https://doi.org/10.5194/hess-2019-565
Preprint. Discussion started: 20 December 2019

Hydrology and Earth System Sciences
Discussions

Two-stage Variational Mode Decomposition and Support Vector Regression for Streamflow Forecasting

Ganggang Zuo, Jungang Luo, Ni wang, Yani Lian, Xinxin He

State Key Laboratory of Eco-hydraulics in Northwest Arid Region, Xi'an University of Technology, Xi'an, Shaanxi 710048, China

Correspondence to: Jungang Luo (jgluo@xaut.edu.cn)

Abstract. Streamflow forecasting is a crucial component in the management and control of water resources. Decomposition-based approaches have particularly demonstrated improved streamflow forecasting performance. However, it is not practical to firstly decompose the entire streamflow into several signal components and then divide the data samples of each component into training and validation sets for signal component prediction. This impracticality is due to the fact that some validation information, that is not available in practical streamflow forecasting, is used in that training process. Unfortunately, firstly dividing the entire streamflow into training and validation sets and then decomposing each set separately lead to undesirable boundary effects and complicated forecasting. Moreover, establishing a model for each signal component is quite laborious and summing the component predictions may lead to error accumulation. In addition, summing the decomposition results may sometimes lead to inaccurate reconstruction of the original streamflow. In order to address these shortcomings of decomposition-based models and improve the forecasting performance in basins lacking meteorological observations (e.g., precipitation and temperature), we propose a two-stage decomposition prediction (TSDP) framework, realize this framework using variational mode decomposition (VMD) and support vector machines (SVR), and refer to this realization as VMD-SVR. In the first stage of the TSDP framework, the entire streamflow data was divided into training and validation sets, each of which was then separately decomposed to avoid the influence of validation information on training. In the second stage, a single model for streamflow prediction was established using a set of mixed shuffled samples. This scheme saves the modelling time and reduces the influence of the boundary effects. We demonstrate experimentally the effectiveness, efficiency and reliability of the TSDP framework and its VMD-SVR realization in terms of the boundary effect reduction, decomposition performance, prediction outcomes, time consumption, overfitting, and forecasting capability for long leading times. Specifically, five comparative experiments were conducted based on the ensemble empirical mode decomposition (EEMD), singular spectrum analysis (SSA), discrete wavelet transform (DWT) and SVR. The experimental results on monthly runoff collected from three stations at the Wei River show the superiority of the TSDP framework compared to benchmark models.

https://doi.org/10.5194/hess-2019-565
Preprint. Discussion started: 20 December 2019

Hydrology and Earth System Sciences
Discussions

Nomenclature			
TSDP	Two-stage decomposition prediction	VMD	Variational mode decomposition
TSDPE	Three-stage decomposition ensemble	SVR	Support vector regression
EEMD	Ensemble empirical mode decomposition	WA	Wavelet transform
SSA	Singular spectrum analysis	IMF	Intrinsic mode function
DWT	Discrete wavelet transform	PACF	Partial autocorrelation coefficient
VMD-SVR	A TSDP model based on VMD and SVR	ML	Maximum likelihood
EEMD-SVR	A TSDP model based on EEMD and SVR	NSE	Nash–Sutcliffe efficiency
SSA-SVR	A TSDP model based on SSA and SVR	NRMSE	Normalized root mean square error
DWT-SVR	A TSDP model based on DWT and SVR	PPTS	Peak percentage of threshold statistic
VMD-SVR-A	A TSDPE model based on VMD and SVR with addition ensemble		
EEMD-SVR-A	A TSDPE model based on EEMD and SVR with addition ensemble		
SSA-SVR-A	A TSDPE model based on SSA and SVR with addition ensemble		
DWT-SVR-A	A TSDPE model based on DWT and SVR with addition ensemble		

1 Introduction

Reliable and accurate streamflow forecasting is of great significance for water resource management (Woldemeskel et al., 2018). The first attempts for streamflow prediction were based on precipitation measurements, and date back to the 19th century (Mulvaney, 1850; Todini, 2007). Since then, streamflow forecasting models have been further developed by analyzing relevant physical processes and incorporating key hydrological terms into the formulation of mathematical models (Kratzert et al., 2018). These hydrological terms include physical characteristics and boundary conditions of the catchments, as well as spatial and temporal variabilities of hydrologic processes (Kirchner, 2006; Paniconi and Putti, 2015). In addition, the computing power and the availability of hydrometeorological and remote sensing data largely drive the development of physically-based models (Singh, 2018; Clark et al., 2015).

However, modelling hydrologic processes with spatial and temporal variabilities at the catchment scale requires a lot of input meteorological data, parameters that define the boundary conditions and physical properties, as well as high-performance computational resources (Binley et al., 1991; Devia et al., 2015). Moreover, current physically-based models do not perform consistently on all scales and datasets because hydrological process models are mostly available for small watersheds only (Kirchner, 2006; Beven, 1989; Grayson et al., 1992; Abbott et al., 1986). Therefore, physically-based models have been rarely used for practical streamflow forecasting (Kratzert et al., 2018). In this context, numerous studies have explored and developed data-driven models such as time-series and machine-learning models (Wu et al., 2009).

Previous studies predicted streamflow through the application of time-series models such as Box-Jenkins (Castellano-Méndez et al., 2004), autoregression (AR), moving average (MA), autoregressive moving average (ARMA) and autoregressive integrated moving average (ARIMA) (Li et al., 2015; Mohammadi et al., 2006; Kisi, 2010; Valipour et al., 2013). However, the linear hypothesis of time-series models makes them unsuitable for forecasting of nonstationary, nonlinear and variable streamflow patterns. Therefore, maximum likelihood (ML) models with nonlinear mapping capabilities have been applied to

https://doi.org/10.5194/hess-2019-565
Preprint. Discussion started: 20 December 2019


Hydrology and Earth System Sciences
Discussions

forecast streamflow. These models include decision trees (DT) (Erdal and Karakurt, 2013; Solomatine et al., 2008; Han et al., 2002), support vector regression (SVR) (Yu et al., 2006; Maity et al., 2010; Hosseini and Mahjouri, 2016), fuzzy inference systems (FIS) (Ashrafi et al., 2017; He et al., 2014; Yaseen et al., 2017) and artificial neural networks (ANN) (Kratzert et al., 2018; Nourani et al., 2009; Tiwari and Chatterjee, 2010; Rasouli et al., 2012).

However, pure ML models cannot always adequately handle highly nonstationary, complex, nonlinear, and multiscale streamflow series in catchments due to the lack of meteorological observations. Therefore, signal processing algorithms have been applied to transform the nonstationary series data into relatively stationary components, which can be analyzed more easily. The most common of these algorithms are those based on wavelet analysis (WA) (Liu et al., 2014; Adamowski and Sun, 2010), empirical mode decomposition (EMD) (Huang et al., 2014; Meng et al., 2019), ensemble empirical mode decomposition (EEMD) (Bai et al., 2016; Zhao and Chen, 2015), singular spectrum analysis (SSA) (Zhang et al., 2015; Sivapragasam et al., 2001), seasonal-trend decomposition based on loess (STL) (Luo et al., 2019) and variational mode decomposition (VMD) (He et al., 2019; Xie et al., 2019). These approaches show improved streamflow forecasting through flow decomposition.

However, the aforementioned decomposition-based methods are based on an implicit assumption that the decomposition results are barely affected by boundary effects (Zhang et al., 2015). The boundary effects cause the decomposition results at the end of a streamflow series to deviate from the real streamflow decomposition data, which was obtained using the future observed data as input. In fact, each of these decomposition approaches firstly decomposed the entire streamflow data and then divided the decomposed components into training and validation sets for streamflow prediction. This generally leads to the incorporation of impractically-available validation information in the training process. The use of the validation information enables the reduction of the boundary effects, which is a crucial step for any operational streamflow forecasting algorithm. However, to avoid using impractically-available validation information, the streamflow time series data must be first divided into training and validation sets, where each set is separately decomposed, while the boundary effects are reduced.

Other relevant research contributions are those of Zhang et al. (2015), Du et al. (2017) , Tan et al. (2018), Quilty and Adamowski (2018), and Fang et al. (2019) who recently pointed out and explicitly criticized the afore-mentioned unpractical (and even incorrect) usage of signal processing techniques. Zhang et al. (2015) evaluated and compared the outcomes of hindcast and forecast experiments (with and without validation information, respectively) for decomposition models based on WA, EMD, SSA, ARMA and ANN. The authors suggested that the decomposition-based models may not be suitable for practical streamflow forecasting. Du et al. (2017) demonstrated that the usage of SSA and the discrete wavelet transform (DWT) directly on the entire hydrological time series is incorrect. Tan et al. (2018) assessed the impractical usage of forecasting scheme based on the EEMD and ANN. Quilty and Adamowski (2018) addressed the incorrect usage of the wavelet-based models for hydrological forecasting. Fang et al. (2019) demonstrated that EMD is not suitable for practical streamflow forecasting. In summary, these contributions have demonstrated that the incorrect usage often leads to much better performance which is practically unachievable.

https://doi.org/10.5194/hess-2019-565
Preprint. Discussion started: 20 December 2019


Hydrology and Earth System Sciences
Discussions

However, the boundary effects constitute a great challenge in practical streamflow forecasting. Zhang et al. (2015) tested several extended mathematical models to reduce the boundary effects and suggested that a properly-designed extension model can improve the forecasting performance. Quilty and Adamowski (2018) proposed a new wavelet data-driven forecasting framework, in which boundary-affected coefficients were removed by adopting either the stationary wavelet transform (SWT) algorithm (also known as "*algorithme à trous*") or the maximal-overlap discrete wavelet transform. Based on the abovementioned streamflow forecasting approaches, we would like to highlight three desirable improvements to attain more reliable and efficient decomposition-based models. In particular, these improvements are: (i) finding simpler and more efficient ways of reducing the boundary effects; (ii) reducing the computational and error accumulation costs that result from using one optimized data-driven model for each signal component and summing up the component predictions; and (iii) decreasing the streamflow reconstruction inaccuracy resulting from inadequate noise removal and summation of predictions from multiple components.

To bridge the gap between the existing and desirable streamflow forecasting characteristics, we designed a two-stage decomposition prediction (TSDP) framework, proposed a TSDP realization based on VMD and SVR, and referred to this realization as VMD-SVR. The proposed scheme can reduce the boundary effects, save the modelling time, and improve the prediction performance. This practical streamflow forecasting framework can be outlined as follows:

(1) Divide the entire streamflow data into training and validation sets and decompose each of these two sets separately into signal components. This procedure avoids using the validation information for training purposes.

(2) Combine the predictors of individual signal components into a final predictors, and select the original streamflow data as the prediction target in order to build only one optimized prediction model.

(3) Generate training and validation samples and divide the validation samples into development and testing samples. Mix and shuffle the training and the development samples to optimize the prediction model, and reduce the boundary effects.

The aim of this paper is to demonstrate the effectiveness, efficiency and reliability of the designed TSDP framework and its VMD-SVR realization. Specifically, we design five comparative experiments which help in the assessment of different performance aspects including the reduction of the boundary effects, decomposition performance, predictability, time consumption, overfitting, and forecasting capabilities for long leading times. In the first experiment, data mixing and shuffling are carried out to reduce the boundary effects. In the second experiment, we compare the EEMD, SSA, DWT, and the VMD methods to demonstrate the superiority of the VMD method for one-month-ahead streamflow forecasting. In the third experiment, we establish a three-stage decomposition ensemble (TSDPE) model, in which one optimized SVR model is built for each signal component. This ensemble model demonstrates that the designed TSDP framework saves the modelling time and improves the prediction performance. For the fourth experiment, we demonstrate that the proposed VMD-SVR realization, which combines the predictors of the individual signal components for the final predictors, barely overfits the original streamflow data. In the last experiment, we present the VMD-SVR prediction capability for long leading times, which extend for more than one month. We use monthly runoff collected at three stations located in the Wei River of China to evaluate the performance of the proposed model against the benchmark models.

https://doi.org/10.5194/hess-2019-565
Preprint. Discussion started: 20 December 2019

Hydrology and Earth System Sciences
Discussions

2 Methodologies

2.1 Variational mode decomposition

The VMD algorithm proposed by Dragomiretskiy and Zosso (2014) concurrently decomposes an input signal ($f(t)$) into K intrinsic mode functions (IMFs) with different center frequencies and limited bandwidths. An IMF ($u_k(t)$) is an amplitude-modulated-frequency-modulated (AM-FM) signal, which is defined as:

$$u_k(t) = A_k(t)\cos\big(\phi_k(t)\big), \tag{1}$$

where $A_k(t)$ and $\phi_k(t)$ are the instantaneous amplitude and phase, respectively.

The VMD decomposition process is mainly divided into two steps, namely (a) constructing a variational problem and (b) solving this problem. The variational problem is constructed by the following scheme:

(1) Use the Hilbert transform to compute the associated analytic signal and obtain a unilateral frequency spectrum.

(2) Shift the frequency spectrum of each mode to a "baseband" by mixing with an exponential function tuned to the respective evaluated center frequency.

(3) Assess the bandwidth by the H^1 Gaussian smoothness of the demodulated signal, i.e., the squared L^2-norm of the gradient.

The constructed variational problem is expressed as follows:

$$\begin{cases} \min\limits_{\{u_k\}\{\omega_k\}} \left\{ \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t}\right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 \right\}, \\ s.t. \quad \sum_k u_k(t) = f(t) \end{cases} \tag{2}$$

where $\{u_k\} = \{u_1, u_2, \cdots, u_k\}$ and $\{\omega_k\} = \{\omega_1, \omega_2, \cdots, \omega_k\}$ are shorthand notations for the set of modes and their center frequencies, respectively. The symbol t denotes time, $j^2 = -1$ is the square of the imaginary unit, * denotes the convolution operator, and δ is the Dirac delta function. In addition, the expression $\left(\delta(t) + \frac{j}{\pi t}\right) * u_k(t)$ is the Hilbert transform of $u_k(t)$, which can transform $u_k(t)$ into an analytical signal to form a one-sided frequency spectrum with only positive frequencies. In this manner, the spectrum of each mode can be shifted to a baseband signal with the index term $e^{-j\omega_k t}$.

To solve this variational problem, a Lagrangian multiplier (λ) and a quadratic penalty term (α) are introduced to transform the constrained optimization problem (2) into an unconstrained problem. The augmented Lagrangian ℓ is defined as follows:

$$\ell(\{u_k\},\{\omega_k\},\lambda) := \alpha \sum_k \left\| \partial_t \left[\left(\delta(t) + \frac{j}{\pi t}\right) * u_k(t) \right] e^{-j\omega_k t} \right\|_2^2 + \|f(t) - \sum_k u_k(t)\|_2^2 + \langle \lambda(t), f(t) - \sum_k u_k(t) \rangle. \tag{3}$$

For the VMD method, the alternate direction method of multipliers (ADMM) is used to solve Eq. (3). The mode $u_k(\omega)$ in the frequency domain, the center frequencies ω_k and the Lagrangian multiplier λ are iteratively and respectively updated by

$$\hat{u}_k^{n+1}(\omega) = \frac{\hat{f}(\omega) - \sum_{i<k} \hat{u}_i^{n+1}(\omega) - \sum_{i>k} \hat{u}_i^{n}(\omega) + \frac{\hat{\lambda}^n(\omega)}{2}}{1 + 2\alpha(\omega - \omega_k)^2}, \tag{4}$$

$$\hat{\omega}_k^{n+1} = \frac{\int_0^\infty \omega \left|\hat{u}_k^{n+1}(\omega)\right|^2 d\omega}{\int_0^\infty \left|\hat{u}_k^{n+1}(\omega)\right|^2 d\omega}, \tag{5}$$

$$\hat{\lambda}^{n+1}(\omega) = \hat{\lambda}^n(\omega) + \tau\big(\hat{f}(\omega) - \sum_k \hat{u}_k^{n+1}(\omega)\big), \tag{6}$$

https://doi.org/10.5194/hess-2019-565
Preprint. Discussion started: 20 December 2019


Hydrology and Earth System Sciences
Discussions

where n is the iteration counter, τ is the noise tolerance, and $\hat{u}_k^{n+1}(\omega)$, $\hat{f}(\omega)$ and $\hat{\lambda}^n(\omega)$ represent the Fourier transforms of $u_k^{n+1}(t)$ $f(t)$ and $\lambda^n(t)$, respectively. The iterative procedure continues until a convergence condition, namely, $\sum_k \|\hat{u}_k^{n+1} - \hat{u}_k^n\|_2^2 / \|\hat{u}_k^n\|_2^2 < \varepsilon$, is reached. The time-domain mode $u_k(t)$ can be obtained as the real part of the inverse Fourier transform of $u_k(\omega)$ expressed by Eq. (4).

The decomposition level (K), the secondary penalty parameter (α), the noise tolerance (τ) and the convergence tolerance (ε), affect the VMD decomposition performance. A value of K that is too small may lead to poor IMF extraction from the input signal, whereas a too-large value of K may cause information redundancy in the IMFs. A too-small value of α may lead to a large bandwidth, information redundancy and additional noise to be included in the IMFs. A too-large value of α may lead to a very small bandwidth and loss of some signal information. As shown in Eq. (6), the Lagrangian multiplier ensures optimal convergence when an appropriate value of $\tau > 0$ is used and the signal has a low noise level. The Lagrangian multiplier hinders the convergence when $\tau > 0$ is used and the signal has a large noise level. This can be avoided by setting τ to 0. However, it is not possible to reconstruct the input signal precisely if τ equals 0. Additionally, the value of ε affects the reconstruction error of the VMD decomposition.

2.2 Support Vector Regression

Support vector machines were first designed by Vapnik (1995) for handling classification problems and later extended by Vapnik et al. (1997) to deal with regression problems. The essentials of support vector regression (SVR) are described here briefly.

For N pairs of training samples $\{\boldsymbol{x_i}, y_i\}_{i=1}^N$, $\boldsymbol{x_i}$ and y_i indicate the input variables and the desired output targets, respectively. Linear regression is replaced by nonlinear regression, through the use of a nonlinear mapping function ϕ, as follows:

$$y_i \approx f(\boldsymbol{x_i}, \boldsymbol{w}) = \langle \boldsymbol{w}, \phi(\boldsymbol{x_i}) \rangle + b, \quad (7)$$

where $\boldsymbol{w}$ and b represent the regression weights and bias, respectively, and $\langle .,. \rangle$ is the inner product of two vectors (Zeng et al., 2011). SVR evaluates the error between y_i and $f(\boldsymbol{x_i}, \boldsymbol{w})$ by using the following ε-insensitive loss function:

$$|y_i - f(\boldsymbol{x_i}, \boldsymbol{w})|_\varepsilon = \begin{cases} 0, & if\ |y_i - f(\boldsymbol{x_i}, \boldsymbol{w})| < \varepsilon \\ |y_i - f(\boldsymbol{x_i}, \boldsymbol{w})| - \varepsilon, & otherwise \end{cases}. \quad (8)$$

Indeed, the error is ignored if the gap between y_i and $f(\boldsymbol{x_i}, \boldsymbol{w})$ is smaller than ε. Otherwise, the excess of this gap above ε is used as the loss value.

Based on the values of $\boldsymbol{w}$ and b, a regularized risk function R is determined as

$$R = \frac{C}{N} \sum_{i=1}^N |y_i - f(\boldsymbol{x_i}, \boldsymbol{w})|_\varepsilon + \frac{1}{2} \|\boldsymbol{w}\|^2, \quad (9)$$

where the first term indicates the empirical risk based on the ε-insensitive loss function. The second term is a regularization term for penalizing the weight vector in order to limit the complexity of the SVR model. In other words, to minimize the risk function R, the regression weights $\boldsymbol{w}$ will be forced to have small values when a large value of C is used.

https://doi.org/10.5194/hess-2019-565
Preprint. Discussion started: 20 December 2019


Hydrology and Earth System Sciences
Discussions

To avoid handling high-dimensional nonlinear features $\phi(\boldsymbol{x})$, SVR uses a kernel trick that substitutes the inner product $\langle\phi(\boldsymbol{x}), \phi(\boldsymbol{x}')\rangle$ in the optimization algorithm with a kernel function, namely, $K(\boldsymbol{x}, \boldsymbol{x}')$. Some Lagrange multipliers, namely, α_i and β, are introduced to solve the constrained optimization problem of minimizing R. The Lagrange form of the regression function is

$$f(\boldsymbol{x}) = \sum_{i=1}^{N} \alpha_i K(\boldsymbol{x}, \boldsymbol{x}') + \beta. \quad (10)$$

The SVR model relies heavily on the hyperparameters (parameters of the prior distribution) and the kernel function. A radial basis function (RBF), namely, $K(\boldsymbol{x}, \boldsymbol{x}') = exp(-\|\boldsymbol{x} - \boldsymbol{x_i}\|^2/2\sigma^2)$, is used as a covariance function in this work. The parameter σ is used to control the RBF width. For more details on SVR, see Vapnik et al. (1997) and Bishop (2006).

2.3 Bayesian optimization based on Gaussian processes

Bayesian optimization (BO) is a sequential model-based optimization (SMBO) approach typically used for the global optimization of black-box functions, for which the real distribution is unknown or the evaluation is extremely expensive. The motivation of the BO algorithm is putting a prior belief on the objective function, which forms the basis for the loss function in a learning model, and sequentially refining this model by gathering function evaluations along with updating the Bayesian posterior (James et al., 2011; Shahriari et al., 2016).

To update our beliefs about the loss function and calculate the posterior expectation, a prior function is applied. Here, we assume that the real loss function distribution can be described by a Gaussian process (GP). Therefore, the loss function values $\{f(\boldsymbol{x_i})\}_{i=1}^{n}$ for an evaluation set $\{x_i\}_{i=1}^{n}$ satisfy the multivariate Gaussian distribution over the function space

$$f_{1:n} \sim N(m(\boldsymbol{x_{1:n}}), \boldsymbol{K}), \quad (11)$$

where $m(\boldsymbol{x_{1:n}})$ is the GP mean function set and $\boldsymbol{K}$ is the kernel matrix given by the covariance function $K(\boldsymbol{x}, \boldsymbol{x}')$. An acquisition function, used to compute the posterior distribution, is applied to describe the utility of candidate points by maximizing this acquisition function. In other words, the point with the highest utility is selected as the candidate for the next evaluation of f. Many acquisition functions have been practically utilized for Bayesian optimization. These functions include the expected improvement (EI), the upper confidence bounds (UCB), the probability of improvement, the Thompson sampling (TS) and the entropy search (ES). However, the EI function is the most commonly used among these functions (James et al., 2011; Shahriari et al., 2016). The expected improvement can be calculated under the GP model as

$$EI(\boldsymbol{x}) = \begin{cases} [\mu(\boldsymbol{x}) - f(\hat{\boldsymbol{x}})]\Phi(\boldsymbol{z}) + \sigma(\boldsymbol{x})\phi(\boldsymbol{z}) & if\ \sigma(\boldsymbol{x}) > 0 \\ 0 \quad if\ \sigma(\boldsymbol{x}) = 0 \end{cases}, \quad (12)$$

$$\boldsymbol{z} = \frac{\mu(\boldsymbol{x}) - f(\hat{\boldsymbol{x}})}{\sigma(\boldsymbol{x})}, \quad (13)$$

where $f(\hat{\boldsymbol{x}})$ is the current lowest loss value, and $\mu(\boldsymbol{x})$ is the expected loss value, while $\Phi(\boldsymbol{z})$ and $\phi(\boldsymbol{z})$ are the cumulative distribution and the probability density function, respectively. The Bayesian optimization based on Gaussian processes (BOGP) might be wrapped up as follows:

https://doi.org/10.5194/hess-2019-565
Preprint. Discussion started: 20 December 2019

Hydrology and Earth System Sciences
Discussions
Open Access
EGU

Step 1 Given the training samples and hyperparameter searching space as input, define the loss function as mean square error (MSE), and initialize the first candidate $\boldsymbol{x_0}$ by random sampling scheme.

Step 2 Given previous evaluations, update the posterior expectation of f using the GP model.

Step 3 Track a new candidate $\boldsymbol{x_{new}}$ that maximizes the EI function, i.e., $\boldsymbol{x_{new}} = \arg\max EI(x)$.

Step 4 Compute the value of f for the new candidate $\boldsymbol{x_{new}}$.

Step 5 Repeat Steps 1-3 until convergence is achieved or a pre-specified number of iterations is reached.

Step 6 Output the last candidate as the optimized hyperparameters.

2.4 The TSDP framework and the proposed VMD-SVR

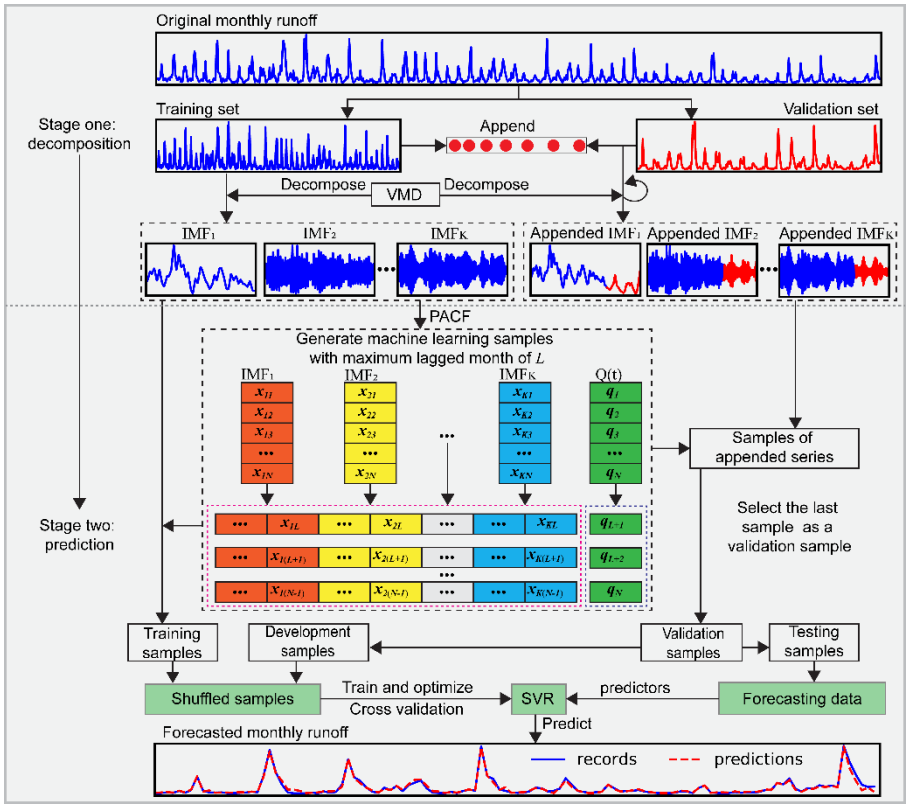


Figure 1: A block diagram of the two-stage VMD-SVR model.

The TSDP framework is designed as follows. *(i) Streamflow decomposition:* the entire streamflow is first divided into training and validation sets, and then the training set is concurrently decomposed while the validation set is sequentially decomposed into several signal components. *(ii) Streamflow prediction:* the optimal lag times (measured in hours, days, months or years) of the decomposed signals components are initially combined as predictors. The original streamflow is selected as the desired predicted target so as to generate training and validation samples. Next, the validation samples are divided into development

https://doi.org/10.5194/hess-2019-565
Preprint. Discussion started: 20 December 2019


Hydrology and Earth System Sciences
Discussions

samples for optimized model selection and testing samples for testing the optimized models. A single learning model is finally trained and tuned using a set of samples obtained by mixing and shuffling the training and development samples. The TSDP streamflow decomposition stage extracts hidden information of the original streamflow data and avoids using validation information that is not available in practical forecasting applications. The TSDP prediction stage builds one learning model only using a set of mixed and shuffled samples to reduce boundary effects, save modelling time, avoid error accumulation, and improve prediction performance.

After each key framework module is discussed separately, the design details of the TSDP framework and its VMD-SVR realization are summarized as follows (See Fig. 1). Steps 1-4 represent the decomposition stage while Steps 5-8 represent the prediction stage. Note that the VMD and SVR schemes can be respectively replaced by other decomposition-based and data-driven prediction models.

Step 1 Collect original streamflow data $Q(t)$ as input of VMD-SVR ($t = 1,2, \cdots N$, N is the length of original streamflow data).

Step 2 Divide the original streamflow data into training and validation sets with 70% and 30% of the data samples, respectively.

Step 3 Concurrently extract K IMF signal components from the training set using the VMD scheme. The optimal K is determined by observing whether the last extracted IMF component exhibits central-frequency aliasing.

Step 4 Sequentially append the validation data to the training set to generate appended sets. Decompose each appended set into K signal components using the VMD scheme.

Step 5 Plot the partial autocorrelation function (PACF) of each signal component for the training set to select the optimal lag period for generating SVR training samples. We assume that the predicted target of the k^{th} signal component is $x_k(t+1)$. If the PACF of the mth lag period lies outside the 95% confidence interval ($\left[-\frac{1.96}{\sqrt{n}}, \frac{1.96}{\sqrt{n}}\right]$, n is the signal component length) and is larger than 0.5, then $x_k(t), x_k(t-1), \cdots, x_k(t+1-m)$ are selected as input predictors for the k^{th} signal component.

Step 6 Combine the input predictors of each signal component as the SVR predictors, and select the original streamflow after the maximum lag period ($Q(t+1)$) as the predicted target.

Step 7 Based on the input variables and output targets obtained in Step 6, generate training samples using signal components, and generate appended samples using the appended signal components created in Step 4. Select the last sample of the appended samples as a validation sample. Divide the validation samples evenly into development and testing samples.

Step 8 Mix and shuffle the training and development samples. Train and optimize the SVR model using the shuffled samples and the BOGP algorithm. For testing, predict the original streamflow by inputting the test sample predictors into the optimized SVR model. The output of VMD-SVR is the testing prediction data.

https://doi.org/10.5194/hess-2019-565
Preprint. Discussion started: 20 December 2019


Hydrology and Earth System Sciences
Discussions

2.5 Comparative experimental setups

We design five comparative experiments to evaluate the effectiveness, efficiency and reliability of the TSDP framework and its VMD-SVR realization in terms of the boundary effect reduction, decomposition performance, predictability, time consumption, overfitting and forecasting capabilities for long leading times. Experimental comparisons are made against benchmark models including the decomposition-based models of EEMD, SSA, DWT and SVR, as well as the no-decomposition baseline SVR model. All these models are optimized by the BOGP.

2.5.1 Experiment 1: Evaluation of the boundary effect reduction

Predicting streamflow without using validation information is quite challenging due to the fact that the validation samples are seriously affected by the boundary effects. The boundary effects lead the deviation of the decomposition results from the real values. This occurs mainly at the end of a time series because the future values used as parameters for decomposing the boundary values of the time series are not observed. We give a detailed example of the boundary effects in Sec. 4.1. As stated in Sec. 2.4, the training set is concurrently decomposed to save time and the validation set is sequentially decomposed to avoid using future validation information. Therefore, only a few decomposition values of the training set are affected by the boundary effects. However, all decomposition values of the validation set are affected by the boundary effect because the last decomposition value of the appended series is selected as a validation decomposition value. Therefore, the boundary effect causes the smooth training samples and the oscillating validation samples to have different distributions. This reduces the generalization performance of the training-data-based optimized models on the validation samples.

To reduce the boundary effects, we design a mixing-and-shuffling step, in which we further divide the validation samples into development and testing samples, training and development shuffled mixed samples. The SVR models are finally optimized using these shuffled samples based on a cross-validation strategy. Note that the mixing-and-shuffling step does not reduce the boundary effects themselves but reduces their ramifications, results and consequences. To demonstrate that this step can reduce the ramifications of the boundary effects, we design a comparative experiment in which we divide the entire streamflow data into a training-and-development set and a testing set with 85% and 15% of the entire streamflow, respectively. We concurrently decomposed the training-and-development set and sequentially decomposed the testing set, and finally generated samples for each set. The training-and-development samples were divided into training and development samples. The remaining procedures are identical to those of the TSDP framework. Note that the numbers of the training, development, and testing samples in this experiment are identical to those in the mixing-and-shuffling step of the TSDP framework. Also, the training and development samples have the same distribution.

2.5.2 Experiment 2: Performance evaluation of the TSDP models

For the second experiment, we test the VMD decomposition performance by comparing the prediction performance of VMD-SVR scheme with those of three other TSDP models which combine EEMD (Wu and Huang, (2009), SSA (Golyandina et al.,

https://doi.org/10.5194/hess-2019-565
Preprint. Discussion started: 20 December 2019


Hydrology and Earth System Sciences
Discussions

(2001) and DWT (Sachindra et al., (2019) with SVR, namely EEMD-SVR, SSA-SVR and DWT-SVR. The modeling processes for the EEMD-SVR, SSA-SVR and DWT-SVR schemes are identical to that of the VMD-SVR scheme except for differences in the decomposition stage.

The EEMD technique is developed based on the EMD one as follows: (1) add white noise with an amplitude of ε to the original series; (2) decompose the noise-contaminated series into IMF components using EMD; (3) repeat the aforementioned two steps with M different levels of white noise; (4) obtain the final results by calculating the ensemble means of the aforementioned IMFs. Therefore, the white noise amplitude (ε) and the number of ensemble members (M) are used to evaluate the standard deviation of error, i.e. the difference between the original time-series and the IMFs, and hence these parameters affect the EEMD decomposition performance. The value of ε should not be too small because a too-small amplitude may not produce the extrema change necessary for EMD (Wu and Huang, 2009). However, increasing M can reduce the white noise effect to a negligible low level (Wu and Huang, 2009). In general, good decomposition results are obtained when M is set to a few hundreds and less than 1% of the error may be caused by the remaining noise (Wu and Huang, 2009). In this work, we set M and ε to 100 and 0.2, respectively, as suggested by Wu and Huang (2009).

Singular spectrum analysis (SSA) decomposes a given time-series into independent trend, periodic, or noise series components. The SSA method has four main steps, namely embedding, singular value decomposition (SVD), grouping and diagonal averaging. The embedding step maps the original series into a trajectory matrix of dimensions W_L by $N - W_L + 1$ with a window length of W_L ($2 \le W_L \le N$), where N is the length of the original series. In the SVD step, the trajectory matrix is decomposed into W_L elementary matrices. For the grouping step, the W_L elementary matrices are divided into m groups (where m is the decomposition level, and $m \le W_L$). Finally, for the diagonal averaging step, m subseries are extracted from the grouped matrices. Clearly, the values of W_L and m affect the SSA decomposition performance. In general, a value of W_L of up to 30-45% of N can adequately separate the underlying periodicities from the overall trend. However, if one of the subseries is periodic, W_L can be set to the period of this subseries to get better decomposition performance (Zhang et al., 2015). The value of W_L is defined as 12 in this work because we perform monthly runoff forecasting. The variance contribution of each subseries to the original series can decide the value of m. However, one can ignore the grouping step if the value of W_L is small (e.g., $W_L \le 20$) because grouping may sometimes cause information hiding in the grouped subseries. Therefore, we do not need to set m, and the grouping step is ignored in this work. In summary, 12 subseries are decomposed herein by SSA.

The discrete wavelet transform (DWT) decomposes a given time-series into several subseries using a discrete mother wavelet function (ψ). This transform decomposes the original series into several detail components which represent the frequencies, and one approximation component which represents the trends. For the first decomposition level, the original series, $Q(t)$, is decomposed into one detail high-frequency component $D_1(t)$, and one approximation low-frequency component $A_1(t)$, such that $Q(t) = D_1(t) + A_1(t)$. For the second decomposition level, the $A_1(t)$ component is further decomposed into one detail component $D_2(t)$ and one approximation component $A_2(t)$, such that $A_1(t) = D_2(t) + A_2(t)$. In such a manner, this decomposition process can be continued until a predefined decomposition level (L) is reached. The final decomposition

https://doi.org/10.5194/hess-2019-565
Preprint. Discussion started: 20 December 2019


Hydrology and Earth System Sciences
Discussions

outcome can be expressed as $Q(t) = D_1(t) + \cdots + D_L(t) + A_L(t)$. Therefore, ψ and L are two vital parameters that affect the DWT decomposition performance. However, there is no theoretical method to select these parameters. Testing numerous combinations of ψ and L is quite laborious. In this work, we set ψ to the Daubechies wavelet function with 10 vanishing moments (db 10) as suggested by Seo et al. (2015). Also, we set L to $int[\log(N)]$ following Nourani et al. (2009).

2.5.3 Experiment 3: Performance evaluation of TSDPE models

In the third experiment, we compare the prediction performance and the computational cost of the TSDP and TSDPE models based on the EEMD, SSA, VMD, DWT and SVR methods. In particular, we investigate four combined schemes, namely EEMD-SVR-A ('A' means the ensemble approach is *addition ensemble*), SSA-SVR-A, VMD-SVR-A and DWT-SVR-A. The TSDPE models include the extra ensemble stage compared to the TSDP models. The decomposition stages of the TSDP and TSDPE models are identical. In the TSDPE prediction stage, the PACF is also used for selecting the predictors and the predicted target for each signal component. However, one optimized SVR model will be trained for each signal component and used to predict this component in the testing phase. The remaining prediction procedures are identical to those of the TSDP models. For testing in the ensemble stage, the prediction results of each signal component are summed to predict the streamflow data. Since the TSDPE models build one SVR model for each signal component, the time cost of any of the TSDPE models is expected to be significantly higher than those of the TSDP models.

2.5.4 Experiment 4: Evaluation of PCA-based dimensionality reduction

Our fourth experiment tests whether dimensionality reduction (i.e. reduction of the number of predictors) improves the prediction performance of the TSDP models of Experiment 2. The TSDP models can save modelling time and sometimes may improve the prediction performance compared with the TSDPE models. However, combining the predictors of each signal component as the input predictors of the TSDP models may lead to an overfitting problem because the predictors increase a lot compared with those of the TSDPE models and predictors might be correlated. Therefore, it is necessary to test whether the reduction of the predictors of the two-stage decomposition models can help these models improve the prediction performance. Principal component analysis (PCA) is used to reduce the input predictors of the TSDP models. This analysis uses an orthogonal transformation to transform the correlated predictors into a set of linearly uncorrelated predictors called the principal components. For further details on PCA, see Jolliffe (2002). The number of principal components, which indicates the number of predictors retained by the PCA procedure, is the main PCA parameter. To find the optimal number of these predictors, numerous values might be tested. We can also guess the number of predictors using the method of Minka (2001). Since the number of predictors varies for different TSDP models, the (guessed) number of predictors is replaced by the (guessed) number of reduced predictors for convenience of comparison. In this paper, the number of reduced predictors ranges from 0 to 15. A zero number of reduced predictors indicates that all the predictors are retained (i.e. dimensionality is not reduced), but the correlated predictors are transformed into uncorrelated ones.

https://doi.org/10.5194/hess-2019-565
Preprint. Discussion started: 20 December 2019

Hydrology and Earth System Sciences
Discussions
Open Access
EGU

2.5.5 Experiment 5: Evaluation of runoff forecasting for long leading times

For the fifth and last experiment, we test the prediction performance of the TSDP models for long leading times. Four levels of leading times are used, namely 3, 5, 7, and 9 months. Except for the sample generation process, the modelling processes of the TSDP models with long leading times (of more than one month) are identical to those of the TSDP models with a 1-month leading time. The predictors obtained in Section 4.2 cannot be used as the predictors for 3-, 5- , 7- and 9-month leading times because these inputs may not be mainly related to the streamflow of these leading times. In this work, we also induce a lag of 20 months and calculate the Pearson correlation coefficient between the monthly runoff of each long leading time and the 20-month lag. A month lag was selected as a predictor under the condition that the absolute value of the Pearson correlation coefficient was larger than a specific threshold. This threshold was selected based on whether or not the number of predictors approximately equals the number of predictors for the one-month-ahead streamflow forecasting.

3 Case study

3.1 Study area and data observations

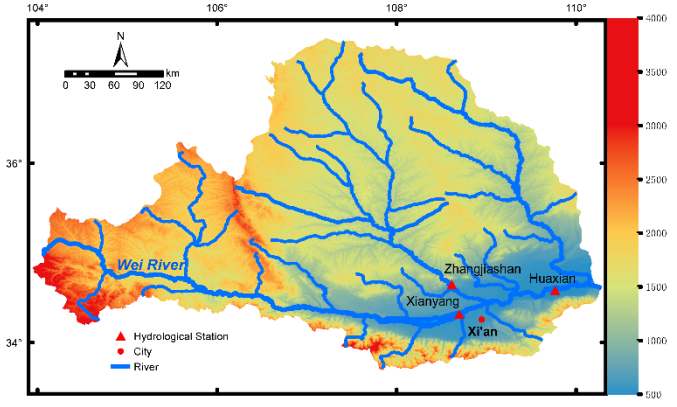


Figure 2: A geographical overview of the Wei River basin.

In this work, the Wei River basin (see Fig. 2) was selected as the location for a case study. The Wei River, the largest tributary of the Yellow River in China, lies between 33.68ºN-37.39ºN and 103.94ºE-110.03ºE and has a drainage area of 135,000 km^2 (Jiang et al., 2019). The Wei River has a total length of 818 km and originates from the Niaoshu Mountains in Gansu province and flows east into the Yellow River (Gai et al., 2019). The associated catchment has a continental monsoon climate with an annual average precipitation of more than 550 mm. The precipitation of the flood season from June to September accounts for 60% of the annual total (Jiang et al., 2019). In the Guanzhong Plain, the Wei River serves as a source of water for agricultural, industrial and domestic purposes (Yu et al., 2016). Therefore, robust monthly runoff prediction in this region plays a vital role in water resource allocation.

https://doi.org/10.5194/hess-2019-565
Preprint. Discussion started: 20 December 2019


Hydrology and Earth System Sciences
Discussions

The historical monthly runoff records from 1/1953 to 12/2018 (792 records) at the Huaxian, Xianyang and Zhangjiashan stations were used to evaluate the proposed model and the other state-of-the-art models. The locations of these hydrological stations are shown in Fig. 2. The records were collected from the hydrological information data center of Shaanxi Hydrographic and the Water Resources Survey Bureau. The monthly runoff records were computed from the instantaneous value (m^3/s) observed at 8 A.M. each day. The entire monthly runoff was divided into a training set covering the period 1953/01-1998/12 which represents approximately 70% of the entire monthly runoff, while the validation set corresponds to the remaining period 1999/01-2018/12. The validation set was further divided evenly into a development set (covering the period 1999/01-2008/12) for selecting the optimal forecasting model and a testing set (covering the period 2009/01-2018/12) for validating the optimal model.

3.2 Data normalization

To make the BOGP algorithm converge faster, all the input variables in this work were normalized to the same scale ranging from -1 to 1 by the following equation:

$$\boldsymbol{y} = 2 \otimes \frac{\boldsymbol{x}-x_{min}}{x_{max}-x_{min}}, \tag{14}$$

where the mathematical multiplication and subtraction are element-wise operations, while $\boldsymbol{x}$ and $\boldsymbol{y}$ are the raw and normalized vectors, respectively, and x_{max} and x_{min} are the maximum and minimum values of $\boldsymbol{x}$, respectively. The reason we used Eq. (14) to normalize the data samples is that the decomposition results contain negative values but no outliers and Eq. (14) will make sure the predictors and prediction target are equally normalized to [-1,1]. Note that x_{max} and x_{min} of the training samples are also used to normalize the development and testing samples to avoid using future information from the development and testing phases when training SVR models, and enforce all samples to follow the same distribution.

3.3 Model evaluation criteria

The Nash–Sutcliffe efficiency (NSE) (Nash and Sutcliffe, 1970), the normalized root-mean-square error (NRMSE), the peak percentage of threshold statistics (PPTS) (Bai et al., 2016; Stojković et al., 2017) and the time cost were employed to evaluate the performance of decomposition-based models and the pure SVR models. The NSE, NRMSE, and PPTS criteria are defined as follows:

$$NSE = 1 - \frac{\sum_{t=1}^{N}\left(x(t)-\hat{x}(t)\right)^2}{\sum_{t=1}^{N}\left(x(t)-\bar{x}(t)\right)^2}, \tag{15}$$

$$NRMSE = \frac{\sqrt{\sum_{t=1}^{N}\left(x(t)-\hat{x}(t)\right)^2/N}}{\sum_{t=1}^{N}x(t)/N}, \tag{16}$$

$$PPTS(\gamma) = \frac{100}{\gamma}\frac{1}{N}\sum_{t=1}^{G}\left|\frac{x(t)-\hat{x}(t)}{x(t)}\right|, \tag{17}$$

where N is the number of samples, $x(t)$, $\bar{x}(t)$ and $\hat{x}(t)$ are the raw, average, and predicted data samples, respectively. The NRMSE facilitates the comparison between datasets or models at different scales. The lower NRMSE values indicate less

https://doi.org/10.5194/hess-2019-565
Preprint. Discussion started: 20 December 2019

Hydrology and Earth System Sciences
Discussions

residual variance. The raw data samples are arranged in a descending order and the predicted data samples are arranged following the same order to calculate the PPTS criterion. The symbol γ denotes a threshold level that controls the percentage of the data selected from the beginning of the arranged data sequence and G is the number of values above this threshold level. For example, PPTS(5) means the top 5% flows, or the peak flows, which are evaluated by the PPTS criterion.

3.4 Open-source software and hardware environments

This work heavily relies on open-source software. We use Pandas (McKinney, 2010) and Numpy (van der Walt et al., 2011) to perform data preprocessing and management, Scikit-Learn (Pedregosa et al., 2011) to establish SVR models for forecasting monthly runoff and perform PCA for dimensionality reduction, Scikit-Optimize (Tim et al., 2018) to tune the SVR models and Matplotlib (Hunter, 2007) to draw the figures. All the models were developed and the time cost of each model was computed based on a 2.50-GHz Intel Core i7-4710MQ CPU with 16.0 GB of RAM.

3.5 Modelling processes

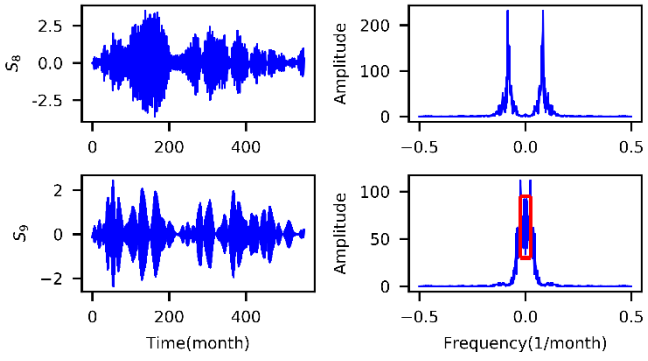
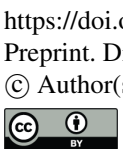


Figure 3: Center-frequency aliasing for the last signal component of the Huaxian station.

The VMD-SVR model is employed as an example to illustrate the modelling processes of the TSDP and TSDPE models. The VMD-SVR decomposition stage obtains multiscale signal components from the original monthly runoff data. As stated in Sec. 2.1, the decomposition level (K), the secondary penalty parameter (α), the noise tolerance (τ) and the convergence tolerance (ε) are the four parameters that influence the VMD decomposition performance, which is particularly very sensitive to K. In this work, the values of α, τ, and ε were set to 2000, 0, and 1e-9, respectively, based on experimental validation. In addition, we proposed an approach of setting the optimal K by observing whether the last IMF had central-frequency aliasing as represented by the red rectangle area in Fig. 3. In other words, we vary K from $K = 2$ with a step size of 1, and if the center-frequency aliasing of the last IMF is first observed when $K = L$, then the optimal K is set to $L - 1$. As shown in Fig. 3, the

https://doi.org/10.5194/hess-2019-565
Preprint. Discussion started: 20 December 2019



Hydrology and Earth System Sciences
Discussions

optimal decomposition level for the Huaxian station is 8. Additionally, we used VMD to extract 8 and 7 signal components from the training sets at the Xianyang and Zhangjiashan stations, respectively.

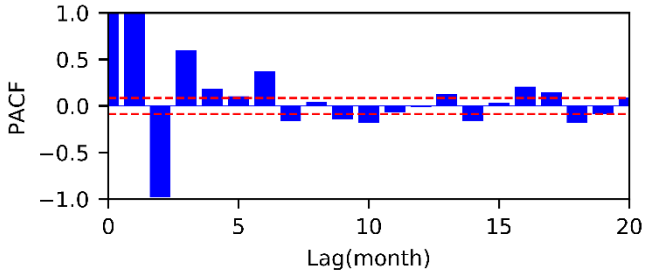


Figure 4: PACF of the first VMD signal component at the Huaxian station.

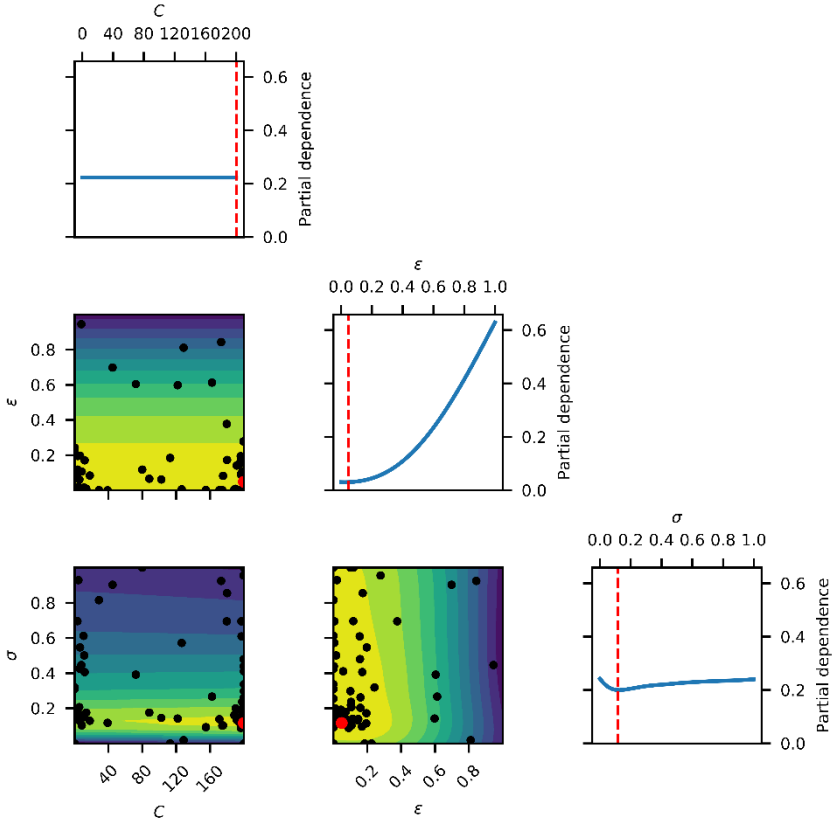


Figure 5: Pairwise partial dependence plot of the objective function (MSE) for the VMD-SVR scheme at the Huaxian station.

https://doi.org/10.5194/hess-2019-565
Preprint. Discussion started: 20 December 2019

Hydrology and Earth System Sciences
Discussions

In the prediction stage, we first determined the predictors and the predicted target to generate the training samples. According to the procedure of Sec. 2.4, PACF is used to determine the optimal predictors for the VMD-SVR scheme. The first VMD IMF component, i.e., S_1, at the Huaxian station is used as an example of tracking the optimal input predictors from the PACF of the S_1. As shown in Fig. 4, the PACF of the third lag month exceeds the boundary of the 95% confidence interval illustrated by the red dashed line, and is larger than 0.5. Moreover, there is no lag month with a PACF larger than 0.5 when the lag time continues to increase. Thus $x_1(t)$, $x_1(t-1)$ and $x_1(t-2)$ are selected as the optimal input predictors for S_1. In such a manner, all the input predictors of each signal component are combined together as the VMD-SVR predictors. Then, the original monthly runoff, i.e., $Q(t+1)$, is selected as the predicted target. Based on this process, the training, development, and testing samples are generated for each of the Huaxian, Xianyang, and Zhangjiashan stations.

We then fine-tuned an SVR model using BOGP. As described in Sec. 2.2, the SVR hyperparameters, i.e. the weight penalty (C), the error tolerance (ε), and the width control coefficient (σ) should be tuned to optimize the VMD-SVR model. To tune these hyperparameters, the maximum number of BOGP iterations was set to 100 and the mean square error (MSE) was selected as the objective function to obtain the SVR optimal parameter set (C, ε and σ). As well, a 6-fold cross-validation scheme was used for tuning these hyperparameters. We ran the BOGP procedure ten times to reduce the impact of random sampling, and the parameters with the lowest MSE in the development phase were selected. As illustrated in Fig. 5, the pairwise partial dependence of the hyperparameters at the Huaxian station shows that the tuned parameters ($C = 200$, $\varepsilon = 0.0486$ and $\sigma = 0.1166$) are globally optimized. This analysis indicates that the BOGP procedure provides reasonable results.

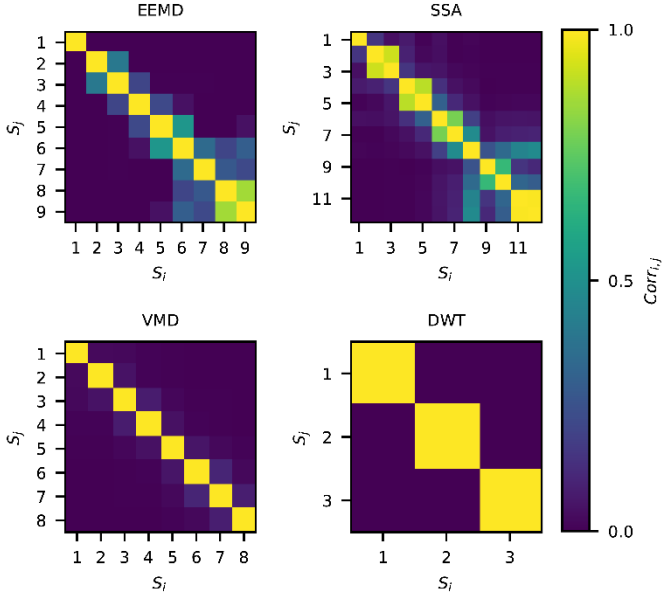


Figure 6: Pearson correlation coefficients of signal components obtained by different decomposition methods. S_i and S_j represent the *i*th and *j*th signal components, respectively.

https://doi.org/10.5194/hess-2019-565
Preprint. Discussion started: 20 December 2019

Hydrology and Earth System Sciences
Discussions

The modelling processes of the DWT-SVR, SSA-SVR and EEMD-SVR schemes are identical to that of VMD-SVR, except for the decomposition stage. According to the parameters settings of Sec. 2.5.2, nine IMFs were extracted from the original monthly runoff data for all the three stations using EEMD, while 12 signal components including several trend, periodic and noise terms were obtained by SSA for all the three stations, and 3 signals components including 2 detail components and an approximation component were extracted by DWT for all the three stations. The Pearson correlation coefficients between the components obtained by these four decomposition methods are illustrated in Fig. 6. The coefficients between the following pairs at the Huaxian station are larger than 0.4: S_2 (the second signal component) and S_3, S_5 and S_6, S_8 and S_9 of EEMD, S_2 and S_3, S_4 and S_5, S_6 and S_7, S_7 and S_8, S_9 and S_{10}, S_8 and S_{12}, S_8 and S_{11}, S_{11} and S_{12} of SSA. This indicates that the signal components of EEMD and SSA have redundant information. The coefficients between the VMD signal components of VMD are less than 0.1, which indicates that these components are barely correlated. The same conclusion can be made for the DWT components whose correlation coefficients are below 0.001. Similar results were obtained at the Xianyang and Zhangjiashan station.

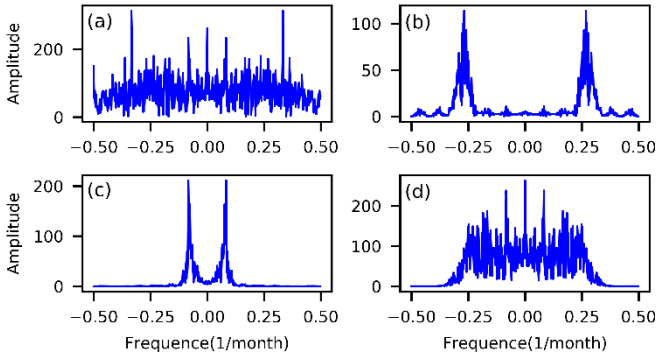


Figure 7: Frequency spectrum for (a) S_1 of EEMD, (b) S_6 of SSA, (c) S_8 of VMD and S_1 of DWT at the Huaxian station.

The modelling processes of the TSDPE models are identical to those of the TSDP models, except for establishing one optimized SVR model for each signal component. Take the first VMD component, S_1, at the Huaxian station as an example: $x_1(t)$, $x_1(t-1)$ and $x_1(t-2)$ are selected as input predictors, while $x_1(t+1)$ is selected as the predicted target. The TSDPE models use the input predictors and the predicted target of each signal component as the inputs and output, respectively. The final TSDPE prediction results are obtained by summing the predictions of each signal component.

The frequency spectrum of the difficultly-predicted signal component for each decomposition algorithm is shown in Fig. 7. Fig. 7(a) shows that the EEMD S_1 component has a large noise level along the entire frequency domain. Fig. 7(b) shows that the SSA S_6 component also has a low noise level along the entire frequency domain. Fig. 7(c) shows that the VMD S_8 component has a low noise level in the low-frequency domain. Fig. 7(d) shows that the DWT S_1 component has a large noise level in the low-frequency domain. These results indicate that the VMD scheme is much more robust to noise than the EEMD, SSA and DWT schemes.

https://doi.org/10.5194/hess-2019-565
Preprint. Discussion started: 20 December 2019

Hydrology and Earth System Sciences
Discussions

485 **4 Experimental Results and Analysis**

4.1 Reduction of the boundary effects in TSDP models

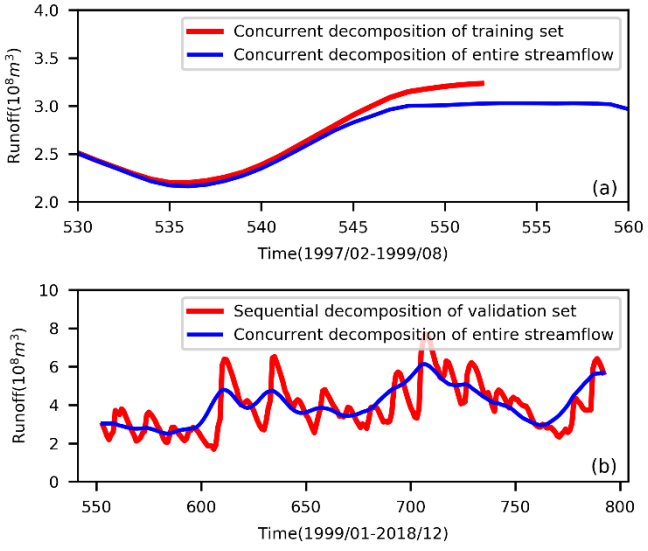
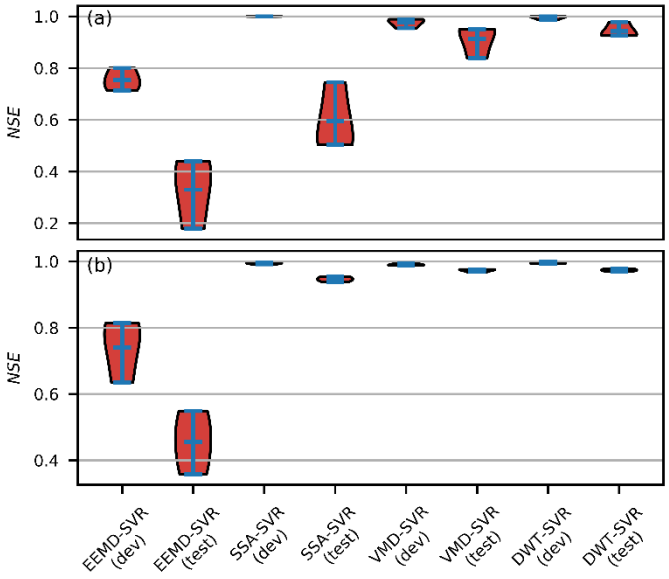


Figure 8: Boundary effect of the VMD S_1 for (a) the training set and (b) the validation set at the Huaxian station.



490

Figure 9: Violin plots of the NSE criterion for TSDP models. (a) The development and testing samples come from different distributions (without mixing and shuffling). (b) The development and testing samples come from the same distribution (with the mixing-and-shuffling step).

https://doi.org/10.5194/hess-2019-565
Preprint. Discussion started: 20 December 2019


Hydrology and Earth System Sciences
Discussions

The boundary effect of the VMD S_1 component at the Huaxian station is illustrated in Fig. 8. Both of the training and the entire sets were concurrently decomposed. The decomposition results for the first component (S_1) of the entire set represent the real decomposition values. Figure 8(a) shows that a deviation occurs at the end of the training period for only about 20 decomposition values. This indicates that the training samples are barely affected by the boundaries. Figure 8(b) shows that the decomposition results of the validation set exhibit time oscillations demonstrating that the validation samples (including the development and testing samples) have different distributions from the training samples. Similar results were obtained for the other signal components obtained by VMD, EEMD, SSA, and DWT at the Huaxian, Xianyang, and Zhangjiashan stations. Comparative analysis results of the TSDP models with and without the mixing-and-shuffling step are illustrated in Fig. 9. Figure 9(a) shows that the NSE of every optimized TSDP model without mixing and shuffling has a larger interquartile range and a smaller mean than the development-phase NSE. This indicates that the optimized model approximates the training distribution reasonably well, though it poorly generalizes to the testing distribution. Figure 9(b) shows that compared with the testing-phase NSE of the TSDP models without the mixing-and-shuffling step, the NSE interquartile range with this step decreased substantially, while the NSE mean increased considerably. This demonstrates the importance of the mixing-and-shuffling step in improving the prediction performance on the test samples, and reducing the consequences of the boundary effects.

4.2 Performance Evaluation for the TSDP models

Figure 10 shows the prediction results during the testing phase for the monthly runoff collected from 2009/01 to 2018/12 for the Huaxian, Xianyang, and Zhangjiashan stations. Quantitative evaluation results are presented in Fig. 11. As shown in Fig. 10 (a1, b1 and c1), the prediction values of the VMD-SVR and DWT-SVR schemes are consistent with the records from the three stations. While the SSA-SVR model generally captures random variations, achieving obvious success in fitting both of the trend and the periodicity components, it suffers from poor tracking for the peak runoff. The EEMD-SVR and SVR models barely capture random variations, and show very poor performance in fitting trend and periodicity components. Fig. 10 (a2, b2 and c2) show that the correlation values of the VMD-SVR and DWT-SVR models are concentrated around the ideal fit, with a small angle between the ideal and linear fits. This indicates that the raw measurements and the predictions of these two models have a high degree of agreement. The correlation values of the EEMD-SVR and SVR models are less concentrated than those of the VMD-SVR, DWT-SVR and SSA-SVR models. However, the correlation values of the EEMD-SVR model are more concentrated than those of the SVR one. This indicates that the decomposition of the original monthly runoff data can sometimes help improve the prediction performance. Overall, these findings indicate that the VMD-SVR and DWT-SVR models possess good prediction performance and satisfactory generalization capabilities for different stations. The SSA-SVR model also enjoys good prediction performance and satisfactory generalization capability but it suffers from slightly poor prediction performance for the peak runoff.

https://doi.org/10.5194/hess-2019-565
Preprint. Discussion started: 20 December 2019

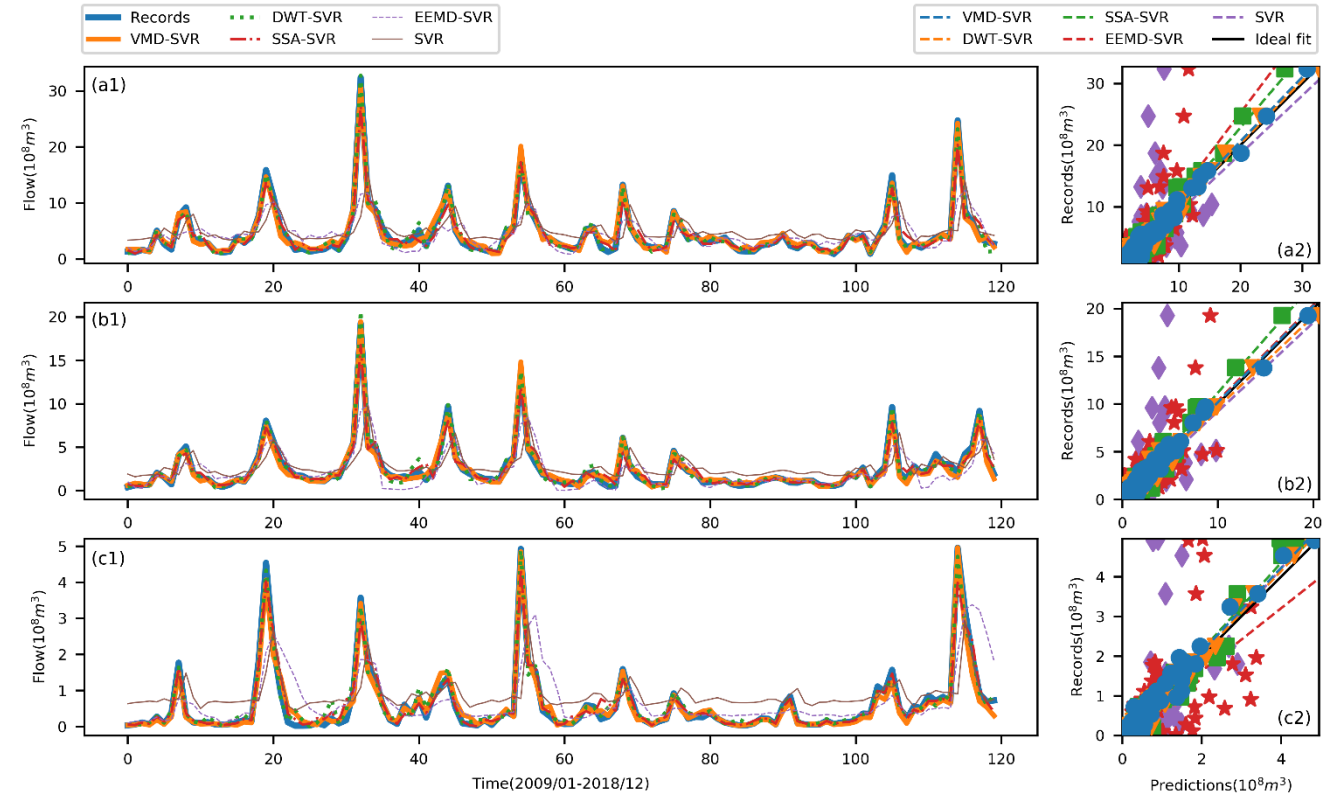


Figure 10: Prediction results of the TSDP framework during the testing phase:(a1 and a2) Huaxian, (b1 and b2) Xianyang and (c1 and c2) Zhangjiashan stations.

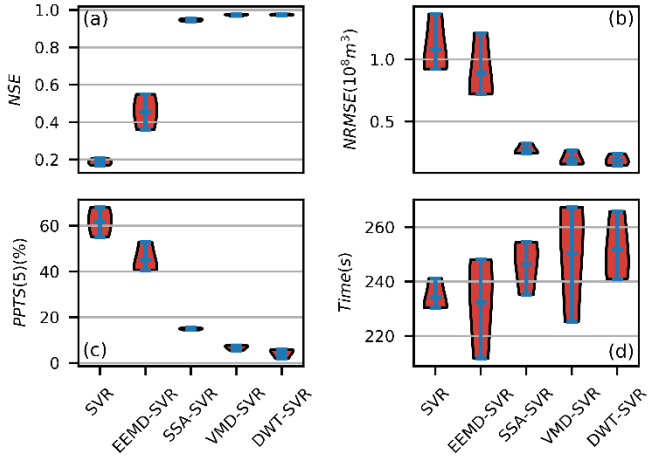


Figure 11: Violin plots of the evaluation criteria for the TSDP models.

https://doi.org/10.5194/hess-2019-565
Preprint. Discussion started: 20 December 2019


Hydrology and Earth System Sciences
Discussions

Figure 11 shows that the VMD-SVR, DWT-SVR, SSA-SVR and EEMD-SVR models perform better than the SVR model in all aspects except for the time cost. Figure 11 also indicates that exploiting the decomposition methods including VMD, DWT, SSA and EEMD improves the prediction accuracy. Fig. 11(a) shows that the mean NSE for the EEMD-SVR, VMD-SVR, DWT-SVR and SSA-SVR model increased by 145%, 408%, 423% and 423% compared to the SVR scheme. This indicates that the VMD-SVR scheme has prediction performance similar to that of the DWT-SVR one, better prediction performance than that of the SSA-SVR one and much better prediction performance than those of the EEMD-SVR and SVR schemes. The mean NSE scores of the SSA-SVR, VMD-SVR and DWT-SVR schemes are larger than 0.94. This illustrates that the predictions of these TSDP models closely match the actual measurements. The mean NSE scores of the EEMD-SVR and SVR schemes are less than 0.8, which is often taken as a threshold value for reasonably well-performing models (Newman et al., 2015). This result indicates that the measurements are not reasonably matched by the predictions of the EEMD-SVR and SVR schemes. Figure 11(b) shows that compared with the SVR model, the mean NRMSE of the EEMD-SVR, SSA-SVR, VMD-SVR and DWT-SVR schemes are reduced by 18%, 75%, 82%, and 82%, respectively. This figure also indicates that the VMD-SVR model has similar prediction performance to the DWT-SVR model and has better performance than that of the SSA-SVR model and much better prediction performance than those of the EEMD-SVR and SVR schemes. Figure 11(c) shows that the values of the mean PPTS(5) of the EEMD-SVR, SSA-SVR, VMD-SVR and DWT-SVR schemes are reduced by 27%, 76%, 89% and 93%, respectively, compared to the SVR scheme. This indicates that the DWT-SVR model has the best prediction performance for the peak runoff and that the VMD-SVR model has a peak streamflow prediction capability, which is slightly worse than that of the DWT-SVR model. However, the VMD-SVR model has a PPTS(5) interquartile range that is smaller than that of DWT-SVR. Thus, the VMD-SVR model has a better generalization ability than that of the DWT-SVR one. Additionally, the interquartile ranges of the NSE, NRMSE and PPTS(5) criteria for the SSA-SVR, VMD-SVR and DWT-SVR schemes are much less than those of the EEMD-SVR and SVR schemes. This indicates that the SSA-SVR, VMD-SVR and DWT-SVR schemes have much better generalization performance. These analyses indicate that the prediction performance of the proposed VMD-SVR model is similar to that of the DWT-SVR model, better than that of the SSA-SVR model, and much better than those of the EEMD-SVR and SVR schemes for the overall streamflow. In addition, Fig. 11(d) shows that the TSDP models and the SVR one have similar time costs ranging from 210 s to 270 s. This indicates that the TSDP models do not obviously increase the modelling time cost.

4.3 Performance gap between the TSDP and TSDPE models

The predictions, records and scatter plots of the TSDPE models are presented in Fig. 12, while the corresponding quantitative evaluation results are illustrated in Fig. 13. Comparisons of prediction scatters for the TSDP and TSDPE models are displayed in Fig. 14.

https://doi.org/10.5194/hess-2019-565
Preprint. Discussion started: 20 December 2019

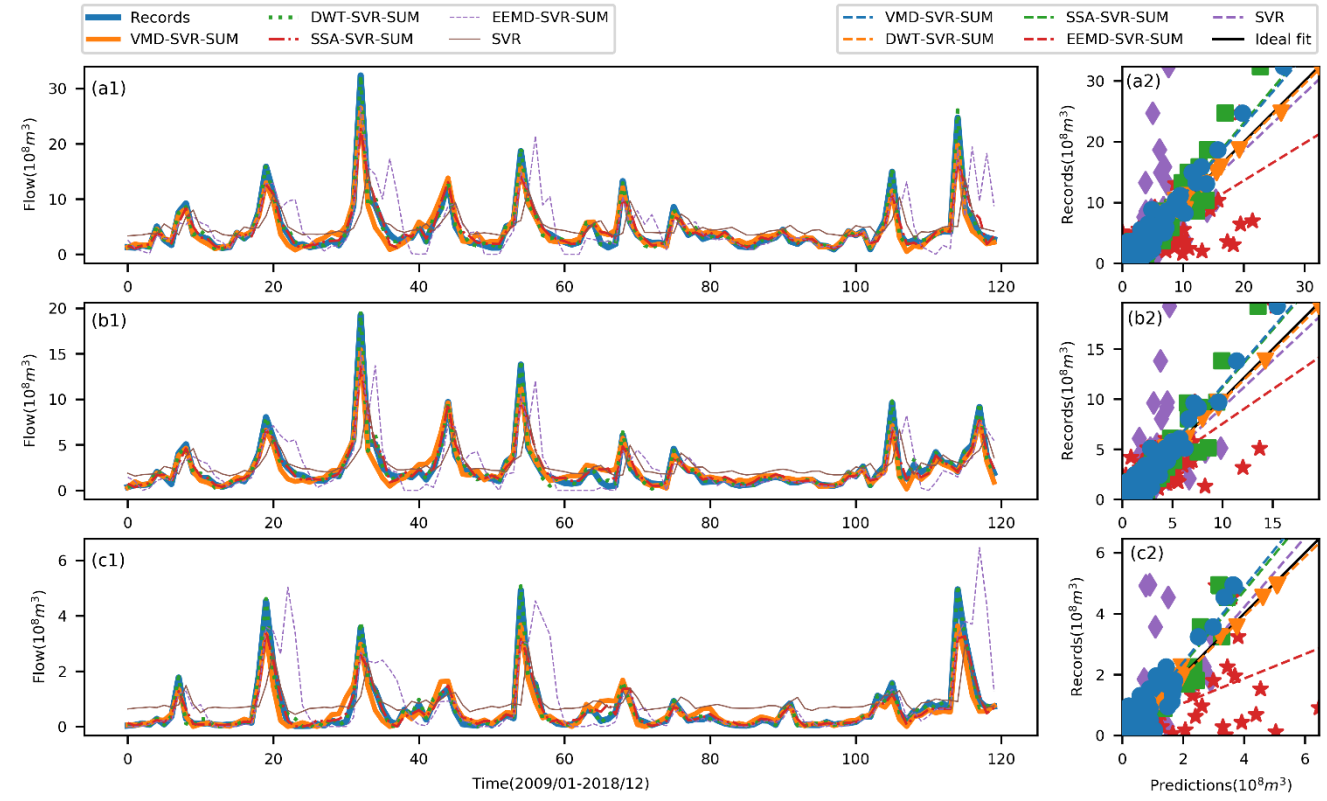


Figure 12: Prediction results of the TSDPE framework during the testing phase:(a1 and a2) Huaxian, (b1and b2) Xianyang and (c1 and c2) Zhangjiashan stations.

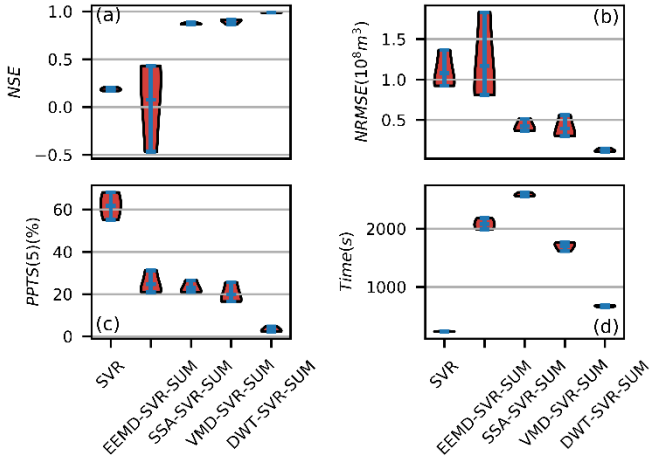


Figure 13: Violin plots of the evaluation criteria for the TSDPE models.

https://doi.org/10.5194/hess-2019-565
Preprint. Discussion started: 20 December 2019


Hydrology and Earth System Sciences
Discussions

As shown in Fig. 12 (a1, b1 and c1), the DWT-SVR-A scheme captures the random variations, with a good capability for fitting both of the trend and periodicity components for all three stations. While the VMD-SVR-A and SSA-SVR-A models generally capture random variations, with a certain capability for fitting both the trend and periodicity components, these models suffer from slightly poor tracking for the peak runoff. However, the predictions of the EEMD-SVR-A and SVR schemes cannot follow variations in the test records. Moreover, the scatter plots in Fig. 12 (a2, b2 and c2) show that the correlation values of the DWT-SVR-A, VMD-SVR-A and SSA-SVR-A schemes have high agreement, and are concentrated near the ideal fit. As well, there is a small angle between the linear and ideal fit curves for the DWT-SVR-A scheme. Also, the correlation values of the VMD-SVR-A and SSA-SVR-A models are concentrated around the linear fit. However, these two models have a relatively large angle between the ideal and linear fit curves. This observation supports the conclusion that the prediction performance of these two models for the peak runoff is slightly poor. The correlation values of the EEMD-SVR-A and SVR schemes are scattered around the ideal fit. While most of the EEMD-SVR-A scatters are under the ideal fit, most of the SVR scatters are above the ideal fit. This behaviour reflects data overestimation by the EEMD-SVR-A method and data underestimation by the SVR one. Overall, these findings indicate that the DWT-SVR-A scheme shows good forecasting and generalization performance for the monthly runoff at different stations. The VMD-SVR-A and SSA-SVR-A models show similar good performance but suffer from poor prediction performance for the peak runoff.

Fig. 13 (a) shows that, the mean NSE scores of the EEMD-SVR-A, SSA-SVR-A, VMD-SVR-A and DWT-SVR-A schemes increased by -58%, 369%, 380%, and 431%, respectively, compared to the SVR one. This indicates that the DWT-SVR-A scheme has better prediction performance than the VMD-SVR-A and SSA-SVR-A ones, and much better prediction performance than the SVR and EEMD-SVR-A schemes. Additionally, the decomposition methods used in the TSDPE models do not always improve the prediction performance. The mean NSE values of the VMD-SVR-A, DWT-SVR-A and SSA-SVR-A schemes are larger than 0.8. This shows that the predictions of these models can match the actual measurements reasonably well. The mean NSE values of the EEMD-SVR-A and SVR models are less than 0.5, illustrating that these two models have worse prediction performance. In addition, the EEMD-SVR-A model has an NSE value less than 0, indicating that the recorded mean is sometimes better than using this model. Fig. 13 (b) shows that the mean NRMSE values of the EEMD-SVR-A, SSA-SVR-A, VMD-SVR-A and DWT-SVR-A models are reduced by -8%, 61%, 63% and 88%, respectively, compared to the SVR model. This indicates that the VMD-SVR-A and SSA-SVR-A models are similar in prediction performance, the EEMD-SVR-A model has the worst performance, while the DWT-SVR-A model has the best prediction performance among these models. This result also demonstrates that the decomposition of the raw streamflow data do not always improve the prediction performance. Fig. 13 (c) shows that the mean PPTS(5) scores of the EEMD-SVR-A, SSA-SVR-A, VMD-SVR-A and DWT-SVR-A models are reduced by 60%, 63%, 68% and 95%, respectively, compared to the SVR model. These results indicate that the DWT-SVR-A model has the best prediction performance for the peak runoff while the VMD-SVR-A model has a better prediction performance for the peak runoff in comparison to the SSA-SVR-A and EEMD-SVR-A models. Overall, these analyses indicate that the DWT-SVR-A scheme has the best prediction performance. Also, the prediction performance of the VMD-SVR-A scheme is similar to that of the SSA-SVR-A scheme, and is much better than those of the EEMD-SVR-A and

https://doi.org/10.5194/hess-2019-565
Preprint. Discussion started: 20 December 2019

Hydrology and Earth System Sciences
Discussions

SVR schemes. In addition, Fig. 13(d) shows that the mean time costs of the EEMD-SVR-A, SSA-SVR-A, VMD-SVR-A and DWT-SVR-A schemes are increased by 790%, 1005%, 627% and 187%, respectively, compared to the SVR scheme. This clarifies that these TSDPE models obviously increase the modelling time cost and that the increase in this cost depends on the decomposition level of the data preprocessing techniques.

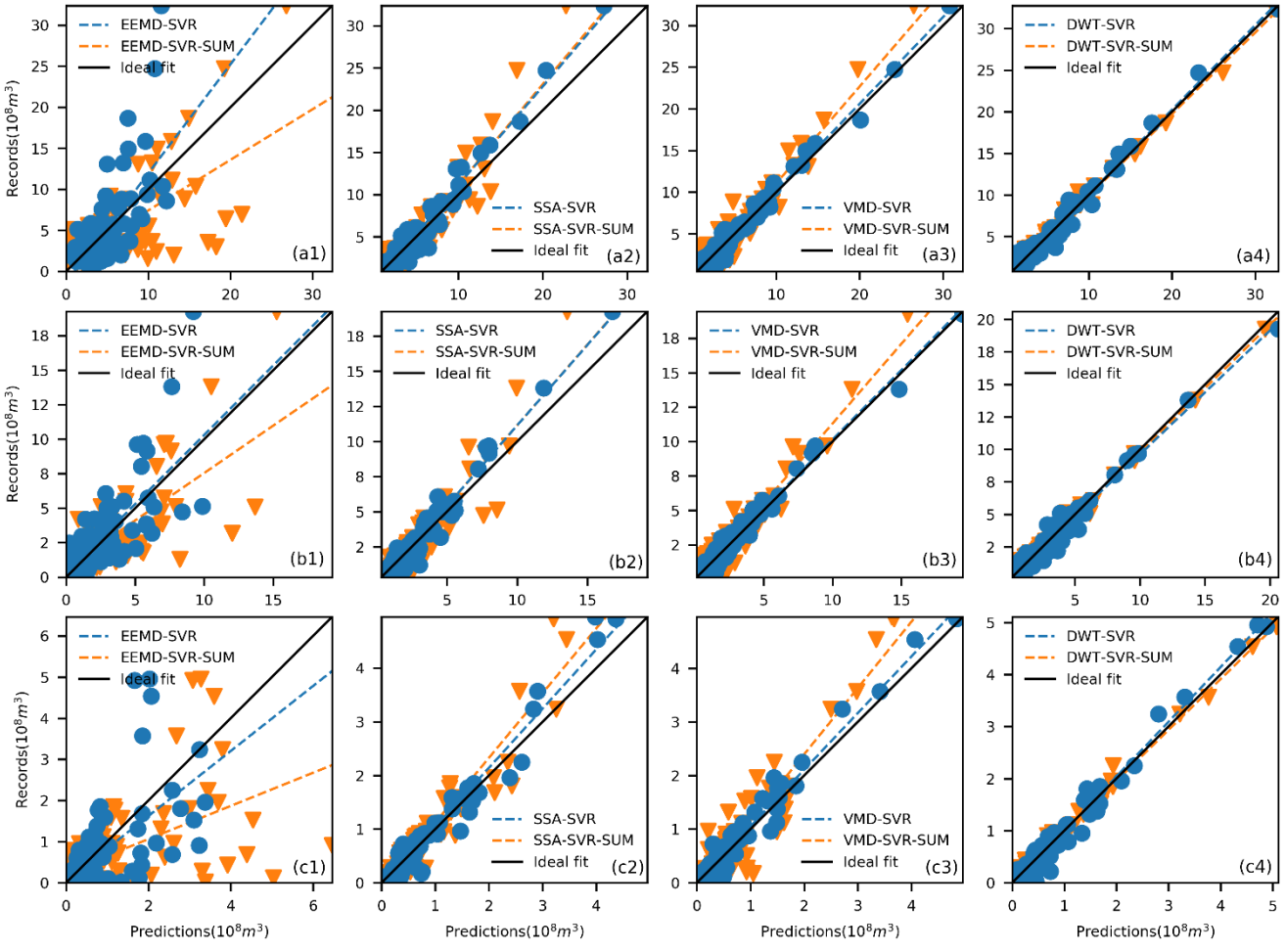


Figure 14: Comparison of the prediction scatters for (a1-a4) Huaxian, (b1-b4) Xianyang and (c1-c4) Zhangjiashan stations.

Figure 14 shows that the correlation values of the VMD-SVR, SSA-SVR and EEMD-SVR schemes are much more concentrated than those of the VMD-SVR-A, SSA-SVR-A and EEMD-SVR-A schemes. The angles between the ideal and linear fit curves of the VMD-SVR, SSA-SVR and EEMD-SVR schemes are smaller than those of the VMD-SVR-A, SSA-SVR-A and EEMD-SVR-A schemes. However, the correlation values of the DWT-SVR scheme are similar to those of the DWT-SVR-A one and the linear fit curves of the DWT-SVR-A scheme at the Xianyang and Zhangjiashan stations are closer to the ideal fit than those of the DWT-SVR scheme. Overall, these analyses indicate that the TSDP models can significantly

https://doi.org/10.5194/hess-2019-565
Preprint. Discussion started: 20 December 2019

Hydrology and Earth System Sciences
Discussions

reduce the modelling time and improve the prediction performance (especially for the peak runoff), except only for the DWT-SVR model.

4.4 Effect of dimensionality reduction on the TSDP models

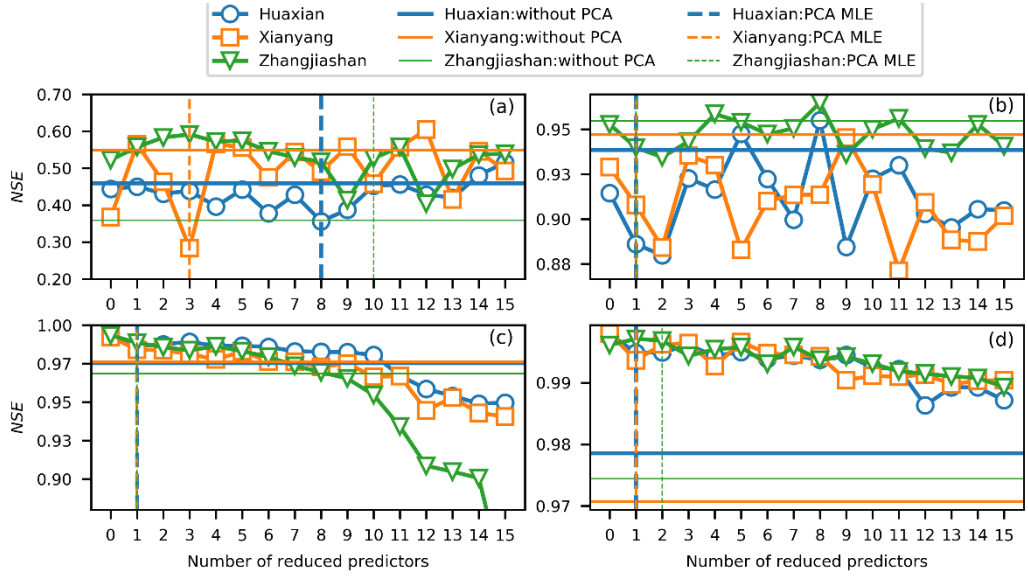


Figure 15: NSE values of different reduced components for the (a) EEMD-SVR, (b) SSA-SVR, (c)VMD-SVR and (d) DWT-SVR schemes.

The NSE values for different numbers of reduced predictors and for all three stations are illustrated in Fig. 15. The number of reduced predictors was set based on Minka's MLE (Minka, 2001) method for estimating the optimal reduced dimension. The results are also presented in Fig. 15. As shown in Fig. 15, dimensionality reduction can significantly improve the NSE score of the TSDP models except for the EEMD-SVR and SSA-SVR models. This indicates that it is difficult to confirm without dimensionality reduction whether the EEMD-SVR and SSA-SVR predictors contain redundant information. The NSE score of the VMD-SVR scheme with dimensionality reduction (the reduced predictors are less than the guessed reduced predictors obtained by MLE method) is larger than that of the same scheme without dimensionality reduction for all three stations. The same conclusion applies to the DWT-SVR scheme. This demonstrates that some of the VMD-SVR and DWT-SVR predictors obtained as discussed in Sec. 4.1 are linearly correlated and that making these predictors uncorrelated improves the prediction performance. However, the NSE scores of these two schemes for the three stations generally decrease as the number of reduced predictors is increased from 0 to 15. This indicates that eliminating predictors leads to some loss of information. In addition, the NSE scores associated with the guessed numbers of reduced predictors of the two schemes are larger than those of the same schemes without PCA. Again, this indicates that the guessed dimensionality reduction also improves the prediction performance. Compared with the NSE score of the VMD-SVR scheme without PCA, the maximal NSE scores at the Huaxian,

https://doi.org/10.5194/hess-2019-565
Preprint. Discussion started: 20 December 2019

Hydrology and Earth System Sciences
Discussions

Xianyang, and Zhangjiashan stations are increased by 1.7%, 1.7% and 2.6%, respectively. Compared with the NSE score of the DWT-SVR scheme without PCA, the maximal NSE scores at the same stations are increased by 2.0%, 2.8% and 2.3%, respectively. The NSE scores of the EEMD-SVR and SSA-SVR schemes have no obvious increase or decrease when the reduced dimension varies from 0 to 15. Overall, these analyses indicate that the VMD-SVR and DWT-SVR predictors contain a small but significant information redundancy.

4.5 Runoff forecasting of long leading times for the TSDP models

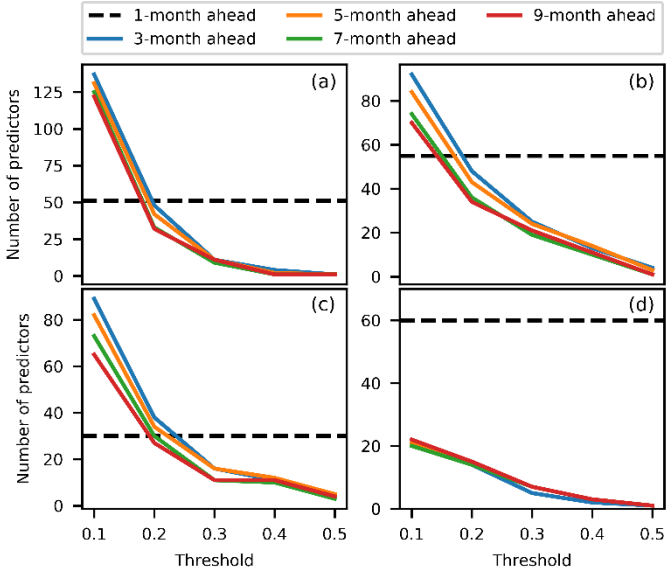


Figure 16: Number of predictors versus threshold at the Huaxian station for the (a) EEMD, (b) SSA, (c) VMD and (d) DWT methods.

We tested five thresholds ranging from 0.1 to 0.5 with a step of 0.1 to select the predictors for long leading times exceeding one month. Figure 16 shows that the number of predictors is reduced as the threshold increases. We can see that the predictor numbers for leading times exceeding one month for the EEMD, SSA and VMD methods are close to that of a 1-month leading time when the threshold is set to 0.2. Therefore, the threshold level was set to 0.2 for convenience of comparison.

The TSDP models for long leading times of 3-, 5-, 7- and 9-month leading times were established to further evaluate the prediction performance of these models. Three performance evaluation criteria, namely NSE, NRMSE and PPTS(5), were used to assess the TSDP model performance for long leading times. The scatter plots and evaluation results for leading times of 1, 3, 5, 7 and 9 months are illustrated in Fig. 17 and Fig. 18, respectively.

Figure 17 shows that as the leading time increases, the correlation values of the TSDP models become gradually more scattered around the ideal fit, and hence the forecasting performance decreases gradually. In addition, the correlation values of the VMD-SVR scheme with leading times exceeding one month are much more concentrated than those of other TSDP models. This indicates that the VMD-SVR scheme has the best prediction performance. The correlation values of the SSA-SVR and DWT-

https://doi.org/10.5194/hess-2019-565
Preprint. Discussion started: 20 December 2019

Hydrology and Earth System Sciences
Discussions

SVR models for 3-, 5- , 7- and 9-month runoff forecasting are much more scattered around the ideal fit compared to the correlation values of the SSA-SVR and DWT-SVR models for 1-month runoff forecasting. This indicates that these models have the worst long-leading-time forecasting performance.

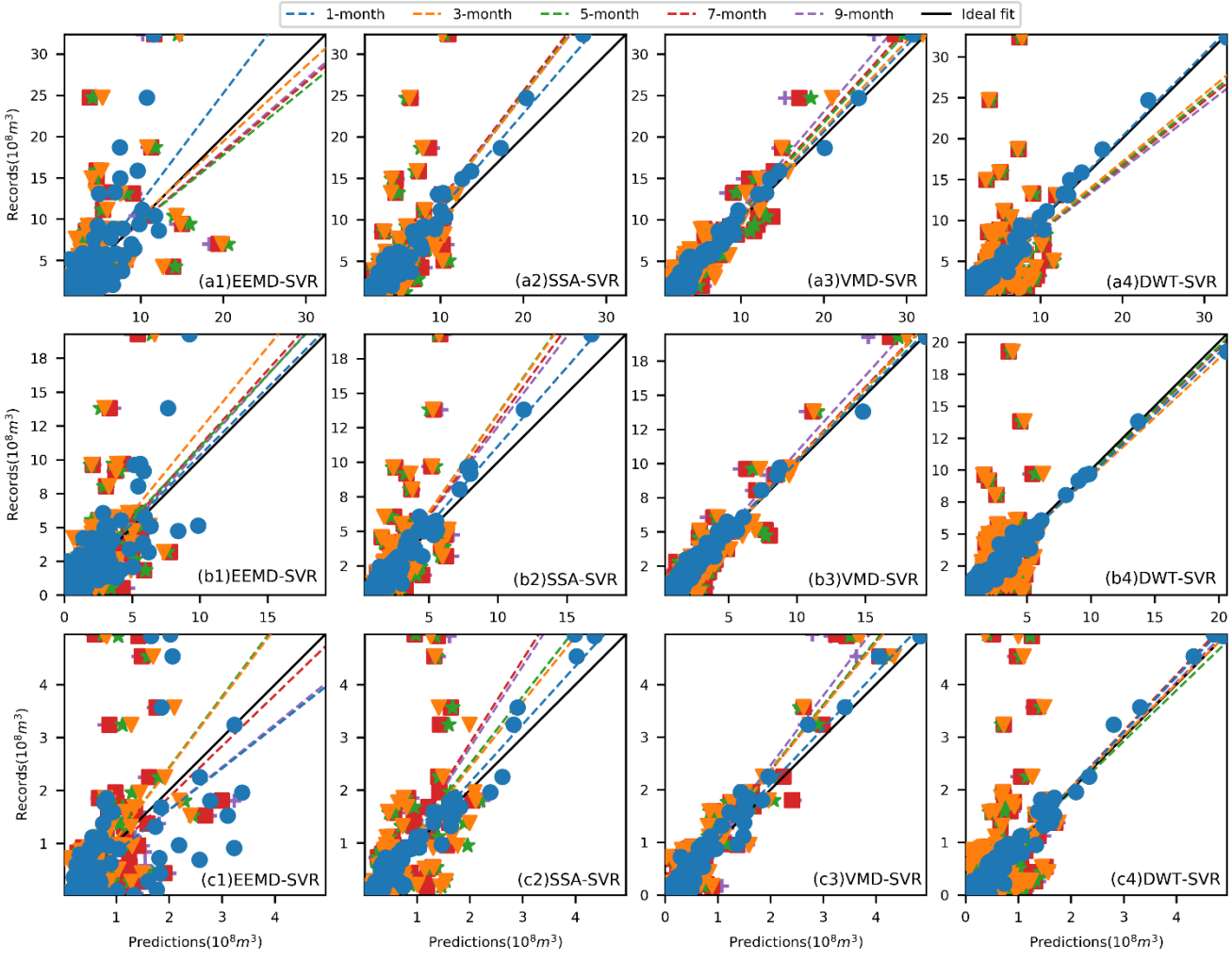


Figure 17: Scatter plots of EEMD (a1,b1 and c1), SSA (a2, b2 and c2), VMD-SVR (a3, b3 and c3) and DWT-SVR (a4, b4 and c4) for different leading times at the Huaxian station (a1-a4), the Xianyang station (b1-b4) and the Zhangjiashan station (c1-c4).

Figure 18 shows that as the leading time increases, the mean NSE gradually decreases, while the mean NRMSE and the mean PPTS(5) gradually increase for a variety of models except for the SSA-SVR one. The results show that the prediction performance for each of the VMD-SVR, DWT-SVR and EEMD-SVR schemes decreases gradually. Compared with the evaluation results of the SSA-SVR scheme for one-month-ahead runoff forecasting, the mean NSE of the SSA-SVR scheme for 3-, 5- ,7- and 9-month-ahead runoff forecasting decreases substantially. However, the mean NSE of the SSA-SVR scheme

https://doi.org/10.5194/hess-2019-565
Preprint. Discussion started: 20 December 2019

Hydrology and Earth System Sciences
Discussions

slightly increases as the leading time increases from 5 to 9 months, indicating that the prediction performance of this scheme for longer leading times is generally worse but not monotonic. The reason is that most SSA signal components are periodic and there might be periodic similarities between the selected predictors and the monthly runoff for long leading times. Similar conclusions can be reached based on the NRMSE and PPTS(5) criteria whose increase and decrease patterns are opposite to those of the NSE criterion.

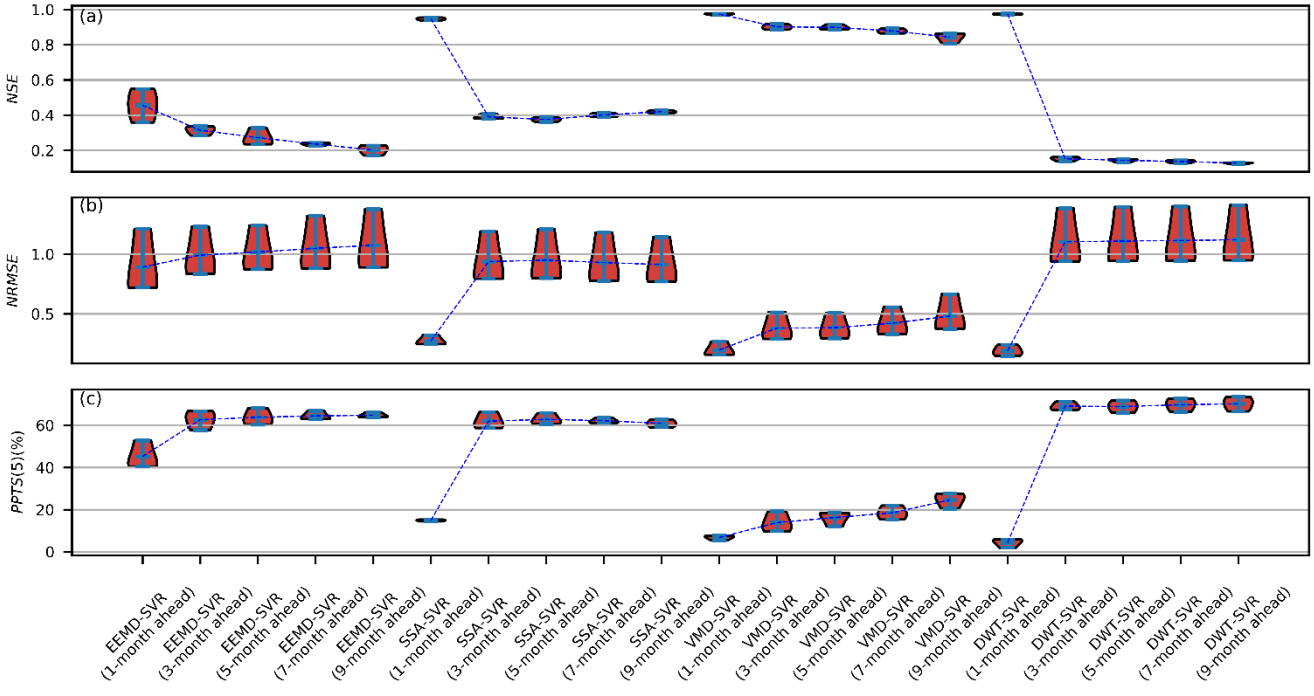


Figure 18: Violin plots of the evaluation criteria with different leading times for the TSDP models.

Figure 18(a) shows that, compared to the VMD-SVR scheme for 1-month ahead runoff forecasting, the mean NSE values of the VMD-SVR, SSA-SVR, EEMD-SVR and DWT-SVR schemes are reduced, respectively, by 7%, 60%, 88% and 84% for a 3-month leading time. For a 5-month leading time, the respective reduction is by 8%, 61%, 72% and 85%. The respective reduction is by 10%, 59%, 76% and 86% for a 7-month leading time. Finally, the corresponding reduction for a 9-month leading time reaches 13%, 57%, 79% and 87%. This analysis indicates that the VMD-SVR scheme has the best prediction performance for leading times exceeding one month. Indeed, the performance is ranked from the highest to the lowest as: VMD-SVR>SSA-SVR>EEMD-SVR>DWT-SVR. The same results could be observed from Fig. 18(b) and Fig. 18(c). Additionally, the VMD-SVR scheme generally has a smaller interquartile range, and also has a good generalization capability for different watersheds. Overall, these findings indicate that the VMD-SVR scheme has the best performance, while the DWT-SVR one has the worst prediction performance for leading times exceeding one month.

https://doi.org/10.5194/hess-2019-565
Preprint. Discussion started: 20 December 2019


Hydrology and Earth System Sciences
Discussions

5 Discussion

As we can see from the experimental results discussed in Sec. 4, the designed TSDP framework and its VMD-SVR realization do achieve the aforementioned desirable features (See Sec. 1). We now discuss why and how the TSDP framework and the VMD-SVR realization are superior to other decomposition-based streamflow forecasting frameworks and models.

The results in Sec. 4.1 indicate that mixing and shuffling the training and development samples in the TSDP framework improve the prediction performance on the testing samples (See Fig. 9). The training samples and the validation samples (which contain development and testing samples) in the designed TSDP framework have quite different distributions. This distribution disparity is caused by the boundary effects (Fig. 8). Mixing and shuffling the training and development samples help the SVR models to be trained using a cross-validation strategy, and enable the assessment of the validation distribution under boundary effects in the training phase without using testing information. Therefore, the mixing-and-shuffling step helps the TSDP framework enhance its generalization capability. However, one may sequentially append the training set to the first streamflow data and decompose the appended sets to force the training and validation samples to follow approximately the same distribution. We refrain from doing this (and strongly advise against it) because the modelling process will become quite laborious and the distribution may not be approximately the same as when one sequentially decomposes the training set.

The results of Sec. 4.2 indicate that the VMD-SVR and DWT-SVR schemes are superior to the SSA-SVR and EEMD-SVR ones. Although the VMD-SVR scheme has a slightly worse peak runoff forecasting capability than the DWT-SVR one, the former possesses a better generalization capability for different watersheds (See Fig. 11). Reasons justifying and leading to these results are: (1) the EEMD and SSA signal components have redundant information (Fig. 6), and therefore chaotically represent the trend, period and noise terms; (2) the EEMD scheme has limitations such as sensitivity to noise and sampling (Dragomiretskiy and Zosso, 2014), which result in larger noise levels for the first EEMD signal component and thus lower predictability (Fig. 7); (3) the number of lag months larger than 0.5 of the DWT method is much more than that of the VMD one (Fig. 16) and the DWT signal components are more orthometric than those of the VMD method (Fig. 6). This sometimes leads to better DWT-SVR prediction performance. However, the VMD method is more robust to sampling and noise. We can manually control the VMD modes to avoid the mode mixing problem (Fig. 3). We can restrict the noise to be included in the signal components by setting the noise tolerance to 0, and therefore easily find the VMD signal components. However, this step does not enable the summation of the VMD signal components to reconstruct the original streamflow. However, the TSDP framework perfectly solves this problem by building a single SVR model to predict the original streamflow instead of summing the predictions of each signal component.

The results in Sec. 4.3 indicate that the TSDP framework saves modelling time and improves prediction performance compared to the TSDPE framework. This is because the TSDP models avoid the error accumulation problem and simulate not only the predictor-runoff relationship but also the decomposition-runoff relationship. Additionally, the DWT-SVR scheme does not improve the performance considerably compared with the DWT-SVR-A scheme due to the fact that only three signal components, which can precisely reconstruct the original monthly runoff, are extracted from the original series such that each

https://doi.org/10.5194/hess-2019-565
Preprint. Discussion started: 20 December 2019


Hydrology and Earth System Sciences
Discussions

signal component is predicted very well. Therefore, the performance gap between the TSDP and TSDPE frameworks is quite small for DWT-based monthly runoff forecasting models. However, the mother wavelet function and the empirical equation of the decomposition level given in Sec. 2.5.2 might be not suitable for daily or annual streamflow forecasting. Therefore, using DWT in the TSDP and TSDPE frameworks needs a substantial prior experience with the mother wavelet function and the decomposition level.

The results in Sec. 4.4 indicate that combining the input predictors of the individual signal components as the ultimate predictors overfits the original streamflow slightly for the VMD-SVR and DWT-SVR schemes (Fig. 15). The reason is that some VMD-SVR and DWT-SVR predictors are correlated. However, since the prediction performance of the VMD-SVR and DWT-SVR schemes for one-month-ahead forecasting is already good enough and dimensionality reduction improves the prediction performance a little bit, we suggest predicting the original streamflow directly based on the VMD-SVR scheme using the combined predictors without dimensionality reduction. However, we suggest employing dimensionality reduction analysis when trying different decomposition algorithms and data-driven models on different streamflow series due to the unknown effect of overfitting.

The results in Sec. 4.5 indicate that the VMD-SVR scheme has the best performance while the DWT-SVR scheme has the worst prediction performance for leading times exceeding one month. The DWT-SVR model performs the worst for 3-, 5- , 7- and 9-month runoff forecasting because the selected predictors with a threshold of 0.2 have a weak correlation with the predicted runoff (Fig. 16). Therefore, we deem that DWT is not suitable for the TSDP framework. The reasons given above in this section also lead to better VMD-SVR prediction performance and worse prediction performance of both of the SSA-SVR and EEMD-SVR schemes for leading times exceeding one month.

In summary, we would like to highlight three major contributions of the TSDP framework and its VMD-SVR realization:

(i) The TSDP framework (or its VMD-SVR variant) first divides the entire monthly runoff into training and validation sets, then concurrently decomposes the training set and sequentially decomposes the validation set, in such a way that the validation information (which is not available in practical hydrological applications) is not included in the training samples. Therefore, the TSDP framework and its VMD-SVR model are practical for streamflow forecasting.

(ii) The TSDP framework (or the VMD-SVR model) only establishes one optimized SVR-based training model instead of an optimized model for each signal component to predict the original streamflow and simultaneously infer the predictor-runoff relationship and the decomposition-runoff relationship. Therefore, the TSDP framework, as well as the VMD-SVR model, are efficient.

(iii) The TSDP framework (or its VMD-SVR variant) mixes and shuffles the training and development samples to reduce the ramifications, results, or consequences of the boundary effects. This is a key step in the TSDP and TSDPE frameworks for improving the prediction performance.

The boundary effect is a barrier for practical streamflow forecasting using decomposition-based models. Previous researchers handled the boundary effects by improving the decomposition algorithms (Quilty and Adamowski, 2018) or fixing the

https://doi.org/10.5194/hess-2019-565
Preprint. Discussion started: 20 December 2019


Hydrology and Earth System Sciences
Discussions

deviation of the boundary decompositions (Zhang et al., 2015). However, as far as we know, approaches of building a forecasting framework that is adapted to the boundary effect never be tried. In this work, we focus on living with rather not eliminating the boundary effects to develop an effective, efficient, and reliable decomposition-based forecasting framework. Note that we do not need a lot of prior experience of signal processing algorithms or mathematical extension methods for correcting boundary deviation. We just make the models assess the validation distribution during the training phase. Hence, this operational streamflow forecasting framework is quite simple and easy to implement.

6 Conclusions

This work investigated the potential of the proposed TSDP framework and its VMD-SVR realization for simulating runoff in basins lacking meteorological observations. The TSDP framework (or its VMD-SVR model) simultaneously simulates the predictor-runoff relationship and the decomposition-runoff relationship to save modelling time and improve prediction performance. With five experiments, we explored the effectiveness, efficiency and reliability of the TSDP framework and its VMD-SVR variant in terms of reducing the boundary effects, decomposition success, prediction performance, time consumption, overfitting, and forecasting capabilities for long leading times. We demonstrated that the TSDP framework (or the VMD-SVR scheme) is able to simulate monthly runoff with competitive performance compared to reference models. With the first experiment, we evaluated the reduction of the boundary effect in the TSDP framework. In the second experiment, we evaluated the TSDP model performance. For the third experiment, we assessed the performance gap between the TSDP and TSDPE models. Our fourth experiment empirically tested the overfitting of the TSDP models. Additionally, we evaluated the prediction performance of the TSDP models for long leading times in the fifth and last experiment.

The goal of this research was to establish a practical, effective, efficient and reliable decomposition-based framework (or model) without focusing on optimizing the system parameters. Therefore, it is likely that better prediction performance can be obtained by exhaustive optimization steps of the BOGP method. However, with 100 iterations of optimization, we were able to obtain reasonably good prediction performance that showed superiority over reference models.

In summary, the major conclusions of this work are the following:

a. The TSDP framework can significantly reduce the boundary effects and improve the prediction efficiency and performance compared to the TSDPE framework (except for the DWT-based TSDP models).
b. The VMD-SVR scheme can simulate and generalize the one-month-ahead runoff very well, and it saves modelling time and improves the prediction performance compared to the VMD-SVR-A scheme.
c. The VMD-SVR scheme overfits the monthly runoff a little bit. Nevertheless, we do believe that the effect of this slight overfitting can be ignored.
d. The VMD-SVR scheme is able to simulate the 3-, 5-, 7- and 9-month-ahead runoff reasonably well.

The boundary effect is a potential barrier for practical streamflow forecasting. We do believe that mixing and shuffling the training and development samples is a promising way to reduce the influences of the boundary effect and improve the VMD-

https://doi.org/10.5194/hess-2019-565
Preprint. Discussion started: 20 December 2019


Hydrology and Earth System Sciences Discussions

SVR prediction performance on the testing samples (i.e. the future monthly runoff). Ultimately, however, the black-box nature of the TSDP framework and the VMD-SVR model (or any data-driven model) is a justifiable barrier of making decisions of water resource management using the prediction results. Further research is needed on the VMD-SVR result interpretability based on statistical analysis.

Author Contributions. Ganggang Zuo, Yani Lian and Ni Wang designed all the experiments. Yani Lian and Xinxin He collected and preprocessed the data. Ganggang Zuo and Yani Lian conducted all the experiments and analyzed the results. Ganggang Zuo wrote the first draft of the manuscript with contributions from Yani Lian and Xinxin He. Ni Wang and Jungang Luo supervised the study and edited the manuscript.

Conflicts of Interest. Declarations of interest: none.

Acknowledgments. This work was supported by the National Natural Science Foundation of China (Grant Nos. 51679186, 51679188, 51979221 and 51709222), the National Key R&D Program of China (Grant No. 2016YFC0401409), and the Research Fund of the State Key Laboratory of Eco-hydraulics in Northwest Arid Region, Xi'an University of Technology (Grant No. 2019KJCXTD-5). Sincere gratitude is extended to the editor and to three anonymous reviewers for their professional comments and corrections.

References

Abbott, M. B., Bathurst, J. C., Cunge, J. A., O'Connell, P. E., and Rasmussen, J.: An introduction to the European Hydrological System — Systeme Hydrologique Europeen, "SHE", 1: History and philosophy of a physically-based, distributed modelling system, Journal of Hydrology, 87, 45–59, doi:10.1016/0022-1694(86)90114-9, 1986.

Adamowski, J. and Sun, K.: Development of a coupled wavelet transform and neural network method for flow forecasting of non-perennial rivers in semi-arid watersheds, Journal of Hydrology, 390, 85–91, doi:10.1016/j.jhydrol.2010.06.033, 2010.

Ashrafi, M., Chua, L. H. C., Quek, C., and Qin, X.: A fully-online Neuro-Fuzzy model for flow forecasting in basins with limited data, Journal of Hydrology, 545, 424–435, doi:10.1016/j.jhydrol.2016.11.057, 2017.

Bai, Y., Chen, Z., Xie, J., and Li, C.: Daily reservoir inflow forecasting using multiscale deep feature learning with hybrid models, Journal of Hydrology, 532, 193–206, doi:10.1016/j.jhydrol.2015.11.011, 2016.

Beven, K.: Changing ideas in hydrology — The case of physically-based models, Journal of Hydrology, 105, 157–172, doi:10.1016/0022-1694(89)90101-7, 1989.

Binley, A. M., Beven, K. J., Calver, A., and Watts, L. G.: Changing responses in hydrology: Assessing the uncertainty in physically based model predictions, Water Resour. Res., 27, 1253–1261, doi:10.1029/91WR00130, 1991.

https://doi.org/10.5194/hess-2019-565
Preprint. Discussion started: 20 December 2019

Bishop, C. M.: Pattern recognition and machine learning, Information science and statistics, Springer, New York, xx, 738, 2006.

Castellano-Méndez, M., González-Manteiga, W., Febrero-Bande, M., Manuel Prada-Sánchez, J., and Lozano-Calderón, R.: Modelling of the monthly and daily behaviour of the runoff of the Xallas river using Box–Jenkins and neural networks methods, Journal of Hydrology, 296, 38–58, doi:10.1016/j.jhydrol.2004.03.011, 2004.

Clark, M. P., Nijssen, B., Lundquist, J. D., Kavetski, D., Rupp, D. E., Woods, R. A., Freer, J. E., Gutmann, E. D., Wood, A. W., Brekke, L. D., Arnold, J. R., Gochis, D. J., and Rasmussen, R. M.: A unified approach for process-based hydrologic modeling: 1. Modeling concept, Water Resour. Res., 51, 2498–2514, doi:10.1002/2015WR017198, 2015.

Devia, G. K., Ganasri, B. P., and Dwarakish, G. S.: A Review on Hydrological Models, Aquatic Procedia, 4, 1001–1007, doi:10.1016/j.aqpro.2015.02.126, 2015.

Dragomiretskiy, K. and Zosso, D.: Variational Mode Decomposition, IEEE Trans. Signal Process., 62, 531–544, doi:10.1109/TSP.2013.2288675, 2014.

Du, K., Zhao, Y., and Lei, J.: The incorrect usage of singular spectral analysis and discrete wavelet transform in hybrid models to predict hydrological time series, Journal of Hydrology, 552, 44–51, doi:10.1016/j.jhydrol.2017.06.019, 2017.

Erdal, H. I. and Karakurt, O.: Advancing monthly streamflow prediction accuracy of CART models using ensemble learning paradigms, Journal of Hydrology, 477, 119–128, doi:10.1016/j.jhydrol.2012.11.015, 2013.

Fang, W., Huang, S., Ren, K., Huang, Q., Huang, G., Cheng, G., and Li, K.: Examining the applicability of different sampling techniques in the development of decomposition-based streamflow forecasting models, Journal of Hydrology, 568, 534–550, doi:10.1016/j.jhydrol.2018.11.020, 2019.

Gai, L., Nunes, J. P., Baartman, J. E.M., Zhang, H., Wang, F., Roo, A. de, Ritsema, C. J., and Geissen, V.: Assessing the impact of human interventions on floods and low flows in the Wei River Basin in China using the LISFLOOD model, Science of The Total Environment, 653, 1077–1094, doi:10.1016/j.scitotenv.2018.10.379, 2019.

Golyandina, N., Nekrutkin, V. V., and Zhigli͡avskiĭ, A. A.: Analysis of time series structure: SSA and related techniques / Nina Golyandina, Vladimir Nekrutkin, and Anatoly Zhigljavsky, Monographs on statistics and applied probability, 90, Chapman & Hall, Boca Raton, Fla., London, 2001.

Grayson, R. B., Moore, I. D., and McMahon, T. A.: Physically based hydrologic modeling: 2. Is the concept realistic?, Water Resour. Res., 28, 2659–2666, doi:10.1029/92WR01259, 1992.

Han, D., Cluckie, I. D., Karbassioun, D., Lawry, J., and Krauskopf, B.: River Flow Modelling Using Fuzzy Decision Trees, Water Resources Management, 16, 431–445, doi:10.1023/A:1022251422280, 2002.

He, X., Luo, J., Zuo, G., and Xie, J.: Daily Runoff Forecasting Using a Hybrid Model Based on Variational Mode Decomposition and Deep Neural Networks, Water Resour Manage, 33, 1571–1590, doi:10.1007/s11269-019-2183-x, 2019.

https://doi.org/10.5194/hess-2019-565
Preprint. Discussion started: 20 December 2019

He, Z., Wen, X., Liu, H., and Du, J.: 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, Journal of Hydrology, 509, 379–386, doi:10.1016/j.jhydrol.2013.11.054, 2014.

Hosseini, S. M. and Mahjouri, N.: Integrating Support Vector Regression and a geomorphologic Artificial Neural Network for daily rainfall-runoff modeling, Applied Soft Computing, 38, 329–345, doi:10.1016/j.asoc.2015.09.049, 2016.

Huang, S., Chang, J., Huang, Q., and Chen, Y.: Monthly streamflow prediction using modified EMD-based support vector machine, Journal of Hydrology, 511, 764–775, doi:10.1016/j.jhydrol.2014.01.062, 2014.

Hunter, J. D.: Matplotlib: A 2D Graphics Environment, Computing in Science & Engineering, 9, 90–95, doi:10.1109/MCSE.2007.55, 2007.

James, B., Bardenet, R., Bengio, Y., and Balázs Kégl (Eds.): Algorithms for Hyper-Parameter Optimization, 2011.

Jiang, R., Wang, Y., Xie, J., Zhao, Y., Li, F., and Wang, X.: Assessment of extreme precipitation events and their teleconnections to El Niño Southern Oscillation, a case study in the Wei River Basin of China, Atmospheric Research, 218, 372–384, doi:10.1016/j.atmosres.2018.12.015, 2019.

Jolliffe, I. T.: Principal Component Analysis, Second Edition, Springer Series in Statistics, Springer-Verlag New York Inc, New York, NY, 2002.

Kirchner, J. W.: Getting the right answers for the right reasons: Linking measurements, analyses, and models to advance the science of hydrology, Water Resour. Res., 42, 2465, doi:10.1029/2005WR004362, 2006.

Kisi, O.: Wavelet regression model for short-term streamflow forecasting, Journal of Hydrology, 389, 344–353, doi:10.1016/j.jhydrol.2010.06.013, 2010.

Kratzert, F., Klotz, D., Brenner, C., Schulz, K., and Herrnegger, M.: Rainfall–runoff modelling using Long Short-Term Memory (LSTM) networks, Hydrol. Earth Syst. Sci., 22, 6005–6022, doi:10.5194/hess-22-6005-2018, 2018.

Li, M., Wang, Q. J., Bennett, J. C., and Robertson, D. E.: A strategy to overcome adverse effects of autoregressive updating of streamflow forecasts, Hydrol. Earth Syst. Sci., 19, 1–15, doi:10.5194/hess-19-1-2015, 2015.

Liu, Z., Zhou, P., Chen, G., and Guo, L.: Evaluating a coupled discrete wavelet transform and support vector regression for daily and monthly streamflow forecasting, Journal of Hydrology, 519, 2822–2831, doi:10.1016/j.jhydrol.2014.06.050, 2014.

Luo, X., Yuan, X., Zhu, S., Xu, Z., Meng, L., and Peng, J.: A hybrid support vector regression framework for streamflow forecast, Journal of Hydrology, 568, 184–193, doi:10.1016/j.jhydrol.2018.10.064, 2019.

Maity, R., Bhagwat, P. P., and Bhatnagar, A.: Potential of support vector regression for prediction of monthly streamflow using endogenous property, Hydrol. Process., 24, 917–923, doi:10.1002/hyp.7535, 2010.

McKinney, W.: Data Structures for Statistical Computing in Python, Proceedings of the 9th Python in Science Conference, 51–56, 2010.

Meng, E., Huang, S., Huang, Q., Fang, W., Wu, L., and Wang, L.: A robust method for non-stationary streamflow prediction based on improved EMD-SVM model, Journal of Hydrology, 568, 462–478, doi:10.1016/j.jhydrol.2018.11.015, 2019.

Minka, T. P.: Automatic Choice of Dimensionality for PCA, Advances in Neural Information Processing Systems, 598–604, 2001.

Mohammadi, K., Eslami, H. R., and Kahawita, R.: Parameter estimation of an ARMA model for river flow forecasting using goal programming, Journal of Hydrology, 331, 293–299, doi:10.1016/j.jhydrol.2006.05.017, 2006.

Mulvaney, T. J.: On the use of self-registering rain and flood gauges in making observations of the relations of rainfall and of flood discharges in a given catchment, Proceedings Institution of Civil Engineers, 4, 18–31, 1850.

Nash, J. E. and Sutcliffe, J. V.: River flow forecasting through conceptual models part I — A discussion of principles, Journal of Hydrology, 10, 282–290, doi:10.1016/0022-1694(70)90255-6, 1970.

Newman, A. J., Clark, M. P., Sampson, K., Wood, A., Hay, L. E., Bock, A., Viger, R. J., Blodgett, D., Brekke, L., Arnold, J. R., Hopson, T., and Duan, Q.: Development of a large-sample watershed-scale hydrometeorological data set for the contiguous USA: data set characteristics and assessment of regional variability in hydrologic model performance, Hydrol. Earth Syst. Sci., 19, 209–223, doi:10.5194/hess-19-209-2015, 2015.

Nourani, V., Komasi, M., and Mano, A.: A Multivariate ANN-Wavelet Approach for Rainfall–Runoff Modeling, Water Resources Management, 23, 2877, doi:10.1007/s11269-009-9414-5, 2009.

Paniconi, C. and Putti, M.: Physically based modeling in catchment hydrology at 50: Survey and outlook, Water Resour. Res., 51, 7090–7129, doi:10.1002/2015WR017780, 2015.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., and Duchesnay, É.: Scikit-learn: Machine Learning in Python, Journal of Machine Learning Research, 12, 2825–2830, 2011.

Quilty, J. and Adamowski, J.: Addressing the incorrect usage of wavelet-based hydrological and water resources forecasting models for real-world applications with best practices and a new forecasting framework, Journal of Hydrology, 563, 336–353, doi:10.1016/j.jhydrol.2018.05.003, 2018.

Rasouli, K., Hsieh, W. W., and Cannon, A. J.: Daily streamflow forecasting by machine learning methods with weather and climate inputs, Journal of Hydrology, 414-415, 284–293, doi:10.1016/j.jhydrol.2011.10.039, 2012.

Sachindra, D. A., Ahmed, K., Rashid, M. M., Sehgal, V., Shahid, S., and Perera, B. J. C.: Pros and cons of using wavelets in conjunction with genetic programming and generalised linear models in statistical downscaling of precipitation, Theor Appl Climatol, 108, D8, doi:10.1007/s00704-019-02848-2, 2019.

Seo, Y., Kim, S., Kisi, O., and Singh, V. P.: Daily water level forecasting using wavelet decomposition and artificial intelligence techniques, Journal of Hydrology, 520, 224–243, doi:10.1016/j.jhydrol.2014.11.050, 2015.

Shahriari, B., Swersky, K., Wang, Z., Adams, R. P., and Freitas, N. de: Taking the Human Out of the Loop: A Review of Bayesian Optimization, Proc. IEEE, 104, 148–175, doi:10.1109/JPROC.2015.2494218, 2016.

Singh, V. P.: Hydrologic modeling: progress and future directions, Geosci. Lett., 5, 1145, doi:10.1186/s40562-018-0113-z, 2018.

https://doi.org/10.5194/hess-2019-565
Preprint. Discussion started: 20 December 2019


Hydrology and Earth System Sciences
Discussions

Sivapragasam, C., Liong, S.-Y., and Pasha, M. F. K.: Rainfall and runoff forecasting with SSA–SVM approach, Journal of Hydroinformatics, 3, 141–152, doi:10.2166/hydro.2001.0014, 2001.

Solomatine, D. P., Maskey, M., and Shrestha, D. L.: Instance-based learning compared to other data-driven methods in hydrological forecasting, Hydrol. Process., 22, 275–287, doi:10.1002/hyp.6592, 2008.

Stojković, M., Kostić, S., Plavšić, J., and Prohaska, S.: A joint stochastic-deterministic approach for long-term and short-term modelling of monthly flow rates, Journal of Hydrology, 544, 555–566, doi:10.1016/j.jhydrol.2016.11.025, 2017.

Tan, Q.-F., Lei, X.-H., Wang, X., Wang, H., Wen, X., Ji, Y., and Kang, A.-Q.: An adaptive middle and long-term runoff forecast model using EEMD-ANN hybrid approach, Journal of Hydrology, 567, 767–780, doi:10.1016/j.jhydrol.2018.01.015, 2018.

Tim, H., MechCoder, Gilles, L., Iaroslav, S., fcharras, Zé Vinícius, cmmalone, Christopher, S., nel215, Nuno, C., Todd, Y., Stefano, C., Thomas, F., rene-rex, Kejia, (K.) S., Justus, S., carlosdanielcsantos, Hvass-Labs, Mikhail, P., SoManyUsernamesTaken, Fred, C., Loïc, E., Lilian, B., Mehdi, C., Karlson, P., Fabian, L., Christophe, C., Anna, G., Andreas, M., and Alexander, F.: Scikit-Optimize/Scikit-Optimize: V0.5.2, Zenodo, 2018.

Tiwari, M. K. and Chatterjee, C.: Development of an accurate and reliable hourly flood forecasting model using wavelet–bootstrap–ANN (WBANN) hybrid approach, Journal of Hydrology, 394, 458–470, doi:10.1016/j.jhydrol.2010.10.001, 2010.

Todini, E.: Hydrological catchment modelling: past, present and future, Hydrol. Earth Syst. Sci., 11, 468–482, doi:10.5194/hess-11-468-2007, 2007.

Valipour, M., Banihabib, M. E., and Behbahani, S. M. R.: Comparison of the ARMA, ARIMA, and the autoregressive artificial neural network models in forecasting the monthly inflow of Dez dam reservoir, Journal of Hydrology, 476, 433–441, doi:10.1016/j.jhydrol.2012.11.017, 2013.

van der Walt, S., Colbert, S. C., and Varoquaux, G.: The NumPy Array: A Structure for Efficient Numerical Computation: A Structure for Efficient Numerical Computation, Comput. Sci. Eng., 13, 22–30, doi:10.1109/MCSE.2011.37, 2011.

Vapnik, V., Golowich, S. E., and Smola, A. J.: Support Vector Method for Function Approximation, Regression Estimation and Signal Processing, Advances in Neural Information Processing Systems, 281–287, 1997.

Vapnik, V. N.: The Nature of Statistical Learning Theory, Springer New York, New York, NY, 1995.

Woldemeskel, F., McInerney, D., Lerat, J., Thyer, M., Kavetski, D., Shin, D., Tuteja, N., and Kuczera, G.: Evaluating post-processing approaches for monthly and seasonal streamflow forecasts, Hydrol. Earth Syst. Sci., 22, 6257–6278, doi:10.5194/hess-22-6257-2018, 2018.

Wu, C. L., Chau, K. W., and Li, Y. S.: Predicting monthly streamflow using data-driven models coupled with data-preprocessing techniques, Water Resour. Res., 45, 1331, doi:10.1029/2007WR006737, 2009.

Wu, Z. and Huang, N. E.: Ensemble Empirical Mode Decomposition: a Noise-Assisted Data Analysis Method, Adv. Adapt. Data Anal., 01, 1–41, doi:10.1142/S1793536909000047, 2009.

https://doi.org/10.5194/hess-2019-565
Preprint. Discussion started: 20 December 2019

Hydrology and Earth System Sciences
Discussions

Xie, T., Zhang, G., Hou, J., Xie, J., Lv, M., and Liu, F.: Hybrid forecasting model for non-stationary daily runoff series: A case study in the Han River Basin, China, Journal of Hydrology, 577, 123915, doi:10.1016/j.jhydrol.2019.123915, 2019.

Yaseen, Z. M., Ebtehaj, I., Bonakdari, H., Deo, R. C., Danandeh Mehr, A., Mohtar, W. H. M. W., Diop, L., El-Shafie, A., and Singh, V. P.: Novel approach for streamflow forecasting using a hybrid ANFIS-FFA model, Journal of Hydrology, 554, 263–276, doi:10.1016/j.jhydrol.2017.09.007, 2017.

Yu, P.-S., Chen, S.-T., and Chang, I.-F.: Support vector regression for real-time flood stage forecasting, Journal of Hydrology, 328, 704–716, doi:10.1016/j.jhydrol.2006.01.021, 2006.

Yu, S., Xu, Z., Wu, W., and Zuo, D.: Effect of land use types on stream water quality under seasonal variation and topographic characteristics in the Wei River basin, China, Ecological Indicators, 60, 202–212, doi:10.1016/j.ecolind.2015.06.029, 2016.

Zeng, Z., Hsieh, W. W., Shabbar, A., and Burrows, W. R.: Seasonal prediction of winter extreme precipitation over Canada by support vector regression, Hydrol. Earth Syst. Sci., 15, 65–74, doi:10.5194/hess-15-65-2011, 2011.

Zhang, X., Peng, Y., Zhang, C., and Wang, B.: Are hybrid models integrated with data preprocessing techniques suitable for monthly streamflow forecasting? Some experiment evidences, Journal of Hydrology, 530, 137–152, doi:10.1016/j.jhydrol.2015.09.047, 2015.

Zhao, X.-h. and Chen, X.: Auto Regressive and Ensemble Empirical Mode Decomposition Hybrid Model for Annual Runoff Forecasting, Water Resources Management, 29, 2913–2926, doi:10.1007/s11269-015-0977-z, 2015.