Journal of Hydrology 552 (2017) 44–51

Contents lists available at ScienceDirect

Journal of Hydrology

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

Research papers

The incorrect usage of singular spectral analysis and discrete wavelet transform in hybrid models to predict hydrological time series

Kongchang Du [a,b], Ying Zhao [a,*], Jiaqiang Lei [a]

[a] Xinjiang Institute of Geography and Ecology, Chinese Academy of Sciences, Urumqi 830011, China
[b] University of Chinese Academy of Sciences, Beijing 100049, China

ARTICLE INFO

Article history:
Received 14 March 2017
Received in revised form 14 May 2017
Accepted 14 June 2017
Available online 19 June 2017
This manuscript was handled by A. Bardossy, Editor-in-Chief, with the assistance of Sheng Yue, Associate Editor

Keywords:
Hydrological time series
Prediction
Discrete wavelet transform
Singular spectrum analysis
Artificial network

ABSTRACT

In hydrological time series prediction, singular spectrum analysis (SSA) and discrete wavelet transform (DWT) are widely used as preprocessing techniques for artificial neural network (ANN) and support vector machine (SVM) predictors. These hybrid or ensemble models seem to largely reduce the prediction error. In current literature researchers apply these techniques to the whole observed time series and then obtain a set of reconstructed or decomposed time series as inputs to ANN or SVM. However, through two comparative experiments and mathematical deduction we found the usage of SSA and DWT in building hybrid models is incorrect. Since SSA and DWT adopt 'future' values to perform the calculation, the series generated by SSA reconstruction or DWT decomposition contain information of 'future' values. These hybrid models caused incorrect 'high' prediction performance and may cause large errors in practice.

1. Introduction

For modeling and predicting hydrological time series, various data driven models have been proposed ranging from traditional statistical models to emerging machine learning methods (Valipour et al., 2013). Traditional time series models such as autoregressive moving average (ARMA) require data to be stationary and have a limited ability to capture nonlinear correlations in data. As machine learning methods, artificial neural network (ANN) and support vectors machine (SVM) are able to model perfectly linear and nonlinear relationships in data even without understanding the underlying physical mechanisms, which make them well suited for building data driven models (Nourani et al., 2014, 2009).

To further improve the prediction performance, data preprocessing techniques are combined with machine learning approaches. Among these techniques, singular spectrum analysis (SSA) and discrete wavelet transform (DWT) are widely used. SSA is thought to be an efficient algorithm to avoid the effect of discontinuous or intermittent values in hydrological time series (Baratta et al., 2003; Sivapragasam et al., 2001). Wavelet transform is thought to be an effective tool for extracting information of different time scales and widely used for various purposes (Kisi and Cimen, 2012, 2009; Nourani et al., 2014, 2009). These hybrid or ensemble methods seem to largely reduce the prediction error (He et al., 2015; Keller et al., 2001; Partal, 2009; Venkata Ramana et al., 2013; Wu et al., 2010, 2009a,b; Wu and Chau, 2013).

The usage of SSA and DWT as preprocessing methods is to apply these techniques to the whole observed raw time series and get a set of reconstructed or decomposed time series. Then the reconstructed time series were partly or entirely chosen as inputs of ANN or SVM models (Baratta et al., 2003; Sivapragasam et al., 2001; Wu et al., 2010, 2009a,b; Wu and Chau, 2013). However, we found these techniques may be widely incorrectly used in literature after the authors' careful validation. The authors' viewpoint is that the reconstructed or decomposed time series would contain information of the future values to be predicted. This is why the combined models seem to largely improve the prediction. This paper aims to verify this viewpoint. We complement the work by two comparisons through a case study to predict monthly rainfall using SSA-ANN and DWT-ANN approach.

2. Material and method

2.1. SSA

The invention of SSA could be traced back to 1986 (Broomhead and King, 1986). At present SSA has proved to be very successful in

* Corresponding author at: Xinjiang Institute of Geography and Ecology, Chinese Academy of Sciences, 818 South Beijing Road, Urumqi, Xinjiang, China.
E-mail address: yzhaosoils@gmail.com (Y. Zhao).

http://dx.doi.org/10.1016/j.jhydrol.2017.06.019
0022-1694/

the analysis of climatic, meteorological and geophysical time series. The main target of SSA is to decompose the raw time series into a sum of time series such that each one can be identified as trend, periodic, quasi-periodic component or noise (Abdollahzade et al., 2015). The standard SSA consists of four steps, which can be briefly described as follows (Golyandina et al., 2001).The first step is the construction of trajectory matrix. Suppose we have a time series $A = (a_1,a_2,\ldots N)$ of length N, and L be an integer. We can define the L-lagged vectors$X_j = (a_j,\ldots a_{j+L-1})^T$,$j = 1,\ldots K$, and the trajectory matrix $X = [X_1:\ldots:X_K]$. The second step is singular value decomposition (SVD) of the matrix X. This can be done through computing eigenvalues and eigenvectors of the matrix $S = XX^T$. This provides a set of L singular values, which are the square roots of the eigenvalues of the matrix S, and the corresponding left and right singular vectors. We thus obtain a representation of X as a sum of rank-one biorthogonal matrices $X_i(i = 1,\ldots,d)$, where $d(d \leqslant L)$ is the number of nonzero singular values of X. At the third step, the set of indices $I = \{1,\ldots,d\}$ is divided into several groups $I_1,\ldots,I_m$ and the matrices X_i within each group are summed. We get the representation

$$X = \sum_{k=1}^{m} X_{Ik}, \quad \text{where } X_{Ik} = \sum_{i \in Ik} X_i$$

At the fourth step, we average the diagonals $i + j = const$ of the matrices X_{Ik}. This gives us the SSA decomposition of the original series A into a sum of series

$$a_n = \sum_{k=1}^{m} a_n^{(k)}, \quad n = 1,\ldots N,$$

where for each k the series $a_n^{(k)}$ is the result of diagonal averaging of the matrix X_{Ik}.

2.2. DWT

Essentially speaking, the wavelet transform is to extract information hidden in the original signal. The continuous wavelet transform (CWT) is defined as the sum over all time of the signal multiplied by scaled, shifted versions of the wavelet $\psi(t)$:

$$W_\psi(a,b) = |a|^{-1/2} \int_{-\infty}^{+\infty} f(t)\bar{\psi}\left(\frac{t-b}{a}\right)dt \quad a \in R, b \in R, a \neq 0$$

where a is the scale parameter which scales the wavelet and b determines the location of the wavelet, $\bar{\psi}(t)$ is the complex conjugate functions of $\psi(t)$. The DWT can be thought as dyadic sampling of continuous wavelet transform. In this case, $a = 2^j$ and, within a given a, $b = k2^j$, where k is a location index and j is referred to as the decomposition level. The discretely scaled and translated wavelet can be written as

$$\psi_{j,k}(t) = 2^{-j/2}\psi(2^{-j}t - k),$$

and the discrete wavelet transform of $f(t)$ can be written as

$$W_\psi(j,k) = 2^{-j/2} \int_{-\infty}^{+\infty} \bar{\psi}(2^{-j}t - k)f(t)dt, \quad j = 0,1,\ldots,k \in Z.$$

For a discrete time series, $x(n)$, the dyadic wavelet transform becomes

$$W_\psi(j,k) = 2^{-j/2}\sum_{n=0}^{N-1}\psi(2^{-j}n - k)x_n$$

where $W_\psi(j,k)$ is wavelet coefficient for the discrete wavelet of scale $a = 2^j$ and location $b = 2^j k$.

The inverse discrete transform is given by

$$x_n = \bar{T} + \sum_{j=1}^{J}\sum_{k=0}^{2^{J-j}-1} W_\psi(j,k)2^{-j/2}\psi(2^{-j}n - k)$$

which can also be written as

$$x_n = \bar{T}(t) + \sum_{j=1}^{J} W_j(t)$$

in which $\bar{T}(t)$ is called the approximation at level J and $W_j(t)$ are details at levels $j = 1,2\ldots,J$.

Note that in reality the Mallat algorithm is used to do discrete wavelets decomposition and reconstruction. In this paper we consider the dyadic wavelet transform for discrete time series. Multiple-level decomposition is performed to get high-scale, low-frequency components (approximations) and low-scale, high-frequency components (details) of the original time series. In related articles these details and approximations were used as inputs of ANN or SVM, instead of the original time series, to further improve prediction.

2.3. ANN

In this study, a multi-layer feed-forward network type of ANN is used to construct the hybrid models. This type of ANN includes input layer, a number of hidden layers and output layer, each layer containing a number of nodes. ANNs with one hidden layer are thought to provide enough complexity to simulate the nonlinear relationships of the hydrological time series data. In this study, the ANN forecast model is formulized as

$$x_t = f(X_t, w, \theta, m, h) = \theta_o + \sum_{j=1}^{h} w_j^{out}\phi\left(\sum_{i=1}^{m} w_{ji}x_{t-i} + \theta_j\right)$$

where x is the hydrological time series of interest, m is the lagged time steps, ϕ denotes transfer functions; w_{ij} are the weights between the ith node of the input layer and the jth node of the hidden layer; h_j are biases associated to the jth node of the hidden layer; w_j^{out} are the weights between the jth node of the hidden layer and the node of the output layer; and h_o is the bias of the output node. These weights and biases are optimized by the common error back-propagation algorithm.

2.4. Experiment design

Suppose we have a monthly rainfall time series F of length N. We firstly compare the SSA (DWT) reconstructed (decomposed) time series of length M using the whole time series and using only the first $M(M < N)$ time series. If the two reconstructed (decomposed) times series are same it can be concluded that using the whole time series did not affect the SSA (DWT) analysis of the first M values, vice versa.

To examine the validity of the hybrid models built as the way in related articles, the next experiment can be divided into two steps. Firstly, the first M time series were preprocessed by SSA (DWT) and then used to construct the hybrid model and perform the 'inner' test. Note that in the SSA (DWT) process the inputs and targets were simultaneously preprocessed, which is also the situation in literature. The performance of the 'inner' test is equivalent to that of hybrid models in relevant articles.

Secondly, to perform the 'outer' test, this model was used to predict values beyond M. To be clear, if we want to predict value $M + K(M + K \leq N)$, we use the values precedent $M + K$ to conduct the SSA (DWT) and then use the resulting values as inputs of the constructed hybrid model to predict. This test is close to reality;