

## Estuary salinity prediction using a coupled GA-SVM model: a case study of the Min River Estuary, China

Yihui Fang, Xingwei Chen and Nian-Sheng Cheng

### ABSTRACT

Estuary salinity predictions can help to improve water safety in coastal areas. Coupled genetic algorithm-support vector machine (GA-SVM) models, which adopt a GA to optimize the SVM parameters, have been successfully applied in some research fields. In light of previous research findings, an application of a GA-SVM model for tidal estuary salinity prediction is proposed in this paper. The corresponding model is developed to predict the salinity of the Min River Estuary (MRE). By conducting an analysis of the time series of daily salinity and the results of simulation experiments, the high-tide level, runoff and previous salinity are considered as the major factors that influence salinity variation. The prediction accuracy of the GA-SVM model is satisfactory, with coefficient of determination ( $R^2$ ) of 0.85, Nash–Sutcliffe efficiency of 0.84 and root mean square error of 119 ( $\mu\text{S}/\text{cm}$ ). The proposed model performs significantly better than the traditional SVM model in terms of prediction accuracy and computing time. It can be concluded that the proposed model can successfully predict the salinity of MRE based on the high-tide level, runoff and previous salinity.

**Key words** | coupled model, genetic algorithm, Min River Estuary, salinity prediction, support vector machine

**Yihui Fang**  
**Xingwei Chen** (corresponding author)  
 School of Geographic Science,  
 Fujian Normal University,  
 Fuzhou,  
 Fujian 350007,  
 China  
 E-mail: [cxwchen215@163.com](mailto:cxwchen215@163.com)

**Yihui Fang**  
 Information Management & Engineer Department,  
 Fujian Commerce College,  
 Fuzhou 350012,  
 China

**Xingwei Chen**  
 Cultivation Base of State Key Laboratory Humid  
 Subtropical Mountain Ecology,  
 Fuzhou 350007,  
 China

**Nian-Sheng Cheng**  
 School of Civil and Environmental Engineering,  
 Nanyang Technological University,  
 Nanyang Ave.,  
 639798, Singapore

### LIST OF ACRONYMS

ANN	artificial neural network	LIBSVM	the open source software for SVM, developed by National Taiwan University
b	scalar threshold	LK	linear kernel
C	the penalty factor, the penalty degree of the sample with error exceeding $\epsilon$	$L_t$	the t-day high-tide level
$\overline{Cf}$	average values of the $Cf(i)$ data series	$K(x_i, x_j)$	kernel function
$Cf(i)$	predicted value	MRE	Min River Estuary
$\overline{Co}$	average values of the $Co(i)$ data series	$n$	size of the data series
$Co(i)$	observed value	OOP	object-oriented programming
$C_t$	the t-day salinity	PK	polynomial kernel
ENS	Nash–Sutcliffe efficiency coefficient	$Q_t$	the t-day runoff
$f(x)$	target function	$R^2$	coefficient of determination
GA	genetic algorithm	RBF	radial basis function
GA-SVM	a coupled model of genetic algorithm and support vector machine	RMSE	root mean square error
$l$	number of input vectors	SVM	support vector machine
		$w$	weight vector

$x$	input vector
$\alpha_i, \alpha_i^*$	Lagrangian multipliers
$\varepsilon$	an error tolerance
$\xi_i, \xi_i^*$	slack variables that specify the upper and the lower training errors subject to $\varepsilon$

## INTRODUCTION

Saltwater intrusion is a common natural phenomenon in tidal estuaries which exerts great impact on the estuary's ecological environment (Soetaert *et al.* 1995; Thomas & Tris 1996) and on urban drinking water safety (Nowroozi *et al.* 1999; Wen *et al.* 2007). Accurate predictions of estuary salinity can alleviate the adverse impacts caused by saltwater intrusion. However, accurate predictions are difficult to ascertain because estuary salinity is directly or indirectly affected by a variety of temporally and spatially variable factors, including runoff, tides, channel topography and human activities (Savenije 1993; Nguyen & Savenije 2006; Wen *et al.* 2007).

There are two approaches mainly used to predict estuary salinity. The first approach develops models based on dynamic salinity variation processes (Reddy & Ghosh 1993; Chevalier *et al.* 2014). The second applies statistical or data mining methods to establish relationships between the salinity and the influencing factors. The first approach can be considered as a 'white-box' method based on a complex process analysis, which requires large amounts of high quality data. In comparison, the second is a 'black-box' method, which requires fewer data and thus is more applicable in practice. Data mining methods based on artificial neural networks (ANN) have been widely used for salinity predictions. For example, Huang & Foo (2002) applied an ANN method to assess salinity variation responding to the multiple forcing functions of freshwater, tide, and wind in Apalachicola River, Florida. However, ANN methods may incorporate various issues, such as over-fitting, a weak generalization ability and lack of suitability for large datasets (Guan *et al.* 2013; Li & Kong 2014). The support vector machine (SVM) attempts to achieve a compromise between complexity (the learning accuracies of certain training samples) and learning capacity (the prediction ability for

samples) according to a limited sample dataset, while obtaining an optimal generalization ability (Vapnik 1998). The SVM can effectively resolve nonlinear problems in small and high dimension samples. Yu *et al.* (2006) adopted the SVM to establish a real-time stage forecasting model. Guan *et al.* (2013) used an SVM model to predict soil electrical conductivity values in an irrigation district. However, SVM models have rarely been used to predict estuary salinity. Liu & Chen (2007) used a coupling model based on the partial least squares and SVM methods to predict saltwater intrusion in the Pearl River Estuary, considering runoff and downstream salinity as the factors that influence the salinity variation.

However, the prediction accuracy of the SVM greatly depends on the right selection of parameters (the penalty factor  $C$  and kernel parameter  $\delta$ ) (Cherkassky & Ma 2004). Generally, the parameters are selected via cross-validation and grid searching, which greatly limit the prediction accuracy and generalization ability. The genetic algorithm (GA) simulates biological evolution processes and provides the optimal solution via global searching. The GA is stochastic in nature and therefore capable of running away from the local optimal. Researchers have used GA to optimize SVM parameters and developed the coupled GA-SVM model, which has been successfully applied in some research fields. Liu & Lu (2014) reported that the GA-SVM model can be used to predict agricultural non-point-source pollution with results better than the ANN method. Li & Kong (2014) successfully applied the GA-SVM model to analyze landslides. Liu *et al.* (2015) adopted the GA-SVM model to discriminate different transgenic cotton seeds with similar characteristics based on terahertz spectroscopy. But it is observed that there are hardly any applications of the GA-SVM model for tidal estuary salinity prediction, so far.

This paper aims to apply the GA-SVM model to predict the salinity of a tidal estuary based on previous research findings (Harish *et al.* 2014; Li & Kong 2014). An application of the GA-SVM model for tidal estuary salinity prediction was proposed and the corresponding model was developed. The Min River Estuary (MRE) was selected as the study area, which is located in the coastal region of southeastern China. The major factors influencing salinity variations were determined by time-series analysis of daily salinity

and simulation experiments. Finally, to evaluate its prediction performance, the proposed model was compared with a traditional SVM model.

## METHODS

### SVM

The principal idea of SVM is to represent the entire sample set with a small number of support vectors (Vapnik 1998). SVM models can be described by the following function:

$$f(x) = w \cdot x + b \quad (1)$$

Considering the existence of some permissible error, Equation (1) can be incorporated in a convex optimization problem as follows:

$$\text{Minimize: } \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l (\xi_i + \xi_i^*) \quad (2)$$

$$\text{Subject to: } \begin{cases} y_i - (w \cdot x_i + b) \leq \varepsilon + \xi_i \\ (w \cdot x_i + b) - y_i \leq \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \geq 0, i = 1, 2, \dots, l \end{cases}$$

where  $\xi_i$  and  $\xi_i^*$  are slack variables that specify the upper and the lower training errors subject to an error tolerance  $\varepsilon$ , and the constant  $C (>0)$  stands for the penalty degree of the sample with error exceeding  $\varepsilon$  and is called the penalty factor. In the feature space, the inner product operations of the linear problem can be substituted for kernel functions. Thus, the dual form of the SVM can be expressed as:

$$\text{Minimize: } \frac{1}{2} \sum_{i,j=1}^l (\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j)K(x_i, x_j) + \varepsilon \sum_{i=1}^l (\alpha_i^* + \alpha_i) - \sum_{i=1}^l y_i(\alpha_i^* - \alpha_i) \quad (3)$$

$$\text{Subject to: } \begin{cases} \sum_{i=1}^l (\alpha_i - \alpha_i^*) = 0 \\ \alpha_i, \alpha_i^* \geq 0, i = 1, 2, \dots, l \end{cases}$$

where  $\alpha_i$  and  $\alpha_i^*$  are the Lagrangian multipliers and  $K(x_i, x_j)$  is a kernel function.

Three primary kernel functions are widely used, including the linear kernel, polynomial kernel and radial basis function (RBF). This paper uses the RBF due to its strong nonlinear mapping ability. The form of the RBF can be expressed as:

$$K(x_i, x_j) = \exp(-\delta \cdot |x_i - x_j|^2) \quad (4)$$

Here,  $\delta$  is the kernel parameter that represents the spatial extent that a particular training sample can reach.

### GA

The GA was designed to simulate genetic evolution mechanisms and random information exchange. The algorithm is guided by the fitness function, which is constructed based on individual specific problems (Whitley 1994). Starting from any initial population, step by step, new better-adapted chromosomes can be generated by selection, copying, crossover and mutation operations and thus a best-adapted chromosome can be acquired finally. Due to its global optimality, implicit parallelism, high stability and wide usability (Li & Kong 2014), the method is used to optimize the SVM model parameters ( $C$  and  $\delta$ ) in this paper.

### GA-SVM model

The performance of the SVM greatly depends on the right selection of parameters, which greatly impact on the efficiency and generalization performance of the SVM model. The GA is a prominent choice for optimizing the parameters of the SVM model according to the results of various studies (Harish et al. 2014; Li & Kong 2014). It can reduce the blindness of human-made choice and thus improve the performance of the model. In this paper, the GA-SVM algorithm was implemented using the C# programming language. The flow chart of the GA-SVM method is illustrated in Figure 1, with the following steps implemented:

- (1) Use the min-max normal to format the dataset and construct vectors.

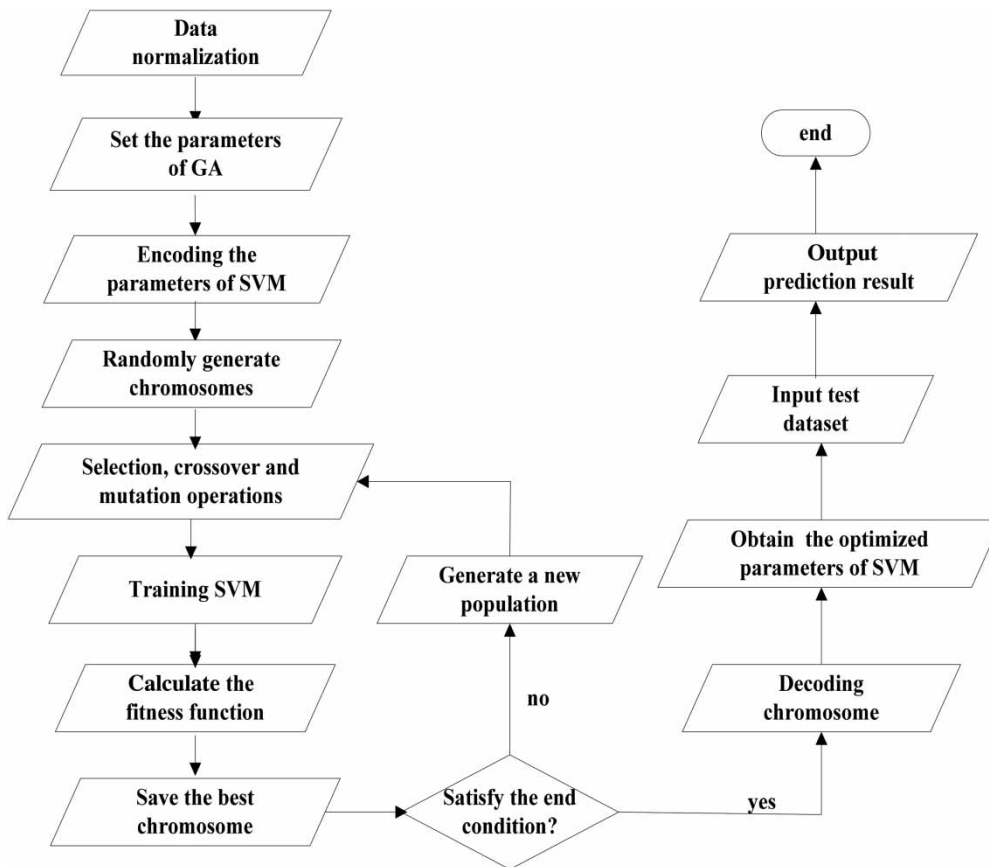


Figure 1 | Flow chart of the GA-SVM algorithm.

- (2) Initialize the GA parameters (the population size, maximum evolution number, population crossover rate and mutation rate) and the value range of the SVM parameters ( $C$  and  $\delta$ ). Then, choose the coefficient of determination ( $R^2$ ) as the GA fitness function.
- (3) Randomly generate a set of SVM parameter value chromosomes with binary coding. A single chromosome is constructed via the binary-string concatenation of  $C$  and  $\delta$ .
- (4) Generate new better-adapted chromosomes using selection, copying, crossover and mutation operations.
- (5) Train the SVM model, and calculate the fitness function value of each individual population, and save the best chromosome.
- (6) Determine if the end conditions are satisfied (the loop number is greater than the maximum evolution number). If it is true, output the optimal individual chromosome and go to step (7). Otherwise, generate a new population and proceed to step (4).

- (7) Decode the optimal chromosome to obtain the optimal SVM parameters  $C$  and  $\delta$ .

The program of the GA-SVM model for salinity prediction was implemented based on the C#.NET platform and the open source software LIBSVM toolkit, which was developed by National Taiwan University (CSIE 2015). The code was organized based on classes using object-oriented programming technology. In order to improve the running speed of the GA-SVM model, multithreading technology was adopted to instantiate threads, which were saved in a multithread pool.

### Model evaluation

The prediction accuracy of the GA-SVM model was evaluated by three indexes, including the coefficient of determination ( $R^2$ ), Nash-Sutcliffe efficiency (ENS) coefficient and root mean square error (RMSE), which are

defined as:

$$R^2 = \left[ \frac{\sum_{i=1}^n (Co(i) - \overline{Co}) (Cf(i) - \overline{Cf})}{\sqrt{\sum_{i=1}^n (Co(i) - \overline{Co})^2} * \sqrt{\sum_{i=1}^n (Cf(i) - \overline{Cf})^2}} \right]^2$$

$$ENS = 1 - \frac{\sum_{i=1}^n (Co(i) - Cf(i))^2}{\sum_{i=1}^n (Co(i) - \overline{Co})^2}$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (Cf(i) - Co(i))^2}$$

where  $Co(i)$  is the observed value,  $Cf(i)$  is the predicted value,  $\overline{Co}$  and  $\overline{Cf}$  represent the average values of the  $Co(i)$  and  $Cf(i)$  data series, respectively, and  $n$  represents the size of the data series. Generally, the higher the  $R^2$  and ENS and the smaller the RMSE, the higher the accuracy of the model is.

## CASE STUDY

### Study area description and dataset

The Min River is located in the Fujian province of China, encompassing a total catchment basin of 60,992 km<sup>2</sup> (Figure 2). The Min River is the main water resource for Fuzhou city, the capital of Fujian province, with a population of over seven million. The catchment is characterized by a typical subtropical monsoon climate. The annual runoff distribution is uneven due to seasonal variations between the wet season, which occurs from June to August, and the dry season from November to February. The MRE salinity significantly rises during the dry season due to a dramatic runoff decrease. The tide in the MRE is an irregular semi-diurnal tide. The annual mean of the daily tidal range is 4.14 m, as recorded at the Baiyantian Tide Station (from 1980 to 2013), which is located near the Min River mouth. Recently, riverbed incision has caused

saltwater intrusion, which poses a threat to drinking water safety in Fuzhou city. The most significant saltwater intrusions occurred in 2009 and 2013.

The Zhuqi Hydrological Station is located at the head of the estuary. It is the major mainstream hydrological observation station available for the Min River. The tidal stations are located at Wenshanli, Jiefang Bridge, Baiyantian and Guantou. The water supply department of Fujian Province has conducted long-term, continuous and simultaneous salinity observations at some river cross-sections, including at the intake of the Changle Water Plant. Large amounts of observational data have been collected over time to provide basic information for controlling saltwater intrusion. The observational data used for this study include the following:

1. Daily salinity series collected at the Changle Water Plant, which is 14 km from the estuary mouth. The salinity is measured as electrical conductivity.
2. Daily tide-level series collected at the Baiyantian Station, which is adjacent to the Changle Water Plant.
3. Average daily runoff series collected at the Zhuqi Station, which is located 41 km upstream of the Changle Water Plant.

The computing time of the GA-SVM model was improved by formatting the dataset using a min-max normal form. Supposing that  $\min_a$  and  $\max_a$  represent the minimum and maximum values for attribute  $a$  respectively, then the model maps a value  $v$  of  $a$  to  $v'$  in the range [0, 1] by computing:

$$v' = \frac{(v - \min_a)}{(\max_a - \min_a)}$$

### Analysis of the factors that influence salinity

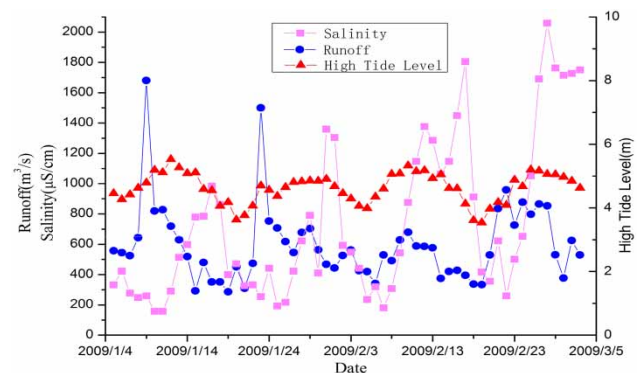
Numerous studies have shown that tides and runoff are the main contributors to salinity variations (Savenije 1993; Nguyen & Savenije 2006; Fei et al. 2011; Chevalier et al. 2014). Savenije suggests that the main factors influencing salinity variations are the tide, runoff and channel topography. In addition, Savenije (1993) successfully applied an empirical model to compute the longitudinal salinity variations



**Figure 2** | Map of the MRE.

in 15 estuaries worldwide. Nguyen & Savenije (2006) adapted Savenije's model to a multi-channel estuary, determining that salinity variations lag behind the tidal and runoff variations by several days. Fei *et al.* (2011) studied the time–frequency characteristics and multiscale correlations between runoff, tidal range and salinity in the Changjiang Estuary. The Zhuqi Station is the chief mainstream hydrological station, which is impacted by a watershed area of 54,500 km<sup>2</sup>, accounting for 89.6% of the entire watershed area. The Baiyantian Tide Station is adjacent to the Changle Water Plant. Therefore, the Baiyantian observational data can accurately reflect the tidal variations of the Changle Water Plant.

The relationships between the salinity, runoff and high-tide level from January 5 to March 4, 2009, are illustrated in Figure 3. There is correlation between high-tide level, runoff and salinity



**Figure 3** | The relationships between the salinity at the Changle Water Plant, runoff at the Zhuqi Station and high-tide level at the Baiyantian Station from January 5 to March 4, 2009.

(correlation coefficient between high-tide level and salinity: 0.51; between runoff and salinity: 0.37). Four salinity variation

processes occur over the time period due to biweekly spring and neap tidal cycles. In general, the salinity rises and falls based on high-tide level fluctuations. The tide is the primary factor influencing salinity variation. However, the highest tide level appears at an early time, while the early-time salinity is lower than the late-time salinity. This relationship is mainly because the early-time runoff is significantly higher than the other processes, and large runoff prevents salinity surges. The late-time runoff is high from January 23 to January 25, which inhibits the salinity. In addition, salinity variation lags behind the corresponding runoff and tide-level variations by 1 to 2 days over the entire time period. Similar findings have been reported in previous studies (Nguyen & Savenije 2006; Fei et al. 2011), but the causes of such a lag are rather complicated and still remain unclear.

An observational dataset collected in 2009, including the salinity at the Changle water plant, runoff at the Zhuqi Station and high tide level at the Baiyantian Station, was used to further determine the time lag between the salinity, runoff, and high-tide level. The previous salinity at the Changle Water Plant was also concerned, because current salinity is strongly correlated with previous salinity. Various combinations of the previous salinity, runoff and high-tide-level time series were fed into the GA-SVM model, which predicted the t-day salinity.

The performance statistics of the model are shown in Table 1, where  $C_t$ ,  $Q_t$  and  $L_t$  denote the t-day salinity, runoff and high-tide level, respectively. As seen in Table 1, the number of influence factors increases from experiments 1 to 4. In addition, the GA-SVM model prediction accuracy improved for experiments 1 to 4 based on the  $R^2$ , ENS and RMSE values. However, the prediction accuracy of the

GA-SVM model decreased after adding the t-3 day salinity, t-3 day runoff and t-2 day high-tide level in experiment 5. Experiment 4 provides the best prediction accuracies based on the  $R^2$ , ENS and RMSE.

Therefore, the model inputs for the t-day salinity prediction at the Changle Water Plant include these six factors:

1. the t-1 and t-2 day salinities at the Changle Water Plant;
2. the t-1 and t-2 day runoffs at the Zhuqi Station;
3. the t and t-1 day high-tide levels at the Baiyantian Station.

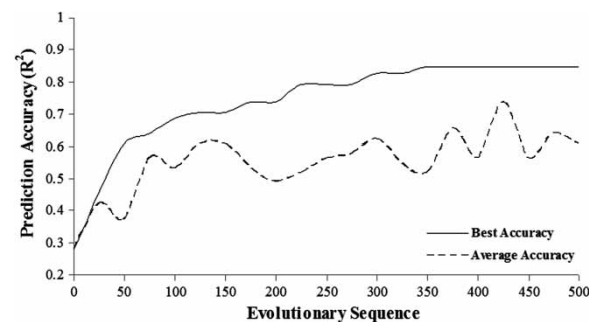
### Experimental validation

The GA is used to optimize the SVM parameters in the GA-SVM model. The optimization process curve is shown in Figure 4. The two curves represent the optimal accuracy rate and the average accuracy rate. The best accuracy increases significantly in the early stage and gradually reaches a stable state when the evolutionary sequence value exceeds 350.

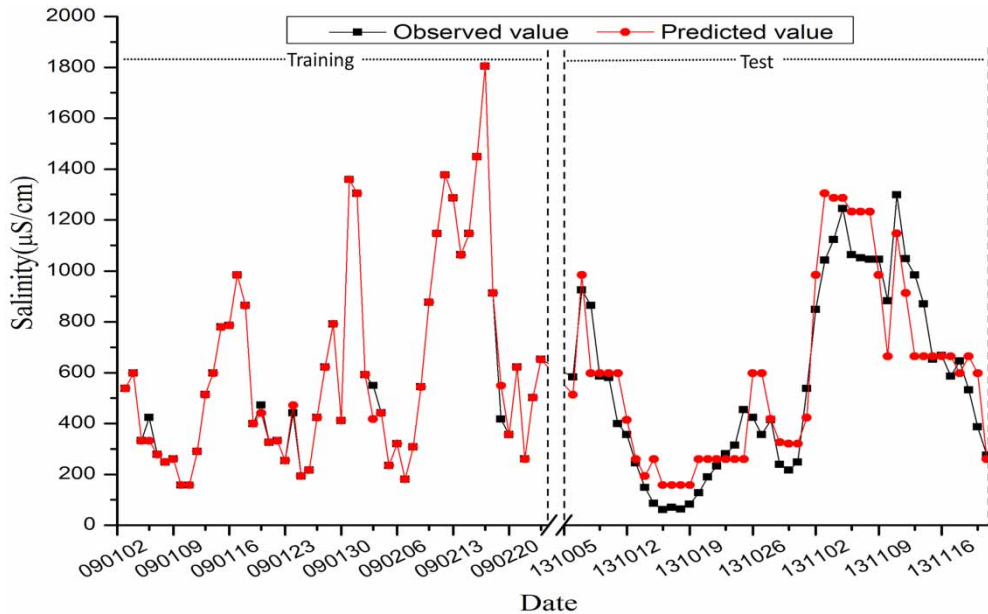
After decoding the best chromosome to obtain the optimal parameter values ( $C = 2.017$ ,  $\delta = 5.953$ ) for the GA-SVM model, the model predicted the salinity of the MRE. The observation data used to validate the model contained the salinity at the Changle Water Plant, runoff at the Zhuqi Station and high-tide level at the Baiyantian Station. Observational data collected in 2009 were chosen for training, while those in 2013 were chosen as the test dataset. The prediction results of the GA-SVM model are shown in Figure 5. The model exhibits a satisfactory accuracy, with  $R^2$ , ENS and RMSE values of 0.99, 0.99 and 7.88 ( $\mu\text{S}/\text{cm}$ ) during the training period, and 0.85, 0.84 and 119 ( $\mu\text{S}/\text{cm}$ ) during the test period, respectively.

**Table 1** | Experimental analysis of the salinity prediction results influenced by different combinations of factors

ID	Influence factors	$R^2$	Evaluation index	
			ENS	RMSE( $\mu\text{S}/\text{cm}$ )
1	$C_{t-1}, Q_{t-2}, L_{t-1}$	0.49	0.53	193
2	$C_{t-1}, Q_{t-1}, L_t$	0.60	0.62	174
3	$C_{t-1}, Q_{t-1}, Q_{t-2}, L_t, L_{t-1}$	0.76	0.78	142
4	$C_{t-1}, C_{t-2}, Q_{t-1}, Q_{t-2}, L_t, L_{t-1}$	0.83	0.84	126
5	$C_{t-1}, C_{t-2}, C_{t-3}, Q_{t-1}, Q_{t-2}, Q_{t-3}, L_t, L_{t-1}, L_{t-2}$	0.78	0.80	138



**Figure 4** | The optimization process curves.



**Figure 5** | The 2009 training and 2013 prediction of the GA-SVM model.

### Comparison of the model performance of GA-SVM with traditional SVM

The prediction performance of the GA-SVM model was evaluated by using a traditional SVM model with optimized grid-search-based parameters. The performance statistics of the two models are shown in Table 2. The computing time for the GA-SVM model is approximately half of that for the SVM model, but the prediction accuracy of the GA-SVM model is significantly better. These results suggest that the GA-SVM model provides improved feasibility and practicability compared to the SVM model.

## CONCLUSIONS

An application of GA-SVM for tidal estuary salinity prediction was proposed in this paper. First, by conducting an

analysis of a time-series of daily salinity in 2009 and the results of simulation experiments, the  $t-1$  and  $t-2$  day salinities,  $t$  and  $t-1$  day high-tide levels and  $t-1$  and  $t-2$  day runoffs were determined to be the major factors that influence the  $t$ -day salinity predictions. Then, a coupled GA-SVM model was developed to predict the typical salinity process of the MRE in 2013. The results show that the model prediction achieves a satisfactory accuracy, with  $R^2$  of 0.85, ENS of 0.84 and RMSE of 119 ( $\mu\text{S}/\text{cm}$ ). The proposed model performs significantly better than the traditional SVM model in terms of prediction accuracy and computing time. It can be concluded that the proposed model can successfully predict the salinity of the MRE based on the high-tide level, runoff and previous salinity.

This coupled GA-SVM modelling approach can be easily applied to other estuary systems. Since the major factors that influence tidal estuary salinity may be different from one estuary to another, efforts should be spent on first identifying the major factors when using the model.

**Table 2** | Prediction results of the GA-SVM compared to the traditional SVM

Year	GA-SVM		Traditional SVM	
	$R^2$	Computing time (s)	$R^2$	Computing time (s)
2009	0.85	1,029.26	0.67	2,062.52
2013	0.85	1,026.64	0.67	2,061.74

## ACKNOWLEDGEMENTS

This study was financially supported by the science and technology major project of Fujian Province, China



(2015Y4002) and the science and technology project of Fujian Provincial Education Department, China (No. JA15731).

## REFERENCES

- Cherkassky, V. & Ma, Y. Q. 2004 Practical selection of SVM parameters and noise estimation for SVM regression. *Neural Networks* **17**, 113–126.
- Chevalier, C., Pagano, M., Corbin, D. & Arfi, R. 2014 The salinity responses of tropical estuaries to changes in freshwater discharge, tidal mixing and geomorphology: case study of the man-affected Senegal River Estuary (West Africa). *Marine and Freshwater Research* **65**, 987–1002.
- CSIE (Department of Computer Science & Information Engineering National Taiwan University). LIBSVM. <https://www.csie.ntu.edu.tw/~cjlin/libsvmtools/> (accessed 30 December 2015).
- Fei, Y. J., Xu, L. L., Du, P. J., Guan, Q. L., Kang, X. & Xiao, W. J. 2011 Analysis of characteristics of time-frequency correlations between runoff, tidal range and salinity in the Changjiang Estuary. *Acta Oceanologica Sinica* **30** (5), 84–93.
- Guan, X. Y., Wang, S. L., Gao, Z. Y. & Lv, Y. 2013 Dynamic prediction of soil salinization in an irrigation district based on the support vector machine. *Mathematical and Computer Modelling* **58**, 719–724.
- Harish, N., Loksha, Mandal, S., Rao, S. & Patil, S. G. 2014 Parameter optimization using GA in SVM to predict damage level of non-reshaped berm breakwater. *The International Journal of Ocean and Climate Systems* **5** (2), 79–88.
- Huang, W. R. & Foo, S. 2002 Neural network modeling of salinity variation in Apalachicola River. *Water Research* **36**, 356–362.
- Li, X. Z. & Kong, J. M. 2014 Application of GA-SVM method with parameter optimization for landslide development prediction. *Natural Hazards and Earth System Sciences* **14**, 525–533.
- Liu, D. D. & Chen, X. H. 2007 Model for prediction of saltwater intrusions based on coupling of support vector machine and partial least square method. *Acta Scientiarum Naturalium Universitatis Sunyatseni* **46** (4), 89–92 (in Chinese).
- Liu, M. & Lu, J. 2014 Support vector machine – an alternative to artificial neuron network for water quality forecasting in an agricultural nonpoint source polluted river? *Environmental Science and Pollution Research International* **21**, 11036–11053.
- Liu, J. J., Li, Z., Hu, F. R., Chen, T. & Zhu, A. J. 2015 A THz spectroscopy nondestructive identification method for transgenic cotton seed based on GA-SVM. *Optical and Quantum Electronics* **47** (2), 313–322.
- Nguyen, A. D. & Savenije, H. H. G. 2006 Salt intrusion in multi-channel estuaries: a case study in the Mekong Delta, Vietnam. *Hydrology and Earth System Sciences* **10**, 743–754.
- Nowroozi, A. A., Horrocks, S. B. & Henderson, P. 1999 Saltwater intrusion into the freshwater aquifer in the eastern shore of Virginia: a reconnaissance electrical resistivity survey. *Journal of Applied Geophysics* **42**, 1–22.
- Reddy, G. S. & Ghosh, S. N. 1993 Aspects of a computational model for salinity variations in a well-mixed tidal reach. *Water Science and Technology* **28** (3–5), 659–667.
- Savenije, H. H. G. 1993 Predictive model for salt intrusion in estuaries. *Journal of Hydrology* **148**, 203–218.
- Soetaert, K., Vincx, M., Wittoeck, J. & Tulkens, M. 1995 Meiobenthic distribution and nematode community structure in five European estuaries. *Hydrobiologia* **311**, 185–206.
- Thomas, A. S. & Tris, H. W. 1996 Ecological responses to reductions in freshwater supply and quality in South Africa's estuaries: lessons for management and conservation. *Journal of Coastal Conservation* **2** (2), 115–130.
- Vapnik, V. N. 1998 *The Nature of Statistical Learning Theory*. Wiley, New York.
- Wen, P., Chen, X. H., Liu, B. & Yang, X. L. 2007 Analysis of tidal saltwater intrusion and its variation in Modaomen Estuary. *Journal of China Hydrology* **27**, 65–67 (in Chinese).
- Whitley, D. 1994 A genetic algorithm tutorial. *Statistics and Computing* **4** (2), 65–85.
- Yu, P. S., Chen, S. T. & Chang, I. F. 2006 Support vector regression for real-time flood stage forecasting. *Journal of Hydrology* **328**, 704–716.

First received 18 January 2016; accepted in revised form 25 May 2016. Available online 28 June 2016