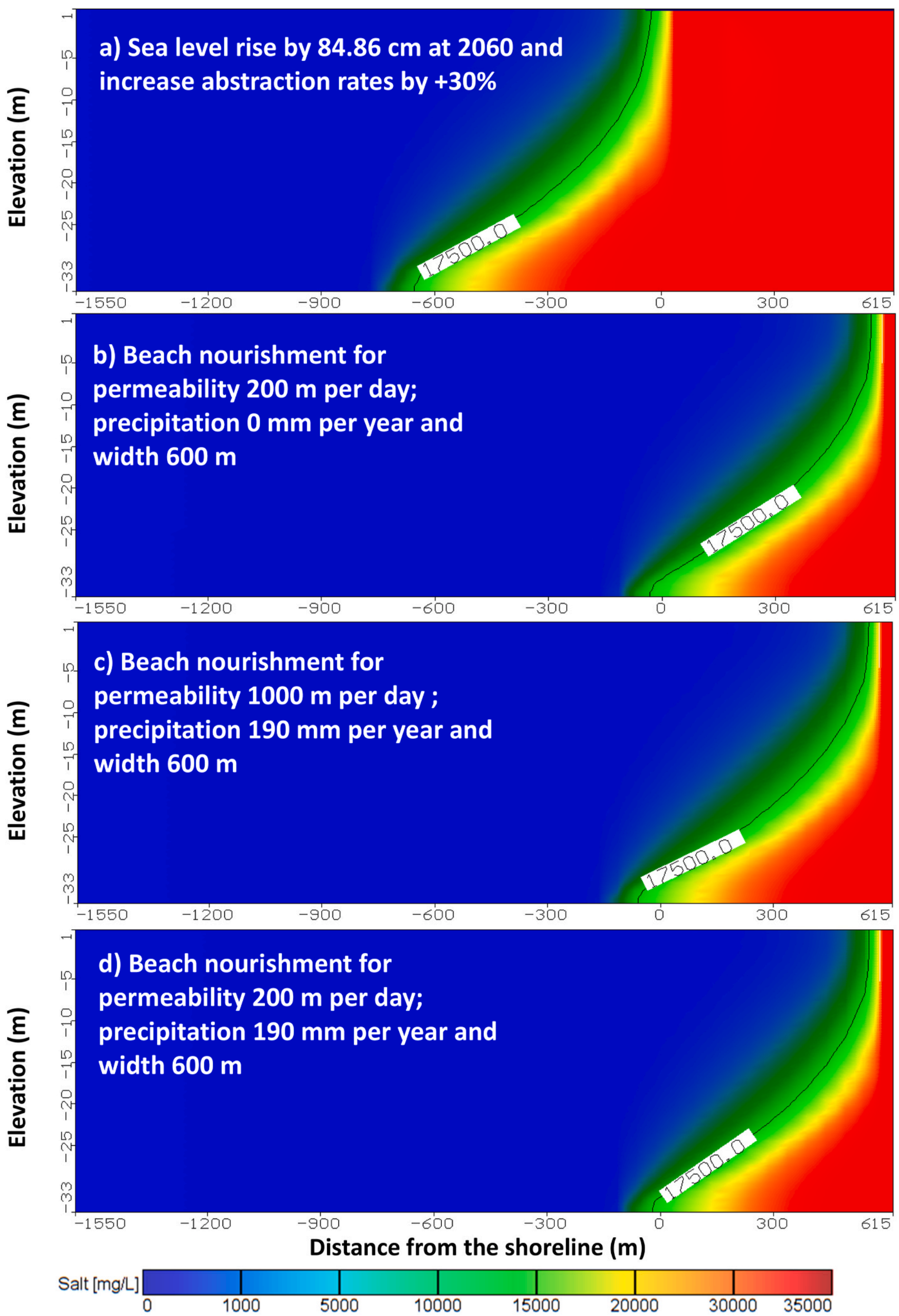


Fig. 5. Biscayne aquifer Vertical distribution of 0.5 isochores (17500 mg/l).

RRSE value is calculated as follows:

$$RRSE = \frac{\sqrt{\sum_{i=1}^{N}(SWI_P^i - SWI_A^i)^2}}{\sqrt{\sum_{i=1}^{N}(SWI_A^i - \overline{SWI_A})^2}} \tag{8}$$

The NSE is determined by subtracting the ratio of the predicted SWI error variance to the observed SWI variance from 1. It ranges from -infinity to 1, with higher values indicating better model performance (Pande et al., 2022). The NSE value closer to 1 is desirable for the best-fit model. The NSE is calculated as follows:

$$NSE = 1 - \left[\frac{\sum_{i=1}^{N}\left(SWI_A^i - SWI_p^i\right)^2}{\sum_{i=1}^{N}(SWI_A^i - \overline{SWI_A})^2}\right] \tag{9}$$

The R^2 assesses the model's capacity to predict SWI in a linear regression context, with its values falling within the range of 0–1. A value of R^2 approaching 1 is considered very good, indicating a strong predictive ability. Conversely, an R^2 value closer to 0 suggests a weaker predictive performance (Elbeltagi et al., 2022a,b). The R^2 is estimated using the following equation.

$$R^2 = 1 - \frac{\sum_{i=1}^{N}\left(SWI_A^i - SWI_p^i\right)^2}{\sum_{i=1}^{N}(SWI_A^i - \overline{SWI_A})^2} \tag{10}$$

3. Results

3.1. SWI under SLR and over-pumping

The projected saltwater intrusion (SWI) under the expected sea level rise (SLR) at Biscayne Bay was investigated to reach 84.86 cm by 2060 (Abiy et al., 2019). The numerical model was simulated for a combination of SLR of 84.86 cm and increased the abstraction rate by 30% due to the expected increase in population. Moreover, Fig. 5a presents the SWI to reach 652 m due to the isochore 17,500 mg/l measured at the aquifer bottom compared to 473 m for the baseline case. According to the findings, rising seawater levels and over-pumping enhanced the SWI in the aquifer.

3.2. Controlling SWI using beach nourishment

Aquifer salinity management was applied by using beach nourishment to check the potential of this mechanism to control SWI, considering three cases.

I. The first case involves changing the beach nourishment width. It was assigned for the simulated model in SEAWAT and extended towards the sea by changing the width from 0 m to 600 m, decreasing the fill permeability to 200 m day^{-1}, and assigning the precipitation of 0 mm year^{-1}. The SEAWAT results showed that the SWI reached 29 m, which is measured from the base of the aquifer to the shoreline (Fig. 5b).
II. The second case is changing the precipitation of beach nourishment from 0 mm year^{-1} to 190 mm year^{-1} at the width 600 m and assigned the permeability of 1000 m day^{-1}. The intrusion reached 62 m from the shoreline at the bottom of the aquifer base (Fig. 5c).
III. The third case is changing the beach nourishment permeability material from 1000 m day^{-1} to 200 m day^{-1} and precipitation of 0 mm year^{-1} to 190 mm year^{-1} at the beach nourishment of 600 m. The simulation of this combination case indicated that the SWI reached 12 m at the aquifer bottom base (Fig. 5d).

After the model simulations using different parameters of fill material properties, the beach nourishment width was simulated from 0 m to 600 by a step of 5 m at a permeability of 200 m day^{-1} and the precipitation of 190 mm year^{-1} to use these results in developing the machine learning model. The study indicated that increasing the beach nourishment widths mitigated the SWI and is considered a good tool for managing the influence of SLR. The SEAWAT results for mitigation of SWI in the current study area are used to develop ML models.

3.3. Machine-learning-based SWI prediction

The seven ML models such as additive regression (AR), support vector machine (SVM), reduced error pruning tree (REPTree), Bagging, random subspace (RSS), random forest (RF), artificial neural network (ANN) were developed and evaluated for the prediction of SWI in Biscayne aquifer, Florida, USA. To attain good performance in the training and testing phases, all ML models undergo extensive training using a trial-and-error method. Table 2 exhibits the ML statistical indices of model performance during training and testing. Fig. 6 depicts the scatter plots of observed and predicted SWI during the testing phase.

According to the performance indices, the AR model showed average performance during both training and testing phases and has statistical performance indicators viz., R^2, NSE, MAE, RMSE, and RRSE as 0.965, 0.964, 27.006, 34.041 and 18.86 %, respectively during training and 0.960, 0.959, 30.723, 40.051 and 20.065%, respectively, during the testing period (Table 2). The AR model failed to capture the temporal pattern of observed SWI values. The scatter plot in Fig. 6 shows the straight dots, which means the model predicted the same or near the same value of SWI in various occurrences. Thus, the model performance is unsatisfactory even though the performance indices are in a good range in the training and testing phases. The SVM model performed well and was identical in the training and testing phases (Table 2). The plots in Fig. 6 show the SVM model has well captured the SWI temporal pattern. However, the model underpredicted the peak values in both the training and testing periods (Fig. 6). Like the AR model, the REPTree model failed to match the temporal pattern and predicted the same SWI values in a few occurrences in the training and testing phases (Fig. 6). Also, the model performance is high in the training phase compared to the testing phase (Table 2). It implies the over-fitting nature of the REPTree model, even after extensive parameter tuning. However, the REPTree model performance is higher than the AR model but less than the SVM model (Table 2).

Bagging and RSS models' performances are excellent in predicting the SWI values in the training and testing periods (Table 2) and matched the temporal pattern of observed values well. Scatter plots of Bagging and RSS models in Fig. 6 revealed that the values are precisely aligned with the 1:1 line with few deviations in the models' predictions. The RSS model performance is comparatively higher than the Bagging model in the training phase; however, the Bagging model performance is higher in the testing phase than the RSS model (Table 2). The RF model performance is excellent in predicting the SWI values, and it is better than the Bagging, ANN and RSS models in both training and testing periods (Table 2). Similarly, the RF model also captured the temporal pattern of SWI values well. However, there are no considerable deviations from the 1:1 line in the scatter plot of the RF compared to the Bagging and RSS models (Fig. 6). It recorded the values of the statistical performance indices as 0.999, 0.999, 0.362, 0.440, and 0.243 %, respectively, for R2, NSE, MAE, RMSE, and RRSE during model training and 0.999. 0.999, 0.324, 0.416, and 0.209 %, respectively, during the model testing period. Furthermore, from the statistical indices like MAE, RMSE, and RRSE, it is evident that the model performance is high in the training phase compared to the testing phase. That means the RF model's over-fitting nature was not eliminated entirely, even after the extensive parameter tuning.

Fig. 7 compares the statistical performance parameters for the developed models. It is clear from Fig. 7a and b that the RF algorithms predicted SWI precisely compared to the other algorithms, had the

Table 2
Statistical indices of ML models in training and testing phases.

Statistical Indicator	AR	SVM	REPTree	Bagging	RSS	RF	ANN
Training							
R^2	0.965	0.999	0.997	0.998	0.996	0.999	0.997
NSE	0.964	0.999	0.997	0.997	0.996	0.999	0.996
MAE	27.006	2.169	6.866	4.422	3.954	0.362	3.569
RMSE (m)	34.041	2.684	8.480	6.353	5.264	0.440	4.430
RRSE (%)	18.86	1.487	4.699	3.519	2.917	0.243	2.455
Testing							
R^2	0.960	0.999	0.995	0.998	0.998	0.999	0.999
NSE	0.959	0.998	0.995	0.997	0.997	0.999	0.998
MAE	30.723	4.633	12.278	5.903	6.431	0.324	3.038
RMSE (m)	40.051	5.585	13.839	7.361	8.021	0.416	3.791
RRSE (%)	20.065	2.798	6.934	3.688	4.049	0.209	1.899

lowest RMSE, MAE, and RRSE values, and had high values of R2 and NSE. However, the AR model failed to predict SWI and showed high error metrics.

Although the above outcomes demonstrate the performance evaluation of all ML models, it is better to summarise the comparison results through the Taylor diagram using the most accurate input structures obtained from the used evaluation criteria. The Taylor diagram presents a high-performance model concerning the observed data in a two-dimensional plot, which includes standard deviation on the y-axis and the RMSE on the radial axis. Taylor diagrams are presented in Fig. 8a and b for predicting SWI values using the ML models in training and testing phases, respectively. Fig. 8a and b shows that the SVM, Bagging, RSS, ANN and RF are superior to the other ML models as they are very near to the reference point in both the training and testing phases. Although the RF model is slightly over-fitted, it is still the best out of the four because it closely matched observed values with the fewest deviations.

4. Qualification of machine learning in the prediction of SWI

4.1. Cost-effective beach nourishment

The SWI is predicted using ML for beach nourishment of 615 m with fill permeability 200 m day^{-1} and precipitation of 190 mm year^{-1}, and the intrusion reached −10m from the shoreline. The cost-effective of using the beach nourishment was discussed by Abd-Elaty et al. (2022); the cost of 0.50 m depth of fill reaching $ 33,387 per acre ($5081.92 per unit meter length), the agriculture $ 790 per acre ($120.25 per unit meter length), the irrigation water cost $100 per acre ($15.22 per unit meter length). The yield reached $1950 per acre ($296.81 per unit meter length), and the precipitation water harvesting by the fill reached 116.85 m^3 year^{-1} For 190 mm year^{-1} ($70.11 per unit meter length). The total revenue for using beach nourishment for a width of 615 m reached ($231.45 per unit meter length, as presented in Fig. 9.

4.2. Practical implications of the present study

Saltwater intrusion (SWI) is a critical issue with multifaceted implications for both the environment and society. In coastal regions, SWI poses a substantial threat to potable water sources and ecosystems, making it an issue of paramount concern. The following points elucidate the practical significance of our findings:

Preservation of Potable Water Sources: SWI can lead to groundwater contamination, which serves as a vital source of drinking water for coastal communities. As sea levels rise, the intrusion of saline water into freshwater aquifers becomes exacerbated. Our research, particularly the accuracy of the Random Forest (RF) model, equips these communities with a tool for early detection and proactive management of SWI. This is instrumental in safeguarding the quality and availability of potable water, reducing the risks associated with salinity for public health.

Agricultural Sustainability: Coastal regions often support agricultural activities that are dependent on freshwater resources. SWI can adversely affect soil salinity, diminish crop yields and threaten food security. The soil salinisation will occur by abstraction and using the groundwater in the irrigation process through the production wells. Climate change caused by rising sea levels and over-pumping from coastal aquifers has led to an increase in coastal groundwater salinisation and soil salinity through agricultural activities. Mazhar et al. (2022) studied the impacts of salinisation caused by SLR on the biological processes of coastal soils. The study showed that the effects of SLR on C and N cycles and, consequently, on GHG emissions from coastal soils are highly variable, depending on the contrasting and concomitant effects of flooding and salinity. Abd-Elaty et al. (2023b) studied the impact of SLR on groundwater salinity of the Nile Delta, Egypt. The results showed that with the projected SLR of 61 cm by 2100, there was an overall salt volume increment of 3 %. Also, Abd-Elaty et al. (2024) showed the hazards of SLR and dam projects built on the downstream River Nile water budget and the salinity of the Nile Delta aquifer. The study showed that the Grand Ethiopian Renaissance Dam reservoir filling could alter the freshwater, in which the aquifer salinity increased by 29.99% for SLR by 100 cm, increasing abstraction rates by 100% and filling the reservoir at elevations 645 m. By employing accurate SWI predictions, farmers can make informed decisions regarding irrigation practices and crop selection, thereby enhancing agricultural sustainability and reducing economic losses.

Infrastructure Resilience: Infrastructure in coastal areas, including roads, buildings, and utilities, is vulnerable to the corrosive effects of SWI. Accurate SWI prediction enables better planning and design of infrastructure projects, considering future sea-level rise scenarios. This, in turn, contributes to increased resilience against the detrimental impacts of saltwater on critical infrastructure components.

Environmental Conservation: Ecosystems in coastal regions are intricately linked to freshwater availability. SWI can disrupt these delicate balances, leading to habitat degradation and loss of biodiversity. Informed conservation efforts contribute to the overall health and sustainability of coastal environments.

Policy Formulation: Accurate SWI predictions can inform decisions regarding land use and sustainable water management practices.

There are many practical implications of this study. Accurate SWI predictions are crucial for coastal communities. The RF model's precision empowers these communities to proactively plan for and mitigate the impacts of salinity on agriculture, drinking water supplies, and infrastructure. The study also supports the effective allocation and conservation of freshwater resources along coastlines. This ensures sustainable agricultural practices and reduces risks associated with SWI, benefiting both the environment and local economies. Policymakers can leverage these findings to enact measures that address SWI. Informed policies on land use, agriculture, and infrastructure can enhance environmental resilience and economic stability in coastal regions. Industries reliant on freshwater resources, such as agriculture and aquaculture, can better plan and adapt with the knowledge provided by accurate SWI predictions. This reduces economic losses due to

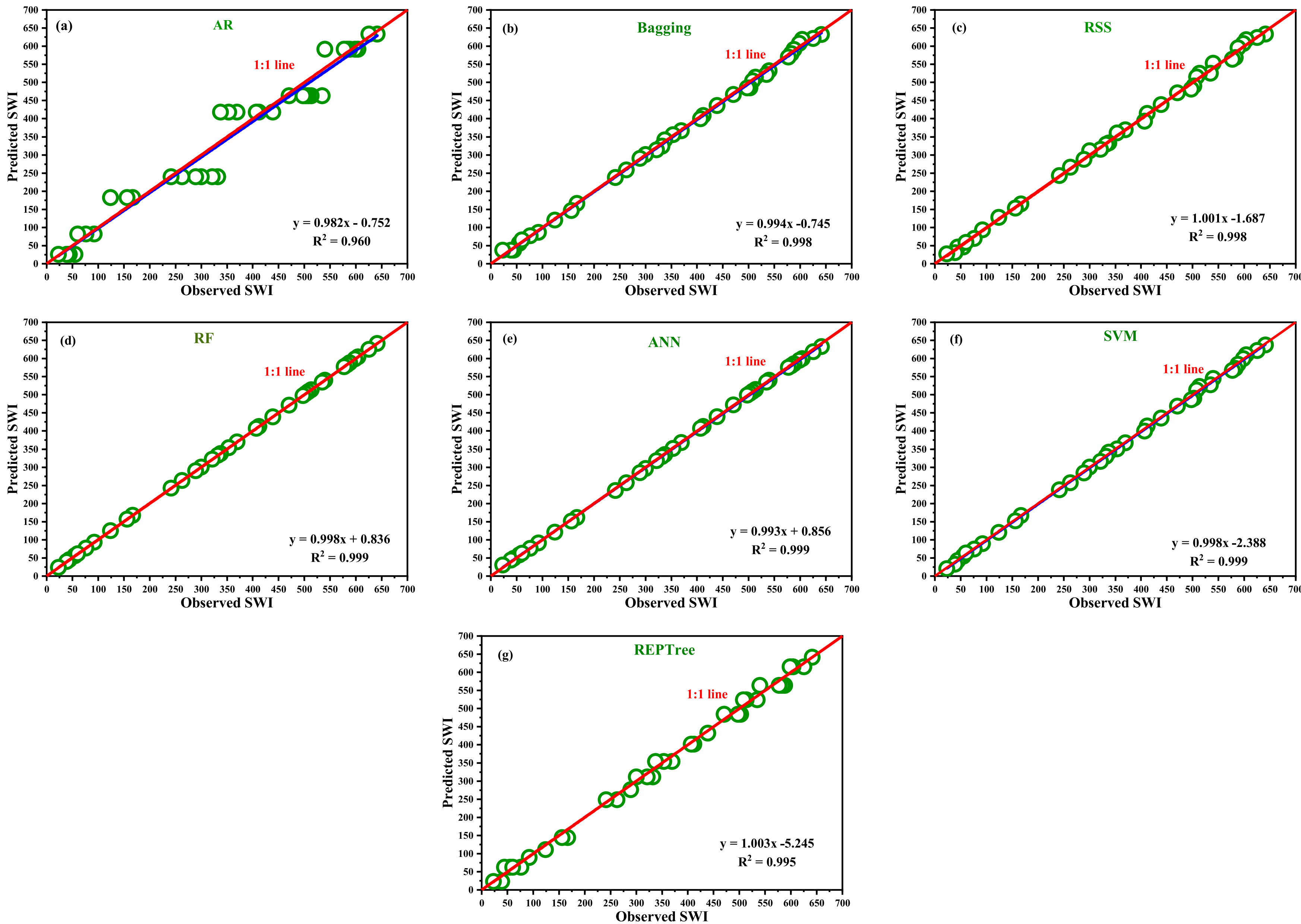


Fig. 6. Scatter plots of observed and predicted SWI using the ML models during testing period.

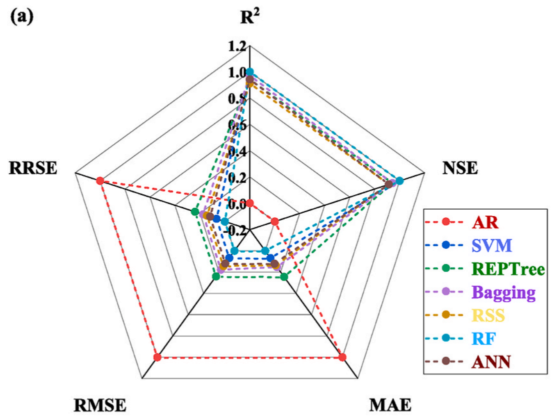


Fig. 7. Radar chart for the best RMSE values (a) the training data and (b) the testing data set.

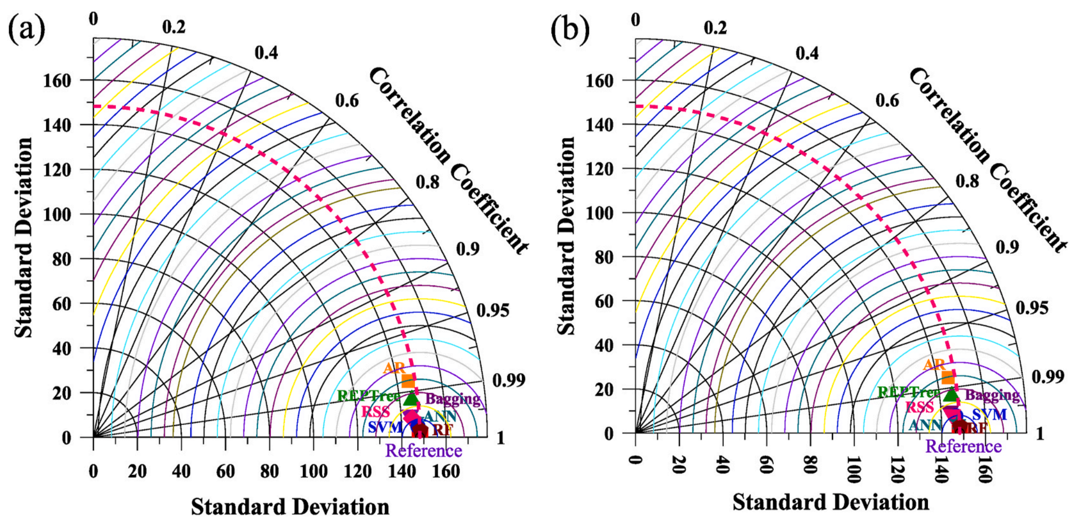


Fig. 8. Taylor diagram for (a) the training data and (b) the testing data set.

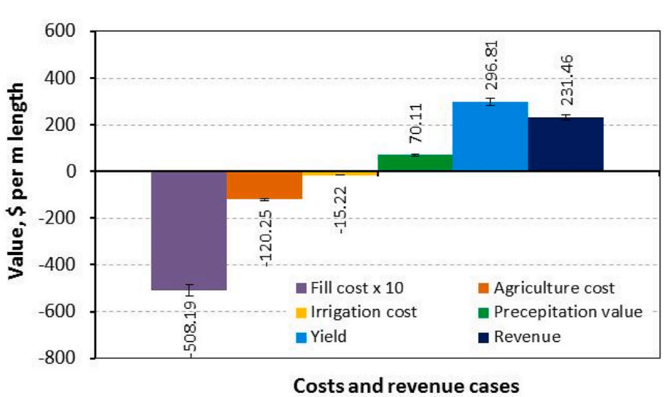


Fig. 9. Costs and revenue for beach nourishment of 615 m in the Biscayne regions.

unexpected shifts in salinity levels.

5. Discussion

Predicting the SWI values is necessary for groundwater management in coastal areas. The comparative performance of additive regression (AR), support vector machine (SVM), reduced error pruning tree (REPTree), Bagging, random subspace (RSS), random forest (RF), artificial neural network (ANN) model revealed that the SVM, Bagging, RSS, ANN and RF models have a great potential in predicting the SWI values with the limited data in the study area. The performance ranking of these four models for the study area is RF > ANN > SVM > RSS > Bagging. Furthermore, the Random Forest (RF) model emerged as the best-fit model out of the seven applied machine learning models in terms of lower values of error metrics and higher values of R2 and NSE for SWI prediction.

The random feature selection at each split helps diversify the learning process, preventing any single feature from dominating the model's predictions. Additionally, the study employs thorough cross-validation and hyperparameter tuning, ensuring the model generalises well to unseen data. With a substantial and diverse dataset, RF is well-equipped to handle the complexity of the prediction task without overfitting. Its advantages include high predictive accuracy, robustness to non-linear relationships, interpretability through variable importance analysis, and resilience in handling missing data, making it an excellent choice for this prediction task. The AR and REPTree, models do not capture the temporal pattern of the observed SWI. This could be attributed to the unavailability of long-time series data for effective training. The AR and REPTree might be more sensitive to the limited data. From the statistical indices (Table 2) and scatter plots (Fig. 6), the RF model is judged as the best ML model to predict the SWI.

Based on Table 3, Pham et al. (2022) used multiple linear regression, RF, and ANN to predict the SWI and concluded that the ANN model is better than the other models. However, the reported NSE value of ANN

Table 3
Summary of different studies on machine learning-based SWI prediction in different parts of world.

Study area	Machine learning algorithms	Statistical performance indicators	Best model and accuracy	Remarks	Reference
Ham Luong River, Ben Tre Province	Multiple Linear Regression (MLR), Random Forest Regression (RFR), Artificial Neural Networks (ANN)	NSE, RMSE, and MAE	ANN model (NSE = 0.842, RMSE = 1.16, MAE = 0.11)	The ANN model showed better result rather than other algorithms	Pham et al. (2022)
Pearl River Delta	Bayesian model averaging (BMA), random forest (RF), support vector machine (SVM), and Elman neural network (ENN)	NSE and percentage of bias (Pbias)	BMA (NSE = 0.79, Pbias = 11.4)	The BMA approach outperformed the individual models (i.e., RF, ENN, and SVM)	Lin et al. (2019)
Coastal aquifer	support vector machine regression (SVMr), genetic programming (GP)	RMSE, mean square error (MSE), relative error (RE), correlation coefficient (Cc), and NSE	SVMr (Cc = 0.996, NSE = 0.99, MSE = 0.149 and RMSE = 0.386)	The SVMr is superior to GP models	Lal and Datta (2018)
Ping Gang water source of Zhuhai city	ANN, alog with Real coding based accelerating genetic algorithm (RAGA) and Back propagation (BP)	NA	NA	The BP-RAGA coupled neural network model proved to be superior to BP neural network model in precision	Dong et al. (2010)

was 0.842, which is less than the presented study. Lin et al. (2019) used Bayesian Model Averaging (BMA) method to ensemble the RF, SVM, and Elman Neural Network (ENN) models for the prediction of SWI. They concluded that the RF model performance is higher than the SVM model. These findings are consistent with the current study. Dong et al. (2010) introduced a novel neural network model known as BP-RAGA (Back-propagation-Real Coding Accelerated Genetic Algorithm). This model was specifically designed to forecast saltwater intrusion in the Ping Gang water source of Zhuhai City. It incorporated essential variables such as the tidal range of the water resources and the observed flows from the upstream hydrological station for the preceding day as key influencing factors in its predictions. To develop an integrated forecast of SWI, Lu et al. (2021) used the BMA approach to merge the forecasting outputs of the RF, SVM, and ENN models. However, the highest NSE value of the proposed method in forecasting is 0.78, which is less than the presented study.

Using input-output patterns from a numerical simulation model, Lal and Datta (2018) employed the SVM model to forecast SWI in a coastal aquifer. The SVM model performance in their study is identical to our model results, with an NSE value of 0.99. Another study (Bagherzadeh et al., 2021a) evaluated seven feature selection methods to enhance total nitrogen (TN) prediction using ML in wastewater treatment plants for efficiency and cost reduction. Random Forest (RF) and Gradient Boosting Machine (GBM) outperformed Artificial Neural Network (ANN), with GBM excelling in generalising patterns on unseen data, showcasing its effectiveness for wastewater component prediction.

The present study also revealed that the RF model is superior to AR, SVM, REPTree, Bagging, RSS, and ANN in predicting SWI under data scarcity. Although we cannot directly compare the model performances with the existing literature due to the varying conditions of the study areas, the results presented in this study align with those of other studies for predicting SWI.

6. Conclusions

The study focuses on the importance of using beach nourishment for saltwater intrusion (SWI) management in the coastal areas as it affects the fresh groundwater especially with sea level rises and over-pumping in the Biscayne aquifer, USA. Therefore, the SEAWAT model was applied to simulate SWI dynamics and develop ML models to forecast SWI, using different models such as additive regression (AR), support vector machine (SVM), reduced error pruning tree (REPTree), Bagging, random subspace (RSS), random forest (RF), and artificial neural network (ANN). The results demonstrated that using beach nourishment, such as landfill width, permeability, and precipitation, can effectively mitigate SWI. Beach nourishment showed the most promising results, suggesting their practical application for managing SWI. The comparative evaluation of seven machine learning models for SWI prediction indicated that the RF model emerged as the best-fit model with statistical measures as 0.999, 0.999, 0.362, 0.440, and 0.243, respectively, for R^2, NSE, MAE, RMSE, and RRSE during the model training phase and 0.999. 0.999, 0.324, 0.416, and 0.209, respectively during model testing phase. The ability of the RF model to capture SWI patterns and predict values with minimal deviations makes it a valuable tool for SWI forecasting. However, the performance of the ANN, SVM, Bagging and RSS models was found to be satisfactory. This research provides practical solutions for managing SWI, emphasises the importance of accurate predictions, and highlights the cost-effectiveness of beach nourishment measures. The findings have significant implications for coastal communities, industries, and policymakers dealing with the challenges of SWI.

Funding

(This study did not receive any funding)

Availability of data and material

(Upon request)

Code availability

(Upon request)

Ethics approval

(Not applicable)

Consent to participate

(Yes)

Consent for publication

(Yes)

CRediT authorship contribution statement

N.L. Kushwaha: Conceptualization, Formal analysis, Investigation, Methodology, Software, Writing – review & editing. **Kallem Sushanth:** Conceptualization, Formal analysis, Resources, Software, Validation. **Abhishek Patel:** Data curation, Formal analysis, Resources, Supervision, Validation. **Ozgur Kisi:** Supervision, Writing – review & editing. **Ashraf Ahmed:** Funding acquisition, Supervision, Writing – review & editing. **Ismail Abd-Elaty:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgment

The authors thank the Department of Water and Water Structures Engineering, Faculty of Engineering, Zagazig University, Zagazig 44519, Egypt, for constant support during the study.

References

Abd-Elaty, I., Kushwaha, N.L., Grismer, M.E., Elbeltagi, A., Kuriqi, A., 2022. Cost-effective management measures for coastal aquifers affected by saltwater intrusion and climate change. Sci. Total Environ. 836, 155656.
Abd-Elaty, I., Abdoulhalik, A., Ahmed, A., 2023a. The impact of future hydrology stresses and climate change on submarine groundwater discharge in arid regions: a case study of the Nile Delta aquifer, Egypt. J. Hydrol.: Reg. Stud. 47, 101395.
Abd-Elaty, I., Sallam, G.A., Pugliese, L., Negm, A.M., Straface, S., Scozzari, A., Ahmed, A., 2023b. Managing coastal aquifer salinity under sea level rise using rice cultivation recharge for sustainable land cover. J. Hydrol.: Reg. Stud. 48, 101466.
Abd-Elaty, I., Kuriqi, A., Ramadan, E.M., Ahmed, A.A., 2024. Hazards of sea level rise and dams built on the River Nile on water budget and salinity of the Nile Delta aquifer. J. Hydrol.: Reg. Stud. 51, 101600.
Abdelgawad, A.M., Abdoulhalik, A., Ahmed, A.A., Moutari, S., Hamill, G., 2018. Transient investigation of the critical abstraction rates in coastal aquifers: numerical and experimental study. Water Resour. Manag. 32, 3563–3577.
Abdoulhalik, A., Abdelgawad, A.M., Ahmed, A.A., 2020. Impact of layered heterogeneity on transient saltwater upconing in coastal aquifers. J. Hydrol. 581, 124393.
Abdoulhalik, A., Ahmed, A.A., Abd-Elaty, I., 2024. Effects of layered heterogeneity on mixed physical barrier performance to prevent seawater intrusion in coastal aquifers. J. Hydrol. 637, 131343.
Abiy, A.Z., Melesse, A.M., Abtew, W., Whitman, D., 2019. Rainfall trend and variability in Southeast Florida: implications for freshwater availability in the Everglades. PLoS One 14, e0212008. https://doi.org/10.1371/journal.pone.0212008.
Ahmed, A.A., Sayed, S., Abdoulhalik, A., Moutari, S., Oyedele, L., 2024. Applications of machine learning to water resources management: a review of present status and future opportunities. J. Clean. Prod., 140715
Alarcon, V.J., Linhoss, A.C., Mickle, P.F., Kelble, C.R., Fine, A., 2022. Estimation of groundwater and salinity for the central Biscayne bay coast, Florida, USA. In: Gervasi, O., Murgante, B., Misra, S., Rocha, A.M.A.C., Garau, C. (Eds.), Computational Science and its Applications – ICCSA 2022 Workshops, Lecture Notes in Computer Science, vol. 13379. Springer, Cham. https://doi.org/10.1007/978-3-031-10545-6_40. ICCSA 2022.
Amrouni, O., Heggy, E., Hzami, A., 2024. Shoreline retreat and beach nourishment are projected to increase in Southern California. Commun Earth Environ 5, 274. https://doi.org/10.1038/s43247-024-01388-6.
Bagherzadeh, F., Mehrani, M.-J., Basirifard, M., Roostaei, J., 2021a. Comparative study on total nitrogen prediction in wastewater treatment plant and effect of various feature selection methods on machine learning algorithms performance. J. Water Process Eng. 41, 102033. https://doi.org/10.1016/j.jwpe.2021.102033.
Bagherzadeh, F., Nouri, A.S., Mehrani, M.-J., Thennadil, S., 2021b. Prediction of energy consumption and evaluation of affecting factors in a full-scale WWTP using a machine learning approach. Process Saf. Environ. Prot. 154, 458–466. https://doi.org/10.1016/j.psep.2021.08.040.
Bahrami, S., Doulati Ardejani, F., Baafi, E., 2016. Application of artificial neural network coupled with genetic algorithm and simulated annealing to solve groundwater inflow problem to an advancing open pit mine. J. Hydrol. 536, 471–484. https://doi.org/10.1016/j.jhydrol.2016.03.002.
Bharti, B., Pandey, A., Tripathi, S.K., Kumar, D., 2017. Modelling of runoff and sediment yield using ANN, LS-SVR, REPTree and M5 models. Nord. Hydrol 48 (6), 1489–1507. https://doi.org/10.2166/nh.2017.153.
Breiman, L., 1996. Bagging predictors. Mach. Learn. 24 (2), 123–140. https://doi.org/10.1007/BF00058655.
Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32. https://doi.org/10.1023/A:1010933404324.
Breiman, L., Friedman, J.H., Olshen, R.A., Stone, C.J., 1984. Classification and Regression Trees. CRC Press, Boca Raton, FL, USA.
Briem, G.J., Benediktsson, J.A., Sveinsson, J.R., 2002. Multiple classifiers applied to multisource remote sensing data. IEEE Trans. Geosci. Rem. Sens. 40 (10), 2291–2299. https://doi.org/10.1109/TGRS.2002.802476.
Cortes, C., Vapnik, V., 1995. Support-vector networks. Mach. Learn. 20 (3), 273–297. https://doi.org/10.1007/BF00994018.
Cristianini, N., Shawe-Taylor, J., 2000. An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods. Cambridge University Press. https://doi.org/10.1017/CBO9780511801389.
Dong, X., Tao, T., Liu, S., Xia, Y., Yu, Y., 2010. Prediction of salt water intrusion using BP-Raga coupled neural network model. In: 2010 3rd International Congress on Image and Signal Processing. Presented at the 2010 3rd International Congress on Image and Signal Processing, pp. 4244–4248. https://doi.org/10.1109/CISP.2010.5646891.
Dong, X., Yu, Z., Cao, W., Shi, Y., Ma, Q., 2020. A survey on ensemble learning. Front. Comput. Sci. 14 (2), 241–258. https://doi.org/10.1007/s11704-019-8208-z.
Dutta, S., Parsons, S.A., Bhattacharjee, C., Bandhyopadhyay, S., Datta, S., 2010. Development of an artificial neural network model for adsorption and photocatalysis of reactive dye on TiO2 surface. Expert Syst. Appl. 37, 8634–8638. https://doi.org/10.1016/j.eswa.2010.06.090.
El, Bilali A., Taleb, A., Brouziyne, Y., 2021. Groundwater quality forecasting using machine learning algorithms for irrigation purposes. Agric. Water Manag. 245, 106625. https://doi.org/10.1016/j.agwat.2020.106625. ISSN 0378-3774.
Elbeltagi, A., Kumar, M., Kushwaha, N.L., Pande, C.B., Ditthakit, P., Vishwakarma, D.K., Subeesh, A., 2022a. Drought indicator analysis and forecasting using data driven models: case study in Jaisalmer, India. Stoch. Environ. Res. Risk Assess. https://doi.org/10.1007/s00477-022-02277-0.
Elbeltagi, A., Kushwaha, N.L., Srivastava, A., Zoof, A.T., 2022b. Chapter 5 - artificial intelligent-based water and soil management. In: Poonia, R.C., Singh, V., Nayak, S.R. (Eds.), Deep Learning for Sustainable Agriculture, Cognitive Data Science in Sustainable Computing. Academic Press, pp. 129–142. https://doi.org/10.1016/B978-0-323-85214-2.00008-2.
Feng, Y., Cui, N., Gong, D., Zhang, Q., Zhao, L., 2017. Evaluation of random forests and generalized regression neural networks for daily reference evapotranspiration modelling. Agric. Water Manag. 193, 163–173.
Guo, W., Langevin, C.D., 2002. User's guide to SEAWAT: a computer program for simulation of three-dimensional variable-density groundwater flow. Techniques of Water-Resources Investigations Book 6, p. 77 (Chapter 7).
Halmy, M.W.A., Gessler, P.E., 2015. The application of ensemble techniques for land-cover classification in arid lands. Int. J. Rem. Sens. 36 (22), 5613–5636.
Ho, T.K., 1998. The random subspace method for constructing decision forests. IEEE Trans. Pattern Anal. Mach. Intell. 20 (8), 832–844. https://doi.org/10.1109/34.709601.
Hoai, Pham Ngoc, Quoc, Pham Bao, Thanh, T.H.A.I., 2022. Tran. Apply machine learning to predict saltwater intrusion in the ham luong river, ben tre province. VNU Journal of Science: Earth and Environmental Sciences, [S.l.] 38 (3). https://doi.org/10.25073/2588-1094/vnuees.4852. ISSN 2588-1094. Available at: https://js.vnu.edu.vn/EES/article/view/4852. (Accessed 13 December 2022).
Hu, L., Jiao, J.J., 2010. Modeling the influences of land reclamation on groundwater systems: a case study in Shekou peninsula, Shenzhen, China. Eng. Geol. 114 (3), 144–153. https://doi.org/10.1016/j.enggeo.2010.04.011.
IPCC, 2007. An Assessment of the Intergovernmental Panel on Climate Change, Adopted Section by Section at IPCC Plenary XXVII (Valencia, Spain, 12-17 November 2007), Represents the Formally Agreed Statement of the IPCC Concerning Key Findings and Uncertainties Contained in the Working Group Contributions to the Fourth Assessment Report.
IPCC, 2014. Climate change 2014: synthesis report. In: Pachauri, R.K., Meyer, L.A. (Eds.), Contribution of Working Groups I, II and III to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Core Writing Team. IPCC, Geneva, Switzerland, p. 151.
Jafarzadeh, H., Mahdianpari, M., Gill, E., Mohammadimanesh, F., Homayouni, S., 2021. Bagging and boosting ensemble classifiers for classification of multispectral, hyperspectral and PolSAR data: a comparative evaluation. Rem. Sens. 13 (21). https://doi.org/10.3390/rs13214405. Article 21.
Joseph, K.S., Ravichandran, T., 2012. A comparative evaluation of software effort estimation using REPTree and K* in handling with missing values. Australian Journal of Basic and Applied Sciences 6, 312–317.
Kohout, F.A., 1960. Cyclic flow of saltwater in the Biscayne aquifer of southeastern Florida. J. Geophys. Res. 65 (7), 2133–2141.
Kohout, F.A., Kolipinski, M.C., 1964. Biological Zonation Related to Groundwater Discharge along the Shore of Biscayne Bay. ESTUARIES, Miami, Florida, pp. 488–499.
Kopsiaftis, G., Kaselimi, M., Protopapadakis, E., Voulodimos, A., Doulamis, A., Doulamis, N., 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. https://doi.org/10.3389/frwa.2023.1195029.
Kushwaha, N.L., Rajput, J., Elbeltagi, A., Elnaggar, A.Y., Sena, D.R., Vishwakarma, D.K., Mani, I., Hussein, E.E., 2021. Data intelligence model and meta-heuristic algorithms-based Pan evaporation modelling in two different agro-climatic zones: a case study from northern India. Atmosphere 12, 1654. https://doi.org/10.3390/atmos12121654.
Kushwaha, N.L., Rajput, J., Sena, D.R., Elbeltagi, A., Singh, D.K., Mani, I., 2022. Evaluation of data-driven hybrid machine learning algorithms for modelling daily reference evapotranspiration. Atmos.-Ocean 62, 1–22. https://doi.org/10.1080/07055900.2022.2087589.

Lal, A., Datta, B., 2018. Development and implementation of support vector machine regression surrogate models for predicting groundwater pumping-induced saltwater intrusion into coastal aquifers. Water Resour. Manag. 32 (7), 2405–2419. https://doi.org/10.1007/s11269-018-1936-2.
Langevin, C.D., 2001. Simulation of Groundwater Discharge to Biscayne Bay, Southeastern Florida. US geological survey, water-resources investigations report 00-4251, Tallahassee, Florida.
Li, H., Lee, Y.-C., Zhou, Y.-C., Sun, J., 2011. The random subspace binary logit (RSBL) model for bankruptcy prediction. Knowl. Base Syst. 24 (8), 1380–1388. https://doi.org/10.1016/j.knosys.2011.06.015.
Li, J., Cheng, J., Shi, J., Huang, F., 2012. Brief introduction of back propagation (BP) neural network algorithm and its improvement. In: Jin, D., Lin, S. (Eds.), Advances in Computer Science and Information Engineering, Advances in Intelligent and Soft Computing. Springer, Berlin Heidelberg, Berlin, Heidelberg, pp. 553–558. https://doi.org/10.1007/978-3-642-30223-7_87.
Lin, K., Lu, P., Xu, C.Y., Yu, X., Lan, T., Chen, X., 2019. Modeling saltwater intrusion using an integrated Bayesian model averaging method in the Pearl River Delta. J. Hydroinf. 21 (6), 1147–1162.
Lu, P., Lin, K., Xu, C.Y., Lan, T., Liu, Z., He, Y., 2021. An integrated framework of input determination for ensemble forecasts of monthly estuarine saltwater intrusion. J. Hydrol. 598, 126225.
Mazhar, S., Pellegrini, E., Contin, M., Bravo, C., De Nobili, M., 2022. Impacts of salinization caused by sea level rise on the biological processes of coastal soils - a review. Front. Environ. Sci. 10, 909415. https://doi.org/10.3389/fenvs.2022.909415.
Miao, X., Heaton, J.S., Zheng, S., Charlet, D.A., Liu, H., 2012. Applying tree-based ensemble algorithms to the classification of ecological zones using multi-temporal multi-source remote-sensing data. Int. J. Rem. Sens. 33 (6), 1823–1849.
Miller, J.A., 1990. Ground Water Atlas of the United States: Alabama, Florida, Georgia, South Carolina, HA. U.S. Geological Survey, Publication, pp. 730–G. https://www.nrc.gov/docs/ML1002/ML100290484.pdf.
Misra, S., Li, H., 2020. Chapter 9 Noninvasive fracture characterization based on the classification of sonic wave travel times. In: Machine Learning for Subsurface Characterization, pp. 243–287. https://doi.org/10.1016/b978-0-12-817736-5.00009-0.
Moore, W.S., Joye, S.B., 2021. Saltwater intrusion and submarine groundwater discharge: acceleration of biogeochemical reactions in changing coastal aquifers. Front. Earth Sci. 9, 600710. https://doi.org/10.3389/feart.2021.600710.
Nosair, A.M., Shams, M.Y., AbouElmagd, L.M., Hassanein, A.E., Fryar, A.E., Abu Salem, H.S., 2022. Predictive model for progressive salinization in a coastal aquifer using artificial intelligence and hydrogeochemical techniques: a case study of the Nile Delta aquifer, Egypt. Environ. Sci. Pollut. Control Ser. 29 (6), 9318–9340. https://doi.org/10.1007/s11356-021-16289-w.
Oude Essink, G.H.P., 2001. Improving fresh groundwater supply—problems and solutions. Ocean Coast Manag. 44 (5), 429–449. https://doi.org/10.1016/S0964-5691(01)00057-6.
Pan, Y., Jiang, J., Wang, R., Cao, H., Cui, Y., 2009. Predicting the auto-ignition temperatures of organic compounds from molecular structure using support vector machine. J. Hazard Mater. 164 (2–3), 1242–1249. https://doi.org/10.1016/j.jhazmat.2008.09.031.
Pande, C.B., Al-Ansari, N., Kushwaha, N.L., Srivastava, A., Noor, R., Kumar, M., Moharir, K.N., Elbeltagi, A., 2022. Forecasting of SPI and meteorological drought based on the artificial neural network and M5P model tree. Land 11, 2040. https://doi.org/10.3390/land11112040.
Pearce, W., Holmberg, K., Hellsten, I., Nerlich, B., 2014. Climate change on twitter: topics, communities and conversations about the 2013 IPCC working group 1 report. PLoS One 9 (4), e94785. https://doi.org/10.1371/journal.pone.0094785.
Pham, B.T., Prakash, I., Tien Bui, D., 2018. Spatial prediction of landslides using a hybrid machine learning approach based on Random Subspace and Classification and Regression Trees. Geomorphology 303, 256–270. https://doi.org/10.1016/j.geomorph.2017.12.008.
Pham, N.H., Pham, B.Q., Tran, T.T., 2022. Apply machine learning to predict saltwater intrusion in the ham luong river, ben tre province. VNU Journal of Science: Earth and Environmental Sciences 38 (3).
Qi, S., He, M., Hoang, R., Zhou, Y., Namadi, P., Tom, B., Sandhu, P., Bai, Z., Chung, F., Ding, Z., et al., 2023. Salinity modeling using deep learning with data augmentation and transfer learning. Water 15, 2482. https://doi.org/10.3390/w15132482.
Rahman, Md S., Islam, A.R. Md T., 2019. Are precipitation concentration and intensity changing in Bangladesh overtimes? Analysis of the possible causes of changes in precipitation systems. Sci. Total Environ. 690, 370–387. https://doi.org/10.1016/j.scitotenv.2019.06.529.
Salzberg, S.L., 1994. C4.5: programs for machine learning by J. Ross quinlan. Morgan kaufmann publishers, inc., 1993. Mach. Learn. 16 (3), 235–240. https://doi.org/10.1007/BF00993309.
Sayadi Shahraki, A., Boroomand Nasab, S., Naseri, A.A., Soltani Mohammadi, A., 2021. Estimation groundwater depth using ANN-PSO kriging and IDW models (case study: salman Farsi Sugarcane Plantation). Cent. Asian J. Environ. Sci. Technol. Innov. 2. https://doi.org/10.22034/CAJESTI.2021.03.01.
Shamshirband, S., Jafari Nodoushan, E., Adolf, J.E., Abdul Manaf, A., Mosavi, A., Wing Chau, K., 2019. Ensemble models with uncertainty analysis for multi-day ahead forecasting of chlorophyll a concentration in coastal waters. Eng. Applic. Comput. Fluid Mech. 13, 91–101. https://doi.org/10.1080/19942060.2018.1553742.
Shinn, E.A., Reese, R.S., Reich, C.D., 1994. Fate and pathways of injection-well effluent in the Florida Keys: U.SGeological Survey Open-File Report 94-276, p. 116.
Singh, A.K., 2020. Coastal agriculture and future challenges. In: Singh, A., Fernando, R.L. S., Haran, N.P. (Eds.), Development in Coastal Zones and Disaster Management. Disaster Research and Management Series on the Global South. Palgrave Macmillan, Singapore. https://doi.org/10.1007/978-981-15-4294-7_5.
Singh, K.P., Basant, N., Gupta, S., 2011. Support vector machines in water quality management. Anal. Chim. Acta 703 (2), 152–162. https://doi.org/10.1016/j.aca.2011.07.027.
Singh, U., Rizwan, M., Alaraj, M., Alsaidan, I., 2021. A machine learning-based gradient boosting regression approach for wind power production forecasting: a step towards smart grid environments. Energies 14, 5196. https://doi.org/10.3390/en14165196.
Skurichina, M., Duin, R.P.W., 2002. Bagging, boosting and the random subspace method for linear classifiers. Pattern Anal. Appl. 5 (2), 121–135. https://doi.org/10.1007/s100440200011.
Taşan, M., Taşan, S., Demir, Y., 2023. Estimation and uncertainty analysis of groundwater quality parameters in a coastal aquifer under seawater intrusion: a comparative study of deep learning and classic machine learning methods. Environ. Sci. Pollut. Control Ser. 30 (2), 2866–2890. https://doi.org/10.1007/s11356-022-22375-4.
Tran, T.T., Pham, N.H., Pham, Q.B., et al., 2022. Performances of different machine learning algorithms for predicting saltwater intrusion in the Vietnamese mekong Delta using limited input data: a study from ham luong river. Water Resour. 49, 391–401. https://doi.org/10.1134/S0097807822030198.
Ubah, J.I., Orakwe, L.C., Ogbu, K.N., Awu, J.I., Ahaneku, I.E., Chukwuma, E.C., 2021. Forecasting water quality parameters using artificial neural network for irrigation purposes. Sci. Rep. 11, 24438. https://doi.org/10.1038/s41598-021-04062-5.
Upchurch, S., Scott, T.M., Alfieri, M.C., Fratesi, B., Dobecki, T.L., 2019. Hydrogeology of Florida. In: The Karst Systems of Florida. Cave and Karst Systems of the World. Springer, Cham. https://doi.org/10.1007/978-3-319-69635-5_4.
Vapnik, V.N., 1998. Statistical Learning Theory. Wiley. Wiley.Com. https://www.wiley.com/en-us/Statistical+Learning+Theory-p-9780471030034.
Wang, Jingzhe, Ding, Jianli, Yu, Danlin, Teng, Dexiong, He, Bin, Chen, Xiangyue, Ge, Xiangyu, Zhang, Zipeng, Wang, Yi, Yang, Xiaodong, Shi, Tiezhu, Fenzhen, Su, 2020a. Machine learning-based detection of soil salinity in an arid desert region, Northwest China: a comparison between Landsat-8 OLI and Sentinel-2 MSI. Sci. Total Environ. 707, 136092. https://doi.org/10.1016/j.scitotenv.2019.136092. ISSN 0048-9697.
Wang, L., Long, F., Liao, W., Liu, H., 2020b. Prediction of anaerobic digestion performance and identification of critical operational parameters using machine learning algorithms. Bioresour. Technol. 298, 122495. https://doi.org/10.1016/j.biortech.2019.122495.
Witten, I.H., Frank, E., Hall, M.A., 2011. Chapter 4 - algorithms: the basic methods. In: Witten, I.H., Frank, E., Hall, M.A. (Eds.), Data Mining: Practical Machine Learning Tools and Techniques, third ed. Morgan Kaufmann, pp. 85–145. https://doi.org/10.1016/B978-0-12-374856-0.00004-3.
Yu, X., He, L., Yao, R., Tu, T., Zhang, Z., Zhao, X., 2024. Effects of beach nourishment on seawater intrusion in layered heterogeneous aquifers. J. Hydrol. 633 (2024), 131018. https://doi.org/10.1016/j.jhydrol.2024.131018. ISSN 0022-1694.
Zanoni, M.G., Majone, B., Bellin, A., 2022. A catchment-scale model of river water quality by Machine Learning. Sci. Total Environ. 838, 156377. https://doi.org/10.1016/j.scitotenv.2022.156377.
Zheng, C., Wang, P.P., P, P., 1999. MT3DMS: a modular three- dimensional multispecies transport model for simulation of advection. In: Dispersion and Chemical Reactions of Contaminants in Groundwater Systems. Vicksburg, Mississippi: Waterways Experiment Station. US Army Corps of Engineers.