A surrogate-based simulation–optimization approach for coastal aquifer management

Zheng Han, Wenxi Lu, Yue Fan, Jin Lin and Qian Yuan

ABSTRACT

This study proposed a pumping-injection (P-I) groundwater management strategy based on a simulation–optimization (S-O) framework to mitigate seawater intrusion (SI). The methodology was applied to a real case in Longkou, China. A three-dimensional variable-density groundwater simulation model was established to simulate and predict the SI process. In the S-O framework, while solving the optimization model, it is required to call the simulation model thousands of times, which leads to enormous computational load. In this case, the Kriging and support vector regression (SVR) surrogate models were established for the simulation model respectively. Furthermore, the ensemble surrogate modeling technique was applied to construct the Kriging-SVR ensemble surrogate model. The most accurate surrogate model was selected as the substitute for the simulation model, saving considerable computing costs. The results show that the ensemble surrogate model performs better than the stand-alone surrogate models in accuracy, indicating that combining stand-alone surrogate models is a potential modeling method for the surrogate model of the variable-density groundwater simulation model. By solving the optimization model, the optimal pumping and injection schemes under different scenarios were obtained. The optimization results demonstrate that the proposed methodology is effective and stable in coastal groundwater management.

Key words | ensemble surrogate model, groundwater resources management, Kriging, seawater intrusion, simulation-optimization, SVR

Zheng Han
Wenxi Lu (corresponding author)
Yue Fan
Qian Yuan
College of New Energy and Environment,
Jilin University,
Changchun,
China;
Key Laboratory of Groundwater Resources and
    Environment,
Jilin University, Ministry of Education,
Changchun,
China;
and
Jilin Provincial Key Laboratory of Water Resources
    and Environment,
Jilin University,
Changchun,
China
E-mail: *luwx999@163.com*

Jin Lin
Nanjing Hydraulic Research Institute,
Nanjing,
China

HIGHLIGHTS

- The P-I strategy based on S-O is efficient in mitigating seawater intrusion.
- An ensemble surrogate model is established for variable-density groundwater simulation model.
- The ensemble surrogate model outperforms stand-alone surrogate models in accuracy.
- The surrogate model can save computational costs during the optimization process.

INTRODUCTION

Seawater intrusion (SI) refers to the phenomenon of saltwater directly invading aquifers due to factors such as sea level rise (SLR) and the decline of freshwater level in coastal aquifers. SI can occur as a result of long-term interaction of various factors of nature and human society. One of the most important reasons is the unreasonable exploitation of groundwater in coastal areas. The occurrence and development of SI can cause a series of environmental geological problems, such as groundwater quality deterioration, soil salinization, etc., which will seriously affect the daily life of local residents and restrict local economic development. SI has occurred in more than 50 countries and regions around the world, and poses grave threats to the ecological environment and

economic development of these areas. Therefore, it is of great significance to perform control measures for SI.

Currently, many measures have been put forward and implemented to prevent and control SI. Among them, forming underground freshwater barriers by artificial recharge is very effective and common (Armanuos *et al.* 2019). The underground freshwater level is raised due to artificial recharge, so that the underground freshwater head is higher than the adjacent saltwater head and the positive hydraulic gradient from the land to the ocean is rebuilt, creating a hydraulic barrier to prevent SI. However, this measure requires a large amount of freshwater supply. In the past, many researchers believed that coastal injection is inefficient because it loses water to the sea. In a study of corrective measures for SI, Abarca *et al.* (2006) proved the efficiency of the hydraulic barrier, overturning a widely held belief. They argued that the net gain obtained from protection of inland pumping more than compensates for the loss. While it is true, sufficient water supply for the injection is still needed. Surface freshwater, like lakes and reservoirs, is usually used as the freshwater source in the traditional injection method. For areas lacking sufficient surface freshwater, it is difficult to implement the mitigation measure of artificial recharge. In this case, the present study proposed a pumping-injection (P-I) groundwater management method to solve this problem. In the P-I method, the groundwater extracted from inland aquifers was utilized as the freshwater source for artificial recharge. The exploitation of groundwater has a significant impact on SI. Improper water allocation of pumping and injection may cause low project returns, or even make SI more serious. Thus, the simulation–optimization (S-O) method was applied in this study to make the P-I measure become safe and profitable.

The S-O approach is widely used for groundwater resources management in coastal aquifers (Yang *et al.* 2018). During the optimization process, it is required to invoke the simulation model repeatedly, causing a huge computational load. Surrogate models, which can be considered as the computationally efficient emulators designed to mimic key characteristics of a simulation model, have great potential to overcome the computational limitations (Asher *et al.* 2015; Song *et al.* 2018). Based on radial basis functions, as surrogate models for the

computationally expensive variable-density flow and salt transport numerical simulations, Christelis & Mantoglou (2016) compared the performance of two adaptive meta-modeling frameworks (the adaptive–recursive approach and the metamodel–embedded evolution strategy) in terms of computational times and optimal solutions of pumping rates. Huang & Chiu (2018) presented an S-O model that aimed to minimize the total injection rate based on the pre-determined locations of injection barriers while solving the SI problem along the coastal aquifers in Pingtung Plain. The SEAWAT code was used to simulate the process of SI and the surrogate model of artificial neural networks was used to approximate the SI numerical model to improve the computational efficiency. Most of the surrogate models built in previous studies were based on only one method. Yet, different stand-alone surrogate models have their own limitations. Thus, constructing an ensemble surrogate model by combining multiple stand-alone surrogate models is a promising surrogate modeling technique. Ouyang *et al.* (2017) applied a new method referred to as the optimal weighted surrogate (OWS), proposed by Viana *et al.* (2009), to construct ensemble surrogate models for the process simulations of surface-enhanced aquifer remediation, and they proved the favorable performance of this method. For the variable-density groundwater simulation model with strong nonlinear characteristics, however, the performance of the ensemble surrogate model has not been investigated. Therefore, the present work introduced the OWS method to construct the ensemble surrogate model of the SI model for the first time, and explored the performance of the ensemble surrogate model in coastal aquifer management.

In this study, a three-dimensional variable-density groundwater numerical simulation model was established for the prediction of SI in Longkou, China. After the calibration and verification of the simulation model, the Kriging surrogate model, support vector regression (SVR) surrogate model, and Kriging-SVR ensemble surrogate model were developed for the SI model respectively. To obtain the optimal P-I groundwater management scheme, the most accurate one of the three surrogate models was selected to couple with the optimization model, in which the objective was to minimize the extent of SI, for calculation. The proposed methodology proves to be a robust

tool for decision-making with respect to coastal groundwater management and the control of SI.

METHODOLOGY

Numerical simulation model for SI

When the mixing zone (transition between freshwater and saltwater zones) is taken into consideration, the problem of SI needs to be described by two equations. One is the equation of groundwater flow that reflects the change of solution density and viscosity (Bear 1972). The other is the equation of water quality that reflects salt transport in groundwater. Because the concentration term is incorporated into the groundwater flow equation and the actual flow velocity in the groundwater solute transport equation needs to be calculated by water head, the above two equations need to be solved by the motion equation. The general mathematical formulation describing variable-density groundwater flow and transport process can be expressed with the following two equations (Miao *et al.* 2019):

$$\frac{\partial}{\partial x}\left(K_{xx}\frac{\partial H}{\partial x}\right) + \frac{\partial}{\partial y}\left(K_{yy}\frac{\partial H}{\partial y}\right) + \frac{\partial}{\partial z}\left[K_{zz}\left(\frac{\partial H}{\partial z} + \eta c\right)\right]$$
$$= \mu_s\frac{\partial H}{\partial t} + n\eta\frac{\partial c}{\partial t} - \frac{\rho}{\rho_0}q \tag{1}$$

$$\frac{\partial}{\partial x}\left(D_{xx}\frac{\partial c}{\partial x}\right) + \frac{\partial}{\partial y}\left(D_{yy}\frac{\partial c}{\partial y}\right) + \frac{\partial}{\partial z}\left(D_{zz}\frac{\partial c}{\partial z}\right)$$
$$- \frac{\partial}{\partial x}(u_x c) - \frac{\partial}{\partial y}(u_y c) - \frac{\partial}{\partial z}(u_z c) \tag{2}$$
$$= \frac{\partial c}{\partial t} + \frac{q}{n}(c - c^*)$$

where K_{xx}, K_{yy}, K_{zz} are the hydraulic conductivity tensors of the aquifer [LT^{-1}]; c denotes the solution concentration [ML^{-3}]; n is effective porosity; ρ_0 is the density of freshwater [ML^{-3}]; $\eta = \frac{\varepsilon}{c_s}$ is the density coupling coefficient [$M^{-1}L^3$], $\varepsilon = \frac{\rho_s - \rho_0}{\rho_0}$ is the density difference, ρ_s is the maximum density [ML^{-3}], c_s is the liquid concentration corresponding to ρ_s [ML^{-3}]; μ_s is equivalent freshwater specific storage [L^{-1}]; q is the volumetric flow rate of sources and sinks per unit volume of the aquifer [T^{-1}]; D_{xx}, D_{yy}, D_{zz} are the hydrodynamic dispersion tensors [L^2T^{-1}]; u_x, u_y, u_z are the seepage velocities

in the respective x, y and z directions [LT^{-1}]; c^* denotes the concentration of salt in the liquid for extraction or injection [ML^{-3}].

The equation of motion for solving the coupling problems is:

$$u_x = -\frac{K_{xx}^0}{n}\frac{\partial h}{\partial x}, \, u_y = -\frac{K_{yy}^0}{n}\frac{\partial h}{\partial y}, \, u_z = -\frac{K_{zz}^0}{n}\left(\frac{\partial h}{\partial z} + \eta c\right) \tag{3}$$

where K^0 is the hydraulic conductivity under the freshwater condition [LT^{-1}]. Equations (1)–(3), together with the initial and boundary conditions, constitute a complete mathematical model of SI. Hydraulic conductivity, specific storage, porosity, and dispersivity are treated as unknown parameters to be determined by inverse modeling in this study.

Surrogate models

The surrogate model has the advantages of a small amount of calculation and fast calculation speed. It functionally approximates the input–output relationship of the simulation model and is often used to replace the simulation model for optimization calculation. Its mathematical essence is using approximation technology to fit or interpolate the information of known sampling points to form a simple mathematical model and thereby realizing the prediction of the response output value of unknown points (Viana *et al.* 2013). The sampling method and the specific modeling method determine the accuracy of the surrogate model. In this study, SVR and Kriging surrogate models were established respectively for the SI model. Based on the two stand-alone surrogate models, an ensemble surrogate model was constructed by giving the corresponding weights of different stand-alone surrogate models.

Kriging surrogate model

The Kriging method is an optimized interpolation approach that was proposed by the French geographer and mathematician Matheron and the South African mining engineer, Krige. According to the information contained in the sample points and the spatial characteristics of variables, the Kriging method can predict the information of the unknown points by establishing approximate functions.

The basic form of the Kriging surrogate model is:

$$\hat{y}(x) = f(x)^{\mathrm{T}}\beta + z(x) \tag{4}$$

where $y(x)$ is the actual value; $\hat{y}(x)$, calculated by the surrogate model, is the predicted value of $y(x)$ and can be divided into two parts: $f(x)^{\mathrm{T}}\beta$ is the linear regression part and $z(x)$ is the local deviation from the regression model; $f(x) = [f_1(x), f_2(x), \cdots, f_k(x)]^{\mathrm{T}}$ is the base function of the known regression model and the quadratic function-type base function is selected in this paper; the undetermined parameter $\beta = [\beta_1, \beta_2, \cdots, \beta_k]^{\mathrm{T}}$ is the coefficient corresponding to the base function, which can be obtained by using the prepared training data. A detailed theory of the Kriging method has been given in the previous work of Guo *et al.* (2019).

SVR surrogate model

SVR was specifically designed to deal with nonlinear regression fitting. It is an extension of support vector machine theory (Vapnik 1995). The key theory of SVR is to project a training sample into a high-dimensional feature space using a non-linear mapping function, and find a linear regression hyperplane in this space to fit the input–output relationship of the training sample. In this way, the input–output relationship of the training samples will change from a complex non-linear relationship in a low-dimensional space to a simple linear relationship in a high-dimensional space (Lahiri & Ghanta 2009).

For a given set of input sample points $\mathbf{X} = [x_1, x_2, \cdots, x_m]^{\mathrm{T}}$ and its corresponding output value $\mathbf{Y} = [y_1, y_2, \cdots, y_m]^{\mathrm{T}}$, the predicted output value of SVR at the test point x_p is as follows:

$$f(x_p) = w \cdot \varphi(x_p) + b \tag{5}$$

where $\varphi(x_p)$ is the nonlinear mapping function, through which the training samples can be mapped to high-dimensional space; w is the weight vector; b is the bias term. A detailed introduction to the SVR surrogate modeling technique can be found in Hou & Lu (2018).

Ensemble surrogate model

By allocating the corresponding weight coefficients of different stand-alone surrogate models and combining them linearly, an ensemble surrogate model is set up. With the advantages of different stand-alone surrogate models, the prediction error can be reduced. Its expression is as follows:

$$\hat{y}_{\mathrm{EN}}(x) = \sum_{i=1}^{n} \omega_i \hat{y}_i(x) \tag{6}$$

where n is the number of stand-alone surrogate models; $\hat{y}_{\mathrm{EN}}(x)$ is the predicted value of the ensemble surrogate model; ω_i is the weight of the ith surrogate model; $\hat{y}_i(x)$ is the predicted value of the ith surrogate model; $\sum_{i=1}^{n} \omega_i = 1$.

The key to constructing an ensemble surrogate model is in determining the weight. In this paper, the method of minimizing the mean square error (MSE) was adopted to calculate weights. The MSE expression of the ensemble surrogate model is as follows:

$$MSE_{\mathrm{EN}} = \frac{1}{V} \int_V e_{\mathrm{EN}}^2(x)dx = \mathbf{w}^{\mathrm{T}}\mathbf{C}\mathbf{w} \tag{7}$$

where V is the variance of the predicted value of the ensemble surrogate model; $\mathbf{w}$ is the weight matrix; MSE_{EN} is the MSE of the ensemble surrogate model; $\mathbf{C}$ is the covariance matrix; e_{EN} is the prediction error of the ensemble surrogate model, $e_{\mathrm{EN}} = y(x) - \hat{y}_{\mathrm{EN}}(x)$, $y(x)$ is the actual value. The elements of $\mathbf{C}$ can be calculated by the following formula:

$$c_{ij} = \frac{1}{p}\widetilde{\mathbf{e}}_i^{\mathrm{T}}\widetilde{\mathbf{e}}_j \tag{8}$$

where p is the number of training samples; i and j represent different stand-alone surrogate models respectively; $\widetilde{\mathbf{e}}$ is the cross-validation error of the surrogate model. In this paper, the leave-one-out cross-validation method is utilized to calculate the cross-validation error, that is, one sample is removed from the training samples in turn, and the remaining samples are used to build the surrogate model and predict the removed sample point. The difference between the actual value and the predicted value of the removed sample point is the cross-validation error of this point.

Based on the calculated covariance matrix, the weights can be obtained by minimizing the MSE of the ensemble surrogate model:

$$\min_w MSE_{\mathrm{EN}} = \mathbf{w}^{\mathrm{T}}\mathbf{C}\mathbf{w}$$
$$s.t. \, \mathbf{1}^{\mathrm{T}}\mathbf{w} = \mathbf{1} \tag{9}$$

The above formula can be solved with the Lagrange multiplier (Viana *et al.* 2009):

$$\mathbf{w} = \frac{\mathbf{C}^{-1}\mathbf{1}}{\mathbf{1}^{\mathrm{T}}\mathbf{C}^{-1}\mathbf{1}} \tag{10}$$

In order to avoid the calculated weights being negative and greater than 1, only the diagonal elements in the covariance matrix $\mathbf{C}$ are used for calculation, so the above formula becomes:

$$\mathbf{w} = \frac{\mathbf{C}_{\mathrm{diag}}^{-1}\mathbf{1}}{\mathbf{1}^{\mathrm{T}}\mathbf{C}_{\mathrm{diag}}^{-1}\mathbf{1}} \tag{11}$$

The optimization model

Establishing a mathematical optimization model is required in the coastal groundwater management approach of P-I. Generally, the optimization model consists of three basic elements: decision variables, objective functions and constraints (Kourakos & Mantoglou 2015). The objective of this study is to minimize the extent of SI expressed by the percentage of mass increase in the aquifer at the end of the management horizon. Decision variables are the pumping rate of each pumping well and the injection rate of each recharge well. Constraints mainly include that total capacity of the selected pumping wells must meet the demands of local water use and injection. By solving the optimization model, the optimal groundwater exploitation and injection scheme can be obtained. The mathematical expression of this optimization management model is written as:

$$\min f = [(mass_{\mathrm{end}} - mass_{\mathrm{ini}})/mass_{\mathrm{ini}}] \times 100\% \tag{12}$$

Subject to:

$$\sum_{i=1}^{m} Q_i^{\mathrm{ex}} = Q_0 + \sum_{j=1}^{n} Q_j^{\mathrm{in}} \tag{13}$$

$$\sum_{j=1}^{n} Q_j^{\mathrm{in}} \leq Q_{\mathrm{max}} \tag{14}$$

$$Q_{\mathrm{min}}^{\mathrm{in}} \leq Q_j^{\mathrm{in}} \leq Q_{\mathrm{max}}^{\mathrm{in}} \tag{15}$$

$$Q_{\mathrm{min}}^{\mathrm{ex}} \leq Q_i^{\mathrm{ex}} \leq Q_{\mathrm{max}}^{\mathrm{ex}} \tag{16}$$

$$mass_{\mathrm{end}} = g(Q, \boldsymbol{p}) \tag{17}$$

where f is the objective function; $mass_{\mathrm{ini}}$ and $mass_{\mathrm{end}}$ are the total solute mass in the coastal aquifer at the beginning and end of the management horizon [M], respectively; Q_j^{in} is the water injection rate of the jth injection well [$\mathrm{L}^3\mathrm{T}^{-1}$]; Q_i^{ex} is the pumping rate of the ith pumping well [$\mathrm{L}^3\mathrm{T}^{-1}$]; Q_0 is the local water demand for selected pumping wells [$\mathrm{L}^3\mathrm{T}^{-1}$]; Q_{max} is the limit of the total injection rate according to the local financial condition [$\mathrm{L}^3\mathrm{T}^{-1}$]; $Q_{\mathrm{min}}^{\mathrm{in}}$, $Q_{\mathrm{max}}^{\mathrm{in}}$, $Q_{\mathrm{min}}^{\mathrm{ex}}$ and $Q_{\mathrm{max}}^{\mathrm{ex}}$ are the minimum allowable water injection rate, the maximum allowable water injection rate, the minimum allowable pumping rate and the maximum allowable pumping rate respectively [$\mathrm{L}^3\mathrm{T}^{-1}$], and their values may vary according to the different condition of each well; g represents the total solute mass in the coastal aquifer at the end of the management horizon calculated by the simulation model [M]; $\boldsymbol{p}$ is the parameter value vector of the simulation model. In the present study, the simulation model was replaced by the most precise one of the above three surrogate models, in order to reduce the computational burden brought by operating the simulation model.

CASE STUDY

Study area

The study area, the coastal part of Longkou City, is located in Shandong Province, the northwest of Jiaodong Peninsula in China, between $120°13'14'' \sim 120°44'46''$ east longitude and $37°27'30'' \sim 37°47'24''$ north latitude (Figure 1). The total area is about 245 km². The south is low hills and the north is the coastal plain. The annual average precipitation in Longkou City is 656.6 mm. The annual precipitation varies greatly, and the regional distribution of precipitation is uneven. The precipitation in the southern mountainous areas is larger than that in the northern plain areas. The average annual evaporation is 1,150–1,250 mm. In this study site, rivers rise in the southern mountains and flow to the northwest. Yongwen River, Balisha River and Beima River are the main rivers.

In Longkou, the thickness of aquifers is small and the shortage of freshwater resources is severe. The long-term overexploitation of groundwater causes the groundwater level to drop sharply, which destroys the original hydraulic
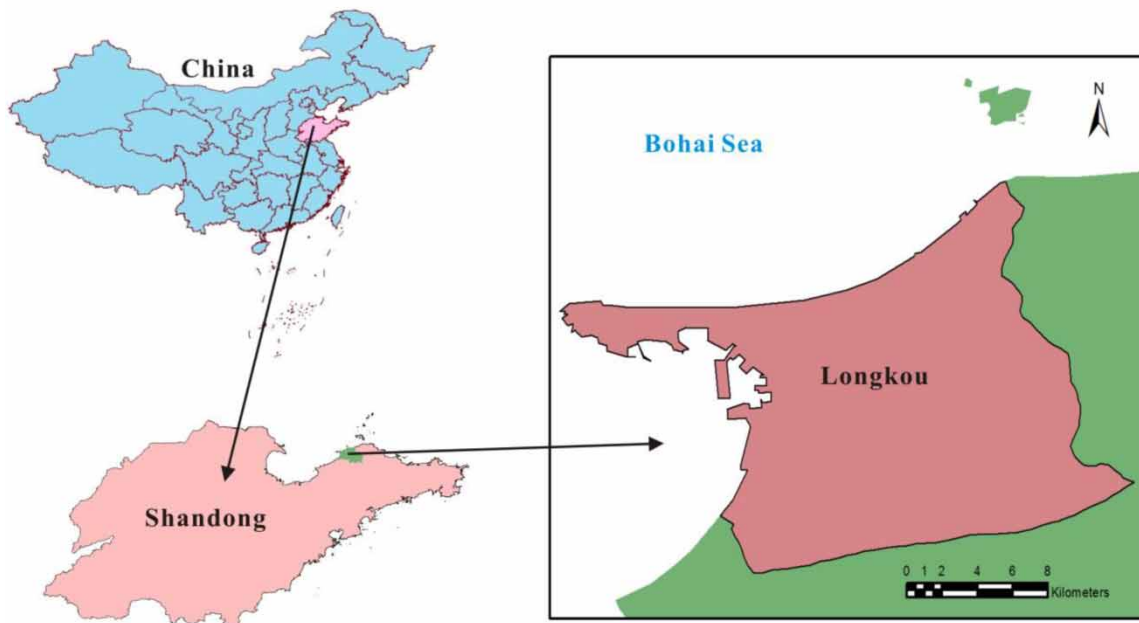
Figure 1 │ Location of the study area.

balance between the salty water body and the fresh water body. At the same time, the marine sediments near the coastline of the Longkou area are fine sand, coarse sand and fine gravel with good water permeability, and there is a good hydraulic connection between the seawater and the coastal aquifer. Therefore, the extent of SI in the Longkou area has been serious for a long time and an efficient groundwater management strategy needs to be proposed urgently. The hydrogeological background of the study area has been described in more detail in the previous work of Miao *et al.* (2019).

The locations of boundaries in this simulation area are shown in Figure 2. Boundary 1 was a sea boundary and could be defined as a known head and constant salt concentration boundary. Boundary 3, the south piedmont boundary of the area, was generalized as a constant flow boundary. Boundary 2 and Boundary 4 represented the watershed and were treated as zero-flux boundaries. The lower boundary, which was the water-proof bedrock, was defined as a zero-flux boundaries. The upper boundary was the groundwater surface, where water exchange between the aquifer and atmosphere took place. The aquifer in this study area can be divided into three layers (Figure 3). The first layer is a phreatic aquifer with sand. This layer is the target unit

for optimization management, and thus the following SI states are all directed at the first layer. Based on the permeability, it was divided into eight zones, as shown in Figure 2. The second layer (C2) is silt or clay with poor permeability, and it gradually thins from west to east. The third layer (C3) is sandstone with very low hydraulic conductivity. Therefore, the aquifer is a heterogeneous anisotropic aquifer, and the flow is generalized as a three-dimensional unsteady flow considering variable density.

SI model

In this study, the SI model was established based on the SEAWAT code. The source code of SEAWAT was developed by combining MODFLOW (Langevin *et al.* 2003) and MT3DMS (Zheng & Wang 1999) into a single program, which can solve the coupled equations of groundwater flow and solute transport.

When the water quality of subsurface aquifers is deteriorating due to SI, the most obvious change is the increase of chloride ion concentration in the water. Therefore, the chloride ion is usually used as an indicator to determine whether SI occurs and the extent of intrusion. In this study, the chloride concentration of the seawater
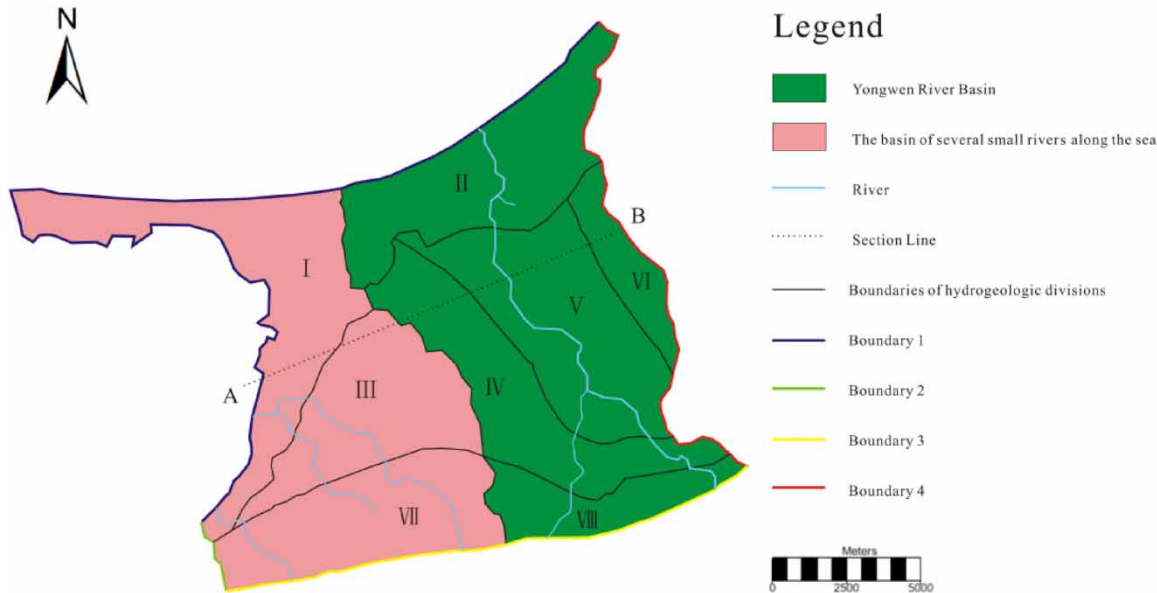
Figure 2 | Boundaries and parameter partitions of the simulation area.
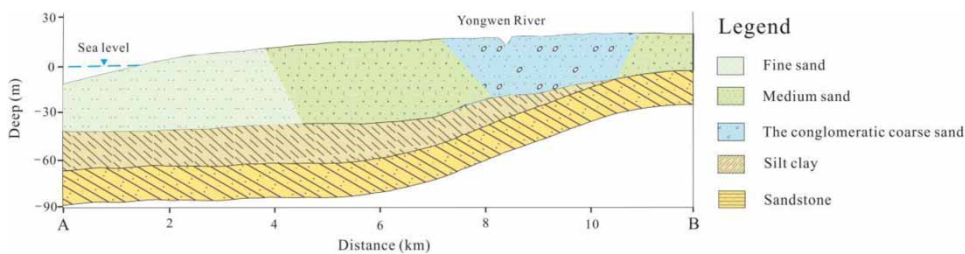

Figure 3 | Schematic diagram of A–B stratigraphic section.

boundary was 19,000 mg/l. According to local requirements of water quality, groundwater chlorine concentration exceeding 250 mg/l was designed as the sign of SI.

The SI model was calibrated and verified based on the groundwater elevation and chloride ion concentration data of the observation wells in the study area over past years. The groundwater elevation and chloride concentration data for the years 2015–2016 were used as the measured data for calibration. For groundwater elevation calibration, the mean relative error (MRE) reached 18.00% (Figure 4(a)). For water quality calibration, the MRE reached 7.20% (Figure 4(b)). The results indicate good agreement between the measured and calculated values in the calibration procedure. Following calibration, the SI model was verified using the observation data of the following years, 2016–2017. Figure 4(c) shows

the measured and the verified groundwater elevations, while Figure 4(d) shows the verification performance of the water quality model. Overall, the MREs between the simulated values and the measured values of the groundwater elevation and chloride concentration were all less than 20%, demonstrating that the SI model can be used to predict the future salinity concentration. The hydrogeological parameters of the calibrated model are shown in Table 1.

Taking June 1, 2017, as the initial time, the above SI model was used to predict the extent of SI in the next 50 years. In the process of prediction, the effect of climate change was taken into account. SLR and precipitation variation caused by climate change are two of the important factors that exert influences on SI (Ketabchi *et al.* 2016). According to the China Sea Level Bulletin 2018, the average
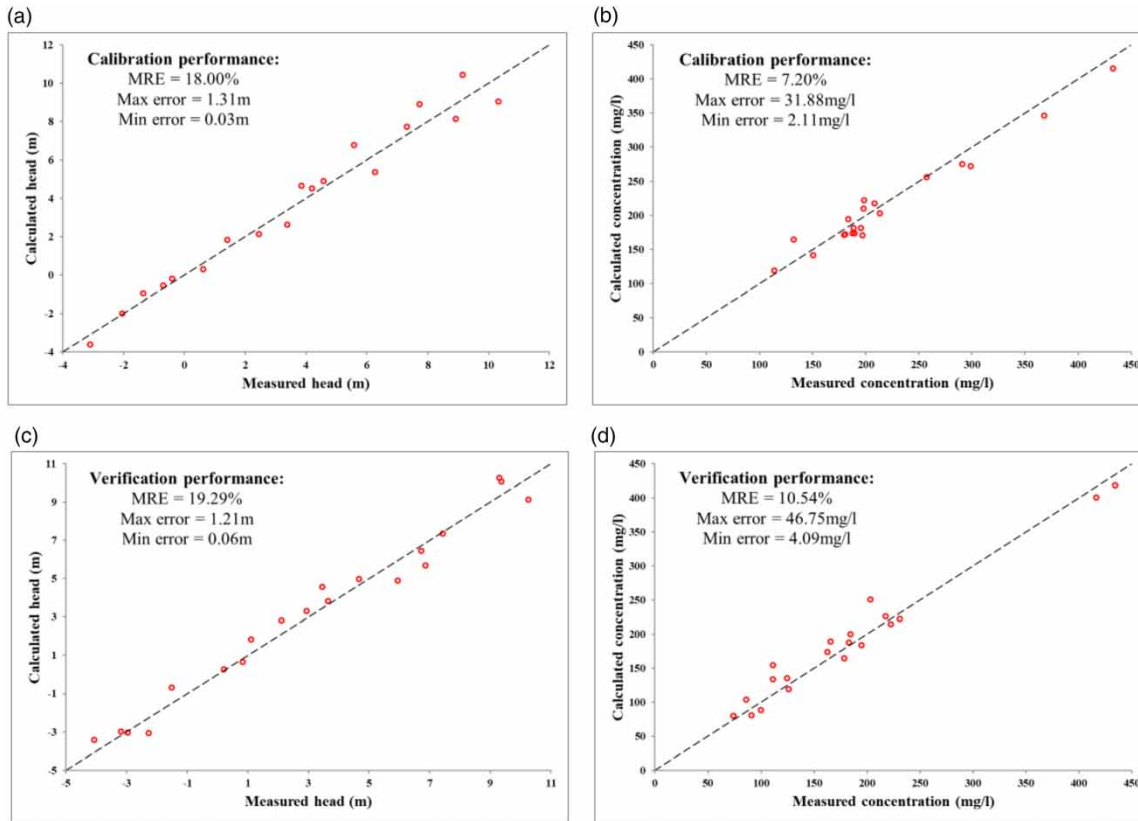
Figure 4 │ Results of model calibration and verification.

Table 1 │ The calibrated parameters of the simulation model

Parameter	Parameter partition									
	I	II	III	IV	V	VI	VII	VIII	C2	C3
Hydraulic conductivity (m/d)	15.5	18	22	20.5	38.5	20	28	28.5	0.25	8.6×10^{-4}
Specific storage (m^{-1})	0.00015	0.00014	0.00018	0.0002	0.00006	0.00013	0.00005	0.00005	0.01	3.3×10^{-6}
Porosity	0.25	0.27	0.25	0.3	0.35	0.3	0.3	0.3	0.4	0.09
Longitudinal dispersivity (m)	54.5	56	58.6	62.6	68	62.6	60	61	60	60
Ratio of horizontal transverse to longitudinal dispersivity	0.1									
Ratio of vertical transverse to longitudinal dispersivity	0.01									

SLR rate in Bohai was predicted to be 3.8 mm/a. For the precipitation variation, a statistical downscaling model (SDSM) was built to simulate and forecast the local response of precipitation in Longkou to global climate change under the RCP4.5 emission scenario. SDSM is a statistical downscaling method coupled with multiple regression and a weather generator (Huang *et al.* 2011), and has been extensively utilized in regional climate change prediction. Due to the wide application of SDSM, this paper will not elaborate this method in detail. The remaining

boundary conditions, groundwater pumping rates, and hydrogeological parameters remain unchanged. Inputting the above data into the SI model, the result of SI in Longkou after 50 years was obtained, as shown in Figure 5(a) (compared with SI on June 1, 2017).

Groundwater resources management

According to Figure 5(a), the extent of SI will significantly deteriorate in the western area of Longkou after 50 years.
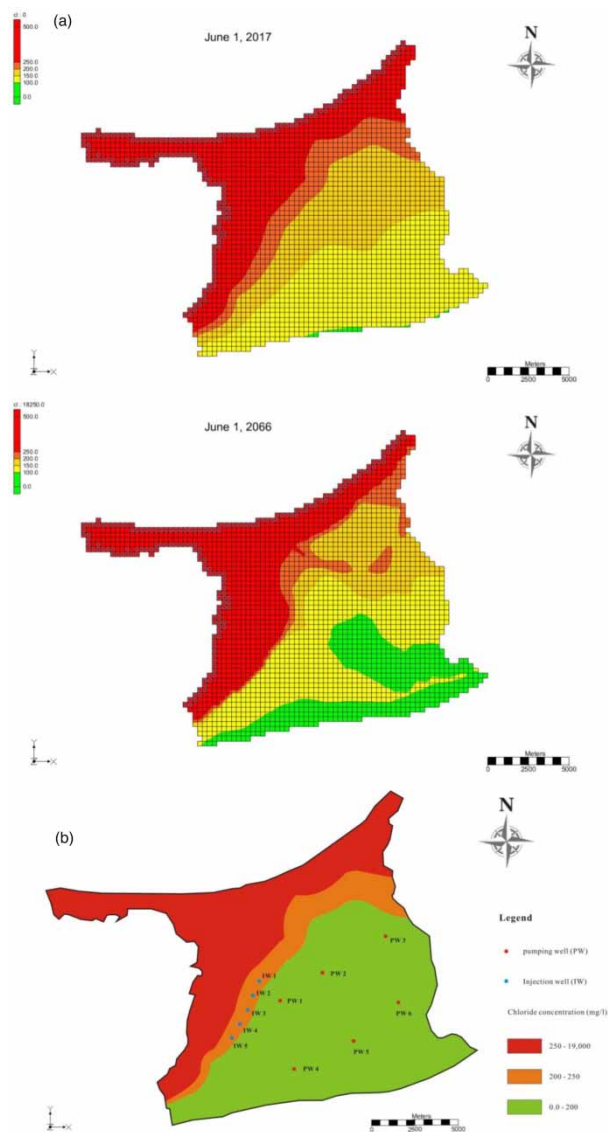

Figure 5 │ (a) States of SI in 2017 and 2066, (b) location of the pumping and injection wells.

This is because the water consumption is huge in the western area where the city center is located. In this case, the P-I groundwater management approach was presented in this study. The management objective is to control and mitigate the extent of SI. The key area to be protected should be the western region, because it is not only the area with the most serious SI, but also with the largest population density. As shown in Figure 5(b), five injection wells were designed to protect the aquifer in the western region from SI and six pumping wells were selected to supply water resources for the injection wells. The sites of the injection wells are in the transition zone of the initial SI state. The water quantity of the six pumping wells can be allocated and the water quality is good. When the water injection wells work, a hydraulic barrier will be formed in the aquifer to isolate the saltwater and the usable groundwater, so that the saltwater cannot further invade the aquifer, and the goal of mitigating the degree of SI is achieved. The total demand of water for the six pumping wells is 5,000 m³/d in order to satisfy local development. The lower pumping rate bound of pumping well 1 (PW1) is set as 500 m³/d considering the need of reducing the pumping rate encouraged by local government, and those of the other five pumping wells are assigned to 900 m³/d. Each of the pumping and injection wells has a maximum pumping or injection capacity of 3,000 m³/d. At this point, the optimization model has been preliminarily established. After solving the optimization model, the optimal groundwater management scheme can be obtained to minimize the extent of SI.

RESULTS AND DISCUSSION

The establishment and evaluation of surrogate models

The input variables are the pumping or injection rates of the eleven wells, and the percentage of chloride mass increase from 2017 to 2066 is taken as the target for the training and verification of the surrogate models. The Latin hypercube sampling (LHS) method (McKay *et al.* 1979) was applied to collect the input data. The established SI simulation model was run to acquire the corresponding output

data. In this way, the input–output data set was obtained to build surrogate models. One hundred sets of input–output data were selected via this approach to train the surrogate models. An additional 20 sets of input–output data were generated by the same approach to verify and evaluate the approximation capability of the established surrogate models to the simulation model.

The principle of LHS is to divide the entire design space equally into non-overlapping sub-intervals with equal probability. In each sub-interval, only one sample point of each variable is selected to ensure that the sampling points cover the entire design space. This method solves the problem of the accumulation of sampling points generated by the simple random sampling method, making the distribution of sampling points in the design space more uniform. The surrogate model established on this basis can more accurately reflect the information of the simulation model (Davey 2008). In the present study, there is an equivalent relationship between the pumped water quantity and the injected water quantity. In this case, the LHS method was used to sample the pumping rates of each pumping well first. Then the sampling data of different pumping wells were combined randomly to form multiple sets of water pumping schemes. Finally, the total quantity of injected water under each set of pumping schemes was calculated and the simple random sampling method was used to sequentially sample the water injection rate of each injection well. For example, one set of the pumping rates sampled by the LHS method was 2,386, 2,354, 1,596, 2,883, 1,567, and 1,004 m^3/d. According to Equation (13), the total quantity of injected water was 6,790 m^3/d. Then the value of 1,093 was sampled from the interval [0, 6,790] for the injection rate of IW1 by the simple random sampling method. Thus, the remaining total quantity of injected water was 5,697 m^3/d. Again, the value of 2,619 was sampled from the interval [0, 5,697] for the injection rate of IW2 by the simple random sampling method. The injection rates of IW3–4 were also obtained based on this rule, and the last total injected water quantity was assigned to IW5. In this way, 120 sets of input–output data were generated and used for the training and verification of the surrogate models, which is summarized in Table 2.

After obtaining the input–output data set, the Kriging and SVR surrogate models were developed and the

Table 2 | Sampling results of training and verification samples

No.	Pumping well (m³/d)						Injection well (m³/d)					Mass increase (%)
	1	2	3	4	5	6	1	2	3	4	5	
1	2,386.20	2,354.17	1,595.81	2,883.31	1,567.32	1,003.93	375.97	1,093.43	2,619.58	163.62	2,538.13	82.81
2	2,516.01	2,696.60	2,045.67	1,979.61	2,352.42	2,623.22	2,028.69	2,912.10	1,501.20	1,650.72	1,120.82	77.07
3	817.25	1,517.07	2,492.07	2,721.03	1,515.05	1,945.67	1,127.27	2,523.73	1,298.29	541.55	517.30	80.62
4	2,591.87	1,127.11	1,090.65	2,392.11	2,127.16	2,260.44	2,590.38	235.64	2,712.85	525.44	525.03	80.44
5	1,644.05	1,371.08	2,533.98	1,835.55	2,254.48	1,389.76	875.93	712.80	1,890.55	167.07	2,382.57	82.10
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
116	622.35	1,862.65	2,653.63	1,105.69	1,054.37	1,494.49	732.26	1,898.86	543.83	184.14	434.10	84.58
117	1,394.49	1,283.37	1,341.02	1,488.73	2,055.75	2,490.16	2,410.16	2,131.28	369.05	65.22	77.82	81.18
118	2,463.99	1,476.87	2,281.28	1,935.17	2,554.58	1,135.02	2,471.91	2,065.99	1,739.91	200.57	368.54	80.43
119	2,343.71	2,194.78	1,038.87	2,251.92	1,505.47	1,323.59	2,556.57	962.84	985.61	1,078.66	74.66	82.78
120	688.95	2,409.82	1,494.81	2,893.35	1,478.09	1,001.20	1,401.76	1,594.95	804.56	80.52	1,084.45	83.08

ensemble model was constructed based on the theory of each method through the MATLAB platform. The parameters of the three surrogate models are summarized in Table 3. The results of verifying the surrogate models are shown in Figure 6, demonstrating the discrepancy between the predicted values of mass increase ratio and the calculated values from the SEAWAT simulation model. The prediction accuracy of the surrogate models was evaluated by root mean square error (RMSE) and determination coefficient (R^2). Their expressions are given by:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{18}$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \tag{19}$$

where $\hat{y}_i$ is the predicted value of the surrogate model for the ith sample; y_i is the exact value of the SI simulation model for the ith sample; $\bar{y}$ is the mean value of exact values; n is the number of the sample. Based on the verification results, the RMSE and R^2 of the three established surrogate models are shown in Table 4. As can be seen from the table, in the two stand-alone surrogate models, the accuracy of the Kriging surrogate model is higher, with a smaller RMSE value and a larger R^2 value, indicating that the Kriging method has a stronger advantage in establishing a surrogate model on this issue. Comparing the Kriging-SVR ensemble
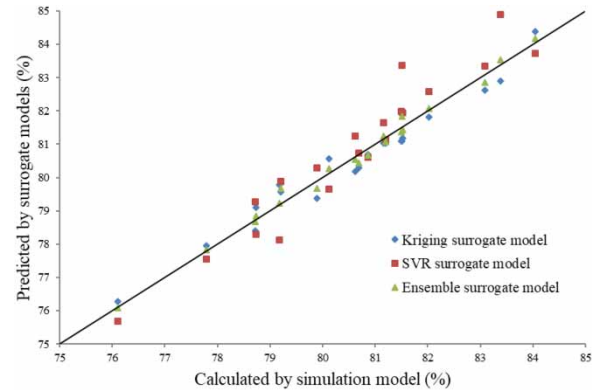

Figure 6 │ Comparison between simulation model and surrogate models.

Table 4 │ Precision comparison between surrogate models

Index	R^2	RMSE
Kriging surrogate model	0.983175	0.003685
SVR surrogate model	0.962183	0.007067
Ensemble surrogate model	0.995397	0.001806

surrogate model with the stand-alone surrogate models, it can be found that the forecast precision of the ensemble surrogate model is better. This demonstrates that the ensemble surrogate model has advantages over the stand-alone surrogate models because it can offset some errors of the stand-alone surrogate models. As a result, the Kriging-SVR ensemble surrogate model was selected in this study for calculations instead of the simulation model during the optimization process.

Table 3 │ The parameters of the three surrogate models

Kriging surrogate model				SVR surrogate model		Ensemble surrogate model	
parameter	value	parameter	value	parameter	value	parameter	value
θ_1	4.87	θ_7	10.00	C	67.907	ω_K	0.6737
θ_2	5.00	θ_8	10.00				
θ_3	3.63	θ_9	10.00	γ	0.075		
θ_4	3.49	θ_{10}	10.00			ω_S	0.3263
θ_5	2.93	θ_{11}	10.00	ε	0.01		
θ_6	2.89	σ^2	0.00003				

ω_K: weight coefficient of Kriging surrogate model; ω_S: weight coefficient of SVR surrogate model.

Optimal management schemes

Three management scenarios with upper limits of the total injection rate of respectively 10,000 m³/d, 8,000 m³/d and 6,000 m³/d were designed so as to provide more choices for decision-makers. In this way, based on the built ensemble surrogate model, an optimization model with the objective of minimizing the extent of SI was constructed. The genetic algorithm (GA) was utilized to solve this optimization model, and the optimal solution results of the three management scenarios are given in Table 5. The SI states of the three management scenarios under the optimal solution condition are shown in Figure 7. It can be seen that as PW1 is closest to the SI area, its pumping rate is obviously the smallest of the six pumping wells. More water used for injection comes from the other five pumping wells, whose water extraction rates have less impact on SI. Among the injection wells, IW5 requires the smallest quantity of water, because the SI rate is relatively slow at the position of IW5 according to Figure 5(a). It is noteworthy that although a larger total injection rate can lead to better SI mitigation, the project cost will also rise due to the increase in the amount of pumping, transporting and injecting of water. Decision-makers can choose the most appropriate management plan according to local conditions.

At the initial stage, the chloride mass in the aquifer of the whole study area is 3.19×10^{12} kg. If without the measure of P-I groundwater management, the percentage of chloride mass increase in the aquifer at the end of the management horizon is 90.75%, much larger than the optimal solution results of the three scenarios, which proves the effectiveness of the P-I management strategy. It not merely protects the groundwater quality of important areas, but also controls the degree of SI in the whole study area. Furthermore, taking the total water injection rate as 10,000 m³/d, another water allocation scheme, in which pumping wells share the required water quantity equally and the injection rate of each water injection well is the same, was designed for comparison. The chloride mass of this scheme at the end of the management horizon was calculated by the SI simulation model. The calculation result shows that the percentage of chloride mass increase is 76.61%, almost equivalent to the percentage of Scenario 2 (8,000 m³/d total water injection rate) under the optimal solution condition, indicating that

Table 5 | The optimal management schemes of three scenarios

No.	Pumping well (m³/d)						Injection well (m³/d)					Mass increase (%)
	1	2	3	4	5	6	1	2	3	4	5	
Scenario 1	1,582.25	2,090.77	2,651.67	2,831.64	2,909.13	2,934.55	2,020.36	2,270.57	2,246.69	2,428.66	1,033.71	74.33
Scenario 2	650.58	1,824.82	1,991.59	2,550.62	2,982.94	2,999.43	1,695.60	1,827.89	1,784.25	1,899.86	792.39	76.32
Scenario 3	547.37	1,457.76	1,686.08	2,192.62	2,202.99	2,913.17	1,273.16	1,340.66	1,303.16	1,429.12	653.90	79.57

Scenario 1: Q_{max} = 10,000 m³/d; Scenario 2: Q_{max} = 8,000 m³/d; Scenario 3: Q_{max} = 6,000 m³/d.
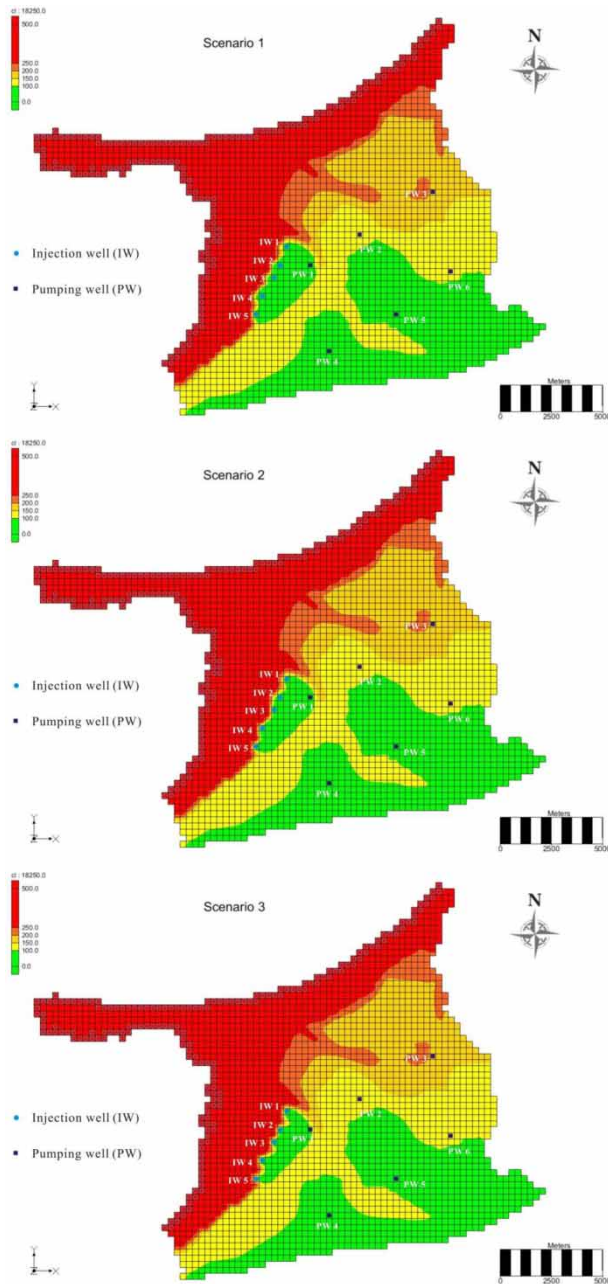
Figure 7 | The SI results of three scenarios under the optimal solution condition.

the government can pay less to achieve the same effect of SI alleviation. It can be seen that the groundwater management strategy of P-I based on the S-O approach can minimize the extent of SI under the constraints of existing conditions by maximizing the positive impact of water injection and minimizing the negative impact of pumping. For the coastal

areas lacking surface freshwater resources, this strategy is very efficient and should be considered in the prevention and control of SI in the future.

Comparison of optimization results from Kriging-SVR and SEAWAT

As mentioned previously, the selected Kriging-SVR ensemble surrogate model was coupled with GA to solve the optimization model. In order to investigate the quality of surrogate-based solutions, a comparison of solutions obtained from the ensemble surrogate model and the simulation model was carried out. For the sake of clarity, GA was combined with the surrogate model and simulation model SEAWAT as GA-KS and GA-SEAWAT, respectively. The optimization results obtained from GA-KS and GA-SEAWAT based on three management scenarios are shown in Figure 8. It can be seen that the solutions of GA-KS, including water allocation of pumping and injection wells and ratio of mass increase, are very close to those of GA-SEAWAT in all three scenarios. The detailed relative errors between the optimization results from GA-KS and GA-SEAWAT are summarized in Table 6.

Operating the SI simulation model each time took 150 s. Therefore, during the process of solving the optimization model, if the simulation model is called directly, the time required for 3,000 iterations of the optimization algorithm is 125 h, while the ensemble surrogate model just requires about 0.5 h. Notably, although a large amount of time has been spent on training, developing and validating the ensemble surrogate model, it is still far less than the time required to repeatedly call the simulation model. The comparison results show that the performance of the Kriging-SVR ensemble surrogate model is almost the same as the original simulation model in coupling with GA to solve the optimization model, while using the Kriging-SVR ensemble surrogate model can reduce the huge computational burden.

CONCLUSIONS

Using the S-O method, this study presented an application of the P-I groundwater management pattern for SI mitigation in Longkou City, China. A three-dimensional variable
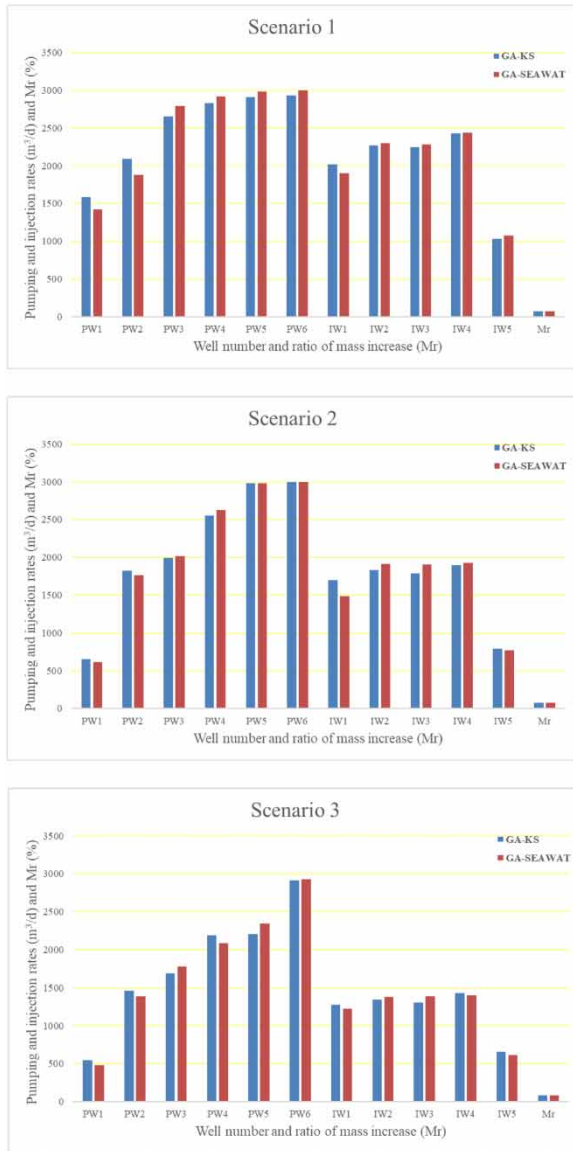
Figure 8 | The optimal solutions based on GA-KS and GA-SEAWAT.

density groundwater mathematical model was built for SI prediction of the Longkou area, considering the effects of SLR and precipitation variation in response to climate change on SI. The Kriging surrogate model, the SVR surrogate model and the Kriging-SVR ensemble surrogate model were established for the SI simulation model. Based on the appropriate indexes, the most accurate surrogate model was selected to couple the optimization management model. The GA was utilized to solve the optimization model, thereby obtaining the optimal pumping and water injection scheme. Two conclusions can be drawn from this research.

First, compared with stand-alone surrogate models, the Kriging-SVR ensemble surrogate model has a higher accuracy. This indicates that combining stand-alone surrogate models is a potential surrogate modeling method for the variable-density groundwater simulation model. Future studies can consider this technology when stand-alone surrogate models show a poor accuracy. Besides, calling the surrogate model instead of the simulation model for calculation during the process of solving the optimization model can tremendously save computing costs while ensuring accuracy.

Second, the P-I management strategy based on the S-O method can effectively control the extent of SI by properly managing groundwater resources. This strategy is especially efficient for regions that lack adequate surface freshwater resources to implement artificial recharge measures. The design of multiple scenarios for the upper limit of total water injection can provide decision-makers with multiple decision-making schemes. That is, decision-makers can make a balance between financial expenditure and the degree of SI mitigation.

Table 6 | The comparison of solutions from GA-KS and GA-SEAWAT

Scenario 1				Scenario 2				Scenario 3			
	RE (%)		RE (%)		RE (%)		RE (%)		RE (%)		RE (%)
PW1	11.16	IW1	6.39	PW1	6.05	IW1	14.38	PW1	14.69	IW1	4.29
PW2	11.14	IW2	1.35	PW2	3.51	IW2	4.51	PW2	5.02	IW2	2.95
PW3	5.17	IW3	1.71	PW3	1.13	IW3	6.27	PW3	5.03	IW3	6.20
PW4	2.94	IW4	0.33	PW4	2.92	IW4	1.51	PW4	5.19	IW4	2.24
PW5	2.44	IW5	3.99	PW5	0.03	IW5	2.81	PW5	6.18	IW5	7.07
PW6	2.18	Mr	0.90	PW6	0.01	Mr	0.57	PW6	0.47	Mr	0.58

RE: relative error; Mr: ratio of mass increase.

ACKNOWLEDGEMENTS

DATA AVAILABILITY STATEMENT

All relevant data are included in the paper or its Supplementary Information.

REFERENCES

Abarca, E., Vázquez-Suñé, E., Carrera, J., Capino, B., Gámez, D. & Batlle, F. 2006 Optimal design of measures to correct seawater intrusion. *Water Resources Research* **42** (9), W09415.

Armanuos, A. M., Ibrahim, M. G., Mahmod, W. E., Takemura, J. & Yoshimura, C. 2019 Analysing the combined effect of barrier wall and freshwater injection countermeasures on controlling saltwater intrusion in unconfined coastal aquifer systems. *Water Resources Management* **33** (4), 1265–1280.

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

Bear, J. 1972 *Dynamics of Fluids in Porous Media*. Dover Publications, Incorporated, Mineola, NY, USA.

Christelis, V. & Mantoglou, A. 2016 Pumping optimization of coastal aquifers assisted by adaptive metamodelling methods and radial basis functions. *Water Resources Management* **30** (15), 5845–5859.

Davey, K. R. 2008 Latin hypercube sampling and pattern search in magnetic field optimization problems. *IEEE Transactions on Magnetics* **44** (6), 974–977.

Guo, J. Y., Lu, W. X., Yang, Q. C. & Miao, T. S. 2019 The application of 0–1 mixed integer nonlinear programming optimization model based on a surrogate model to identify the groundwater pollution source. *Journal of Contaminant Hydrology* **220**, 18–25.

Hou, Z. Y. & Lu, W. X. 2018 Comparative study of surrogate models for groundwater contamination source identification at DNAPL-contaminated sites. *Hydrogeology Journal* **26** (3), 923–932.

Huang, P. S. & Chiu, Y. C. 2018 A simulation–optimization model for seawater intrusion management at Pingtung Coastal Area, Taiwan. *Water* **10** (3), 251.

Huang, J., Zhang, J. C., Zhang, Z. X., Xu, C. Y., Wang, B. L. & Yao, J. 2011 Estimation of future precipitation change in the Yangtze River basin by using statistical downscaling method. *Stochastic Environmental Research and Risk Assessment* **25** (6), 781–792.

Ketabchi, H., Mahmoodzadeh, D., Ataie-Ashtiani, B. & Simmons, C. T. 2016 Sea-level rise impacts on seawater intrusion in coastal aquifers: review and integration. *Journal of Hydrology* **535**, 235–255.

Kourakos, G. & Mantoglou, A. 2015 An efficient simulation–optimization coupling for management of coastal aquifers. *Hydrogeology Journal* **23** (6), 1167–1179.

Lahiri, S. K. & Ghanta, K. C. 2009 Support vector regression with parameter tuning assisted by differential evolution technique: study on pressure drop of slurry flow in pipeline. *Korean Journal of Chemical Engineering* **26** (5), 1175–1185.

Langevin, C. D., Shoemaker, W. B. & Guo, W. 2003 *MODFLOW-2000, the U.S. Geological Survey Modular Ground-Water Model – Documentation of the SEAWAT2000 Version with the Variable-Density Flow Process (VDF) and the Integrated MT3DMS Transport Process (IMT)*, Open-File Report 03-426. US Geological Survey, Denver, CO, USA.

McKay, M. D., Beckman, R. J. & Conover, W. J. 1979 A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics* **21** (2), 239–245.

Miao, T. S., Lu, W. X., Guo, J. Y., Lin, J. & Fan, Y. 2019 Modeling and uncertainty analysis of seawater intrusion based on surrogate models. *Environmental Science and Pollution Research* **26** (25), 26015–26025.

Ouyang, Q., Lu, W. X., Miao, T. S., Deng, W. B., Jiang, C. L. & Luo, J. N. 2017 Application of ensemble surrogates and adaptive sequential sampling to optimal groundwater remediation design at DNAPLs-contaminated sites. *Journal of Contaminant Hydrology* **207**, 31–38.

Song, J., Yang, Y., Wu, J. F., Wu, J. C., Sun, X. M. & Lin, J. 2018 Adaptive surrogate model based multiobjective optimization for coastal aquifer management. *Journal of Hydrology* **561**, 98–111.

Vapnik, V. N. 1995 *The Nature of Statistical Learning Theory*. Springer, New York, USA.

Viana, F. A. C., Haftka, R. T. & Steffen, V. 2009 Multiple surrogates: how cross-validation errors can help us to obtain the best predictor. *Structural and Multidisciplinary Optimization* **39** (4), 439–457.

Viana, F. A. C., Haftka, R. T. & Watson, L. T. 2013 Efficient global optimization algorithm assisted by multiple surrogate techniques. *Journal of Global Optimization* **56** (2), 669–689.

Yang, Y., Wu, J. F., Lin, J., Wang, J. G., Zhou, Z. F. & Wu, J. C. 2018 An efficient simulation–optimization approach for controlling seawater intrusion. *Journal of Coastal Research* **84** (Sp 1), 10–18.

Zheng, C. & Wang, P. P. 1999 *MT3DMS: A Modular Three-Dimensional Multispecies Transport Model for Simulation of Advection, Dispersion, and Chemical Reactions of Contaminants in Groundwater Systems; Documentation and User's Guide*. US Army Engineer Research and Development Center, Vicksburg, MS, USA.