Optimising the modelling of eutrophication for Bohai Bay based on the cellular automata – support vector machine method

Zheng Dongsheng, Xiang Xianquan and Tao Jianhua

ABSTRACT

With the development of marine economy, eutrophication has become one of the key issues in the marine environment. In this paper, a eutrophication model for Bohai Bay based on the cellular automata-support vector machine (CA-SVM) has been established by applying the soft computing approach with a large quantity of remote sensing data to the marine environment. In order to optimise the coupled model further, two main tasks have been done in this study. First, to choose reasonable influence factors as the input parameters of the model, nine series of training and simulation exercises were conducted based on nine different types of input parameter combinations. A reasonable input parameter combination was selected, and the eutrophication model (the basic model) was established by the comparative analysis of the simulation results. Second, according to Shelford's Law of Tolerance, an optimised model was developed. It is combined of nine special models and each model corresponds to a stage of sea surface temperature and the chlorophyll-a concentration, respectively. The comparison between the optimised model and the basic model indicated that prediction accuracy was improved by the optimised model. By this study, it can be observed this model could provide a scientific basis for the prediction and management of the aquatic environment of Bohai Bay.

Key words | Bohai Bay, cellular automata, chlorophyll-a concentration, eutrophication model, hydroinformatics, support vector machine

Zheng Dongsheng
Tao Jianhua (corresponding author)
School of Mechanical Engineering,
Tianjin University,
Tianjin 300072,
China
E-mail: *jhtao@tju.edu.cn*

Xiang Xianquan
The National Marine Data and Information Service,
Tianjin 300171,
China

INTRODUCTION

In recent years, with the rapid development of the economy in the Bohai Bay region, pollution of the coastal waters has become more and more serious, the degree of the water eutrophication has increased and red tide has frequently been observed. Bohai Bay, which is a semi-enclosed shallow muddy bay, is located in the western part of the Bohai Sea. The water exchange and self-purification capacity of Bohai Bay are notably weak (Sun & Tao 2006). Establishing a eutrophication model for Bohai Bay has practical significance to respond to the early warning of the eutrophication and red tide for environment protection.

Water eutrophication has become increasingly serious since the beginning of the twentieth century and, gradually, people have paid more attention to it. A massive amount of research has been carried out on eutrophication and in the early period could be divided into two types. One type of research was conducted through experimental methods. The relationship between the biological growth and the environment could be determined through the study of the growth environment of aquatic plants and animals. Another type of research was the numerical simulation method. As part of the mechanism of eutrophication is not fully understood, it was difficult to get a good predicted result. In summary, a significant contribution had been made in studying eutrophication and exploring the mechanism by the two types of methods.

doi: 10.2166/hydro.2014.098

In recent years, with the development of modern monitoring technology and the diversification of monitoring tools, large amounts of data can be gathered quickly, not only in quantity, but also in coverage and complexity, and quantities of measured data are greater than ever before. Meanwhile, information technology has also developed rapidly accompanied by the rise of data-driven technologies that require computing intelligence as the main direction for development (such as genetic algorithms (GA) (Yao *et al.* 2007), artificial neural networks (Islam 2010), fuzzy systems (Salski & Sperlbaum 1991; Chen *et al.* 2003) and cellular automata (CA)). All of these circumstances provide the opportunity for the development of a model for eutrophication.

At the beginning of the 1990s, hydroinformatics was gradually applied to the study of ecological forecasting and eutrophication. The development of the eutrophication model was accelerated. Friedrich *et al.* (2002) compared potentials and achievements of artificial neural networks and GA in terms of forecasting and understanding of algal blooms in Lake Kasumigaura in Japan. Lee *et al.* (2003) used an artificial neural network to predict the algal bloom dynamics of the coastal waters of Hong Kong. Chau (2004) developed and implemented a prototype expert system for model manipulation for flow and water quality by employing an expert system shell. Chen & Mynett (2004) successfully applied fuzzy logic to model algal biomass. Chau *et al.* (2005) employed two hybrid models based on artificial intelligence technology and adaptive-network-based fuzzy inference system for flood forecasting in a channel reach of the Yangtze River in China. Cheng *et al.* (2005) used several artificial neural network models to forecast daily and monthly river flow discharges in Manwan Reservoir. Muttil & Chau (2006) selected genetic programming for modelling and prediction of algal blooms with the artificial neural network in Tolo Harbour. An & Tao (2007) applied fuzzy algorithms in a eutrophication model and used the two-grade fuzzy synthesis assessment method to assess the eutrophication level in Bohai Bay. Ma (2007) used the information technologies of chaos theory, neural networks and GA, etc., to study the eutrophication assessment model of Bohai Bay. Chau & Muttil (2007) carried out a spatial-temporal analysis of Tolo Harbour (in Hong Kong), analysing coastal water quality data by using data mining and multivariate statistical analysis. Wu *et al.* (2009) proposed a crisp distributed support vectors regression model for monthly streamflow prediction in comparison with four other models. Tao & Xiang (2009) adopted GA to adjust the internal parameters of ANFIS (adaptive network-based fuzzy inference system) for establishing a eutrophication prediction model for Bohai Bay. Yu (2011) proposed a model of the back propagation (BP) neural network, which was based on the particle swarm optimisation algorithm and explored its application to nonlinear prediction of eutrophication in Ming Lake. Adamowski *et al.* (2012) developed multivariate adaptive regression spline and wavelet transform artificial neural network models with limited data for runoff forecasting applications in the mountainous watershed of Sainji in the Himalayas. Riccardo *et al.* (2012) presented an application of feed forward neural networks (FFNs) for long period simulations of hourly groundwater levels in a coastal unconfined aquifer sited in the Venice Lagoon. Liang *et al.* (2013) developed a generic canal model to simulate the hydro-environmental processes for inland navigational canals. Xiang *et al.* (2013) used several hybrid soft computing algorithms including a support vector machine (SVM), GA and CA to predict the concentration of chlorophyll-a in Bohai Bay, in which a conclusion was reached that hybrid soft computing algorithms constituted the preferred tool to predict the concentration of chlorophyll-a in Bohai Bay. Chen & Mynett (2006) noted that the future of ecohydraulics and eco-environmental modelling seemed to lie in the integration of different paradigms and techniques.

Although simulation study has developed quickly recently, it has some limitations. Due to the fact that the measured data are limited and distribution is uneven, the influence of data information in the neighbour region has not often been considered. Also, the generalisation property needs to be further improved.

In this situation, the remote sensing data and the flow field data were introduced to study eutrophication. The abundant remote sensing data has better continuity over time, and its distribution is wide. Also, the flow field data were supplied by the hydrodynamic model which has been strictly verified. In order to consider the influence of the neighbour region and improve the generalisation property of the eutrophication model, the CA and the SVM were coupled to do the simulation.

CA constitute a grid dynamic model in which time, space and state are all discrete, the spatial grid exhibits

mutual effects and causal relations are mostly local. The use of CA is similar to the Eulerian approach in classical fluid mechanics (Abbott & Minns 1998). CA take the spatial unit as the object, study its time evolution and the space motion and, finally, produce the spatial-temporal pattern of the system, a typical pattern of discrete mode and spatially explicit mode. Chen *et al.* (2009) noted that CA seemed to be more advantageous than partial different equations (PDEs) due to the inherent characteristics of ecosystems in ecological modelling. As a spatially explicit model, CA have been widely used in ecological simulation studies and have obtained good results. Examples of some good results include the forest dynamic simulation (David & Ricard 2000), aquatic plant competitiveness growth simulation (Chen *et al.* 2002), dynamic simulation of vegetation on the shore (Chen & Ye 2008; Guan *et al.* 2009) and wetland plant simulation (Ellison & Bedford 1995).

SVM, which can simulate the complex nonlinear relationships, is based on statistical learning theory. SVM takes the Vapnik–Chervonenkis dimension (VC-dimension) theory and structural risk minimisation principles as the basic theories. SVM can determine a global optimal solution and overcome the shortcomings of empirical risk minimisation exhibited by neural networks. SVM has a good ability to extrapolate in small sample situations. In recent years, SVM has been gradually applied to eutrophication forecast model research and has achieved results in certain aspects (Feng *et al.* 2007; Behzad *et al.* 2009). Many comparisons between neural network and SVM have been done by many researchers. The results have shown that the SVM had better prediction accuracy and generalisation property than neural network. Francis & Cao (2001) used SVM and a multi-layer BP neural network to do financial time series forecasting. The experiment showed that SVM outperformed the BP neural network based on the criteria of normalised mean square error (NMSE), mean absolute error (MAE), directional symmetry (DS) and weighted directional symmetry (WDS). The results showed that, to take the prediction data of CME-SP, compared with the BP neural network, SVM gave a reduction of 10.7% for NMSE, reduction of 7.6% for MAE, increase of 5.5% for DS and decrease of 10.8% for WDS. Srinivas *et al.* (2002) used neural networks and SVM to construct an intrusion detection system based on a set of benchmark DARPA data. The results showed that SVM had great potential to be used to replace neural networks due to its scalability and faster training and running time. For example, the training time for SVM is significantly shorter than the BP neural network (17.77 s versus 18 min). Doniger *et al.* (2002) used the multilayer perception neural network and a SVM to predict the blood–brain barrier permeability of different classes of molecules. The results showed that the SVM outperformed the neural network. The SVM can predict up to 96% of the molecules correctly, averaging 81.5% over 30 test sets, compared with the neural network's average performance of 75.7% with the same test sets. More comparisons can be referred to in Evgeny *et al.* (2003) and Samsudin *et al.* (2010).

Phytoplankton is the first level of the food chain. The production capacity of phytoplankton is the material base of marine life and is also the primary factor that leads to eutrophication. As there are many different types of phytoplankton with significant individual differences, the chlorophyll-a content is taken as the main indicator of the phytoplankton biomass and its dynamic changes (Li *et al.* 2005). In the assessment and monitoring of marine environments, the chlorophyll-a content is always used as the main factor to determine whether eutrophication and red tide occur. Therefore, the study of the eutrophication of Bohai Bay was conducted by choosing chlorophyll-a as the representative indicator.

Before this study, a series of researches was done by our research group in this field. Two aspects were included in the previous work: (1) the chlorophyll-a prediction model of Bohai Bay was established by using the fuzzy pattern recognition, genetic algorithm and artificial neural network; (2) the soft computing method was comprehensively studied in the chlorophyll-a prediction model. First, the GA was used to optimise the parameters of the SVM. Second, the coupled model of CA and the SVM was preliminarily used to predict the chlorophyll-a concentration.

Based on these, the purpose of this study is to optimise the coupled model of the CA and the SVM. Two tasks were mainly carried out. (1) More comprehensive influence factors were considered. The hydrodynamic parameters flow field, wind field were introduced as the influence factors. Through the comparison and analysis of multiple sets of scheme, the best combination of influence factors was produced. (2) Shelford's Law of Tolerance (Shelford 1913) and its supplement (Odum 1983) in ecology was introduced to improve the

model prediction accuracy. According to Shelford's Law of Tolerance and its supplement in ecology, the training data were divided into nine groups and each group corresponded to a stage of sea surface temperature and the chlorophyll-a concentration, respectively. After training each group respectively, nine models were set up. Through the method, the optimised model composed of nine small models is constructed. In the process of application, one small model was selected first according to the index which is calculated by the prediction basic data. Finally, the prediction accuracy was improved through choosing the pertinent model.

MODEL

Cellular automata

The composition of CA is introduced in six parts: cell, cellular state, cellular space, neighbour, rules and time.

- *Cell:* The most basic component of CA is termed 'cell'. It is also called unit or element. Cell is distributed on the lattice point of one-dimensional, two-dimensional or multi-dimensional Euclid space.
- *Cellular state:* The cellular state is the description of cellular properties. It can be the binary form of {0, 1} or integer form of {S1, S2, …, Sn}.
- *Cellular space:* The spatial grid system on which cells are distributed is called cellular space.
- *Neighbour:* Neighbours are the combination of cells that could affect the evolution of the central cell state. Von Neumann type, Moore type and Extended Moore type were included in the usual neighbour types. As shown in Figure 1, the grid which has vertical stripes is the central cell, and the grids which have cross stripes represent neighbours.
- *Rules:* Rule is also known as the conversion rule, it is a state transfer function, namely the dynamic function that determines the cellular state of next moment according to its current state and its neighbours' current state. The key for the CA to work successfully is whether the designed conversion rule is reasonable, and whether the conversion rule can reflect the inherent characteristic of objective things.
- *Time:* CA are a dynamic system. Its change in the time dimension is discrete, namely the time t is an integer value, continuous and the interval is equal. In the conversion function, the cellular state of time $t+1$ is determined

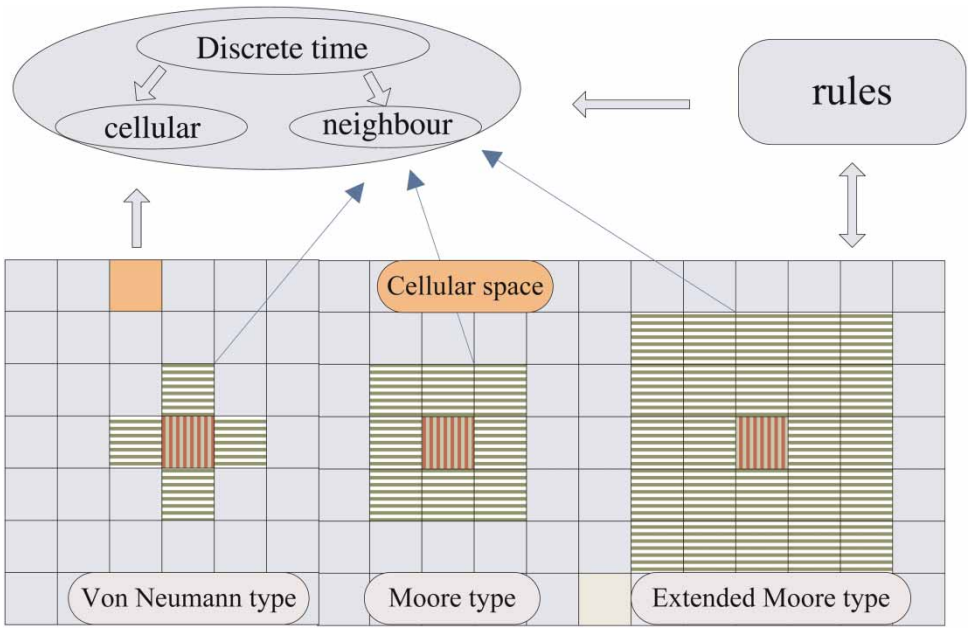


Figure 1 | The CA model and the common neighbour types.

by the cellular state of time t and its neighbours' state of time t.

The cellular space could be expressed as follows (Amoroso & Patt 1972):

$$A = (L_d, S, N, f) \tag{1}$$

where A represents a CA system, L represents cellular space, d is an integer and represents the number of dimension, S represents the combination of limited discrete states, N represents the combination of all states of centre cell and its neighbours, and f is the conversion rules.

A single cellular model could be expressed as the following form:

$$S^{t+1} = f(S^t, N) \tag{2}$$

where t, $t+1$ is the discrete time, S^{t+1} is the cellular state at time $t+1$, and S^t is the cellular state at time t. The formula indicates that the state of time $t+1$ is determined by the state of time t, the combination of neighbour states and the conversion rules.

Before using CA, an assumption was made that each cell was uniform, and the spatial heterogeneity may be affected. Due to the fact that the cell was small and the amount of the cells is large, the spatial heterogeneity of the whole of Bohai Bay could be still reflected.

Support vector machine

SVM was developed by Vapnik *et al.* (Boser *et al.* 1992; Vapnik 1998; Cortes & Vapnik 1995) at Bell Laboratories. SVM is a novel machine learning methodology that is based on more than 30 years of research on statistical learning theory by Vapnik *et al.*

The algorithm principle of SVM is as follows: suppose the sample set $\{(x_i, y_i), i = 1, 2, \ldots m\}$, $x_i \in R^N$ is the input value, $y_i \in R^N$ is the corresponding target value. ε is defined as the insensitive loss function. The training error is defined as follows:

$$L_\varepsilon(Y, f(x)) = \max(0, |y - f(x)| - \varepsilon) \tag{3}$$

The insensitive loss function ε reflects the regression model sensitivity to the noise that the input variable contained. The insensitive loss function can control the fitting precision of the model.

First, the linear regression function $f(x) = \omega x + b$ is considered. According to the structural risk minimisation principle, the function estimation is to obtain the function $f(x)$ that makes the function R_{reg} the smallest

$$R_{\text{reg}} = \frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{m} L_\varepsilon(y_i, f(x_i)) \tag{4}$$

where $1/2\|\omega\|^2$ reflects the complexity of the model, $\sum_{i=1}^{m} L_\varepsilon(y_i, f(x_i))$ reflects the training errors, namely, the experience risk, and C is set as the error penalty parameter, used to control the degree of punishment for misclassifying samples and realising the compromise between misclassifying samples and the model complexity.

Through transforming this risk minimisation question to the solving of the constrained optimisation question and taking the duality theory and the Lagrange multiplier method to transform the form, the dual form of this optimised question can be obtained and is described as follows:

$$\begin{cases} \min \dfrac{1}{2}\displaystyle\sum_{i,j=1}^{m}(a_i^* - a_i)(a_j^* - a_j)(x_i \cdot x_j) + \varepsilon\sum_{i=1}^{m}(a_i^* + a_i) \\ \quad - \displaystyle\sum_{i=1}^{m} y_i(a_i^* - a_i) \\ \text{subject to } \displaystyle\sum_{i=1}^{m} y_i(a_i - a_i^*) = 0 \end{cases} \tag{5}$$

$$0 \le a_i, a_i^* \le C \tag{6}$$

The Lagrange multipliers a and a^* can be obtained by using the quadratic programming method of the optimisation theory. Only a small part of a and a^* are not 0, and they are called support vectors. The bias b can be obtained according to the Karush–Kuhn–Tucker condition in convex quadratic programming. Thus the decision function of SVM can be determined:

$$f(x) = \sum (a_i^* - a_i)(x_i, x) + b \tag{7}$$

When the sample set is nonlinear, a nuclear function could be introduced to map the sample set to high dimensional Hilbert space and change the nonlinear issues to linear high-dimensional separable issues. The issue is finally solved. The detailed solution process can be found in Deng & Tian (2009).

The coupling of CA and SVM

In order to take advantage of both the CA and SVM, the coupled model was established.

In the coupled model, CA is used to establish the frame. Thus, Bohai Bay was divided into 150×150 cells. The status of each cell is the description of hydrodynamic, water quality and ecological parameters of each cellular region, such as the sea surface temperature, the chlorophyll-a concentration, wind speed, flow field and so on. The status of each cell is shown in Figure 2.

For the CA, the rule can be defined by the model builder, or could be obtained through training the test data by using appropriate soft computing methods. For example, it is obtained by using the SVM in this study.

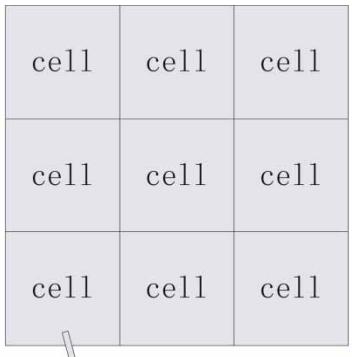
Each cell has its own input and output, and it is related to a set of input and one output. The training data of all the cells are trained together to get a rule. When the prediction is taken, this rule is applied to all the cells for the prediction.

The chlorophyll-a concentration of each cell at time step $t+1$ is determined by the status of the neighbour cells and the central cell itself at time step t. The data flow of the coupled model is explained in Table 1, and the parameters selected in the procedure are described in the section The choice of reasonable input parameters.

Parameters are chosen by the relevancy to the change of the chlorophyll-a concentration. In this study, the parameters were chosen by the prediction results comparison, and these are introduced in the section The establishment and optimisation of the eutrophication model.

SVM is based on the VC-dimension theory and the structural risk minimisation principle. SVM has a rigorous theoretical foundation. Before this study, a contrast between the SVM and the BP neural network was done. The results showed that SVM has better generalisation property than the BP neural network. More details can be referenced in Xiang (2011).

Therefore in order to construct the eutrophication prediction model, the CA and the SVM were coupled. The CA were used to set up the framework, and the SVM was used to extract rules from the training. These rules take full advantage of the CA characteristics of dealing well with the spatial heterogeneity and the SVM characteristics of modelling the nonlinear problem well.



Sea surface temperature	wind speed	location	the chlorophyll-a concentration	flow field	Other parameters
15℃	U_wind = 5 m/s V_wind = 2 m/s	(117.6, 38)	3 μg/L	U = 3 m/s V = 2 m/s	...

Figure 2 | The status of each cell.

Table 1 | The data flow of the coupled model

	Model input (time t)					Model output (time $t+1$)
Cellular space	Input param.1	Input param.2	...	Input param.n		chlorophyll-a concentration
Cell 1	*	*	*	*	SVM	?
Cell 2	*	*	*	*		?
Cell 3	*	*	*	*		?
...	*	*	*	*		?
Cell n	*	*	*	*		?

In this table, '*' shows the data are known and '?' shows the data would be predicted.

The overall running process of the coupled model is shown in Figure 3.

The theoretical basis of model optimisation

Temperature is the essential factor of biological vital activity. All creatures live in the external environment at a certain temperature and are affected by the change of that temperature. When the temperature is higher or lower than the temperature range that the creature could tolerate, the biological growth would be hampered and the creature would even die.

Shelford (1913) proposed that biological adaptation to its environment has ecological minimum and maximum limits within which organisms could survive. Only between these two limits can organisms survive. The minimum and maximum limits are called the biological tolerance range. Odum (1983) supplemented the Law of Tolerance by stating that the adaptation range of the same organisms to the ecological factors is different at different stages of growth. The requirement of ecological conditions is usually stricter in the reproductive stage. In a certain temperature range, the phytoplankton growth would be promoted. In a certain temperature range, the phytoplankton reproduction may be affected by the temperature, and thereby phytoplankton growth rates may be affected, but the life of the organism is not threatened. At a certain temperature, the temperature causes the death of phytoplankton and, finally, causes the concentration of chlorophyll-a to be reduced. Similarly, the population density of phytoplankton is affected by the population density itself. When the population density is in a certain range, there is no competition within the species, and the population grows rapidly. When the population density grows to a certain range, factors such as nutrition and moisture are saturated, so competition would occur between the populations. At this time, the population growth rate would be slowed or even stopped. When the population density exceeds a certain boundary, the population competition would increase and thus cause partial

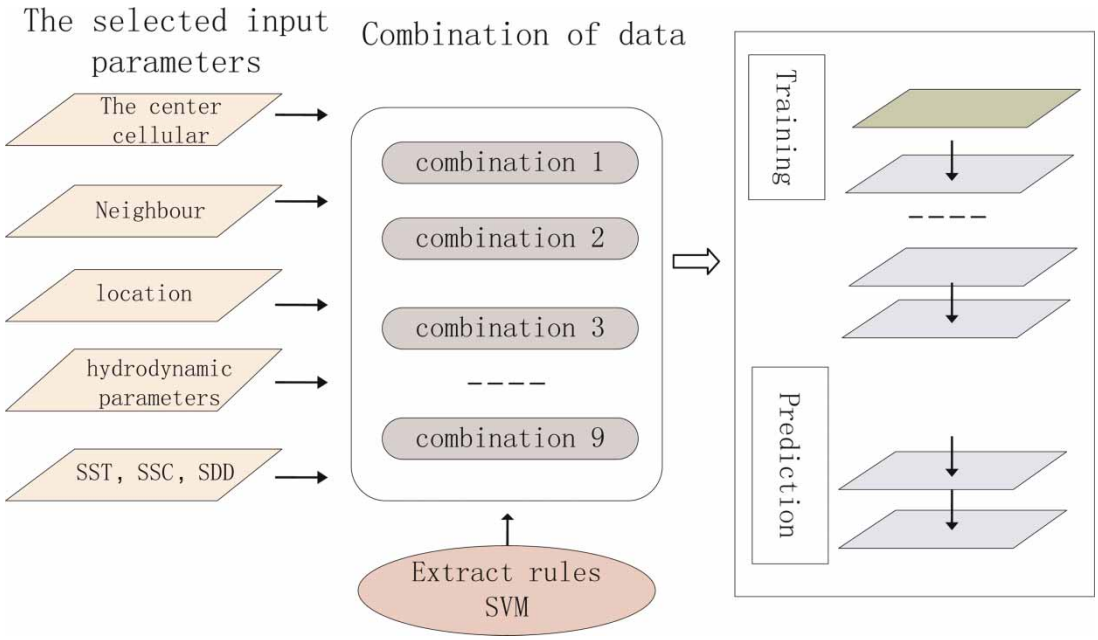


Figure 3 | The overall running process of the coupled model.

biological individual death. Competition causes the population density to be reduced. Therefore, the size of biological growth rates is closely related to the stage of the growing conditions, especially the temperature and the population density itself.

OBJECT OF STUDY AND DATA

Object of study

Study area

The framework of the Bohai Bay eutrophication model is established by using CA. In this model, the Bohai Bay is divided into 150×150 grids. The grid spatial resolution is 0.01° latitude by 0.01° longitude (about 1,100 m $\times$ 1,100 m). Each cell corresponds to a grid, and the cellular space corresponds to all grids. The cellular space is shown in Figure 4.

Neighbour and rules

Due to multidirectional sea water movement and the biological interactions, the Moore type (one of the neighbour types) was chosen as the neighbour type of Bohai Bay cellular space.

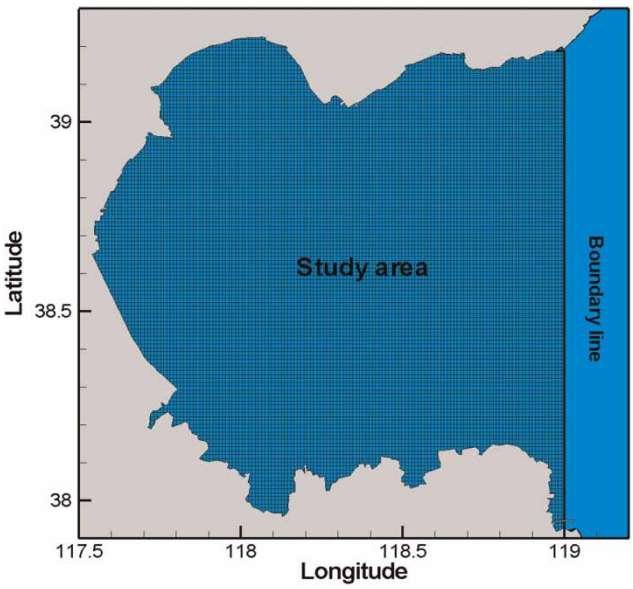


Figure 4 | The study area and the cellular space of Bohai Bay.

The conversion rules are obtained from training the sample data. These rules were applied to carry on the model forecast.

Data

Data sources

Data such as the concentration of chlorophyll-a (CHL), suspended sediment concentration (SSC), Secchi disk depth (SDD), sea surface temperature (SST), satellite wind field data and hydrodynamic flow field data were used in this study.

The CHL, SSC, SDD and SST data were obtained from the China National Satellite Ocean Application Service, which provides remote sensing data from the national marine-1 satellite. The national marine-1 satellite (HY-1B), launched on April 11, 2007, is the satellite follow-up star of the national marine-1 satellite (HY-1A). The data mainly include CHL, SSC, SDD and SST. The measurement units are μg/L, mg/L, m and °C, respectively. The period of data is from April 14, 2007 to May 15, 2010.

The satellite wind field data come from the further analysis of information from the European Centre for Medium-Range Weather Forecasts.

Hydrodynamic flow field data were provided by the two-dimensional hydrodynamic model of our research group. The detailed description and validation of the model are found in Sun (2007). These data could be used under the assumption that the flow field is uniform in the vertical direction. As the depth of Bohai Bay was shallow and the flow field is relatively uniform in the vertical direction, the flow field data were used as the alternative data.

Data processing

As the growth and extinction process of phytoplankton was relatively slow, all the data were processed in an average of a week. The time interval was set to 1 week. Finally, data were obtained in 162 weeks after processing.

Due to rainy or cloudy weather, the satellite could not detect all the real-time and full range of information needed, so that data for some days and some space were missing. The rejection method was used to address the missing data in this study.

Selection of input parameters

Like the multi-directional movement of sea water, the chlorophyll-a concentration for the next time step must be affected by that of the centre cell and the neighbour cells at the last time step. Therefore, the chlorophyll-a concentration of centre cell and neighbour cells at last time step are taken as the optional input parameters of the eutrophication model.

As the differences in spatial location, for example, nutrient distribution, the distance to the pollution sources, solar radiation, etc., could affect the concentration of chlorophyll-a, the spatial location was also chosen as one of the optional input parameters for the eutrophication model.

The growth of the algae was affected by photosynthesis, and the light is an important factor. Therefore, SDD was chosen as one of the optional input parameters of the eutrophication model.

As the photosynthesis of phytoplankton is affected by the luminous intensity which is directly affected by SSC, SSC was chosen as one of the optional input parameters of the eutrophication model.

When the light intensity reaches a saturated value, the temperature itself clearly affects the photosynthesis of algae in a way that the rate of photosynthesis increases as the temperature rises. In the beginning, the rate increases rapidly, then the rate decreases, and finally, the rate declines rapidly (Zhu *et al*. 2010). Therefore, SST was chosen as one of the optional input parameters of the eutrophication model.

As the phytoplankton may be moved under the influence of a flow field and a wind field, the hydrodynamic flow field and the wind field were chosen as optional input parameters for the eutrophication model.

Finally, the chlorophyll-a concentration of the centre cell and its neighbour cells at the last moment, the location, SDD, SST, SSC, the flow field and the wind field were taken as the optional input parameters of the eutrophication model.

THE ESTABLISHMENT AND OPTIMISATION OF THE EUTROPHICATION MODEL

The establishment and optimisation of the eutrophication model requires two steps:

(1) Establish the eutrophication model by training through choosing a reasonable influent factor combination as the input parameters.
(2) Optimise the eutrophication according to Shelford's Law of Tolerance (Shelford 1913) and its supplement (Odum 1983).

The flow chart of the establishment and optimisation of the eutrophication model is shown in Figure 5.

The choice of reasonable input parameters

There are several factors influencing the concentration of chlorophyll-a. Choosing the reasonable influence factors as the input factors of the eutrophication model could improve the precision of the prediction of the model.

Nine types of combinations have been obtained in this study by choosing different input factors. Through the contrast in the nine groups of prediction results, the best combination of input factors was chosen. The corresponding model is called the basic model.

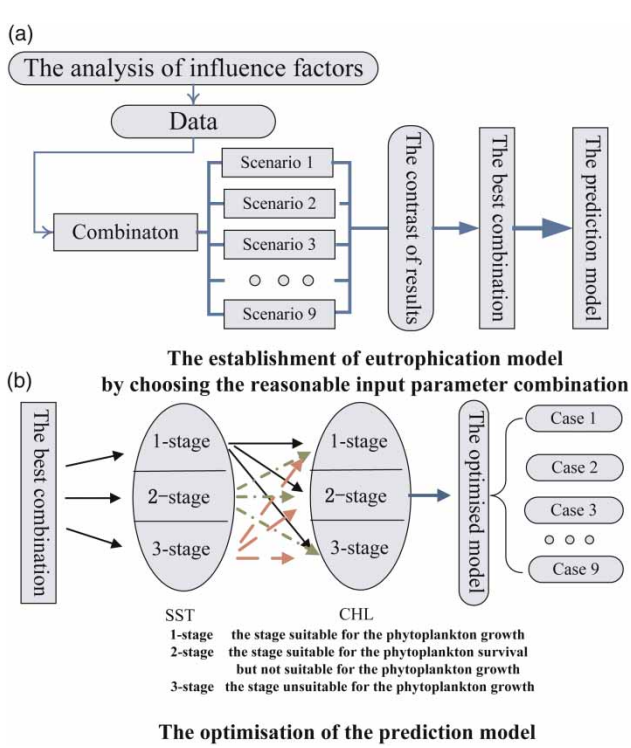


Figure 5 | The flow chart of the establishment and optimisation of the eutrophication model.

The nine scenarios of combinations of influence factors are shown in Table 2.

Annotation: The centre cell denotes the centre cell chlorophyll-a concentration. The neighbour denotes the chlorophyll-a concentration of eight neighbour cells. U denotes the flow velocity in the X direction; V denotes the flow velocity in the Y-direction.

The choice of the model parameters

The Vapnik (1998) research indicated that the performance of SVM was not affected obviously by the choice of the kernel function. RBF kernel function is a universal kernel function, and it has only one controllable parameter γ. Thus the RBF kernel function was chose as the kernel function of the SVM.

There are two parameters that needed to be determined while using RBF kernels: penalty parameter C and kernel function parameter γ. The grid research method was chosen as the method of parameter optimisation after comprehensively considering the time and the effect. The five-fold cross validation was used to carry out parameter optimisation. The scopes of parameters C and γ are set $[2^{-3}, 2^{3}]$, and the step is 0.01. The results of nine scenarios' parameter optimisation are shown in Table 3.

Table 2 | The nine scenarios of combinations of influence factors

	Factors
Scenario 1	The centre cell, the neighbour
Scenario 2	The centre cell, the neighbour, the coordinate
Scenario 3	The centre cell, the neighbour, the coordinate, SDD, SSC, SST
Scenario 4	The centre cell, the neighbour, the coordinate, the wind
Scenario 5	The centre cell, the neighbour, the coordinate, U, V
Scenario 6	The centre cell, the neighbour, the coordinate, SDD, SSC, SST, the wind
Scenario 7	The centre cell, the neighbour, the coordinate, SDD, SSC, SST, U, V
Scenario 8	The centre cell, the neighbour, the coordinate, the wind, U, V
Scenario 9	The centre cell, the neighbour, the coordinate, SDD, SSC, SST, U, V, the wind

Table 3 | The results of parameter optimisation

	Best C	Best γ	Mean σ (μg/L)
Scenario 1	1.0718	1.0718	0.1549
Scenario 2	1.0718	2.1435	0.1194
Scenario 3	1.1487	1.4142	0.0839
Scenario 4	1.0718	1.0718	0.0600
Scenario 5	1.0718	1.7411	0.0928
Scenario 6	1.0718	1.0718	0.0638
Scenario 7	1.0718	1.0718	0.0736
Scenario 8	1.0718	1.0718	0.0564
Scenario 9	1.0718	1.0718	0.0607

Comparison of the training results

The comparison of the training results for nine types of combinations with the remote sensing data is shown in Figure 6. The comparison of mean squared error and squared correlation coefficient is shown in Table 4.

The location is strongly correlated with chlorophyll-a concentration, and the location is a key influence factor which must be considered. At the same time, SST, SSC and SDD play a role in improving prediction accuracy, so SST, SSC and SDD are also key influence factors that affect the concentration of chlorophyll-a. Flow field and wind field also promote the improvement of the prediction accuracy.

Comparison of the prediction results

Predictions for 12 weeks starting from May 30, 2009 were conducted, and the comparative analyses were performed between the results and the remote sensing data. The comparison of mean square error and square correlation coefficient for the nine scenarios are shown in Tables 5 and 6, respectively. The comparison of spatial autocorrelation coefficient Moran's I between the simulation results and the remote sensing data is shown in Figure 7.

The average mean square error is the smallest and the precision is the highest in Scenario 7, as shown in Table 5. Meanwhile, the square correlation coefficient in Scenario 7 is the highest, as presented in Table 6.

Figure 7 shows that the spatial autocorrelation coefficient Moran's I of Scenario 7 is closest to the Moran's I

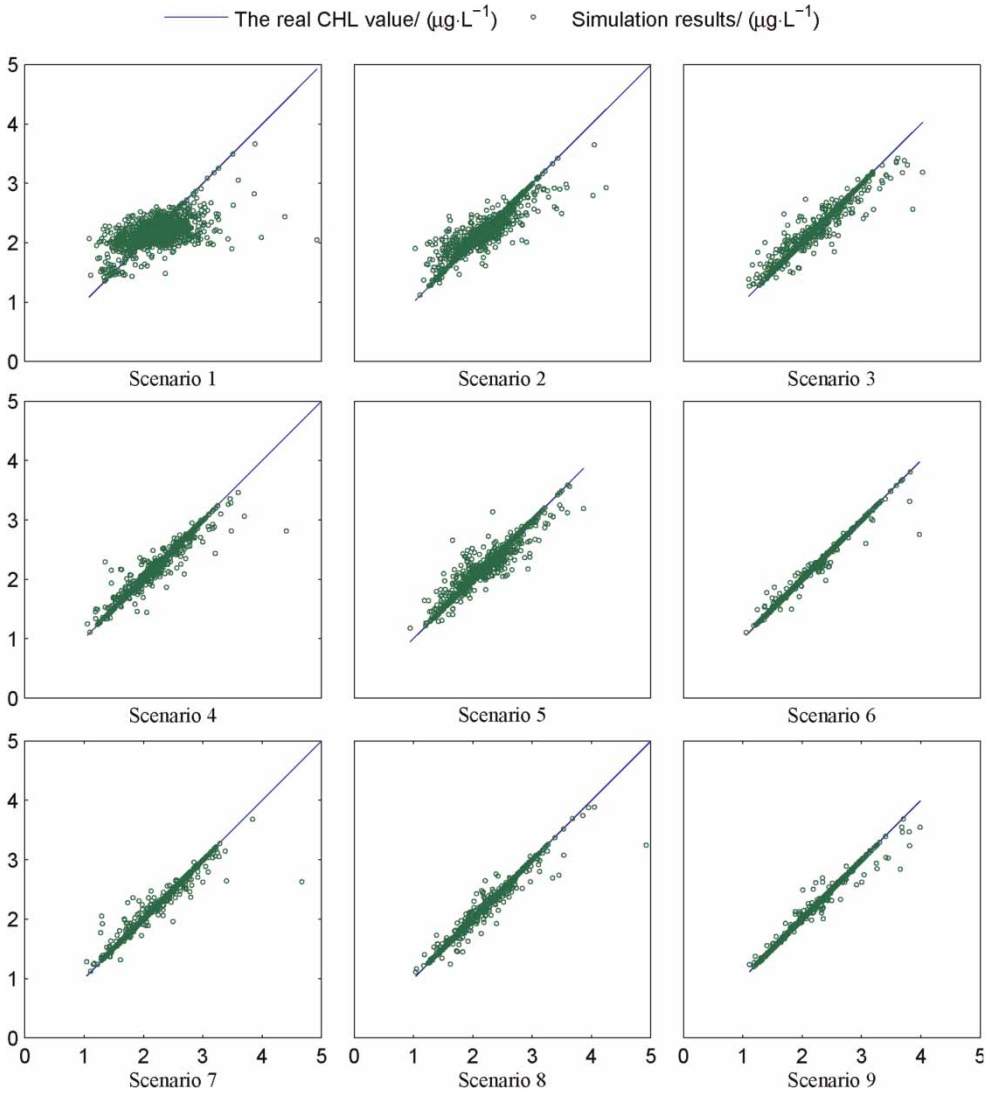


Figure 6 | The comparison of the training results of nine types of combinations with the remote sensing data.

value from the remote sensing. This indicates that the prediction result of Scenario 7 and remote sensing data had the best spatial correlation.

Comprehensively considering the training results and prediction results, the prediction error of Scenario 7 is smaller, the square correlation coefficient is the best and the prediction result of Scenario 7 and remote sensing data have the best spatial correlation. Scenario 7 was therefore deemed as the best combination of input parameters.

Above all, the combination in Scenario 7 was chosen as the best combination of input parameters. The corresponding model is called the basic model.

Parameter sensitivity analysis

The parameter sensitivity analysis was carried out in order to know the parameter influence on the prediction accuracy. Through changing the penalty parameter C and the kernel parameter γ, respectively (increase 10% or decrease 10%), the parameter sensitivity analysis was carried out. The mean square error comparison of each situation is shown in Table 7.

It could be seen that the mean square error sensitivity to the penalty parameter C is larger than the kernel parameter γ.

Table 4 | Analysis of the training results

	Mean squared error	Squared correlation coefficient
Scenario 1	0.1332	28.28
Scenario 2	0.0360	80.59
Scenario 3	0.0123	93.51
Scenario 4	0.0188	89.94
Scenario 5	0.0147	91.92
Scenario 6	0.0101	94.60
Scenario 7	0.0100	94.30
Scenario 8	0.0142	92.36
Scenario 9	0.0086	95.39

The comparison

In order to prove the coupled model has better prediction accuracy and generalisation property, the comparison of predictive results was carried out among radical basis function neural network, SVM and the cellular automata-support vector machine (CA-SVM) model using the best parameter combination Scenario 7 in the section The choice of the model parameters. The comparison of mean square error is shown in Table 8.

It was found in Table 8 that the CA-SVM coupled model has higher prediction accuracy and better generalisation property. Meanwhile, the comparison of predictive effect between the SVM and the BP neural network can be referenced in Xiang (2011), in which the SVM outperformed the BP neural network in the procedure of chlorophyll-a concentration prediction.

Optimisation of the eutrophication model

For the coupled model in this study, there are two main methods about the data information that could be used to improve the prediction accuracy: (1) choose the best influence factors combination; (2) extract more information from the available data. The discretisation method could be used to extract more information through improving the pertinence. The discretisation method includes the same width discretisation, the equal frequency discretisation, the clustering discretisation and the discretisation according to its special feature. In this study, Shelford's Law of Tolerance in ecology was chosen in order to do the discretisation.

Optimisation of the model

According to the three stages of SST and CHL separately, namely, the stage suitable for phytoplankton growth, the stage suitable for phytoplankton survival but not suitable for phytoplankton growth and the stage unsuitable for phytoplankton survival, the training data composed of the reasonable input parameters are divided into nine parts. Each part corresponds to a special stage of SST and CHL

Table 5 | Comparison of mean square error for the nine scenarios

	Sce. 1	Sce. 2	Sce. 3	Sce. 4	Sce. 5	Sce. 6	Sce. 7	Sce. 8	Sce. 9
Week 1	0.0553	0.0354	0.0313	0.0356	0.0340	0.0300	0.0341	0.0342	0.0292
Week 2	0.0589	0.0580	0.0316	0.0808	0.0418	0.0510	0.0266	0.0701	0.0472
Week 3	0.2293	0.1062	0.0897	0.1279	0.0750	0.1463	0.0533	0.1288	0.1443
Week 4	0.2128	0.1466	0.0950	0.0820	0.1179	0.0855	0.1185	0.0798	0.0851
Week 5	0.1026	0.0833	0.0687	0.1639	0.0692	0.1422	0.0641	0.1562	0.1417
Week 6	0.1958	0.2031	0.1497	0.4399	0.1392	0.3867	0.1115	0.4259	0.3829
Week 7	0.3361	0.1162	0.1000	0.0887	0.0924	0.0901	0.0893	0.0864	0.0879
Week 8	0.1376	0.1329	0.1689	0.2127	0.1370	0.1925	0.1378	0.2046	0.1899
Week 9	0.3632	0.4262	0.4223	0.5556	0.5662	0.5239	0.4880	0.5516	0.5259
Week 10	0.1420	0.0968	0.1001	0.1332	0.1121	0.1145	0.1029	0.1264	0.1112
Week 11	0.6385	0.5804	0.6179	0.4220	0.4750	0.4702	0.4961	0.4337	0.4766
Week 12	0.3171	0.3741	0.3664	0.2761	0.3469	0.2965	0.3302	0.2815	0.2976
Mean	0.2324	0.1966	0.1868	0.2182	0.1839	0.2108	0.1710	0.2149	0.2099

Table 6 | Comparison of square correlation coefficient for the nine scenarios

	Sce. 1	Sce. 2	Sce. 3	Sce. 4	Sce. 5	Sce. 6	Sce. 7	Sce. 8	Sce. 9
Week 1	11.88	40.14	47.04	38.57	39.38	43.22	39.73	40.15	44.23
Week 2	1.73	30.21	44.94	44.57	42.02	54.11	48.59	53.38	59.93
Week 3	0.73	32.31	33.80	0.19	38.19	0.05	52.82	0.23	0.35
Week 4	19.01	39.99	43.73	47.23	37.08	48.11	36.05	50.62	51.07
Week 5	26.95	39.22	45.77	3.08	47.19	11.72	49.02	7.69	12.32
Week 6	9.13	18.40	22.61	16.75	29.38	11.85	33.15	16.09	12.38
Week 7	3.89	13.84	7.25	5.84	9.00	5.28	4.28	6.88	5.65
Week 8	8.62	11.70	3.43	19.17	17.59	7.93	9.51	19.68	7.23
Week 9	14.40	10.47	6.39	5.15	1.04	4.90	3.34	3.67	3.25
Week 10	3.42	7.98	2.66	1.13	4.20	0.08	1.35	1.49	0.05
Week 11	16.18	15.22	3.77	18.59	11.39	8.23	8.47	13.74	6.57
Week 12	3.72	5.92	4.28	4.39	5.50	1.29	4.43	6.82	0.64
Mean	9.97	22.12	22.14	17.06	23.50	16.40	24.23	18.37	16.97

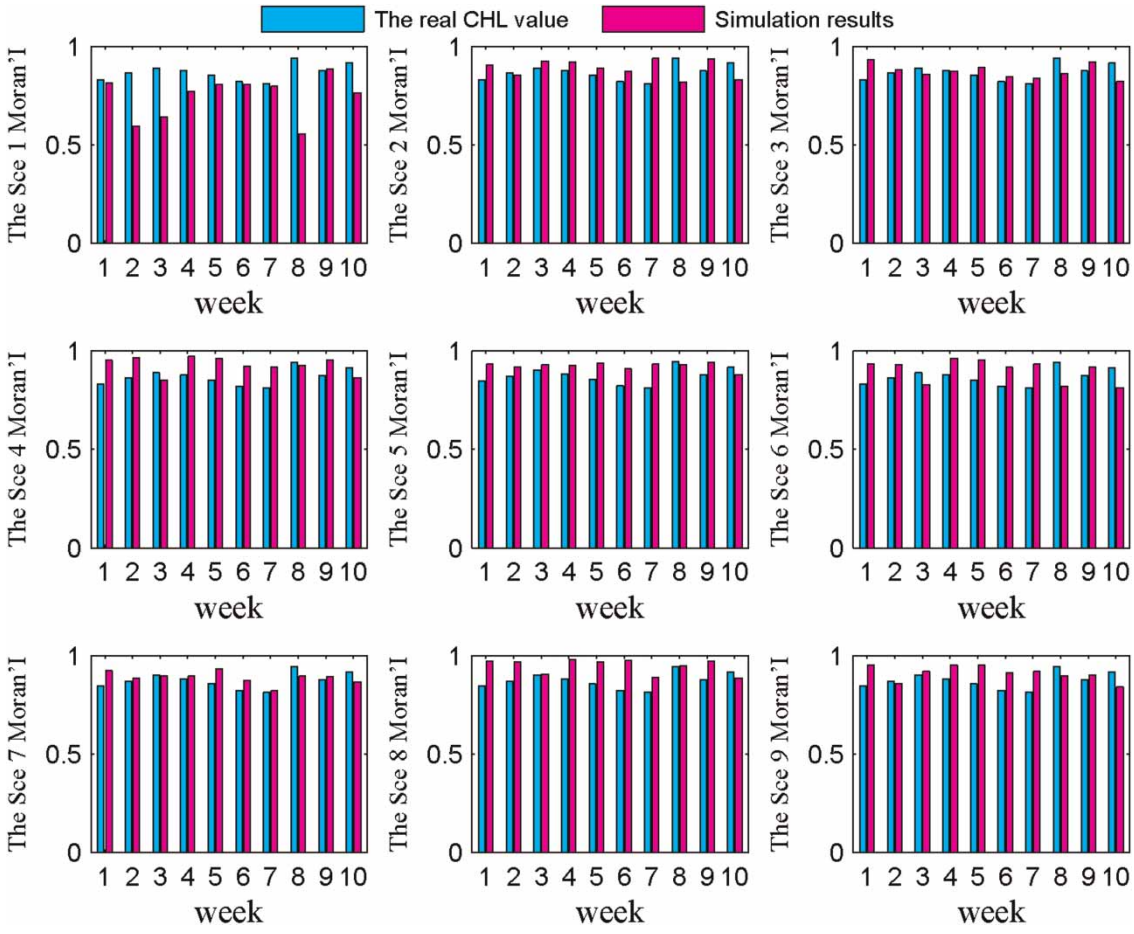


Figure 7 | Comparison of the spatial autocorrelation coefficient Moran's I between the simulation results and the remote sensing data.

Table 7 | The sensitivities of parameters

MSE	(C, γ)	(1.1*C, γ)	(0.9*C, γ)	(C, 1.1*γ)	(C, 0.9*γ)
1	0.0341	0.0397	0.0386	0.0322	0.0405
2	0.0266	0.0292	0.0315	0.0257	0.0359
3	0.0533	0.0633	0.0598	0.0599	0.0704
4	0.1185	0.1035	0.0984	0.1002	0.1161
5	0.0641	0.0728	0.0663	0.0653	0.0629
6	0.1115	0.1305	0.1368	0.1112	0.1907
7	0.0893	0.0892	0.1146	0.0858	0.1406
8	0.1378	0.1609	0.1810	0.1400	0.0962
9	0.4880	0.5062	0.5328	0.5185	0.4118
10	0.1029	0.1031	0.1233	0.0984	0.0923
11	0.4961	0.5369	0.5240	0.5258	0.4672
12	0.3302	0.3304	0.3480	0.3348	0.4068
Mean	0.1710	0.1805	0.1879	0.1748	0.1776

C is the penalty parameter and γ is the parameter of the SVM kernel function parameter.

Table 8 | Comparison of mean square error

	ANN	SVM	CA-SVM
Week 1	0.1962	0.0386	0.0341
Week 2	0.1725	0.0315	0.0266
Week 3	0.2361	0.0598	0.0533
Week 4	0.3291	0.0984	0.1185
Week 5	0.4045	0.0663	0.0641
Week 6	0.4457	0.1368	0.1115
Week 7	2.8957	0.1146	0.0893
Week 8	0.6476	0.1810	0.1378
Week 9	1.2153	0.5328	0.488
Week 10	2.6786	0.1233	0.1029
Week 11	1.1814	0.5240	0.4961
Week 12	3.0639	0.3480	0.3302
Mean	1.1222	0.1879	0.1710

(shown in Table 5). Through the separate training of the nine parts, the nine special models were developed, and each special model corresponds to a specific stage. The optimised model is established by the series of the nine special models. When the prediction was conducted, a suitable model was chosen according to the situation. The SST and CHL were finally discreted according to the same width and the real situation of the remote sensing data. The detail of discretisation can be seen in Table 9. More elaborated and more in-depth discretisation should be done in future research. Finally, the training data were divided into nine parts according to the three stages of SST and CHL separately, nine parts of data were trained respectively and nine models were produced.

The scope of each stage is shown in Table 9.

Comparison and analysis of the prediction results

Data covering a wide range were chosen to do the simulation to validate the model. The results of the basic model and the results of the optimised model were compared and analysed.

The comparison of the mean square error and square correlation coefficient is shown in Table 10; the comparison of

Table 9 | The scope of each stage

	SST(0 °C–9 °C)	SST(9 °C–18 °C)	SST(18 °C–)
CHL (0 – 2 mg/L)	Case_1_model	Case _2_model	Case _3_model
CHL (2 mg/L – 3 mg/L)	Case _4_model	Case _5_model	Case _6_model
CHL (3 mg/L –)	Case _7_model	Case _8_model	Case _9_model

Table 10 | The comparison of prediction result between the optimised model and the basic model

	Mean squared error		Squared correlation coefficient	
	Basic model	Optimised model	Basic model	Optimised model
Week 20	0.5104	0.0143	62.36	98.39
Week 29	0.1260	0.0246	5.39	59.27
Week 31	0.1736	0.0605	17.75	60.35
Week 34	0.1354	0.0599	15.07	64.61
Week 43	0.2307	0.0767	24.6	58.88
Week 49	0.3822	0.1601	30.59	93.43
Week 50	0.3205	0.1092	36.14	75.35
Week 53	0.0557	0.0183	28.51	65.15
Week 116	0.0399	0.0125	36.4	61.24
Week 119	0.0816	0.0384	46.69	70.22
Mean	0.1220	0.0369	30.35	70.69

the spatial autocorrelation coefficient Moran's I between the basic model and the optimised model is shown in Figure 8.

The prediction accuracy and square correlation coefficient of the optimised model were improved to a great extent based on the basic model as presented in Table 10.

The Moran's I of the optimised model is closer to the Moran's I of the remote sensing data than the basic model, as shown in Figure 8.

In brief, the prediction accuracy was improved by the optimised model, and the purpose of optimising the model was realised.

Error analysis and solutions

Although the prediction effect of the optimised model has been improved to a certain stage, the error between the prediction results and the remote sensing data still exists, and sometimes the error is large. The possible reasons are presented in the following:

(1) The set of influence factors may not be completely considered. For example, the change of the chlorophyll-a concentration has a very close relationship with the nutrients, but the related data might not be obtained due to various constraints.
(2) The data used in the model came from the remote sensing data. There might be some error in the interpretation of the remote sensing data.

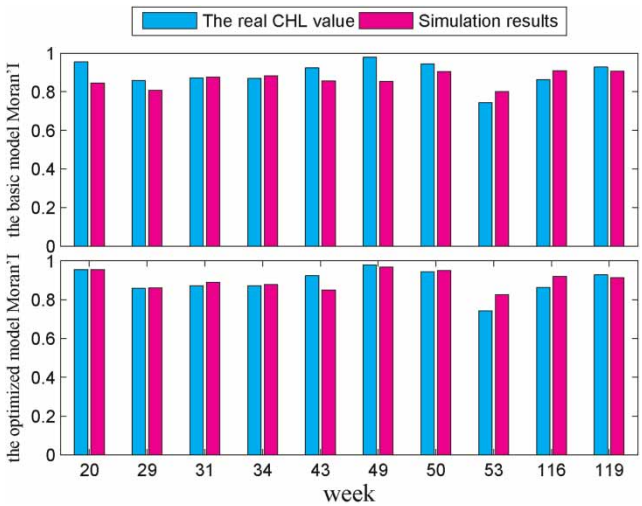


Figure 8 | The comparison of the spatial autocorrelation coefficient Moran's I between the basic model and the optimised model.

(3) A singular value point may exist in the remote sensing data, and the training result may be affected.
(4) Due to the difference of weather and air quality, there is much diversity in the process of data interpretation, and this diversity may lead to a proportionality coefficient existing between the remote sensing data at different times. If the time is different, and the proportionality coefficient is also different, the model is also unstable.
(5) The scope of the three stages of the key influence factors has not been determined. At present, the standards could only be preliminarily simplified, leading to the decline of the predictive accuracy of the model.

There are several steps which should be undertaken in future research to improve the model accuracy. First, obtain and use real-time test data as much as possible in the future to improve the accuracy. Second, perform research on the scope of the three stages to realise the purpose of more precise prediction.

CONCLUSIONS

In order to optimise the eutrophication model of Bohai Bay which is based on the coupling of the CA with the SVM, two objectives were achieved in this study. First, the more reasonable influence factors combination was chosen through the comparison of prediction results of nine sets of influence parameters combinations, and then the basic model was established. Second, the model was optimised further according to Shelford's Law of Tolerance (Shelford 1913) and its supplement (Odum 1983). Through the optimisation, the prediction accuracy was improved.

Nevertheless, there are still some limitations in this study. Some essential factors were not considered in the process of parameter choice for some reason, for example the nutrients. Moreover, the boundary of each stage is not clear up to now. Therefore, there are several suggested tasks that should be done in future work: (1) identify more reasonable influence factors (for example, nutrients) and add these factors to the model to realise a more accurate simulation; (2) perform research on the scope of the three stages of the key environmental factors to realise the objective of more precise prediction.

ACKNOWLEDGEMENTS

The financial support from National Science and Technology Support Program of China (No. 2010BAC68B04) and the open foundation of Key Laboratory of Marine Spill Oil Identification and Damage Assessment Technology, SOA (No. 201215) is greatly acknowledged. This research is supported partly by the National Nature Science Foundation of China (11002099). The authors are also grateful to National Satellite Ocean Application Service of China for providing the remote sensing data.

REFERENCES

Abbott, M. B. & Minns, A. W. 1998 *Computational Hydraulics*. 2nd en., Ashgate Publishing, Burlington, Vermont, USA.

Adamowski, J., Chan, H. F., Prasher, S. O. & Sharda, V. N. 2012 Comparison of multivariate adaptive regression splines with coupled wavelet transform artificial neural networks for runoff forecasting in Himalayan micro-watersheds with limited data. *J. Hydroinformat.* **14** (3), 731–744.

Amoroso, S. & Patt, Y. 1972 Decision procedures for surjectivity and injectivity of parallel maps for tessellation structures. *Comput. Syst. Sci.* **6**, 448–464.

An, F. & Tao, J. 2007 Method of two-grade fuzzy synthesis assessment for seawater eutrophication and its application in Bohai Bay. *Marine Environ. Sci.* **27** (4), 366–369.

Behzad, M., Asghari, K., Eazi, M. & Palhang, M. 2009 Generalization performance of support vector machines and neural networks in runoff modeling. *Exp. Syst. Applicat.* **36**, 7624–7629.

Boser, B., Guyon, I. & Vapnik, V. 1992 A training algorithm for optimal margin classifiers. In: *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*, 27–29 July, Pittsburgh, PA, pp. 144–152.

Chau, K. & Muttil, N. 2007 Data mining and multivariate statistical analysis for ecological system in coastal waters. *J. Hydroinformat.* **9** (4), 305–317.

Chau, K. W. 2004 Intelligent manipulation of calibration parameters in numerical m modeling. *Adv. Environ. Res.* **8**, 467–476.

Chau, K. W., Wu, C. L. & Li, Y. S. 2005 Comparison of several flood forecasting models in Yangtze river. *J. Hydrol. Eng.* **10**, 485–491.

Chen, Q. & Mynett, A. 2004 A robust fuzzy logic approach to modelling algal biomass. *J. Hydraul. Res.* **42** (3), 303–309.

Chen, Q. & Mynett, A. E. 2006 Hydroinformatics techniques in eco-environmental modelling and management. *J. Hydroinformat.* **8** (3), 297–316.

Chen, Q. & Ye, A. 2008 Unstructured cellular automata and the application to model river riparian vegetation dynamics. *Lect. Notes Comput. Sci.* **5191**, 337–344.

Chen, Q., Mynett, A. & Minns, A. 2002 Application of cellular automata to modelling competitive growths of two underwater species *Chara aspera* and *Potamogeton pectinatus* in Lake Veluwe. *Ecol. Modell.* **147**, 253–265.

Chen, Q., Mynett, A. & Wang, F. 2003 Integration of data mining techniques with heuristic knowledge in a fuzzy logic modelling of eutrophication in Taihu Lake. *Ecol. Modell.* **162**, 55–67.

Chen, Q., Ye, F. & Li, W. 2009 Cellular-automata-based ecological and ecohydraulics modeling. *J. Hydroinformat.* **11** (3–4), 252–265.

Cheng, C. T., Chau, K. W., Sun, Y. G. & Lin, J. Y. 2005 Long-term prediction of discharges in Manwan Reservoir using artificial neural network models. *Lect. Notes Comput. Sci.* **3498**, 1040–1045.

Cortes, C. & Vapnik, V. 1995 Support vector networks. *Mach. Learn.* **20** (3), 273–297.

David, A. & Ricard, V. 2000 The DivGame simulator: a stochastic cellular automata model of rainforest dynamics. *Ecol. Modell.* **133**, 131–141.

Deng, N. & Tian, Y. 2009 *The Support Vector Machine – Theory Algorithm and Development*. Science Press, Beijing.

Doniger, S., Hofman, T. & Yeh, J. 2002 Predicting CNS permeability of drug molecules: Comparison of neural network and support vector machine algorithms. *J. Computat. Biol.* **9** (6), 849–864.

Ellison, A. & Bedford, B. 1995 Response of a wetland vascular plant community to disturbance: a simulation study. *Ecol. Appl.* **5**, 109–123.

Evgeny, B., Uli, F., Jens, S. & Gisbert, S. 2003 Comparison of support vector machine and artificial neural network systems for drug/nondrug classification. *Artificial Neural Network Systems* **43** (6), 1882–1889.

Feng, J., Wang, H. & Li, S. 2007 Research on prediction of phytoplankton's density using support vector machines. *Marine Environ. Sci.* **26** (5), 438–441.

Francis, E. H., Tay & Lijuan Cao 2001 Application of support vector machines in financial time series forecasting. *Omega* **29** (4), 309–317.

Friedrich, R., Jason, B., Peter, W. & Hugh, W. 2002 Comparative application of artificial neural networks and genetic algorithms for multivariate time-series modelling of algal blooms in freshwater lakes. *J. Hydroinformat.* **4** (2), 125–133.

Guan, Y., Zhang, L. & Chen, C. 2009 Study of salt marsh vegetation spread dynamics model based on conditions optimized CA. *Geomatics and Information Science of Wuhan University* **34** (6), 701–705.

Islam, A. 2010 Improving flood forecasting in Bangladesh using an artificial neural network. *J. Hydroinformat.* **12** (3), 351–364.

Lee, J., Huang, Y., Dickman, M. & Jayawardena, A. 2003 Neural network modelling of coastal algal blooms. *Ecol. Modell.* **159**, 179–201.

Li, C., Zhang, F., Shen, X., Yang, B., Shen, Z. & Sun, S. 2005 Concentration, distribution and annual fluctuation of chlorophyll-a in the JiaoZhou Bay. *Oceanologia ET Limnologia Sinica* **36** (6), 499–506.

Liang, D., Zeckoski, R. W. & Wang, X. 2013 Development of a hydro-environmental model for inland navigational canals. *J. Hydroinformat.* doi:10.2166/hydro.2013.021.

Ma, G. 2007 Study on Eutrophication Model of BohaiBay Based on Chaos Optimization Neural Network. Master's Thesis, Tianjin University, Tianjin.

Muttil, N. & Chau, K. W. 2006 Neural network and genetic programming for modelling coastal algal blooms. *Int. J. Environ. Pollut.* **28**, 3–4.

Odum, E. P. 1983 *Basic Ecology*. CBS College Publishing, New York, pp. 223–224.

Riccardo, T., Chau, K. W. & Sethi, R. 2012 Artificial neural network simulation of hourly groundwater levels in a coastal aquifer system of the Venice lagoon. *Eng. Appl. Artif. Intell.* **25**, 1670–1676.

Salski, A. & Sperlbaum, C. 1991 Fuzzy logic approach to modelling in ecosystem research. *Lect. Notes Comput. Sci.* **521**, 520–527.

Samsudin, R., Shabri, A. & Saad, P. 2010 A comparison of time series forecasting using support vector machine and artificial neural network model. *J. Appl. Sci.* **10** (11), 950–958.

Shelford, V. E. 1913 Animal Communities in Temperate America. University of Chicago Press, Chicago, pp. 302–303.

Srinivas, M., Guadalupe, J. & Andrew, S. 2002 Intrusion detection using neural networks and support vector machines. *Neural Networks* **2**, 1702–1707.

Sun, J. 2007 The Relation Matrix Method and Its Application of Water Exchange Study for Sea Bay and Coastal Water. PhD Thesis, Tianjin University, Tianjin.

Sun, J. & Tao, J. 2006 Relation matrix of water exchange for sea Bays and its application. *China Ocean Eng.* **20** (4), 529–544.

Tao, J. & Xiang, X. 2009 Eutrophication predict modelling of Bohai Bay based on hydroinformatics technology. In: *The 8th International Conference on Hydroinformatics*, Concepción, Chile.

Vapnik, V. 1998 *Statistical Learning Theory*. Wiley, New York.

Wu, C. L., Chau, K. W. & Li, Y. S. 2009 Predicting monthly streamflow using data-driven models coupled 1 with data-preprocessing techniques. *Water Resour. Res.* **45** (8), w08432.

Xiang, X. 2011 Study on Characteristics and Modelling Methods for Aquatic Ecological Environment of Bohai Bay using Hydroinformatics Technology. PhD Thesis, Tianjin University, Tianjin.

Xiang, X., Xu, X. & Tao, J. 2013 Modelling chlorophyll-a in Bohai Bay based on hybrid soft computing approach. *J. Hydroinformat.* **15** (4), 1099–1108.

Yao, Z., Fei, M., Li, K., Kong, H. & Zhao, B. 2007 Recognition of blue-green algae in lakes using distributive genetic algorithm based neural networks. *Neurocompution* **70** (4–6), 641–647.

Yu, Y. 2011 Prediction of EutroPhieation by BP Neural Network Model Based on PSO Algorithm. Master's Thesis, Jinan University, Guangdong.

Zhu, W., Jiang, M., Zhao, L. & Tian, T. 2010 Field survey and analysis of influence of suspended sediment on algae growth. *Adv. Water Sci.* **21** (2), 241–247.

First received 21 August 2013; accepted in revised form 4 February 2014. Available online 11 March 2014