



Contents lists available at ScienceDirect

## Marine Pollution Bulletin

journal homepage: [www.elsevier.com/locate/marpolbul](http://www.elsevier.com/locate/marpolbul)

# Real-time eutrophication status evaluation of coastal waters using support vector machine with grid search algorithm

Xianyu Kong, Yuyan Sun, Rongguo Su\*, Xiaoyong Shi

Key Laboratory of Marine Chemistry Theory and Technology, Ministry of Education Ocean University of China, Qingdao 266100, China

## ARTICLE INFO

### Keywords:

Eutrophication assessment  
Easily measured parameters  
CDOM  
Fluorescence  
TRIX  
Support vector machine

## ABSTRACT

The development of techniques for real-time monitoring of the eutrophication status of coastal waters is of great importance for realizing potential cost savings in coastal monitoring programs and providing timely advice for marine health management. In this study, a GS optimized SVM was proposed to model relationships between 6 easily measured parameters (DO, Chl-*a*, C1, C2, C3 and C4) and the TRIX index for rapidly assessing marine eutrophication states of coastal waters. The good predictive performance of the developed method was indicated by the  $R^2$  between the measured and predicted values (0.92 for the training dataset and 0.91 for the validation dataset) at a 95% confidence level. The classification accuracy of the eutrophication status was 86.5% for the training dataset and 85.6% for the validation dataset. The results indicated that it is feasible to develop an SVM technique for timely evaluation of the eutrophication status by easily measured parameters.

## 1. Introduction

Eutrophication is a water enrichment in nutrients, especially compounds of nitrogen (N) and phosphorus (P), that causes an accelerated growth of algae and higher forms of plant life to generate an undesirable disturbance to the quality of the water and to the balance of aquatic organisms (Ferreira et al., 2010). It usually leads to the depletion of dissolved oxygen, reduction in light transparency, occurrence of toxic algal blooms and a loss of biodiversity in marine environments (Teixeira and Rosa, 2006; Tekile et al., 2015; Chen et al., 2016a). In recent decades, there has been an increasing tendency toward the symptoms of eutrophication in the coastal and offshore areas triggered by terrestrial input, including industrial activities, agriculture production, transport, energy production, fishing and tourism (Bricker et al., 2008; Selman et al., 2008; García-Nieto et al., 2016b). Eutrophication has aroused an increasing concern in marine environmental protection worldwide due to the severe threat to marine ecosystems and public health (Xue and Landis, 2010; Cabrita et al., 2015; Schmoker et al., 2016). Therefore, it is indispensable to develop the techniques for real-time evaluation of marine eutrophication and regularly implementing marine health management and eutrophication monitoring programs.

However, some procedures of data collection and analysis for the assessment of eutrophication, such as nutrients, biological oxygen demand (BOD), chemical oxygen demand (COD), and dissolved organic carbon (DOC), are dependent on field sampling and laboratory analysis,

strenuous, costly and time-consuming, which may not sufficiently facilitate rapid eutrophication assessment. To meet the requirements for real-time monitoring, the application of easily measured parameters would assist in developing the technology for timely evaluation of eutrophication. Chlorophyll *a* (Chl-*a*), dissolved oxygen (DO), temperature and salinity, easily measured in the field using a multi-parameter water quality probe, are most often used for characterizing eutrophication (Gibson et al., 2000; Alonso Fernández et al., 2014; Park et al., 2015; Xiong et al., 2016; Yan et al., 2016). Many opinions have been stated in the literature concerning the use of the aforementioned physicochemical or biological parameters for eutrophication assessment. For example, Alonso Fernández et al. (2014) modeled eutrophication and risk prevention in a reservoir in the Northwest of Spain from biological and physicochemical parameters (turbidity, DO, temperature, total nitrogen (TN), total phosphorus (TP), salinity, and so on) by using multivariate adaptive regression spline analysis. Kuo et al. (2007) used an artificial neural network to combine the key factors that influence water quality, such as DO, Chl-*a*, TP, and the Secchi disk depth, for eutrophication prediction in a reservoir in central Taiwan. Pinto et al. (2012) developed a two-level discriminant function analysis model for rapidly assessing the eutrophication risk from three easy-to-measure parameters (saturated dissolved oxygen, turbidity and temperature), providing approximately 72% prediction accuracy.

Apart from the above variables, chromophoric dissolved organic matter (CDOM), the optically active fraction of bulk dissolved organic matter (DOM), plays an important part in the biogeochemistry cycle of

\* Corresponding author.

E-mail address: [surongguo@ouc.edu.cn](mailto:surongguo@ouc.edu.cn) (R. Su).

<http://dx.doi.org/10.1016/j.marpolbul.2017.04.022>

Received 24 October 2016; Received in revised form 11 April 2017; Accepted 12 April 2017  
0025-326X/ © 2017 Published by Elsevier Ltd.

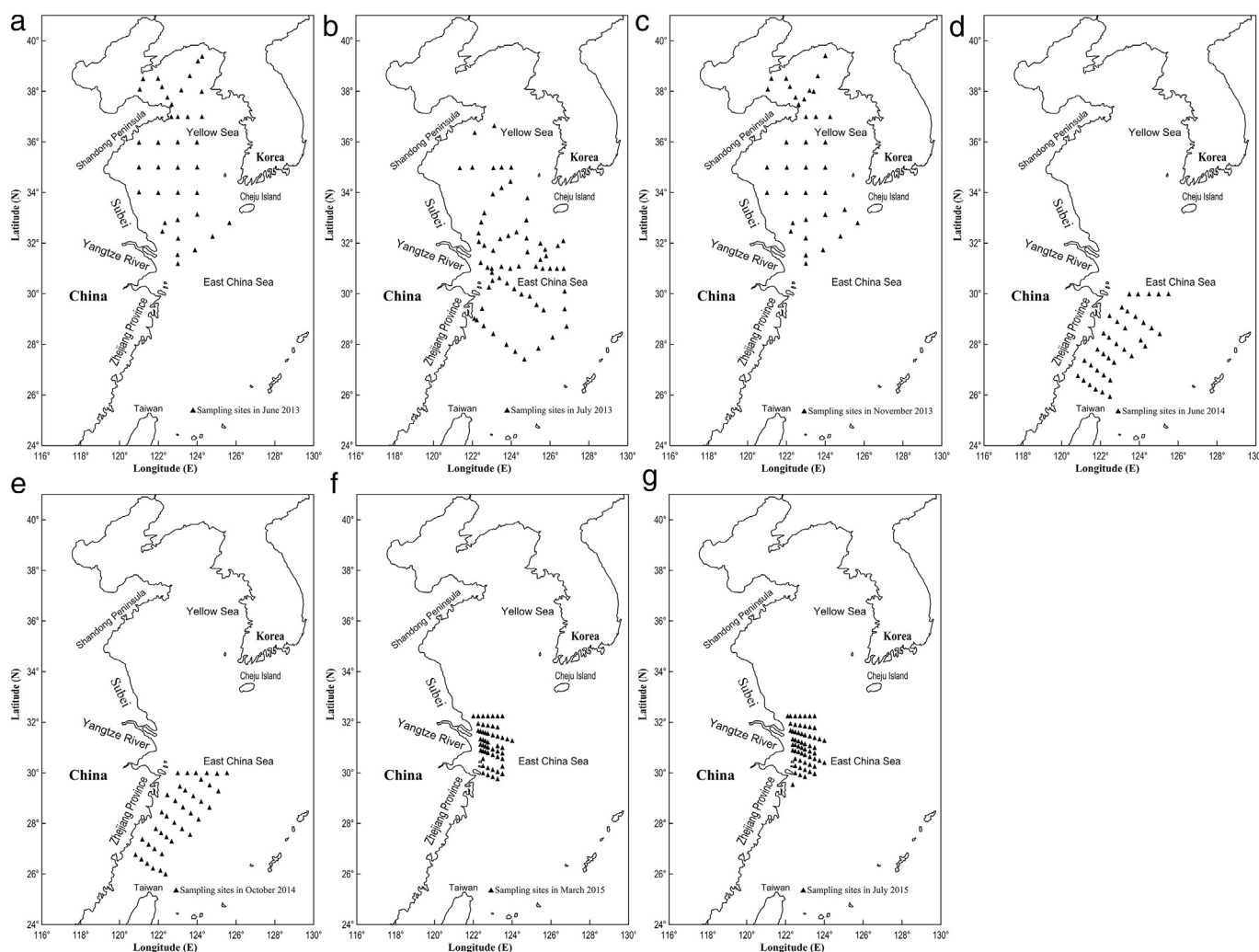


Fig. 1. Map of sampling sites in the Yellow Sea and the East China Sea.

nutrients and is often coupled with nutrients from which biologically available nitrogen or phosphorus is released for phytoplankton growth (Stedmon et al., 2007a; Zhang et al., 2011; Slonecker et al., 2016). Excitation emission matrix fluorescence coupled with parallel factor analysis (EEM-PARAFAC) has been considered the most effective tool for characterizing CDOM and assessing its sources and dynamics in a wide range of aquatic environments (Jørgensen et al., 2011; Yamashita et al., 2013; DeVilbiss et al., 2016). Recently, CDOM fluorescence sensors have been developed to meet real-time field measurement requirements at a high frequency and low cost (Foden et al., 2008; Kowalczyk et al., 2010b). Particular attention has been focused on determining the correlations between the CDOM fluorescence parameters and some water quality parameters associated with eutrophication (Zhang et al., 2010b; Osburn et al., 2011). For example, Liu et al. (2014) developed the techniques for real-time monitoring several water quality parameters (including TN, TP, COD, DOC and so on) from chromophoric dissolved organic matter (CDOM) fluorescence. Hur and Cho (2012) developed a real-time monitoring tool for the prediction of BOD, COD, and TN concentrations using PARAFAC-EEM indices in a typical urban river, and demonstrated an enhancement in the estimation capability with Spearman's rho values of 0.948, 0.977 and 0.984, respectively. Moreover, the fluorescence indices (such as the humification index HIX and the biological index BIX) often serve as composition indicators of CDOM (Huguet et al., 2009; Birdwell and Engel, 2010; Zhou et al., 2016b).

Although a wide range of physical, chemical, and biological

variables contributes to the understanding of marine eutrophication processes, some parameters are highly correlated and, undoubtedly, not entirely necessary for development of the eutrophication assessment method. Currently, multivariate statistical approaches have been developed and applied to assess the interrelations among the various parameters related to eutrophication and to combine eutrophic effects with different aspects of the marine environment. For example, principal component analysis (PCA) has been used to determine the main variables that affect eutrophication processes from a wide number of water quality parameters including TN, TP, oxygen, Chl-*a*, Secchi depth, phosphate, nitrate, nitrite, and ammonia (Lundberg et al., 2005; Primpas et al., 2010); Cluster analysis (CA) has been used to classify waters into the three eutrophication statuses including the oligotrophic, mesotrophic, and eutrophic state using several variables (Chl-*a*, phosphate, nitrate, nitrite, and ammonia) (Stefanou et al., 2000; Primpas et al., 2008); Discriminant factor analysis (DFA) has been used to identify different variables (nitrate, phosphate, Chl-*a*, DO, turbidity and temperature) that can differentiate sampling sites and to group them according to their eutrophication conditions (Tsirtsis and Karydis, 1999; Pinto et al., 2012); Artificial neural network (ANN) mode has been used for prediction of eutrophication conditions with reasonable accuracy by a wide range of variables (TP, TN, COD, the Secchi disk depth, DO and Chl-*a*) (Jiang et al., 2006; Kuo et al., 2007).

Support vector machine (SVM) (Vapnik, 1995) is a promising power approach used to reflect the nonlinearity between responsive indicator and input variables using stochastic error minimization approaches

(Zhou et al., 2016a) and is an effective tool to predict values from a wide variety of environmental fields (Ribeiro and Torgo, 2008; Farfani et al., 2015; Kisi et al., 2015). The grid search (GS) algorithm is straight forward to determine the optimized parameter values for the SVM (Sajan et al., 2015; Gao and Hou, 2016). Thus, the main objective of this study was to develop a rapid and low-cost method for evaluating the eutrophication status of coastal and offshore waters with four water quality parameters and CDOM fluorescence parameters by support vector machine (SVM).

## 2. Material and methods

### 2.1. The study area

As shown in Fig. 1, a total of 324 sampling sites were collected in seven field sampling campaigns conducted in the Yellow Sea (YS) and the East China Sea (ECS) in June ( $n = 37$ ), July ( $n = 64$ ) and November ( $n = 36$ ) 2013, in June ( $n = 36$ ) and October ( $n = 38$ ) 2014, and in March ( $n = 53$ ) and July ( $n = 60$ ) 2015.

The study area ( $25^{\circ} 56'30''\text{S}$ ,  $120^{\circ} 49'40''\text{E}$  to  $39^{\circ} 24'49''\text{S}$ ,  $126^{\circ} 51'0.4''\text{E}$ ) is a combination of two major marginal seas in the northwest Pacific Ocean, the Yellow Sea (YS) and the East China Sea (ECS), that are topographically connected, but divided subjectively by a line from the mouth of the Yangtze River to the Cheju Island (Sun et al., 2015; Ning et al., 2011; Qu et al., 2015). This region is mainly influenced by a series of circulation patterns, including the ECS Coastal Current and the YS Coastal Current that run along the mainland Chinese coast, the YS Cold Water Mass, the YS Warm Current, the Taiwan Warm Current, the Kuroshio Current and the Changjiang Diluted Water (Zhou et al., 2008; Yuan et al., 2008; Pang et al., 2016). In addition, the Changjiang River, one of the largest rivers worldwide, flows from the west into the ECS and YS with large amounts of nutrients that support high primary productivity (Li et al., 2007; Gan et al., 2016). Therefore, the YS and ECS are highly biologically active areas with complex hydrological variations and are strongly influenced by land-ocean interactions (Ning et al., 1998; Shi and Wang, 2012; Wang et al., 2016). In recent decades, as a result of the terrigenous input, the YS and ECS have received a high loading of anthropogenic nutrients resulting in increasingly serious eutrophication and frequent, harmful algal blooms (Gao et al., 2010; Zhang et al., 2010a; Chen et al., 2016b).

### 2.2. Data collection and analysis

#### 2.2.1. Chemical analysis

The dataset used was from 1198 samples collected at standard depths, i.e., surface, 10, 20, 30 m and bottom according to the water depth in the YS and ECS. Field measurement of salinity, temperature and DO were simultaneously determined during cruises by a CTD multi-parameter probe. Water samples for the determination of the Chl-*a*, TN, and TP concentrations were all collected using Niskin bottles mounted on a Seabird CTD Rosette. Water samples for the determination of the TN and TP concentrations were stored in 150 mL acid-cleaned plastic bottles at  $4^{\circ}\text{C}$  in the field until transport to the laboratory. TN and TP concentrations were determined using unfiltered aliquots of samples according to the Valderrama ((Koroleff, 1983a, b) method. Samples (500 mL–2 L) for determining Chl-*a* were filtered through 25 mm glass fiber filters (Whatman GF/F,  $0.7\ \mu\text{m}$  pore size) under low vacuum ( $< 0.3\ \text{kPa}$ ) and dim lighting to prevent the degradation of pigments. Chl-*a* concentrations were determined using a Shimadzu 2550 UV–Vis spectrophotometer after grinding the fresh filters in 10 mL of 90% acetone and extracting in the dark at above freezing temperatures for 12 h (Jeffrey and Humphrey, 1975).

#### 2.2.2. Three-dimensional fluorescence measurement

Excitation–emission matrix spectroscopy (EEMs) provided a three-dimensional contour plot with a set of fluorescence wavelength datasets

(Coble, 1996). Samples for spectroscopic analyses were filtered through a glass filtration unit (Whatman GF/F,  $0.7\ \mu\text{m}$ , combusted at  $450^{\circ}\text{C}$  for 4.5 h) to remove phytoplankton cells and large particles. The EEMs were determined on a fluorescence spectrophotometer (Fluorolog3-11) equipped with a 450-W Xe arc lamp (USHIO INC. Japan) and an R928P detector. EEMs was conducted using excitation wavelengths ranging from 240 to 480 nm in 5 nm intervals and emission wavelengths ranging from 250 to 580 nm in 5 nm intervals at a scanning speed of  $1200\ \text{nm min}^{-1}$ , and the resulting spectra were normalized to quinine sulfate units using  $0.01\ \text{mg/L}$  quinine sulfate monohydrate in a solution of  $0.05\ \text{mol/L}$   $\text{H}_2\text{SO}_4$  at  $E_x/E_m$  wavelengths of 350/450 nm (Hoge et al., 1993).

#### 2.2.3. Fluorescence indices

The humification index (HIX) and the biological index (BIX) were calculated from the EEM spectral data to determine the relative degree of humification and autotrophic productivity of fluorescent CDOM (Huguet et al., 2009; Birdwell and Engel, 2010). HIX was the ratio of the area under the emission spectra at 435–480 nm to that at 300–345 nm with an excitation wavelength of 254 nm. In this study, we used an excitation wavelength of 255 nm. BIX was determined as the ratio of the fluorescence intensity at an emission of 380 nm (characteristic of autochthonous production) and 430 nm (characteristic of humic-like materials) with an excitation wavelength of 310 nm.

#### 2.2.4. PARAFAC modeling

The parallel factor analysis (PARAFAC) technique was used to simulate the EEMs results into individual fluorophore peaks (Stedmon et al., 2003; Zhou et al., 2016b). It utilizes an alternating least-squares algorithm to minimize the sum of the squared residuals across the dataset and estimate the underlying structure of the EEMs (Harshman and Lundy, 1994). As a result, the EEMs fluorescence was resolved into a series of trilinear terms and a residual array:

$$x_{ijk} = \sum_{f=1}^F a_{if} b_{jf} c_{kf} + e_{ijk}$$

$$i = 1, 2, \dots, I \quad j = 1, 2, \dots, J \quad k = 1, 2, \dots, K \quad (1)$$

where  $x_{ijk}$  is the fluorescence intensity of the  $i$ th sample at the  $j$ th variable (emission wavelengths) and at the  $k$ th variable (excitation wavelengths) in the F-component model;  $a_{if}$  is directly proportional to the concentration of the  $f$ th fluorescent component of the  $i$ th sample;  $b_{jf}$  and  $c_{kf}$  are scaled estimates of the emission and excitation spectra at wavelengths  $j$  and  $k$ , respectively, for the  $f$ th fluorescent component; and  $e_{ijk}$  is the residual of the variability not accounted for by the model.

In this study, the fluorescence intensity of each component was represented by  $F_{\max}$  (QSU, i.e. quinine sulfate units) and split-half analysis was further performed to validate the reliability of the model results (Stedmon et al., 2003).

### 2.3. Trophic index TRIX

The composite trophic index TRIX, based on several biological, chemical, and physical indicators, offers a suitable and acceptable method for evaluating coastal eutrophication. It was chosen as a target variable for coastal eutrophication in this research. The following formula was used to calculate the coastal eutrophication levels (Vollenweider et al., 1998):

$$\text{TRIX} = [\log_{10} \text{Chl-}a \times \text{aD}\%O \times \text{TN} \times \text{TP} + k]/m \quad (2)$$

where Chl-*a* = chlorophyll *a* concentration in  $\mu\text{g/L}$ , aD%O = oxygen as an absolute percentage deviation from saturation, TN = total nitrogen in  $\mu\text{g/L}$ , TP = total phosphorus in  $\mu\text{g/L}$ . The parameters  $k = 1.5$  and  $m = 1.2$  were scale coefficients that were included to fix the lower limit value of the index.

As an integrated evaluation of indicators of eutrophication, the

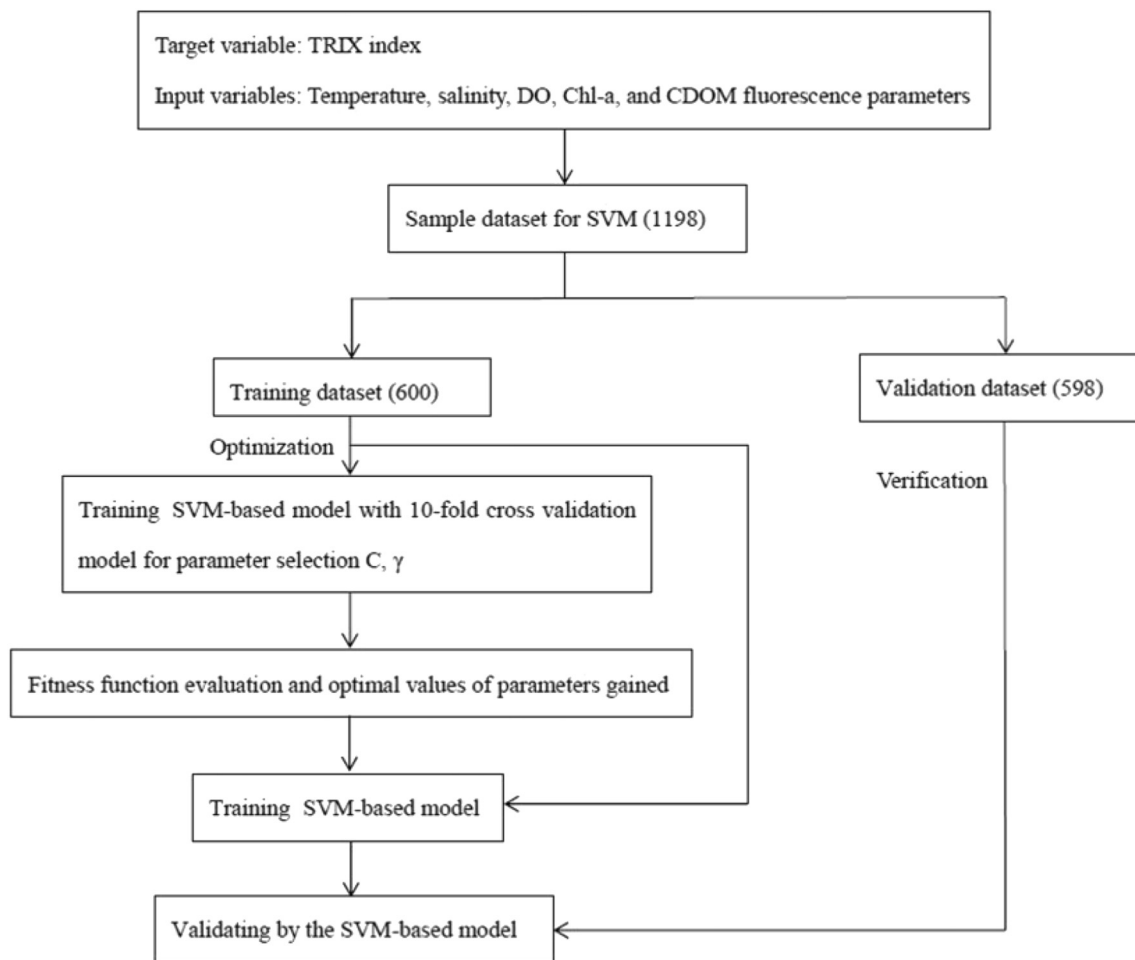


Fig. 2. Flowchart of the SVM optimization procedure. All the SVM algorithms were performed with Matlab 2012. One-way ANOVA was performed with the SPSS software package, Version 16.0.

TRIX index is a linear combination of the logarithms of four state variables (Chl-*a*, TN, TP, and ad%O), allowing synthesis of the key indicators into a simple numeric expression to effectively make information over a large array of spatial and temporal trophic conditions. The TRIX index suitably adapts to marine water features and provides useful metrics for the assessment of the trophic status of coastal waters (Pettine et al., 2007; Primpas and Karydis, 2011; Cabrita et al., 2015). This index has been scaled and used routinely to assess the trophic status of coastal waters. The proposed eutrophication scaling is as follows:  $TRIX < 4$ : pressure high water quality,  $4 \leq TRIX < 5$ : good water quality,  $5 \leq TRIX < 6$ : moderate water quality, and  $6 \leq TRIX$ : poor water quality. Higher TRIX values represent highly productive waters and trophic levels (Penna et al., 2004). The TRIX index has been used for assessing eutrophication in the Baltic Sea (Pettine et al., 2007; Primpas and Karydis, 2011), the Black Sea (Moncheva et al., 2002; Parkhomenko et al., 2003), the Caspian Sea (Nasrollahzadeh et al., 2008; Shahrban and Etemad-Shahidi, 2010) and the Adriatic Sea (Giovannardi and Vollenweider, 2004; Mozetič et al., 2008).

#### 2.4. SVM regression theory and model development

SVMs are a set of related machine learning methods that work on the principle of structural risk minimization exploration to reduce the upper limit of the generalization error (Taboada et al., 2007; García Nieto et al., 2015b). SVMs were first proposed for classification and were later popularized to settle regression problems (Suárez Sánchez et al., 2011; Sajan et al., 2015). The latter method is called support

vector regression (SVR).

Based on SVR, the target regression function could be defined in the feature space as follows (García Nieto et al., 2016a):

$$f(x) = w \cdot \varphi(x) + b \quad (3)$$

where  $w$  is the weight vector,  $\varphi(x)$  is the nonlinear mapping from the input space to the output space, and  $b$  is a bias. Through the nonlinear mapping function  $\varphi(x)$ , the input variable  $x$  is mapped into a high dimensional linear feature space, where an inner product was defined based on the kernel trick  $K$  (Subasi, 2013; García Nieto et al., 2016a):

$$\langle \varphi(x_i), \varphi(x_j) \rangle = \varphi(x_i) \cdot \varphi(x_j) = K(x_i, x_j) \quad (4)$$

where  $\langle , \rangle$  denotes the dot product in  $x$ ,  $\varphi(x_i) \cdot \varphi(x_j)$  is the product of inner vectors in the high dimensional space,  $K(x_i, x_j)$  is the kernel function.

Kernel function replaced the dot product operation in high feature space without calculating mapping function  $\varphi$ . The problem could be established by standard dualization principle utilizing Lagrange multipliers ( $\alpha_i, \alpha_i^*$ ) that define the relative importance of the training datasets for the output results (Smola and Schölkopf, 2004; Aich and Banerjee, 2016). The final function can be solved by a set of sample data points called support vectors (s.v.) as follows (de Cos Juez et al., 2010; García Nieto et al., 2015a):

$$w = \sum_{s.v.} \beta_i \varphi(x_i) \quad (5)$$

where  $\beta_i = (\alpha_i - \alpha_i^*)$  is the Lagrange coefficient corresponding to support vector.



Finally, a decision function is obtained based on Eqs. (3), (4), (5).

$$f_{w,b}(x) = \sum_{S.V.} \beta_i < \varphi(x_i), \varphi(x) > + b = \sum_{S.V.} \beta_i K(x_i, x) + b \quad (6)$$

The most commonly used types of kernel function in SVR model are linear kernel function (LKF), polynomial kernel function (PKF) and radial basis function (RBF), as expressed in Eqs. (7)–(9).

Linear kernel function (LKF) (Sajan et al., 2015):

$$K(x_i, x) = x_i^T x \quad (7)$$

Polynomial kernel function (PKF) (Fletcher, 2009):

$$K_{x_i, x} = \gamma x_i^T x + r^d, \gamma > 0 \quad (8)$$

where  $\gamma$ ,  $r$  and  $d$  are parameters defining the kernel's behavior.

Radial basis function (RBF) (Liu and Zhou, 2015):

$$K_{x_i, x} = \exp[-\gamma \|x - x_i\|^2], \gamma > 0 \quad (9)$$

where  $\gamma$  is the kernel function parameter and  $\|\cdot\|$  is the modulus of the high dimensional real vector.

These parameters must be selected accurately because they control the complexity of the final function and manage the structure of the high dimensional feature space (Hsu et al., 2003; Park et al., 2015; Sajan et al., 2015). Therefore, some algorithms should be applied to identify the optimal parameters. The grid search (GS) is a straightforward and simple algorithm to find the optimal parameter values for the SVM regression (Gao and Hou, 2016).

Fig. 2 shows the technical flowchart of the SVM-based model development. The TRIX index was used as target variable (out variable), while CDOM fluorescence parameters and four water quality parameters were chosen as input variables. Generally speaking, a regression task involves a training dataset and a validation dataset. The 1198 samples were randomly divided into two datasets, i.e., 600 samples in the training dataset and 598 samples in the validation dataset (see Table S1 and Table S2 in the Supplementary material). The training dataset was used to train the model by recognizing the patterns between input and output data. The validation dataset was used to verify the performance and effectiveness of the established model. In the present work, the GS module was proposed to select the optimal parameters of the SVM-based model, and then these optimal parameters were used to construct the SVM-based model in order for proceeded prediction.

### 3. Results and discussion

#### 3.1. Fluorescence component characteristics

Four fluorescence components were identified by PARAFAC analysis. Table 1 and Fig. 3 show excitation and emission wavelength pairs of

the main peaks of the four components in comparison with the results of other studies.

C1 exhibited a single peak at excitation/emission wavelengths of 360/450 nm (Fig. 3a, e), which was similar to the humic-like component in the visible region (peak C) introduced by Coble (1996). C1 represents the terrestrial humic-like substance that originates from microbial degradation and terrestrial sediments. This component was a typical fluorophore in terrestrial and coastal environments (Coble, 1996; Stedmon and Markager, 2005a; Murphy et al., 2008).

C2 presented a single peak at excitation and emission wavelengths of 330/395 nm (Fig. 3b, f), which was similar to those of the M peak (Coble, 1996). C2 was categorized as the marine humic-like substance derived from microbial or phytoplankton degradation processing (Stedmon and Markager, 2005b; Kowalczyk et al., 2010a; Zhang et al., 2011). This component was also identified as a terrestrial humic-like component (Guéguen et al., 2011) and has been found in both fresh and marine waters (Dainard and Guéguen, 2013).

C3 showed primary (secondary) fluorescence peaks at excitation/emission wavelengths of 415(< 315 nm)/495 nm (Fig. 3c, g), which could not be typically determined by traditional peaks (Coble, 1996) but was identified as a terrestrial humic-like substance (Yamashita et al., 2010). C3 fluorophores showed the longest emission wavelengths and the broadest excitation band in this study. This spectrum was related to aromatic terrestrial and high-molecular weight organic matter (Murphy et al., 2008; Kowalczyk et al., 2010a).

C4 exhibited a single well-defined peak at excitation/emission wavelengths of 295/340 nm (Fig. 3d, h), which was typically categorized as a fluorescent protein-like (tryptophan-like) component (peak T) (Coble, 1996). This component was considered to have relevance to biological production in bodies of water and was previously observed in many different aquatic environments (Coble, 1996; Coble et al., 1998; Yamashita and Tanoue, 2003).

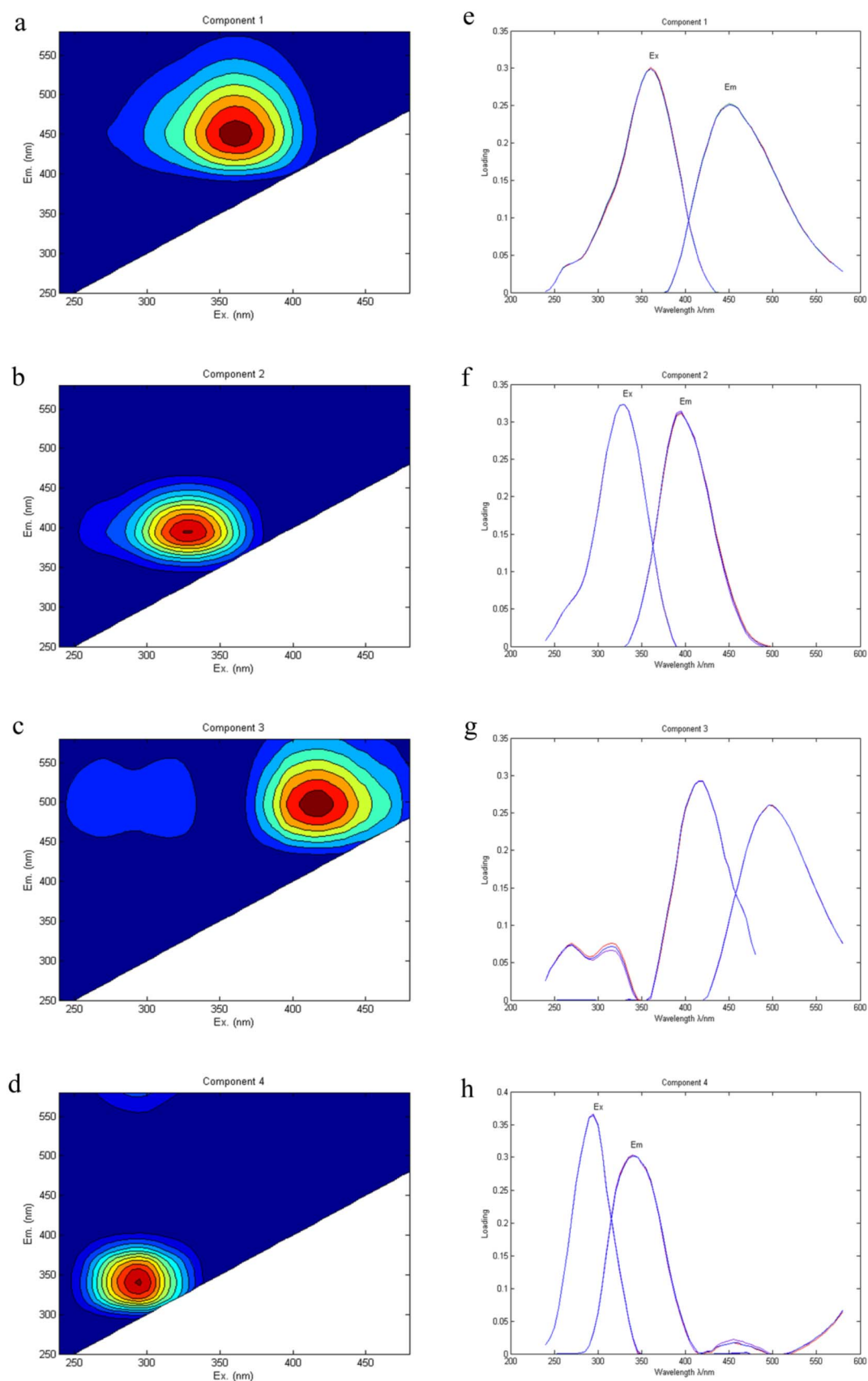
The fluorescence intensities of three humic-like components (C1, C2 and C3) were strongly negatively correlated with salinity ( $r = -0.719$  for C1,  $r = -0.810$  for C2,  $r = -0.799$  for C3) (Table 3). These relationships suggested that all three humic-like components were of terrestrial origin, even though the spectral characteristics of C2 are similar to the microbial humic-like “peak M” which has been reported to correlate with chlorophyll *a* in the ocean (Coble, 1996). In the present study, however, C2 showed strongly positive correlations with other humic-like components ( $r = 0.951$  for C3,  $r = 0.825$  for C1) but not with chlorophyll *a* ( $r = 0.069$ ) (Table 3), which further indicated that humic-like C2 was possibly of terrestrial origin.

The protein-like (tryptophan-like) substance C4 is considered to be both produced and consumed by marine bacteria, and influenced by anthropogenic activity and autochthonous production related to the biological degradation of CDOM (Romera-Castillo et al., 2011; Yao

**Table 1**

Characteristics of the four components identified in the present study compared with the results of previous studies.

Component	$E_x/E_m$ /nm	Description	Reference
Component 1	360/450	Terrestrial humic-like materials	Peak C: 350/420–480 (Coble, 1996) C3: 270 (360)/478 (Stedmon et al., 2003) C2: 345/433 (Yamashita et al., 2008)
Component 2	330/395	Marine and terrestrial humic-like materials	Peak M: 312/380–420 (Coble, 1996) C1: 325/404 (Kim and Kim, 2015) C3: 250(310)/400 (Kowalczyk et al., 2010a) C2: 320/410 (Guéguen et al., 2011)
Component 3	(< 315) 415/495	Terrestrial humic-like materials	C4: 270 (390)/508 (Kowalczyk et al., 2010a) C3: 260 (370)/490 (Murphy et al., 2008) C7: 420 (275)/488 (Stedmon and Markager, 2005b) C3: (300)410/510 (Yamashita et al., 2011)
Component 4	295/340	Tryptophan-like fluorescence	T park, 275/340 (Coble, 1996) C7: 280/344 (Stedmon and Markager, 2005a) C8: 280/330–340 (Fellman et al., 2008) C3: 280/340 (Wu et al., 2011)



**Fig. 3.** The Contour plots (a, b, c and d) and the excitation and emission loadings from the split-half validation (e, f, g and h) of the four components determined by the PARAFAC model. The contour plots show the spectral shapes of the excitation and emission. The loadings were estimated from the complete dataset and two of the independent halves.

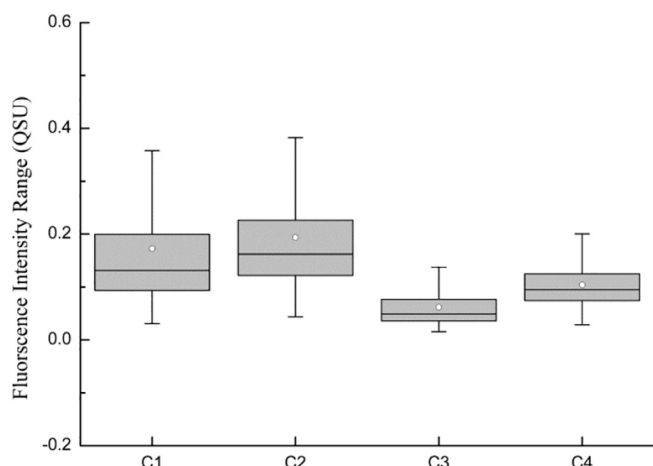


Fig. 4. Range of fluorescence intensities of four PARAFAC components.

et al., 2011). Previous studies indicated that the protein-like components were labile or semi-labile and related to primary production in the euphotic zone (Jørgensen et al., 2011).

The relative contributions of humic-like or protein-like fluorophores to CDOM can be calculated from the scores of each PARAFAC component. Each component score was shown in Fig. 4. The box plots showed the variability and mean values of the fluorescence intensities of all water samples analyzed. Humic-like C1 and C2 showed higher fluorescence intensities than humic-like C3, thus humic-like C1 and C2 had the higher relative contribution to CDOM than humic-like C3. Moreover, humic-like components are more abundant than protein-like components (Fig. 4) and terrestrial humic-like material significantly contributes to the total CDOM fluorescence intensity in the study area. Mostofa et al. (2009) indicated that humic substances of terrestrial origin were the dominant CDOM fractions in coastal seawater.

### 3.2. Descriptive statistical analyses

The full dataset consists of 10 input variables (including temperature, salinity, DO, Chl-*a* and 6 CDOM fluorescence parameters (HIX, BIX, C1, C2 C3 and C4)) and one target variable (TRIX index). The descriptive statistics for the data are shown in Table 2. The coefficient of variation (CV) values for all input parameters were from 0.07 to 0.92, which indicated low variation. Among these data, the TRIX index ranged from 3.9 to 7.4. The TRIX index of water samples was chosen as the target variable for the model development, and 10 parameters, temperature, salinity, DO, Chl-*a*, HIX, BIX, C1, C2 C3 and C4 were used as input variables for this analysis.

**Table 2**  
Descriptive statistics of the data collected in the Yellow Sea and East China Sea.

	Minimum	Maximum	Mean	Std. error	Std. deviation	CV
aD%O	0.32	82.91	26.47	0.63	21.80	0.82
Chl- <i>a</i> (μg/L)	0.12	32.59	2.94	0.08	2.72	0.92
Temperature (°C)	3.93	28.95	19.26	0.16	5.47	0.28
Salinity (PSU)	16.49	34.75	32.15	0.07	2.36	0.07
HIX	0.59	7.50	1.61	0.02	0.79	0.49
BIX	0.66	2.45	1.13	0.006	0.19	0.17
C1 (QSU)	0.03	1.93	0.17	0.004	0.15	0.89
C2 (QSU)	0.04	1.12	0.19	0.003	0.12	0.62
C3 (QSU)	0.02	0.38	0.06	0.001	0.04	0.64
C4 (QSU)	0.03	0.38	0.10	0.001	0.04	0.41
TN (μg/L)	75.53	1691.23	352.29	5.21	180.35	0.51
TP (μg/L)	3.54	196.91	20.95	0.36	12.53	0.60
TRIX	3.89	7.39	5.67	0.02	0.62	0.11

Std. deviation and CV represent standard deviation and coefficient of variation, respectively.

### 3.3. Correlation analyses

TN and TP have been typically used to estimate the quality of natural water bodies and their degree of eutrophication (Bourgeois et al., 2001), and they are independent variables of the TRIX formula. To verify the rationality of the adopted variables, the relationships between the 10 input variables (including temperature, salinity, DO, Chl-*a* and 6 CDOM fluorescence parameters (HIX, BIX, C1, C2 C3 and C4)) and the 3 target variables (TRIX index, TP and TN) were examined by the Spearman's correlation coefficients (Table 3).

Among the qualitative parameters, it could be seen that salinity and temperature had significant negative correlations with the nutrients because they represent terrigenous discharge and biological activity that could cause changes in nutrient conditions, which was analogous to previous studies (Gasiūnaitė et al., 2005; Siswanto et al., 2008; Wei et al., 2016; Xiong et al., 2016). DO and Chl-*a* showed positive correlations with TN ( $r = 0.121$  and  $0.103$ ,  $p < 0.01$ , for DO and Chl-*a*, respectively) and TP ( $r = 0.353$  and  $0.125$ ,  $p < 0.01$ , for DO and Chl-*a*, respectively). DO was the main factor influencing the nutrients cycle under certain environmental conditions, which has been proved in many previous studies (Boström and Pettersson, 1982; Kim et al., 2004; Zhang et al., 2014). Nitrogen (N) and phosphorus (P) commonly determine the growth of algae and excessive nutrients often result in nuisance algal blooms (Busse et al., 2006; Liu et al., 2010). Recent studies indicated that Chl-*a* can be used to predict TN and TP in various water environments (Meeuwig et al., 2000; Hoyer et al., 2002).

The four fluorescence components (C1, C2, C3 and C4) also exhibited very strong correlations with TN and TP. Indeed, some researchers found that nutrients can be significantly altered by either dissolved organic phosphorus (DOP) or dissolved organic nitrogen (DON) (Vahatalo and Jarvinen, 2007; Li et al., 2008), which might favor the occurrence of photosynthesis and then accelerate growth of algae in aquatic environments (McCarthy et al., 2009; Mohlin and Wulff, 2009). Previous studies showed close interrelations between CDOM fluorescence parameters and nutrients (Zhang et al., 2009; Liu et al., 2014). In addition, the humification index HIX was significantly correlated with TN ( $r = 0.552$ ,  $p < 0.01$ ) and TP ( $r = 0.441$ ,  $p < 0.01$ ), and the autochthonous index BIX exhibited a certain negative relationship with TN and TP ( $r = -0.265$  and  $-0.137$ ,  $p < 0.01$ , for TN and TP, respectively).

However, the Spearman's correlation coefficient just indicates the linear correlation between the different variables. If the correlation of variables is nonlinear, the Spearman's correlation coefficient is not appropriate to estimate the interrelation. Thus, the Spearman's correlation coefficient cannot be used to fully analyze the complex situations of marine eutrophication. SVM is an effective method to further explore the nonlinearity between target variable and input variables. Therefore, SVM was employed to predict eutrophication status (TRIX index) using

**Table 3**  
Correlation matrix of all parameters.

	aD%O	Chl- <i>a</i>	Temperature	Salinity	HIX	BIX	C1	C2	C3	C4	TN	TP	TRIX
aD%O	1												
Chl- <i>a</i>	−0.146**	1											
Temperature	−0.267**	0.135**	1										
Salinity	−0.109**	−0.144**	0.150**	1									
HIX	0.340**	0.003	−0.285**	−0.708**	1								
BIX	−0.259**	0.106**	0.222**	0.363**	−0.509**	1							
C1	0.267**	0.242**	−0.179**	−0.719**	0.750**	−0.371**	1						
C2	0.289**	0.069*	−0.248**	−0.810**	0.845**	−0.413**	0.825**	1					
C3	0.258**	0.139**	−0.221**	−0.799**	0.833**	−0.420**	0.868**	0.951**	1				
C4	0.039	0.150**	−0.230**	−0.664**	0.371**	−0.008	0.579**	0.698**	0.682**	1			
TN	0.121**	0.103*	−0.118**	−0.486**	0.552**	−0.265**	0.454**	0.490**	0.460**	0.260**	1		
TP	0.353**	0.125**	−0.236**	−0.271**	0.441**	−0.137**	0.367**	0.392**	0.361**	0.224**	0.468**	1	
TRIX	0.633**	0.365**	−0.299**	−0.284**	0.444**	−0.184**	0.411**	0.413**	0.401**	0.196**	0.511**	0.645**	1

\*\* Correlation is significant at the 0.01 level.

\* Correlation is significant at the 0.05 level.

the aforementioned 10 input variables (referred to as easy-to-measure parameters in this paper).

### 3.4. Relative importance of the input variables in SVM-based model

In fact, the multidimensional property of marine eutrophication means that no single variable can represent the eutrophication status and it cannot be simply evaluated by a single physical, chemical, or biological parameter (Xu et al., 2001; Kitsiou and Karydis, 2011; Cabrita et al., 2015). Each variable individually may not explain the trophic status accurately (Parinet et al., 2004; Kitsiou and Karydis, 2011). The 10 input variables of the SVM-based model are correlated with the eutrophication conditions in multiple ways with varying degrees of impact. It is essential to investigate the importance of input variables to the prediction of the TRIX index and to select the most appropriate input variables for optimizing the model.

Firstly, a GS optimized SVM-based model with the radial basis function (RBF) was constructed to predict the TRIX index from the following 10 easily measured variables: Chl-*a*, DO, salinity, temperature, four fluorescence components (C1, C2, C3 and C4) and CDOM fluorescence indices (HIX and BIX). Then, the weight of each input variable, which was calculated with support vectors and a Lagrange coefficient corresponding to support vector based on Eq. (5), was available for evaluating the influence of each variable on the eutrophication prediction (García Nieto et al., 2015b). In this way, the importance ranking of the ten input variables to predict the TRIX index (target variable) in this high nonlinear complex problem was shown in Table 4. The most significant variable in the TRIX index prediction is the Chl-*a*, followed by DO, C1, C2, C3, C4, HIX, temperature, salinity and finally BIX.

The most commonly occurring eutrophic symptom was an accelerated growth of algae. Chl-*a* is the major photosynthetic pigment of

**Table 4**  
Evaluation of input variables in importance according to their weights and standardized weights (using the absolute value) in the fitted SVM-based model.

Input variable	Weights	Standardized weights
Chl- <i>a</i>	5.14675	100
aD%O	3.555034	69.1
C1	2.077745	40.4
C2	1.548919	30.1
C3	1.459119	28.4
C4	1.122524	21.8
HIX	0.816254	15.9
T	−0.48753	−9.5
S	−0.15049	−2.9
BIX	0.016585	0.3

algae and is usually employed as an estimator for algae growth (Gibson et al., 2000). Excessive nutrients bring about a progression of eutrophic symptoms that often begin with observations of high concentrations of chlorophyll *a*, suggesting that Chl-*a* is an important biological response closely related to water eutrophication (McCarthy et al., 2009; Mohlin and Wulff, 2009). Chlorophyll *a*, at high levels, always represents the water quality degradation associated with eutrophication (Ferreira et al., 2007; Bricker et al., 2008). This would explain why Chl-*a* was the most important variable in eutrophication prediction in this study.

The DO levels in waters are very complicated and greatly dependent on temperature, salinity, depth, turbulence, degradation of organic matter and the photosynthesis and respiratory metabolism of most aquatic organisms (Badran, 2001; Wheeler et al., 2003; Manasrah et al., 2006). DO was the main factor influencing the solubility and availability of nutrients and the release of nutrients from the sediments into the water, which can act strongly on the eutrophication status (Xie et al., 2003; Kim et al., 2004; García Nieto et al., 2013). Therefore, DO is an indispensable environmental condition for algal growth and the production of biodegradable organic matter and is an important environmental disturbance indicator in the water quality and eutrophication status (Pettine et al., 2007; Yan et al., 2016).

CDOM fluorescence components (C1, C2, C3 and C4) had a significant effect on predicting the eutrophication status (i.e., Rank 3, Rank 4, Rank 5, and Rank 6). CDOM is the optically active components of dissolved organic matter (DOM) consisting of a complicated mixture of organic compounds of both autochthonous and allochthonous origin (Coble, 2007; Massicotte and Markager, 2016). Indeed, nutrients were significantly produced from autochthonous or allochthonous DOM, which may impact eutrophication (Bushaw et al., 1996). Excessive nutrients can promote the growth of phytoplankton, which increases the fluorescence components of the CDOM derived from phytoplankton degradation (Zhang et al., 2009), whereas the increasing of CDOM might lead to an increase in the amount of nutrients through microbial degradation and photodegradation (Piccini et al., 2009). Additionally, CDOM may influence phytoplankton dynamics by limiting nutrients in the form of dissolved organic nitrogen (DON) or phosphorus (DOP) (Glibert et al., 2004; Maie et al., 2012). The significant correlations between CDOM and nutrients have been proven and utilized to predict nutrient concentrations (Hur and Cho, 2012; Liu et al., 2014).

The HIX and BIX also play a role in the TRIX index prediction. The HIX was applied to research the degree of humification at which highly humified organic substance is thought to be less labile than lowly humified organic material (Ohno, 2002). This also indicated the higher degree humified organic material could maintain longer than lower degree humified matter in water environments. BIX presents an estimation of the recently produced or autochthonous production of organic matter DOM and autochthonous biological activity in water



environments (Wang et al., 2014). An increase of BIX index is associated with an increase of the concentration of the autochthonous DOM in aquatic environments (Huguet et al., 2009). The HIX index for the humified content of CDOM with high values ( $> 10$ ) for CDOM indicates a significant humified feature or mainly terrigenous origin, but low values ( $< 4$ ) represents non-humified CDOM of aquatic bacterial or biological origin (Huguet et al., 2009). The BIX index for the autochthonous DOM with high values ( $> 0.8$ ) indicates a significant autochthonous contribution from bacterial or biological origin, but low values ( $< 0.6$ – $0.7$ ) presents a weekly autochthonous contribution (Huguet et al., 2009; Birdwell and Engel, 2010).

In this study, the HIX values were generally  $< 4$  (only 26 values exceeded this) and the BIX index with a mean value of  $1.13 \pm 0.63$  ranged between 0.66 and 2.45 and only eight of the 1198 values were  $< 0.8$  (Table 2), which was indicative of weakly humified and labile organic material.

In the coastal area, terrigenous inputs were a dominant CDOM source and the organic pollutants from anthropogenic activity often significantly contribute to an increase in the concentration of CDOM, especially in a densely populated and developed coastal area (Stedmon et al., 2007b).

As reported by previous studies, temperature, mostly controlled by seasons, is often considered to be an important variable related to algae growth and aquatic organism activity (García-Nieto et al., 2016b; Wei et al., 2016; Xiong et al., 2016). Salinity, associated with freshwater discharge in the coastal systems, is one of the abiotic factors that can be utilized as a simple freshwater tracer (Gasiūnaitė et al., 2005; Yoon and Woo, 2015). Salinity variations can also be thought of as major environmental factors impacting the distribution and existence of organisms in estuaries and coastal regions (Yoon and Woo, 2015). It can be used as an indicator of trophic situations and related to eutrophication control (Gasiūnaitė et al., 2005). Hence, temperature and salinity are external factors that mostly indirectly impact the eutrophication status, and thus they became relatively insignificant variables in the TRIX index prediction by the SVM-based model (Table 4).

The standardized weights have high values ( $> 20$ ) for Chl-*a*, DO and CDOM fluorescence components, but low values ( $< 20$ ) for salinity, temperature, HIX and BIX. Specifically, they identified Chl-*a*, DO and CDOM fluorescence components (C1, C2, C3 and C4) as more meaningful variables than salinity, temperature, HIX and BIX. This was one of the main findings in this work. For this, a new SVM-based model was further explored using the 6 easily measured variables including Chl-*a*, DO and CDOM fluorescence components (C1, C2, C3 and C4). Obviously, this could assist in designing an optimal SVM-based model to identify the eutrophication status and save costs by reducing the number of input parameters.

### 3.5. Model development and performance evaluation

A GS optimized SVM with the radial basis function (RBF) was proposed to model relationships between the 6 easily measured parameters (DO, Chl-*a*, C1, C2, C3 and C4) and the TRIX index, for rapid assessment of marine eutrophication states in the YS and ECS. SVM can correctly determine the global optimum regardless of input dimensionality and thus avoid over-fitting the output, resulting in better generalization performance. The application of those easily measured parameters would meet the requirements for real-time monitoring and facilitate the rapid comprehensive assessment of eutrophication. Based on the training dataset, a GS optimized SVM-based model was constructed. To avoid under-fitting and over-fitting of the SVM-based model, an exhaustive 10-fold cross validation technique was used to select the optimal SVM parameters that can simulate as many real situations as possible so that the model can adapt the new observations (Picard and Cook, 1984; Sajan et al., 2015). The training dataset was randomly divided into 10 equal subsets. Each subset was

employed once as a validation dataset, whereas the other 9 subsets were grouped into a new training dataset, and the optimal parameters of the SVM-based model were found with the grid search (GS) technique. This process was performed 10 times for calculating and testing the average error (Cristianini and Shawe-Taylor, 2000; Zhong et al., 2016). In this way, all the possible variability of the SVM-based model has been estimated to obtain the optimal parameters that minimize the average error.

As a consequence, the mean squared error (MSE) for cross-validation was 0.010, and the suitable optimal parameters of the SVM-based model were calculated as penalty parameter  $C = 5.66$ , kernel function parameter  $\gamma = 0.35$ . Using the above parameters, 202 support vectors were selected. The SVM function of the TRIX index is expressed as Eq. (10).

$$f(x) = \sum_{i=1}^{202} \beta_i \exp \left( -0.35 \left\| x - x_i \right\|^2 \right) + b \quad (10)$$

where  $x$  is the input variable,  $f(x)$  is the target variable,  $\beta_i$  is the Lagrange coefficient corresponding to support vector, and  $b$  is a bias.

Predictive performance of the proposed model was measured for training and validation datasets by the coefficient of determination ( $R^2$ ) and the mean squared error (MSE). The  $R^2$  represents the correlation between the measured and the predicted values, and thus determines accuracy of the fitting model. The MSE indicates the relative improvement in accuracy, and thus a smaller value is indicative of better accuracy.

Figs. 5 and 6 illustrate the differences between the measured and predicted values of TRIX index on the training and validation datasets. The coefficient of determination for the SVM-based model was 0.92 for the training dataset and 0.91 for the validation dataset, which indicated an approximately positive relationship with the linear correlations between the measured and the predicted TRIX index.

The differences between the measured and the predicted TRIX index were also evaluated by a one-way analysis of variance (ANOVA), followed by Dunnett's test at the 95% confidence level ( $p = 0.05$ ) with  $F$  values to determine a significant difference (Table 5). Differences were considered to be significant if  $F$  value  $\geq F$  critical value ( $p < 0.05$ ), whereas not significant if  $F$  value  $< F$  critical value ( $p > 0.05$ ) (Zhang et al., 2016).

One-way ANOVA showed no significant difference between the measured and the predicted TRIX index for the training dataset and the validation dataset (ANOVA followed by Dunnett's test:  $F = 0.015$ ,  $p > 0.05$  and  $F = 0.007$ ,  $p > 0.05$ , respectively) and confirmed the validity of the SVM-based model.

To determine whether the eutrophication status was well evaluated

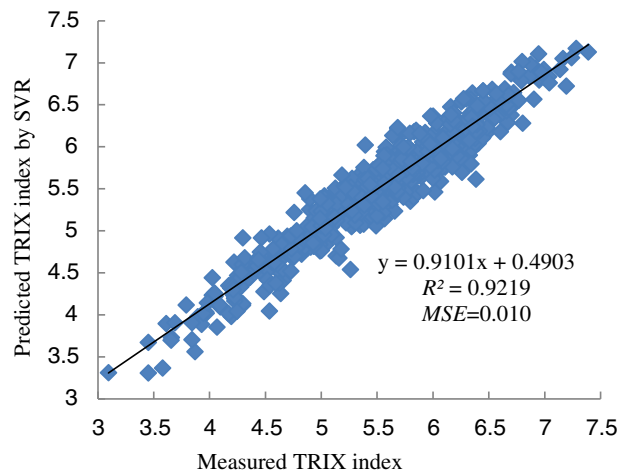


Fig. 5. Scatter plots of measured and predicted TRIX index for training dataset using the SVM.

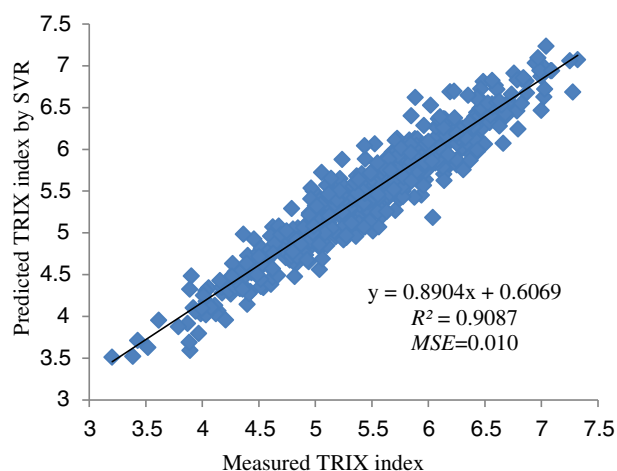


Fig. 6. Scatter plots of measured and predicted TRIX index for validation dataset using the SVM.

by the SVM-based model, each sample was categorized in relation to the proposed TRIX eutrophication scaling. There were only 34 samples scaling  $TRIX < 4$ . Thus, high and good status of original TRIX was redefined as good status and the proposed eutrophication scaling is as follows:  $TRIX < 5$ : pressure good water quality,  $5 \leq TRIX < 6$ : moderate water quality, and  $6 \leq TRIX$ : poor water quality. As a result, the accuracies of the SVM-based model for good water quality, moderate water quality, poor water quality and overall data cases for the training dataset and the validation dataset are presented in Tables 6 and 7.

For the training dataset, the overall classification accuracy was 86.5%. Good water quality and the moderate trophic dataset cases were well predicted by the eutrophication model (91.8% and 88.7%, respectively), whereas poor water quality was poorly predicted by the eutrophication model (77.5%). The incorrectly classified samples included 12 good status samples, 33 moderate status samples and 36 poor status samples according to TRIX classification results, in which 8 of 12 good status samples were with TRIX values of 4.9–5.0, 20 of 33 moderate status samples with TRIX values of 5.0–5.1 or 5.9–6.0, and 20 of 36 poor status samples with TRIX values of 6.0–6.1. Most incorrectly classified samples had TRIX values near the boundaries (Table 6).

For the validation dataset, the overall classification accuracy was 85.6%. The samples of good, moderate and the poor trophic status were all well classified by the SVM model (83.3%, 88.7% and 81.8%, respectively). The incorrectly classified samples included 24 good status samples, 34 moderate status samples and 28 poor status samples according to TRIX classification results, in which 17 of 24 good status samples were with TRIX values of 4.9–5.0, 17 of 34 moderate status samples with TRIX values of 5.0–5.1 or 5.9–6.0, and 16 of 28 poor status samples with TRIX values of 6.0–6.1. Most incorrectly classified samples also had TRIX values near the boundaries (Table 7). Those results indicated a good agreement between the measured and the predicted TRIX index. Therefore, it is feasible to use the SVM-based model as an effective approach to solve the problem of nonlinearities of the eutrophication status.

Table 5

Differences identified by one-way analysis of variance (ANOVA) between the measured and the predicted TRIX index for the training dataset and the validation dataset.

Mean square				F value		F critical value		p Value	
Between group		Within group							
Training dataset	Validation dataset	Training dataset	Validation dataset	Training dataset	Validation dataset	Training dataset	Validation dataset	Training dataset	Validation dataset
0.008	0.004	0.553	0.524	0.015	0.007	3.850	3.850	0.902	0.935

Table 6

The classification accuracy of the SVM-based model for the training dataset.

Data	Number of samples classified by the measured TRIX index (%)	Accuracy (%)	Number of samples classified by the predicted TRIX index (%)		
			Good	Moderate	Poor
Good	147	91.8	135	12	0
Moderate	293	88.7	12	260	21
Poor	160	77.5	0	36	124
Total	600	86.5	147	308	145

Table 7

The classification accuracy of the SVM-based model for the validation dataset.

Data	Number of samples classified by the measured TRIX index (%)	Accuracy (%)	Number of samples classified by the predicted TRIX index (%)		
			Good	Moderate	Poor
Good	144	83.3	120	24	0
Moderate	300	88.7	15	266	19
Poor	154	81.8	0	28	126
Total	598	85.6	135	318	145

### 3.6. The performance of linear kernel function (LKF) and polynomial kernel function (PKF)

The kernel function played an important role in the SVM regression progress since the kernel function could simplify the learning process by defining the feature space. The performances for LKF and PKF were evaluated. Firstly, the GS optimized SVM-based models were developed with LKF and PKF and the hyperparameters were given. Then, the two GS optimized SVM-based models were implemented to predict the TRIX values by Chl-*a*, DO and four fluorescence components (C1, C2 C3 and C4), the predictive performances were assessed by the coefficient of determination ( $R^2$ ) and the mean squared error (MSE) (Table 8).

The penalty parameter  $C$  is the most important parameter which determines the tradeoff between minimizing the training error and minimizing the model's complexity. If  $C$  was too small, the model would place insufficient stress on the fitting of training data. On the contrary, an overfitted output might be obtained when a large  $C$  value was employed (Sajan et al., 2015; Xu et al., 2015). The SVM model with LKF had a large penalty parameter  $C$  equal to 256 and led to an overfitted output with a high  $R^2$  of 0.97 for the training dataset and a low  $R^2$  of 0.11 for the validation dataset. The SVM model with RBF had a higher  $R^2$  and a lower mean squared error (MSE) for the validation dataset than that with PKF, which indicated a better agreement between the predicted TRIX values and the measured TRIX values. Recent works also suggested that the use of RBF in the SVM regression model was appropriate and sufficient because of its effectiveness, robustness, the acceptable generalization ability and the promising results (Chou et al., 2014; Sajan et al., 2015; Huang et al., 2016; Mehraein and Riahi, 2017).

**Table 8**

The parameters of the GS optimized SVM-based models with different kernel functions.

Kernel function	Values of optimal parameters	The training dataset		The validation dataset	
		$R^2$	MSE	$R^2$	MSE
Linear kernel function (LKF)	Penalty parameter $C = 256$	0.97	0.008	0.11	0.166
Polynomial kernel function (PKF)	Penalty parameter $C = 0.25$ ; Kernel function parameter $\gamma = 2.83$	0.92	0.010	0.88	0.015
Radial basis function (RBF)	Penalty parameter $C = 5.66$ ; Kernel function parameter $\gamma = 0.35$	0.92	0.010	0.91	0.010

#### 4. Conclusions

In this paper, a GS optimized SVM-based model has been developed to rapidly and timely evaluate marine eutrophication status of the coastal waters with easily measured parameters (DO, Chl-*a* and the four fluorescence components). The predicted results are high consistent with the measured TRIX index at a 95% confidence level, the coefficients of determination equal to 0.92 for the training dataset and 0.91 for the validation dataset, and the overall classification accuracy was 86.5% for the training dataset and 85.6% for the validation dataset. The SVM-based model would be an effective approach for in-situ monitoring of eutrophication of the coastal waters. When it is fully explored, the integrated SVM regression method and framework has extensive potential applications, not only in water quality evaluation but also in other environmental areas.

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.marpolbul.2017.04.022>.

#### Conflicts of interest

The authors declare no conflict of interest.

#### Acknowledgments

This study was supported by the Natural Science Foundation of China (No. 41376106), the National Key Research and Development Program (No. 2016YFC1402101) and the Research Program from National Marine Hazard Mitigation Service (No. 2014AA060).

#### References

- Aich, U., Banerjee, S., 2016. Application of teaching learning based optimization procedure for the development of SVM learned EDM process and its pseudo Pareto optimization. *Appl. Soft Comput.* 39, 64–83.
- Alonso Fernández, J.R., García Nieto, P.J., Díaz Muñoz, C., Álvarez Antón, J.C., 2014. Modeling eutrophication and risk prevention in a reservoir in the Northwest of Spain by using multivariate adaptive regression splines analysis. *Ecol. Eng.* 68, 80–89.
- Badran, M.I., 2001. Dissolved oxygen, chlorophyll *a* and nutrients: seasonal cycles in waters of the Gulf of Aquaba, Red Sea. *Aquat. Ecosyst. Health Manag.* 4 (2), 139–150.
- Birdwell, J.E., Engel, A.S., 2010. Characterization of dissolved organic matter in cave and spring waters using UV–vis absorbance and fluorescence spectroscopy. *Org. Geochem.* 41 (3), 270–280.
- Boström, B., Pettersson, K., 1982. Different patterns of phosphorus release from lake sediments in laboratory experiments. *Hydrobiologia* 91, 415–429.
- Bourgeois, W., Burgess, J.E., Stuetz, R.M., 2001. On line monitoring of wastewater quality: a review. *J. Chem. Technol. Biotechnol.* 76 (4), 337–348.
- Bricker, S.B., Longstaff, B., Dennison, W., Jones, A., Boicourt, K., Wicks, C., Woerner, J., 2008. Effects of nutrient enrichment in the nation's estuaries: a decade of change. *Harmful Algae* 8 (1), 21–32.
- Bushaw, K.L., Zepp, R.G., Tarr, M.A., Schulz-Jander, D., Bourbonniere, R.A., Hodson, R.E., Miller, W.L., Bronk, D.A., Moran, M.A., 1996. Photochemical Release of Biologically Available Nitrogen From Aquatic Dissolved Organic Matter.

- Busse, L.B., Simpson, J.C., Cooper, S.D., 2006. Relationships among nutrients, algae, and land use in urbanized southern California streams. *Can. J. Fish. Aquat. Sci.* 63 (12), 2621–2638.
- Cabrita, M.T., Silva, A., Oliveira, P.B., Angélico, M.M., Nogueira, M., 2015. Assessing eutrophication in the Portuguese continental exclusive economic zone within the European marine strategy framework directive. *Ecol. Indic.* 58, 286–299.
- Chen, C.W., Ju, Y.R., Chen, C.F., Dong, C.D., 2016a. Evaluation of organic pollution and eutrophication status of Kaohsiung Harbor, Taiwan. *Int. Biodeterior. Biodegrad.* 113, 318–324.
- Chen, Y., Yang, G.P., Liu, L., Zhang, P.Y., Leng, W.S., 2016b. Sources, behaviors and degradation of dissolved organic matter in the East China Sea. *J. Mar. Syst.* 155, 84–97.
- Chou, J.S., Cheng, M.Y., Wu, Y.W., Pham, A.D., 2014. Optimizing parameters of support vector machine using fast messy genetic algorithm for dispute classification. *Expert Syst. Appl.* 41 (8), 3955–3964.
- Coble, P.G., 1996. Characterization of marine and terrestrial DOM in seawater using excitation-emission matrix spectroscopy. *Mar. Chem.* 51 (4), 325–346.
- Coble, P.G., 2007. Marine optical biogeochemistry: the chemistry of ocean color. *Chem. Rev.* 107 (2), 402–418.
- Coble, P.G., Del Castillo, C.E., Avril, B., 1998. Distribution and optical properties of CDOM in the Arabian Sea during the 1995 Southwest Monsoon. *Deep-Sea Res. II Top. Stud. Oceanogr.* 45 (10), 2195–2223.
- de Cos Juez, F.J., Nieto, P.G., Torres, J.M., Castro, J.T., 2010. Analysis of lead times of metallic components in the aerospace industry through a supported vector machine model. *Math. Comput. Model.* 52 (7), 1177–1184.
- Cristianini, N., Shawe-Taylor, J., 2000. An Introduction to Support Vector Machines and Other Kernel-based Learning Methods. Cambridge University Press.
- Dainard, P.G., Guéguen, C., 2013. Distribution of PARAFAC modeled CDOM components in the North Pacific Ocean, Bering, Chukchi and Beaufort seas. *Mar. Chem.* 157, 216–223.
- DeVilbiss, S.E., Zhou, Z., Klump, J.V., Guo, L., 2016. Spatiotemporal variations in the abundance and composition of bulk and chromophoric dissolved organic matter in seasonally hypoxia-influenced Green Bay, Lake Michigan, USA. *Sci. Total Environ.* 565, 742–757.
- Farfani, H.A., Behnamfar, F., Fathollahi, A., 2015. Dynamic analysis of soil-structure interaction using the neural networks and the support vector machines. *Expert Syst. Appl.* 42 (22), 8971–8981.
- Fellman, J.B., D'Amore, D.V., Hood, E., Boone, R.D., 2008. Fluorescence characteristics and biodegradability of dissolved organic matter in forest and wetland soils from coastal temperate watersheds in southeast Alaska. *Biogeochemistry* 88 (2), 169–184.
- Ferreira, J.G., Bricker, S.B., Simas, T.C., 2007. Application and sensitivity testing of a eutrophication assessment method on coastal systems in the United States and European Union. *J. Environ. Manag.* 82 (4), 433–445.
- Ferreira, J.G., Andersen, J.H., Borja, A., Bricker, S.B., Camp, J., Cardoso da Silva, M., Garcés, E., Heiskanen, A.S., Humborg, C., Ignatiades, L., Lancelot, C., Menesguen, A., Tett, P., Hoepffner, N., Claussen, U., 2010. Marine strategy framework directive–task group 5 report eutrophication. *EUR* 24338, 49.
- Fletcher, T., 2009. Support vector machines explained: introductory course. In: Internal Report. University College London (UCL), London.
- Foden, J., Sivy, D.B., Mills, D.K., Devlin, M.J., 2008. Spatial and temporal distribution of chromophoric dissolved organic matter (CDOM) fluorescence and its contribution to light attenuation in UK waterbodies. *Estuar. Coast. Shelf Sci.* 79 (4), 707–717.
- Gan, S., Wu, Y., Zhang, J., 2016. Bioavailability of dissolved organic carbon linked with the regional carbon cycle in the East China Sea. *Deep-Sea Res. II Top. Stud. Oceanogr.* 124, 19–28.
- Gao, X., Hou, J., 2016. An improved SVM integrated GS-PCA fault diagnosis approach of Tennessee Eastman process. *Neurocomputing* 174, 906–911.
- Gao, L., Fan, D., Li, D., Cai, J., 2010. Fluorescence characteristics of chromophoric dissolved organic matter in shallow water along the Zhejiang coasts, southeast China. *Mar. Environ. Res.* 69 (3), 187–197.
- García Nieto, P.J., Alonso Fernández, J.R., de Cos Juez, F.J., Sánchez Lasheras, F., Díaz Muñoz, C., 2013. Hybrid modelling based on support vector regression with genetic algorithms in forecasting the cyanotoxins presence in the Trasona reservoir (Northern Spain). *Environ. Res.* 122, 1–10.
- García Nieto, P.J., Alonso Fernández, J.R., González Suárez, V.M., Díaz Muñoz, C., García-Gonzalo, E., Mayo Bayón, R., 2015a. A hybrid PSO optimized SVM-based method for predicting of the cyanotoxin content from experimental cyanobacteria concentrations in the Trasona reservoir: a case study in Northern Spain. *Appl. Math. Comput.* 260, 170–187.
- García Nieto, P.J., García-Gonzalo, E., Sánchez Lasheras, F., de Cos Juez, F.J., 2015b. Hybrid PSO–SVM-based method for forecasting of the remaining useful life for aircraft engines and evaluation of its reliability. *Reliab. Eng. Syst. Saf.* 138, 219–231.
- García Nieto, P.J., García-Gonzalo, E., Alonso Fernández, J.R., Díaz Muñoz, C., 2016a. A hybrid PSO optimized SVM-based model for predicting a successful growth cycle of the *Spirulina platensis* from raceway experiments data. *J. Comput. Appl. Math.* 291, 293–303.
- García-Nieto, P.J., García-Gonzalo, E., Alonso Fernández, J.R., Díaz Muñoz, C., 2016b. Using evolutionary multivariate adaptive regression splines approach to evaluate the eutrophication in the Pozón de la Dolores lake (Northern Spain). *Ecol. Eng.* 94, 136–151.
- Gasiūnaitė, Z.R., Cardoso, A.C., Heiskanen, A.S., Henriksen, P., Kauppi, P., Olenina, I., Pilkaityte, R., Purina, I., Razinkovas, A., Sagert, S., Schubert, H., Wasmund, N., 2005. Seasonality of coastal phytoplankton in the Baltic Sea: influence of salinity and eutrophication. *Estuar. Coast. Shelf Sci.* 65 (1), 239–252.
- Gibson, G., Carlson, R., Simpson, J., Smeltzer, E., Gerritsen, J., Chapra, S., Heiskary, S., Jones, J., Kennedy, R., 2000. Nutrient Criteria Technical Guidance Manual: Lakes and



- Reservoirs (EPA-822-B-00-001). United States Environment Protection Agency, Washington DC.
- Giovanardi, F., Vollenweider, R.A., 2004. Trophic conditions of marine coastal waters: experience in applying the Trophic Index TRIX to two areas of the Adriatic and Tyrrhenian seas. *J. Limnol.* 63 (2), 199–218.
- Glibert, P.M., Heil, C.A., Hollander, D.J., Revilla, M., Hoare, A., Alexander, J., Murasko, S., 2004. Evidence for dissolved organic nitrogen and phosphorus uptake during a cyanobacterial bloom in Florida Bay. *Mar. Ecol. Prog. Ser.* 280, 73–83.
- Guéguen, C., Granskog, M.A., McCullough, G., Barber, D.G., 2011. Characterisation of colored dissolved organic matter in Hudson Bay and Hudson Strait using parallel factor analysis. *J. Mar. Syst.* 88 (3), 423–433.
- Harshman, R.A., Lundy, M.E., 1994. PARAFAC: parallel factor analysis. *Comput. Stat. Data Anal.* 18 (1), 39–72.
- Hoge, F.E., Vodacek, A., Blough, N.V., 1993. Inherent optical properties of the ocean: retrieval of the absorption coefficient of chromophoric dissolved organic matter from fluorescence measurements. *Limnol. Oceanogr.* 38 (7), 1394–1402.
- Hoyer, M.V., Frazer, T.K., Notestein, S.K., Canfield Jr., D.E., 2002. Nutrient, chlorophyll, and water clarity relationships in Florida's nearshore coastal waters with comparisons to freshwater lakes. *Can. J. Fish. Aquat. Sci.* 59 (6), 1024–1031.
- Hsu, C.W., Chang, C.C., Lin, C.J., 2003. A Practical Guide to Support Vector Classification.
- Huang, H.B., Li, R.X., Huang, X.R., Lim, T.C., Ding, W.P., 2016. Identification of vehicle suspension shock absorber squeak and rattle noise based on wavelet packet transforms and a genetic algorithm-support vector machine. *Appl. Acoust.* 113, 137–148.
- Huguet, A., Vacher, L., Relexans, S., Saubusse, S., Froidefond, J.M., Parlanti, E., 2009. Properties of fluorescent dissolved organic matter in the Gironde Estuary. *Org. Geochem.* 40 (6), 706–719.
- Hur, J., Cho, J., 2012. Prediction of BOD, COD, and total nitrogen concentrations in a typical urban river using a fluorescence excitation-emission matrix with PARAFAC and UV absorption indices. *Sensors* 12, 972–986.
- Jeffrey, S.W., Humphrey, G.F., 1975. New spectrophotometric equations for determining chlorophylls a, b, C1 and C2 in higher-plants, algae and natural phytoplankton. *Biochem. Physiol. Pflanz.* 167 (2), 191–194.
- Jiang, Y., Zuxin, X., Hailong, Y., 2006. Study on improved BP artificial neural networks in eutrophication assessment of China eastern lakes. *J. Hydrodyn., Ser. B* 18 (3), 528–532.
- Jørgensen, L., Stedmon, C.A., Kragh, T., Markager, S., Middelboe, M., Søndergaard, M., 2011. Global trends in the fluorescence characteristics and distribution of marine dissolved organic matter. *Mar. Chem.* 126 (1), 139–148.
- Kim, J., Kim, G., 2015. Importance of colored dissolved organic matter (CDOM) inputs from the deep sea to the euphotic zone: results from the East (Japan) Sea. *Mar. Chem.* 169, 33–40.
- Kim, L.H., Choi, E., Gil, K.I., Stenstrom, M.K., 2004. Phosphorus release rates from sediments and pollutant characteristics in Han River, Seoul, Korea. *Sci. Total Environ.* 321 (1), 115–125.
- Kisi, O., Shiri, J., Karimi, S., Shamshirband, S., Motamedi, S., Petković, D., Hashim, R., 2015. A survey of water level fluctuation predicting in Urmia Lake using support vector machine with firefly algorithm. *Appl. Math. Comput.* 270, 731–743.
- Kitsiou, D., Karydis, M., 2011. Coastal marine eutrophication assessment: a review on data analysis. *Environ. Int.* 37 (4), 778–801.
- Koroleff, F., 1983a. Determination of phosphorus. In: Grasshoff, K., Ehrhardt, M., Kremling, K. (Eds.), *Methods of Seawater Analysis*. Weinheim, Verlag Chemie, pp. 125–139.
- Koroleff, F., 1983b. Total and organic nitrogen. In: Grasshoff, K., Ehrhardt, M., Kremling, K. (Eds.), *Methods of Seawater Analysis*. Weinheim, Verlag Chemie, pp. 162–173.
- Kowalczyk, P., Cooper, W.J., Durako, M.J., Kahn, A.E., Gonsior, M., Young, H., 2010a. Characterization of dissolved organic matter fluorescence in the South Atlantic Bight with use of PARAFAC model: relationships between fluorescence and its components, absorption coefficients and organic carbon concentrations. *Mar. Chem.* 118 (1), 22–36.
- Kowalczyk, P., Zablocka, M., Sagan, S., Kulinski, K., 2010b. Fluorescence measured in situ as a proxy of CDOM absorption and DOC concentration in the Baltic Sea. *Oceanologia* 52 (3), 431–471.
- Kuo, J.T., Hsieh, M.H., Lung, W.S., She, N., 2007. Using artificial neural network for reservoir eutrophication prediction. *Ecol. Model.* 200 (1), 171–177.
- Li, M., Xu, K., Watanabe, M., Chen, Z., 2007. Long-term variations in dissolved silicate, nitrogen, and phosphorus flux from the Yangtze River into the East China Sea and impacts on estuarine ecosystem. *Estuar. Coast. Shelf Sci.* 71 (1), 3–12.
- Li, W., Wu, F., Liu, C., Fu, P., Wang, J., Mei, Y., Wang, L., Guo, J., 2008. Temporal and spatial distributions of dissolved organic carbon and nitrogen in two small lakes on the Southwestern China Plateau. *Limnology* 9 (2), 163–171.
- Liu, F., Zhou, Z., 2015. A new data classification method based on chaotic particle swarm optimization and least square-support vector machine. *Chemom. Intell. Lab. Syst.* 147, 147–156.
- Liu, Y., Guo, H., Yang, P., 2010. Exploring the influence of lake water chemistry on chlorophyll a: a multivariate statistical model analysis. *Ecol. Model.* 221 (4), 681–688.
- Liu, X., Zhang, Y., Shi, K., Zhu, G., Xu, H., Zhu, M., 2014. Absorption and fluorescence properties of chromophoric dissolved organic matter: implications for the monitoring of water quality in a large subtropical reservoir. *Environ. Sci. Pollut. Res.* 21 (24), 14078–14090.
- Lundberg, C., Lönnroth, M., Von Numers, M., Bonsdorff, E., 2005. A multivariate assessment of coastal eutrophication. Examples from the Gulf of Finland, northern Baltic Sea. *Mar. Pollut. Bull.* 50 (11), 1185–1196.
- Maie, N., Yamashita, Y., Cory, R.M., Boyer, J.N., Jaffé, R., 2012. Application of excitation emission matrix fluorescence monitoring in the assessment of spatial and seasonal drivers of dissolved organic matter composition: sources and physical disturbance controls. *Appl. Geochem.* 27 (4), 917–929.
- Manasrah, R., Raheed, M., Badran, M.I., 2006. Relationships between water temperature, nutrients and dissolved oxygen in the northern Gulf of Aqaba, Red Sea. *Oceanologia* 48 (2), 237–253.
- Massicotte, P., Markager, S., 2016. Using a Gaussian decomposition approach to model absorption spectra of chromophoric dissolved organic matter. *Mar. Chem.* 180, 24–32.133.
- McCarthy, M.J., James, R.T., Chen, Y., East, T.L., Gardner, W.S., 2009. Nutrient ratios and phytoplankton community structure in the large, shallow, eutrophic, subtropical lakes Okeechobee (Florida, USA) and Taihu (China). *Limnology* 10 (3), 215–227.
- Meeuwig, J.J., Kauppi, P., Pitkänen, H., 2000. Predicting coastal eutrophication in the Baltic: a limnological approach. *Can. J. Fish. Aquat. Sci.* 57 (4), 844–855.
- Mehraei, I., Riahi, S., 2017. The QSPR models to predict the solubility of CO<sub>2</sub> in ionic liquids based on least-squares support vector machines and genetic algorithm-multi linear regression. *J. Mol. Liq.* 225, 521–530.
- Mohlin, M., Wulff, A., 2009. Interaction effects of ambient UV radiation and nutrient limitation on the toxic cyanobacterium *Nodularia spumigena*. *Microb. Ecol.* 57 (4), 675–686.
- Moncheva, S., Dontcheva, V., Shtereva, G., Kamburska, L., Malej, A., Gorinstein, S., 2002. Application of eutrophication indices for assessment of the Bulgarian Black Sea coastal ecosystem ecological quality. *Water Sci. Technol.* 46 (8), 19–28.
- Mostofa, K.M.G., Wu, F.C., Yoshioka, T., Sakugawa, H., Tanoue, E., 2009. Dissolved Organic Matter in the Aquatic Environments. Natural Organic Matter and Its Significance in the Environment. Science Press, Beijing, pp. 3–66.
- Mozetič, P., Malačič, V., Turk, V., 2008. A case study of sewage discharge in the shallow coastal area of the Northern Adriatic Sea (Gulf of Trieste). *Mar. Ecol.* 29 (4), 483–494.
- Murphy, K.R., Stedmon, C.A., Waite, T.D., Ruiz, G.M., 2008. Distinguishing between terrestrial and autochthonous organic matter sources in marine environments using fluorescence spectroscopy. *Mar. Chem.* 108 (1), 40–58.
- Nasrollahzadeh, H.S., Din, Z.B., Foong, S.Y., Makhloogh, A., 2008. Trophic status of the Iranian Caspian Sea based on water quality parameters and phytoplankton diversity. *Cont. Shelf Res.* 28 (9), 1153–1165.
- Ning, X.R., Liu, Z.L., Cai, Y.M., Fang, M., Chai, F., 1998. Physicobiological oceanographic remote sensing of the East China Sea: satellite and in situ observations. *J. Geophys. Res. Oceans* 103 (C10), 21623–21635.
- Ning, X., Lin, C., Su, J., Liu, C., Hao, Q., Le, F., 2011. Long-term changes of dissolved oxygen, hypoxia, and the responses of the ecosystems in the East China Sea from 1975 to 1995. *J. Oceanogr.* 67 (1), 59–75.
- Ohno, T., 2002. Fluorescence inner-filtering correction for determining the humification index of dissolved organic matter. *Environ. Sci. Technol.* 36 (4), 742–746.
- Osburn, C.L., Wigdahl, C.R., Fritz, S.C., Saros, J.E., 2011. Dissolved organic matter composition and photoreactivity in prairie lakes of the US Great Plains. *Limnol. Oceanogr.* 56 (6), 2371–2390.
- Pang, C., Li, K., Hu, D., 2016. Net accumulation of suspended sediment and its seasonal variability dominated by shelf circulation in the Yellow and East China seas. *Mar. Geol.* 371, 33–43.
- Parinet, B., Lhote, A., Legube, B., 2004. Principal component analysis: an appropriate tool for water quality evaluation and management—application to a tropical lake system. *Ecol. Model.* 178 (3), 295–311.
- Park, Y., Cho, K.H., Park, J., Cha, S.M., Kim, J.H., 2015. Development of early-warning protocol for predicting chlorophyll-a concentration using machine learning models in freshwater and estuarine reservoirs, Korea. *Sci. Total Environ.* 502, 31–41.
- Parkhomenko, A.V., Kuftarkova, E.A., Subbotin, A.A., Gubanov, V.I., 2003. Results of hydrochemical monitoring of Sevastopol Black Sea's offshore waters. *J. Coast. Res.* 907–911.
- Penna, N., Capellacci, S., Ricci, F., 2004. The influence of the Po River discharge on phytoplankton bloom dynamics along the coastline of Pesaro (Italy) in the Adriatic Sea. *Mar. Pollut. Bull.* 48 (3), 321–326.
- Pettine, M., Casentini, B., Fazi, S., Giovanardi, F., Pagnotta, R., 2007. A revisitation of TRIX for trophic status assessment in the light of the European Water Framework Directive: application to Italian coastal waters. *Mar. Pollut. Bull.* 54 (9), 1413–1426.
- Picard, R.R., Cook, R.D., 1984. Cross-validation of regression models. *J. Am. Stat. Assoc.* 79 (387), 575–583.
- Piccini, C., Conde, D., Pernthaler, J., Sommaruga, R., 2009. Alteration of chromophoric dissolved organic matter by solar UV radiation causes rapid changes in bacterial community composition. *Photochem. Photobiol. Sci.* 8 (9), 1321–1328.
- Pinto, U., Maheshwari, B., Shrestha, S., Morris, C., 2012. Modelling eutrophication and microbial risks in peri-urban river systems using discriminant function analysis. *Water Res.* 46 (19), 6476–6488.
- Primpas, I., Karydis, M., 2011. Scaling the trophic index (TRIX) in oligotrophic marine environments. *Environ. Monit. Assess.* 178 (1–4), 257–269.
- Primpas, I., Karydis, M., Tsirtsis, G., 2008. Assessment of clustering algorithms in discriminating eutrophic levels in coastal waters. *Global NEST J.* 10 (3), 359–365.
- Primpas, I., Tsirtsis, G., Karydis, M., Kokkoris, G.D., 2010. Principal component analysis: development of a multivariate index for assessing eutrophication according to the European water framework directive. *Ecol. Indic.* 10 (2), 178–183.
- Qu, B., Song, J., Yuan, H., Li, X., Li, N., Duan, L., Chen, X., Lu, X., 2015. Summer carbonate chemistry dynamics in the Southern Yellow Sea and the East China Sea: regional variations and controls. *Cont. Shelf Res.* 111, 250–261.
- Ribeiro, R., Torgo, L., 2008. A comparative study on predicting algae blooms in Douro River, Portugal. *Ecol. Model.* 212 (1), 86–91.
- Romera-Castillo, C., Sarmento, H., Álvarez-Salgado, X.A., Gasol, J.M., Marrasé, C., 2011. Net production and consumption of fluorescent colored dissolved organic matter by natural bacterial assemblages growing on marine phytoplankton exudates. *Appl.*



- Environ. Microbiol. 77 (21), 7490–7498.
- Sajan, K.S., Kumar, V., Tyagi, B., 2015. Genetic algorithm based support vector machine for on-line voltage stability monitoring. Int. J. Electr. Power Energy Syst. 73, 200–208.
- Schmoker, C., Russo, F., Drillet, G., Trottet, A., Mahjoub, M.S., Hsiao, S.H., Larsen, O., Tun, K., Calbet, A., 2016. Effects of eutrophication on the planktonic food web dynamics of marine coastal ecosystems: the case study of two tropical inlets. Mar. Environ. Res. 119, 176–188.
- Selman, M., Greenhalgh, S., Diaz, R., Sugg, Z., 2008. Eutrophication and Hypoxia in Coastal Areas: A Global Assessment of the State of Knowledge. 284. World Resources Institute. 1–6.
- Shahrban, M., Etemad-Shahidi, A., 2010. Classification of the Caspian Sea coastal waters based on trophic index and numerical analysis. Environ. Monit. Assess. 164 (1–4), 349–356.
- Shi, W., Wang, M., 2012. Satellite views of the Bohai Sea, Yellow Sea, and East China Sea. Prog. Oceanogr. 104, 30–45.
- Siswanto, E., Nakata, H., Matsuoka, Y., Tanaka, K., Kiyomoto, Y., Okamura, K., Zhu, J., Ishizaka, J., 2008. The long-term freshening and nutrient increases in summer surface water in the northern East China Sea in relation to Changjiang discharge variation. J. Geophys. Res. Oceans 113 (C10).
- Slonecker, E.T., Jones, D.K., Pellerin, B.A., 2016. The new Landsat 8 potential for remote sensing of colored dissolved organic matter (CDOM). Mar. Pollut. Bull. 107 (2), 518–527.
- Smola, A.J., Schölkopf, B., 2004. A tutorial on support vector regression. Stat. Comput. 14 (3), 199–222.
- Stedmon, C.A., Markager, S., 2005a. Resolving the variability in dissolved organic matter fluorescence in a temperate estuary and its catchment using PARAFAC analysis. Limnol. Oceanogr. 50 (2), 686–697.
- Stedmon, C.A., Markager, S., 2005b. Tracing the production and degradation of autochthonous fractions of dissolved organic matter by fluorescence analysis. Limnol. Oceanogr. 50 (5), 1415–1426.
- Stedmon, C.A., Markager, S., Bro, R., 2003. Tracing dissolved organic matter in aquatic environments using a new approach to fluorescence spectroscopy. Mar. Chem. 82 (3), 239–254.
- Stedmon, C.A., Markager, S., Tranvik, L., Kronberg, L., Slätis, T., Martensen, W., 2007a. Photochemical production of ammonium and transformation of dissolved organic matter in the Baltic Sea. Mar. Chem. 104 (3), 227–240.
- Stedmon, C.A., Thomas, D.N., Granskog, M., Kaartokallio, H., Papadimitriou, S., Kuosa, H., 2007b. Characteristics of dissolved organic matter in Baltic coastal sea ice: allochthonous or autochthonous origins? Environ. Sci. Technol. 41 (21), 7273–7279.
- Stefanou, P., Tsirtsis, G., Karydis, M., 2000. Nutrient scaling for assessing eutrophication: the development of a simulated normal distribution. Ecol. Appl. 10 (1), 303–309.
- Suárez Sánchez, A., García Nieto, P.J., Riesgo Fernández, P., del Coz Díaz, J.J., Iglesias-Rodríguez, F.J., 2011. Application of an SVM-based regression model to the air quality study at local scale in the Avilés urban area (Spain). Math. Comput. Model. 54 (5), 1453–1466.
- Subasi, A., 2013. Classification of EMG signals using PSO optimized SVM for diagnosis of neuromuscular disorders. Comput. Biol. Med. 43 (5), 576–586.
- Sun, S., Zhang, F., Li, C., Wang, S., Wang, M., Tao, Z., Wang, Y., Zhang, G., Sun, X., 2015. Breeding places, population dynamics, and distribution of the giant jellyfish *Nemopilema nomurai* (Scyphozoa: Rhizostomeae) in the Yellow Sea and the East China Sea. Hydrobiologia 754 (1), 59–74.
- Taboada, J., Matías, J.M., Ordóñez, C., García, P.J., 2007. Creating a quality map of a slate deposit using support vector machines. J. Comput. Appl. Math. 204 (1), 84–94.
- Teixeira, M.R., Rosa, M.J., 2006. Comparing dissolved air flotation and conventional sedimentation to remove cyanobacterial cells of *Microcystis aeruginosa*: part I: the key operating conditions. Sep. Purif. Technol. 52 (1), 84–94.
- Tekile, A., Kim, I., Kim, J., 2015. Mini-review on river eutrophication and bottom improvement techniques, with special emphasis on the Nakdong River. J. Environ. Sci. 30, 113–121.
- Tsirtsis, G., Karydis, M., 1999. Application of discriminant analysis for water quality assessment in the Aegean. In: Proceedings of the 6th Conference on Environmental Science and Technology. Univ. of the Aegean.
- Vahatalo, A.V., Jarvinen, M., 2007. Photochemically produced bioavailable nitrogen from biologically recalcitrant dissolved organic matter stimulates production of a nitrogen-limited microbial food web in the Baltic Sea. Limnol. Oceanogr. 52 (1), 132.
- Vapnik, V.N., 1995. A Nature of Statistical Learning Theory. Springer-Verlag, New York.
- Vollenweider, R.A., Giovanardi, F., Montanari, G., Rinaldi, A., 1998. Characterization of the trophic conditions of marine coastal waters, with special reference to the NW Adriatic Sea: proposal for a trophic scale, turbidity and generalized water quality index. Environmetrics 9 (3), 329–357.
- Wang, Y., Zhang, D., Shen, Z., Chen, J., Feng, C., 2014. Characterization and spatial distribution variability of chromophoric dissolved organic matter (CDOM) in the Yangtze Estuary. Chemosphere 95, 353–362.
- Wang, L., Zhang, G., Zhu, Z., Li, J., Liu, S., Ye, W., Han, Y., 2016. Distribution and sea-to-air flux of nitrous oxide in the East China Sea during the summer of 2013. Cont. Shelf Res. 123, 99–110.
- Wei, Q.S., Yu, Z.G., Wang, B.D., Fu, M.Z., Xia, C.S., Liu, L., Ge, R.F., Wang, H.W., Zhan, R., 2016. Coupling of the spatial-temporal distributions of nutrients and physical conditions in the southern Yellow Sea. J. Mar. Syst. 156, 30–45.
- Wheeler, P.A., Huyer, A., Fleischbein, J., 2003. Cold halocline, increased nutrients and higher chlorophyll off Oregon in 2002. Geophys. Res. Lett. 30 (15).
- Wu, J., Zhang, H., He, P.J., Shao, L.M., 2011. Insight into the heavy metal binding potential of dissolved organic matter in MSW leachate using EEM quenching combined with PARAFAC analysis. Water Res. 45 (4), 1711–1719.
- Xie, L.Q., Xie, P., Tang, H.J., 2003. Enhancement of dissolved phosphorus release from sediment to lake water by *Microcystis* blooms—an enclosure experiment in a hyper-eutrophic, subtropical Chinese lake. Environ. Pollut. 122 (3), 391–399.
- Xiong, J., Wang, X.C., Zhang, Q., Duan, R., Wang, N., 2016. Characteristics of a landscape water with high salinity in a coastal city of China and measures for eutrophication control. Ecol. Indic. 61, 268–273.
- Xu, F.L., Tao, S., Dawson, R.W., Li, B.G., 2001. A GIS-based method of lake eutrophication assessment. Ecol. Model. 144 (2), 231–244.
- Xu, Y., Ma, C., Liu, Q., Xi, B., Qian, G., Zhang, D., Huo, S., 2015. Method to predict key factors affecting lake eutrophication—a new approach based on support vector regression model. Int. Biodeterior. Biodegrad. 102, 308–315.
- Xue, X., Landis, A.E., 2010. Eutrophication potential of food consumption patterns. Environ. Sci. Technol. 44 (16), 6450–6456.
- Yamashita, Y., Tanoue, E., 2003. Chemical characterization of protein-like fluorophores in DOM in relation to aromatic amino acids. Mar. Chem. 82 (3), 255–271.
- Yamashita, Y., Jaffé, R., Maie, N., Tanoue, E., 2008. Assessing the dynamics of dissolved organic matter (DOM) in coastal environments by excitation emission matrix fluorescence and parallel factor analysis (EEM-PARAFAC). Limnol. Oceanogr. 53 (5), 1900–1908.
- Yamashita, Y., Maie, N., Briceño, H., Jaffé, R., 2010. Optical characterization of dissolved organic matter in tropical rivers of the Guayana Shield, Venezuela. J. Geophys. Res. Biogeosci. 115 (G1).
- Yamashita, Y., Panton, A., Mahaffey, C., Jaffé, R., 2011. Assessing the spatial and temporal variability of dissolved organic matter in Liverpool Bay using excitation–emission matrix fluorescence and parallel factor analysis. Ocean Dyn. 61 (5), 569–579.
- Yamashita, Y., Boyer, J.N., Jaffe, R., 2013. Evaluating the distribution of terrestrial dissolved organic matter in a complex coastal ecosystem using fluorescence spectroscopy. Cont. Shelf Res. 66, 136–144.
- Yan, H.Y., Zhang, X.R., Dong, J.H., Shang, M.S., Shan, K., Wu, D., Yuan, Y., Wang, X., Meng, H., Huang, Y., Wang, G.Y., 2016. Spatial and temporal relation rule acquisition of eutrophication in Da'ning River based on rough set theory. Ecol. Indic. 66, 180–189.
- Yao, X., Zhang, Y., Zhu, G., Qin, B., Feng, L., Cai, L., Gao, G., 2011. Resolving the variability of CDOM fluorescence to differentiate the sources and fate of DOM in Lake Taihu and its tributaries. Chemosphere 82 (2), 145–155.
- Yoon, B.I., Woo, S.B., 2015. The along-channel salinity distribution and its response to river discharge in tidally-dominated Han River Estuary, South Korea. Process. Eng. 116, 763–770.
- Yuan, D., Zhu, J., Li, C., Hu, D., 2008. Cross-shelf circulation in the Yellow and East China seas indicated by MODIS satellite observations. J. Mar. Syst. 70 (1), 134–149.
- Zhang, Y., van Dijk, M.A., Liu, M., Zhu, G., Qin, B., 2009. The contribution of phytoplankton degradation to chromophoric dissolved organic matter (CDOM) in eutrophic shallow lakes: field and experimental evidence. Water Res. 43 (18), 4685–4697.
- Zhang, M., Tang, J., Dong, Q., Song, Q., Ding, J., 2010a. Retrieval of total suspended matter concentration in the Yellow and East China seas from MODIS imagery. Remote Sens. Environ. 114 (2), 392–403.
- Zhang, Y., Zhang, E., Yin, Y., Van Dijk, M.A., Feng, L.Q., Shi, Z.Q., Liu, M.L., Qin, B., 2010b. Characteristics and sources of chromophoric dissolved organic matter in lakes of the Yungui Plateau, China, differing in trophic state and altitude. Limnol. Oceanogr. 55 (6), 2645–2659.
- Zhang, Y., Yin, Y., Feng, L., Zhu, G., Shi, Z., Liu, X., Zhang, Y., 2011. Characterizing chromophoric dissolved organic matter in Lake Tianmuhu and its catchment basin using excitation–emission matrix fluorescence and parallel factor analysis. Water Res. 45 (16), 5110–5122.
- Zhang, L., Wang, S., Wu, Z., 2014. Coupling effect of pH and dissolved oxygen in water column on nitrogen release at water–sediment interface of Erhai Lake, China. Estuar. Coast. Shelf Sci. 149, 178–186.
- Zhang, G., Bai, J., Xi, M., Zhao, Q., Lu, Q., Jia, J., 2016. Soil quality assessment of coastal wetlands in the Yellow River Delta of China based on the minimum data set. Ecol. Indic. 66, 458–466.
- Zhong, J., Peter, W.T., Wang, D., 2016. Novel Bayesian inference on optimal parameters of support vector machines and its application to industrial survey data classification. Neurocomputing.
- Zhou, M.J., Shen, Z.L., Yu, R.C., 2008. Responses of a coastal phytoplankton community to increased nutrient input from the Changjiang (Yangtze) River. Cont. Shelf Res. 28 (12), 1483–1489.
- Zhou, C., Yin, K., Cao, Y., Ahmed, B., 2016a. Application of time series analysis and PSO-SVM model in predicting the Bazimen landslide in the Three Gorges Reservoir, China. Eng. Geol. 204, 108–120.
- Zhou, Y., Zhang, Y., Jeppesen, E., Murphy, K.R., Shi, K., Liu, M., Liu, X., Zhu, G., 2016b. Inflow rate-driven changes in the composition and dynamics of chromophoric dissolved organic matter in a large drinking water lake. Water Res. 100, 211–221.