



Forecasting riverine total nitrogen loads using wavelet analysis and support vector regression combination model in an agricultural watershed

Xiaoliang Ji¹ · Jun Lu^{1,2}

Received: 10 January 2018 / Accepted: 3 July 2018 / Published online: 7 July 2018
© Springer-Verlag GmbH Germany, part of Springer Nature 2018

Abstract

In the context of non-point source pollution management and algal blooms control, the reliable nutrient forecasting is of critical importance. Considering the highly stochastic, non-linear, and non-stationary natures involved in riverine total nitrogen (TN) load time series data, some traditional statistical and artificial intelligence models are inherently unable to give accurate nutrient forecasts due to their mechanism and structure characteristics. In this study, based on the wavelet analysis (WA) and support vector regression (SVR), a promising combined WA-SVR model was proposed for forecasting riverine TN loads. The data pre-processing tool WA was employed to decompose the time series data of riverine TN load for revealing its dominator. Subsequently, all wavelet components were used as inputs to SVR for WA-SVR model. The continuous riverine TN loads during 2004–2012 in the ChangLe River watershed of eastern China were estimated by using a calibrated Load Estimator model. Performance criteria, namely, determination coefficient (R^2), Nash-Sutcliffe model efficiency (NS), and mean square error (MSE) were applied to assess the performance of the developed models. The effects of different mother wavelets on the efficiency of the conjunction model were investigated. The results demonstrated that the mother wavelet played a crucial role for the successful implementation of the WA-SVR model. Among the 23 selected mother wavelet functions, dmey wavelet performed best in forecasting the daily and monthly TN loads. Furthermore, the performance of the optimal WA-SVR model was compared with that of single SVR model without wavelet decomposition. The comparison indicated that the hybrid model provided better accuracy than that of single SVR model. For daily riverine TN loads, the R^2 , NS, and MSE values of WA-SVR model during the test stage were 0.9699, 0.9658, and 0.4885×10^7 kg/day, respectively. For monthly riverine TN loads, the R^2 , NS, and MSE values of the model during the test stage were 0.9163, 0.9159, and 0.3237×10^{10} kg/month, respectively. The overall results strongly suggested that the combined WA-SVR method can successfully forecast riverine TN loads in agricultural watersheds.

Keywords Forecasting · Riverine TN loads · Non-point source pollution · Wavelet analysis · Support vector regression model

Introduction

Surface water quality impairment is one of the most prevalent environmental problems worldwide (Houser and

Richardson 2010; Li and Zhang 2010; Morse and Wollheim 2014). In the last two decades, the increasing nutrients in the receiving water have been commonly attributed to non-point source (NPS) pollution from agricultural and human living systems (Wang et al. 2011). The excess nitrogen and phosphorus in rivers decrease the water quality, degrade aquatic ecosystem health, and induce anoxia and harmful algal blooms in several fresh water and coastal ecosystems (Bowes et al. 2010; Gao and Zhang 2010; Howarth et al. 2012; Ryusuke et al. 2002). The accurate nutrient load forecasts are particularly important to inform future policy and management decisions in terms of prioritizing water quality management and algal blooms control. But now, riverine total nitrogen (TN) load time series forecasting is still one of

Responsible editor: Marcus Schulz

✉ Jun Lu
jlu@zju.edu.cn

¹ College of Environment and Natural Resources, Zhejiang University, Hangzhou 310058, Zhejiang Province, China

² China Ministry of Education Key Lab of Environment Remediation and Ecological Health, Zhejiang University, Hangzhou 310058, China

the most difficult and challenging issues for water resource managers and government agencies because riverine TN loads are closely related to anthropogenic activities as well as dynamic, non-linear, and complex natural processes involved in NPS pollution.

Complex mechanistic models such as Hydrological Simulation Program-Fortran (HSPF), Soil and Water Assessment Tool (SWAT), and Agricultural Non-Point Source Pollution Model (AGNPS) have been developed in the past few decades to obtain quantitative information of NPS pollution loads (Arnold and Fohrer 2005; Borah and Bera 2004; Donigian et al. 1984). These watershed scale models have been widely used for forecasting nutrient loads under various climate, management, and pollutant loads scenarios. For example, Shen et al. (2014) used SWAT model to simulate the spatial and temporal distribution of NPS pollution loads in the Three Gorges Reservoir Region (China). The results exhibited that the NPS pollution loads in the western area were the highest and an upward trend existed in recent years. Coffey and Line (1998) predicted nutrient loads export from dairies to Cane Creek Reservoir (USA) through AGNPS model, and the results indicated that annual TN and total phosphorus (TP) export from entire farms varied from 1.0 to 3.7 kg/ha and 0.7 to 1.8 kg/ha, respectively. Chou et al. (2007) employed HSPF model to estimate watershed pollutant loads from NPS in the Feitsui Reservoir watershed (China), and found that NPS pollution contributed ~85% of the annual average TP loads into the reservoir. However, these models are unable to provide defensible and quantifiable future nutrient loads which are required for developing early warning systems.

Traditionally, researchers have utilized statistical models on the basis of time series analysis (e.g., autoregressive, autoregressive moving average, and autoregressive integrated moving average) to cope with forecasting problems (Belayneh et al. 2014; Huang et al. 2004; Kothyari and Singh 1999; Mishra and Desai 2005; Wu et al. 2009a). Nevertheless, stochastic models have linear structures that prevent these models from dealing with highly non-linear processes. In the past two decades, artificial neural network (ANN) has gained significant attention and been widely used in various configuration to perform a range of tasks including forecasting, classification, pattern recognition, data mining, and non-linear process modeling (Chen and Chau 2016; Nabavi-Pelesaraei et al. 2017; Sefeedpari et al. 2016; Wang et al. 2014; Wu et al. 2009b; Yoon et al. 2011). It should be noted that ANN has shortcomings inherent in its architecture, like over-fitting, slow training speed, and vulnerability to being trapped in the local optimum (Chau 2017). The support vector regress (SVR) method, presented by Vapnik (1995, 1998), has been attracting a great deal of interest because of the following advantages: (1)

simultaneous minimization of model complexity and prediction error since the use of kernel trick in building expert knowledge, (2) good generalization ability can be achieved due to the adoption of structural risk minimization (SRM) principle (Chau and Wu 2010), (3) good performance can be obtained with relatively small data sets, and (4) prevention of over-fitting problem which is the critical drawback of ANN. SVR model utilizes appropriate kernel function to map the original data sets from the input space into a higher dimensional space, from where the non-linear problem changed into linear problem. Consequently, SVR model is capable of modeling non-linear data. Recently, the SVR model has been widely applied to various scenarios, such as stream flow forecasting (Lin et al. 2006), water level prediction (Khan and Coulibaly 2006; Yoon et al. 2011), and water quality parameters modeling (Noori et al. 2015).

The aforementioned data-driven models are advantageous because they require less information, do not necessitate a full understanding of underlying processes in mathematical forms, and are easy to implement. Furthermore, they are capable of addressing forecasting problems. Yet, their ability is limited when dealing with non-stationary data (Adamowski and Sun 2010; Cannas et al. 2006). To overcome this drawback, some researchers explored wavelet analysis (WA) in hydrologic forecasting (Belayneh et al. 2014). WA is recognized as an effective tool for analyzing trends, periodicities, and variations in time series (Chou and Wang 2002; Lu 2002; Partal and Kucuk 2006; Simith et al. 1998). A non-stationary time series signal can be decomposed into several different resolution levels by WA to reveal the useful information involved in original time series. Because of this, WA has a significant advantage in handling non-stationary processes. In addition, the wavelet method is robust since any potentially parametric testing procedures or erroneous assumptions are included (Kisi and Cimen 2011). WA is often used as a data pre-processing tool for revealing hidden information and capturing both the periodic and chaotic behaviors of a time series otherwise not captured by other signal processing techniques (Adamowski and Sun 2010; Shoaib et al. 2014). In recent years, several studies have demonstrated the successful applications of WA combined with data-driven models (e.g., ANN and SVR) in hydrologic forecasting. For example, Wang and Ding (2003) first explored the ability of ANN model coupled with WA (WA-ANN) in short and long-term prediction of hydrological time series. They found that the suggested WA-ANN model absorbed several advantages of WA and ANN and was capable of forecasting daily river discharge with high accuracy. Similarly, Liu et al. (2013) proposed a WA-ANN model to predict suspended sediment concentration (SSC) in the Kuye River, a representative hyper-concentrated river in the middle Yellow River catchments of China. Results revealed that the proposed model

obtained better SSC predicts than ANN model or sediment rating curve (SRC) model in a hyper-concentrated river setting, with highly non-linear and non-stationary time series. Olyaie et al. (2015) compared the accuracy of three different soft computing methods, namely ANN, adaptive neuro-fuzzy inference system (ANFIS), WA-ANN, and conventional SRC approach for estimating the daily SSC in two gauging stations in USA. Results indicated that the WA-ANN was the most accurate model in SSC estimation in comparison to other models. Kisi and Cimen (2011) used WA and SVR conjunction model (WA-SVR) to illustrate that the forecast accuracy was higher in monthly stream flow forecasting compared with the single SVR model. Kalteh (2013) applied two hybrid models (WA-ANN and WA-SVR) for monthly river flow prediction and investigated their accuracy, indicating the wavelet decomposition and artificial intelligence (AI) combination models could produce better efficiency than conventional forecasting models. Results also demonstrated that the WA-SVR model outperformed the WA-ANN model. Despite the applications of coupled WA-ANN or WA-SVR model to a multitude of hydrologic cases, no attempt has been performed on NPS load forecasting with the WA-ANN or WA-SVR model in highly agricultural watersheds for guiding NPS pollution remediation efforts and algal bloom control until now. Rather, only few mother wavelets were included in the abovementioned studies, which rarely consider the effect of different mother wavelets on accuracy of the combined models. Thus, forecasting of NPS loads by using WA-SVR model and investigation of the different mother wavelets' impact on established model performance are required in terms of NPS pollution management.

NPS pollution processes are influenced by both natural and human factors; as such, riverine TN load time series often display highly stochastic, non-linear, and non-monotonic characteristics. Therefore, the supporting assumptions of TN load data cannot match well with the traditional stochastic and single ANN or SVR model. This might yield invalid results when these data-driven models are employed to forecast riverine TN loads. Through combining WA with SVR model, the WA-SVR model has the merits of both techniques and thus can effectively address the stochastic, non-linear, and non-stationary issues. For this reason, WA-SVR model was first proposed for pollutant load estimation in this study, with the purpose to overcome the difficulties associated with riverine TN load forecasting. Specifically, the main objectives of this study were to (1) obtain the riverine TN load time series in the ChangLe River from the years 2004 to 2012, (2) apply the WA-SVR hybrid model to forecast the short-term (daily) and long-term (monthly) riverine TN loads, (3) examine the effects of mother wavelets on the performance of constructed models, and (4) validate the performance of WA-SVR model by comparing it

with that of single SVR model. It is expected that the results can provide accurate forecasts of riverine TN loads and early warning for water quality agencies and thus facilitate algal bloom control and water pollution remediation.

Study area and data collection

Study area

The ChangLe River watershed (29° 27' 98"–29° 35' 12" N, 120° 35' 56"–120° 49' 03" E) has a drainage area of 864 km² in Zhejiang Province that lies in eastern China (Fig. 1). This region is characterized by subtropical monsoon climate, receiving 1228 mm of annual precipitation with 70% falls during summer (April–September). The average monthly air temperature ranges from 4 °C (in January) to 28 °C (in July). The river traverses a total distance of 70.5 km before merging with the Cao-E River, which ultimately drains into the East China Sea. The width of the river varies from 40 to 70 m, with an average value of 55 m. The ChangLe River watershed is located in the most intensive agricultural regions in China (Chen et al. 2011, 2013). The primary land use categories are woodland and farmland, accounting for around 48 and 42% of the entire watershed, respectively. Approximately 92% of the river water comes from catchment runoff, and the remainder originates as the headwater streams (Nanshan Reservoir) drainage. The ChangLe River watershed was selected because (1) this region represents a typically agricultural watershed in eastern China. The NPS pollution is the primary source of nitrogen in the river systems, and (2) the long period of observations for TN concentration and stream flow are available.

Data collection

River water quality samples were collected every month ($n = 106$ sampling times total) at Yazhi station (Fig. 1) from January 2004 to December 2012. Water samples were collected at a depth of ~20 cm in the middle of the river and stored in 0.5 L high-density polyethylene bottles. Subsequently, the samples were acidified with sulfuric acid (H₂SO₄) and immediately chilled by placing ice packs in the cold closet. The concentration of TN in water samples was determined within 24 h of sampling campaign by alkaline potassium persulfate method (State Environment Protection Bureau of China 2002). The continuous daily stream discharges at Yazhi station were obtained from the hydrological bureau of Zhejiang Province, China. The time series of stream flow and observed TN concentrations are illustrated in Fig. 2.

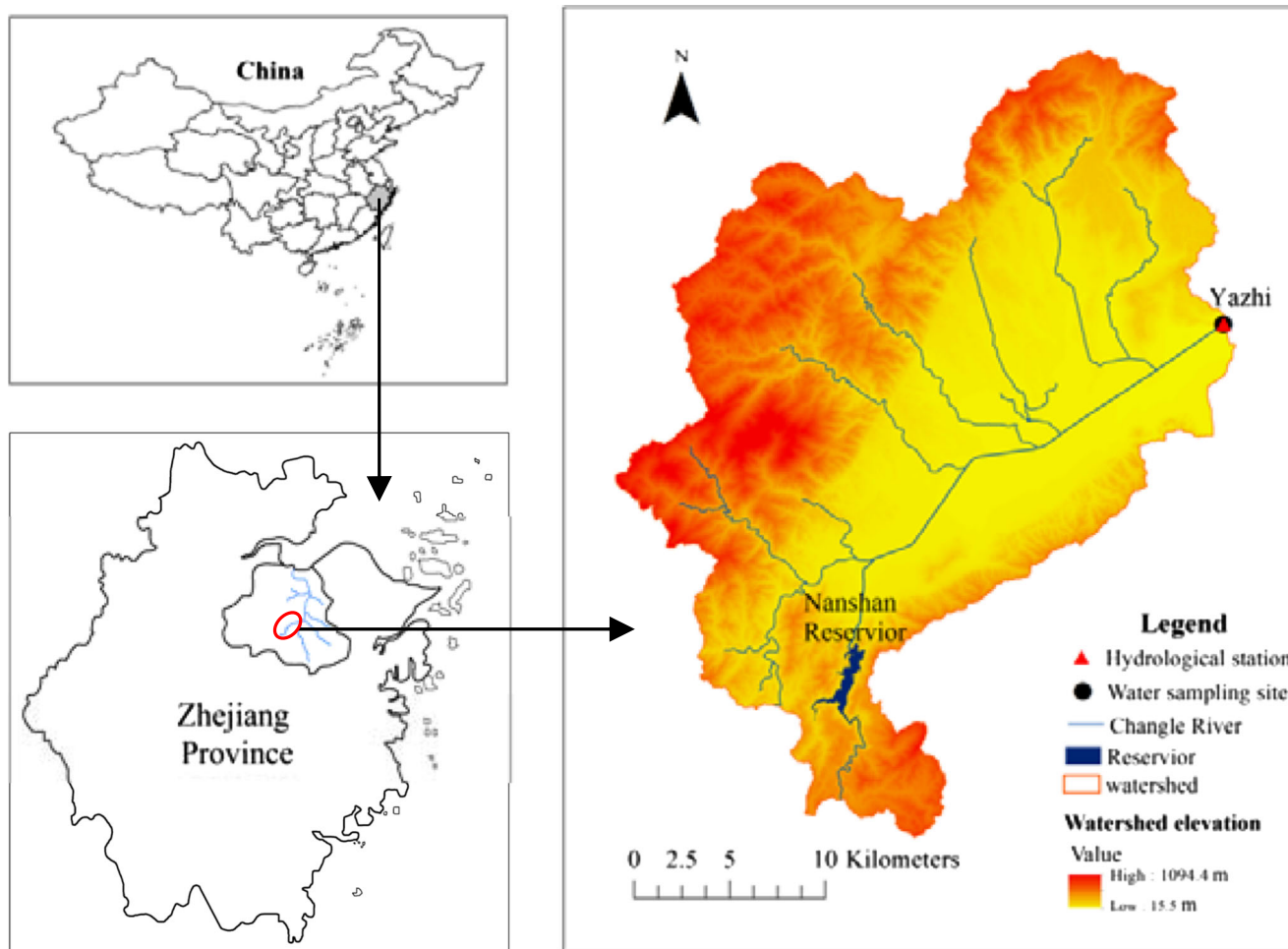


Fig. 1 Location of ChangLe River watershed in China and Zhejiang Province and the river hydrology and water quality sampling site

Model development

Continuous time series data of riverine TN loads

Generally, river discharge is daily recorded by hydrology monitoring stations. Nevertheless, monthly or bimonthly

sampling is often employed as the standard sampling frequency for water quality monitoring since the limitations of cost and time. Given this dilemma, the Load Estimator (LOADEST) model was presented to estimate daily nutrient loads over time from relatively infrequent discrete water quality samples and high frequency discharge measurements

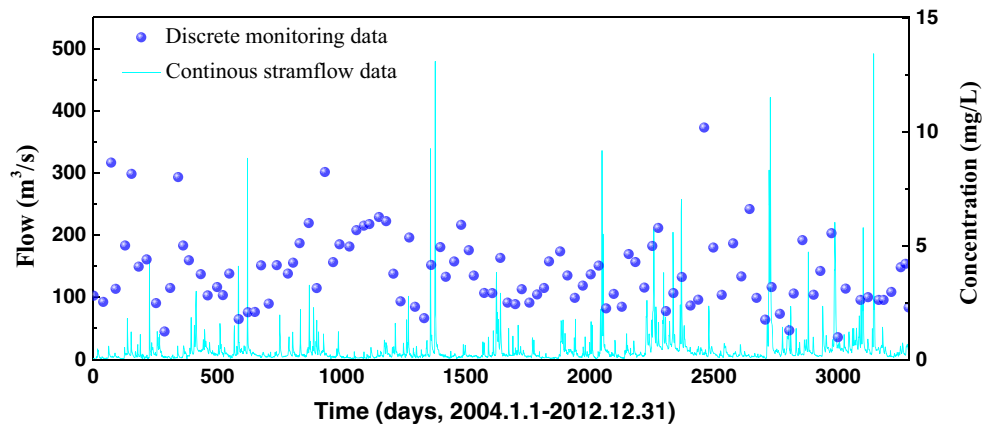


Fig. 2 Stream flow and observed TN concentrations in the ChangLe River during 2004–2012

(Runkel et al. 2004). This method sufficiently considers the correlation between the pollutant loads and concentration, discharge, and time. It has many advantages such as low data dependency, stable result, and ease of operation. Therefore,

$$\ln(\text{Load}) = a_0 + a_1 (\ln(Q) \ln(Q)_c) + a_2 (\ln(Q) - \ln(Q)_c)^2 + a_3 \sin(2\pi (t - t_c)) + a_4 \cos(2\pi (t - t_c)) + a_5 (t - t_c) + a_6 (t - t_c)^2 \tag{1}$$

where \ln is the natural logarithm function; Load is the measured riverine TN loads (kg/day); Q is the measured daily average discharge (m³/s); $\ln(Q)_c$ is a centering term of the study period (a constant) to ensure that the linear and quadratic discharge terms are independent (m³/s); t is time in decimal days; t_c is a centering time of the study period (a constant) to ensure that the linear and quadratic time terms are independent; $\alpha_0 \dots \alpha_6$ are the fitted parameters in the multiple regression model; α_1 and α_2 describe the relation between loads and discharge; α_3 and α_4 describe seasonal variation in loads data; α_5 and α_6 describe the relation between loads and time.

SVR model

The SVR model was chosen since it exhibited better results than the ANN model, as indicated by the results of previous works (He et al. 2014; Kalteh 2013; Kisi 2012; Mohammadpour et al. 2015). The SVR model, which employed the principle of SRM, was developed on the basis of machine-learning process (Mohammadpour et al. 2015). A more detailed information on SVR can be found in published literature (Cristianine and Taylor 2000; Raghavendra and Deka 2014; Vapnik 1998). Thereby, only a short explanation regarding SVR was given as below, and the network architecture of SVR model is shown in Fig. 3.

Consider a data set $\{(x_i, y_i), i = 1, \dots, n\}$, where x_i refers to the input vector, y_i is the desired value, and n refers to the total number of data patterns, the regression function of SVR is formulated as follows:

$$f(x) = w \cdot \phi(x) + b \tag{2}$$

where w and b denote the weight vector and bias, respectively, and $\phi(x)$ denotes the non-linear mapping function. Parameters w and b are estimated by solving the following optimization problem:

$$\text{Minimize : } \frac{1}{2} \|w\|^2 + C \sum_1^n (\xi_i + \xi_i^*) \tag{3}$$

$$\text{Subject to } \begin{aligned} y_i - w \cdot \phi(x) - b &\leq \varepsilon + \xi_i \\ w \cdot \phi(x) + b - y_i &\leq \varepsilon + \xi_i^* \\ \xi_i &\geq 0, \xi_i^* &\geq 0 \end{aligned} \tag{4}$$

LOADEST model software was adopted in this study. LOADEST has 11 regression models, and one of the most common models is listed as follows:

where ξ_i and ξ_i^* are slack variables, and C is the penalty parameter.

Equation (3) can be solved by using Lagrangian theory and imposing Karush-Kuhn-Tucker optimality condition.

The final form of the SVR model can be expressed as the following:

$$f(x) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) \cdot K(x_i, x) + b \tag{5}$$

There are four possible choices for the kernel function of SVR model, involved sigmoid, linear, polynomial, and radial basis function (RBF). RBF is by far the most popular kernel function (Zhang et al. 2008) due to the following reasons: (1) RBF has the ability to model non-linear relationships by mapping inputs vectors into a high-dimensional feature space in a non-linear fashion; (2) the number of tuning parameters in RBF is fewer than those in sigmoid and polynomial kernels, making RBF more easy-to-use (Keerthi and Lin 2001); and (3) the superior performances of RBF have been demonstrated in numerous studies (Dibike et al. 2001; Keerthi and Lin 2001). Thus, RBF was taken as the kernel function for riverine TN flux forecasting in this study. The RBF is defined as follows:

$$K(x_i, x) = \exp(-\gamma \|x_i - x\|^2) \tag{6}$$

where γ is the adjustable kernel parameter.

Data normalization is a common step prior to processing AI model for scaling raw data into a more usable form for the model to utilize. There are four commonly used methods for performing data normalization, such as min-max, median, logarithmic function, and z-score transformation (Antanasijević et al. 2014; Feng et al. 2015; He et al. 2014; Najah et al. 2014; Ravansalar et al. 2017). The four normalization techniques were tested using the single SVR model for daily TN load forecasting and the values of the performance metrics are illustrated in Fig. 11. Obviously, the model constructed using z-score normalized input values yielded the best performance; as such, the z-score normalization was selected to perform transformation on all time

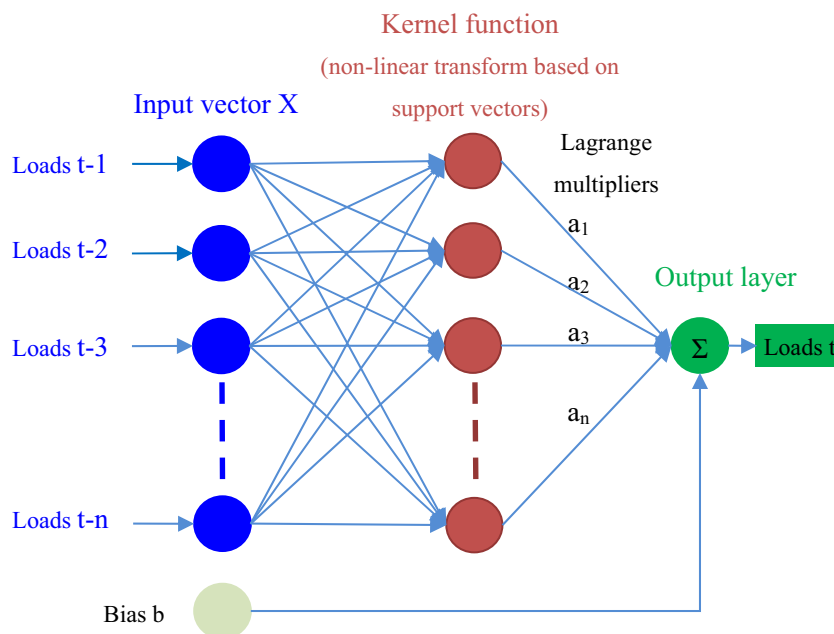


Fig. 3 Network architecture of the SVR model for riverine TN flux forecasting

series data in this study. The mathematical formula for z-score method is the following:

$$x_n = \frac{x - x_{\text{mean}}}{x_{\text{SD}}} \tag{7}$$

where x_n is the normalized value; x is the original value; x_{mean} and x_{SD} are the mean and standard deviation of the original data sets.

Wavelet analysis

WA is regarded as a powerful tool for extracting the useful information in stationary or non-stationary data (Nourani et al. 2014). Short intervals are generally used to capture high frequency information, whereas long intervals are typically used to capture low frequency information inherent to the time series during WA (Shoab et al. 2014). The WA decomposes a time series into approximation (e.g., a3) and details (e.g., d1, d2, and d3) by translating, stretching, or squeezing the wavelet (Seo et al. 2015). It can provide good localization characteristics in both time and frequency domains. Additionally, the choice of the mother wavelet in wavelet decomposition is flexible in terms of the time series properties (Adamowski and Sun 2010). Two main wavelet transforms present, namely, the continuous wavelet transform (CWT) and discrete wavelet transform (DWT) (Wei et al. 2013). The CWT of signal $x(t)$ can be expressed as the following:

$$\text{CWT}_x^\psi(\tau, s) = \frac{1}{\sqrt{|s|}} \int_{-\infty}^{+\infty} x(t) \psi^* \left(\frac{t-\tau}{s} \right) dt \tag{8}$$

where ψ is the mother wavelet; s is the scale parameter; τ is the translation parameter; $*$ is the complex conjugate. The mathematical formula for the DWT is listed in the following equation:

$$\psi_{j,k}(t) = \frac{1}{\sqrt{|s_0^j|}} \psi \left(\frac{t-k\tau_0 s_0^j}{s_0^j} \right) \tag{9}$$

where j and k are the integers that control the wavelet dilation and translation, respectively. $s_0 > 1$ is the fixed dilation step, and τ_0 is the location parameter. The CWT needs a significant amount of computation time since it calculates wavelet coefficients at every possible scale. In addition to time consuming, CWT might produce abundant and invalid data. Compare with CWT, DWT is simpler to implement and requires less computation time and resources (Kalteh 2013). Therefore, DWT was adopted to couple with SVR to develop the hybrid WA-SVR model. The decomposition result of the original signal(s) by DWT satisfies the following: $s = d1 + a1 = d1 + d2 + a2 = d1 + d2 + d3 + a3$ and so on (Shoab et al. 2014).

WA-SVR model

The flowchart of the WA-SVR model for forecasting riverine TN loads is illustrated in Fig. 4. The hybrid model was completed in five steps that were listed as follows: (i) estimation of daily and monthly TN load time series, (ii) selection of suitable input variables and identification of the optimal lag times for the inputs, (iii) implementation of DWT to decompose

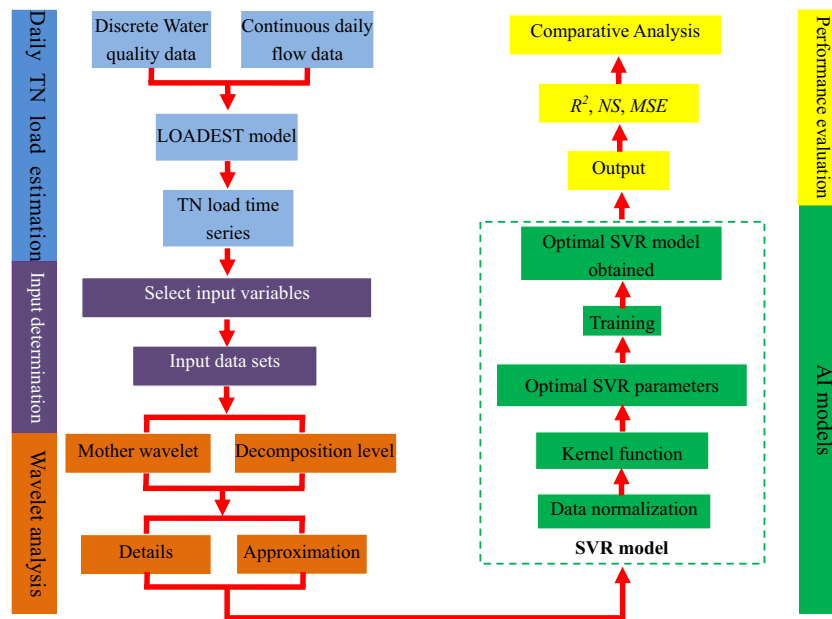


Fig. 4 Flowchart of the WA-SVR model for riverine TN load forecasting

original time series into approximation (ai) and details (d1, d2,..., di), where *i* is the number of decomposition level, (iv) SVR model training and testing wherein approximation and details are used as inputs, and (v) evaluation of the developed WA-SVR model performance.

Model performance and execution

The performances of the established model can be evaluated using the following criteria: efficiency coefficient, determination coefficient (R^2), agreement index, Nash-Sutcliffe model efficiency (NS), mean absolute error, mean square error (MSE), standard error of prediction, mean absolute relative error, and so on. A good examination of model performance should contain at least one absolute error measure (e.g., MSE) and at least one goodness-of-fit or relative error measure (e.g., NS) (Legates and McCabe Jr 1999). Additionally, R^2 , as the widely applied statistical score metric, is often used to estimate the model performance. A model can be adequately assessed by R^2 , NS, and MSE, and thus, the model performances in this study were evaluated using these performance indexes. R^2 indicates the percentage of variability that can be explained by the model (Singh et al. 2011). NS, an indicator of the model fit, is typically used for assessing the performance of hydrological models (Adamowski et al. 2012; Shoaib et al. 2016). MSE measures the absolute error regarding dependent variable forecasting (Legates and McCabe Jr 1999). If the $R^2 = 1$, $NS = 1$, and $MSE = 0$, the model presents the perfect performance. These criteria can be computed as the following:

$$R^2 = \left(\frac{\sum_{i=1}^n (P_i - \bar{P})(O_i - \bar{O})}{\sqrt{\sum_{i=1}^n (P_i - \bar{P})^2 \sum_{i=1}^n (O_i - \bar{O})^2}} \right)^2 \tag{10}$$

$$NS = 1 - \frac{\sum_{i=1}^n (O_i - P_i)^2}{\sum_{i=1}^n (O_i - \bar{O})^2} \tag{11}$$

$$MSE = \frac{1}{n} \sum_{i=1}^n (O_i - P_i)^2 \tag{12}$$

where *n* is the number of observations; O_i and P_i denote the observed and forecasted *i*th values of TN loads, respectively; \bar{O} and \bar{P} are the mean values of observed and forecasted TN loads, respectively.

In this study, LOADEST model software (United States Geological Survey 2004) was used to estimate the continuous daily riverine TN loads. Wave toolbox in MATLAB 2014a (Mathworks Inc., Natick, USA) was used to perform WA. The SVR model was constructed using the LIBSVM toolbox, which is a software package of support vector machine algorithms working in MATLAB environment (Chang and Lin 2011). All the other computations were also conducted in MATLAB.

Results and discussion

Estimation of daily riverine TN loads

LOADEST model was employed to estimate the continuous daily riverine TN loads in ChangLe River from 2004 to 2012 using the discrete monitoring data and continuous daily flow data. The available data for a period of 6 years from 2004 to 2009 was used for calibrating LOADEST model parameters (i.e., $\alpha_0 \dots \alpha_6$), whereas the remaining data for a period of 3 years from 2010 to 2012 was used for validation purpose. The calibration and estimation procedures within LOADEST are based on the adjusted maximum likelihood estimation (AMLE), maximum likelihood estimation (MLE), and least absolute deviation (LAD) methods (Runkel et al. 2004). AMLE and MLE are appropriate when the calibration model residuals are normally distributed. Of the two, AMLE is the suitable method when the calibration data set contains censored data. If the censored data is not contained in calibration data set, MLE is equal to AMLE. LAD can be applied when the residuals are not normally distributed (Runkel et al. 2004). Therefore, the appropriate estimation method can be determined by evaluating the statistical distributions of residuals and the types of calibration data. The probability plot correlation coefficient (PPCC) was calculated to estimate the residual distribution. A PPCC of 1.00 represents a perfect normal probability of residuals (Helsel and Hirsch 2002). In this work, the PPCC value was 0.988 (Table 1), indicating the model residuals were normally distributed. For the above reasons, AMLE was adopted for the calibration and estimation procedures within LOADEST.

The plot of the modeled and observed daily riverine TN loads in the ChangLe River is illustrated in Fig. 5. High R^2 and NS implied that the established LOADEST model can be

successfully applied to estimate daily riverine TN loads in the ChangLe River. The monthly TN loads were calculated via summation of the corresponding daily TN loads.

The performances of WA-SVR and single SVR models were examined on riverine TN loads. To achieve this, the available riverine TN loads data was divided into a training data set from 2004 to 2009 and testing data set from 2010 to 2012. All the models were first trained using the training data set to obtain the optimum values of the parameters (C and γ) and then tested using the testing data set. The statistical parameters of the riverine TN loads data are given in Table 2.

Application of the WA-SVR model

Model inputs

The previous riverine TN loads were considered as input variables in this paper. Appropriate lag times must be determined in advance for developing WA-SVR model. This is because the response of a hydrological system is inherently relied upon their previous states, so the use of time lagged data is necessarily required with aim to encode temporal features of the input data (Shoab et al. 2016). The optimal lag time for the inputs can be identified through the statistical analysis of data series such as auto-correlation function, partial auto-correlation function (PACF), cross-correlation function, and average mutual information. PACF has the capability for removing the dependence on intermediate elements (those within lags interpreted as a regression of time series against its past lagged value) and identifying the extent to which current TN loads is correlated to past days (Yaseen et al. 2016). At present, this approach has been successfully applied in many hydrological

Table 1 The calibrated LOADEST parameters for riverine TN loads in the ChangLe River

Parameters	Values	Std. Dev.	t-ratio ^a	<i>P</i> value ^b	Performance assessment
α_0	7.6854	0.0717	107.14	0.0000	AIC ^c = 0.929
α_1	1.0083	0.0572	17.62	0.0000	SPPC ^d = -41.424
α_2	0.0339	0.0290	1.17	0.2216	PPCC ^e = 0.988
α_3	0.1666	0.0664	2.51	0.0099	
α_4	0.1071	0.0624	1.72	0.0739	
α_5	-0.0116	0.0255	-0.46	0.6313	
α_6	-0.0183	0.0160	-1.14	0.2311	

^a When the absolute value of t-ratio is less than 1, the estimated coefficient is highly uncertain

^b High value of *P* value indicated that the estimated coefficient may not be significantly difference than zero

^c Akaike information criterion

^d Schwarz posterior probability criterion

^e Probability plot correlation coefficient, if PPCC = 1, the residuals in the model follow a perfect normally distributed

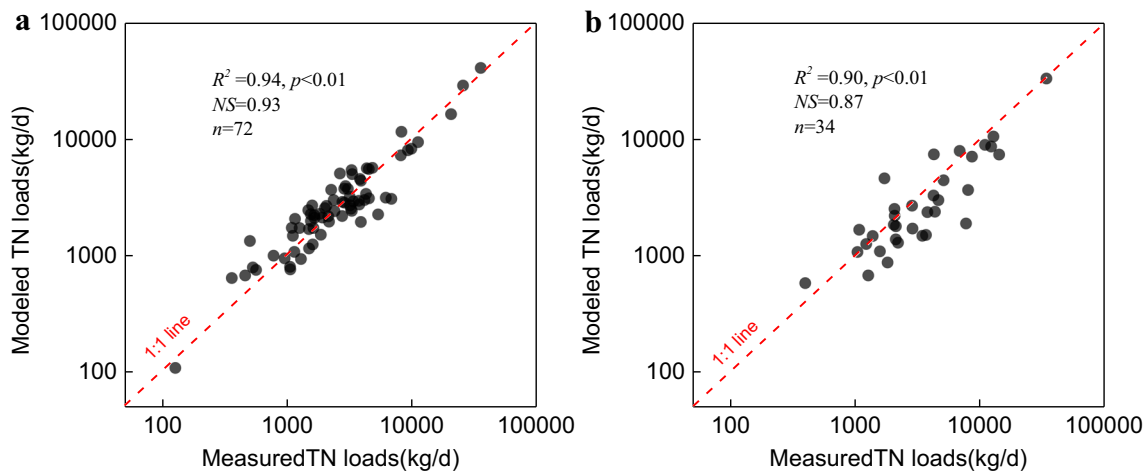


Fig. 5 Plot of the modeled and observed daily riverine TN loads in the ChangLe River in **a** 2004–2009 and **b** 2010–2012

studies including He et al. (2014), Kalteh (2016), Seo et al. (2015), Shiri and Kisi (2010), Yaseen et al. (2016), etc. Hence, the PACF method was employed here to determine the numbers of lags. The PACF values, calculated for a lag range of 1–16, are presented in Fig. 6. It could be seen from the figure that first 4 lags and 10 lags had the most amount of information that could be used to forecast short-term and long-term riverine TN loads, respectively. Consequently, 4 and 10 previous lags for daily and monthly forecasting, respectively, were selected as inputs to the WA-SVR model. It may be argued that the selection of 4 lags and 10 lags was not optimum. Nevertheless, the aim of this paper was limited to evaluation of the WA-SVR model performance and examination of the effects of mother wavelets on developed model efficiency rather than the selection of optimal inputs. This might be a subject of another study.

Table 2 The riverine TN loads statistical parameters of each data set

	L_{mean}	L_{min}	L_{max}	L_{sd}	L_{ske}
Daily					
Training data	4324.54	26.06	233,030.00	9632.92	9.55
Testing data	6190.23	438.73	140,100.00	11,964.19	5.70
Whole data	4946.43	26.06	233,030.00	10,502.92	9.55
Monthly					
Training data	131,658.16	24,951.54	602,944.60	112,936.08	2.18
Testing data	188,458.03	24,675.72	855,061.02	198,933.60	1.99
Whole data	150,591.45	24,675.72	855,061.02	148,767.94	2.43

L_{mean} , L_{min} , L_{max} , L_{sd} , and L_{ske} denote the mean, minimum, maximum, standard deviation, and skewness coefficient of the riverine TN load data. The units of statistical parameters (i.e., L_{mean} , L_{min} , L_{max} , L_{sd}) for daily riverine TN load data set are kilograms per day, for monthly riverine TN load data set are kilograms per month

Wavelet analysis

The riverine TN load time series were decomposed into approximation (low frequency, large scale) and details (high frequency, small scale) by DWT, which requires less computational effort and simpler implementation than the CWT (Adamowski et al. 2012). The selection of the efficient mother wavelet and determination of the decomposition level are two important issues in WA (Nourani et al. 2014). Many types of wavelet families exist, including Haar wavelet (haar), Daubechies wavelet (dbN), Coiflet wavelet (coifN), Mexican hat wavelet, Gaussian wavelet, Morlet wavelet, and so on (Minu et al. 2010). Haar, which was invented by Haar (1910), is a single symmetry orthogonal and discontinuity wavelet (Stolajescu 2012) and the simplest of all available wavelets (Shoaib et al. 2014). The dbN was proposed by Ingrid Daubechies. It is one of the most commonly used wavelets. In dbN, N denotes the number of vanishing moments. Symmlet wavelet (SymN), which was also proposed by Ingrid Daubechies, is a modification to the dbN and has improved symmetry (Seo et al. 2015). The coifN was named after R. Coifman who demanded Ingrid Daubechies to construct this wavelet (Daubechies 1992). The wavelet and scaling functions both considered in coifN have vanishing moments. Moreover, the Coiflet is a continuous, compact supported, orthogonal, and nearly symmetric mother wavelet. Meyer wavelet, also called mey, was designed by Meyer (1985). The dmey denotes the discrete approximation of the Meyer wavelet. In order to cover maximum range of wavelets in the current study, we surveyed effects of 23 mother wavelets which are from the 5 most frequently used wavelet families on the performance of developed WA-SVR model. A summary of mathematical properties of these mother wavelets is presented in Table 3.

Another important issue that should be considered in WA is the selection of suitable decomposition level. To the authors’

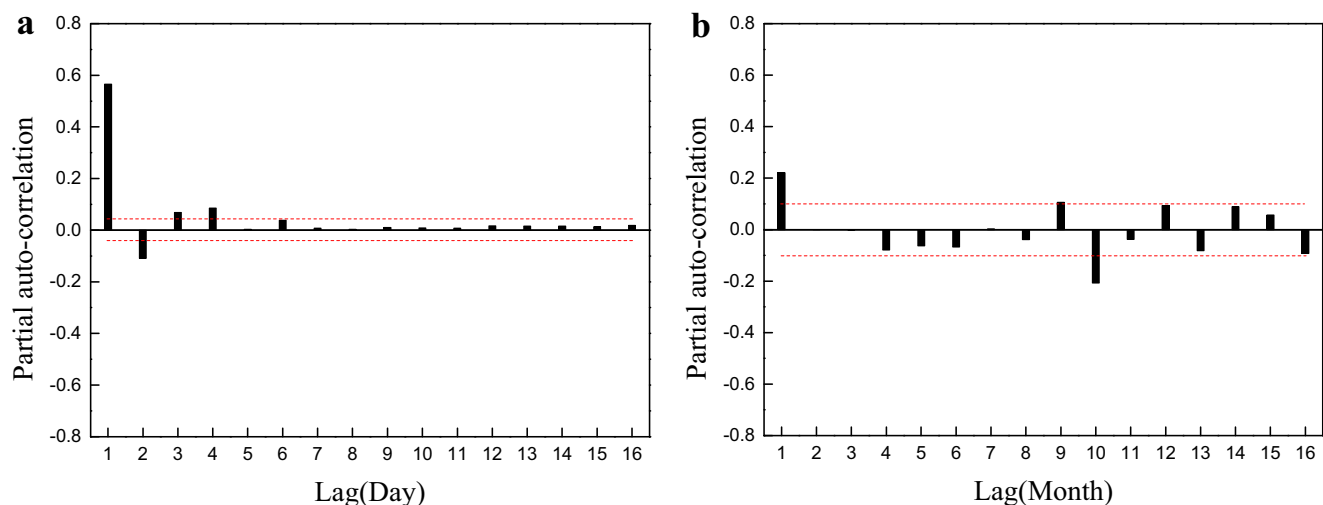


Fig. 6 Partial auto-correlation function of **a** daily and **b** monthly riverine TN load time series

best knowledge, there is no standard method for determining the decomposition level. Partal and Kisi (2007), Adamowski and Sun (2010), and Kisi and Shiri (2011) used trial and error method to select decomposition level and employed level 10, level 8, and level 3 decomposition, respectively. According to regression analysis, Shoaib et al. (2014, 2016) favored the use of decomposition level 9. Yet, Kisi (2010) suggested that using a large number of inputs should be avoided in order to save computational time and effort. Furthermore, the higher level of decomposition, the less likely the transformed signals represent the original time series (Belayneh et al. 2014). At present, an effective and operable empirical formula has been used in many studies to choose suitable decomposition level (Adamowski and Chan 2011; Aussem et al. 1998; Nourani et al. 2009a, 2009b; Tiwari and Chatterjee 2010). Here, we also used the following empirical formula to obtain the decomposition level:

$$L = \text{int}[\log(N)] \tag{13}$$

where int is the integer-part function; L is the number of decomposition level; N is the number of time series data. According to the time series number, L was equal to 3 and 2 for daily and monthly riverine TN load forecasting, respectively. Consequently, daily and monthly riverine TN load time series were decomposed into sub-series at three and two resolution levels using DWT, respectively. The original riverine TN load time series, as well as the approximations and details are shown in Fig. 7. Considering that all sub-series contain information related to the original time series (Adamowski and Sun 2010), the new inputs to SVR for the WA-SVR model were set by using all wavelet components.

Parameter optimization of the SVR model

The key step for constructing a high-performance SVR model is the determination and optimization of the penalty parameter (C) as well as adjustable kernel parameter (γ). Parameter C controls the empirical error in optimization problems (Singh et

Table 3 Mathematical properties of selected mother wavelets

Mother wavelets	Abbreviations	Number	Mathematical properties						
			A	B	C	D	E	F	G
Haar	haar	1	√	√	1	√	√	1	2
Daubenchies	dbN	9	√	–	N	√	√	2N-1	2N
Symlets	SymN	7	√	√*	N	√	√	2N-1	2N
Coiflers	CoifN	5	√	√*	2N	√	√	6N-1	6N-1
Dmeyer	dmey	1	–	√	–	√	√	Finite length	[–8,8]

A compact support, B symmetry, C vanishing moments, D orthogonality, E biorthogonality, F supporting length, G filter length. √ have the property, √* have the nearly property, – does not have the property

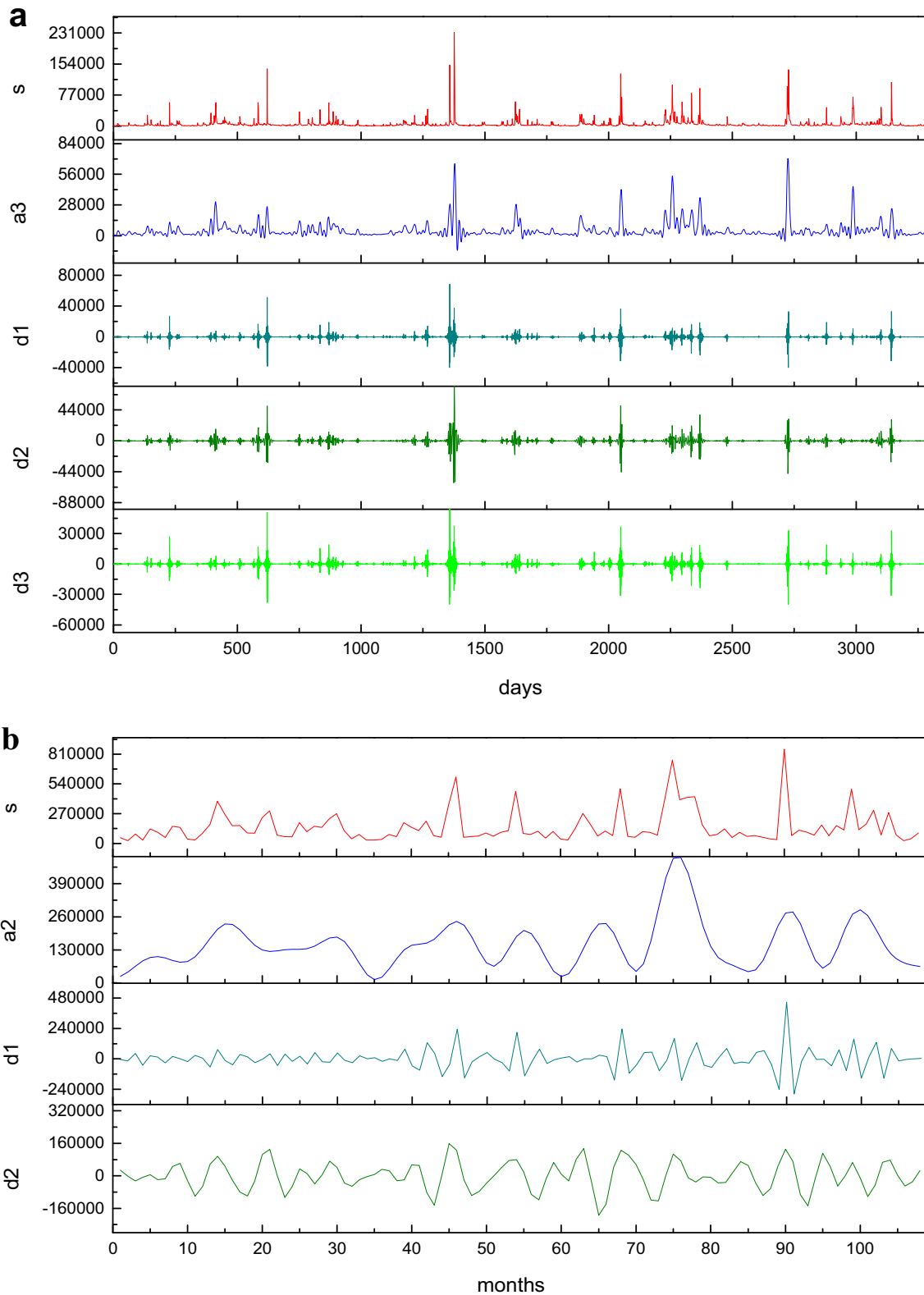


Fig. 7 Original(s) and decomposed time series (a2, a3, d1, d2, and d3) of **a** daily and **b** monthly riverine TN loads using dmev wavelet

al. 2011). Too small value will lead to an under-fitting problem, whereas too large value will result in an over fit of training data (Wang et al. 2007). Parameter γ , which determines

the amplitude of kernel function, is the only parameter in the RBF kernel function (Noori et al. 2011). The LIBSVM toolbox provides cross validation with grid search algorithm to

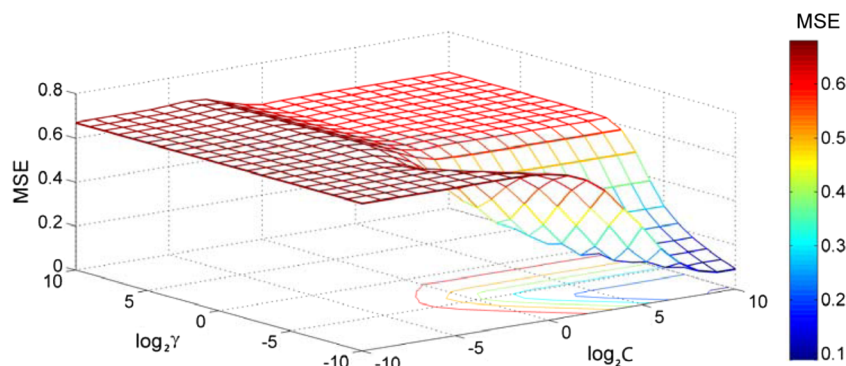


Fig. 8 Three dimension view of the optimization results for parameters C and γ (take the WA-SVR with dmey for monthly riverine TN load forecasting for example, best $C = 512$, $\gamma = 0.00097656$)

find the optimal pairwise C and γ , and the MSE value is used as the criterion. In fivefold cross validation, the data points in the training set are divided into five equal size sub-sets. In the next stage, four sub-sets are used to train the model, and the rest one sub-set is used for the testing. All instances in the entire training set can be forecasted once (Singh et al. 2014). The advantage and characteristic of the cross validation method is that it avoids the over-fitting problem (Hsu and Chang 2003). The algorithm of grid search divides the search range to

be optimized into grids. All grid points are then traversed to derive the best results. The parameter ranges along with interval size determine the accuracy of grid search optimization (Singh et al. 2011). A higher accuracy can be achieved by decreasing step size and increasing parameters range (Wang et al. 2007). In this work, the most appropriate C and γ were obtained through the grid search over a range from 2^{-10} to 2^{10} with a step size of 2^1 . The optimal parameters were subsequently employed to train the SVR model. Figure 8 illustrates

Table 4 Comparison of performance of the WA-SVR model with different mother wavelets during testing stage

	Daily			Monthly		
	R^2	NS	MSE (10^7 kg/day)	R^2	NS	MSE (10^{10} kg/month)
coif1	0.8613	0.8452	2.2133	0.6990	0.6457	1.3633
coif2	0.9057	0.9016	1.4078	0.8039	0.7934	0.7948
coif3	0.9219	0.9084	1.3093	0.8623	0.8550	0.5577
coif4	0.9323	0.9291	1.0142	0.8837	0.8767	0.4744
coif5	0.9439	0.9346	0.9357	0.8935	0.8870	0.4348
db2	0.9027	0.8955	1.4938	0.5939	0.5359	1.7858
db3	0.8999	0.8977	1.4629	0.7269	0.7080	1.1234
db4	0.8845	0.8818	1.6902	0.8355	0.8350	0.6349
db5	0.9000	0.8898	1.5765	0.7465	0.7432	0.9882
db6	0.9066	0.9047	1.3635	0.7609	0.7554	0.9413
db7	0.9364	0.9281	1.0278	0.8490	0.8273	0.6645
db8	0.9259	0.9259	1.0594	0.6442	0.6005	1.5370
db9	0.9556	0.9529	0.6731	0.6336	0.6322	1.4153
db10	0.9643	0.9640	0.5141	0.8458	0.8406	0.6134
sym2	0.8946	0.8927	1.5343	0.5939	0.5359	1.7858
sym3	0.8999	0.8977	1.4629	0.7269	0.7080	1.1234
sym4	0.9052	0.8932	1.5277	0.8105	0.7946	0.7902
sym5	0.9197	0.9197	1.1482	0.6310	0.5757	1.6324
sym6	0.9126	0.9033	1.3823	0.8145	0.8140	0.7156
sym7	0.9427	0.9422	0.8267	0.6838	0.6698	1.2704
sym8	0.9318	0.9302	0.9988	0.8545	0.8499	0.5774
harr	0.8125	0.7835	3.0959	0.0117	-0.0726	4.1268
demy	<i>0.9690</i>	<i>0.9651</i>	<i>0.4997</i>	<i>0.8928</i>	<i>0.8906</i>	<i>0.4208</i>

The best-fit results are marked in italics

an example of the optimization of pairwise C and γ during the grid search.

Mother wavelet selection

The performance of different combined WA-AI models, such as WA-SVR model, is extremely sensitive to the mother wavelet. In this regard, 23 selected mother wavelets, including haar, db2, db3, db4, db5, db6, db7, db8, db9, db10, sym2, sym3, sym4, sym5, sym6, sym7, sym8, coif1, coif2, coif3, coif4, coif5, and dmey, were utilized to examine their impacts on the performance of WA-SVR model. The values of the R^2 , NS, and MSE statistics of different mother wavelets during testing stage are listed in Table 4. It was clearly that the proposed WA-SVR model provided different accuracies for various mother wavelets. For short-term riverine TN load forecasting, the performance of the WA-SVR for the 23 mother wavelets followed the order of dmey > db10 > db9 > sym7 > coif5 > sym8 > coif4 > db7 > db8 > sym5 > coif3 > db6 > sym6 > coif2 > db3 = sym3 > db2 > sym4 > sym2 > db5 > db4 > coif1 > haar. Normally, if the NS is greater than 0.8, a model can be considered as accurate, if the NS is greater than 0.65, a model can be considered as satisfactory (Borah and Bera 2004; Shu and Ouarda 2008). According to the NS, all the selected mother wavelets provided excellent results (NS > 0.8), except for haar (NS = 0.7835). The WA-SVR model with dmey performed best among all tested mother wavelets, and the corresponding increase in NS and R^2 were about 20%, whereas the decrease in MSE was over 80% relative to the worst performance mother wavelet (haar). A relatively weaker efficiency of the WA-SVR model was achieved for long-term forecasting. This phenomenon was likely due to previous monthly variations having smaller impacts on the current riverine TN loads compared to daily variation (Fig. 6). And besides that, it should be kept in mind that the monthly riverine TN load values in the training set did not cover the entire range in the testing set (Table 2), which had negative influences on the model efficiency. The performance of WA-SVR model for the 23 mother wavelets followed the order of dmey > coif5 > coif4 > coif3 > sym8 > db10 > db4 > db7 > sym6 > sym4 > coif2 > db6 > db5 > db3 = sym3 > sym7 > coif1 > db9 > db8 > sym5 > db2 > sym2 > haar. The NS value of WA-SVR model with dmey, coif5, coif4, coif3, sym8, db10, db4, db7, and sym6 wavelets were all greater than 0.8, indicating accurate performance, whereas with 7 of the 23 wavelets, the NS values were less than 0.65, demonstrating relatively inferior results. In this study, dmey mother wavelet was selected for further WA-SVR model

performance validation as it outperformed other mother wavelets.

Comparison between the WA-SVR and single SVR

To assess and validate the effect of data pre-processing tool WA on the SVR model's efficiency, the single SVR model without wavelet decomposition was adopted as the benchmark for comparison. The optimal WA-SVR model results were compared to that of the single SVR model. The generation of a single SVR model, including data partition and associated parameter optimization, was accomplished in exactly the same way as the WA-SVR model.

In fact, although the observed minimum riverine TN loads may be close to zero, they cannot be negative. Unfortunately, some negative values existed in the outputs of the established models (i.e., WA-SVR and SVR). These values are not unusual but serve no purpose with regard to riverine TN flux forecasting. Accordingly, all negative forecasts were adjusted to minimum values of observations in testing data set (i.e., 438.73 kg/day for daily and 24,675.72 kg/month for monthly riverine TN load forecasting). After changing the negative forecasts, the WA-SVR model performance improved slightly. For example, the R^2 and NS of the optimal WA-SVR model for monthly riverine TN load forecasting were increased by 2.63% from 0.8928 to 0.9163 and 2.84% from 0.8906 to 0.9159, respectively.

A demonstration of the comparison between the observed daily riverine TN loads and the proposed WA-SVR as well as the SVR model results in testing period is provided in Fig. 9 via the hydrograph and scatter plots. The hydrograph demonstrated that the WA-SVR results were closer to the corresponding observed values than the SVR forecasts. The WA-SVR and SVR models both generally underestimated the corresponding peak daily riverine TN loads. In the field of stream flow forecasting, several previous research exhibited that black-box data-driven models failed to produce good forecasting accuracy for extremely high values of stream flow. The authors of these studies pointed out the reason behind this phenomenon could be lack of data in the high value region in training data set, as the black-box data-driven models' training required large number of input-output data sets (Kasiviswanathan et al. 2016; Kisi and Cimen 2011; Ravansalar et al. 2017; Yaseen et al. 2016). Obviously, the WA-SVR or SVR model constructed here also cannot be sufficiently trained with only a small number of riverine TN load peaks and consequently compromised the model performance in forecasting peak values of riverine TN load. However, the WA-SVR model had greater improvement than single SVR. For example, the WA-SVR estimated the maximum peak value as 112,829 kg/d instead of the observed 140,100 kg/d with an underestimation of 19.47%, while the SVR result was 18,708 kg/d with an underestimation of 86.65%. In scatter

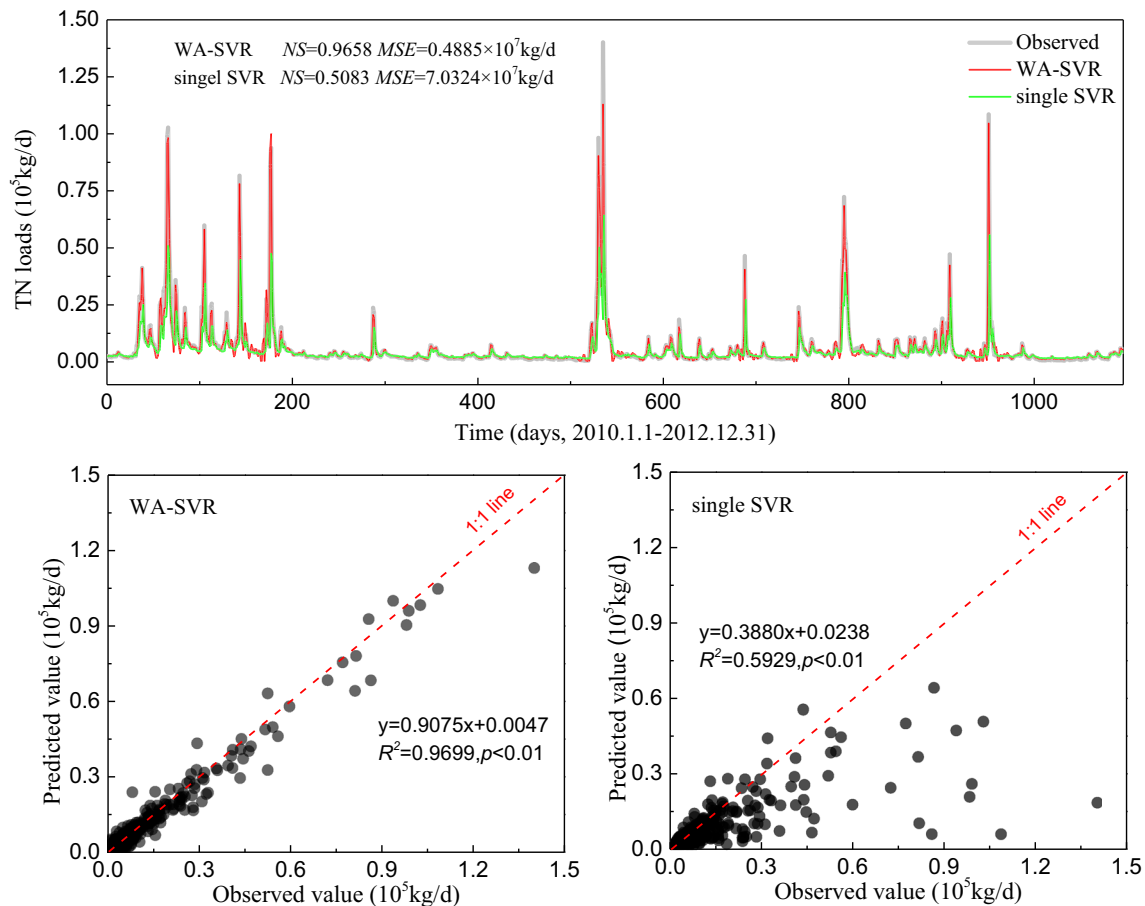


Fig. 9 Observed and forecasted daily riverine TN loads by the WA-SVR and SVR models in testing period

plots, the 1:1 line represents perfect results, in which when the points are closer to it, the model yields the better results (Wu et al. 2015). As seen from the scatter plots, a standard deviation around 1:1 line for the WA-SVR model was lower than that of the single SVR model. The fit line equations, which assume that the mathematical equations expressed as $y = ax + b$, revealed that the coefficient of a and b for the WA-SVR model are close to 1 and 0, respectively. The comparison between the WA-SVR and SVR test results illustrated that the hybrid model provided better performance than the single SVR model with regard to TN load forecasting. The WA-SVR model yielded smaller MSE and higher R^2 and NS. Concretely, the R^2 , NS, and MSE for the WA-SVR were 0.9699, 0.9658, and 0.4885×10^7 kg/day, respectively, while those of SVR were 0.5929, 0.5083, and 7.0324×10^7 kg/day, respectively.

Figure 10 demonstrates the observed and forecasted monthly riverine TN loads by the WA-SVR and SVR models in the testing phase. As shown in the hydrograph, the WA-SVR forecasts were closer to the corresponding observed values than those of the single SVR results. The SVR model failed to forecast riverine TN loads 1 month in advance. In contrast, the WA-SVR model performed much better as WA considerably improved the performance of the SVR model.

The R^2 and NS of the WA-SVR model increased from 0.0439 to 0.9163 and from -0.0779 to 0.9159, respectively. While the MSE decreased from 4.1474×10^{10} to 0.3237×10^{10} kg/month compared with that of the SVR model. For peaks, the hybrid model offered less error than the single SVR model. The WA-SVR estimated the maximum peak value (855,061 kg/month) as 759,221 kg/month with an underestimation of 11.21%, whereas the SVR result was 136,405 kg/month with an underestimation of 84.05%. The WA-SVR forecasted the second maximum peak value as 736,429 kg/month instead of the observed 757,419 kg/month with an underestimation of 2.77%, and the SVR forecasted as 195,138 kg/month with an underestimation of 74.24%. The WA-SVR and SVR computed the third maximum peak value (492,082 kg/month) as 468,663 and 135,960 kg/month, respectively, with underestimation of 4.76 and 72.37%. The scatter plots also demonstrated that the SVR model suffered more from scattering, which indicated lower accuracy, whereas the WA-SVR model had a better fitting effect between the forecasted and observed data considering that all of the data points clustered closely to the 1:1 line.

In the light of these findings, the joint application of the SVR and WA methods seemed to be more adequate than the

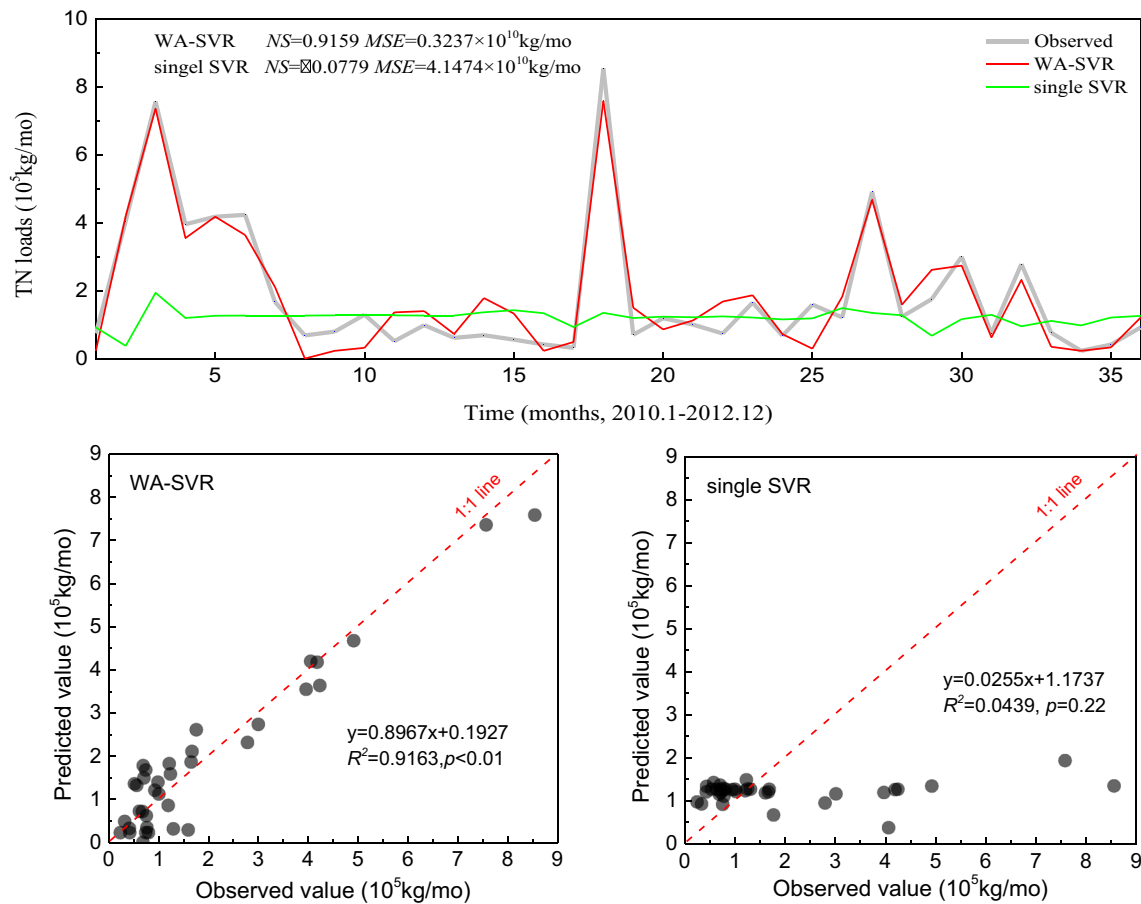


Fig. 10 Observed and forecasted monthly riverine TN loads by the WA-SVR and SVR models in testing period

single SVR without WA for NPS load forecasting in agricultural river systems. This finding may be primarily attributed to the WA decomposing the complex original time series into several simple wavelet components, which in turn exhibits the features (such as periodically) more clearly than the original signal (Kisi and Cimen 2011). In addition, the wavelet transformed data improved the ability of the SVR model by capturing useful information on various resolutions (Adamowski and Sun 2010).

Conclusions

A combined WA and SVR model was developed and explored for riverine TN load forecasting in the ChangLe River watershed based on antecedent riverine TN load values. The riverine TN load time series obtained through the LOADEST model were decomposed into sub-series by DWT. These sub-series were then used as conjunction model inputs. In the WA-SVR, the effects of mother wavelet functions on the model efficiency were evaluated and the results revealed that the WA-SVR model was sensitive to the mother wavelets. For daily forecasting, all the selected mother wavelets could provide accurate performance, except haar. The model with the dmey

mother wavelet function yielded the best results. For monthly forecasting, dmey, coif5, coif4, coif3, sym8, db10, db4, db7, and sym6 wavelets gave accurate results. Additionally, the models with sym4, coif2, db6, db5, db3, sym3, and sym7 yielded satisfactory results, whereas the other 7 wavelets performed weakly. The comparison between WA-SVR and single SVR models indicated that WA can significantly improve the efficiency of the SVR model. For daily riverine TN load forecasting, the R^2 , NS, and MSE in the testing period were 0.9699, 0.9658, and 0.4885×10^7 kg/day with WA-SVR, and 0.5929, 0.5083, and 7.0324×10^7 kg/day with the SVR model. For monthly riverine TN load forecasting, the R^2 , NS, and MSE in the testing period were 0.9163, 0.9159, and 0.3237×10^{10} kg/month with the WA-SVR, and 0.0439, -0.0779, and 4.1474×10^{10} kg/month with the SVR model.

This study, for the first time, explored the WA-SVR combination model for riverine TN load forecasting. The results highlighted that the WA-SVR model provided a promising and effective method to address the problem of riverine nutrient load forecasting. However, some unsolved problems require future investigations and improvements for reinforcing the conclusions in this study. First, negative values which served no purpose as forecasts occurred in the results of the proposed model. The attempt to present an appropriate

method for dealing with negative values yielded by the WA-SVR model is recommended. Second, the NPS loads are affected by complex factors, e.g., flow velocity, temperature, precipitation, and fertilization. Nevertheless, for the WA-SVR model built in this research, only previous riverine TN load data was taken into account. In the future, other data regarding the riverine TN loads will be needed to reinforce the results drawn from this work. Third, the optimum decomposition level was selected by empirical equation, which could not guarantee obtaining the best performance. A comparative investigation on the impact of different decomposition level on the combined model can be also encouraged to improve the efficiency of the constructed model. Besides, the results exhibited the limitations in the use of WA-SVR model for forecasting peaks. Thereby, it is also suggested the future studies

to integrate the WA-SVR model with physically based models or empirical relationships between influencing factors and peak values for enhancing forecasting of TN load peak values.

Acknowledgements The authors would like to express appreciation to hydrological bureau of Zhejiang Province for the data provided for the ChangLe River.

Funding This work was supported by the National Natural Science Foundation of China (Grant No. 41571216) and the Chinese National Key Technology R&D Program (Grant No. 2016YFD0801103).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Appendix

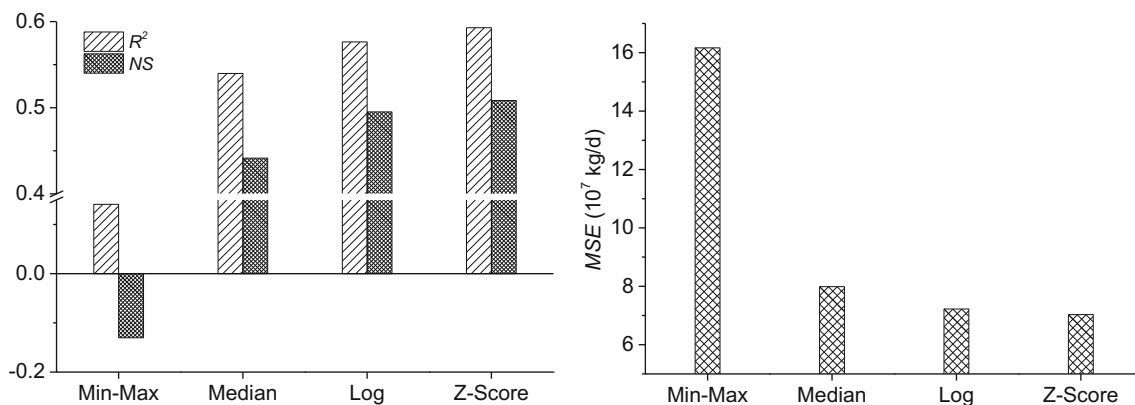


Fig. 11 Single SVR model performance with different data normalization for daily riverine TN load forecasting in testing stage

References

- Adamowski J, Chan HF (2011) A wavelet neural network conjunction model for groundwater level forecasting. *J Hydrol* 407(1–4):28–40
- Adamowski J, Sun K (2010) Development of a coupled wavelet transform and neural network method for flow forecasting of non-perennial rivers in semi-arid watersheds. *J Hydrol* 390:85–91
- Adamowski J, Chan HF, Prasher SO, Ozga-Zielinski B, Sliusariva A (2012) Comparison of multiple linear and nonlinear regression, autoregressive integrated moving average, artificial neural network, and wavelet artificial neural network methods for urban water demand forecasting in Montreal, Canada. *Water Resour Res* 48: W01528
- Atanasijević DZ, Pocajt VV, Povrenović DS, Perić-Grujić AA, Ristić MD (2014) Modelling of dissolved oxygen content Danube River using artificial neural networks and Monte Carlo simulation uncertainty analysis. *J Hydrol* 519:1895–1907
- Arnold JG, Fohrer N (2005) SWAT2000: current capabilities and research opportunities in applied watershed modelling. *Hydrol Process* 19: 563–572
- Aussem A, Campbell J, Murtagh F (1998) Wavelet-based feature extraction and decomposition strategies for financial forecasting. *J Comput Intell Finance* 6(2):5–12
- Belayneh A, Adamowski J, Khalil B, Ozga-Zielinski B (2014) Long-term SPI drought forecasting in the Awash River Basin in Ethiopia using wavelet neural network and wavelet support vector regression models. *J Hydrol* 508:418–429
- Borah DK, Bera M (2004) Watershed-scale hydrologic and nonpoint-source pollution models: reviews of application. *Trans ASAE* 47: 789–803
- Bowes MJ, Neal C, Jarvie HP, Smith JT, Davies HN (2010) Predicting phosphorus concentrations in British rivers resulting from the introduction of improved phosphorus removal from sewage effluent. *Sci Total Environ* 408:4239–4250

- Cannas B, Fanni A, See L, Sias G (2006) Data preprocessing for river flow forecasting using neural networks: wavelet transforms and data partitioning. *Phys Chem Earth* 31:1164–1171
- Chang CC, Lin CJ (2011) LIBSVM: a library for support vector machines. *ACM Trans Intell Syst Technol* 2:1–27
- Chau KW (2017) Use of meta-heuristic techniques in rainfall-runoff modeling. *Water* 9(3):186–191
- Chau KW, Wu CL (2010) A hybrid model coupled with singular spectrum analysis for daily rainfall prediction. *J Hydroinf* 12(4):458–473
- Chen XY, Chau KW (2016) A hybrid double feedforward neural network for suspended sediment load estimation. *Water Resour Manag* 30(7):2179–2194
- Chen DJ, Lu J, Shen YN, Gong DQ, Deng OP (2011) Spatio-temporal variations of nitrogen in an agricultural watershed in eastern China: catchment export, stream attenuation and discharge. *Environ Pollut* 159:2989–2995
- Chen DJ, Dahlgren RA, Lu J (2013) A modified load apportionment model for identifying point and diffuse source nutrient inputs to rivers from stream monitoring data. *J Hydrol* 501:25–34
- Chou CM, Wang RY (2002) On-line estimation of unit hydrographs using the wavelet-based LMS algorithm. *Hydrol Sci J* 47(5):721–738
- Chou WS, Lee TC, Lin JY, Shaw LY (2007) Phosphorus load reduction goals for Feitsui Reservoir watershed, Taiwan. *Environ Monit Assess* 131:395–408
- Coffey SW, Line DE (1998) Simulation of dairy best management practices using the AGNPS model. *J Lake Reserv Manage* 14(4):417–427
- Cristianine N, Taylor JS (2000) An introduction to support vector machine and other kernel based learning methods. Cambridge University Press, Cambridge
- Daubechies I (1992) Ten lectures on wavelets (CBMS-NSF regional conference series in applied mathematics). Soc Indust Appl Math
- Dibike YB, Velickov S, Solomatine DP, Abbott MB (2001) Model induction with support vector machines: introduction and applications. *J Comput Civ Eng* 15:208–216
- Donigian AS, Imhoff JC, Bicknell BR, Kittle JL (1984) Application guide for the hydrological simulation program FORTRAN. GA7 Environmental Research Laboratory, US Environmental Protection Agency, Athens
- Feng Q, Wen XH, Li JG (2015) Wavelet analysis-support vector machine coupled models for monthly rainfall forecasting in arid regions. *Water Resour Manag* 29:1049–1065
- Gao C, Zhang TL (2010) Eutrophication in a Chinese context: understanding various physical and socio-economic aspects. *Ambio* 39:385–393
- Haar A (1910) Zur Theories der orthogonalen Funktionensysteme. *Math Ann* 69(3):331–371
- He ZB, Wen XH, Liu H, Du J (2014) A comparative study of artificial neural network, adaptive neuro fuzzy inference system and support vector machine for forecasting river flow in the semiarid mountain region. *J Hydrol* 509:379–386
- Helsel DR, Hirsch RM (2002) Statistical methods in water resources. Techniques of water resources investigations, book 4, chapter A3. US Geological Survey, Reston
- Houser JN, Richardson WB (2010) Nitrogen and phosphorous in the Upper Mississippi River: transport, processing, and effects on the river ecosystem. *Hydrobiologia* 640:71–88
- Howarth RW, Swaney D, Billen G, Garnier J, Hong B, Humborg C, Johnes P, Mörth CM, Marino R (2012) Nitrogen fluxes from the landscape are controlled by net anthropogenic nitrogen inputs and by climate. *Front Ecol Environ* 10:37–43
- Hsu CW, Chang CC (2003) A practical guide to support vector classification. <http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf>
- Huang WR, Xu B, Chan-Hilton A (2004) Forecasting flows in Apalachicola River using neural networks. *Hydrol Process* 18:2545–2564
- Kalteh AM (2013) Monthly river flow forecasting using artificial neural network and support vector regression models coupled with wavelet transform. *Comput Geosci-UK* 54:1–8
- Kalteh AM (2016) Improving forecasting accuracy of streamflow time series using least squares support vector machine coupled with data-preprocessing techniques. *Water Resour Manag* 30:747–766
- Kasiviswanathan KS, He J, Sudheer KP, Tay JH (2016) Potential application of wavelet neural network ensemble to forecast streamflow for flood management. *J Hydrol* 536:161–173
- Keerthi SS, Lin CJ (2001) Asymptotic behaviors of support vector machines with Gaussian kernel. *Neural Comput* 15:1667–1689
- Khan MS, Coulbaly P (2006) Application of support vector machine in lake water level prediction. *J Hydrol Eng* 11:199–205
- Kisi O (2010) Daily suspended sediment estimation using neuro-wavelet models. *Int J Earth Sci* 99:1471–1482
- Kisi O (2012) Modeling discharge-suspended sediment relationship using least square support vector machine. *J Hydrol* 456–457:110–120
- Kisi O, Cimen M (2011) A wavelet-support vector machine conjunction model for monthly streamflow forecasting. *J Hydrol* 399:132–140
- Kisi O, Shiri J (2011) Precipitation forecasting using wavelet-genetic programming and wavelet-neuro-fuzzy conjunction models. *Water Resour Manag* 25(13):3135–3152
- Kothiyari UC, Singh VP (1999) A multiple-input single-output model for flow forecasting. *J Hydrol* 220:12–26
- Legates DR, McCabe GJ Jr (1999) Evaluating the use of goodness-of-fit measures in hydrologic and hydroclimatic model validation. *Water Resour Res* 35(1):233–241
- Li S, Zhang Q (2010) Spatial characterization of dissolved trace elements and heavy metals in the upper Han River (China) using multivariate statistical techniques. *J Hazard Mater* 176:579–588
- Lin JY, Cheng CT, Chau KW (2006) Using support vector machines for long term discharge prediction. *Hydrol Sci J* 51:599–612
- Liu KJ, Shi ZH, Fang NF, Zhu HD, Ai L (2013) Modeling the daily suspended sediment concentration in a hyper concentrated river on the Loess Plateau, China, using the Wavelet-ANN approach. *Geomorphology* 186:181–190
- Lu RY (2002) Decomposition of interdecadal and interannual components for north China rainfall in rainy season. *Chinese J Atmos* 26:611–624 (In Chinese)
- Meyer Y (1985) Principe d'incertitude, bases hilbertiennes et algebres d'operateurs., Asterisque, Societe Mathematique de France, Paris, France
- Minu KK, Lineesh MC, John CJ (2010) Wavelet neural networks for nonlinear time series analysis. *Appl Math Sci* 4(50):2485–2495
- Mishra AK, Desai VR (2005) Drought forecasting using stochastic models. *Stoch Env Res Risk A* 19:326–339
- Mohammadpour R, Shaharuddin S, Chang CK, Zakaria NA, Ghani AA, Chan NW (2015) Prediction of water quality index in constructed wetlands using support vector machine. *Environ Sci Pollut Res* 22:6208–6219
- Morse NB, Wollheim WM (2014) Climate variability masks the impacts of land use change on nutrient export in a suburbanizing watershed. *Biogeochemistry* 121:45–59
- Nabavi-Pelesaraei A, Bayat B, Hosseinzadeh-Bandbafha H, Afrasyabi H, Chau KW (2017) Modeling of energy consumption and environmental life cycle assessment for incineration and landfill systems of municipal solid waste management—a case study in Tehran Metropolis of Iran. *J Clean Prod* 148:427–440
- Najah A, El-Shafie A, Karim OA, El-Shafie AH (2014) Performance of ANFIS versus MLP-NN dissolved oxygen prediction models in water quality monitoring. *Environ Sci Pollut Res* 21(13):1658–1670
- Noori R, Karbassi AR, Moghaddamnia A, Han D, Zokaei-Ashtiani MH, Farokhnia A, Ghafari Gousheh M (2011) Assessment of input variables determination on the SVM model performance using PCA, gamma test, and forward selection techniques for monthly stream flow prediction. *J Hydrol* 401:177–189

- Noori R, Yeh HD, Abbasi M, Kachoosangi FT (2015) Uncertainty analysis of support vector machine for online prediction of five-day biochemical oxygen demand. *J Hydrol* 527:833–843
- Nourani V, Alami MT, Aminfar MH (2009a) A combined neural-wavelet model for prediction of Ligvanchai watershed precipitation. *Eng Appl Artif Intell* 22:466–472
- Nourani V, Komasi M, Mano A (2009b) A multivariate ANN-wavelet approach for rainfall–runoff modeling. *Water Resour Manag* 23(14):2877–2894
- Nourani V, Baghanam AH, Adamowski J, Kisi O (2014) Applications of hybrid wavelet-artificial intelligence models in hydrology: a review. *J Hydrol* 514:358–377
- Olyae E, Banejad H, Chau KW, Melesse AM (2015) A comparison of various artificial intelligence approaches performance for estimating suspended sediment load of river systems: a case study in United States. *Environ Monit Assess* 187:189
- Partal T, Kisi O (2007) Wavelet and neuro-fuzzy conjunction model for precipitation forecasting. *J Hydrol* 342(1):199–212
- Partal T, Kucuk M (2006) Long-term trend analysis using discrete wavelet components of annual precipitations measurements in Marmara region (Turkey). *Phys Chem Earth* 31:1189–1200
- Raghavendra NS, Deka PC (2014) Support vector machine applications in the field of hydrology: a review. *Appl Soft Comput* 19:372–386
- Ravansalar M, Rajaei T, Kisi O (2017) Wavelet-linear genetic programming: a new approach for modeling monthly streamflow. *J Hydrol* 549:461–475
- Runkel RL, Crawford CG, Cohn TA (2004) Load Estimator (LOADEST): A FORTRAN program for estimating constituent loads in streams and rivers, techniques and methods report 4-A5 U.S. Geological Survey, Reston
- Ryusuke H, Takuro S, Zheng TG, Masahiko O, Li ZW (2002) Nitrogen budgets and environmental capacity in farm systems in a large-scale karst region, southern China. *Nutr Cycl Agroecosyst* 63:139–149
- Sefeedpari P, Rafiee S, Akram A, Chau KW, Pishgar-Komleh SH (2016) Prophesying egg production based on energy consumption using multi-layered adaptive neural fuzzy inference system approach. *Comput Electron Agric* 131:10–19
- Seo WM, Kim SW, Kisi O, Singh VP (2015) Daily water level forecasting using wavelet decomposition and artificial intelligence techniques. *J Hydrol* 520:224–243
- Shen ZY, Qiu JL, Hong Q, Chen L (2014) Simulation of spatial and temporal distributions of non-point source pollution load in the three gorges reservoir region. *Sci Total Environ* 493:138–146
- Shiri J, Kisi O (2010) Short-term and long-term streamflow forecasting using a wavelet and neuro-fuzzy conjunction model. *J Hydrol* 394:486–493
- Shoaib M, Shamseldin AY, Melville BW (2014) Comparative study of different wavelet based neural network models for rainfall–runoff modeling. *J Hydrol* 515:47–58
- Shoaib M, Shamseldin AY, Melville BW, Khan MM (2016) A comparison between wavelet based static and dynamic neural network approaches for runoff prediction. *J Hydrol* 535:211–225
- Shu C, Ouarda TBMJ (2008) Regional flood frequency analysis at ungauged sites using the adaptive neuro-fuzzy inference system. *J Hydrol* 349:31–43
- Simith LC, Turcotte DL, Isacks B (1998) Stream flow characterization and feature detection using a discrete wavelet transform. *Hydrol Process* 12:233–249
- Singh KP, Basant N, Gupta S (2011) Support vector machines in water quality management. *Anal Chim Acta* 703:152–162
- Singh KP, Gupta S, Rai P (2014) Predicting dissolved oxygen concentration using kernel regression modeling approaches with nonlinear hydro-chemical data. *Environ Monit Assess* 186:2749–2765
- State Environment Protection Bureau of China (2002) Water and wastewater analysis method. China Environmental Science Press, Beijing (in Chinese)
- Stolojescu CL (2012) A wavelets based approach for time series mining, Ph.D. Dissertation, Telecom Bretagne, France
- Tiwari MK, Chatterjee C (2010) Development of an accurate and reliable hourly flood forecasting model using wavelet-bootstrap-ANN (WBANN) hybrid approach. *J Hydrol* 394(3–4):458–470
- Vapnik VN (1995) The nature of statistical learning theory. Springer, New York
- Vapnik VN (1998) Statistical learning theory. Wiley, New York
- Wang W, Ding J (2003) Wavelet network model and its application to the prediction of hydrology. *Nat Sci* 1:67–71
- Wang J, Du HY, Liu HX, Yao XJ, Hu ZD, Fan BT (2007) Prediction of surface tension for common compounds based on novel methods using heuristic method and support vector machine. *Talanta* 73:147–156
- Wang FE, Tian P, Yu J, Lao GM, Shi TC (2011) Variations in pollutant fluxes of rivers surrounding Taihu Lake in Zhejiang Province in 2008. *Phys Chem Earth* 36(9–11):366–371
- Wang WC, Xu DM, Chau KW, Lei GJ (2014) Assessment of river water quality based on theory of variable fuzzy sets and fuzzy binary comparison method. *Water Resour Manag* 28(12):4183–4200
- Wei S, Yang H, Song JX, Abbaspour K, Xu ZX (2013) A wavelet-neural network hybrid modelling approach for estimating and predicting river monthly flows. *Hydrol Sci J* 58:374–389
- Wu CL, Chau KW, Li YS (2009a) Predicting monthly streamflow using data-driven models coupled with data-preprocessing techniques. *Water Resour Res* 45:W08432
- Wu CL, Chau KW, Li YS (2009b) Methods to improve neural network performance in daily flows prediction. *J Hydrol* 372(1):80–93
- Wu ZZ, Xu EB, Long J, Wang F, Xu XM, Jin ZY, Jiao AQ (2015) Rapid measurement of antioxidant activity and γ -aminobutyric acid content of Chinese rice wine by fourier-transform near infrared spectroscopy. *Food Anal Methods* 8:2541–2553
- Yaseen ZM, Jaafar O, Deo RC, Kisi O, Adamowski J, Quilty J, El-Shafie A (2016) Stream-flow forecasting using extreme learning machines: a case study in a semi-arid region in Iraq. *J Hydrol* 542:603–614
- Yoon H, Jun SC, Hyun Y, Bae GO, Lee KK (2011) A comparative study of artificial neural networks and support vector machines for predicting groundwater levels in a coastal aquifer. *J Hydrol* 396:128–138
- Zhang Y, Cong Q, Xie YF, Yang JX, Zhao B (2008) Quantitative analysis of routine chemical constituents in tobacco by near-infrared spectroscopy and support vector machine. *Spectrochimica Acta Part A* 71:1408–1413