Wavelet based relevance vector machine model for monthly runoff prediction

Fang Ruiming

ABSTRACT

In this study, wavelet transform (WT) and a relevance vector machine (RVM) are integrated to predict monthly runoff. First, the WT method is adopted to decompose the monthly runoff time series into subsequences of different scales, and the variation characteristics, especially the periodicity of the runoff, are analyzed. Then, the regression model of RVM is established in each subsequence. Finally, the prediction results of each subsequence are integrated to obtain the final predicted values of monthly runoff through wavelet inverse transform. The proposed model was tested using the historical data of Minjiang River; the results show that compared with the RVM model, the WT-RVM model has better precision and can be applied in the prediction of monthly runoff.

Key words | hybrid model, Mallat wavelet transform, monthly runoff prediction, relevance vector machine

INTRODUCTION

Runoff prediction is an important part of hydrological prediction, which is of great significance to the comprehensive development and utilization of water resources, including reservoir optimization, flood control, drought protection, shipping management and water resource allocation (Freer et al. 1996). Besides that, runoff prediction is also the basis of optimal operation and decision for hydropower stations (Zhou et al. 2009). Therefore, selecting a suitable runoff prediction model is vital for runoff prediction.

The medium- and long-term runoff prediction methods mainly include a flow driven model based on physical concept and a data driven model based on statistics (Gangopadhyay et al. 2009). The flow driven model requires the comprehensive consideration of the atmospheric circulation, the evolution of the long-term weather process and the physical condition of the underlying surface of the basin to simulate the physical process of runoff formation. It requires high dimensional input data and the modeling process is complex. The statistics based data driven model is a black box model, which is based on the hydrological time series to study the nonlinear relationship among hydrologic variables. This model needs less data and has a simple structure, so it is widely used (Chua et al. 2008). The most used data driven models include multivariate linear regression (MLR) (Rezaie-Balf et al. 2017), the differential autoregressive moving average model (ARIMA) (Valipour et al. 2013), and the autoregressive moving average model (ARMA) (Sunkwon et al. 2009). In recent years, some intelligent methods have been applied to runoff prediction, such as the support vector machine method (SVM) (Shahraiyni et al. 2015), and artificial neural network (ANN) (Elshorbagy et al. 2000). He et al. (2014) simulated the monthly runoff sequence by SVM model and compared the results with the results of AMRM and ANN models, and the results show that the predictive effect of SVM is best (He et al. 2014). However, SVM has many problems, such as some parameters to be determined, and the kernel function is restricted severely, which affects the practical application of SVM in runoff prediction.
Based on the Bayesian principle, Markov theory, automatic correlation decision priori and maximum likelihood, the relevance vector machine (RVM) algorithm is proposed by Tipping (2001) and Bishop & Tipping (2000). RVM uses a given training sample set to optimize the model parameters through Bayesian inference, and establishes the nonlinear mapping relationship between the input and output vectors. Compared with the SVM algorithm, the RVM algorithm is sparse, and its probabilistic output can provide more decision information. Moreover, the algorithm has few undetermined parameters and the kernel function is not restricted by Mercer conditions. Since the runoff time series data are always affected by many factors, there are often some mutation data, and these data are the main causes of runoff prediction error. Wavelet transform (WT) is a time-frequency analysis method, which overcomes the defects of the single resolution of short time Fourier transform, and has the ability to characterize local information in the time and frequency domain (Shoaib et al. 2014). WT can effectively separate the low-frequency trend term and high-frequency noise term in runoff, so that the object of study can be transformed into a subsequence with stronger regularity and correlation.

In this paper, by adapting the WT method, the runoff time series data are decomposed to separate the low frequency trend component and the high frequency noise component in runoff, so the runoff time series data can be transformed to a more regular and more relevant subsequence. Then some RVM prediction models are built with the different decomposed subsequence components as input. The prediction results of each RVM models are reconstructed by wavelet inverse transformation, and the final runoff prediction results are obtained. Finally, the method is verified by simulation and compared with other methods to testify the effectiveness and feasibility of the given model.

METHODOLOGY

Discrete wavelet transform and Mallat algorithm

WT can express the time and frequency information of time series simultaneously. Wavelet function is a type of function with concussion characteristics. It calculates individual spectral components by stretching translation transformation, so it can analyze non-stationary data.

The equations of continuous WT are generally defined as:

\[ W_j(a, b) = (f(t), \psi_{a,b}(t)) = |a|^{-1/2} f(t) \psi^*(\frac{t-b}{a}) dt \]  

where \( f(t) \) is a time series; \( a \) is a scale factor, and \( b \) is a translation factor; \( \psi(t) \) is a wavelet function, while \( \psi^*(t) \) is the complex conjugate of wavelet function.

In this paper, the Morlet wavelet is chosen as the wavelet function, which is a single frequency complex sine function under Gauss envelope. Its expression is:

\[ \psi(t) = e^{i\omega_0 t}e^{-t^2/2} \]  

where \( \omega_0 \) is a constant, and its frequency domain expression is \( \psi(\omega) = \sqrt{2\pi}e^{-(\omega-\omega_0)^2/2} \). When using the Morlet wavelet function, the relationship between the scaling scale \( a \) and the periodic \( T \) in Fourier transformation can be described as:

\[ T = \left[ \frac{4\pi}{\omega_0\sqrt{2 + \omega_0^2}} \right] \times a \]  

when \( \omega_0 = 6.2, T = a \).

In most cases, the given time series are always discrete. So, the Mallat algorithm is used to deal with the discrete data. The Mallat algorithm is a fast WT based on multiresolution analysis, including the decomposition algorithm and the reconstruction algorithm. For an original time series \( c_0 \), the Mallat wavelet fast decomposition algorithm is defined as:

\[
\begin{align*}
&\{ c_{j+1} = Hc_j, j = 1, 2, \ldots, J \\
&d_{j+1} = Gc_j, j = 1, 2, \ldots, J
\end{align*}
\]

The Mallat reconstruction algorithm is:

\[ c_j = Hc_{j+1} + Gd_{j+1}, j = J - 1, J - 2, \ldots, 0 \]  

where \( H \) is a lowpass filter, \( G \) is a high pass filter, and \( J \) is a scale.

By adapting Equation (4), \( c_0 \) can be decomposed into \( d_1, d_2, \ldots, d_J \) and \( c_j \), where \( c_j \) is called the low-frequency
part (approximate part) of original series, and \( d_j \) is called the high-frequency part (detail part) of the original series.

The frequency and time domain characteristics of hydrological time series can be reflected by \( W_j(j, k) \). When \( J \) is small, the resolution in frequency domain is low and the resolution in time domain is high, and \( d_j \) can be obtained, which is the wavelet coefficient of the high frequency part. It is mainly composed of random parts and reflects the effect of various determinate factors on the hydrological time series. When \( J \) is increased, the resolution of the frequency domain is high and the resolution of the time domain is low, and \( c_j \) is obtained, which is the wavelet coefficient of the low frequency part. It is composed of deterministic components and reflects the periodicity and trend of hydrological time series.

**Relevance vector machine**

RVM is a pattern recognition algorithm based on the kernel function method, which can realize nonlinear transformation between feature space and data space. RVM has the advantages of automatically setting parameters and using any kernel function. For a regression problem, the correspondence between the target value and the input can be expressed as:

\[
y_n = f(x_n; \omega) + \varepsilon_n
\]

where \( x_n, y_n \) is the input and target value of the \( n \)-th data; \( \omega = [\omega_0, \omega_1, \cdots, \omega_M]^T \) is the weight vector for identification; \( \varepsilon_n \) is an independent distributed error term and it obeys the normal distribution \( N \cdot (0, \sigma^2) \), so \( y_n \) obeys the normal distribution with a mean \( f(x_n; \omega) \) and a variance of \( \sigma^2 \).

Based on the theory of kernel method, \( f(x_n; \omega) \) can be expressed as:

\[
f(x_n; \omega) = \omega^T \phi(x_n) = \sum_{i=1}^{M} \omega_i K(x_n, x_i) + \omega_0
\]

where \( K(x_n, x_i) \) is a kernel function, and the Gauss kernel function is usually chosen in the regression problem; \( M \) is the sum of kernel function; \( \phi(x_n) = [1, K(x_n, x_1), \cdots, K(x_n, x_M)]^T \).

From Equation (6), it can be drawn that for the parameter \( \omega \), the model is linear while \( K(x_n, x_i) \) can be highly nonlinear, which also gives the model the ability to express the nonlinear problem, especially when the number of kernel functions is higher, and the model can be used flexibly.

According to the distribution of \( y_n \), in the independent test of \( N \) times, the probability of the occurrence of \( y|y_1, y_2, \cdots, y_n \) is:

\[
p(y|\omega, \sigma^2) = (2\pi\sigma^2)^{-\frac{N}{2}} \exp\left(-\frac{1}{2\sigma^2} \|y - \omega \|^2\right)
\]

where \( \phi = [\phi(x_1), \phi(x_2), \cdots, \phi(x_N)]^T \).

The values of \( \omega \) and \( \sigma^2 \) can be obtained by maximum likelihood estimation. To avoid over-learning situations, based on the Bayes principle, RVM considers \( \omega_i \) and \( \sigma^2 \) as random variables. The prior distribution of \( \omega_i \) is set to obey the normal distribution with 0 as mean and \( \alpha_i^{-1} \) as variance, and they are independent of each other. Furthermore, the prior distribution of the super parameter \( \alpha_i \) and \( \sigma^2 \) is the gamma distribution with both the shape parameter and the scale parameter as 0. By choosing the prior distribution, the model has a good ability to adapt to training data and can guarantee the sparsity and generalization ability. A posteriori distribution of \( \omega, \alpha \) and \( \sigma^2 \) is obtained based on Bayes formula as:

\[
p(\omega, \alpha, \sigma^2) \propto p(y|\omega, \sigma^2)p(\omega, \alpha, \sigma^2)\]

\[
p(\omega, \alpha, \sigma^2) \propto p(y|\omega, \sigma^2)p(\omega, \alpha, \sigma^2)\]

\[
p(\omega, \alpha, \sigma^2) \propto p(y|\omega, \sigma^2)p(\omega, \alpha, \sigma^2)
\]

\[
p(\omega, \alpha, \sigma^2) \propto p(y|\omega, \sigma^2)p(\omega, \alpha, \sigma^2)
\]

The left side of Equation (9) can be decomposed:

\[
p(\omega, \alpha, \sigma^2) = p(y|\omega, \sigma^2)p(\omega, \alpha, \sigma^2)
\]

By using Equation (10), the original problem is transformed into a posterior distribution of \( \alpha, \sigma^2 \) in the case of \( y \) and \( \omega \) in the case of \( \alpha, \sigma^2, y \). In fact, for the simplified calculation, it is considered that the posterior distribution of \( \alpha, \sigma^2 \) is Fermi–Dirac distribution \( \delta(\alpha_{MP}, \sigma_{MP}^2) \), where \( \alpha_{MP}, \sigma_{MP}^2 \) respectively represent the maximum likelihood estimation of \( \alpha, \sigma^2 \). Then, the value of \( \omega \) which obeys the
normal distribution can be calculated, that is, the mean and variance are:

\[ E = (\sigma^2 \mathbf{\phi}^T \mathbf{\phi} + A)^{-1} \quad (11) \]
\[ \mu = \sigma^2 E \mathbf{\phi}^T y \quad (12) \]

where \( A = \text{diag}(\alpha_0, \alpha_1, \cdots, \alpha_M) \).

In the computation process, many posterior distributions of weights approach zero. For the relevance vector regression model, these non-zero vectors represent the prototype samples in the data, so they are called ‘relevance vectors’. The relevance vector embodies the core features of data centralization. After obtaining the model parameters through training data, for the new input vector \( x^* \), the distribution density of the target value \( y^* \) is:

\[ p(y^* | y) = \int p(y^* | \omega, \alpha, \sigma^2) p(\omega, \alpha, \sigma^2 | y) \text{d}\omega \text{d}\alpha \text{d}\sigma^2 \quad (13) \]

From Equation (12), \( y^* \) is a normal distribution. The expected value \( \overline{y^*} \) and variance \( \sigma^2 \) are:

\[ \begin{align*}
\overline{y^*} &= \mu^T \mathbf{\phi}(x^*) \\
\sigma^2 &= \sigma^2_{\text{MP}} + \phi^T(x^*) E \phi(x^*) \end{align*} \quad (14) \]

### Wavelet based relevance vector machine model for monthly runoff prediction

The block diagram of runoff prediction model based on the wavelet RVM is shown in **Figure 1**. The detailed steps are as follows:

1. The original runoff time series is decomposed by WT, and the non-stationary time series is decomposed into subsequences with different scale characteristics.
2. The decomposed runoff subsequence components are inputted separately into different RVM models, and each RVM model is used for regression prediction to obtain the prediction components of each subsequence.
3. The predicted components of each group were reconstructed by wavelet, and the final runoff prediction results were obtained.

### APPLICATION ANALYSIS AND DISCUSSION

In this paper, the monthly runoff prediction in Minjiang basin of Fujian province, China is taken as an example. The Minjiang basin is rich in rainfall. The historical data of monthly runoff data from January 1967 to December 2017 are shown in **Figure 2**.

### Time-varying characteristics analysis of runoff time series based on WT

Based on WT, some specific scales \( a = 16, 32, 64, 128 \) are selected, and the time change process diagram of the wavelet real part is drawn as in **Figure 3**, which can be used to study the periodic variation of the monthly mean runoff with time.

The following points can be analyzed from **Figure 3**:

1. When \( a = 16 \) months, the positive and negative phases and abrupt points appear more frequently because of the smaller time scale.
2. When \( a = 64 \) months, the alternation of abundance and withering is still very frequent. At this time scale, the...
monthly average runoff mainly experienced 18 alternations of abundance and withering.

(3) When $n = 128$ months, the monthly average runoff is more regular, and at this time scale, the monthly average runoff mainly experienced nine alternations of abundance and withering.

For the time series prediction problem, it is very important to choose the appropriate number of hysteresis input variables. Uncorrelated inputs will lead to poor accuracy and high complexity of models. It is assumed that there is an unknown functional relationship between the monthly mean runoff time series as follows:

$$Q(t) = f(Q(t - 1), Q(t - 2), \ldots, Q(t - \text{Lag}))$$  \hspace{1cm} (15)

where $Q(t)$ is the flow of current period, and $Q(t - \text{Lag})$ is the flow before the lag period. The correlation between the current flow and the early flow is reflected in the lag. In essence, it reflects the inherent characteristics of the basin, and how to select the parameter lag will greatly affect the prediction results. In this paper, the optimum lag value is determined by the analysis of the autocorrelation coefficient.

It can be seen from Figure 4 that the autocorrelation coefficient is maximum when lag = 12, while the partial autocorrelation coefficients all fall within the 95% confidence interval. Therefore, the Lag is chosen as 12, that is, the runoff data of 12 months before the predicted month are chosen as the input factor of the prediction model.

**Prediction results and discussion**

Using the wavelet transformation, the normalized runoff time sequence is decomposed by the DB wavelet function in five layers, as shown in Figure 5.

In Figure 5, D1, D2, \ldots, D5 is the high-frequency component decomposed, and A5 is the low frequency component. Each subsequence is used to train each RVM model, and the prediction results of each RVM model are reconstructed to obtain the final prediction results, as shown in Figure 6(b). For comparison, Figure 6(a) also provides the results predicted by using the RVM model only.

In this study, the accuracy of the model is determined by using the root mean square error (RMSE), the mean
absolute error (MAE), and the relative error (RE). The formulas are as follows:

\[
RMSE = \sqrt{\frac{1}{n} \sum_{t=1}^{n} (x_t - y_t)^2} \tag{16}
\]

\[
MAE = \frac{1}{n} \sum_{t=1}^{n} |x_t - y_t| \tag{17}
\]

\[
RE = \frac{\left( \sum_{t=1}^{n} y_t - \sum_{t=1}^{n} x_t \right)}{\sum_{t=1}^{n} x_t} \times 100\% \tag{18}
\]
where $x_t$ is the measured value of monthly runoff, $y_t$ is the predicted value of monthly runoff, and $n$ is the number of months to be predicted.

In order to analyze the prediction results, the accuracy of the prediction is judged by RMSE, MAE, and RE. The smaller RMSE and MAE are, and the closer RE is to 0, the higher the precision obtained.

A single support vector machine model is also given for prediction monthly runoff of Minjiang River with the same data. A comparison of the two models is shown in Table 1.

As shown in Table 1, the prediction accuracy of the WT + RVM model is higher than that of the single RVM model. The RMSE, MAE, and RE of the WT + RVM model are smaller, indicating a better prediction accuracy.
model are, respectively, 19.45 m$^3$/s, 5.78 m$^3$/s, and 4.8%. All indexes are superior to those of the single RVM model.

According to the specifications for hydrological information and forecasting (Water Resource Information Center of the Water Resources Ministry 2001), when the prediction RE is less than 20%, it can be regarded as a qualified standard. The RE between the predicted results obtained by the two models and the measured values is analyzed, and the percentage of prediction results with relative error less than 20%, and the relative error less than 30%, are calculated respectively, as shown in Table 2.

It can be seen from Table 2 that the qualified rate of the WT + RVM model is 84.33%, which is higher than 80%, indicating that the prediction accuracy of the WT + RVM model meets the requirements and can be applied to the monthly runoff prediction of the Minjiang River basin.

### CONCLUSIONS

In this paper, the WT and the RVM are combined to predict the monthly runoff of Minjiang River. The WT is used to decompose the non-stationary of the monthly runoff time series data. The RVM method is used to map the nonlinear relationship between the input and output variables and to form a regression prediction model. This method satisfies the requirement that the relative errors should be less than 30% in monthly runoff prediction. Compared with the RVM model, the WTRVM model has a higher precision prediction effect.

### REFERENCES

Bishop, C. M. & Tipping, M. E. 2000 Variational relevance vector machine. In: The 16th Conference on Uncertainty in Artificial Intelligence, Morgan Kaufmann, USA.


