THE UNIVERSITY of EDINBURGH

Edinburgh Research Explorer

A Hybrid Wavelet de-noising and Rank-Set Pair Analysis approach for forecasting hydro-meteorological time series

Citation for published version:
Borthwick, A, Wang, D, He, H, Zhu, J, Lu, Y, Zeng, X, Wu, J, Wang, L, Zou, X, Wang, Y, Xu, P, Liu, J, Zou, Y & He, R 2018, 'A Hybrid Wavelet de-noising and Rank-Set Pair Analysis approach for forecasting hydro-meteorological time series' Environmental Research, pp. 269-281. DOI: 10.1016/j.envres.2017.09.033

Digital Object Identifier (DOI):
10.1016/j.envres.2017.09.033

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Peer reviewed version

Published In:
Environmental Research

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 05. Apr. 2019

Title: Hybrid Wavelet De-noising and Rank-Set Pair Analysis approach for forecasting hydro-meteorological time series

Authors: Dong Wang[*1], Handan He[1], Alistair G. Borthwick[2], Jieyu Zhu[1], Yuan Lu[1], Yuankun Wang[1], Xiankui Zeng[1], Jichun Wu[1], Lachun Wang[3], Xinqing Zou[3]

[1]Key Laboratory of Surficial Geochemistry, MOE, Department of Hydrosciences, School of Earth Sciences and Engineering, Collaborative innovation center of South China Sea Studies, State Key Laboratory of Pollution Control and Resource Reuse, Nanjing University, Nanjing, P.R. China

[2]School of Engineering, The University of Edinburgh, Edinburgh EH9 3JL, U.K; St Edmund Hall, Queen's Lane, Oxford OX1 4AR, UK.

[3]School of Geographic and Oceanographic science, Collaborative innovation center of South China Sea Studies, Nanjing University, Nanjing, P.R China

(*Corresponding author: Dong Wang, wangdong@nju.edu.cn)

Key points:

- ✧ A new hybrid approach is proposed for forecasting hydro-meteorological time series.
- ✧ Rank-Set Pair Analysis combined with wavelet de-noising markedly improves forecasting accuracy.
- ✧ The performance of this approach proves best amongst its present competitors even when the extreme value occurs.

ABSTRACT

Accurate, convenient forecasting of hydro-meteorological time series is presently a major challenge. This paper proposes a hybrid approach, WD-RSPA (Wavelet De-noising and Rank-Set Pair Analysis), that takes full advantage of a combination of Wavelet De-noising and Rank-Set Pair Analysis to improve forecasts of hydro-meteorological time series. Wavelet de-noising permits decomposition and reconstruction of series by the wavelet transform, and hence separation of noise from the original series. Rank-Set Pair Analysis (RSPA), a more reliable and efficient version of Set Pair Analysis, is integrated in the hybrid WD-RSPA approach. Two kinds of hydro-meteorological data sets with different characteristics and different influences from human activity at for representative stations are used to illustrate the WD-RSPA approach. The approach is also compared with three other generic methods: the conventional Auto Regressive Integrated Moving Average (ARIMA) method; Artificial Neural Networks (ANNs) (BP, MLP and RBF); and single Rank-Set Pair Analysis (RSPA). Nine error measures are used to evaluate model performance. The results show that WD-RSPA is accurate, feasible, and effective. Moreover, WD-RSPA is found to be the best of the various generic methods compared, even when the extreme value occurs within the series. The theoretical approach developed here could be applied more generally to forecasting time series in other relevant areas.

Index Terms: 1872, 1988, 1869, 3270

Key words: Data-driven model, Forecasting, Hydro-meteorological series, Rank-Set Pair Analysis, Wavelet De-noising

1 Introduction

Water is a prerequisite for life, and so its availability is fundamentally important for human society and the environment. However, many countries worldwide experience water problems related to the overabundance or lack of water, and deterioration in water quality; such problems include water shortages, droughts, floods, damage to aquatic eco-systems, and can be exacerbated by economic development and climate change [Whitworth *et al.* 2012; Mehran *et al.* 2015]. A major challenge is presently faced in how to ensure that water resources remain sustainable, and this is made harder by the insufficiency of hydrologic data in developing countries [Leung *et al.*, 2013; Qian and Leung, 2007]. Obviously, effective rainfall and runoff forecasting techniques are needed that provide scientific evidence and significant reference data to underpin water resources planning, design and management.

The hydrometeorological process is particularly complicated, partly owing to climate and anthropogenic drivers, and is associated with large uncertainties which can have random, fuzzy, fractal, and chaotic characteristics [(Sivakumar *et al.*, 1999; Kavvas *et al.*, 2013; Wang *et al.*, 2015]. Existing hydrological forecasting methods fall into two broad categories: (1) physically-based models; and (2) data-driven models [Shoaib *et al.,* 2015]. Physically-based models require a substantial amount of data to simulate the various constituent physical processes within a watershed. Data-driven models include stochastic methods and machine learning methods, and may have certain advantages over fully distributed models [Nourani *et al.,* 2013]. The most popular data-driven stochastic methods are the auto regressive integrated moving average (ARIMA) method, ARIMA with exogenous input (ARIMAX), and Multiple Linear Regression (MLR) [Pulido-Calvo and Portela, 2007; Zhang *et al.*, 2011]. Machine learning methods, such as supervised learning methods, are essentially based on statistical techniques for developing predictive models using training data. Unlike physics-based models, machine learning methods rely almost exclusively on information embedded in training datasets [Sun *et al.*, 2014]. The most commonly applied of these methods are Artificial Neural Network (ANN) and support vector machine (SVM) algorithms [Ghosh and Mujumdar, 2008; Wu and Chau 2011; Valipour *et al*. 2013; He *et al*. 2014]. The input data used in these two categories of hydrological forecasting models, which include calibration data for physically-based models and training data for data-driven models, help determine the

accuracy and reliability of the forecasting results.

Two critical issues arise. One concerns noise which contaminates input data derived from hydro-meteorological observations [Wang *et al*. 2014]. The presence of such noise alters the characteristics of the input time series, and limits the performance of identification, simulation, parameter estimation and prediction techniques [Minville *et al*., 2008]. Self-similarity, phase-space reconstruction at small length scales, prediction error, and period identification [Elshorbagy *et al*., 2002; Stevenson *et al*., 2010] may also be undermined. If noise-contaminated observed data are input to a forecasting model, there will undoubtedly be a negative impact on the predictions. Therefore, it is necessary to remove noise from observed data before input so as to enhance the accuracy and reliability of forecasts of meteorological and hydrologic time series. To achieve this, we propose preconditioning the observed data by wavelet de-noising, a technique based on wavelet analysis (WA) which has been found very effective in the multi-scale analysis of time series and has been widely applied to noise reduction [Labat, 2005; Schaefli *et al*., 2007; Nalley *et al*., 2012].

The second critical issue concerns the calculation methodology and applicability of the models. Currently, many different kinds of forecasting models have been developed. However, the mathematical complexity of these models has hindered further development and applicability. To overcome this drawback, a simple, effective hydro-meteorological forecasting approach is needed based on clear concepts, convenient calculations, and which is feasible to apply in practice. Herein, we use Rank-Set Pair Analysis which is a modification of Set Pair Analysis (SPA), a powerful uncertainty analysis method which analyzes the degree of connection of a set pair, aspects related to identity, discrepancy and contradiction. Following Zhao (2000), SPA has seen widespread applications in mathematics, physics, information science, economy, resource assessment, and environmental science. In the context of hydrology and environmental science, SPA has been used for urban ecosystem health assessment (Su *et al*., 2009), water resources system assessment (Wang *et al*., 2009), river health evaluation (Xu *et al*., 2011), landslide hazard degree assessment (Wang and Li, 2012), selection of a reference basin in ungauged regions (Wang *et al*., 2013), risk assessment and forewarning for regional water resources (Zhao *et al*., 2013), evaluation of drought index at multi-time scales (Zhang *et al*., 2013), water resources

trends (Feng *et al*., 2014), river basin resource compensation characteristics (Chen *et al.*, 2014), river eco-system assessment and restoration (Jiang *et al*. 2015), waterlog disaster risk evaluation (Jin *et al.*, 2015), and sustainability assessment of a water resources system (Du *et al*. 2015).

Several recent advances have improved the reliability and efficiency of SPA. Jin *et al.* (2012) established a forewarning model for sustainable water resources based on a BP neural network coupled with SPA. Yang *et al*. (2012) established an optimal weight combination model, involving rank-SPA, RBF and AR sub-models, which provided more accurate precipitation forecasts. Zou *et al.* (2013) proposed a model for comprehensive flood risk assessment based on SPA and variable fuzzy set (VFS) theory. Su *et al*. (2013) constructed an evaluation model of sea dike safety based on a modified SPA method. Zhang *et al.* (2013) established a SPA phase-space reconstruction (SPA-PSR) model that improved forecasting precision. Guo *et al.* (2014) presented a modified SPA to compute the relative membership degree functions of variable fuzzy set (VFS) theory used in flood risk assessment. Yang *et al.* (2014a) examined the relative performance of SPA and modified SPA in regional debris flow hazard assessment. Yang *et al*. (2014b) established an improved SPA model for systematic assessment of water resources vulnerability to climate change. Chou (2014) applied SPA with similarity forecast and wavelet de-noising to forecast annual runoff. Zhang and Wang (2015) used an entropy-weighted SPA model to evaluate the water resource security of a city. Wang et al. (2015b) utilized entropy weighted-SPA to identify dam leaks.

The present study proposes a hybrid approach, WD-RSPA, which takes full advantage of both wavelet de-noising (WD) and rank set-pair analysis (RSPA) in achieving accurate, convenient forecasts of meteorological and hydrologic time series. The performance of the WS-RSPA approach is examined using annual precipitation time series from stations in Zhengzhou (1951–2009) and Beijing (1951–2010), and annual runoff time series from the lower Yellow River at Huayuankou (1950–2007) and Sanmenxia (1956–2010). Results from the WD-RSPA approach are compared against those from three alternative methods: (1) conventional Auto Regressive Integrated Moving Average (ARIMA); (2) Artificial Neural Networks (ANNs) with BP (error Back Propagation), MLP (Multilayer Perceptron) and RBF (Radial Basis Function); and (3) single Rank-Set Pair Analysis (RSPA). Nine error metrics are used to evaluate

model performance. The results demonstrate that WD-RSPA is accurate, feasible and effective, and better synthetically than conventional ARIMA, ANNs, and single RSPA methods.

The paper is organized as follows. Section 2 briefly introduces the basic theory behind wavelet analysis and de-noising, Set Pair Analysis and Rank-Set Pair Analysis. Section 3 outlines the proposed WD-RSPA hybrid approach, coupling discrete wavelet de-noising with Rank-Set Pair Analysis. Section 4 describes the application of WD-RSPA to observed hydro-meteorological data, and the results are discussed in the context of alternative forecasting approaches. Section 5 lists the main conclusions.

2 Methodology

2.1 Wavelet analysis and de-noising

First, we describe certain key features of wavelet analysis; for a detailed discussion see Labat (2005), Chanerley and Alexander (2007), and Schaefli *et al*. (2007). Wavelet analysis defines a mother wavelet function, denoted $\psi(t)$ where t is time, that must satisfy the following admissibility condition in the frequency domain:

$$C_\psi = \int_{-\infty}^{+\infty} \frac{|\psi_F(\omega)|^2}{|\omega|} d\omega < \infty \quad , \tag{1}$$

where $\psi_F(\omega)$ is the Fourier transform of the wavelet function $\psi(t)$ at frequency ω. Wavelet functions are obtained by translating and expanding the mother wavelet function to give

$$\psi_{a,b}(t) = |a|^{-1/2} \psi\left(\frac{t-b}{a}\right) \quad a,b \in R, a \neq 0 \quad , \tag{2}$$

where $\psi_{a,b}(t)$ is the wavelet function; a is a temporal scale factor which reflects the periodic length of a wavelet; and b is a time position factor. To carry out the wavelet transform of a signal, one needs to define $L^2(R)$ as a measurable square integral function space on the real axis. Then, the continuous wavelet transform (CWT) of a signal $f(t) \in L^2(R)$ can be written,

$$W_f(a,b) = |a|^{-1/2} \int_{-\infty}^{+\infty} f(t) \psi^*\left(\frac{t-b}{a}\right) dt \quad , \tag{3}$$

where $\psi^*(t)$ is the complex conjugate of $\psi(t)$, and $W_f(a,b)$ are the so-called wavelet coefficients.

Meteorological or hydrologic time series are normally expressed as discrete signals, $f(k\Delta t)$ (k = 1, 2,…, n in which Δt is the time interval), with a and b also given discrete values. The discrete wavelet

transform (DWT) of signal $f(k\Delta t)$ is expressed as

$$W_{j,k} = a_0^{-j/2} \int_{-\infty}^{+\infty} f(t)\psi^*(a_0^{-j}t - kb_0)dt \tag{4}$$

where a_0 ($a_0 > 1$) and b_0 are constants. Integer j is a temporal scale factor analogous to parameter a in equation (3), and kb_0 is a time position factor analogous to parameter b. In practice, the dyadic DWT is usually implemented by assigning $a_0 = 2$ and $b_0 = 1$.

The wavelet de-noising method proposed by Donoho *et al*. (1995) consists of three steps:

1. Decomposition: After choosing an appropriate wavelet function and resolution level M, the original data are decomposed into approximate coefficients at level M and detailed coefficients at various resolutions using DWT.

2. Threshold (T_j): Detailed coefficients $W_{j,k}$ from levels 1 to M undergo threshold selection, leading to decomposed coefficients. Choice of a suitable threshold of wavelet coefficient is undertaken using soft-threshold processing,

$$W'_{j,k} = \begin{cases} \mathrm{sgn}(W_{j,k})\left(|W_{j,k}| - T_j\right) & |W_{j,k}| > T_j \\ 0 & |W_{j,k}| < T_j \end{cases}. \tag{5}$$

3. Reconstruction: The decomposed coefficients of levels from 1 to M and the approximate coefficients at level M are reconstructed to de-noise the data.

At the core of wavelet de-noising is the selection of a reasonable threshold value, which relates directly to the quality of de-noising. Conventional methods for choosing a suitable threshold value include FT, SURE, or MINMAX.

2.2 Set Pair Analysis (SPA) and Rank-Set Pair Analysis (RSPA)

Set Pair Analysis (SPA) (see e.g. Zhao (2000); Su *et al.* (2009); Wang *et al.* (2009) for further details) simultaneously examines certainty and uncertainty links between objective things as an integrated determinate–indeterminate system. Properties of SPA are determined by means of identity, discrepancy and contradiction of a set pair, according to which the connection degree can be established. The principle of SPA is as follows. First the set pair for two relative sets in a uncertainty system is constructed; then its properties are determined according to identity, discrepancy and contradiction, namely I, D and C; finally the connection degree of the set pair is established according to I, D and C. In short, the basis of SPA is a

set pair, and its key is connection degree.

The set pair H(A, B) is constructed from set A and relative set B whose characteristics are given by coefficients a_1, a_2, …, a_n, and b_1, b_2, …, b_n. The degree of connection of the set pair H(A, B) is defined by

$$\mu_{A-B} = \frac{s}{n} + \frac{f}{n} i + \frac{p}{n} j \tag{6}$$

where n is the total number of characteristics of the set pair, s is the number of identity characteristics, f is the number of contrary characteristics, p is the number of the characteristics for which the set pair is neither identity nor contrary, i is the uncertainty coefficient of discrepancy degree (which has condition-dependent values in the range [−1, 1] or may be considered solely as a marker of discrepancy), and j is the uncertainty coefficient of contradiction degree (which has value of −1 or may be considered solely as a marker of contradiction). Equation (6) can be rewritten as

$$\mu_{A-B} = a + bi + cj \tag{7}$$

where $a = \frac{s}{n}$ is the identity degree, $b = \frac{f}{n}$ is the discrepancy degree, and $c = \frac{p}{n}$ is the contradiction degree and $a + b + c = 1$. Equations (6) and (7) describe the three-element connection degree. A multi-element connection degree can be obtained by expanding bi in (7) into $bi = b_1 i_1 + b_2 i_2 + \ldots + b_k i_k$. For example, when $k = 3$, the 5-element connection degree can be obtained as

$$\mu_{A-B} = a + b_1 i_1 + b_2 i_2 + b_3 i_3 + cj \tag{8}$$

where $a + b_1 + b_2 + b_3 + c = 1$. The discrepancy degree components, b_1, b_2, b_3, may be viewed as expressing whether discrepancy is mild, moderate, or severe; i_1, i_2, i_3, are uncertainty component coefficients of discrepancy degree. Choice of a, b (or b_1, b_2 and b_3) and c models the internal subtle structure of sets of A and B, feeding into μ_{A-B} which reflects the overall connectivity of sets A and B. Here, the connection degree overcomes certain drawbacks of conventional relationships such as correlation coefficient, subordinate degree, or grey correlation degree, each of which involve a single index. SPA has the following advantages: (1) clear exposition of the relationship structure; (2) quantification of three or more characteristics of a complex relationship; (3) determination of the

changeable value of a comprehensive relationship, which may depend on required different standards or properly selected values for i or i_1, i_2, i_3, ….

Rank-Set Pair Analysis (RSPA) [Ou *et al.*, 2009] is more reliable and efficient than SPA, and is based on the principle of similarity forecasting, taking full advantage of a combination of Rank, a classic stochastic concept, and Set Pair Analysis. RSPA is implemented as follows. The meteorological or hydrologic time series is denoted as $x_1, x_2, \cdots, x_n$, with an underlying assumption that x_t exists dependency with its previous T historical value. A moving method is used to obtain the historical sets, $\mathrm{B}_i (i = 1, 2, \mathrm{L}\ , n-T)$ corresponding to the subsequent value x_{T+i} (Table 1). Subsequent values x_{T+i} of the current set, denoted Y in Table 1, are obtained by (1) identifying the set or set group similar to Y in the historical set $B_1, B_2, \cdots, B_{n-T}$, and the corresponding subsequent values, and (2) using the weighted average method to obtain the forecasting value.

Insert Table 1 here

The detailed procedure is as follows.

(1) Undertake rank transformation. Mark the elements in $B_1, B_2, \cdots$ and Y from 1 to T according to the rank of elements in the sets to which they belong. Mark elements of the same rank, according to their rounded average value. Thus, obtain the rank set $B_1', B_2', \cdots$ and Y'.

(2) Construct $n-T$ rank set pairs (B_i', Y') (i=1, 2, . . . , $n-T$) and calculate the difference d between the corresponding elements of B_i' and Y'. If $|d| = 0$, $|d| > T-2$, or $0 \le |d| \le T-2$, the elements are respectively marked “identical”, “contrary”, or “discrepant.”

(3) Count the total number of “identical,”“contrary,” and “discrepant” elements in each rank set pair.

(4) Calculate value of connection degree for each rank set pair using equation (6).

(5) Find similar set B_i' (or several similar sets under certain circumstances) of Y in accordance with the maximum principle. B_i' is the counterpart of B_i whose subsequent value is $x_{\mathrm{T+i}}$.

(6) Evaluate

$$\hat{x}_{n+1} = \frac{1}{m}\sum_{k=1}^{m} \omega_k x_{T+k} \quad , \tag{9}$$

where ω_k is the ratio of the average value of the elements in *Y* and the average value of the elements in B_k, *m* is the number of the similar sets of *Y*.

3 Wavelet De-noising and Rank-Set Pair Analysis forecasting approach

The Wavelet De-noising and Rank-Set Pair Analysis (WD-RSPA) procedure is now described. Where appropriate, we alter the WD-RSPA acronym to identify the particular de-noising function selected; for example, if the wavelet de-noising function "db9" is used, then the established WD-RSPA model is referred to as db9-RSPA. WD-RSPA is implemented as follows:

(1) Select a time series of *n* consecutive years, $\{rf_i\}$ (i = 1, 2, …, n), from the observed hydro-meteorological time series as input data to the model.

(2) Use an appropriate wavelet function and a suitable decomposition level to compute the de-noised series $\{xi\}$.

(3) Determine set dimension, *T*, of the time series and establish historical rank sets $\{B_j\}$ (j = 1, 2, …, n-T) and current set *Y*, according to the length of the series (n) and Table 1.

(4) Make rank transformation of the sets from Step (3) to obtain the rank historical set $\{Bj'\}$ (j = 1, 2, ..., N-T) and a rank current set *Y'*, and so constitute the rank set pairs $(B_i', Y')\,(i = 1, 2, \cdots, n - T)$.

(5) Calculate connection degree of each rank set pair $(B_i', Y')\,(i = 1, 2, \cdots, n - T)$ using equation (6).

(6) Use equation (9) to obtain forecasting value for year n + 1.

The overall process involves one-step forecasting to take full advantage of the available information. Figure 1 provides a flow chart illustrating the above steps used to implement WD-RSPA for forecasting hydro-meteorological data series.

Insert Figure 1 here

4 Application and discussion

To evaluate the effectiveness of the WD-RSPA approach, we consider hydro-meteorological data sets from four stations in China. Figure 2 shows the locations of these representative stations, one in Beijing and three along the lower Yellow River. Case 1 involves the annual runoff series at Huayuankou from 1950 to 2007. Case 2 considers the annual runoff series at Sanmenxia station from 1956 to 2010. Cases 3 and 4 consider annual precipitation series at Zhengzhou from 1951 to 2009, and at Beijing from 1951 to 2010.

Insert Figure 2 here

The cases are used to compare WD-RSPA against three generic alternatives: Auto Regressive Integrated Moving Average (ARIMA) method with model selection based on the AIC criterion; three types of Artificial Neural Networks (ANNs) namely ANN-BP(Back propagation), ANN-MLP (Multilayer Perception) and ANN-RBF (Radial Basis Function); and single Rank- Set Pair Analysis (RSPA). The following nine measures are used for model evaluation: Relative Error (RE); percentage pass rate for RE < 10% (P_{10}); percentage pass rate for RE < 20% (P_{20}); Maximum Relative Error (Max_{RE}); Minimum Relative Error (Min_{RE}); Mean Relative Error (MRE); Standard Deviation of Relative Error (SD-RE); Root Mean Square Error (RMSE); and Thiel Inequality Coefficient (TIC).

4.1 Case 1: Annual runoff series at Huayuankou Station, lower Yellow River, China

Huayuankou Station is located in the lower Yellow River, China, near where the riverbed rises above the level of the surrounding land (the so-called perched Yellow River). We consider the observed annual runoff series at Huayuankou Station from 1950 to 2007, a period lasting 58 years. Using input data from 1950 to 1997, model forecasts of the annual runoff series from 1998 to 2007, are compared against observations.

Here, mother wavelet functions “coif3” and “bior2.4” are used with one resolution level during wavelet de-noising. The set dimension, *T*, used in the WD-RSPA and the single RSPA models is variously prescribed to be 4, 5, or 6.

Figure 3 presents the observed annual runoff time series with superimposed forecasts made by the AR(4), ANN-BP, coif3-RSPA, and bior2.4-RSPA models for Huayuankou, taking $T = 5$. Table 2 lists values of the observed annual runoff and nine performance measures obtained for the auto-regressive AR(4), ANN-BP, coif3-RSPA, and bior2.4-RSPA models for set dimension $T = 5$. In this case, the AR(4) model performed best out of the ARIMA models, and the ANN-BP model performed better than either the ANN-RBF or ANN-MLP models.

Insert Figure 3 here

Insert Table 2 here

It can be seen from Table 2 and Figure 3 that:

(1) Annual runoff forecasts by WD-RSPA depend on choice of wavelet de-noising function. This is particularly evident in years 2000 and 2003, when forecasts by the bior2.4-RSPA model (186.65 and 229.84) are much closer to observed values (165.30 and 272.70) than forecasts by the coif3-RSPA model (205.59 and 201.25).

(2) The WD-RSPA and ANN-BP models simulate correctly the changing characteristics of the observed series, unlike the AR(4) model. Forecasts by the AR(4) model are closer to the mean value of the observed series than those by the other models.

(3) The maximum value of Min_{RE} is obtained using the AR(4) model indicating that it gives the worst forecast over the 10 year period of interest.

(4) The minimum value of SD-RE is associated with the AR(4) model indicating that its forecasting errors are relatively concentrated in comparison with the other models whose forecasting errors are relatively dispersed.

(5) The maximum P_{10} and P_{20} measures are obtained for the bior2.4-RSPA model indicating it is most accurate.

(6) The other measures, Max_{RE}, MRE, RMSE and TIC, invariably have minimum values for the bior2.4-RSPA model, which indicates it provides the best comprehensive forecasts in this case of the methods considered.

Taking bior2.4 as the de-noising function, we now study the influence of the set dimension, *T*, used in WD-RSPA and single RSPA models. Table 3 compares the observed and forecast annual runoff series statistics by listing the performance measures of the single RSPA and bior2.4-RSPA models for T = 4, 5, and 6, applied to the data from Huayuankou.

Insert Table 3 here

Figure 4 compares the relative errors between observed and forecast annual runoff time series at Huayuankou, the latter obtained using the single RSPA and bior2.4-RSPA models for T = 4, 5, and 6.

Insert Figure 4 here

It can be seen from Table 3 and Figure 4 that:

(1) The performance measures of the WD-RSPA and single RSPA models are affected by the choice of set dimension, *T*.

(2) For each *T*, the P_{10} and P_{20} values obtained using bio2.4-RSPA tend to be greater than those using the single RSPA model, indicating that bio2.4-RSPA provides a better forecast in this case.

(3) For each *T*, the remaining six measures, Max_{RE}, Min_{RE}, MRE, SD-RE, RMSE and TIC, are invariably smaller for the bior2.4-RSPA model than for the single RSPA model, implying that that the forecasting capability of the bior2.4-RSPA model is better than the single RSPA model.

(4) For eight performance measures (P_{10}, P_{20}, MRE, Max_{RE}, Min_{RE}, SD-RE, RMSE and TIC), selection of the bior2.4-RSPA model with T = 5 invariably results in a minimum or maximum value corresponding to the most accurate, comprehensive forecasts of annual runoff.

4.2 Case 2: Annual runoff series at Sanmenxia Station, lower Yellow River, China

Sanmenxia Station is located close to a major dam upstream of Huayuankou, and monitors runoff in the upper reaches of the lower Yellow River. The observed annual runoff series from Sanmenxia station covers 55 years from 1956 to 2010. Here, annual forecasts are made for the period from 2002 to 2010 using observed data from 1956 to 2001, and the results compared against the remaining observed data. During wavelet de-noising of the time series data, mother wavelet functions "db6" and "dmey" are used, with a resolution level of 1. Again, the set dimension, *T*, used in WD-RSPA and single RSPA models, is varied, taking values of 4, 5, and 6.

Tables 4–6 list the observed annual runoff series at Sanmenxia and the resulting values of the different performance measures obtained for the AR(4), ANN-BP, single RSPA, db6-RSPA, and dmey-RSPA models. In this case, the AR(4) model performs best among the ARIMA models and the ANN-RBF is most accurate of the ANN models considered, and so results from the other ARIMA and ANN models are not included in Tables 4–6.

Insert Table 4 here

Insert Table 5 here

Insert Table 6 here

Insert Figure 5 here

It can be seen from Tables 4–6 and Figure 5 that:

(1) The WD-RSPA and AR(4) models provide the best estimates of the changing characteristics of the annual runoff series in comparison with observations, with the ANN-BP model giving the worst forecast in this case. The forecast annual runoff series by the single RSPA model almost invariably has smaller values than the corresponding observations.

(2) Forecasts of annual runoff at Sanmenxia by the WD-RSPA models are influenced by the choice of wavelet de-noising function. For example, considering 2009 and 2010 with *T* set to 6, the RE values associated with the db6-RSPA model are 0.003 and 0.13, whereas those of the dmey-RSPA model are 0.03 and 0.28.

(3) Overall, the 8 statistical measures, P_{10}, P_{20}, Max_{RE}, Min_{RE}, MRE, SD-RE, RMSE, and TIC, show that WD-RSPA usually attains a minimum or maximum value, indicating that WD-RSPA (especially dmey-RSPA with $T = 5$) provides the most accurate, comprehensive forecasts for Sanmenxia over the ten year period under consideration.

(4) In terms of all 9 performance measures, the annual runoff forecasts made by the WD-RSPA and single RSPA models are affected by the value assigned to *T*. For example, RMSE for the db6-RSPA model forecasts alters from 41.56 to 35.61 to 48.36 as T changes from 4 to 5 to 6. Similarly, RMSE for the single RSPA model appears to increase monotonically from 40.09 to 55.94 as T increases from 4 to 6.

4.3 Case 3: Annual precipitation series at Zhengzhou Station, lower Yellow River, China

We now consider annual precipitation series data from Zhengzhou Station in Henan Province, China. Observed precipitation data are available for a 59-year period from 1951 to 2009. We fit the models to data from 1951 to 2000, obtain forecasts from 2001 to 2009, and compare the results against observations. For de-noising, the mother wavelet functions used are "db9" and "rbio3.5". Again, the resolution level is 1, and the set dimension, *T*, varied from 4 to 6 in the WD-RSPA and single RSPA models.

Tables 7–9 compare the relative performances of the AR(1), ANN-BP, single RSPA, db9-RSPA, and rbio3.5-RSPA models in forecasting precipitation series at Zhengzhou for $T = 4$, 5, and 6 . Results from other ARIMA and ANN models are not reproduced in Tables 7–9 because they give poorer forecasts than AR(1) and ANN-RBF respectively.

Insert Table 7 here

Insert Table 8 here

Insert Table 9 here

Insert Figure 6 here

It can be seen from Tables 7–9 and Figure 6 that:

(1) The WD-RSPA and ANN-BP models are best at forecasting the changing characteristics of the annual precipitation series, by comparison with observations at Zhengzhou, whereas the AR(1) model and single RSPA model are much less accurate.

(2) Choice of wavelet de-noising function in WD-RSPA is important. For example, when T = 4, the RE values of db9-RSPA are 0.007 and 0.7, and rbio3.5-RSPA are 0.0043 and 0.12 respectively for 2008 and 2009.

(3) Interpretation of the values of the remaining 8 measures (P_{10}, P_{20}, Max_{RE}, Min_{RE}, MRE, SD-RE, RMSE and TIC) shows that WD-RSPA gives maxima or minima associated with best performance in forecasting annual precipitation at Zhengzhou.

(4) Again, the selected value of set dimension has an impact on the forecasts made by the WD-SPA and RSPA models, as can been seen in the variation in all nine performance measures with T. For example, the RMSE values obtained for the forecasts made by the db9-RSPA model are 1659.06, 1444.62, and 1653.06 for T = 4, 5, and 6, respectively.

4.4 Case 4: Annual precipitation series at Beijing Station, China

Finally, we consider a 60-year annual precipitation record from 1951 to 2011 for Beijing Station, located in the capital city of China. Forecasts are provided for the years from 2002 to 2002. Mother wavelets, "db6" and "dmey" are used for de-noising the raw data. The number of the resolution level is 1. Table 10 lists values of the 9 performance measures, obtained by comparing the forecast and observed annual precipitation time series using AR(3), ANN-RBF, db6-RSPA, and dmey-RSPA models, taking T = 4. Results from the other ARIMA and ANN models are not presented owing to the poorer performance of these models.

Insert Table 10 here

Figure 7 presents the annual precipitation observations and forecasts for Beijing Station from 2002 to 2010, with the forecasts obtained using AR(3), ANN-RBF, db6-RSPA, and dmey-RSPA models, and T = 4.

Insert Figure 7 here

It can be seen from Table 10 and Figure 7 that:

(1) The annual precipitation forecasts obtained for Beijing by WD-RSPA are affected by choice of wavelet de-noising function. For example, the forecasts for 2009 and 2010 by the db6-RSPA model (4971.8 and 5274.1) are much closer to the observed values (4806 and 5225) than those by the dmey-RSPA model (5544.4 and 4616.3).

(2) The WD-RSPA and AR(3) models are better at simulating the changing annual precipitation behavior than the ANN-RBF model.

(3) The values of the P_{10} and P_{20} measures confirm that the WD-RSPA model gives forecasts that are more accurate and comprehensive than those of the other models considered.

(4) Six statistical measures, Max_{RE}, Min_{RE}, MRE, SD-RE, RMSE, and TIC, show that dmey-RSPA model consistently obtains the minimum value, indicating that the dmey-RSPA model is most applicable in this case.

Now we take "db6" as the de-noising function, and investigate the influence of set dimension, used in the established WD-RSPA and single RSPA models, for $T = 4$, 5, and 6.

Table 11 lists the observed annual precipitation forecasts and nine performance metrics for Beijing Station obtained using the single RSPA and db6-RSPA models, setting $T = 4$, 5, and 6. Figure 8 plots the corresponding RE values obtained for both models.

Insert Table 11 here

Insert Figure 8 here

It can be seen from Table 11 and Figure 8 that:

(1) Different annual precipitation time series are forecast using WD-RSPA and the single RSPA models, with RSPA tending to give higher estimates. The results are sensitive to the value of T, with extreme values obtained for either $T = 4$ or 6.

(2) For each T, the values obtained for P_{10} and P_{20} are always larger for db6-RSPA than RSPA, indicating that db6-RSPA is better at forecasting the precipitation in Beijing than the single RSPA model in this case.

(3) For each T, the remaining six measures, Max_{RE}, Min_{RE}, MRE, SD-RE, RMSE, and TIC, have consistently lower values for db6-RSPA than single RSPA, confirming the better forecasting efficacy of db6-RSPA.

(4) Examination of the values of P_{10}, Max_{RE}, SD-RE, RMSE, and TIC, shows that db6-RSPA with $T = 4$ always provides a minimum or maximum value, demonstrating that this tuned model is best in this case.

4.5 Further discussion

(1) By de-noising the raw time series data, a better representation is achieved of the actual characteristics of hydro-meteorological time series. By taking full advantage of wavelet de-noising, WD-RSPA can improve the accuracy of hydro-meteorological time series forecasts. For example, in Case 3 with $T = 5$, MRE obtained using the db9-RSPA model is 0.12, a value less than half that obtained using the single RSPA model. Examining all nine performance metrics in Cases 1 to 4, it can be seen that the de-noised WD-RSPA approach almost always performs better than any of the models without de-noising.

(2) The type of wavelet de-noising function utilized is a key factor influencing the performance of the WD-RSPA approach. In all cases, the forecasts by the WD-RSPA models depended on the choice of wavelet de-noising function. For example, in Case 2 the RMSE values obtained using the db6-RSPA and dmey-RSPA models are 41.56 and 40.86 at $T = 4$, and 35.61 and 34.20 at $T = 5$, respectively. It should be noted that, though the present work has studied selection of appropriate wavelet de-noising function for four specific cases in China, the problem of matching wavelet de-noising functions to generic cases remains a major challenge in practice.

(3) The set dimension, T, is another key factor influencing the performance of the WD-RSPA approach. In particular, WD-RSPA model forecasts differ according to T in all cases. Taking Case 3 as an example, the MRE values for the db9-RSPA model results are 0.15, 0.12, and 0.18, for T = 4, 5 and 6, respectively. Further study is needed on selection of the set dimension.

(4) In all cases, and by all measures, WD-RSPA performs best even when the extreme value is encountered, provided a suitable wavelet function and set dimension are selected. In general, forecasting the extreme values is a major problem. For example, in Case 1 the minimum value of runoff occurs in 2000, over the range of forecast-years considered; from Table 2, RE of the bior2.4-RSPA (T = 5) forecast is 0.13, the least value of all the models considered. The runoff of 2001 is the second lowest value in all the forecast-years considered, and the corresponding RE of the bior2.4-RSPA model (T = 5) is 0.01, again the smallest value of all models. As to the runoff of 2006, the maximum value in all the forecast-years, RE of the bior2.4-RSPA model (T = 4) is 0.10, below that of the single RSPA model (0.16) and the AR(4) model (0.17), but much larger than the minimum of 0.04 from the ANN-BP model.

(5) Observed hydro-meteorological time series are very complicated, being influenced by factors such as atmospheric circulation, geographical features, land surface conditions, and human activities. The accuracy of hydro-meteorological forecasts could be restricted if a single method is solely used; it is prudent to combine two or more methods to improve the accuracy of such forecasts.

5 Conclusion

This paper has dealt with two critical issues that arise in forecasting meteorological and hydrologic time series, namely: noise contamination of input data; and the over-complexity of calculation approaches at present. To overcome these drawbacks, a hybrid WD-RSPA approach is proposed to take full advantage of both Wavelet De-noising and Rank-Set Pair Analysis in improving the accuracy and ease of hydro-meteorological forecasting. Analyses of annual runoff and precipitation time series from four representative stations in China are used to examine the effectiveness of the WD-RSPA approach by comparison with other standard techniques, including the conventional Auto Regressive Integrated Moving Average (ARIMA) method, Artificial Neural Networks (ANNs), and single Rank-Set Pair Analysis (RSPA). Using nine statistical measures to evaluate model performance it is found that

WD-RSPA approach is accurate, feasible and effective, and almost invariably the best amongst the various methods compared, even when the extreme value occurs. The improved accuracy of WD-RSPA should make it a useful technique in the study of hydro-meteorological (and other) time series.

References

Chanerley, A.A., and N.A. Alexander (2007), Correcting data from an unknown accelerometer using recursive least squares and wavelet de-noising. Computers & Structures, 85: 1679-1692.

Chen, Q.W., J. Li, R.N. Li, W.D. Wei, and L.M. Wang (2014), River basin water resource compensation characteristics by set pair analysis: the Dongjiang example. Frontiers of Earth Science, 8(1): 64-69. DOI: 10.1007/s11707-013-0389-4

Chou, C.M. (2014), Application of Set Pair Analysis-based similarity forecast model and wavelet denoising for runoff forecasting. Water, 6(4): 912-928. DOI: 10.3390/w6040912

Donoho, D.L. (1995), De-noising by soft-thresholding. IEEE Transactions on Information Theory, 41: 613-627. DOI: 10.1109/18.382009.

Du, C.Y., J.J. Yu, H.P. Zhong, and D.D. Wang (2015), Operating mechanism and set pair analysis model of a sustainable water resources system. Frontiers of Environmental Science & Engineering, 9(2): 288-297. DOI: 10.1007/s11783-014-0642-4

Elshorbagy, A., S.P. Simonovic, and U.S. Panu (2002), Noise reduction in chaotic hydrologic time series: Facts and doubts. Journal of Hydrology, 256, 147–165.

Feng, L.H., G.S. Sang, and W.H. Hong (2014), Statistical prediction of changes in water resources trends based on set pair analysis. Water Resources Management, 28(6): 1703-1711. DOI: 10.1007/s11269-014-0581-7

Ghosh, S., and P.P. Mujumdar (2008), Statistical downscaling of GCM simulations to streamflow using relevance vector machine. Advances in Water Resources, 31(1): 132-146.

Guo, E.L., J.Q. Zhang, X.H. Ren, Q. Zhang, and Z.Y. Sun (2014), Integrated risk assessment of flood disaster based on improved set pair analysis and the variable fuzzy set theory in central Liaoning Province, China. Natural Hazards, 74(2): 947-965. DOI: 10.1007/s11069-014-1238-9

He, Z., X. Wen, H. Liu, and J. Du (2014), A comparative study of artificial neural network, adaptive neuro fuzzy inference system and support vector machine for forecasting river flow in the semiarid mountain region. Journal of Hydrology, 509, 379-386.

Jiang, X., S.G. Xu, Y.Y. Liu, and X.D. Wang (2015), River ecosystem assessment and application in ecological restorations: A mathematical approach based on evaluating its structure and function. Ecological Engineering, 76: 151-157. DOI: 10.1016/j.ecoleng.2014.04.027

Jin, J.L., J. Fu, Y.M.Wei, S.M. Jiang, Y.L. Zhou, L. Liu, Y.Z. Wang, and C.G. Wu (2015), Integrated risk assessment method of waterlog disaster in Huaihe River Basin of China. Natural Hazards, 75(s2): S155-S178. DOI: 10.1007/s11069-014-1307-0

Jin, J.L., Y.M. Wei, L.L. Zou, L. Liu, W.W. Zhang, and Y.L. Zhou (2012), Forewarning of sustainable utilization of regional water resources: a model based on BP neural network and set pair analysis. Natural Hazards, 62(1): 115-127. DOI: 10.1007/s11069-011-0037-9

Kavvas, M., S. Kure, Z. Chen, N. Ohara, and S. Jang (2013), WEHY-HCM for modeling interactive atmospheric-hydrologic processes at watershed scale. I: model description. Journal of Hydrologic Engineering, 18(10): 1262-1271. DOI: 10.1061/(ASCE)HE.1943-5584.0000724.

Labat, D. (2005), Recent advances in wavelet analyses: Part I. A review of concepts. Journal of Hydrology, 314: 275-288. DOI: 10.1016/j.jhydrol.2005.04.003.

Leung, L.Y.R., T. Ringler, W.D. Collins, M. Taylor, and M. Ashfaq (2013), A hierarchical evaluation of regional climate simulations. Eos, 94(34):297-298. DOI: 10.1002/2013EO340001

Mehran, A., O. Mazdiyasni, and A. AghaKouchak (2015), A hybrid framework for assessing socioeconomic drought: Linking climate variability, local resilience, and demand. Journal of Geophysical Research. D. (Atmospheres), 120, 7520–7533. doi: 10.1002/2015JD023147.

Minville, M., F. Brissette, and R. Leconte (2008), Uncertainty of the impact of climate change on the hydrology of a nordic watershed. Journal of Hydrology, 358, 70–83, doi:10.1016/j.jhydrol.2008.05.033.

Nalley D., J. Adamowski, and B. Khalil (2012), Using discrete wavelet transforms to analyze trends in streamflow and precipitation in Quebec and Ontario (1954–2008). Journal of Hydrology, 475, 204-228.

Nourani, V., A.H. Baghanam, J. Adamowski, and M. Gebremichael (2013), Using self-organizing maps and wavelet transforms for space–time pre-processing of satellite precipitation and runoff data in neural network based rainfall–runoff modeling. Journal of Hydrology, 476, 228-243.

Ou, Y., Q. Zhang, W.S. Wang, and X.B. Song (2009), Annual runoff forecasting model based on Rank-Set Pair Analysis. Yangtze River, 40(3): 63-65.

Pulido-Calvo, I., and M.M. Portela (2007), Application of neural approaches to one-step daily flow forecasting in Portuguese watersheds. Journal of Hydrology, 332, 1–15.

Qian Y., and L.R. Leung (2007), A Long-term Regional Simulation and Observations of the Hydroclimate in China. Journal of Geophysical Research. D. (Atmospheres), 112, D14, doi:10.1029/2006JD008134

Schaefli B., D. Maraun, and M. Holschneider (2007), What drives high flow events in the Swiss Alps? Recent developments in wavelet spectral analysis and their application to hydrology. Advances in Water Resources, 30: 2511-2525.

Shoaib, M., A.Y. Shamseldin, B.W. Melville, and M.M. Khan (2015). Runoff forecasting using hybrid Wavelet Gene Expression Programming (WGEP) approach. Journal of Hydrology, 527, 326-344.

Sivakumar, B., K.K. Phoon, S.Y. Liong, and C.Y. Liaw (1999), A systematic approach to noise reduction in chaotic hydrological time series. Journal of Hydrology, 219: 103-135. DOI:

10.1016/s0022-1694(99)00051-7.

Stevenson, J. A., X. Sun, and N. C. Mitchell (2010), Despeckling SRTM and other topographic data with a denoising algorithm, Geomorphology, 114, 238–252

Su, H.Z., P. Qin, and Z.H. Qin (2013), A method for evaluating sea dike safety. Water Resources Management, 27(15): 5157-5170. DOI: 10.1007/s11269-013-0459-0

Su, M.R., Z.F. Yang, B. Chen, and S. Ulgiati (2009), Urban ecosystem health assessment based on emergy and set pair analysis-A comparative study of typical Chinese cities. Ecological Modelling, 220(18): 2341-2348. DOI: 10.1016/j.ecolmodel.2009.06.010

Sun, A.Y., D. Wang, and X. Xu (2014). Monthly streamflow forecasting using Gaussian process regression. Journal of Hydrology, 511, 72-81.

Valipour, M., M. E. Banihabib, and S.M.R. Behbahani (2013). 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.

Wang, D., H. Ding, V.P. Singh, X.S. Shang, D.F. Liu, Y.K. Wang, X.K. Zeng, J.C. Wu, L.C. Wang, and X.Q. Zou (2015), A hybrid wavelet analysis–cloud model data-extending approach formeteorologic and hydrologic time series. Journal of Geophysical Research. D. (Atmospheres), 120, 4057–4071

Wang, D., V.P. Singh, X.S. Shang, H. Ding, J.C. Wu, L.C. Wang, X.Q. Zou, Y.F. Chen, X. Chen, S.C. Wang, and Z.L. Wang (2014), Sample entropy-based adaptive wavelet de-noising approach for meteorologic and hydrologic time series. Journal of Geophysical Research. D. (Atmospheres), 119, 8726–8740

Wang, T., J.S. Chen, T. Wang, and S. Wang (2015), Entropy weight-set pair analysis based on tracer techniques for dam leakage investigation. Natural Hazards, 76(2): 747-767. DOI: 10.1007/s11069-014-1515-7

Wang, W.S., and Y.Q. Li (2012), Hazard degree assessment of landslide using set pair analysis method. Natural Hazards, 60(2): 367-379. DOI: 10.1007/s11069-011-0017-0

Wang, W.S., J.L. Jin, J. Ding, and Y.Q. Li (2009), A new approach to water resources system assessment - set pair analysis method. Science in China Series E: Technological Sciences, 52(10): 3017-3023. DOI: 10.1007/s11431-009-0099-z

Wang, W.S., S.X. Hu, Y.Q. Li, and S.Y. Cao (2013), How to select a reference basin in the ungauged regions. Journal of Hydrologic Engineering, 18(8): 941-947. DOI: 10.1061/(ASCE)HE.1943-5584.0000680

Whitworth, K. L., D. S. Baldwin, and J. L. Kerr (2012), Drought, floods and water quality: drivers of a severe hypoxic blackwater event in a major river system (the southern Murray–Darling Basin, Australia). Journal of Hydrology, 450, 190-198.

Wu, C.L., and K.W. Chau (2011). Rainfall–runoff modeling using artificial neural network coupled with

singular spectrum analysis. Journal of Hydrology, 399(3), 394-409.

Xu, F., Y.W. Zhao, Z.F. Yang, and Y. Zhang (2011), Multi-scale evaluation of river health in Liao River Basin, China. Frontiers of Environmental Science and Engineering in China, 5(2): 227-235. DOI: 10.1007/s11783-010-0219-9

Yang, F.G., Y. Liang, V.P. Singh, W.S. Wang, X.Q. Zhou, X.N. Liu, S.Y. Cao, E. Huang, and Y.H. Wu (2014a), Debris flow hazard assessment using set pair analysis models: Take Beichuan county as an example. Journal of Mountain Science, 11(4): 1015-1022. DOI: 10.1007/s11629-013-2495-x

Yang, S.S., X.H. Yang, R. Jiang, and Y.C. Zhang (2012), New optimal weight combination model for forecasting precipitation. Mathematical Problems in Engineering, 376010. DOI: 10.1155/2012/376010

Yang, X.H., J. He, C.L. Di, and J.Q. Li (2014b), Vulnerability of assessing water resources by the improved set pair analysis. Thermal Science, 18(5): 1531-1535. DOI: 10.2298/TSCI1405531Y

Zhang, J.P., B.T. Guo, and Z.H. Ding (2013), Research on the drought index of irrigation district with multi-time scales. Environmental Monitoring and Assessment, 185(10): 8749-8757. DOI: 10.1007/s10661-013-3209-0

Zhang, J.Y., and L.C.Wang (2015), Assessment of water resource security in Chongqing City of China: What has been done and what remains to be done? Natural Hazards, 75(3): 2751-2772. DOI: 10.1007/s11069-014-1460-5

Zhang, Q., B.D. Wang, B. He, Y. Peng, and M.L. Ren (2011), Singular spectrum analysis and ARIMA hybrid model for annual runoff forecasting. Water Resources Management, 25, 2683–2703.

Zhang, Y., X.H. Yang, L. Zhang, W.Y. Ma, and L.X. Qiao, 2013. Set pair analysis based on phase space reconstruction model and its application in forecasting extreme temperature. Mathematical Problems in Engineering, 516150.DOI: 10.1155/2013/516150

Zhao, J., Z.P. Huang, J.L. Jin, B.H. Lu, X.M. Zhang, and Y.Q. Chen (2013), Risk assessment of regional water resources and forewarning model at different time scales. Journal of Hydrologic Engineering, 18(9): 1114-1121. DOI: 10.1061/(ASCE)HE.1943-5584.0000490

Zhao, K.Q. (2000), Set Pair Analysis and its Preliminary Applications. Hangzhou: Science and Technology Press of Zhejiang.

Zou, Q., J.Z. Zhou, C. Zhou, L.X. Song, J. Guo (2013), Comprehensive flood risk assessment based on set pair analysis-variable fuzzy sets model and fuzzy AHP. Stochastic Environmental Research and Risk Assessment, 27(2): 525-546. DOI: 10.1007/s00477-012-0598-5

List of Tables

List of Figures

Table 1 Set pairs composed of elements from the time series

Sets	Elements in sets				Subsequent values
B_1	x_1	x_2	…	x_T	x_{T+1}
B_2	x_2	x_3	…	x_{T+1}	x_{T+2}
…	…	…	…	…	…
B_i	x_i	x_{i+1}	…	x_{T+i-1}	x_{T+i}
…	…	…	…	…	…
B_{n-T}	x_{n-T}	x_{n-T+1}	…	x_{n-1}	x_n
Y	x_{n-T+1}	x_{n-T+2}	…	x_n	x_{n+1}

Table 2 Annual runoff observations and forecasts, with nine performance metrics for AR(4), ANN-BP, coif3-RSPA, and bior2.4-RSPA models (T = 5): Case 1, Huayuankou Station, lowerYellow River, China.

Year	MV	AR(4) FV/RE	ANN-BP FV/RE	coif3-RSPA FV/RE	bior2.4-RSPA FV/RE
1998	217.9	168.51/0.22	172.83/0.21	206.64/0.05	230.68/0.06
1999	208.7	219.84/0.05	229.02/0.10	200.46/0.04	203.51/0.02
2000	165.3	190.43/0.15	188.06/0.14	205.59/0.24	186.65/0.13
2001	165.5	194.73/0.18	138.82/0.16	160.31/0.03	163.28/0.01
2002	195.6	218.84/0.12	191.10/0.02	154.96/0.21	155.81/0.20
2003	272.7	216.25/0.21	236.97/0.13	201.25/0.26	229.84/0.16
2004	240.5	218.50/0.09	273.46/0.14	238.36/0.009	234.49/0.03
2005	257.0	230.29/0.10	316.23/0.23	244.49/0.05	243.73/0.05
2006	281.1	234.30/0.17	269.42/0.04	247.85/0.11	247.42/0.12
2007	269.7	237.72/0.12	269.11/0.002	295.15/0.09	293.90/0.09
P_{10}		0.20	0.40	0.60	0.60
P_{20}		0.80	0.80	0.70	0.90
Max_{RE}		0.22	0.23	0.26	0.20
Min_{RE}		0.05	0.002	0.009	0.01
MRE		0.14	0.12	0.11	0.09
SD-RE		0.16	0.23	0.28	0.19
RMSE		34.91	31.15	32.43	24.51
TIC		0.0784	0.0669	0.0720	0.0540

Note:

1. MV: measured value, $10^8 m^3$; FV: forecast value, $10^8 m^3$.

2. (1) AR(4): auto-regression method; (2) ANN-BP: artificial neural network – error back propagation method; (3) coif3-RSPA: WD-RSPA method using de-noising wavelet function-coif3; (4) bior2.4-RSPA: WD-RSPA method using de-noising wavelet function-bior2.4.

3. (1) RE: relative error; (2) P_{10}: percentage pass rate for relative error < 10%; (3) P_{20}: percentage pass rate for relative error < 20%; (4) Max_{RE}: maximum relative error; (5) Min_{RE}: minimum relative error; (6) MRE: mean relative error; (7) SD-RE: standard deviation of relative error; (8) RMSE: root mean square error; (9) TIC: Thiel inequality coefficient.

Table 3 Annual runoff observations, with nine performance metrics for single RSPA and bior2.4-RSPA models (T = 4, 5, 6): Case 1, Huayuankou Station, lower Yellow River,China.

Year	MV	RE (T=4)		RE (T=5)		RE (T=6)	
		RSPA	bior2.4-RSPA	RSPA	bior2.4-RSPA	RSPA	bior2.4-RSPA
1998	217.9	0.12	0.02	0.22	0.06	0.63	0.08
1999	208.7	0.09	0.087	0.06	0.02	0.14	0.03
2000	165.3	0.08	0.17	0.16	0.13	0.06	0.25
2001	165.5	0.07	0.04	0.10	0.01	0.07	0.05
2002	195.6	0.09	0.22	0.17	0.20	0.10	0.24
2003	272.7	0.36	0.20	0.36	0.16	0.56	0.21
2004	240.5	0.29	0.05	0.26	0.03	0.20	0.06
2005	257.0	0.03	0.02	0.30	0.05	0.18	0.16
2006	281.1	0.16	0.10	0.23	0.12	0.33	0.27
2007	269.7	0.07	0.11	0.04	0.09	0.18	0.12
P_{10}		0.60	0.50	0.20	0.60	0.20	0.40
P_{20}		0.80	0.80	0.50	0.90	0.60	0.60
Max_{RE}		0.36	0.22	0.36	0.20	0.63	0.27
Min_{RE}		0.03	0.02	0.04	0.01	0.06	0.03
MRE		0.14	0.10	0.19	0.09	0.24	0.15
SD-RE		0.32	0.22	0.31	0.19	0.60	0.28
RMSE		42.91	28.24	53.58	24.51	76.78	40.51
TIC		0.0977	0.0624	0.1191	0.0540	0.1734	0.0906

Note:

1. MV: measured value, $10^8 m^3$.

2. bior2.4-RSPA: WD-RSPA method using de-noising wavelet function-bior2.4.

3. (1) RE: relative error; (2) P_{10}: percentage pass rate for relative error < 10%; (3) P_{20}: percentage pass rate for relative error < 20%; (4) Max_{RE}: maximum relative error; (5) Min_{RE}: minimum relative error; (6) MRE: mean relative error; (7) SD-RE: standard deviation of relative error; (8) RMSE: root mean square error; (9) TIC: Thiel inequality coefficient.

Table 4 Annual runoff observations, with nine performance metrics for RSPA, AR(4), ANN-BP, db6-RSPA, and dmey-RSPA models (T = 4): Case 2, Sanmenxia Station, lowerYellow River, China

Year	MV	RSPA~RE	AR(4)~RE	ANN-BP~RE	db6-RSPA~RE	dmey-RSPA~RE
2002	152.1	0.16	0.11	0.08	0.03	0.08
2003	236.1	0.34	0.34	0.29	0.40	0.40
2004	168.7	0.003	0.22	0.16	0.06	0.07
2005	211.4	0.18	0.11	0.07	0.09	0.24
2006	212.0	0.10	0.07	0.28	0.08	0.15
2007	242.7	0.28	0.17	0.40	0.08	0.07
2008	210.8	0.10	0.15	0.24	0.13	0.03
2009	219.7	0.08	0.02	0.001	0.01	0.02
2010	250.1	0.01	0.07	0.88	0.28	0.16
	P_{10}	0.33	0.33	0.33	0.67	0.56
	P_{20}	0.78	0.78	0.44	0.78	0.78
	Max_{RE}	0.34	0.34	0.88	0.40	0.40
	Min_{RE}	0.003	0.02	0.001	0.01	0.02
	MRE	0.14	0.14	0.26	0.13	0.14
	SD-RE	0.33	0.28	0.74	0.36	0.35
	RMSE	40.09	36.75	87.97	41.56	40.86
	TIC	0.0997	0.0874	0.1948	0.1019	0.1010

Note:

1. MV: measured value, $10^8 m^3$.

2. (1) AR(4): auto-regression method; (2) ANN-BP: artificial neural network – error back propagation method; (3) db6-RSPA: WD-RSPA method using de-noising wavelet function-db6; (4) dmey-RSPA: WD-RSPA method using de-noising wavelet function-dmey.

3. (1) RE: relative error; (2) P_{10}: percentage pass rate for relative error < 10%; (3) P_{20}: percentage pass rate for relative error < 20%; (4) Max_{RE}: maximum relative error; (5) Min_{RE}: minimum relative error; (6) MRE: mean relative error; (7) SD-RE: standard deviation of relative error; (8) RMSE: root mean square error; (9) TIC: Thiel inequality coefficient.

Table 5 Annual runoff observations, with nine performance metrics for RSPA, AR(4), ANN-BP, db6-RSPA, and dmey-RSPA models (T = 5): Case 2, Sanmenxia Station, lowerYellow River, China

Year	MV	RSPA~RE	AR(4)~RE	ANN-BP~RE	db6-RSPA~RE	dmey-RSPA~RE
2002	152.1	0.08	0.11	0.08	0.03	0.03
2003	236.1	0.47	0.34	0.29	0.36	0.34
2004	168.7	0.10	0.22	0.16	0.22	0.14
2005	211.4	0.08	0.11	0.07	0.05	0.05
2006	212.0	0.22	0.07	0.28	0.07	0.23
2007	242.7	0.23	0.17	0.40	0.07	0.04
2008	210.8	0.29	0.15	0.24	0.15	0.07
2009	219.7	0.13	0.02	0.001	0.06	0.03
2010	250.1	0.07	0.07	0.88	0.12	0.11
P_{10}		0.44	0.33	0.33	0.56	0.56
P_{20}		0.56	0.78	0.44	0.78	0.78
Max_{RE}		0.47	0.34	0.88	0.36	0.34
Min_{RE}		0.07	0.02	0.001	0.03	0.03
MRE		0.19	0.14	0.27	0.13	0.11
SD-RE		0.38	0.28	0.74	0.30	0.30
RMSE		50.70	36.75	87.97	35.61	34.20
TIC		0.1290	0.0874	0.1948	0.0854	0.0836

Note:

1. MV: measured value, $10^8 m^3$.

2. (1) AR(4): auto-regression method; (2) ANN-BP: artificial neural network – error back propagation method; (3) db6-RSPA: WD-RSPA method using de-noising wavelet function-db6; (4) dmey-RSPA: WD-RSPA method using de-noising wavelet function-dmey.

3. (1) RE: relative error; (2) P_{10}: percentage pass rate for relative error < 10%; (3) P_{20}: percentage pass rate for relative error < 20%; (4) Max_{RE}: maximum relative error; (5) Min_{RE}: minimum relative error; (6) MRE: mean relative error; (7) SD-RE: standard deviation of relative error; (8) RMSE: root mean square error; (9) TIC: Thiel inequality coefficient.

Table 6 Annual runoff observations, with nine performance metrics for RSPA, AR(4), ANN-BP, db6-RSPA, and dmey-RSPA models ($T = 6$): Case 2, Sanmenxia Station, lowerYellow River, China

Year	MV	RSPA~RE	AR(4)~RE	ANN-BP~RE	db6-RSPA~RE	dmey-RSPA~RE
2002	152.1	0.07	0.11	0.08	0.01	0.15
2003	236.1	0.44	0.34	0.29	0.39	0.45
2004	168.7	0.02	0.22	0.16	0.57	0.18
2005	211.4	0.007	0.11	0.07	0.07	0.09
2006	212.0	0.26	0.07	0.28	0.12	0.16
2007	242.7	0.28	0.17	0.40	0.14	0.17
2008	210.8	0.30	0.15	0.24	0.07	0.005
2009	219.7	0.17	0.02	0.001	0.003	0.03
2010	250.1	0.31	0.07	0.88	0.13	0.28
P_{10}		0.33	0.33	0.33	0.44	0.33
P_{20}		0.44	0.78	0.44	0.78	0.78
Max_{RE}		0.44	0.34	0.88	0.57	0.45
Min_{RE}		0.007	0.02	0.001	0.003	0.005
MRE		0.20	0.14	0.27	0.17	0.17
SD-RE		0.41	0.28	0.74	0.54	0.38
RMSE		55.94	36.75	87.97	48.36	48.48
TIC		0.1342	0.0874	0.1948	0.1156	0.1213

Note:

1. MV: measured value, $10^8 m^3$.

2. (1) AR(4): auto-regression method; (2) ANN-BP: artificial neural network – error back propagation method; (3) db6-RSPA: WD-RSPA method using de-noising wavelet function-db6; (4) dmey-RSPA: WD-RSPA method using de-noising wavelet function-dmey.

3. (1) RE: relative error; (2) P_{10}: percentage pass rate for relative error < 10%; (3) P_{20}: percentage pass rate for relative error < 20%; (4) Max_{RE}: maximum relative error; (5) Min_{RE}: minimum relative error; (6) MRE: mean relative error; (7) SD-RE: standard deviation of relative error; (8) RMSE: root mean square error; (9) TIC: Thiel inequality coefficient.

Table 7 Annual precipitation observations, with nine performance metrics for RSPA, AR(1), ANN-BP, db9-RSPA, and rbio3.5-RSPA models ($T = 4$): Case 3, Zhengzhou Station, lowerYellow River, China

Year	MV	RSPA~RE	AR(1)~RE	ANN-BP~RE	db9-RSPA~RE	rbio3.5-RSPA~RE
2001	4018	0.47	0.54	0.47	0.50	0.42
2002	5993	0.03	0.08	0.16	0.16	0.09
2003	9539	0.44	0.35	0.07	0.45	0.33
2004	7674	0.06	0.16	0.16	0.0044	0.07
2005	7288	0.02	0.14	0.11	0.08	0.10
2006	6926	0.18	0.07	0.05	0.005	0.06
2007	5964	0.24	0.05	0.08	0.08	0.18
2008	6582	0.02	0.02	0.37	0.007	0.0043
2009	7625	0.12	0.18	0.17	0.07	0.12
P_{10}		0.44	0.44	0.33	0.67	0.44
P_{20}		0.67	0.78	0.78	0.78	0.78
Max_{RE}		0.47	0.54	0.47	0.50	0.42
Min_{RE}		0.02	0.02	0.05	0.0044	0.0043
MRE		0.18	0.18	0.18	0.15	0.15
SD-RE		0.50	0.48	0.40	0.54	0.39
RMSE		1688.98	1521.34	1287.35	1659.06	1350.12
TIC		0.1227	0.1142	0.0898	0.1221	0.0983

Note:

1. MV: measured value, 0.1mm.

2. (1) AR(1): auto-regression method; (2) ANN-BP: artificial neural network – error back propagation method; (3) db9-RSPA: WD-RSPA method using de-noising wavelet function-db9; (4) rbio3.5-RSPA: WD-RSPA method using de-noising wavelet function- rbio3.5.

3. (1) RE: relative error; (2) P_{10}: percentage pass rate for relative error < 10%; (3) P_{20}: percentage pass rate for relative error < 20%; (4) Max_{RE}: maximum relative error; (5) Min_{RE}: minimum relative error; (6) MRE: mean relative error; (7) SD-RE: standard deviation of relative error; (8) RMSE: root mean square error; (9) TIC: Thiel inequality coefficient.

Table 8 Annual precipitation observations, with nine performance metrics for RSPA, AR(1), ANN-BP, db9-RSPA, and rbio3.5-RSPA models (T = 5): Case 3, Zhengzhou Station, lowerYellow River, China

Year	MV	RSPA~RE	AR(1)~RE	ANN-BP~RE	db9-RSPA~RE	rbio3.5-RSPA~RE
2001	4018	0.41	0.54	0.47	0.26	0.39
2002	5993	0.24	0.08	0.16	0.13	0.18
2003	9539	0.49	0.35	0.07	0.42	0.36
2004	7674	0.23	0.16	0.16	0.01	0.04
2005	7288	0.19	0.14	0.11	0.01	0.03
2006	6926	0.28	0.07	0.05	0.02	0.09
2007	5964	0.36	0.05	0.08	0.15	0.06
2008	6582	0.87	0.02	0.37	0.01	0.12
2009	7625	0.004	0.18	0.17	0.05	0.16
P_{10}		0.22	0.44	0.33	0.56	0.44
P_{20}		0.33	0.78	0.78	0.78	0.78
Max_{RE}		0.49	0.54	0.47	0.42	0.39
Min_{RE}		0.004	0.02	0.05	0.01	0.03
MRE		0.25	0.18	0.18	0.12	0.16
SD-RE		0.47	0.48	0.40	0.40	0.37
RMSE		2114.96	1521.34	1287.35	1444.62	1425.49
TIC		0.1611	0.1142	0.0898	0.1061	0.1060

Note:

1. MV: measured value, 0.1mm.

2. (1) AR(1): auto-regression method; (2) ANN-BP: artificial neural network – error back propagation method; (3) db9-RSPA: WD-RSPA method using de-noising wavelet function-db9; (4) rbio3.5-RSPA: WD-RSPA method using de-noising wavelet function- rbio3.5.

3. (1) RE: relative error; (2) P_{10}: percentage pass rate for relative error < 10%; (3) P_{20}: percentage pass rate for relative error < 20%; (4) Max_{RE}: maximum relative error; (5) Min_{RE}: minimum relative error; (6) MRE: mean relative error; (7) SD-RE: standard deviation of relative error; (8) RMSE: root mean square error; (9) TIC: Thiel inequality coefficient.

Table 9 Annual precipitation observations, with nine performance metrics for RSPA, AR(1), ANN-BP, db9-RSPA, and rbio3.5-RSPA models ($T = 6$): Case 3, Zhengzhou Station, lowerYellow River, China.

Year	MV	RSPA~RE	AR(1)~RE	ANN-BP~RE	db9-RSPA~RE	rbio3.5-RSPA~RE
2001	4018	0.59	0.54	0.47	0.50	0.47
2002	5993	0.08	0.08	0.16	0.13	0.09
2003	9539	0.60	0.35	0.07	0.43	0.33
2004	7674	0.20	0.16	0.16	0.05	0.04
2005	7288	0.30	0.14	0.11	0.05	0.01
2006	6926	0.47	0.07	0.05	0.08	0.02
2007	5964	0.33	0.05	0.08	0.14	0.39
2008	6582	0.0095	0.02	0.37	0.09	0.10
2009	7625	0.05	0.18	0.17	0.15	0.07
P_{10}		0.22	0.44	0.33	0.44	0.67
P_{20}		0.22	0.78	0.78	0.78	0.67
Max_{RE}		0.60	0.54	0.47	0.50	0.47
Min_{RE}		0.0095	0.02	0.05	0.05	0.01
MRE		0.29	0.18	0.18	0.18	0.17
SD-RE		0.63	0.48	0.40	0.47	0.50
RMSE		2580.80	1521.34	1287.35	1653.06	1483.93
TIC		0.1972	0.1142	0.0898	0.1212	0.1066

Note:

1. MV: measured value, 0.1mm.

2. (1) AR(1): auto-regression method; (2) ANN-BP: artificial neural network – error back propagation method; (3) db9-RSPA: WD-RSPA method using de-noising wavelet function-db9; (4) rbio3.5-RSPA: WD-RSPA method using de-noising wavelet function- rbio3.5.

3. (1) RE: relative error; (2) P_{10}: percentage pass rate for relative error < 10%; (3) P_{20}: percentage pass rate for relative error < 20%; (4) Max_{RE}: maximum relative error; (5) Min_{RE}: minimum relative error; (6) MRE: mean relative error; (7) SD-RE: standard deviation of relative error; (8) RMSE: root mean square error; (9) TIC: Thiel inequality coefficient.

Table 10 Annual precipitation observations and forecasts, with nine performance metrics for AR(3), ANN-RBF, db6-RSPA, and dmey-RSPA models ($T = 4$): Case 4, Beijing Station, China

Year	MV	AR(3) FV/RE	ANN-RBF FV/RE	db6-RSPA FV/RE	dmey-RSPA FV/RE
2002	3704	3375.20/0.09	4614.30/0.25	3578.80/0.03	3713.40/0.003
2003	4449	3747.30/0.16	5791.70/0.30	4251.20/0.04	4073.70/0.08
2004	4835	4053.90/0.16	4443.70/0.08	4427.70/0.08	4481.20/0.07
2005	4107	4480.20/0.09	5170.80/0.26	4504.40/0.10	4659.90/0.13
2006	3180	4457.70/0.40	4468.50/0.41	4248.80/0.34	3436.50/0.08
2007	4839	3998.30/0.17	4485.50/0.07	3859.40/0.20	3420.50/0.29
2008	6263	4320.90/0.31	4397.20/0.30	4429.30/0.29	5067.80/0.19
2009	4806	5048.10/0.05	4422.10/0.08	4971.80/0.03	5544.40/0.15
2010	5225	5194.70/0.006	4495.00/0.14	5274.10/0.009	4616.30/0.12
P_{10}		0.44	0.33	0.67	0.44
P_{20}		0.78	0.44	0.67	0.89
Max_{RE}		0.40	0.41	0.34	0.29
Min_{RE}		0.006	0.07	0.009	0.003
MRE		0.16	0.21	0.1261	0.1255
SD-RE		0.36	0.34	0.34	0.23
RMSE		914.11	1046.64	807.80	744.90
TIC		0.1015	0.1114	0.0888	0.0822

Note:

1. MV: measured value, 0.1mm.

2. (1) AR(3): auto-regression method; (2) ANN-RBF: artificial neural network – Radial Basis Function method; (3) db6-RSPA: WD-RSPA method using de-noising wavelet function-db6; (4) dmey-RSPA: WD-RSPA method using de-noising wavelet function- dmey.

3. (1) RE: relative error; (2) P_{10}: percentage pass rate for relative error < 10%; (3) P_{20}: percentage pass rate for relative error < 20%; (4) Max_{RE}: maximum relative error; (5) Min_{RE}: minimum relative error; (6) MRE: mean relative error; (7) SD-RE: standard deviation of relative error; (8) RMSE: root mean square error; (9) TIC: Thiel inequality coefficient.

Table 11 Annual precipitation observations, with nine performance metrics for for single RSPA and db6-RSPA models (T = 4, 5, 6): Case 4, Beijing Station, China

Year	MV	RE (T=4)		RE (T=5)		RE (T=6)	
		RSPA	db6-RSPA	RSPA	db6-RSPA	RSPA	db6-RSPA
2002	3704	0.04	0.03	0.03	0.04	0.14	0.01
2003	4449	0.23	0.04	0.01	0.007	0.02	0.15
2004	4835	0.34	0.08	0.24	0.10	0.27	0.09
2005	4107	0.03	0.10	0.20	0.13	0.26	0.01
2006	3180	0.47	0.34	0.49	0.34	0.51	0.09
2007	4839	0.14	0.20	0.11	0.14	0.14	0.24
2008	6263	0.30	0.29	0.32	0.37	0.43	0.42
2009	4806	0.10	0.03	0.16	0.10	0.20	0.04
2010	5225	0.03	0.009	0.14	0.02	0.10	0.03
	P_{10}	0.44	0.67	0.22	0.44	0.22	0.67
	P_{20}	0.56	0.67	0.67	0.78	0.44	0.78
	Max_{RE}	0.47	0.34	0.49	0.37	0.51	0.42
	Min_{RE}	0.03	0.009	0.01	0.007	0.02	0.01
	MRE	0.19	0.13	0.19	0.14	0.23	0.12
	SD-RE	0.44	0.34	0.42	0.38	0.44	0.38
	RMSE	1065.76	807.80	1053.19	940.50	1271.21	1007.37
	TIC	0.1206	0.0888	0.1195	0.1030	0.1472	0.1150

Note:

1. MV: measured value, 0.1mm.

2. db6-RSPA: WD-RSPA method using de-noising wavelet function-db6.

3. (1) RE: relative error; (2) P_{10}: percentage pass rate for relative error < 10%; (3) P_{20}: percentage pass rate for relative error < 20%; (4) Max_{RE}: maximum relative error; (5) Min_{RE}: minimum relative error; (6) MRE: mean relative error; (7) SD-RE: standard deviation of relative error; (8) RMSE: root mean square error; (9) TIC: Thiel inequality coefficient.

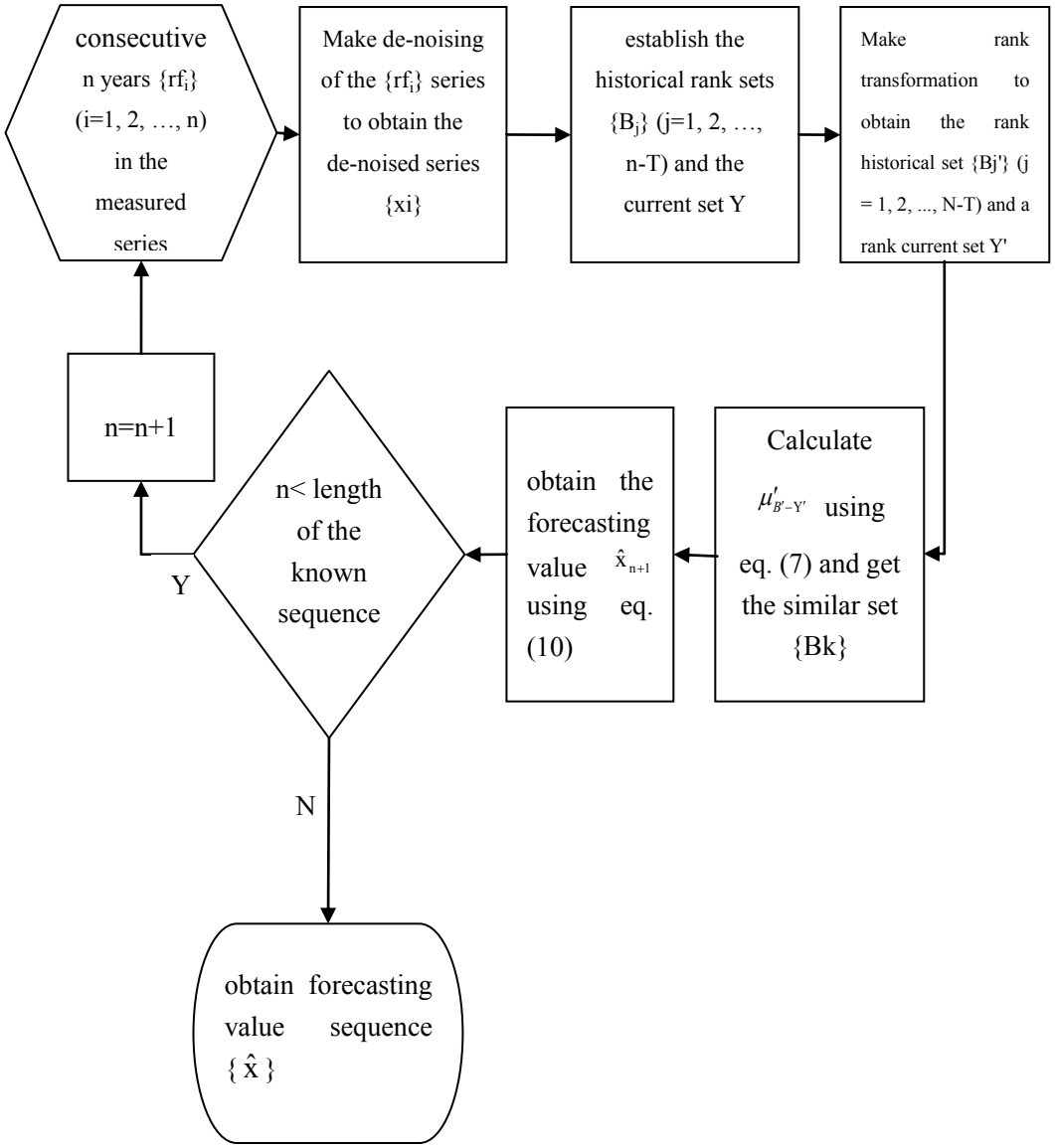


Figure 1 Flow chart outlining WD-RSPA (Wavelet De-noising and Rank-Set Pair Analysis) procedure for forecasting hydro-meteorological data series

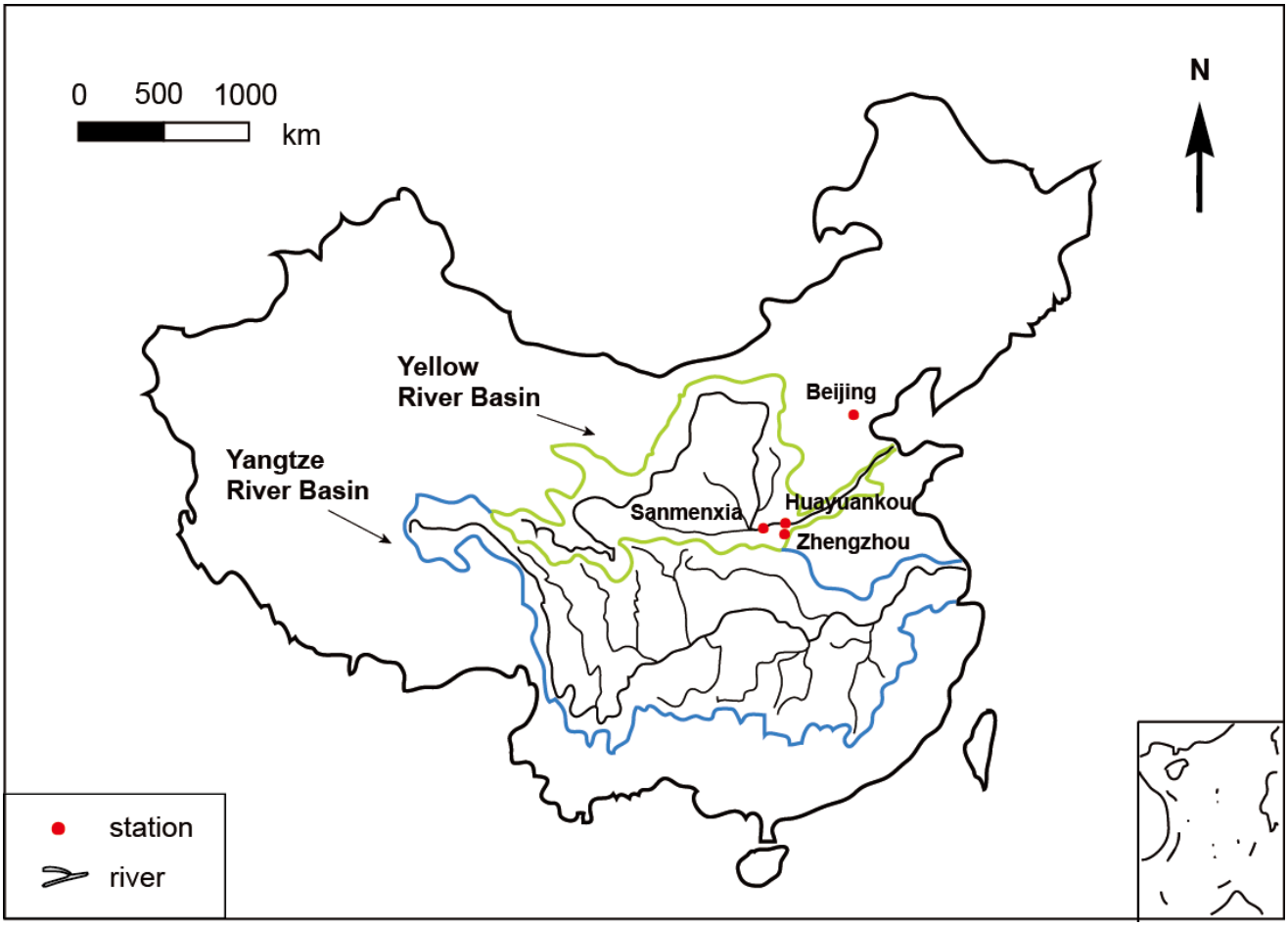


Figure 2 Four representative hydro-meteorological stations in China

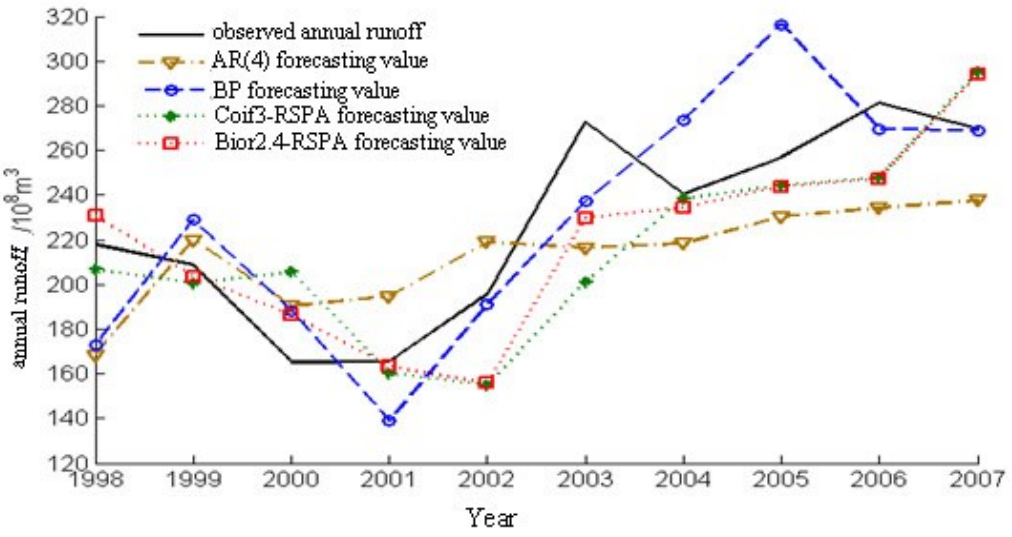


Figure 3 Observed and forecast annual runoff time series, the latter obtained using AR(4), ANN-BP, coif3-RSPA, and bior2.4-RSPA models (T = 5): Case 1, Huayuankou Station, lower Yellow River, China

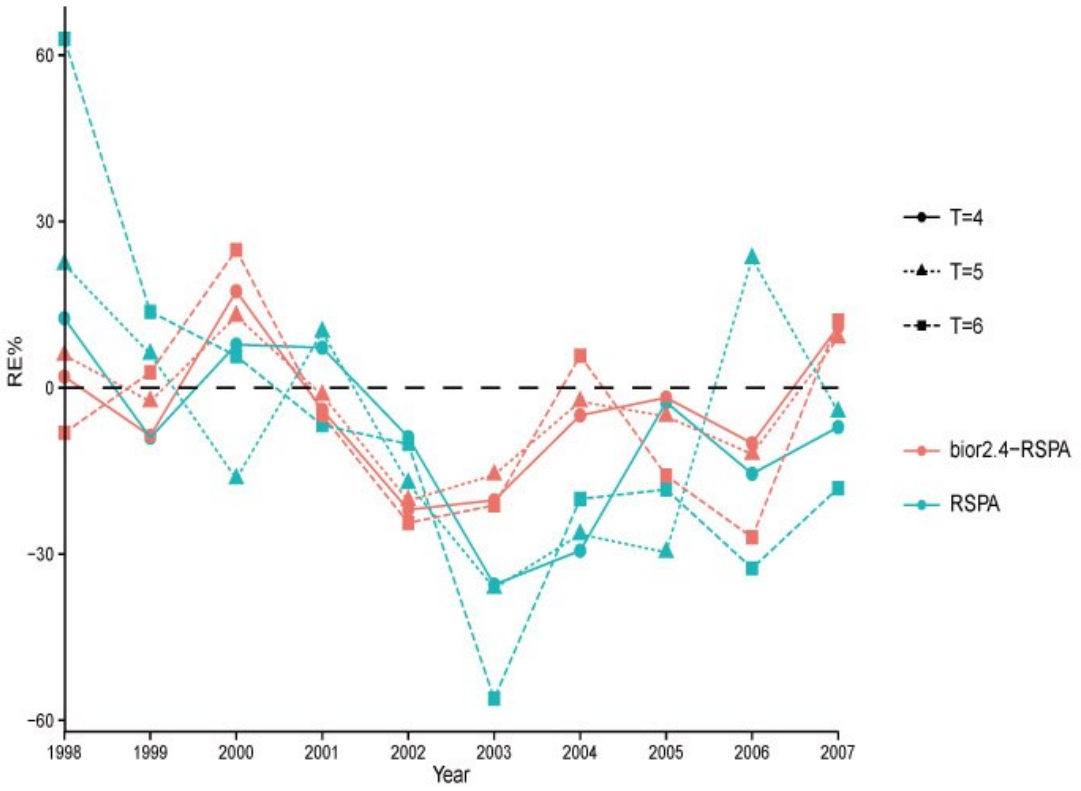
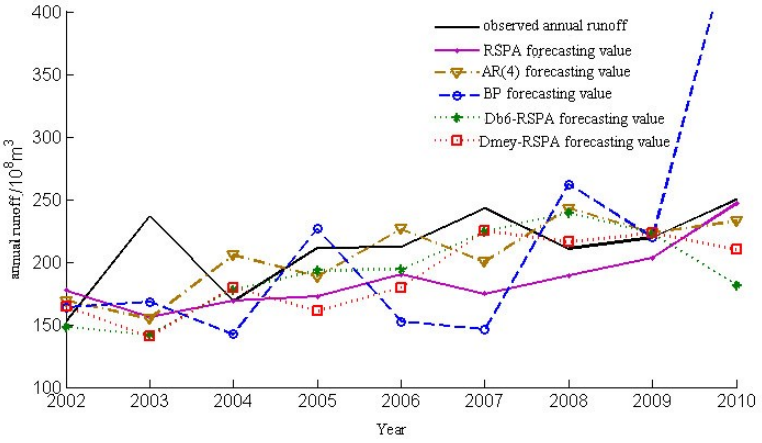
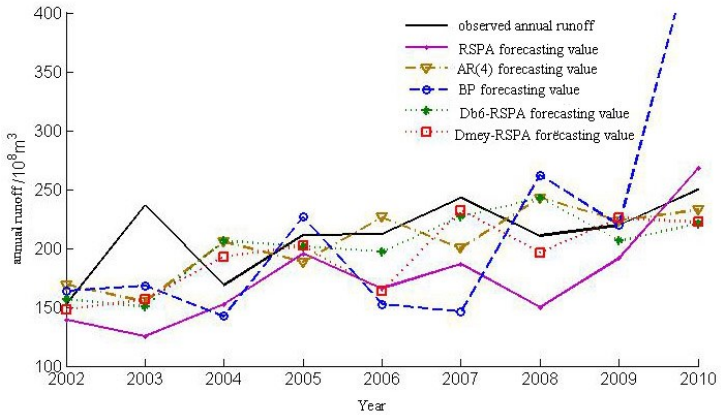


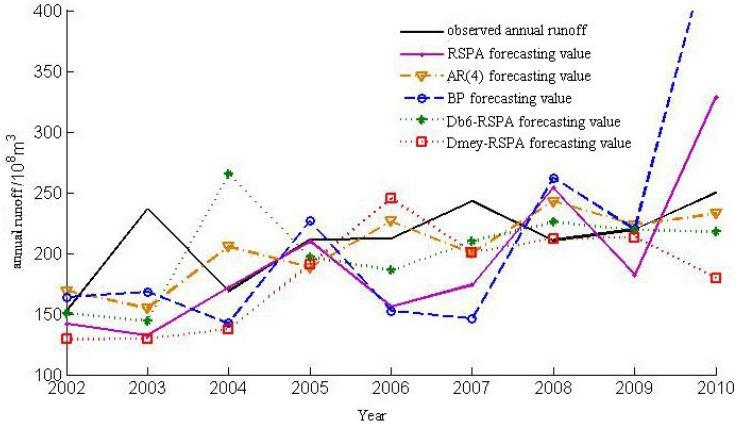
Figure 4 Relative errors between observed and forecast annual runoff time series, the latter obtained using single RSPA and bior2.4-RSPA models (T = 4, 5, 6): Case 1, Huayuankou Station, lower Yellow River, China



(a) T=4

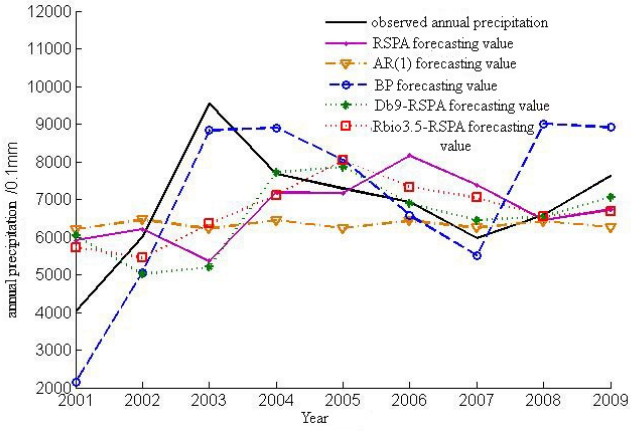


(b) T=5

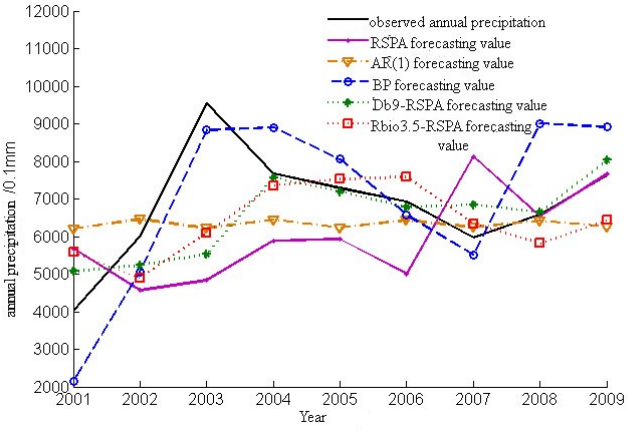


(c) T=6

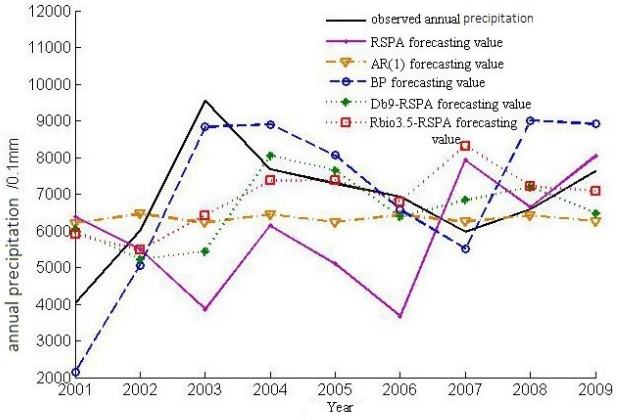
Figure 5 Observed and forecast annual runoff time series, the latter obtained using AR(4), ANN-BP, single RSPA, db6-RSPA, and dmey-RSPA models, for T = (a) 4,(b) 5 and (c) 6: Case 2, Sanmenxia Station, lower Yellow River, China



(a) T=4



(b) T=5



(c) T=6

Figure 6 Observed and forecast annual precipitation time series, the latter obtained using AR(1), ANN-BP, single RSPA, db9-RSPA, and rbio3.5-RSPA models, for T = (a) 4,(b) 5 and (c) 6: Case 3, Zhengzhou Station, Yellow River, China

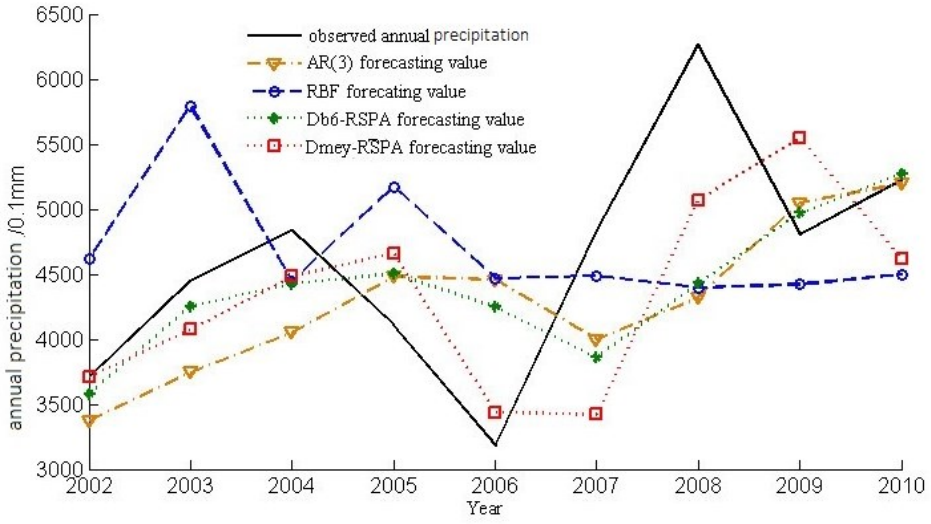


Figure 7 Observed and forecast precipitation time series, the latter obtained using AR(3), ANN-RBF, db6-RSPA, and dmey-RSPA models, for $T = 4$: Case 4, Beijing Station, China

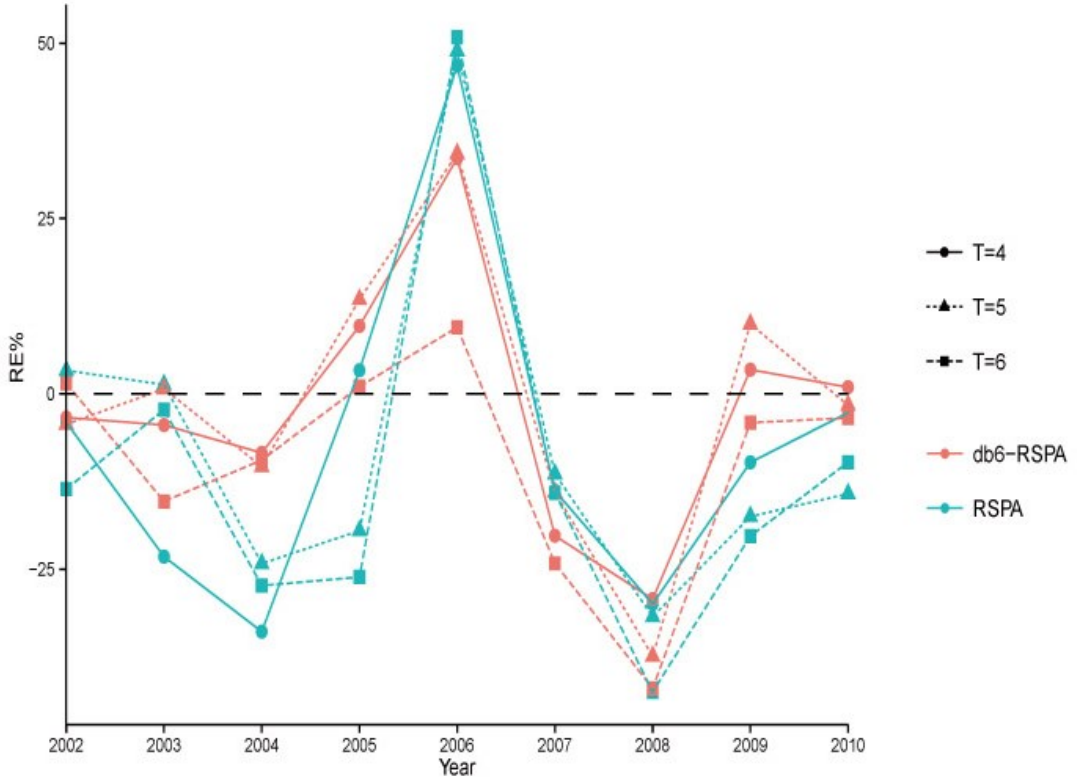


Figure 8 Relative errors between observed and forecast annual precipitation time series, the latter obtained using single RSPA and db6-RSPA models (T = 4, 5, 6): Case 4, Beijing Station, China