Evaluating a coupled discrete wavelet transform and support vector regression for daily and monthly streamflow…

Article *in* Journal of Hydrology · July 2014

DOI: 10.1016/j.jhydrol.2014.06.050

CITATIONS
18

READS
180

4 authors, including:

Zhiyong Liu
NOAA, STAR / CIRA, CSU
15 PUBLICATIONS **49** CITATIONS
SEE PROFILE

Ping Zhou
Guangdong Academy of Forestry
17 PUBLICATIONS **239** CITATIONS
SEE PROFILE

Journal of Hydrology 519 (2014) 2822–2831

Contents lists available at ScienceDirect

Journal of Hydrology

journal homepage: www.elsevier.com/locate/jhydrol

Evaluating a coupled discrete wavelet transform and support vector regression for daily and monthly streamflow forecasting

Zhiyong Liu [a], Ping Zhou [b,*], Gang Chen [c], Ledong Guo [b]

[a] Institute of Geography, Heidelberg University, Heidelberg 69120, Germany
[b] Research Station of Dongjiangyuan Forest Ecosystem, Guangdong Academy of Forestry, Guangzhou 510520, China
[c] Institute of Forest Resource Information Techniques, Chinese Academy of Forestry, Beijing 100091, China

ARTICLE INFO

Article history:
Available online 6 July 2014

Keywords:
Wavelet analysis
Support vector regression
Streamflow forecasting
Model averaging
Indiana

SUMMARY

This study investigated the performance and potential of a hybrid model that combined the discrete wavelet transform and support vector regression (the DWT–SVR model) for daily and monthly streamflow forecasting. Three key factors of the wavelet decomposition phase (mother wavelet, decomposition level, and edge effect) were proposed to consider for improving the accuracy of the DWT–SVR model. The performance of DWT–SVR models with different combinations of these three factors was compared with the regular SVR model. The effectiveness of these models was evaluated using the root-mean-squared error (RMSE) and Nash–Sutcliffe model efficiency coefficient (NSE). Daily and monthly streamflow data observed at two stations in Indiana, United States, were used to test the forecasting skill of these models. The results demonstrated that the different hybrid models did not always outperform the SVR model for 1-day and 1-month lead time streamflow forecasting. This suggests that it is crucial to consider and compare the three key factors when using the DWT–SVR model (or other machine learning methods coupled with the wavelet transform), rather than choosing them based on personal preferences. We then combined forecasts from multiple candidate DWT–SVR models using a model averaging technique based upon Akaike's information criterion (AIC). This ensemble prediction was superior to the single best DWT–SVR model and regular SVR model for both 1-day and 1-month ahead predictions. With respect to longer lead times (i.e., 2- and 3-day and 2-month), the ensemble predictions using the AIC averaging technique were consistently better than the best DWT–SVR model and SVR model. Therefore, integrating model averaging techniques with the hybrid DWT–SVR model would be a promising approach for daily and monthly streamflow forecasting. Additionally, we strongly recommend considering these three key factors when using wavelet-based SVR models (or other wavelet-based forecasting models).

1. Introduction

Streamflow is a fundamental and critical component of global and regional hydrological cycles (Makkeasorn et al., 2008). It is also strongly associated with human water supply, the agricultural and industrial sectors, and natural disasters (e.g., droughts and floods). Therefore, reliable short and long-term forecasts of streamflow are crucial for appropriate and effective water resource planning and management, especially in drought and flood-prone regions (Kisi and Cimen, 2011). Over the last few decades, streamflow prediction has become more important and has received significant attention, because the fluctuations of global climate change are causing frequent and extreme drought and flood events (Adamowski and Sun, 2010). Hydrologic phenomena (e.g., streamflow) can be forecasted using either physical, conceptual, or data-driven approaches. The data-driven approach can be developed quickly, is easy to implement in real-time, and requires minimum information when compared with physically based hydrological models. Therefore, it may be an ideal tool for watersheds where other climatological and hydrogeological data are limited, and where it is more important to provide precise forecasts than understand physical catchment processes (Adamowski, 2008; Adamowski and Sun, 2010). A variety of data-driven models have been developed and used for streamflow forecasting in different interesting regions. They include traditional statistical models such as multiple linear regression (MLR) and autoregressive integrated moving average (ARIMA) models (McKerchar and Delleur, 1974), and machine learning techniques such as artificial neural networks (ANN) and support vector machine (SVM) (Kim and Barros, 2001; Sivapragasam et al., 2001; Kisi and Cimen, 2011).

* Corresponding author. Tel.: +86 20 87033527; fax: +86 020 87031245.
E-mail address: zhoupinger@gmail.com (P. Zhou).

http://dx.doi.org/10.1016/j.jhydrol.2014.06.050
0022-1694/

To date, SVM has attracted a great deal of interest as a soft computational technique (Kisi and Cimen, 2011). The theory of SVM was introduced by Vapnik and co-workers on the basis of a separable bipartition problem at AT&T Bell Laboratories in 1992. This prediction tool uses machine learning theory to maximize predictive accuracy while automatically avoiding over-fitting to the data (Vapnik, 1995). The SVM is trained using the structural risk minimization (SRM) principle, rather than the traditional empirical risk minimization (ERM) principle (Yu et al., 2006; Wu et al., 2012). The ERM principle can only minimize the training error, but SRM minimizes an upper bound of the generalization error. The SVM model can thus achieve an optimum network structure and better generalization using the SRM principle (Lin et al., 2006). SVM was originally used to solve the classification problem, before and then another version of SVM for regression problem was proposed by Vapnik et al. (1997). This new version is called support vector regression (SVR), which has become the most common application form of SVM. More recently, it has been further improved and successfully applied in different problems of prediction such as signal processing, stock price forecasting and traffic flow prediction (Cao and Tay, 2001; Chang and Lin, 2001; Hong, 2011; Kao et al., 2013). In addition to those applications, SVR has also gained popularity in hydrological field. Liong and Sivapragasam (2002) used a SVR for flood stage forecasting in Dhaka, Bangladesh. Bray and Han (2004) identified an appropriate model structure and relevant parameters when using SVR to accurately forecast streamflows. Yu et al. (2006) used an SVR to establish a real-time flood stage forecasting model in Lan-Yang River, Taiwan, and stated that the proposed models can effectively predict flood stages 1–6 h ahead. Wang et al. (2009) compared the performance of several methods for forecasting monthly discharge time series, and revealed that the SVR performed better than the ANN and ARIMA models.

Although the SVR model presents flexibility and usefulness in forecasting hydrological time series, it has limitations regarding highly non-stationary hydrological responses that vary over a range of scales (e.g., from daily to multi-decadal) (Cannas et al., 2006; Adamowski and Chan, 2011). Recent developments in wavelet theory pave the path to reliably obviate SVR (or other data-driven models) shortcomings in dealing with the non-stationary behavior of hydrological signals. Wavelet transform has the ability to provide a time–frequency representation of a signal at various scales in the time domain. It can decompose a given hydrological time series into various periodic components, providing considerable information about the physical structure of the data (Daubechies, 1990). It is thus possible to generate better forecasts by combining the strengths of wavelet transform and SVR (or other data-driven models) (Kisi, 2008, 2009; Nourani et al., 2009, 2011; Remesan et al., 2009; Adamowski and Sun, 2010; Pramanik et al., 2010; Shiri and Kisi, 2010; Li, 2011; Tiwari and Chatterjee, 2011; Kisi and Cimen, 2012; Rasouli et al., 2012; Adamowski, 2013; Sang, 2013). For instance, Kisi and Cimen (2011) proposed a hybrid wavelet and SVR model for hydrological forecasting and demonstrated that the new approach provided a better prediction than the regular SVR model. Kalteh (2013) predicted monthly streamflows using wavelet-based data-driven models (including SVR), and concluded that the coupled model provided more accurate forecasts than the non-coupled data-driven model.

However, some key factors of the hybrid wavelet–SVR model still need to be explored in detail. These factors involve the choice of an appropriate wavelet, the decomposition level, and the effect of boundary problems in the wavelet decomposition phase of a wavelet–SVR model. Although such issues are essential for wavelet-based hydrological forecasting, they are often overlooked. In most practical streamflow forecasting applications using wavelet-based SVR models (or other wavelet-based forecasting models), the selection of wavelets, decomposition levels, and edge effects were based on personal preferences or subjective assumptions. In fact, different settings of these key factors in the wavelet decomposition phase of a wavelet-based SVR model may result in relatively significant differences in the accuracy of a certain streamflow forecast. It is therefore desirable to establish an overall assessment regarding the choice of wavelets, decomposition levels, and edges when using a wavelet-based SVR model (or other wavelet-based forecasting models) for streamflow prediction. An insight into the influence of these proposed key factors on model performance would also be welcome. Furthermore, given a set of forecasts generated from different wavelet-based SVR models developed using varying settings of these key factors, it is interesting to explore the ensemble prediction given by combining individual forecasts from candidate models, instead of using a single best model.

Therefore, this study aims to develop a framework that evaluates the performance discrepancies resulting from different mother wavelets, decomposition levels, and edge effects in a wavelet-based SVR model, and provides an effective ensemble streamflow prediction based on a model averaging technique. We first apply the wavelet-based SVR model in one-step ahead forecasting for both daily and monthly streamflows, and compare the results with those from the regular SVR model. We then explore the potential of multi-model ensemble prediction using an averaging technique based on Akaike's information criterion (AIC). In addition, we implement multi-step ahead forecasting for both daily (lead times of 2–3 days) and monthly (lead time of two months) streamflows.

2. Theoretical background

2.1. Support vector regression (SVR)

Support vector regression (SVR) is derived from the support vector machine (SVM) (Vapnik, 1995). SVR is used to solve regression problems with SVM. SVR uses a hypothesis space of linear functions in a high-dimensional feature space, and is trained by an algorithm from optimization theory that implements a learning bias derived from statistical learning theory (Yu et al., 2006). In SVR, the input vector is mapped to a high-dimensional feature space using a nonlinear mapping function (Wu et al., 2012). The learning goal of SVR is to find a regression function that estimates the functional dependence between a set of sampled points $x = \{x_1, x_2, \ldots, x_n\}$ (the input vector) and desired values $y = \{y_1, y_2, \ldots, y_n\}$ (Kisi and Cimen, 2011) (here, the input and desired vectors refer to the daily or monthly streamflow records and n is the total number of data points). The regression function of SVR is formulated as follows:

$$f(x) = (w \cdot \Phi(x)) + b, \tag{1}$$

where w and b are the weight vector and bias terms which are the coefficients in this regression function, and $\Phi(x)$ is a nonlinear mapping function. By mapping the input vector onto a high-dimensional space, the nonlinear separable problem becomes linearly separable in space (Maity et al., 2010).

The coefficients of a traditional regression model are determined by minimizing the square error, which can be regarded as an empirical risk based on the loss function (a measure of the quality of estimation) (Kao et al., 2013). The SVR uses a new type of loss function, called the ε-insensitivity loss function (L_ε). It is defined as

$$L_\varepsilon(f(x), y) = \begin{cases} |f(x) - y| - \varepsilon & \text{for } |f(x) - y| \geq \varepsilon \\ 0 & \text{otherwise} \end{cases}, \tag{2}$$

where y is the desired (target) output, and ε is a user-determined parameter which defines the region of ε-insensitivity. There is zero

loss when the predicted value falls into the band area. On the contrary, if the predicted value falls out the band area, the loss is equal to the difference between the predicted value and the margin (Kao et al., 2013)

Using the SRM principle, the weight vector, w, in Eq. (1) can be estimated by minimizing the following regularized risk function:

$$R_{reg} = C\frac{1}{n}\sum_{i=1}^{n} L_{\varepsilon}(f(x_i), y_i) + \frac{1}{2}\|w\|^2, \tag{3}$$

where $\frac{1}{2}\|w\|^2$ is the regularization term that determines the trade-off between the complexity and approximation accuracy of the regression model and ensures the model possesses an improved generalized performance, and C is a user-defined regularization constant that influences the trade-off between the regularization terms and empirical risk.

Considering empirical risk and structure risk synchronously, Eq. (3) is transformed into the following constrained form with slack variables (Vapnik, 1995, 2000):

$$\text{minimize } \frac{1}{2}||w||^2 + C\sum_{i=1}^{n}(\xi_i + \xi_i^*) \tag{4}$$

$$\text{subject to} \begin{cases} y_i - (w \cdot \Phi(x_i) + b) \le \varepsilon + \xi_i \\ (w \cdot \Phi(x_i) + b) - y_i \le \varepsilon + \xi_i^* \\ \xi_i \ge 0, \xi_i^* \ge 0, \quad i = 1, \ldots, n \end{cases} \tag{5}$$

where ξ_i and ξ_i^* are the positive slack variables that are used to measure the deviation of training samples outside the ε-insensitivity zone. Eq. (4) is a quadratic programming problem, which can be estimated using Lagrangian theory and the Karush–Kühn–Tucker condition (Haykin, 2003; Yu et al., 2006; Azamathulla and Wu, 2011). Finally, the general form of the SVR regression function can be expressed as (Vapnik, 2000; Kao et al., 2013)

$$f(x) = \sum_{i=1}^{n}(\alpha_i - \alpha_i^*)K(x, x_i) + b, \tag{6}$$

where α_i, $\alpha_i^* \ge 0$ are the Lagrangian multipliers that satisfy the equality $\alpha_i^*\alpha_i = 0$, and $K(x, x_i)$ is the kernel function. Kernel functions are used to change the dimensionality of the input space, resulting in a more confident regression (Azamathulla and Wu, 2011). The kernel function plays a critical role in SVR and its performance. There are several choices of kernel function (e.g., linear, polynomial, radial basis function, multi-layer perception, sigmoid, etc.), but the most commonly used is the radial basis function (RBF),

$$K(x, x_i) = \exp(-\gamma\|x - x_i\|^2), \tag{7}$$

here, $\gamma > 0$ is the kernel specific parameter (Hua et al., 2007; Azamathulla and Wu, 2011). This study used the RBF.

In the current study, the SVR parameters were optimized using a two-step grid search method. This can calibrate the parameters more effectively and systematically than the traditional method (e.g., trial-and-error) (Chang and Lin, 2001; Hsu et al., 2010). The first step applies a coarse grid to determine the "better" region of the grid. A finer grid search on this "better" region is then conducted to find the optimal parameters. Detailed information about this optimization method and the SVR model is given by Yu et al. (2006) and Hsu et al. (2010). We implemented the SVR models based on the LIBSVM program developed by Chang and Lin (2001).

2.2. Discrete wavelet transform (DWT)

Wavelet transform has been developed by the mathematics community over the last two decades. They give a timescale representation of time series and their relationships, and can be used to analyze series that contain non-stationarities (Torrence and Compo, 1998; Zhang et al., 2007; Adamowski and Chan, 2011). Time series data can be decomposed into different components at different resolution levels using the wavelet function (called the mother wavelet) (Tiwari and Chatterjee, 2010). The wavelet function has a wave shape with a mean value (equal to zero), and a limited, but flexible, length. It is localized in both the time and frequency domains (Liu et al., 2013). Additionally, wavelet transform can capture many properties of a time series that may not be revealed by other signal analysis techniques, such as trends, change points, self-similarity, and discontinuities (Nalley et al., 2012). There are two approaches for wavelet transform: continuous wavelet transform (CWT) and discrete wavelet transform (DWT). The CWT can detect and decompose signals on all scales, and thus requires a significant amount of computation time (Partal and Küçük, 2006). Moreover, the CWT generates too much data and includes a large amount of redundant information, resulting in more difficult data analyses (Percival, 2008). In contrast, the DWT process is simple and requires less computation time. However, it still produces a very efficient and accurate analysis (Partal and Küçük, 2006). This is attributed to the fact that the DWT is normally based on dyadic calculations (i.e., integer powers of two) of the position and scale of a signal (Chou, 2007). DWT is especially useful when the signal contains jump or shifts (Nalley et al., 2012). The wavelet function for DWT is

$$\psi_{(j,k)}\left(\frac{t-\gamma}{s}\right) = \frac{1}{s_0^{j/2}}\psi\left(\frac{t - k\gamma_0 s_0^j}{s_0^j}\right), \tag{8}$$

where ψ is the wavelet function, t is the time and γ is the translation factor (time step) of the wavelet over the time series, s indicates the scale (scale factor), j is an integer that determines the dilation (dilation factor), k is an integer that determines the translation, s_0 is a specified fixed dilation step greater than 1, and γ_0 denotes the location parameter (must be greater than zero). The most common and simplest choice for the parameters are $s_0 = 2$ and $\gamma_0 = 1$ (Mallat, 1989). For a discrete time series x_t, occurring at a discrete time t, the wavelet coefficient ($W_\psi(j, k)$) for the DWT is

$$W_\psi(j,k) = \frac{1}{2^{j/2}}\sum_{t=0}^{N-1} x_t \psi\left(\frac{t}{2^j} - k\right), \tag{9}$$

where the wavelet coefficient is evaluated at scales $s = 2^j$ and locations $\gamma = 2^j k$ to account for the variation of signals at different scales and locations (Partal and Küçük, 2006; Nalley et al., 2012). Because most streamflow time series are sampled in discrete internals, the DWT has been used for decomposing the original data in this study.

2.3. Model averaging based on Akaike's information criterion (AIC)

Model averaging produces forecasts that are not conditional on any single best model, but are instead derived from different competing models. A common approach to combine forecasts from multiple models is to consider fitting the multivariate linear regression model:

$$Y = \sum_{i=1}^{k}\beta_i X_i + \varepsilon, \tag{10}$$

where Y is the observation, X_i denotes forecasts from k different competing models (here, the different wavelet-based SVR models), and β_i are the regression parameters. ε is an error vector that is assumed to be independent and have a normal distribution with mean 0 and unknown variance (Diks and Vrugt, 2010).

In the AIC scheme, the (Akaike) weight ω_m for a given AIC model m is estimated as follows (Buckland et al., 1997; Burnham and Anderson, 2002; Symonds and Moussalli, 2011):

$$\omega_m = \frac{\exp(-I_m/2)}{\sum_{n=1}^{R}\exp(-I_n/2)}, \quad (11)$$

where I_m ($m \in R$, R is a finite family of different AIC models) is the AIC describing the fit of the model, and n denotes other AIC models ($n \in R$). The form of I_m is defined as $I_m = -2\log(L_m) + 2p$, where L_m is the (maximized) likelihood function for the AIC model and p is the number of parameters in the model. Thus, Akaike weights generate an information-theoretic measure of the evidence for each AIC model, relative to the other AIC models in R (Burnham and Anderson, 2002).

Based on the AIC technique, the averaged regression parameters β in Eq. (10) can be calculated using the following full-model averaging form (Lukacs et al., 2009):

$$\beta = \sum_{m=1}^{R}\omega_m\hat{\beta}_m, \quad (12)$$

where $\hat{\beta}_m$ is the estimated parameter in AIC model m, and ω_m is the corresponding Akaike weight. Further details on AIC averaging are given in Burnham and Anderson (2002) and Symonds and Moussalli (2011).

3. Data used

In this study, we used observed streamflow data from two US Geological Survey (USGS) stations: Station I (03371500, East Fork White River, near Bedford, Indiana, United States) and Station II (03328500, Eel River, near Logansport, Indiana, United States) (Fig. 1). Daily and monthly streamflow data from 1960 to 2009 are available for each station. These stations were chosen because they are minimally impacted by human activities (Maity et al., 2013). Table 1 gives some statistical information about the streamflow data. In this table, the terms X_{mean}, X_{min}, X_{max}, S_d, and C_{sx} denote the mean, minimum, maximum, standard deviation, and skewness, respectively. The data were divided into training and testing datasets. The training set comprised data from 1960 to 2000, whereas the testing set comprised data from 2001 to 2009.

4. Model development and application

Based on the study by Kisi and Cimen (2011), the DWT–SVR model was developed by combining the DWT and SVR methods. However, unlike previous studies, the influences of several key factors of the wavelet decomposition (wavelet choice, decomposition level, and edge effect) have been considered to improve the performance of the DWT–SVR model.

4.1. Autocorrelation and partial autocorrelation analysis

The autocorrelation and partial autocorrelation statistics were estimated (with corresponding 95% confidence levels) from lag-1 to lag-10 for the daily and monthly streamflow data from the two stations, as illustrated in Figs. 2 and 3. The statistics indicated that lag-1 had the highest correlation (significant at 95% confidence level) for the daily and monthly data of both stations. Therefore, one previous lag ($t-1$) was first considered for the input vectors to the SVR and DWT–SVR models, and the output was the streamflow for the current day or month (t).

4.2. Time series decomposition via DWT

In the proposed DWT–SVR model, the DWT was used to decompose the original daily and monthly streamflow time series for each station into several sub-series of details and approximations at different resolution levels. The details indicate the high-frequency components of the original signal, and the approximations indicate the low-frequency components. Thus, these sub-series represent the different periodic components (e.g., daily, monthly, seasonal, annually) of the original time series (Wang and Ding,

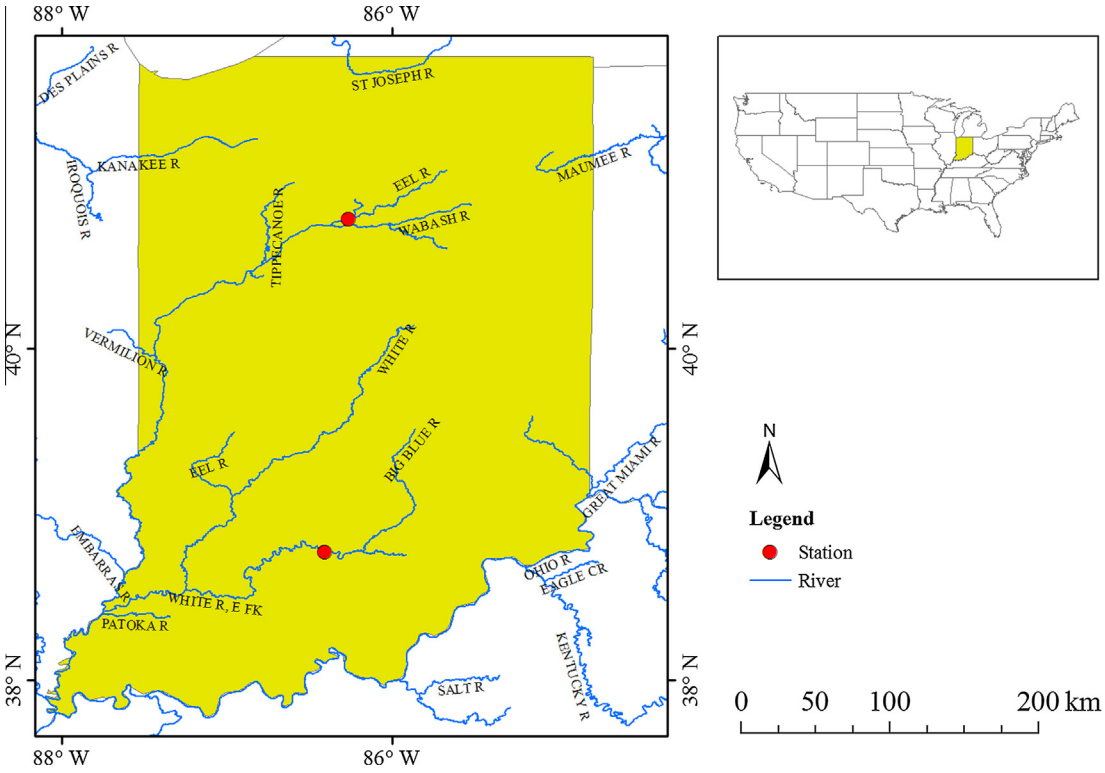

Fig. 1. Location of the two selected stations.

Table 1
Statistical information of the streamflow data.

Station	Dataset	X_{mean}	X_{min}	X_{max}	S_d	C_{sx}
Station I	Daily (m^3/s)	130.173	6.230	2579.660	178.307	3.967
	Monthly (m^3/s)	130.441	7.367	868.048	125.355	1.810
Station II	Daily (m^3/s)	22.680	1.980	470.060	32.019	4.374
	Monthly (m^3/s)	22.731	2.693	130.588	18.870	1.482

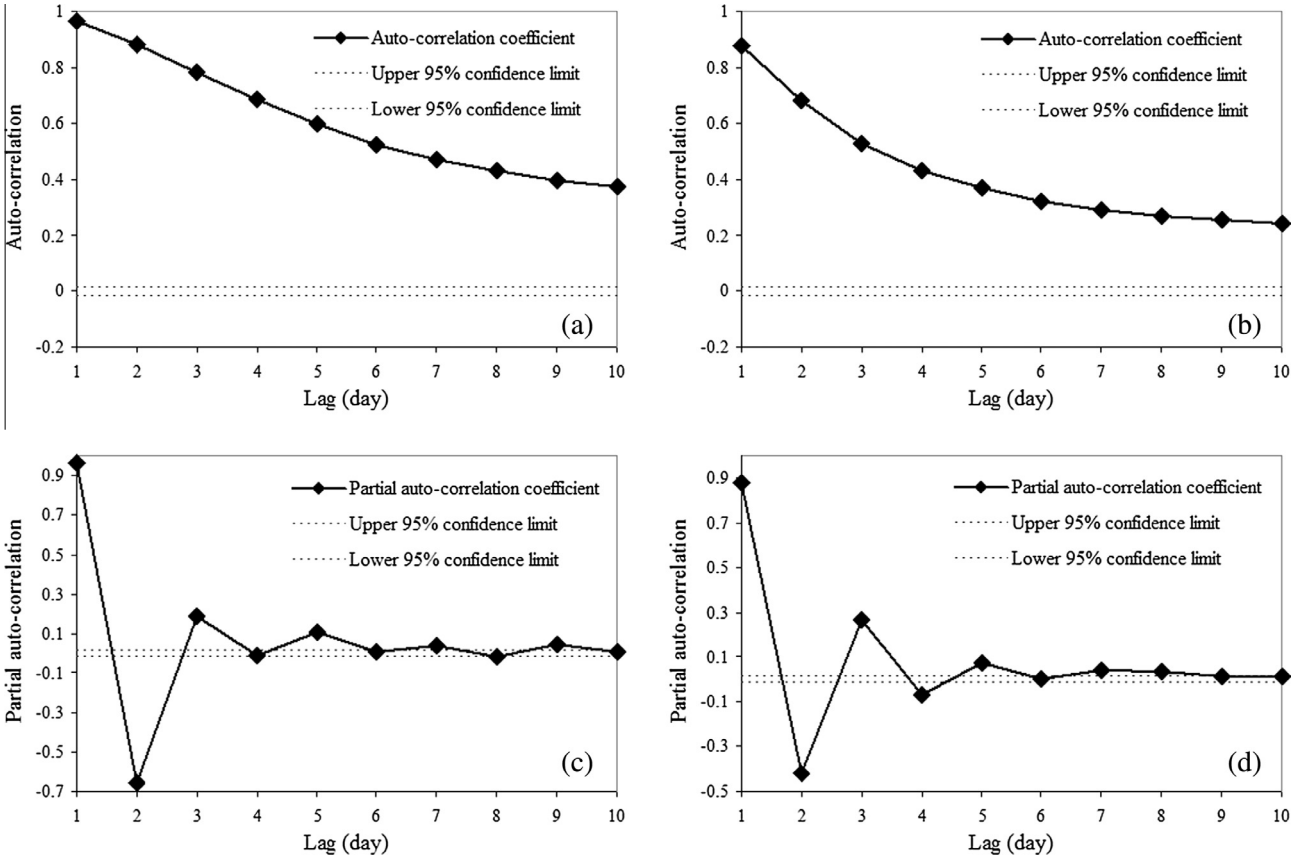

Fig. 2. Autocorrelation and partial autocorrelation functions of the daily streamflow time series of Station I (left) and Station II (right).

2003; Kisi and Cimen, 2011). The effective components were determined based on their correlation with the original streamflows. In this study, only the sub-series of details and approximations with a correlation coefficient of greater than 0.4 were considered to add as the inputs to the DWT–SVR model.

Additionally, several key factors (decomposition level, mother wavelet, and edge effect) must be appropriately chosen (Sang et al., 2013; Nalley et al., 2012), rather than arbitrarily based on personal preferences. This has been ignored by most studies when using wavelet-based methods for streamflow forecasting. To attain more reliable and accurate results, these factors were considered in the comparisons and verifications.

The choice of an appropriate mother wavelet is the foremost task for research related to wavelet analysis (Torrence and Compo, 1998; Sang et al., 2013). At present, there is no existing criterion that determines the correct mother wavelet for a specific application. In this study, a set of Daubechies (db) wavelets were considered, because they have certain features that are very important for localizing events in the time-dependent signals (Daubechies, 1990; Popivanov and Miller, 2002). For example, Daubechies wavelets provide compact support (Vonesch et al., 2007). This means that the wavelets have non-zero basis functions over a finite interval. They also provide full scaling and translational orthonormality properties (Popivanov and Miller, 2002; de Artigas et al., 2006). They are the commonly used mother wavelets for DWT analysis of hydrological–meteorological time series. Daubechies wavelets db5–db10 were used for each daily and monthly-based time series to identify the influences of different mother wavelets on streamflow forecasting.

An optimal decomposition level is important in discrete wavelet-based analysis to preserve information and reduce distortion of the original time series. In most studies of streamflow forecasting using DWT, the decomposition level is arbitrarily selected because of the lack of reliable basis (Sang et al., 2013). The decomposition level depends on the length of the time series (the number of data points) and the mother wavelet. The highest decomposition level should correspond to the data point where the last subsampling becomes smaller than the filter length. Therefore, according to the studies of de Artigas et al. (2006) and Nalley et al. (2012), the highest decomposition level, J, for the daily and monthly series of the two stations is

$$J = \frac{\mathrm{Log}\left(\frac{N}{2k-1}\right)}{\mathrm{Log}(2)}, \tag{13}$$

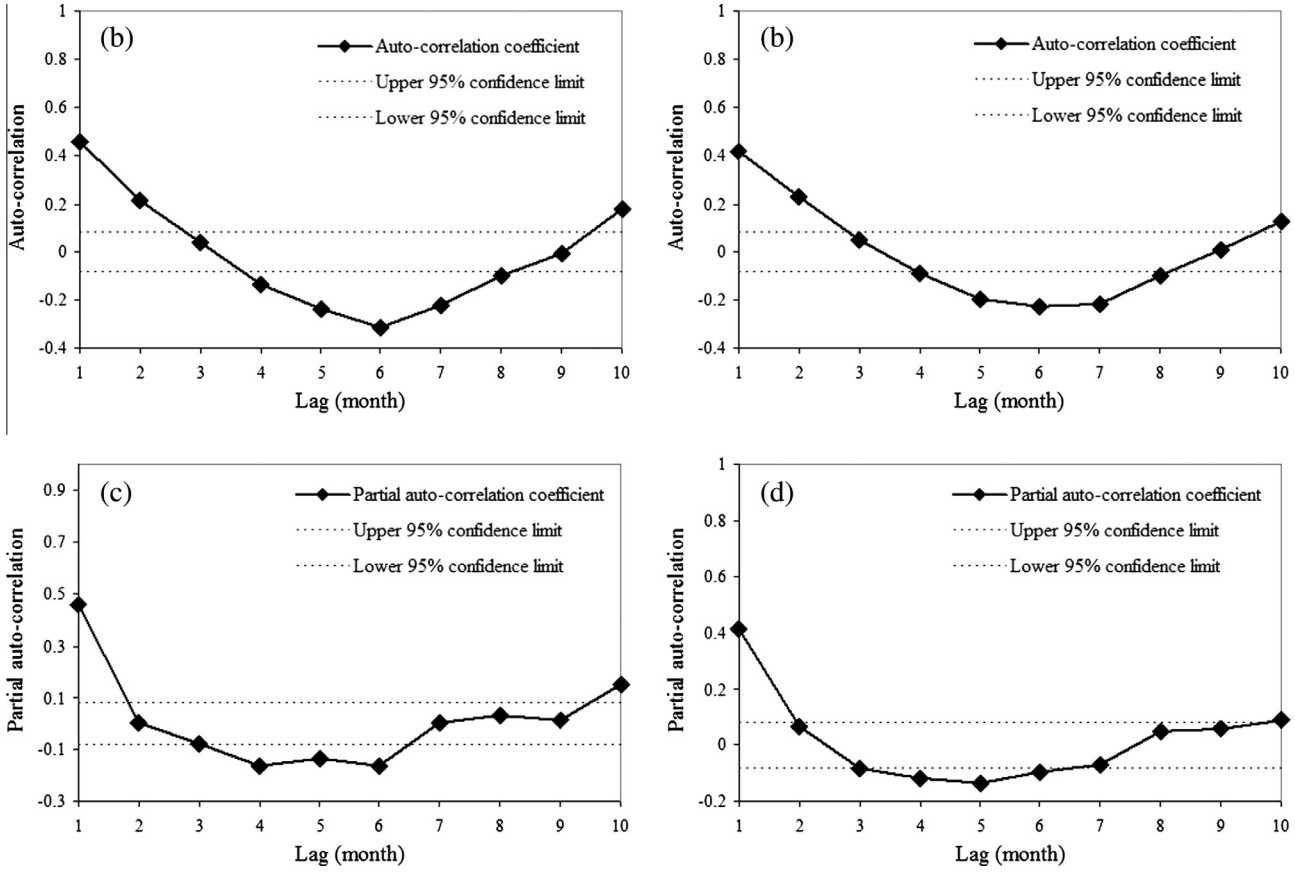

Fig. 3. Autocorrelation and partial autocorrelation functions of the monthly streamflow time series of Station I (left) and Station II (right).

where N is the length of the monthly series, and k is the number of vanishing moments of a db wavelet. This study covers the period of 50 years. There are 18,263 data points in the daily series and 600 data points in the monthly series. Daubechies wavelets (db5–db10) were used for each monthly series. The maximum levels for different Daubechies wavelets (db5–db10) were between 9.9 and 11.0 for the daily series, and between 5.0 and 6.1 for the monthly series. Therefore, 10 and 11 levels were used for daily data, and 5 and 6 levels for monthly data.

The edge effect is another factor that affects the accuracy of discrete wavelet-based analyses. The convolution process (in DWT) cannot proceed at both ends of the signal as it lacks information outside these boundaries (Nalley et al., 2012). Thus, the signal must be extended at both ends. Three common extension modes (border conditions) are: symmetrization (assuming that signals outside the original support can be recovered by symmetric boundary replication), periodic padding (assuming that signals can be recovered outside of the original support by periodic extension), and zero-padding (padding the signal with zeros beyond the original support of the wavelet) (de Artigas et al., 2006; Nalley et al., 2012). There are advantages and disadvantages to all modes, so all of them have been considered.

4.3. Coupled DWT–SVR model and ensemble prediction

In the DWT–SVR model, the streamflow time series was modeled in a similar manner as for regular SVR. The only difference is that the inputs were derived from the selected sub-series mentioned above. In this way, we obtained multiple DWT–SVR models using different settings of the proposed key factors at each station. We then used the AIC averaging technique to generate an ensemble prediction by combining individual forecasts from selected DWT–SVR models according to their forecasting performance.

4.4. Model performance measures

The performance of the SVR and DWT–SVR models was assessed using two statistical measures of goodness of fit: root-mean-squared error (RMSE) and Nash–Sutcliffe model efficiency coefficient (NSE). The RMSE is used to measure the average error magnitude. It is given by:

$$\mathrm{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2} \tag{14}$$

where N is the number of data points, $\hat{y}_i$ is the predicted value, y_i is the observed value. The smaller RMSE value, the better the performance of the model.

The NSE indicates how well the plot of the observed values versus the simulated values fits the 1:1 line. NSE ranges from $-\infty$ to 1, with larger values signifying better model performance. It is calculated as:

$$\mathrm{NSE} = 1 - \frac{\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{N}(y_i - \bar{y}_i)^2} \tag{15}$$

with the variables defined as before. $\bar{y}_i$ is the mean of the observed data values for the entire time period.

5. Results and discussions

5.1. Evaluation of DWT–SVR models for 1-day lead time forecasting

After establishing the forecasting models, the NSE and RMSE statistics of the SVR and the DWT–SVR models were compared over the testing period. Table 2 shows the performance statistics of the models at Station I for a 1-day lead time. It is interesting to note that some DWT–SVR models were not more accurate than the regular SVR model, and that there were very noticeable differences in the NSE and RMSE statistics of the different DWT–SVR models. However, the highest accuracy (NSE: 0.928, RMSE: 58.443 m^3/s) was obtained using 11 decomposition levels and db8 in periodic extension mode. The results highlight the importance of considering different decomposition levels, mother wavelets, and border conditions when using the wavelet transform. The NSE and RMSE statistics of the SVR and DWT–SVR models at Station II are given in Table 3. Again, it can be seen that the DWT–SVR models are worse than the SVR model in most cases. However, the highest NSE (0.765) and lowest RMSE (17.418 m^3/s) values were obtained by the DWT–SVR model with db6 in the periodic extension mode at 11 levels.

5.2. Evaluation of DWT–SVR models for 1-month lead time forecasting

With respect to the monthly data, the 1-month lead time forecasts over the testing period for both stations demonstrated the varying accuracy of the different DWT–SVR models. As shown in Tables 4 and 5, the NSE and RMSE significantly differed among the different border conditions, various db wavelet types and decomposition levels. At Station I (Table 4), the DWT–SVR model with the best accuracy (using periodic extension mode with 6 levels and db8) had a NSE of 0.666 and RMSE of 86.571 m^3/s. It was superior to the other DWT–SVR models and the regular SVR model. At Station II (Table 5), the highest NSE (0.541) and the smallest RMSE (14.337 m^3/s) in the testing period were also obtained in the DWT–SVR model (5 decomposition levels with zero-padding mode and db9 wavelet). It is also worthwhile to note that the DWT–SVR models for monthly streamflow forecasting at both stations were substantially superior to the SVR model in most cases.

Overall, the best DWT–SVR models provided more accurate forecasts than the SVR model for daily and monthly forecasting. In addition, the accuracy of DWT–SVR models was noticeably different when they used different border conditions, db wavelet types, and decomposition levels for both the 1-day and 1-month lead time forecasts. The conjunction of the wavelet decomposition could improve forecasting ability, or result in uncertain estimation (an even worse prediction than the SVR model). Therefore, it is important to consider the three key factors proposed in this study when using the DWT–SVR model for streamflow forecasting.

5.3. Multi-model ensemble streamflow prediction

Given the above results, we explored whether an ensemble of the forecasts from different selected DWT–SVR models with "good" performance could further improve the model performance for both 1-day and 1-month ahead predictions. The basic idea of an ensemble prediction is to combine forecasts from several candidate models. Because the performance of the candidate members is extremely important in determining the overall accuracy of multi-model ensemble forecasting, we selected the best three DWT–SVR models based on their NSE scores as candidate models (Pokhrel et al., 2013). In this study, the AIC averaging approach was used to combine the candidate models and generate the ensemble forecast. The summary statistics of 1-day ahead streamflow forecasting from the best DWT–SVR model and final ensemble model for both stations are given in Table 6. It can be observed that the ensemble model was generally superior, to some extent, to the best single DWT–SVR model at both stations. With regard to 1-month ahead streamflow forecasting, the proposed ensemble model outperformed the single best DWT–SVR model at both stations (Table 6), indicating once again the superiority of ensemble

Table 2
Performance statistics of the regular SVR and DWT–SVR models with different mother wavelets, decomposition levels, and border conditions. A 1-day lead time was considered, forecasting in the testing period (2001–2009) at Station I. The best DWT–SVR model is indicated in bold font.

Wavelet	Symmetrization				Periodic extension				Zero-padding			
	Level 10		Level 11		Level 10		Level 11		Level 10		Level 11	
	NSE	RMSE (m^3/s)	NSE	RMSE (m^3/s)	NSE	RMSE (m^3/s)	NSE	RMSE (m^3/s)	NSE	RMSE (m^3/s)	NSE	RMSE (m^3/s)
db5	0.875	77.031	0.915	63.625	0.866	79.642	0.911	65.191	0.871	78.346	0.908	66.218
db6	0.913	64.145	0.837	88.085	0.834	88.782	0.838	87.759	0.901	68.474	0.883	74.714
db7	0.853	83.488	0.835	88.496	0.914	63.798	0.894	70.920	0.859	81.857	0.918	62.542
db8	0.920	61.760	0.804	96.530	0.919	61.904	**0.928**	**58.443**	0.919	61.919	0.871	78.428
db9	0.717	115.863	0.772	103.981	0.683	122.777	0.750	109.010	0.723	114.822	0.744	110.260
db10	0.924	59.984	0.923	60.514	0.876	76.889	0.912	64.816	0.924	60.212	0.921	61.271
Regular SVR	0.920	61.667										

Table 3
Performance statistics of the regular SVR and DWT–SVR models with different mother wavelets, decomposition levels, and border conditions. A 1-day lead time was considered, forecasting in the testing period (2001–2009) at Station II. The best DWT–SVR model is indicated in bold font.

Wavelet	Symmetrization				Periodic extension				Zero-padding			
	Level 10		Level 11		Level 10		Level 11		Level 10		Level 11	
	NSE	RMSE (m^3/s)	NSE	RMSE (m^3/s)	NSE	RMSE (m^3/s)	NSE	RMSE (m^3/s)	NSE	RMSE (m^3/s)	NSE	RMSE (m^3/s)
db5	0.753	17.873	0.686	20.145	0.736	18.479	0.726	18.828	0.754	17.834	0.693	19.926
db6	0.753	17.850	0.764	17.461	0.752	17.916	**0.765**	**17.418**	0.754	17.826	0.749	18.001
db7	0.764	17.466	0.764	17.470	0.754	17.833	0.743	18.235	0.762	17.533	0.763	17.485
db8	0.728	18.758	0.701	19.652	0.698	19.765	0.762	17.543	0.727	18.789	0.703	19.579
db9	0.754	17.816	0.457	26.498	0.752	17.899	0.669	20.691	0.756	17.754	0.757	17.716
db10	0.752	17.890	0.736	18.483	0.754	17.845	0.739	18.374	0.744	18.188	0.753	17.873
Regular SVR	0.761	17.575										

Table 4
Performance statistics of the regular SVR and DWT–SVR models with different mother wavelets, decomposition levels, and border conditions. A 1-month lead time was considered, forecasting in the testing period (2001–2009) at Station I. The best DWT–SVR model is indicated in bold font.

Wavelet	Symmetrization				Periodic extension				Zero-padding			
	Level 5		Level 6		Level 5		Level 6		Level 5		Level 6	
	NSE	RMSE (m^3/s)	NSE	RMSE (m^3/s)	NSE	RMSE (m^3/s)	NSE	RMSE (m^3/s)	NSE	RMSE (m^3/s)	NSE	RMSE (m^3/s)
db5	0.578	97.358	0.540	101.556	0.591	95.758	0.583	96.778	0.566	98.680	0.577	97.488
db6	0.435	112.627	0.425	113.583	0.590	95.899	0.564	98.884	0.415	114.611	0.446	111.536
db7	0.096	142.411	0.165	136.918	0.300	125.333	0.345	121.282	0.050	146.026	0.186	135.161
db8	0.431	113.032	0.441	112.001	0.664	86.822	**0.666**	**86.571**	0.437	112.435	0.454	110.711
db9	0.261	128.771	0.186	135.164	0.330	122.615	0.365	119.385	0.325	123.089	0.281	127.009
db10	0.606	94.057	0.607	93.864	0.179	135.705	0.170	136.444	0.599	94.827	0.610	93.557
Regular SVR	0.177	135.880										

Table 5
Performance statistics of the regular SVR model and the DWT–SVR models with different mother wavelets, decomposition levels, and border conditions. A 1-month lead time was considered, forecasting in the testing period (2001–2009) at Station II. The best DWT–SVR model is indicated in bold format.

Wavelet	Symmetrization				Periodic extension				Zero-padding			
	Level 5		Level 6		Level 5		Level 6		Level 5		Level 6	
	NSE	RMSE (m^3/s)	NSE	RMSE (m^3/s)	NSE	RMSE (m^3/s)	NSE	RMSE (m^3/s)	NSE	RMSE (m^3/s)	NSE	RMSE (m^3/s)
db5	0.393	16.477	0.408	16.281	0.380	16.654	0.395	16.457	0.400	16.394	0.403	16.349
db6	0.402	16.364	0.418	16.144	0.366	16.843	0.198	18.952	0.411	16.243	0.418	16.137
db7	0.308	17.599	0.354	17.010	0.471	15.386	0.488	15.139	0.305	17.641	0.391	16.506
db8	0.376	16.713	0.511	14.790	0.452	15.661	0.490	15.103	0.429	15.981	0.496	15.015
db9	0.500	14.960	0.533	14.462	0.371	16.779	0.449	15.702	**0.541**	**14.337**	0.540	14.349
db10	0.300	17.698	0.362	16.902	0.322	17.426	0.300	17.707	0.356	16.974	0.413	16.215
Regular SVR	0.280	17.955										

Table 6
Summary statistics of the best DWT–SVR models and ensemble models for 1-day and 1-month ahead forecasting in the testing period at Station I and Station II.

Station	1-day ahead forecasting				1-month ahead forecasting			
	Best DWT–SVR model		Ensemble model		Best DWT–SVR model		Ensemble model	
	NSE	RMSE (m^3/s)	NSE	RMSE (m^3/s)	NSE	RMSE (m^3/s)	NSE	RMSE (m^3/s)
Station I	0.928	58.443	0.929	58.065	0.666	86.571	0.671	85.894
Station II	0.765	17.418	0.783	16.729	0.541	14.337	0.637	12.745

forecasting. This suggests the proposed ensemble model is able to provide an effective means of combining forecasts from multiple models.

The ensemble prediction extends the range of streamflow forecasting applications when using the hybrid DWT–SVR model (or other wavelet-based data-driven models). It is superior to non-ensemble DWT–SVR models in both daily and monthly streamflow forecasting. Therefore, the multi-model ensemble prediction is a promising alternative to the single best DWT–SVR model. It also provides helpful insights into other hybrid wavelet-based data-driven methods (e.g., wavelet–ANN models) regarding applications in streamflow forecasting. In addition to the model averaging technique used in this study, there are some other feasible approaches for model averaging such as Bayesian information criterion averaging, Granger–Ramanathan averaging, ANN-based weight estimation, and Bayesian model averaging (BMA) (Granger and Ramanathan, 1984; Ajami et al., 2007; Diks and Vrugt, 2010). These approaches could also be applied to combine forecasts from different candidate models and would potentially improve the forecasting performance of hybrid DWT–SVR models.

5.4. Longer lead time forecasting of both daily and monthly streamflows

The DWT–SVR model was then applied to 2- and 3-day ahead daily streamflow forecasting, and 2-month ahead monthly forecasting. Again, different border conditions, various db wavelets, and different decomposition levels were considered for the DWT–SVR models, and an ensemble prediction was generated by combining candidate competing models using the AIC averaging approach. For the daily streamflow forecasts, the NSE and RMSE statistics of different models at both stations are shown in Table 7. It is clear that the ensemble models were consistently more accurate than the best DWT–SVR and regular SVR models. Table 8 gives the monthly streamflow forecasts for a lead time of two months. These results indicated that the ensemble models outperformed the others to different magnitudes. However, it should be noted that the increased lead times diminished the forecast reliability, and significantly affected model performance, particularly for monthly streamflow forecasting. This is expected in data-driven time series forecasting, because the autocorrelation in the time series decreases and the time series becomes much less predictable as the lead time increases, regardless of the forecasting technique used (Tiwari and Chatterjee, 2010; Dhanya and Nagesh Kumar, 2011). It can also be seen that the DWT–SVR model exhibited different forecasting skills at each station. This suggests that the predictability also depends on the nature of the streamflow time series, which could be strongly affected by the factors including local micrometeorological conditions (e.g., rainfall and soil moisture) and topographic features in a certain sub-climate zone (Dhanya and Nagesh Kumar, 2011).

Overall, the ensemble model increases the forecasting skill of DWT–SVR models. However, further efforts are still required to improve the model performance for streamflow forecasting,

Table 7
Summary statistics of the regular SVR models, the best DWT–SVR model and ensemble models for daily streamflow forecasting over the lead times of 2–3 days in the testing period at Station I and Station II.

Lead time (days)	Regular SVR model		Best DWT–SVR model		Ensemble model	
	NSE	RMSE (m^3/s)	NSE	RMSE (m^3/s)	NSE	RMSE (m^3/s)
Station I						
2	0.754	108.141	0.765	105.573	0.781	102.117
3	0.572	142.665	0.631	132.371	0.639	131.026
Station II						
2	0.409	27.640	0.494	25.562	0.508	25.221
3	0.208	31.987	0.436	27.004	0.453	26.586

Table 8
Summary statistics of the regular SVR model, the best DWT–SVR model and ensemble model for monthly streamflow forecasting over the lead time of 2 months in the testing period at Station I and Station II.

Lead time (months)	Regular SVR model		Best DWT–SVR model		Ensemble model	
	NSE	RMSE (m^3/s)	NSE	RMSE (m^3/s)	NSE	RMSE (m^3/s)
Station I						
2	0.001	149.761	0.216	132.663	0.227	131.753
Station II						
2	−0.211	23.285	0.180	19.163	0.297	17.745

particularly for monthly streamflow predictions. Considering additional predictors (e.g., local rainfall, temperature, relative humidity, climate indices) as the model inputs would help to increase the forecasting skill of the DWT–SVR and ensemble models.

6. Conclusions

This study first investigated the capability and effectiveness of the SVR model coupled with wavelet transform (DWT) for daily and monthly streamflow forecasting. We considered three important factors of the wavelet decomposition phase (decomposition levels, mother wavelets, and edge effect) that strongly affect the performance of the DWT–SVR model. The DWT–SVR models that use different combinations of these factors were evaluated using RMSE and NSE. The forecasting skill of these DWT–SVR models was tested using daily and monthly streamflow data from two hydrological stations located in Indiana, United States. The test results were compared with the regular SVR model. The 1-day and 1-month ahead streamflow forecasts for both stations demonstrated noticeable differences in the DWT–SVR models with different combinations of decomposition levels, mother wavelets, and edge effect. In addition, they were not always more accurate than the regular SVR model. Therefore, we strongly recommended that future applications combining the SVR model (or other machine learning methods such as artificial neural networks) with wavelet transform consider these three factors when decomposing the time series using DWT.

The AIC averaging approach was then used to combine forecasts from the candidate DWT–SVR models with "good" forecasting performance. This provided a more reliable prediction for both daily and monthly streamflows. Longer lead-time predictions (i.e., 2- and 3-day and 2-month lead times) were also conducted using the DWT–SVR models and the ensemble model. The ensemble models consistently outperformed the regular SVR model and best single DWT–SVR model for all lead times and at both stations. This highlights the strength of the proposed ensemble approach for streamflow prediction. In future work, more predictors (e.g., local rainfall, temperature, relative humidity, and climate indices) will be considered as the model inputs. We expect this to further improve the forecasting performance of the DWT–SVR and ensemble models. Additionally, we will investigate and compare different model averaging strategies (e.g., Granger–Ramanathan averaging and BMA) in terms of their potential to increase the forecasting skill of wavelet-based data-driven models.

Acknowledgements

The study was financially supported by the State Forestry Administration Public Benefit Research Foundation of China (201204104, 201104005) and State Forestry Administration 948 Innovative Significant Project of China (2011-76). Special appreciation is given to the USGS for providing the streamflow data. We gratefully acknowledge the helpful comments of the editor and anonymous reviewers.

References

Adamowski, J.F., 2008. Development of a short-term river flood forecasting method for snowmelt driven floods based on wavelet and cross-wavelet analysis. J. Hydrol. 353 (3–4), 247–266.
Adamowski, J., 2013. Using support vector regression to predict direct runoff, base flow and total flow in a mountaineous watershed with limited data. Land Reclam. 45 (1), 71–83.
Adamowski, J., Chan, H.F., 2011. A wavelet neural network conjunction model for groundwater level forecasting. J. Hydrol. 407, 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 (1–2), 85–91.
Ajami, N.K., Duan, Q., Sorooshian, S., 2007. An integrated hydrologic Bayesian multimodel combination framework: confronting input, parameter, and model structural uncertainty in hydrologic prediction. Water Resour. Res. 43, W01403. http://dx.doi.org/10.1029/2005WR004745.
Azamathulla, H.M., Wu, F.C., 2011. Support vector machine approach to for longitudinal dispersion coefficients in streams. Appl. Soft Comput. 11, 2902–2905.
Bray, M., Han, D., 2004. Identification of support vector machines for runoff modeling. J. Hydroinform. 6 (4), 265–280.
Buckland, S.T., Burnham, K.P., Augustin, N.H., 1997. Model selection: an integral part of inference. Biometrics 53, 603–618.
Burnham, K.P., Anderson, D.R., 2002. Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach, second ed. Springer, New York, USA.
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 (18), 1164–1171.
Cao, L.J., Tay, F.E.H., 2001. Financial forecasting using support vector machines. Neural Comput. Appl. 10, 184–192.
Chang, C.C., Lin, C.J., 2001. LIBSVM: A Library for Support Vector Machines. <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>.
Chou, C.M., 2007. Applying multi-resolution analysis to differential hydrological grey models with dual series. J. Hydrol. 332 (1–2), 174–186.

Daubechies, I., 1990. The wavelet transform, time–frequency localization and signal analysis. IEEE Trans. Inform. Theory 36 (5), 6–7.
de Artigas, M.Z., Elias, A.G., de Campra, P.F., 2006. Discrete wavelet analysis to assess long-term trends in geomagnetic activity. Phys. Chem. Earth 31 (1–3), 77–80.
Dhanya, C.T., Nagesh Kumar, D., 2011. Predictive uncertainty of chaotic daily streamflow using ensemble wavelet networks approach. Water Resour. Res. 47, W06507. http://dx.doi.org/10.1029/2010WR010173.
Diks, C.G.H., Vrugt, J.A., 2010. Comparison of point forecast accuracy of model averaging methods in hydrologic applications. Stoch. Environ. Res. Risk Assess. 24 (6), 809–820.
Granger, C.W.J., Ramanathan, R., 1984. Improved methods of combining forecast accuracy. J. Forecast. 3, 197–204.
Haykin, S., 2003. Neural Networks: A Comprehensive Foundation. Fourth Indian Reprint. Pearson Education, Singapore, pp. 842.
Hong, W.C., 2011. Traffic flow forecasting by seasonal SVR with chaotic simulated annealing algorithm. Neurocomputing 74 (12–13), 2096–2107.
Hsu, C.W., Chang, C.C., Lin, C.J., 2010. A Practical Guide to Support Vector Classification. <http://www.csie.ntu.edu.tw/~cjlin/papers/guide/guide.pdf>.
Hua, X., Ni, Y., Ko, J., Wong, K., 2007. Modeling of temperature–frequency correlation using combined principal component analysis and support vector regression technique. J. Comput. Civ. Eng. 21 (2), 122–135.
Kalteh, A.M., 2013. Monthly River flow forecasting using artificial neural network and support vector regression models coupled with wavelet transform. Comput. Geosci. 54, 1–8.
Kao, L.J., Chiu, C.C., Lu, C.J., Yang, J.L., 2013. Integration of nonlinear independent component analysis and support vector regression for stock price forecasting. Neurocomputing 99, 534–542.
Kim, G., Barros, A.P., 2001. Quantitative flood forecasting using multisensor data and neural networks. J. Hydrol. 246, 45–62.
Kisi, O., 2008. Stream flow forecasting using neuro-wavelet technique. Hydrol. Process. 22 (20), 4142–4152.
Kisi, O., 2009. Neural networks and wavelet conjunction model for intermittent stream flow forecasting. J. Hydrol. Eng. 14 (8), 773–782.
Kisi, O., Cimen, M., 2011. A wavelet-support vector machine conjunction model for monthly streamflow forecasting. J. Hydrol. 399, 132–140.
Kisi, O., Cimen, M., 2012. Precipitation forecasting by using wavelet-support vector machine conjunction model. Eng. Appl. Artif. Intell. 25, 783–792.
Li, L., 2011. Water resource requirement prediction based on the wavelet–bootstrap–Svm hybrid model. J. Inf. Comput. Sci. 8 (13), 2563–2568.
Lin, J.Y., Cheng, C.T., Chau, K.W., 2006. Using support vector machines for long-term discharge prediction. Hydrol. Sci. J. 51 (4), 599–612.
Liong, S., Sivapragasam, C., 2002. Flood stage forecasting with support vector machines. J. Am. Water Resour. Assoc. 38 (1), 173–186.
Liu, Z., Zhou, P., Zhang, F., Liu, X., Chen, G., 2013. Spatiotemporal characteristics of dryness/wetness conditions across Qinghai Province, Northwest China. Agric. For. Meteorol. 182–183, 101–108.
Lukacs, P.M., Burnham, K.P., Anderson, D.R., 2009. Model selection bias and Freedman's paradox. Ann. Inst. Stat. Math. 62, 117–125.
Maity, R., Bhagwat, P.P., Bhatnagar, A., 2010. Potential of support vector regression for prediction of monthly streamflow using endogenous property. Hydrol. Process. 24, 917–923.
Maity, R., Ramadas, M., Govindaraju, R.S., 2013. Identification of hydrologic drought triggers from hydroclimatic predictor variables. Water Resour. Res. 49, 4476–4492.
Makkeasorn, A., Chang, N.B., Zhou, X., 2008. Short-term streamflow forecasting with global climate change implications—a comparative study between genetic programming and neural network models. J. Hydrol. 352, 336–354.
Mallat, S.G., 1989. A theory for multiresolution signal decomposition: the wavelet representation. IEEE Trans. Pattern Anal. Mach. Intell. 11 (7), 674–693.
McKerchar, A.I., Delleur, J.W., 1974. Application of seasonal parametric linear stochastic models to monthly flow data. Water Resour. Res. 10, 246–255.
Nalley, D., Adamowski, J., Khalil, B., 2012. Using discrete wavelet transforms to analyze trends in streamflow and precipitation in Quebec and Ontario (1954–2008). J. Hydrol. 475, 204–228.
Nourani, V., Komasi, M., Mano, A., 2009. A multivariate ANN–wavelet approach for rainfall–runoff modeling. Water Resour. Manage. 23 (14), 2877–2894.
Nourani, V., Kisi, O., Komasi, M., 2011. Two hybrid artificial intelligence approaches for modeling rainfall–runoff process. J. Hydrol. 402 (1–2), 41–59.
Partal, T., Küçük, M., 2006. Long-term trend analysis using discrete wavelet components of annual precipitations measurements in Marmara region (Turkey). Phys. Chem. Earth 31 (18), 1189–1200.
Percival, D.B., 2008. Analysis of geophysical time series using discrete wavelet transforms: an overview. In: Donner, R.V., Barbosa, S.M. (Eds.), Nonlinear Time Series Analysis in the Geosciences – Applications in Climatology, Geodynamics, and Solar-terrestrial Physics. Springer, Berlin/Heidelberg.
Pokhrel, P., Wang, Q.J., Robertson, D.E., 2013. The value of model averaging and dynamical climate model predictions for improving statistical seasonal streamflow forecasts over Australia. Water Resour. Res. 49, 6671–6687. http://dx.doi.org/10.1002/wrcr.20449.
Popivanov, I., Miller, R.J., 2002. Similarity search over time-series data using wavelets. In: Proceedings 18th International Conference on Data Engineering, pp. 212–221.
Pramanik, N., Panda, R.K., Singh, A., 2010. Daily river flow forecasting using wavelet ANN hybrid models. J. Hydroinform. 13 (1), 49–63.
Rasouli, K., Hsieh, W.W., Cannon, A.J., 2012. Daily streamflow forecasting by machine learning methods with weather and climate inputs. J. Hydrol. 414–415, 284–293.
Remesan, R., Shamim, M.A., Han, D., Mathew, J., 2009. Runoff prediction using an integrated hybrid modelling scheme. J. Hydrol. 372, 48–60.
Sang, Y.F., 2013. A review on the applications of wavelet transform in hydrology time series analysis. Atmos. Res. 122, 8–15.
Sang, Y.F., Wang, Z., Liu, C., 2013. Discrete wavelet-based trend identification in hydrologic time series. Hydrol. Process. 27, 2021–2031.
Shiri, J., Kisi, O., 2010. Short-term and long-term stream flow forecasting using a wavelet and neuro-fuzzy conjunction model. J. Hydrol. 394 (3–4), 486–493.
Sivapragasam, C., Liong, S.Y., Pasha, M.F.K., 2001. Rainfall and runoff forecasting with SSA–SVM approach. J. Hydroinform. 3 (3), 141–152.
Symonds, M.R.E., Moussalli, A., 2011. A brief guide to model selection, multimodel inference and model averaging in behavioural ecology using Akaike's information criterion. Behav. Ecol. Sociobiol. 65, 13–21.
Tiwari, M.K., Chatterjee, C., 2010. Development of an accurate and reliable hourly flood forecasting model using wavelet–bootstrap–ANN (WBANN) hybrid approach. J. Hydrol. 1 (394), 458–470.
Tiwari, M.K., Chatterjee, C., 2011. A new wavelet–bootstrap–ANN hybrid model for daily dischargeforecasting. J. Hydroinform. 13 (3), 500–519.
Torrence, C., Compo, G.P., 1998. A practical guide to wavelet analysis. Bull. Am. Meteorol. Soc. 79, 61–78.
Vapnik, V., 1995. The Nature of Statistical Learning Theory. Springer Verlag, New York, USA.
Vapnik, V.N., 2000. The Nature of Statistical Learning Theory. Springer, Berlin.
Vapnik, V., Golwich, S., Smola, A.J., 1997. Support vector method for function approximation, regression estimation, and signal processing. In: Mozer, M., Jordan, M., Petsche, T. (Eds.), Advances in Neural Information Processing Systems 9. MIT Press, Cambridge, Massachusetts, USA, pp. 281–287.
Vonesch, C., Blu, T., Unser, M., 2007. Generalized Daubechies wavelet families. IEEE Trans. Signal Process. 55 (9), 4415–4429.
Wang, W., Ding, J., 2003. Wavelet network model and its application to the prediction of the hydrology. Nat. Sci. 1 (1), 67–71.
Wang, W.C., Chau, K.W., Cheng, C.T., Qiu, L., 2009. A comparison of performance of several artificial intelligence methods for forecasting monthly discharge time series. J. Hydrol. 374 (3–4), 294–306.
Wu, M.C., Lin, G.F., Lin, H.Y., 2012. Improving the forecasts of extreme streamflow by support vector regression with the data extracted by self-organizing map. Hydrol. Process. http://dx.doi.org/10.1002/hyp.9584.
Yu, P.S., Chen, S.T., Chang, I.F., 2006. Support vector regression for real-time flood stage forecasting. J. Hydrol. 328 (3–4), 704–716.
Zhang, Q., Xu, C.Y., Jiang, T., Wu, Y.J., 2007. Possible influence of ENSO on annual maximum streamflow of the Yangtze River, China. J. Hydrol. 333, 265–274.