Modelling chlorophyll-*a* in Bohai Bay based on hybrid soft computing approach

Xiang Xianquan, Xu Xiaofu and Tao Jianhua

ABSTRACT

Algae bloom has become a serious problem of global concern. Scientists have managed to study it using various mathematical models with different degrees of complexity. However, these conventional modelling approaches are limited due to the complexity of the processes involved, the scarcity of data and spatial heterogeneity. In this study, hybrid soft computing (SC) algorithms, including support vector machine (SVM), genetic algorithms (GA) and cellular automata (CA) are illustrated and then employed to model chlorophyll-*a* (Chl-*a*) in Bohai Bay, China, with two models. In the first model, SVM tuned by GA (GA-SVM) is developed to model Chl-*a* for its greater capacity in dealing with nonlinear complex relationships. Then, in order to take into account the spatial heterogeneity and local interaction of the blooms, an integrated model of CA and SVM (CA-SVM) has been developed to model Chl-*a* in Bohai Bay using remote sensing data. Through this study, it can be observed that the hybrid SC approach is the prefered tool to predict the concentration of Chl-*a* in Bohai Bay, and capture the non-linear information in ecological processes. This work should be helpful in understanding marine ecological processes, protecting coastal aquatic ecosystem and marine management.

Key words | Bohai Bay, cellular automata, chlorophyll-*a*, genetic algorithm, hybrid soft computing, support vector machine

Xiang Xianquan (corresponding author)
The National Marine Data and Information Service,
Tianjin,
China 300171
E-mail: *xxquan@tju.edu.cn*

Xu Xiaofu
School of Environmental Science and Engineering,
Tianjin University,
Tianjin,
China 300072

Tao Jianhua
Department of Mechanics,
Tianjin University,
Tianjin,
China 300072

INTRODUCTION

The Bohai Bay is a semi-enclosed sea bay with mild-slope mud beach and shallow water, located at the west end of the Bohai Sea in north China. The water exchange and self-purifying capacity of this sea bay are very weak (Sun & Tao 2006). With the increasing discharge of nutrients and pollutants into the bay, the coastal water of Bohai Bay has encountered frequent red tides. It is very important to mitigate the impacts from such population outbreaks. Thus, investigation and modelling of algae population outbreaks is of great significance in providing important insights into the characteristics of the region and eco-environmental management.

Aquatic population growth involves a number of physical, chemical, ecological and biological processes. To describe the spatio-temporal evolution as in propagation, diffusion or transport phenomena, these complex systems have most often been modelled based on conservation of mass, Newton's second law of motion and the first and second laws of thermodynamics, and employ partial difference equations (PDEs) to describe physical processes (Abbott & Minns 1998). The variables of the models are continuous in time and space. Such model paradigms have played important roles in the progress of ecological research and are still fundamental tools (Jørgensen 1994).

However, modelling of algal bloom is rather difficult and still far from being accurate. On the one hand, limited availability of adequate spatial and temporal measurement data for water quality and ecological parameters makes the validation and calibration of any model very challenging. While on the other hand, the lack of process understanding makes it difficult to model using physically-based modelling tools, and the process-based mathematical

models mostly fail to take into account the effects of spatial heterogeneity and local interactions, which can be crucial to the ecosystem dynamics (Chen & Mynett 2004a, 2006a).

These problems present a great opportunity for the application of soft computing (SC) technologies. SC is a term originally coined by Zadeh (1994) to denote systems that '…exploit the tolerance for imprecision, uncertainty, and partial truth to achieve tractability, robustness, low solution cost, and better rapport with reality'. SC is an association of computing methodologies that includes as its principal members fuzzy logic (FL), neuro-computing (NC), evolutionary computing (EC), probabilistic computing (PC) and support vector machine (SVM). The main reason for the popularity of SC is the synergy derived from its components. SC's main characteristic is its intrinsic capability to create hybrid systems that are based on a (loose or tight) integration of constituent technologies. This integration provides complementary reasoning and searching methods that allow us to combine domain knowledge and empirical data to develop flexible computing tools and solve complex problems.

Over the past few years, we have seen an increasing number of hybrid algorithms, in which two or more SC technologies have been integrated to leverage the advantages of individual approaches. SC models have been widely used as a surrogate for the conceptual models, for their ability to capture the behaviour of the underlying physical or other processes. Chen & Mynett (2004b) successfully applied FL to model algal biomass. Muttil & Lee (2005) showed that genetic programming correctly demonstrated the relationship between the auto-regressive nature of bloom dynamics and flushing time. An artificial neural network (Lee *et al.* 2003) was proposed to predict the algal bloom dynamics of the coastal waters of Hong Kong. Tao & Xiang (2009) adopted a genetic algorithm (GA) to adjust internal parameters of ANFIS (adaptive neuro-fuzzy inference system) for establishing eutrophication prediction modelling of Bohai Bay.

Moreover, dynamic models of ecological communities should not neglect the spatial aspect, although most of them do so (Green 1989). The simplification of using a non-spatial model may lead to invalid forecasts, because spatial relationships can radically change conditions for persistence and coexistence (Czárán & Bartha 1992). Following the wide use of advanced survey technologies, it is possible to collect large scale and high resolution spatial data. Therefore, spatially explicit models have been rapidly developed and applied in broad disciplines (Perry & Enright 2007). Spatially explicit models are expected to increase our ability to accurately model populations subject to complex processes, to create basic ecological knowledge and to improve the application of landscape ecology to conservation and land management problems (Dunning *et al.* 1995; Kareiva & Wennergren 1995). Cellular automata (CA) are spatially explicit dynamic systems, and are discrete in space and time to integrate local information to elucidate complex dynamics in a comprehensive manner without losing computational coherence in the systems. Abbott & Minns (1998) pointed out CA could be similar to the Eulerian approach in classical fluid mechanics. According to Chen *et al.* (2009), in ecological modelling, CA seem to be more advantageous than PDEs due to the inherent characteristics of ecosystems.

After application of the CA in ecological modelling by Jeltsch & Wissel (1994) for modelling forest ecosystems, the CA have been additionally applied in ecological sciences dealing with spatial and temporal dynamics: landscape dynamics (Darwen & Green 1996; Li & Archer 1997; Fall & Fall 2001; Ostendorf *et al.* 2001), population dynamics of the plant species (Balzter *et al.* 1998), prey–predator population dynamics (Qu *et al.* 2008), prediction of forest fire (Karafyllidis & Thanailakis 1997; Berjak & Hearne 2002; Favier *et al.* 2004), modelling forest insect infestations (Bone *et al.* 2006), nutrient dynamics (Ito 2005), and prediction of fish stock (Dreyfus-León 2006).

However, to apply correctly the rules of CA requires searching a potentially massive parameter space. Data-driven technology can provide a fast solution for unclear phenomena, which is nowadays more popular because of its capability of knowledge discovery. A novel methodology of SC, SVM (Vapnik 2000) has proved to be able to avoid trapping in a local minimum and overfitting problems. Liu *et al.* (2009) demonstrated that compared with artificial neural network (ANN) and back-propagation-ANN (BP-ANN), SVM can generalize better to unseen data and make the running time considerably shorter with the same or higher accuracy. Yang *et al.* (2008) tested the SVM, which is good at dealing with nonlinear complex relationships, as a method for constructing nonlinear transition rules for CA.

Therefore, this paper studies the hybrid SC algorithms and demonstrates their application in modelling the concentration of chlorophyll-*a* (Chl-*a*) of Bohai Bay by two models: a model of the concentration of Chl-*a* in Bohai Bay based on SVM tuned by GA (GA-SVM) (Xiang & Tao 2010) using *in situ* data, and a model of the concentration of Chl-*a* in Bohai Bay based on an integrated CA and SVM (CA-SVM) model using remote sensing data.

It can be observed that the hybrid SC approach can be preferably used to predict the concentration of Chl-*a* in Bohai Bay, and capture the non-linear information in ecological processes. This work has great significance for understanding marine ecological processes, protecting coastal aquatic ecosystem and marine management.

MODEL OF THE CHLOROPHYLL-*a* BASED ON GA-SVM USING *IN SITU* DATA

Study area and data sets

A regular monitoring area in Bohai Bay was selected as the study area. During June to October in both 2006 and 2007, water quality of the monitoring area was surveyed twice a

month. Physical parameters such as water temperature, Secchi-depth and salinity were measured. Chemical analyses of water samples in the laboratory included chemical oxygen demand (COD), dissolved oxygen (DO), Chl-*a*, total inorganic nitrogen (TIN), total inorganic phosphorus (TIP) and silicate.

In this model, the monitoring data at No.5 station are identified as the verification data, because No.5 station is located relatively at the centre of the monitoring area, which can be considered to be representative of the entire area. The remaining stations numbered 1, 2, 3, 4 and 6 are selected for model construction (see Figure 1).

Methodology

GA-SVM model

SVM, which is a novel machine learning methodology based on statistical learning theory, VC (Vapnik–Chervonenkis) dimension theory and structural risk minimization (SRM), has been extensively approved because it can conquer the over-fitting problem effectively, decrease computation amount and avoid 'dimension disaster' (Xiang *et al.* 2010). SVM algorithms can be used for solving various data mining problems such as classification, regression
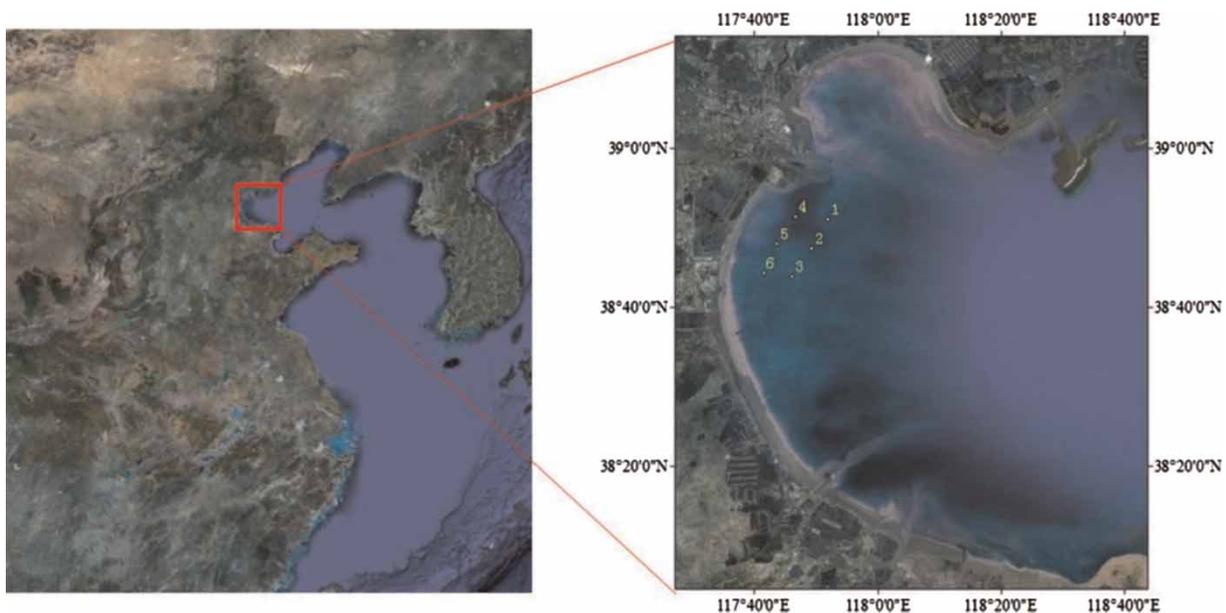

Figure 1 │ Location of Bohai Bay and the six stations.

estimation and novelty detection. In this section, we introduce some basic ideas behind SVM for regression.

Given a set of data points:

$$\{x_1, y_1\}, \ldots \{x_m, y_m\} \subseteq \{X \subseteq R^n \times Y \subseteq R\} \tag{1}$$

A loss function is required to measure the errors between the estimated values and the actual values. In this algorithm, a specific loss function called ε-insensitive loss is developed to create a sparseness property for SVR (support vector regression). It is defined by:

$$L_\varepsilon(y, f(x)) = \max(0, |y - f(x)| - \varepsilon) \tag{2}$$

where $f(x)$, computed by the SVR, is the estimated value of the y. The corresponding errors being less than ε boundary (ε-tube) are not penalized. Given a linear decision function which has the following form:

$$f(x) = \omega \cdot x + b \tag{3}$$

SVM estimates the function by minimizing the regularized risk function:

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

The first term $||\omega||^2/2$ is called the regularized term. Minimizing it will make a function as flat as possible. The second term is the empirical error which is measured by the ε-insensitive loss function. Coefficient C (a penalty parameter) is referred to as the regularization constant which plays a trading-off between the regularized term and the empirical error.

By using positive margin slack variables $\xi_i^{(*)}$ (() denotes variables with and without *), Equation (4) is equivalent to:

$$\min \frac{1}{2}||\omega||^2 + C \cdot \sum_{i=1}^{m} (\xi_i + \xi_i^*)$$
$$\text{subject to } ((\omega \cdot x_i) + b) - y_i \leq \varepsilon + \xi_i, \tag{5}$$
$$y_i - ((\omega \cdot x_i) + b) \leq \varepsilon + \xi_i^*, \xi_i^{(*)} \geq 0.$$

Equation (5) is called the primal objective function. It is solved by using the method of Lagrange multipliers. The

following dual optimization problem is obtained:

$$\min \frac{1}{2}\sum_{i,j=1}^{m} (\alpha_i^* - \alpha_i)(\alpha_j^* - \alpha_j)Q_{ij} + \varepsilon \sum_{i=1}^{m} (\alpha_i^* + \alpha_i)$$
$$- \sum_{i=1}^{m} y_i(\alpha_i^* - \alpha_i) \tag{6}$$
$$\text{subject to } \sum_{i=1}^{m} (\alpha_i - \alpha_i^*) = 0, \alpha_i, \alpha_i^* \in \left[0, \frac{C}{m}\right].$$

where $Q_{ij} = (x_i \cdot x_j)$. Finally by solving the above optimization problem and deriving the coefficients α^* and α, the estimation function is calculated by:

$$f(x) = \sum_{i=1}^{m} (\alpha_i^* - \alpha_i)(x_i, x) + b \tag{7}$$

In general, complex real-world applications like ecological problems require more expressive hypothesis spaces than linear functions. Thus, the use of the Kernel function K is motivated by exploring the inner product expression in both Equations (5) and (6). A kernel function can be illustrated as following:

$$K(x_i, x_j) = \phi(x_i)^T \cdot \phi(x_j) \tag{8}$$

The functions which satisfy the Mercer theory can be exploited as a Kernel function. For example, linear Kernel function, polynomial function, sigmoid function and radial basis function (RBF) are some common Kernels.

The RBF function is calculated by:

$$K(x, y) = \exp(-\gamma||x - y||^2) \tag{9}$$

where γ is a constant user-defined factor. In the RBF kernel, only one variable needs to be determined. In addition, SVM constructed from the RBF kernel has an excellent nonlinear forecasting performance. Therefore, the RBF kernel was used in SVM in this model. There are two parameters while using RBF kernels: C and γ. It is not known beforehand which C and γ are the best for one problem.

The cross-validation procedure is followed, which can prevent the overfitting problem. In k-fold cross-validation,

firstly the training data are separated to k folds. Sequentially one fold is considered as the validation set and the rest are for training. The average of error on predicting the validation sets is the cross validation error. The parameter k is traditionally set to 3 in this research. Consequently, some kind of model selection (parameter search) like trial and error must be done. LIBSVM (a library for SVMs) provides a parameter selection tool: cross validation via parallel grid search (Chang & Lin 2001).

Another tool, GA-SVM is developed in this study. GA is one of the best optimization strategies, which possesses global superiority, parallelity, rapidity and robusticity. In this program, GA was employed to select the optimal parameters of SVM, i.e. C and γ. We set population size $= 40$, generations $= 200$, generation gap $= 0.9$, probability of mutation $= 0.1$ and crossover probability $= 0.9$.

Definition of model input and output

As a comprehensive indictor to reflect phytoplankton biomass in water, Chl-*a* is one of the most important indicators of eutrophication. Therefore, Chl-*a* concentration is adopted as the representative token of eutrophication and selected as the model output. Principal component analysis (PCA), which defines new attributes (i.e. principal components) as mutually-orthogonal linear combinations of the original attributes, was identified model inputs. Using PCA analysis, T, COD, DO, Chl-*a*, TIN, TIP and Si were selected as the model inputs.

It is very important to scale the field measured data before applying SVM. The main advantage is to avoid the attributes in greater numeric ranges which dominate those in smaller numeric ranges. Another advantage is to avoid numerical difficulties during the calculation. Each attribute is linearly scaled to the range [0,1] in this model.

Results and discussion

To assess the performance of the optimized SVM, the correlation between observed and predicted Chl-*a* concentration was evaluated by using the squared correlation coefficient R^2 and the root mean square error (RMSE) in the same units but not the variables to quantify the error. The optimizing process of GA was illustrated in Figure 2. As can be seen, after 80
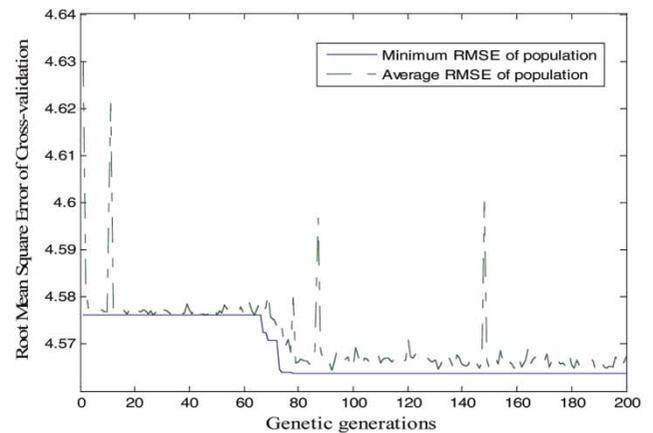

Figure 2 │ Change of minimum RMSE and average RMSE of each population.

generations, the minimum RMSE of each population changes slightly, which suggests that the best performance is achieved.

In Table 1, the optimal parameters C and γ, RMSE and R^2 for LIBSVM and GA-SVM models are listed, including the results of using default value of parameters. It can be seen that the optimized SVM models show better performance, which implies that parameter optimization is very important and necessary for SVM. In addition, the GA-SVM produces slightly better results than LIBSVM.

The independent data (i.e. data from No.5 station) were used to validate the models, and are not included in the training process. The validation results of the optimized SVM models are presented in Table 2. The validation results demonstrate that both optimized models are acceptable for predicting the concentration of Chl-*a* and the GA-SVM model obtains higher accuracy and generalization

Table 1 │ Results of SVM models on the modelling data

Pattern	*C*	γ	RMSE	R^2 (%)
SVM	1	0.1	4.4667	15.028
LIBSVM	8.069	35.898	0.8035	97.495
GA-SVM	3.2249	52.0815	0.6723	98.337

Table 2 │ Validation results of the optimized SVM models

Pattern	RMSE	R^2 (%)
LIBSVM	6.960	67.799
GA-SVM	6.305	85.840

performance than LIBSVM, which suggests that GAs possess better searching optimization ability.

MODEL OF THE CHLOROPHYLL-*a* IN BOHAI BAY BASED ON CA-SVM USING REMOTE SENSING DATA

Study area and remote sensing data sets

Satellite remote sensing data derived from the first Chinese ocean colour satellite HY-1 are adopted. Main observation factors of HY-1 are optical characters of ocean water: concentration of Chl-*a*, Secchi disk depth (SDD), suspended sediment concentration (SSC), sea surface temperature (SST), and so on. With the remote sensing interpretation, concentration of Chl-*a* could be available with a $0.01°$ latitude by $0.01°$ longitude grid spatial resolution (about $1,100 \times 1,100$ m). Using Arcmap function modules developed by ArcGIS software of Environmental Systems Research Institute (ESRI), these attribute data of Bohai Bay are extracted from the HDF files and planar vector file with (*.shp) format are generated. Thus, 11,106 grids covered Bohai Bay are generated (see Figure 3).

Methodology

Cellular automata

CA are dynamic models, discrete in space, time and state. They consist of a regular lattice of cells which interact with their neighbours. The cell states are synchronously updated in time according to local rules, which calculate the new state of a cell at time $t + 1$ using its state and those of neighbouring cells at time t.

A simple 'elementary' cellular automaton A is defined by a lattice L, a state space Q, a neighbourhood template δ and a local transition function f (Adamatzky 1994):

$$A = <L, Q, \delta, f> \tag{10}$$

where L is a lattice, based on an uniform discretization (several identical cells) of the spatial domain Q; Q is a finite set, containing values characterizing all the possible states that each cell can take; δ is a neighbourhood characterized by the neighbouring cells which have an influence on the cell. Several types of neighbourhoods are possible, but in two dimensional lattice, the most usual are Von Neumann containing the central cell and four adjacent cells and Moore neighbourhoods consisting of the central cell and eight adjacent cells; f is the transition rule, making it possible to define the state change from state at time t, to state at time $t + 1$ for each cell.

CA-SVM model setup

An integrated model of CA and SVM is developed to model the Chl-*a* of Bohai Bay using remote sensing data. The RBF kernel was selected in SVM in the CA-SVM model. There are two parameters while using RBF kernels: C and γ.
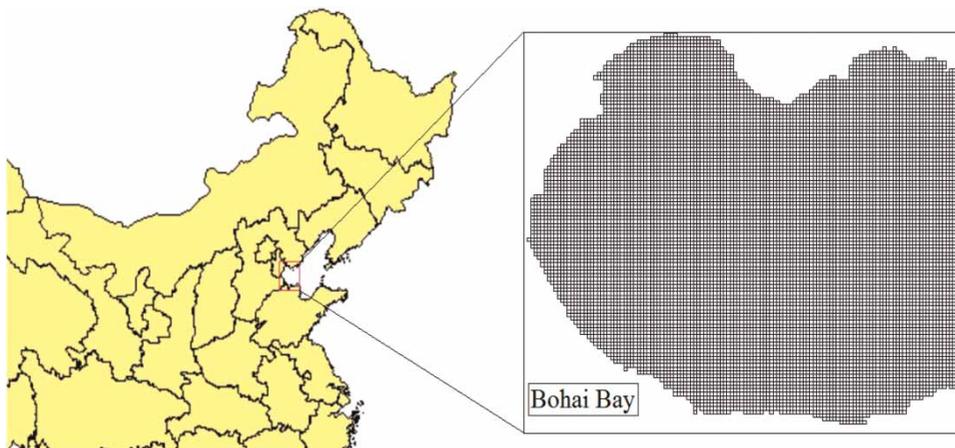

Figure 3 │ Location of Bohai Bay and its grids.

Pairs of (C and γ) are tried and the one with the best cross-validation accuracy is chosen.

The cell size is identified as the grid spatial resolution (about $1{,}100\,\text{m} \times 1{,}100\,\text{m}$). One of the main problems with CA concerns boundary conditions, generally including periodic boundary, reflective boundary, and constant boundary. The boundary conditions of Bohai Bay are set up as shown in Figure 4.

The yellow vertical line represents the outside sea, and the green line represents the boundary part of Bohai Bay (the full colour version of Figure 4 is available online at http://www.iwaponline.com/jh/toc.htm). The boundary connecting Bohai Bay with the outside sea is defined as the open boundary and the boundary connecting Bohai Bay with the land part is defined as the solid boundary, which means the state of land part value is identified as 0. The cell configuration adopts Moore neighbourhoods consisting of the central cell and eight adjacent cells, and the simulation time step of the CA model is 1 week.

The proposed model is a synchronous two-dimensional CA where data belonging to each grid-cell are the basis for the state of a cell in the cellular lattice. More precisely, each cell's state can be split into three parts. The first one is the concentration of Chl-*a*, which will be typical of a CA in that it will be recalculated at each time step according to the current rules. The second denotes the space location: latitude and longitude; both remain constant for the duration of each simulation run. The third is SDD, SSC and SST, which are regarded to be external forcing,
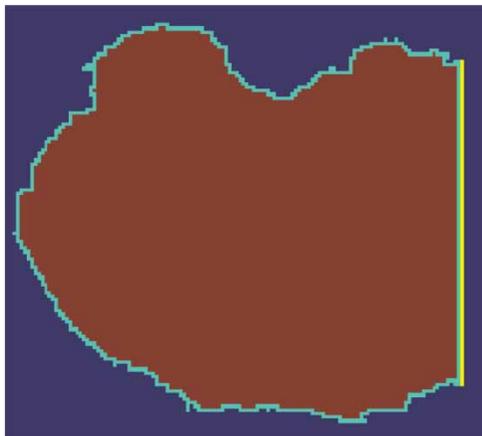

Figure 4 │ Boundary conditions of Bohai Bay of the CA-SVM model. The full colour version of this figure is available online at http://www.iwaponline.com/jh/toc.htm.

Table 3 │ Cases of different SVM inputs

	Inputs
Case 1	Chl-*a*, 8-neighbouring cell Chl-*a*
Case 2	Chl-*a*, 8-neighbouring cell Chl-*a*, Lat, Lon
Case 3	Chl-*a*, 8-neighbouring cell Chl-*a*, Lat, Lon, SDD, SSC, SST

input from remote data as known condition of each simulation run.

SVM is used to extract the CA transition rule by training the remote sensing data. Three cases of different SVM inputs are set up shown in Table 3. Lat and Lon are latitude and longitude, respectively, which reflect the effect of the spatial position of the cell.

Results and discussion

The time span of the CA-SVM model covered 14 weeks. The initial conditions were configured through the remote sensing data of May 30, 2009. The data of 10 weeks before were selected for model construction. SVM was used to extract the CA transition rule for three cases in Table 1. The established CA-SVM model was applied on the whole 14 weeks.

The results of model for three cases were evaluated by calculating Kappa Index of Agreement (KIA), which measures the degree of agreement on a scale from zero to one. If the two images are completely different, the KIA equals 1. The KIA of model results in different cases is shown in Figure 5. The comparison of Chl-*a* distribution of model results in three cases and remote sensing of some weeks is given in Figure 6.

The two figures show that the performance of case 3 is the best as a whole, and the performance of case 2 is better than that of case 1. The results also reflect that the spatial position, and external factors such as SDD, SSC, SST have a great effect on the distribution of Chl-*a*.

CONCLUSIONS

In this paper, the hybrid SC algorithms are studied with their application in modelling the concentration of Chl-*a* in Bohai Bay. In GA-SVM model, SVM tuned by GA was developed to model the concentration of Chl-*a* in Bohai
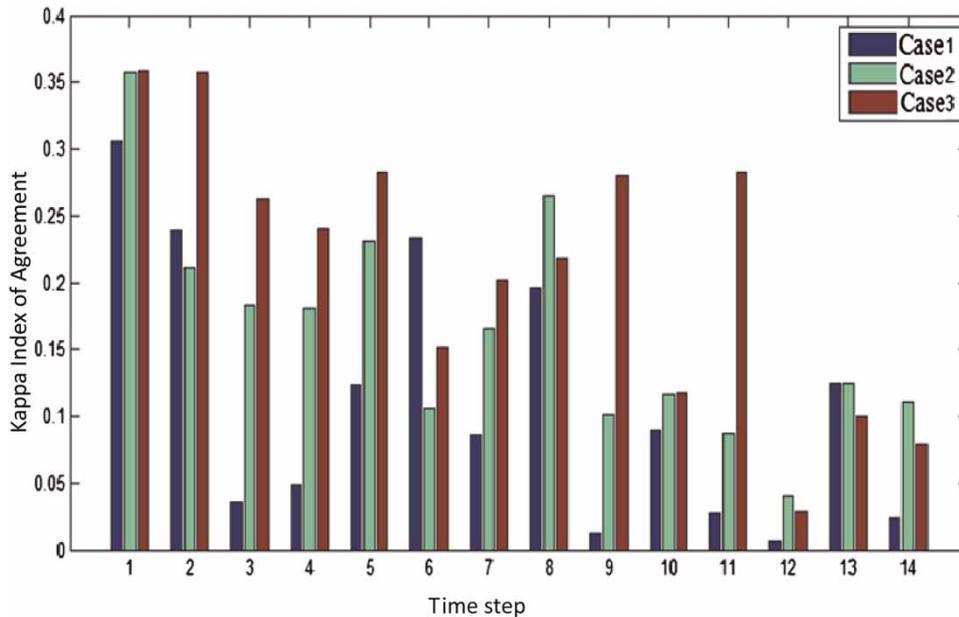
Figure 5 │ KIA of model results in different cases.

Bay using *in situ* data. The results from the simulating and verifying reveal that the GA-SVM model shows better performance than the SVM and LIBSVM model, which suggests that GAs possess better searching optimization ability. In the CA-SVM model, an integrated model of CA and SVM has been developed to model the concentration of Chl-*a* in Bohai Bay using remote sensing data. Through this study of three cases, it is shown that the spatial position and external factors such as SDD, SSC and SST have a great effect on the distribution of chlorophyll-*a*, and the proposed CA-SVM model can be preferable to simulate the spatial heterogeneity, local interaction and non-linear information in ecological processes.

Although the hybrid SC algorithms show great performance of modelling the non-linear information in ecological processes, the issue is still far from being resolved. It is greatly constrained by the availability of data.

In general, aquatic ecosystems are characterized by high dimensionality, complexity and non-linearity. Hybrid SC techniques have been being applied more and more widely in ecological modelling, and some advantages of these techniques have been illustrated in this paper. However, this does not mean that such an approach can take over conventional methods that are mainly characterized by physically-based formulations. An integrated numerical

and fuzzy CA model was developed to predict possible algal blooms in Dutch coastal waters basing on the irradiance, nutrients and neighbourhood conditions (Chen & Mynett 2006b). The key issue nowadays is to select proper methods from a variety of available tools according to the understanding of the problem at hand and the availability of data. The future of ecological modelling then lies in the integration of different paradigms and techniques at multiple and plausible spatio-temporal scales and actual processes.
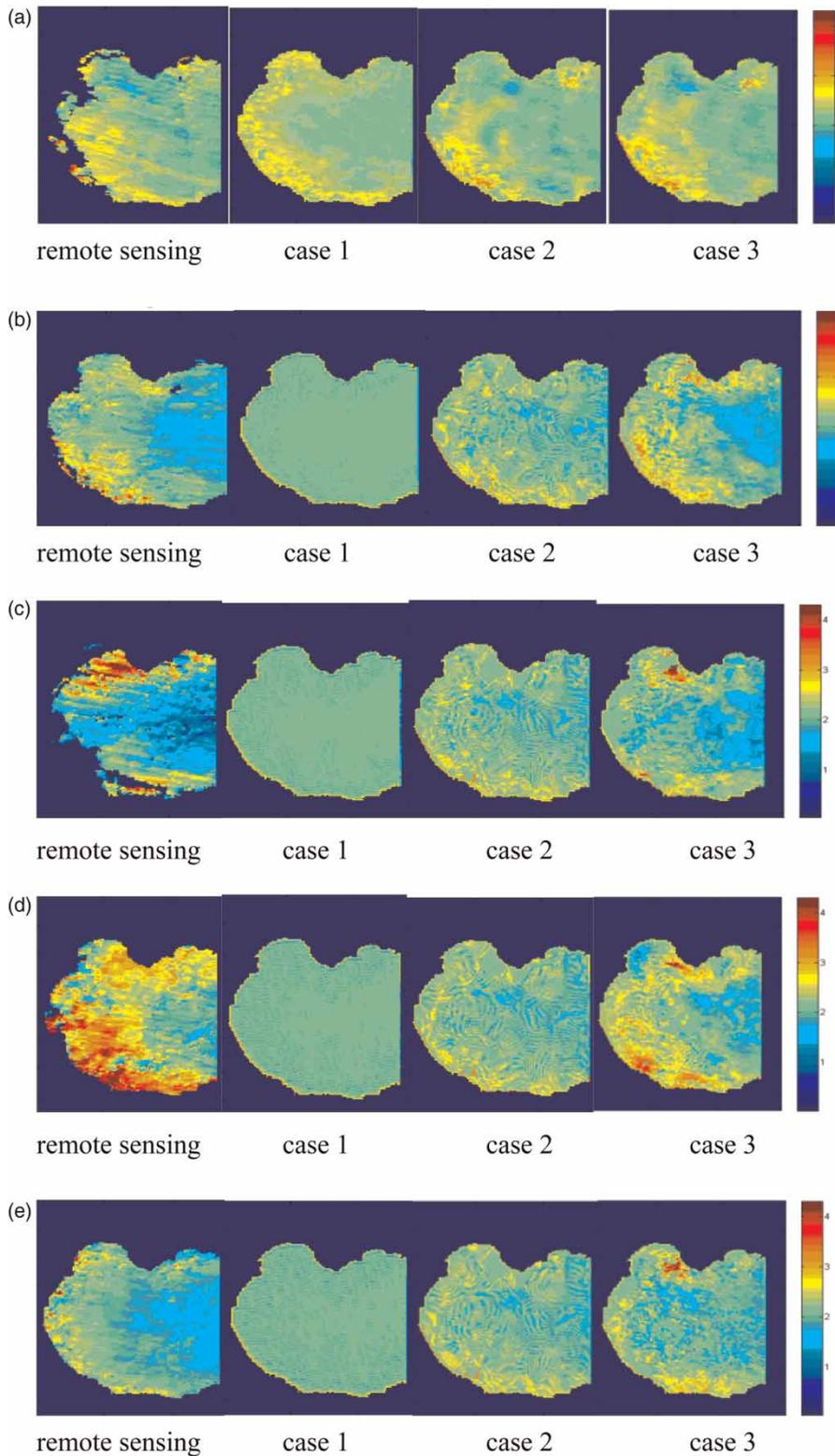
ACKNOWLEDGEMENTS

Figure 6 │ Comparison of Chl-*a* distribution of model results and remote sensing (μg L^{-1}): (a) results of week 1; (b) results of week 5; (c) results of week 9; (d) results of week 11; (e) results of week 13.

REFERENCES

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

Adamatzky, A. 1994 *Identification of Cellular Automata*. Taylor and Francis, London.

Balzter, H., Braun, P. & Köhler, W. 1998 Cellular automata models for vegetation dynamics. *Ecol. Model.* **107**, 113–125.

Berjak, S. G. & Hearne, J. W. 2002 An improved cellular automaton model for simulating fire in a spatially heterogeneous savanna system. *Ecol. Model.* **148**, 133–151.

Bone, C., Dragicevic, S. & Roberts, A. 2006 A fuzzy-constrained cellular automata model of forest insect infestations. *Ecol. Model.* **192**, 107–125.

Chang, C. & Lin, C. 2001 LIBSVM: A Library for Support Vector Machines. Software available at <http://www.csie.ntu.edu.tw/~cjlin>.

Chen, Q. & Mynett, A. E. 2004a Comparison of spatial patterns between cellular automata model simulations and remote sensing observations. *IAHS Red Book* **296**, 363–370.

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

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

Chen, Q. & Mynett, A. E. 2006b Modelling algal blooms in the Dutch coastal waters by integrated numerical and fuzzy cellular automata approaches. *Ecol. Model.* **199**, 73–81.

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

Czárán, T. & Bartha, S. 1992 Spatio-temporal dynamic models of plant populations and communities. *TREE* **7**, 38–42.

Darwen, P. J. & Green, D. G. 1996 Viability of populations in a landscape. *Ecol. Model.* **85**, 165–171.

Dreyfus-León, M. 2006 Modelling cooperation between fishermen with a cellular automaton. A framework for fishing effort spatial dynamics. *Ecol. Inform.* **1**, 101–105.

Dunning, J. B., Stewart, D. J., Danielson, B. J., Noon, B. R., Root, T. L., Lamberson, R. H. & Stevens, E. E. 1995 Spatially explicit population models: current forms and future uses. *Ecol. Appl.* **5**, 3–11.

Favier, C., Chave, J., Fabing, A., Schwartz, D. & Dubois, M. A. 2004 Modelling forest–savanna mosaic dynamics in man-influenced environments: effects of fire, climate and soil heterogeneity. *Ecol. Model.* **171**, 85–102.

Fall, A. & Fall, J. 2001 A domain-specific language for models of landscape dynamics. *Ecol. Model.* **137**, 1–21.

Green, D. G. 1989 Simulated effects of fire, dispersal and spatial pattern on competition within forest mosaics. *Vegetatio* **82**, 139–153.

Ito, A. 2005 Modelling of carbon cycle and fire regime in an east Siberian larch forest. *Ecol. Model.* **187**, 121–139.

Jeltsch, F. & Wissel, C. 1994 Modelling dieback phenomena in natural forests. *Ecol. Model.* **75–76**, 111–121.

Jørgensen, S. E. 1994 *Fundamentals of Ecological Modelling*, 2nd edition. Elsevier, Amsterdam.

Karafyllidis, I. & Thanailakis, A. 1997 A model for predicting forest fire spreading using cellular automata. *Ecol. Model.* **99**, 87–97.

Kareiva, P. & Wennergren, U. 1995 Connecting landscape patterns to ecosystem and population processes. *Nature* **373**, 299–302.

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

Li, B. L. & Archer, S. 1997 Weighted mean patch size: a robust index for quantifying landscape structure. *Ecol. Model.* **102**, 353–361.

Liu, J., Zhang, Y. & Qian, X. 2009 Modeling chlorophyll-*a* in Taihu Lake with machine learning models. *2009 3rd International Conference on Bioinformatics and Biomedical Engineering*, Beijing, China, pp. 1–6.

Muttil, N. & Lee, J. H. W. 2005 Genetic programming for analysis and real-time prediction of coastal algal blooms. *Ecol. Model.* **189**, 363–376.

Ostendorf, B., Hilbert, D. W. & Hopkins, M. S. 2001 The effect of climate change on tropical rainforest vegetation pattern. *Ecol. Model.* **145**, 211–224.

Perry, G. L. W. & Enright, N. J. 2007 Contrasting outcomes of spatially implicit and spatially explicit models of vegetation dynamics in a forest-shrub land mosaic. *Ecol. Model.* **207**, 327–338.

Qu, S., Chen, Q. & Recknagel, F. 2008 Cellular automata based simulation of random versus selective harvesting strategies in predator–prey systems. *Ecol. Inform.* **3**, 252–258.

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

Tao, J. & Xiang, X. 2009 Eutrophication Predict Modelling of Bohai Bay Based on Hydroinformatics Technology, *The 8th International Conference on Hydroinformatics*, Concepción, Chile.

Vapnik, V. N. 2000 *The Nature of Statistical Learning Theory*, 2nd ed. Springer-Verlag, New York.

Xiang, X. & Tao, J. 2010 Study on eutrophication model of Bohai Bay based on hybrid soft computing approach. *The 9th International Conference on Hydroinformatics*, Tianjin, China.

Xiang, X., Yuan, D. & Tao, J. 2010 Eutrophication prediction model of Bohai Bay based on optimized support vector machine. *4th International Conference on Bioinformatics and Biomedical Engineering*, Chengdu, China.

Yang, Q., Li, X. & Shi, X. 2008 Cellular automata for simulating land use changes based on support vector machines. *Comput. Geosci.* **34**, 592–602.

Zadeh, L. A. 1994 Fuzzy logic and soft computing: Issues, contentions and perspectives. In *Proceedings of IIZUKA'94: 3rd Int. Conf. Fuzzy Logic, Neural Nets and Soft Computing*, Iizuka, Japan, pp. 1–2.