


Modeling saltwater intrusion using an integrated Bayesian model averaging method in the Pearl River Delta

Kairong Lin, Pengyu Lu, Chong-Yu Xu , Xuan Yu, Tian Lan and Xiaohong Chen

ABSTRACT

The reverse flow of seawater causes salinity in inland waterways and threatens water resources of the coastal population. In the Pearl River Delta, saltwater intrusion has resulted in a water crisis. In this study, we proposed a tailored approach to chlorinity prediction at complex delta systems like the Pearl River Delta. We identified the delayed predictors prior to optimization based on the maximal information coefficient (MIC) and Pearson's correlation coefficient (r). To achieve an ensemble simulation, a Bayesian model averaging (BMA) method was applied to integrate temporally sensitive empirical model predictions given by random forest (RF), support vector machine (SVM), and Elman neural network (ENN). The results showed that: (a) The ENN performed the worst among the three; (b) The BMA approach outperformed the individual models (i.e., RF, ENN, and SVM) in terms of Nash–Sutcliffe efficiency (NSE), and the percentage of bias (Pbias). The BMA weights reflect the model performance and the correlation of the predictions given by its ensemble models. (c) Our variable selection method resulted in a stronger model with greater interpretability.

Key words | Bayesian model averaging, chlorinity, maximal information coefficient, Modaomen waterway, saltwater intrusion

Kairong Lin (corresponding author)

Pengyu Lu

Tian Lan

Xiaohong Chen

Center for Water Resources and Environment,
Sun Yat-sen University,
Guangzhou,
China

E-mail: linkr@mail.sysu.edu.cn

Kairong Lin

Xiaohong Chen


Southern Marine Science and Engineering
Guangdong Laboratory (Zhuhai),
China

Kairong Lin

Xuan Yu

Xiaohong Chen

School of Civil Engineering,
Sun Yat-sen University,
Guangzhou,
China

Chong-Yu Xu 

Department of Geosciences,
University of Oslo,
Oslo,
Norway

INTRODUCTION

Saltwater intrusion is becoming a global issue in coastal zones due to sea-level rise (Nicholls & Cazenave 2010), storm surges (Howes *et al.* 2010), and water withdrawal (Wu *et al.* 2016). Deltas are highly sensitive to saltwater intrusion risk due to local human activities and complex dilution and mixing mechanisms (Tessler *et al.* 2015). It is challenging to understand intrusion processes for salinity prediction and water resource management (Passeri *et al.* 2015).

This is an Open Access article distributed under the terms of the Creative Commons Attribution Licence (CC BY-NC-ND 4.0), which permits copying and redistribution for non-commercial purposes with no derivatives, provided the original work is properly cited (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

doi: 10.2166/hydro.2019.073

Physics-based numerical models are widely applied to saltwater prediction. Wang *et al.* (2012) explored the dynamic mechanism of saltwater intrusion and the influencing factors with a high-resolution three-dimensional (3D) numerical model based on the Finite Volume Coastal Ocean Model (FVCOM), which covered the entire river network, the Pearl River Estuary, and the adjacent sea. Gingerich & Voss (2005) designed a 3D-SUTRA model of groundwater flow and solute transport in the Pearl Harbor aquifer in southern Oahu, Hawaii by incorporating a two-dimensional (2D) cross-sectional model with a realistic areal distribution of the hydrogeological structure, recharge, and pumping. Yu *et al.* (2016) explored the role of a river network in saltwater intrusion due to storm surges using a coupled surface-subsurface

model HydroGeoSphere. The performance of numerical models relies on large amounts of available observed chlorinity and hydrological data, high-resolution estuarine topography data, as well as significant efforts for setting up, code compiling, and computational resources.

Salinity prediction of saltwater intrusion has benefited from recent advances in machine learning techniques. Various machine learning techniques have been applied to empirical saltwater intrusion prediction models (Liu *et al.* 2014). Dong *et al.* (2010) developed a coupled neural network model combining backpropagation and a real coding-based accelerating genetic algorithm (BP-RAGA) to predict saltwater intrusion in the Ping Gang water source of Zhuhai city and used the tidal range of the water resources and the observed flows in the upstream hydrological station for the last day as influencing factors. Lal & Datta (2018) investigated the feasibility of using support vector machine (SVM) for predicting salinity concentrations at selected monitoring wells in an illustrative aquifer under variable groundwater pumping conditions. Unlike sophisticated physics-based numerical models, empirical models provide a fast response based on observations specific to the study area in a quick manner (Rohmer & Brisset 2017), which can identify the highly complex interactions between influencing factors and output variables.

Empirical models require the selection of key factors affecting chlorinity. Usually, the influencing factors are the antecedent conditions. The chlorinity in areas of saltwater intrusion not only has its own periodic characteristics but also is subject to many external forcings, including river discharge, tides, and wind (e.g., Leonard 1976; Mao *et al.* 2004; Ralston *et al.* 2008; Becker *et al.* 2010; Uncles & Stephens 2011). Previous studies on empirical saltwater prediction models have used the influencing factors for the current day or the last day as predictors (e.g., Dong *et al.* 2010; Yoon *et al.* 2017) and neglected to quantify the inner periodic characteristics of chlorinity, as well as the delay for the chlorinity to respond to its factors in the preprocessing stage. It has been shown that the model interpretability is improved and the generalization is enhanced by reducing overfitting when the predictors most directly linked to the predictions are selected in the preprocessing stage based on a correlation analysis (Tan *et al.* 2014). Relationship coefficients are usually used in a correlation

analysis to measure the attribute similarity. Pearson's correlation coefficient is widely used because it is easy to calculate. However, Pearson's coefficient (r) is only appropriate for linear relationships (Zhao *et al.* 2013). As a result of multiple impacts from the tide, upstream discharges, wind, and other factors, saltwater intrusion is a highly complex process containing both linear and nonlinear signals (Liu *et al.* 2014). The maximal information coefficient (MIC), which was proposed by Reshef *et al.* (2011), can measure a wide range of possible dependencies, including linear and nonlinear associations. Consequently, we expect that the MIC and Pearson's correlation coefficient can be used to capture the nonlinear and linear relationships between saltwater intrusion and the influencing factors.

Although the random forest (RF), SVM, and Elman neural network (ENN) have been successfully used for salt tide prediction, they are black-box models and have their own limitation and uncertainty. A multi-model ensemble strategy is required to exploit and integrate the diversity of a skillful prediction from different models and to reduce their uncertainty. Bayesian model averaging (BMA) is a weighted average of a number of individual models. Various studies have shown that BMA outperformed individual models. Neto *et al.* (2018) used BMA to merge different model configuration results into a single streamflow probabilistic prediction; the BMA simulation showed higher accuracy and precision than all simulations produced by the ensemble members. Zhang *et al.* (2015) analyzed the uncertainty of three reservoir operating rules using the BMA method; the results showed that BMA reduced the uncertainty of the operating rules, which is of great potential benefit for evaluating the confidence interval of the decisions. The expectation-maximization (EM) method is easy to implement for the estimation of the BMA weights (e.g., Raftery *et al.* 2016; Liu & Merwade 2018). Therefore, we expect that BMA can be used to combine machine learning empirical models for better chlorinity prediction.

The Pearl River Delta is considered one of the world's most complex river networks. The river network density is around 0.8 km/km². The river channels of Pearl River Delta have changed rapidly due to climate change and human activities, which have a significant impact on the hydrological processes and the environment (Lin *et al.* 2017). Due to the complexity of the estuary circulation and

the density of its river network, it is difficult to develop process-driven numerical models covering the entire estuary for a wide range of spatial and temporal scales and account for a variety of physical processes.

The main goal of the study is to develop a correlation-based selection process and flexible chlorinity predicting approach tailored for complex delta systems like the Pearl River Delta. To achieve this objective, we: (a) use the MIC and Pearson's correlation coefficient (r) to quantify the periodic characteristic of chlorinity and the time-lag for chlorinity in response to its influencing factors in an important waterway of the Pearl River Delta; (b) develop four flexible model series with different sets of predictors to predict chlorinity for the river delta and for each model series, SVM, RF, and ENN are used; (c) use BMA to integrate SVM, RF, and ENN to generate an integrated and robust empirical statistical

saltwater prediction model with high accuracy. Our study demonstrates the advantages of our integrated temporally sensitive empirical model for chlorinity prediction, which may be applicable in other complex delta systems.

METHODOLOGY

The framework of the BMA ensemble approach

The framework of the BMA ensemble approach is summarized as follows (Figure 1). The details about MIC, r , BMA, RF, ENN, SVM are presented in the Supplementary material.

1. The MIC and r are used to determine the periodic characteristic of chlorinity time series and the time-lag for

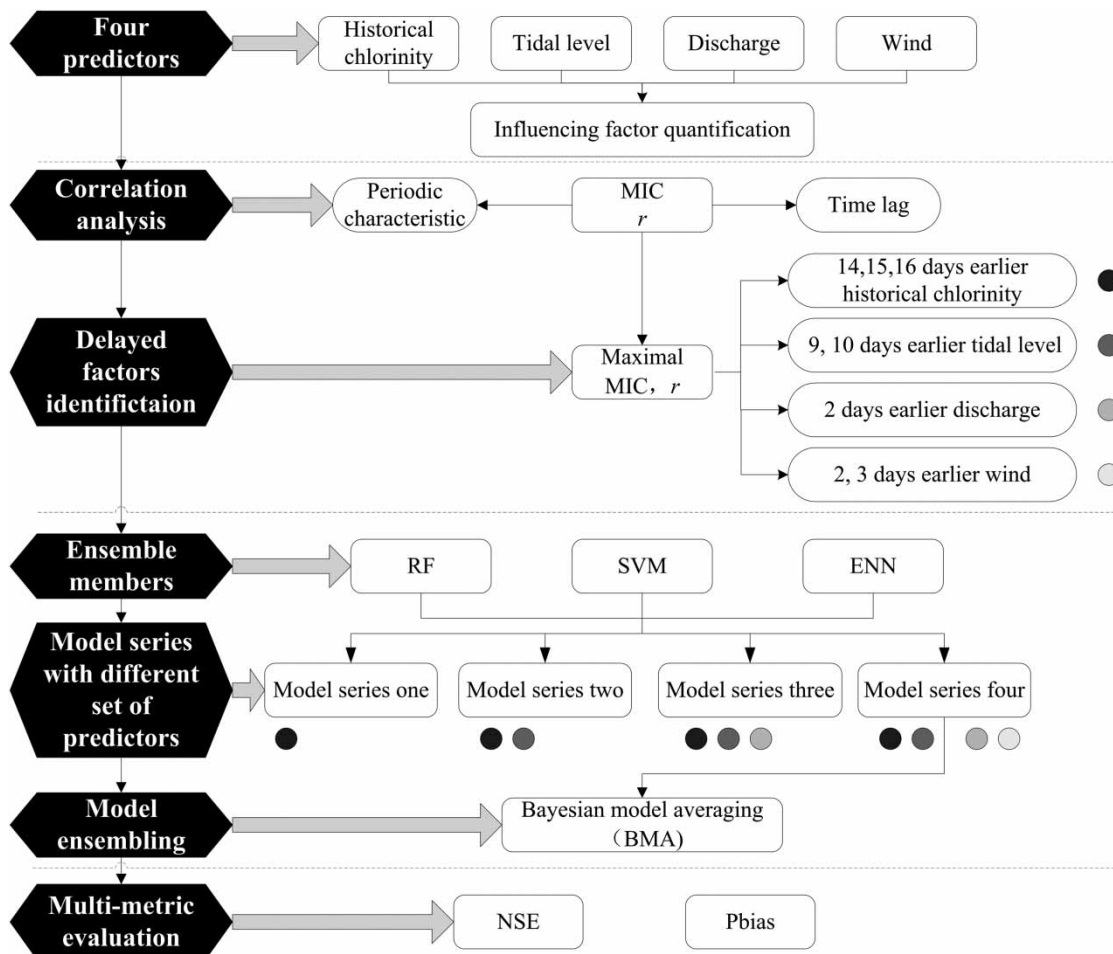


Figure 1 | Flowchart of the chlorinity ensemble-based forecast generation using Bayesian model averaging method.

chlorinity to respond to its influencing factors. It is assumed that the delayed predictor with the maximal MIC or the maximal r value has the strongest correlation with the target chlorinity.

2. Time series of historical chlorinity, tidal level, discharge, and wind are selected as model inputs. To clarify the effect of the influencing factors on the chlorinity prediction, four series of models are developed: (1) Model one using historical chlorinity as the only factor; (2) Model two using historical chlorinity and tidal level as factors; (3) Model three using historical chlorinity, tidal level, and discharge as factors; and (4) Model four using historical chlorinity, tidal level, discharge, and wind as factors. For each model, the RF, SVM, and ENN models are used to predict the chlorinity at Pinggang station.
3. The BMA ensemble model is derived based on the predictions given by RF, SVM, and ENN.

Model training and testing criteria

In this study, model parameters are optimized to maximize Nash and Sutcliffe efficiency (NSE) between observed and calculated values, which are shown in Equation (1). NSE is a widely used efficiency criterion for hydrological model assessment (e.g., Nash & Sutcliffe 1970; Lin et al. 2014). Grid search approach is used to derive the optimal parameter set for each model. Five-fold cross-validation is used to account for the bias caused by data randomness. Prior to model training, the model inputs are normalized to $[-1, 1]$.

Models are evaluated using NSE and percent bias (Pbias), which are shown in Equations (1) and (2). Pbias and NSE have been commonly used in pairs in hydrological model evaluation (Viney et al. 2009; Vaze et al. 2010; Yang et al. 2018). NSE is not very sensitive to systematic model over-prediction or under-prediction (Krause et al. 2005), so an additional criterion, Pbias, is needed to measure the overall adequacy between distributional statistics of predicted and observed data (Hauduc et al. 2015):

$$NSE = 1 - \frac{\sum_{i=1}^n (P_i - O_i)^2}{\sum_{i=1}^n (O_i - \bar{O})^2} \quad (1)$$

$$Pbias = 100 \times \frac{\bar{P} - \bar{O}}{\bar{O}} \quad (2)$$

where O_i represents the observed chlorinity and P_i represents the simulated chlorinity.

The modeling processes were all programmed on Matlab.

STUDY AREA AND DATA

Study site

The Modaomen waterway (Figure 2(b)), one of the eight outlets in the Pearl River Delta (Figure 2(a)), is located downstream of the West River between latitudes 22°N and $22^\circ40'\text{N}$ and longitudes 113°E and $113^\circ40'\text{E}$. Among eight outlets, the Modaomen waterway runs from north-northwest to south-southeast, and has the largest (about one-third) river discharge of the runoff from the Pearl River system (Zhang et al. 2010).

In the Pearl River Delta of China, the Modaomen waterway is the main source of drinking water for the cities of Jiangmen, Zhongshan, Zhuhai, and Macao. Saltwater intrusion occurs in the dry season, such as on November 1, 2005, to February 28, 2006, and greatly affects the residential and industrial water supply.

Due to the importance of water supply security for the Pearl River Delta, the Modaomen waterway in the Pearl River Delta was selected as the study area to demonstrate the applicability of the BMA ensemble approach using the RF, SVM, and ENN-based empirical models.

Data collection

Saltwater intrudes upstream during the dry season from November to February as the river discharge decreases and tidal mixing increases. Hence, the data sets in this study were collected during the dry seasons from 2005 to 2008 (Table 1).

The field data consist of daily chlorinity levels at Pinggang station, daily tidal levels at Denglongshan gauging station, daily discharges at the Makou hydrologic station, and daily wind measurements at Zhongshan station. The total length of the continuous monitored target chlorinity time series was 360 days (2005/11/1–2006/2/28, 2006/11/1–2007/2/28, 2007/11/1–2008/2/28). Using a 7:3 split (Klemeš 1986), the first 254 observations were used as the training set and the last 30%, 106 observations, were used as the testing set.

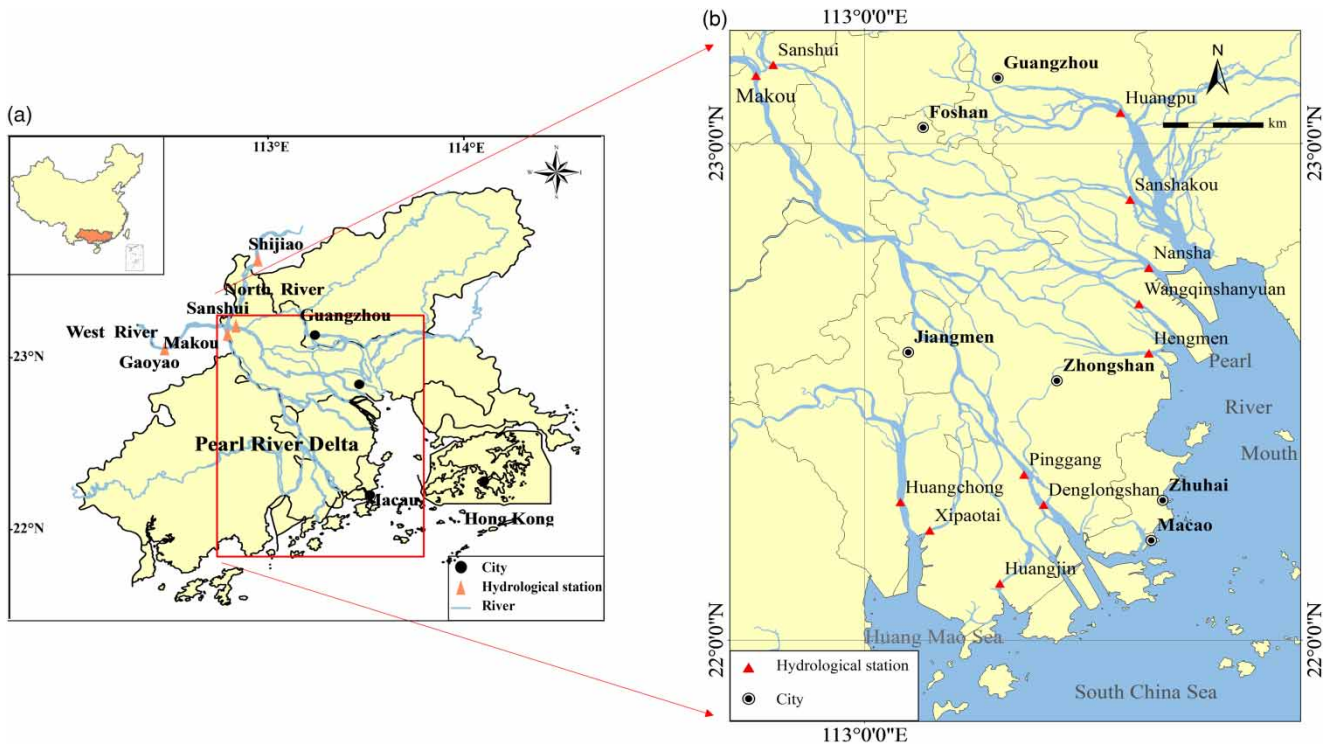


Figure 2 | (a) Map of the Pearl River Delta area; (b) map of the Modaomen waterway and the location of the hydrological stations. Note: (1) The area within the rectangle with two arrows in (a) represents the area in (b). (2) The daily chlorinity levels are collected at Pinggang station, daily tidal levels are collected at Denglongshan gauging station, daily discharges are collected at the Makou hydrologic station, and daily wind measurements are collected at Zhongshan station.

Table 1 | Data collection for salt tide analysis

	Target	Predictors for correlation analysis and model construction			
Salt tide factors	Chlorinity	Chlorinity (delayed)	Tidal level (delayed)	Discharge (delayed)	Wind (delayed)
Sample type	Continuous monitored daily observations				
Station	Pinggang	Pinggang	Denglongshan	Makou	Zhongshan
Time series	2005/11/1–2006/2/28	2005/10/1–2006/2/28			
	2006/11/1–2007/2/28	2006/10/1–2007/2/28			
	2007/11/1–2008/2/28	2007/10/1–2008/2/28			

To quantify the wind effect on the saltwater intrusion, a polar coordinate system with the positive direction north-northwest was used because the saltwater intrusion is decreased by the downstream winds and increased by the upstream winds. The wind effect on the saltwater intrusion was quantified by mapping the wind vector (F, x) with regard to the polar coordinate:

$$F' = F \times \cos\left(\frac{\pi}{8} \times x\right) \tag{3}$$

where F denotes the magnitude of the wind vector and $\frac{\pi}{8} \times x$ is the polar angle of the wind vector.

RESULTS AND DISCUSSION

Correlation analysis results

The delayed predictors were selected using the MIC and r for the correlation analysis. The auto-correlation analysis

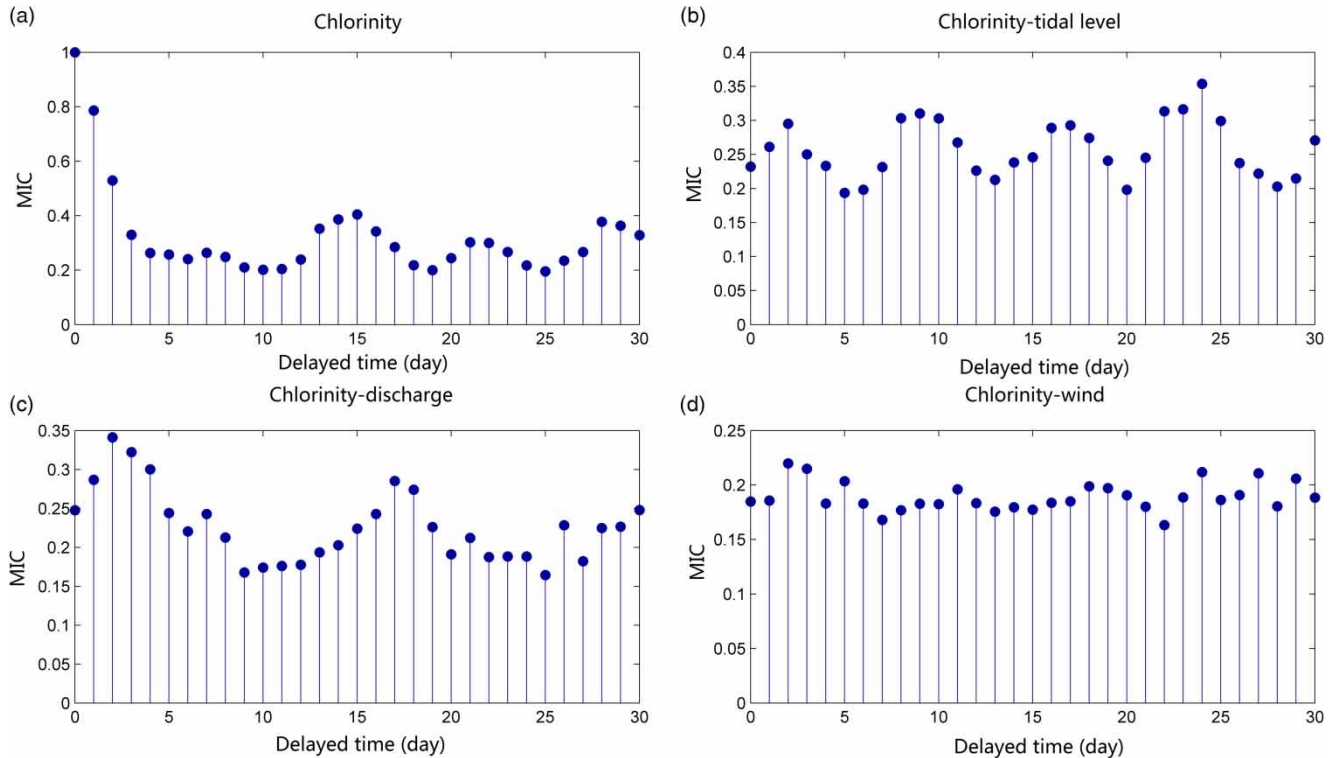


Figure 3 | Results of maximal coefficient information (MIC) correlation analysis of (a) chlorinity and between (b) chlorinity and tidal level time series; (c) chlorinity and upstream discharge time series; (d) chlorinity and wind time series. Note: The MIC correlated analysis is conducted between observed chlorinity (2005/11/1–2006/2/28, 2006/11/1–2007/2/28, 2007/11/1–2008/2/28) with 0–31 days earlier observed influencing factors (historical chlorinity, tidal level, discharge, and wind).

results of the chlorinity time series at Pinggang station and the cross-correlation analysis results between the chlorinity at Pinggang station and the tidal level at Denglongshan station, the upstream discharge at Makou station, and the wind data at Zhongshan station are shown in Figures 3 and 4.

Predictors with the maximal correlation to the target chlorinity were selected based on the correlation analysis results of the periodic characteristics and the time-lag relationships (Table 2). According to the results of MIC analysis, 14, 15, 16 days earlier chlorinity, 9 days earlier tidal level, 2 days earlier discharge, and 2 days earlier wind were selected as delayed predictors. Based on r , 10 days earlier tidal level and 3 days earlier wind were also selected.

SVM model training results

The main parameters of the SVM models were the error penalty for C and γ with ε fixed at 0.01. The NSE

distributions of C and γ for the four models are shown in Figure 3. For the x-axis and y-axis, the $\log_2 C$ and $\log_2 \gamma$ are in the range of $[-10, 10]$ with increments of 0.5. The variations of the NSE are presented in the supplementary material (Figure S1). The paired parameters (C , γ) used for the SVM models are (2.00, 0.25), (1.00, 0.13), (0.71, 0.50), and (1.44, 0.71), respectively.

RF model training results

In the RF models, $mtry$ is the number of variables randomly selected at each splitting node and $ntree$ is the number of ensemble regression trees. To search for the optimal values of $mtry$ and $ntree$, we set $ntree$ to range from 500 to 3,000 with an increment of 100 at each step. $mtry$ varies from 1 to M , where M is the number of predictors for each model, which means 1 to 3 for Model one, 1 to 4 for Model two, 1 to 5 for Model three, and 1 to 6 for Model four. The $ntree$, $mtry$ pair had 78 (26×3), 104 (26×4), 130

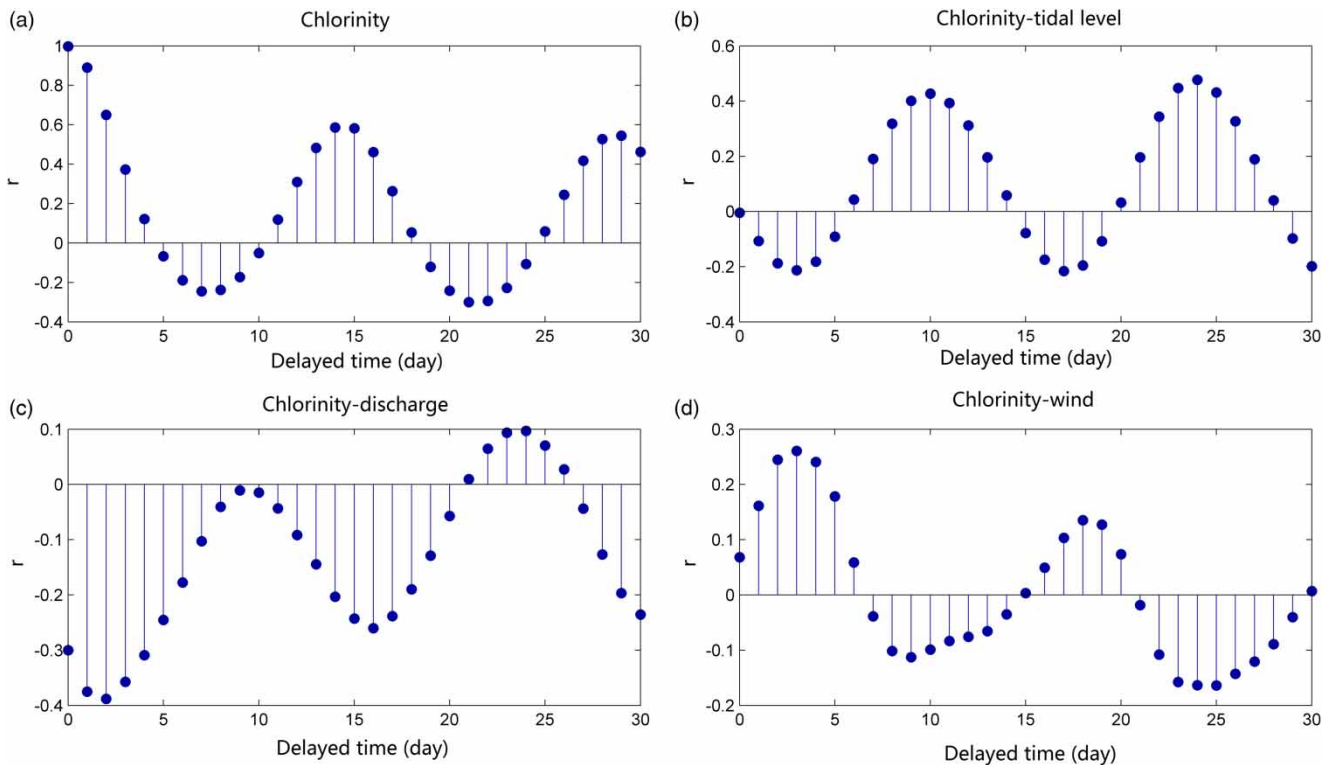


Figure 4 | Pearson's correlation coefficient (r) correlation analysis of: (a) chlorinity and between (b) chlorinity and tidal level time series; (c) chlorinity and upstream discharge time series; (d) chlorinity and wind time series. Note: The Pearson's correlation coefficient (r) correlated analysis is conducted between observed chlorinity (2005/11/1–2006/2/28, 2006/11/1–2007/2/28, 2007/11/1–2008/2/28) with 0–31 days earlier observed influencing factors (historical chlorinity, tidal level, discharge and wind).

Table 2 | Time series of model input and output

Response variable	Station	Time series	Influencing factors	Stations	Time series
Chlorinity	Pinggang	2005/11/1–2006/2/28; 2006/11/1–2007/2/28; 2007/11/1–2008/2/28	Historical chlorinity	Pinggang	13/14/15 days earlier
			Tidal level	Denglongshan	9,10 days earlier
			Discharge	Makou	2 days earlier
			Wind	Zhongshan	2, 3 days earlier

(26×5), and 156 (26×6) trails for the four models, respectively. The variation of the NSE is presented in the Supplementary material (Figure S2) and the paired parameters ($mtry$, $ntree$) used for the RF models are (3, 500), (1, 1,700), (1, 1,600) and (2, 2,400), respectively.

ENN model training results

The number of the neural nodes is the major parameter to train ENN. A neural network with one hidden layer is sufficient to describe any nonlinear function with precision (Yao

2010). The training algorithm Traingdx was used during model implementations. Traingdx is a network training function that updates the weight and bias values based on the gradient descent momentum and adaptive learning rate. After several empirical tests, the following training parameters of traingdx were used: the maximum number of epochs to train = 5,000, learning rate = 0.5, performance goal = 0.06. The hyperbolic tangent transfer function (Tansig) and Purelin Transfer function (purelin) (Dorofki et al. 2012) were employed as neutral transfer functions of the hidden layer and output layer, respectively.

The number of the neural nodes in the input layer was the number of predictors in each model, which means 3 for Model one, 4 for Model two, 5 for Model three, 6 for Model four. The number of output neural nodes was fixed at 1. The number of neurons for the hidden layer could be changed to obtain the optimum prediction result. The variation in of the NSE is presented in the Supplementary material (Figure S3). The NSE fluctuated with a general decreasing trend as the neuron numbers increased from 5 to 20. Hence, we created networks with 3-19-1, 4-19-1, 5-20-1, and 6-19-1 structures for the four models, respectively.

BMA model training results

The MBA models are trained with the predictions given by the previously mentioned three models as input and the observed chlorinity as output. The BMA weights for each model members are trained using Box–Cox power transformation method (Box & Cox 1964; Osborne 2010) and expectation maximization (EM) algorithm. The Box–Cox power transformation method is used to transform the original chlorinity data into a normal distribution. The transformed chlorinity data are used to find the BMA weights for each model members. The EM algorithm is used to find the maximum likelihood. The BMA is applied to Model four and the trained weights are 0.5443, 0.3317, and 0.1240 for RF, SVM, and ENN, respectively.

Analysis of periodic characteristics and time-lags

The time-lag was determined based on MIC and r . The MIC and r charts (Figures 3 and 4) show that the chlorinity exhibited a stable 15-day period, which was in agreement with the findings in the study of Wang *et al.* (2012), who determined that the salinity at Pinggang station showed a clear semi-monthly variation with three peaks in 45 days in dry seasons. The 15-day period of the chlorinity can be attributed to the semi-monthly periodic rise and fall of the tide. The correlation between the chlorinity and the tidal level at Denglongshan exhibited quarter-monthly periodicity, which was the combined result of the semi-monthly periodicity of the chlorinity and tidal level, and the 3-day to 5-day delay between the tidal level and the chlorinity. The phase difference between the maximum salinity and the tidal range

represented an abnormal characteristic of the saltwater intrusion in the Modaomen waterway in many previous studies (e.g., Wang *et al.* 2012; Liu *et al.* 2014). There was a 2- to 3-day delay for the chlorinity at Pinggang to respond to the upstream discharge from Makou and the wind at Zhongshan.

To verify the feasibility and effectiveness of the MIC and r as correlation analysis methods for predictor selection, the target response chlorinity was compared with the phase-shifted time series of chlorinity by 15 days earlier, the tidal level by 9, 10 days earlier, the upstream discharge by 2 days earlier, and the wind by 2, 3 days earlier, respectively (Figure 5). The results show that the fluctuations of the response chlorinity and its phase-shifted influencing factors are in agreement with the results of the MIC and r analysis. After the 15-day phase shift, the waveforms of the historical chlorinity reached in phase with those of the target chlorinity, confirming the 15-day periodicity of chlorinity (Figure 5(a)). Shifted by 9 and 10 days earlier, the series of tidal level and target chlorinity matched in both peaks and valleys (Figure 5(b)), which was in agreement with the findings shown in Figures 3 and 4, i.e., the chlorinity reached its peak 9 or 10 days after the tidal maximum. The magnitude of the phase-shifted discharge reached a peak and the phase-shifted wind reached its minimum while the chlorinity wave reached its minimum (Figure 5(c) and 5(d)). This is in accordance with the salt tide mechanism that saltwater intrusion is inversely correlated to river discharge and is exacerbated by upstream wind.

These results indicate that the MIC and r are effective methods to measure the strength of the complex associations between chlorinity and its influencing factors.

Delayed predictor selecting method by correlation analysis

This variable selection technique balances the goodness-of-fit with simplicity and provides three main benefits when developing predictive models, i.e., improved model interpretability, shorter training times, and enhanced generalization by reducing overfitting (Tan *et al.* 2014). To verify the improvement ability of the MIC- and r -based predictor selection method, we used 0 to 31 days earlier values of each factor as model inputs for RF, SVM, and ENN. In each trial, we changed the delay of one of the factors and

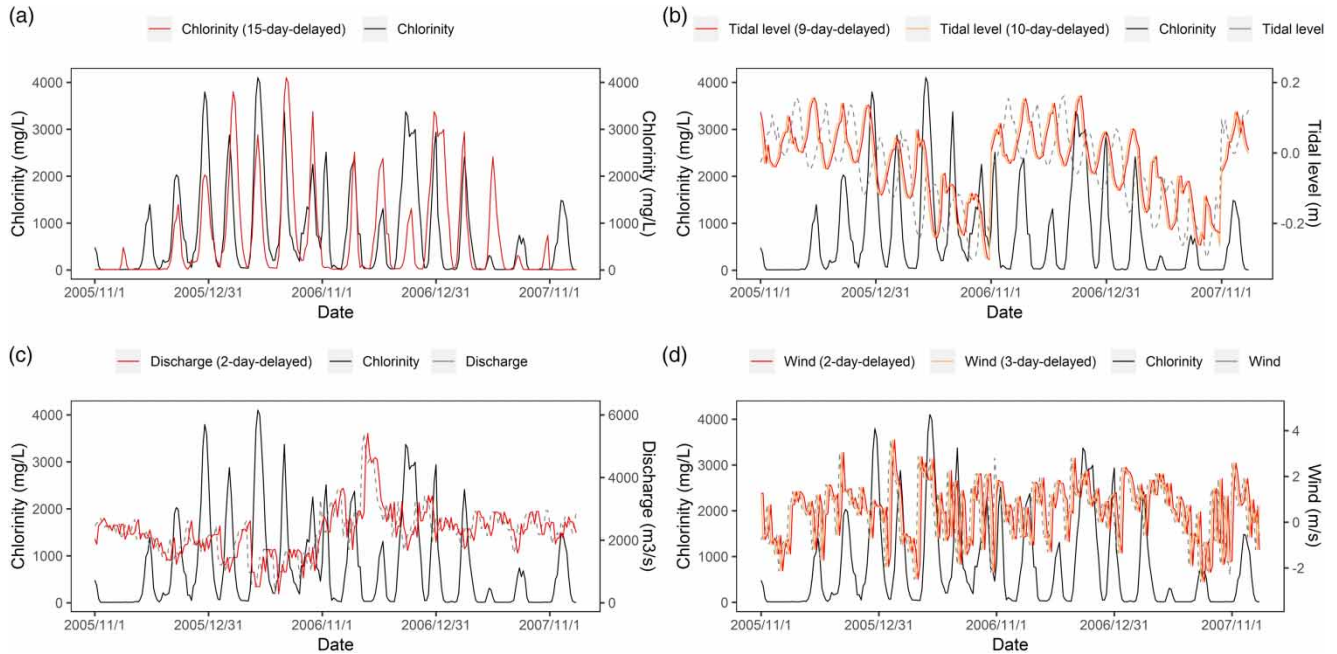


Figure 5 | Curves of: (a) response chlorinity and 15 days earlier chlorinity; (b) response chlorinity, tidal level, and 9, 10 days earlier tidal level; (c) response chlorinity, discharge, and 2 days earlier discharge; (d) response chlorinity, wind, and 2, 3 days earlier wind. *Note:* the black solid lines indicate the time series of chlorinity during the calibration period, the gray dashed lines indicate the influencing factors of the day, the solid lines indicate the phase-shifted time series of the influencing factors. The time series of chlorinity is shifted 15 days earlier, the time series of the tidal level is shifted 9, 10 days earlier, the time series of the upstream runoff is shifted 2 days earlier, and the time series of the wind is shifted 2, 3 days earlier.

fixed the others. The results are shown in Figure 6(a)–6(c). NSE reached its maximum and Pbias reached its minimum when 1–3 days earlier chlorinity, 14–16 days earlier chlorinity, 7–9 days earlier tidal level, 1–3 earlier discharge, and 1–3 earlier wind were used as model inputs, which is in accordance with the MIC and r analysis results.

These results indicate that the delayed predictors with the maximal MIC and r (i.e., 14, 15, 16 days earlier chlorinity, 9 and 10 days earlier tidal level, 2 days earlier discharge and 2, 3 days earlier wind) to the response chlorinity should be selected.

We compared the training and testing results using both MIC and r for delayed predictors' selection to build the models with those using only either one of them. Based on the MIC correlation only, 14, 15, 16 days earlier chlorinity, 9 days earlier tidal level, 2 days earlier discharge and wind were selected. Based on the r correlation only, 14, 15, 16 days earlier chlorinity, 10 days earlier tidal level, 2 days earlier discharge and 3 days earlier wind were selected. The results are shown in Table 3.

The models constructed by the predictors selected by both MIC and r outperformed the models constructed by

the predictors selected using only either one of them. Based on this result, we inferred that the 9-day delayed tidal level contains more nonlinear signals related to salt tide and 10-day delayed tidal level contains more linear signals; the 2-day delayed wind contains more nonlinear signals related to salt tide and 3-day delayed wind contains more linear signals; the delayed chlorinity and delayed discharge contain both nonlinear and linear signals.

The complex associations are confirmed by Figure 7 where chlorinity vs each (delayed) influencing factors is plotted. We inferred that linear and nonlinear signals are both contained by the influencing factors. Considering this complex association, we recommend that both MIC and r should be used when selecting the predictors rather than trusting only one of them.

Prediction of chlorinity using individual machine learning techniques

The predictions given by the four model series based on the RF, SVM, and ENN models are shown in Figure 8, Table 4, and Figure S4.

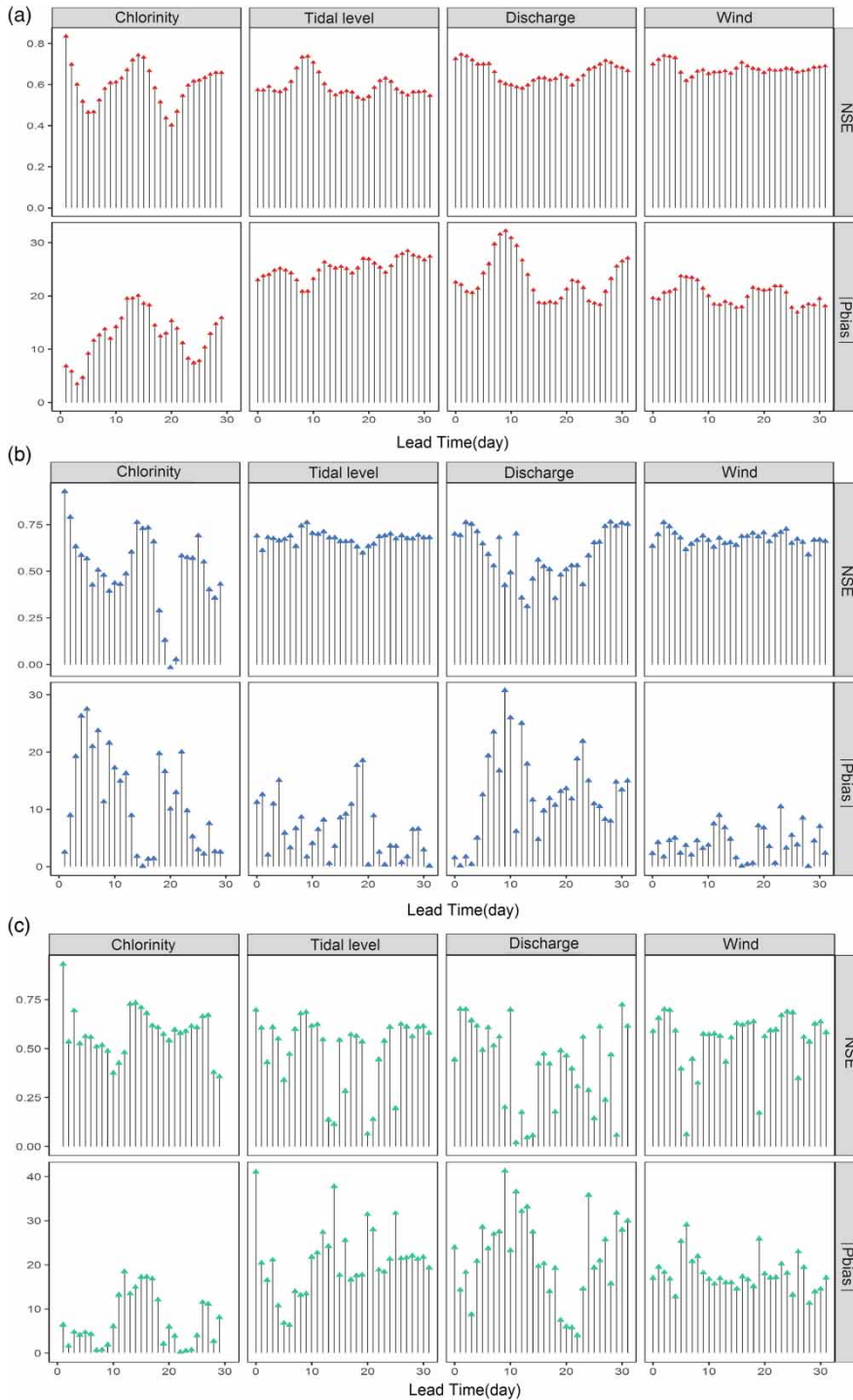


Figure 6 | Evaluation of estimation performance given by RF (a), SVM (b), and ENN (c) against chlorinity observation for testing period in terms of Nash-Sutcliffe efficiency coefficient (NSE) and absolute percent bias (Pbias) changing the delayed time of one of the predictors while fixing other predictors of model series 4.

The accuracy of the predictions, especially the predicted values of peaks and valleys, increased from Model one to Model four as more influencing factors

were taken into consideration (Figure 8). The high peaks were strongly underestimated and the valleys were overestimated by Model one (Figure 8(a)). When the tidal level,

Table 3 | Training and testing results using both MIC and r for predictor selecting and those using only one of them

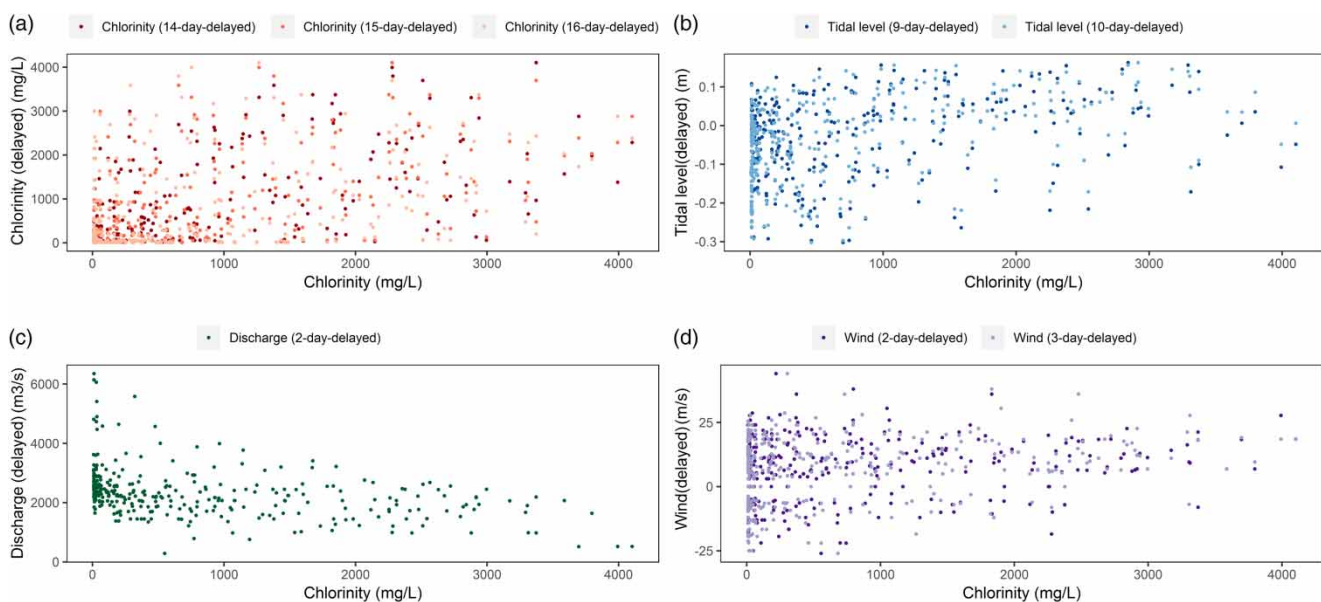
Predictor selecting methods	Models	Training		Testing	
		NSE	Pbias	NSE	Pbias
MIC + r	RF_4	0.91	0.8	0.76	19.8
	SVM_4	0.72	-15.8	0.74	-8.8
	ENN_4	0.74	-1.9	0.72	10.5
	BMA_4	0.86	-4.7	0.79	11.4
MIC	RF_4	0.91	0.7	0.71	20.1
	SVM_4	0.70	-16.6	0.73	-4.4
	ENN_4	0.74	2.8	0.68	14.4
	BMA_4	0.85	-4.7	0.76	12.8
r	RF_4	0.91	1.0	0.74	20.64
	SVM_4	0.70	-14.2	0.76	-2.3
	ENN_4	0.74	0.3	0.70	14.6
	BMA_4	0.85	-3.8	0.78	14.0

upstream discharge, and wind were taken into consideration, the accuracies of the high peaks and valleys improved. The wind was an indispensable predictor for chlorinity prediction. When the wind was taken into consideration, the NSE values of the RF, SVM, and ENN in Model four increased by 9%, 12%, and 4%, respectively, compared to the values in Model three. Generally, the errors decreased as more influencing factors were taken

into consideration and Model four achieved the highest accuracy.

The data set chosen in this study is a time series of chlorinity during dry seasons and it includes some abnormally high peaks and a small number of valleys. Figure 9 shows the scatter plots for the testing data set in Model four. For the high peaks, SVM and ENN yielded the best performance and provided similar predictions in the testing period. The errors at valleys of SVM were significantly lower than those of the RF and ENN. The RF resulted in the worst performance for predicting the valleys. Generally, the RF predictions fall above the optimal chlorinity curve on the lower limb of the curve and below the optimal chlorinity curve on the upper limb of the curve. This means that the RF predictions were often larger than the observed values when the observed values were relatively low; the predictions were often lower than the observed values when the observed values were relatively high.

We assumed that the large errors at the peaks and valleys can be explained by the imbalanced chlorinity data set, in which there were no sufficient peak and valley samples to train the models and the highly data-dependent characteristic (a large number of samples is required) of machine-learning algorithms; a similar result was also reported in Yoon *et al.* (2017). We inferred that SVM was

**Figure 7** | Chlorinity vs each (delayed) influencing factors: (a) chlorinity vs 14, 15, and 16 days earlier chlorinity; (b) chlorinity vs 9 and 10 days earlier tidal level; (c) chlorinity vs 2 days earlier discharge; (d) chlorinity vs 2 and 3 days earlier wind.

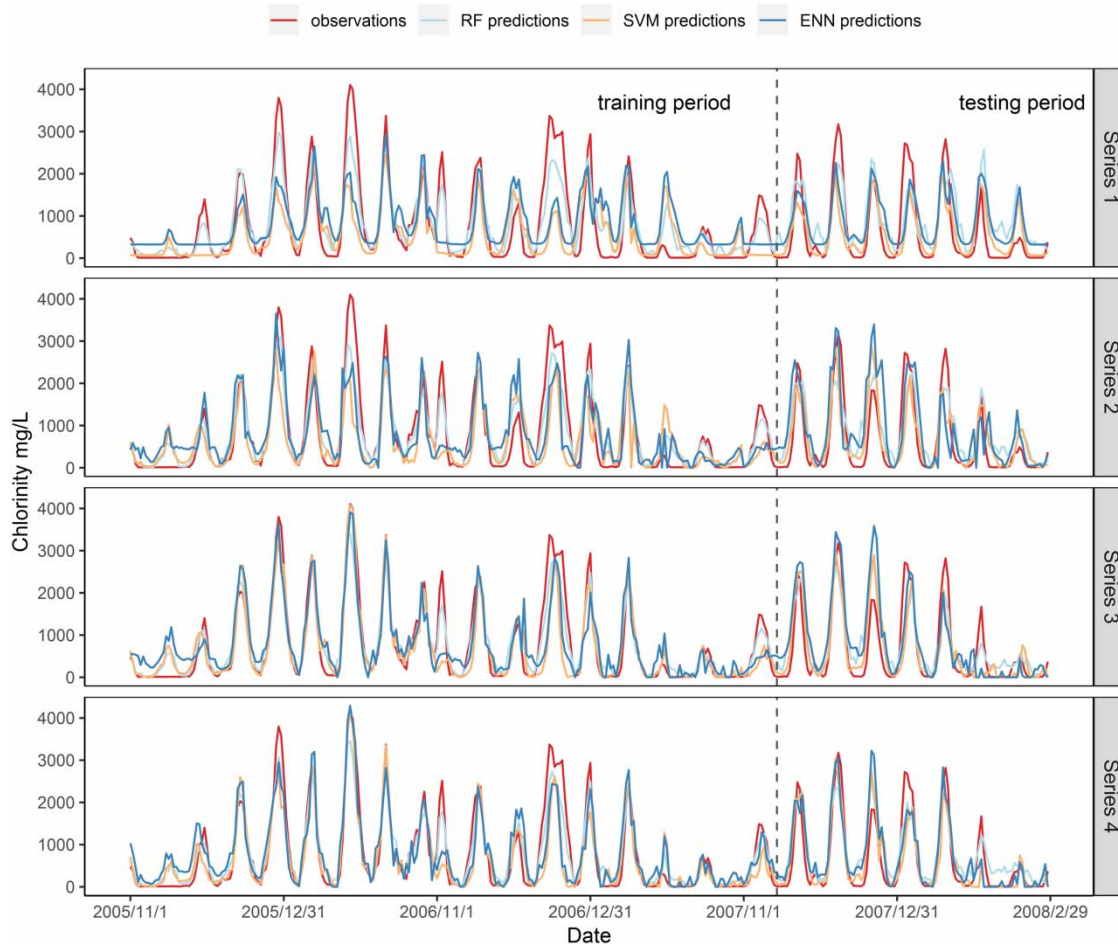


Figure 8 | Curves of observed chlorinity and simulations, predictions given by random forest (RF), support vector machine (SVM), and Elman neural network (ENN) during the training period and testing period for series one, series two, series three and series four. Note: the gray dashed line divides the time series into training period and testing period.

able to compensate for the highly imbalanced chlorinity data set to a certain extent for its property of being able to select a small and suitable subset of support vectors to build the models (Behzad *et al.* 2009). All the training data were used for the ENN and RF to simulate the complex associations between the input and output. Since there was a lack of sufficient peak and valley samples, the algorithm was better trained for the chlorinity between the valleys and peaks. As a result, the peaks were underestimated and the valleys were overestimated. The RF exhibited the worst performance for dealing with imbalanced data in this study, which was also the case in the study of Dudoit & Fridly (2003), who found that the RF performed poorly for an imbalanced data set. The findings indicate that RF model systematically over-predicts low chlorinity values and

under-predicts high chlorinity values. We believe that this explains why the RF achieved higher NSE values than the ENN although the ENN has the highest Pbias among the three models. The NSE value was not sensitive to the quantification of systematic errors (Krause *et al.* 2005).

The performance of the ENN was not as stable as those of the RF and SVM for the unexpected peaks and the oscillation during the low chlorinity period; this may be attributed to the lack of a global search ability and falling into a local minimum (Zhou *et al.* 2013). This finding was consistent with the results of Behzad *et al.* (2010), who reported that the SVM offered a better alternative to the artificial neural network (ANN) with its global optimization algorithm.

In terms of the performance criteria scores for the testing set, ENN performed the worst among the three. ENN

Table 4 | Performance criteria scores of the RF, SVM, ENN, and BMA for calibration and evaluation in the chlorinity prediction models

	Training		Testing	
	NSE	Pbias	NSE	Pbias
RF_1	0.83	1.2	0.48	28.9
SVM_1	0.27	-49.5	0.55	-12.4
ENN_1	0.37	0.2	0.52	-5.8
RF_2	0.85	1.1	0.65	22.9
SVM_2	0.47	-22.5	0.65	4.3
ENN_2	0.55	0.01	0.48	-15.6
RF_3	0.90	0.70	0.70	16.6
SVM_3	0.75	-13.1	0.66	-2.9
ENN_3	0.70	0.70	0.69	8.3
RF_4	0.91	0.81	0.76	19.8
SVM_4	0.72	-15.8	0.74	-8.8
ENN_4	0.74	-1.9	0.72	10.5
BMA_4	0.86	-4.7	0.79	11.4

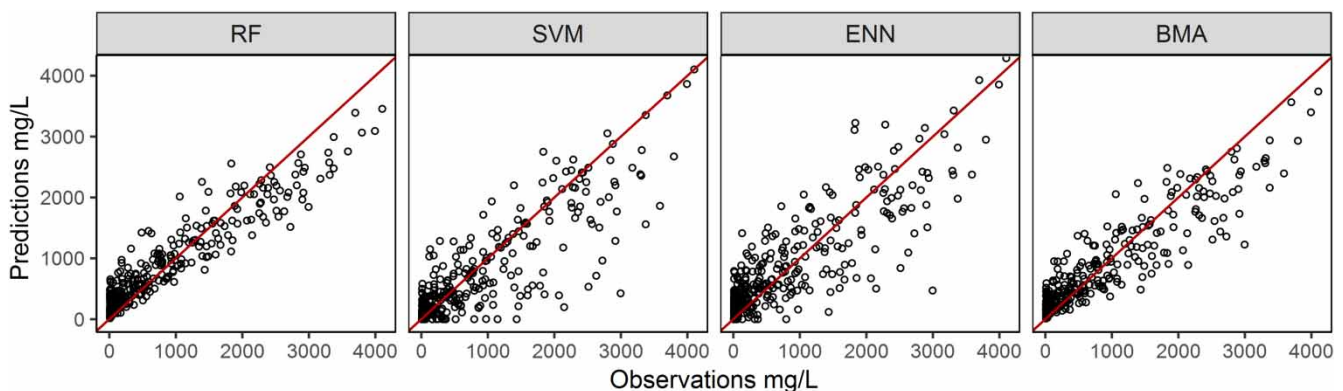
achieved much better NSE and Pbias than SVM in the training period but performed the worst among the three models in the testing period. This may be attributed to the difference in model complexity between the three models. RF and SVM only have two parameters while ENN is much more complex, which may have led to overfitting and poorer performance in the testing period compared to RF and SVM. SVM exhibited the highest tendency to be resistant to over-fitting for its regulated risk function. The regulated risk function refers to the structural risk minimization principle that the SVMs is based on, which minimizes a bound

on a generalized risk (error), as opposed to the empirical risk minimization principle exploited by conventional regression techniques (e.g., ANNs) (Behzad *et al.* 2009). RF over-fitted in this study may be attributed to noise in the chlorinity data set, which is also the case in the study of Segal (2004).

A multi-model ensemble strategy was needed to exploit the diversity of the three models. The lack of physical laws in empirical model approaches and the fact that these models have many non-defined parameters that require optimization make them vulnerable to errors (De Vos & Rientjes 2005). Since these three algorithms were grounded in different frameworks of the statistical learning theory, their performance exhibited different characteristics and their errors showed different distributions. Hence, the BMA ensemble approach was necessary for the three models.

Application of BMA method for a model ensemble approach

The BMA resulted in better prediction performance than its ensemble members. The BMA approach was applied to the final model, i.e., the regression models with historical chlorinity, tidal level, discharge, and wind as predictors; the results are shown in Table 4. The BMA performed better than any of the other three models in terms of the highest NSE and achieved similar Pbias with SVM and ENN. The BMA resulted in the maximum NSE (0.79). The BMA had the best goodness-of-fit among the ensemble members. Figure 9 also shows that the predicted chlorinity values

**Figure 9** | Scatter plot of chlorinity predicted in series 4 by random forest (RF), support vector machine (SVM), Elman neural network (ENN) and Bayesian model averaging method (integrated forecast based on RF, SVM, and ENN) vs chlorinity observed.

are less dispersed for the BMA than the SVM and ENN, and contain less systematic errors than those of the RF. Thus, the BMA provided the most accurate and stable prediction of chlorinity among the three models. The outputs of RF, SVM, and ENN exhibited the different extent of dispersion and systematic errors.

The model inferences were made without a priori knowledge of which model would better predict salinity; and thus, we recommend using BMA to give integrated forecasts instead of relying upon only one specific model.

Instead of weighting the ensemble members based on only one specific criterion in the training period, the BMA method weights its ensemble members based on how much information each member can provide and how accurate the information is. The weight rank order of the member models is not completely the same as that of NSE or the inverse order of the Pbias in the training period. The BMA weights of the RF and SVM are both higher than 0.3, indicating that both the RF and SVM are necessary for providing forecast information. The ENN had a similar NSE and Pbias with the SVM in the training period but a much lower BMA weight, at 0.1240. This suggested that once the RF and SVM are known, there is little additional information provided by the ENN. One can understand why this occurred by looking again at the prediction given by the three models in Figures 8 and 9; the ENN and SVM gave a similar prediction for predicting the high peak values and the scatter plot of their prediction versus observations showed a similar extent of dispersion. On the other hand, although the SVM had worse NSE and Pbias value than the ENN in the training period, the SVM contributed more additional information because it was less correlated with the others.

CONCLUSION

In this study, we developed an integrated temporally sensitive empirical model of chlorinity and tested it on the Pearl River Delta. The approach that we proposed was tailored to the salt tide mechanisms and complex delta systems. Our approach used MIC and r correlations to explore the inner correlation and to adjust the lag time between measurement of influencing factors of dilution and mixing (upstream discharge, tide, wind) and their effect on chlorinity.

The following conclusions can be drawn:

1. The MIC and r are effective tools for analyzing the correlation and time lag between chlorinity and its influencing factors. The chlorinity at Pinggang station and the tidal level at Denglongshan station both have a stable 15-day (semi-monthly) periodicity. There exists a 9, 10-day time delay for chlorinity at Pinggang station to respond to the fluctuation of the tidal level at Denglongshan station and a 2, 3-day time delay to respond to both the variation in the upstream discharge at Makou station and the wind disturbance at Zhongshan station.
2. The BMA model had better prediction accuracy than its ensemble members (i.e., RF, ENN, and SVM) and the BMA weights reflected the model performance and the correlation of the predictions given by the ensemble models.

Furthermore, our approach provides new insights into the prediction of chlorinity of complex delta systems. First, by selecting predictors with delayed influences based on correlation analysis prior to model building improved model interpretability, shortened training times, and enhanced generalization. Second, empirical models based on machine learning algorithms are suitable for modeling complex delta systems where chlorinity and hydrological data and high-resolution estuarine topography data are limited. Our approach is also practical for identifying factors affecting freshwater as a source for drinking water and predicting in advance when and where sources for drinking water will become too saline. In particular, this application benefitted from the temporal scale of this model. Although the current study predicts a very broad range of chlorinity, it is possible that models focused on the chlorinity range for sources of potable water could be even more useful. We will consider using two or three models tailored to different pre-specified chlorinity ranges and multiple BMA weight sets to be examined for different chlorinity ranges in our future research.

In summary, the BMA ensemble approach enabled us to develop a strong reliable model to predict salt water intrusion because it was able to leverage information contained in the three diverse models. It is likely that other data-driven modeling efforts would also benefit by using the BMA approach for model development.

ACKNOWLEDGMENTS

This research was supported by the National Key R&D Program of China (2017YFC0405900), the National Natural Science Foundation of China (51822908, 51779279), the open research foundation of Dynamics and the Associated Process Control Key Laboratory in the Pearl River estuary, the Ministry of Water Resources (2017KJ12), Baiqianwan project's young talents plan of special support program in Guangdong Province (42150001), and the Fundamental Research Funds for the Central Universities (20187614031620001). All are gratefully acknowledged.

SUPPLEMENTARY DATA

The Supplementary Data for this paper is available online at <http://dx.doi.org/10.2166/hydro.2019.073>.

REFERENCES

- Becker, M. L., Luettich Jr, R. A. & Mallin, M. A. 2010 Hydrodynamic behavior of the Cape Fear River and estuarine system: a synthesis and observational investigation of discharge salinity intrusion relationships. *Estuarine Coastal and Shelf Science* **88**, 407–418.
- Behzad, M., Asghari, K., Eazi, M. & Palhang, M. 2009 Generalization performance of support vector machines and neural networks in runoff modeling. *Expert Systems with Applications* **36** (4), 7624–7629.
- Behzad, M., Asghari, K. & Coppola, E. A. 2010 Comparative study of SVMs and ANNs in aquifer water level prediction. *Computing in Civil Engineering* **24**, 408–413.
- Box, G. E. P. & Cox, D. R. 1964 An analysis of transformations. *Journal of the Royal Statistical Society* **26** (2), 211–252.
- De Vos, N. J. & Rientjes, T. H. M. 2005 Constraints of artificial neural networks for rainfall-runoff modelling: trade-offs in hydrological state representation and model evaluation. *Hydrology and Earth System Sciences* **9**, 111–126.
- Dong, X., Tao, T., Liu, S., Liu, S., Xia, Y. & Yu, Y. 2010 Prediction of salt water intrusion using BP-RAGA coupled neural network model. In: *3rd International Congress on Image and Signal Processing IEEE*, Yantai, China, pp. 4244–4248.
- Dorofki, M., Elshafie, A. H., Jaafar, O., Karim, O. A. & Mastura, S. 2012 Comparison of Artificial Neural Network transfer functions abilities to simulate extreme runoff data. In: *International Conference on Environment, Energy and Biotechnology*. IACSIT Press, Singapore 33, 39–44.
- Dudoit, S. & Fridly, J. 2003 Introduction to classification in microarray experiments. *A Practical Approach to Microarray Data Analysis* (D. P. Berrar, W. Dubitzky & M. Granzow, eds). Springer, Boston, MA, Chapter 7, 132–149.
- Gingerich, S. B. & Voss, C. I. 2005 Three-dimensional variable-density flow simulation of a coastal aquifer in southern Oahu, Hawaii, USA. *Hydrogeology Journal* **13** (2), 436–450.
- Hauduc, H., Neumann, M. B., Muschalla, D., Gamerith, V., Gillot, S. & Vanrolleghem, P. A. 2015 Efficiency criteria for environmental model quality assessment: a review and its application to wastewater treatment. *Environmental Modelling & Software* **68** (C), 196–204.
- Howes, N. C., FitzGerald, D. M., Hughes, Z. J., Georgiou, I. Y., Kulp, M. A., Miner, M. D., Smith, J. M. & Barras, J. A. 2010 Hurricane-induced failure of low salinity wetlands. *Proceedings of the National Academy of Science of the USA* **107**, 14014. <https://doi.org/10.1073/pnas.0914582107>.
- Klemeš, V. 1986 Operational testing of hydrological simulation models. *International Association of Scientific Hydrology Bulletin* **31** (1), 13–24.
- Krause, P., Boyle, D. P. & Se, F. B. 2005 Comparison of different efficiency criteria for hydrological model assessment. *Advances in Geosciences* **5** (5), 89–97.
- 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 Resources Management* **32**, 2405–2419.
- Leonard, W. H. 1976 The effect of the spring neap tidal cycle on the vertical salinity structure of the James, York and Rappahannock Rivers, Virginia, USA. *Estuarine & Coastal Marine Science* **5**, 485–496.
- Lin, K., Liu, P., He, Y. & Guo, S. 2014 Multi-site evaluation to reduce parameter uncertainty in a conceptual hydrological modeling within the GLUE framework. *Journal of Hydroinformatics* **16** (1), 60–73.
- Lin, K. R., Lin, Y. Q., Xu, Y. M., Chen, X. H., Chen, L. & Singh, V. P. 2017 Inter- and intra- annual environmental flow alteration and its implication in the Pearl River Delta, South China. *Journal of Hydro-Environment Research* **15**, 27–40.
- Liu, Z. & Merwade, V. 2018 Accounting for model structure, parameter and input forcing uncertainty in flood inundation modeling using Bayesian model averaging. *Journal of Hydrology* **565**, 138–149.
- Liu, B. J., Yan, S. L., Chen, X. H., Lian, Y. Q. & Xin, Y. B. 2014 Wavelet analysis of the dynamic characteristics of saltwater intrusion – a case study in the pearl river estuary of China. *Ocean & Coastal Management* **95** (4), 81–92.
- Mao, Q. W., Shi, P., Yin, K. D., Gan, J. & Qi, Y. 2004 Tides and tidal currents in the Pearl River Estuary. *Continental Shelf Research* **24**, 1797–1808.

- Nash, J. E. & Sutcliffe, J. V. 1970 River flow forecasting through the conceptual models, 1: a discussion of principles. *Journal of Hydrology* **10** (3), 282–290.
- Neto, A. A. M., Oliveira, P. T. S., Rodrigues, D. B. B. & Wendland, E. 2018 Improving streamflow prediction using uncertainty analysis and Bayesian model averaging. *Journal of Hydrologic Engineering* **23** (5), 05018004.
- Nicholls, R. J. & Cazenave, A. 2010 Sea-level rise and its impact on coastal zones. *Science* **328** (5985), 1517–1520.
- Osborne, J. W. 2010 Improving your data transformations: applying the Box-Cox transformation. *Practical Assessment Research and Evaluation* **15** (12), 1–9.
- Passeri, D. L., Hagen, S. C., Medeiros, S. C., Bilskie, M. V., Alizad, K. & Wang, D. 2015 The dynamic effects of sea level rise on low-gradient coastal landscapes: a review. *Earth's Future* **3** (6), 159–181.
- Raftery, A. E., Gneiting, T., Balabdaoui, F. & Polakowski, M. 2016 Using Bayesian model averaging to calibrate forecast ensembles. *Monthly Weather Review* **133** (5), 1155–1174.
- Ralston, D. K., Geyer, W. R. & Lerczak, J. A. 2008 Subtidal salinity and velocity in the Hudson River estuary: observations and modeling. *Journal of Physical Oceanography* **28**, 753–770.
- Reshef, D. N., Reshef, Y. A., Finucane, H. K., Grossman, S. R., Mcvean, G., Turnbaugh, P. J., Lander, E. S., Mitzenmacher, M. & Sabeti, P. 2011 Detecting novel associations in large datasets. *Science* **334** (6062), 1518.
- Rohmer, J. & Brisset, N. 2017 Short-term forecasting of saltwater occurrence at La Comté River (French Guiana) using a kernel-based support vector machine. *Environmental Earth Sciences* **76** (6), 246.
- Segal, M. R. 2004 *Machine Learning Benchmarks and Random Forest Regression*. Center for Bioinformatics & Molecular Biostatistics, University of California, San Francisco, CA, USA.
- Tan, Q., Jiang, H. & Ding, Y. 2014 Model selection method based on maximal information coefficient of residuals. *Acta Mathematica Scientia (English Series)* **34** (2), 579–592.
- Tessler, Z. D., Vorosmarty, C. J., Grossberg, M., Gladkova, I., Aizenman, H., Syvitski, J. P. M. & Fofoula-Georgiou, E. 2015 Profiling risk and sustainability in coastal deltas of the world. *Science* **349** (6248), 638–643.
- Uncles, R. J. & Stephens, J. A. 2011 The effects of wind, runoff and tides on salinity in a strongly tidal sub-estuary. *Estuaries and Coasts* **34**, 758–774.
- Vaze, J., Post, D. A., Chiew, F. H. S., Perraud, J., Viney, N. R. & Teng, J. 2010 Climate nonstationarity-Validity of calibrated rainfall-runoff models for use in climate change studies. *Journal of Hydrology* **394**, 447–457.
- Viney, N. R., Perraud, J., Vaze, J., Chiew, F. H. S., Post, D. A. & Yang, A. 2009 The usefulness of bias constraints in model calibration for regionalization to ungauged catchments. In: *Proceedings of the 18th World IMACS/MODSIM Congress*. International Environmental Modelling and Software Society, Cairns, Australia, pp. 3421–3427.
- Wang, B., Zhu, J. R., Wu, H., Yu, F. J. & Song, X. J. 2012 Dynamics of saltwater intrusion in the Modaomen Waterway of the Pearl River Estuary. *Science China Earth Sciences* **55** (11), 1901–1918.
- Wu, S., Cheng, H., Xu, Y. J., Li, J. & Zheng, S. 2016 Decadal changes in bathymetry of the Yangtze River estuary: human impacts and potential saltwater intrusion. *Estuarine Coastal & Shelf Science* **182**, 158–169.
- Yang, X., Magnusson, J. & Xu, C. Y. 2018 Transferability of regionalization methods under changing climate. *Journal of Hydrology* **568**, 67–81.
- Yao, X. 2010 A review of evolutionary artificial neural networks. *International Journal of Intelligent Systems* **8** (4), 539–567.
- Yoon, H., Kim, Y., Ha, K., Lee, S. H. & Kim, G. P. 2017 Comparative evaluation of ANN- and SVM-time series models for predicting freshwater-saltwater interface fluctuations. *Water* **9** (5), 323.
- Yu, X., Yang, J., Graf, T., Koneshloo, M., O'Neal, M. A. & Michael, H. A. 2016 Impact of topography on groundwater salinization due to ocean surge inundation. *Water Resources Research* **52** (8), 5794–5812.
- Zhang, Z., Cui, B., Zhao, H., Fan, X. & Zhang, H. 2010 Discharge-salinity relationships in Modaomen waterway, Pearl River estuary. *Procedia Environmental Sciences* **2** (1), 1235–1245.
- Zhang, J., Liu, P., Wang, H., Lei, X. & Zhou, Y. 2015 A Bayesian model averaging method for the derivation of reservoir operating rules. *Journal of Hydrology* **528**, 276–285.
- Zhao, X., Deng, W. & Shi, Y. 2013 Feature selection with attributes clustering by maximal information coefficient. *Procedia Computer Science* **17** (2), 70–79.
- Zhou, C., Ding, L. Y. & He, R. 2013 PSO-based Elman neural network model for predictive control of air chamber pressure in slurry shield tunneling under Yangtze River. *Automation in Construction* **36** (15), 208–212.

First received 4 April 2019; accepted in revised form 19 August 2019. Available online 27 September 2019