Establishment and application of a fractional difference-autoregressive model for daily runoff time series forecasting based on wavelet analysis

Jie Zhang\textsuperscript{a}, Meili Wang\textsuperscript{b} and JieLong Hu\textsuperscript{a}

\textsuperscript{a}College of Hohai, Chongqing Jiaotong University, Chongqing, People’s Republic of China; \textsuperscript{b}College of architecture and urban planning, Chongqing Jiaotong University, Chongqing, People’s Republic of China

ABSTRACT

The daily runoff is a complicated hydrological time-series, its extreme changes will affect not only the safety of the ships underway but also the dispatching of water transportation. Daily runoff forecast is a hot topic in hydrologic analysis and water transport planning. The fractional difference-autoregressive model based on wavelet analysis (WFIAR) is used to forecast the daily runoff. The daily runoff time series from 2002 to 2013 of Hankou Hydrologic Station are decomposed by the orthogonal db4 wavelet function into a series of stationary high-frequency sub-sequences and a non-stationary low frequency subsequence. According to the Hurst exponent of the low frequency subsequence, the appropriate difference order is determined and fractional difference is made. The differenced low-frequency part and high-frequency parts set up autoregressive models separately. These models are combined to forecast the daily runoff time series in 2014 and 2015 separately. The first-order difference autoregression (DIAR) model and the first-order autoregression (WDIAR) model based on wavelet analysis are also established to compare the forecasting accuracy of different models. The results show that the prediction accuracy of WFIAR is higher than the other two models that reveal the advantage of WDIAR model in daily runoff forecast.

1. Introduction

Hydrologic time series is a sample of observations with complex variation characteristics. Hydrologic time series analysis aims at indicating the complex hydrology process which is the basis of hydrologic simulation and prediction, water resources and transport planning and many other water activities. The task is not easy to fulfill due to its various influence factors (Sang, Dong, and Wu, 2009). Establishing an auto-regressive model is a traditional time series prediction method with a convenient and intuitive advantage (Lin, Xia, and Jie, 2013; Wang, Liang, and Ding, 2016). However, hydrological time series are usually non-stationary and do not meet the requirements of establishing an auto-regressive model. The existing studies mostly adopt the first-order difference method to eliminate some volatility of the time series and make it become a stationary time series. Although the method is computationally simple, it is prone to over-differentiation problems, resulting in the loss of some important detail-changing features in the time series. Fractional difference can be to some extent avoid this problem. In recent years, some domestic and foreign scholars have carried out some researches on the application of fractional difference in time series prediction. Hu, Xu, and Wang (2006) proposed a method on the basis of fractional difference Fuzzy-AR method for network traffic simulation and prediction, the fractional difference method was used to eliminate the long-time correlation of the time series. Zuo, Wang, and Xu (2009) established the coefficient auto-regressive predictive model based on fractional difference method, automatically selected fractional-order difference or integer-order difference to deal with the original data and estimate the optimal modelling parameters. The example verification shows that the DFAR model improves the prediction accuracy of fault feature quantities for non-stationary and non-linear development. GADA Pereira and Souza (2014) developed a study of the effects of long dependence in the series of runoff natural energy in the South subsystem, in order to estimate a long memory model capable of generating synthetic hydrologic series. The above attempts prove the validity of fractional order difference in processing time series.

Wavelet analysis theory developed from the Fourier analysis. As it can reveal the local characteristics of time series simultaneously (Kumar and Foufoula-Georgiou,
difference method and autoregressive model. To highlight the advantages of this model, an autoregressive prediction model (DIAR) on the basis of integer difference method and a wavelet-autoregressive prediction model (WDIAR) on the basis of fractional difference method are also established.

The innovation point of this study is as follows.

(1) The discrete wavelet transform method is used in establishing an autoregressive prediction model.
(2) To solve the strong non-stationarity of the daily runoff sequence, the fractional difference method was used after the original sequence had been decomposed.

2. Methodology

2.1. Wavelet analysis

Wavelet analysis is a method of time–frequency multi-resolution analysis. In the book Hydrological Wavelet Analysis (Wang, Ding, & Li, 2005), the continuous wavelet transform is defined as:

\[ W_f(a, b) = \langle f(t), \psi_{a,b}(t) \rangle = |a|^{-1/2} f(t) \tilde{\psi} \left( \frac{t - b}{a} \right) \, dt \tag{1} \]

where, \( W_f(a, b) \) is wavelet transform coefficient; \( f(t) \) is the original signal; \( \psi(t) \) is called the mother wavelet; \( \langle \cdot, \cdot \rangle \) is the inner product; \( a \) is the scaling factor; \( b \) is the time translation factor; \( \tilde{\psi} \) is the complex conjugation mother wavelet function, \( \psi_{a,b}(t) \) is a family of functions which are flexed and translated from \( \psi(t) \).

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

Where, \( \psi_{a,b}(t) \) is called the analysing wavelet or the continuous wavelet. Hydrological time series are mostly discrete; therefore, the continuous wavelet transform must be discretized.

Make \( a = d_0^j b = k_0 b_0 \) and \( a_0 > 0, a_0 \neq 1, b_0 \in R, k \) and \( j \) are integers. The wavelet transform coefficient can be expressed by the discrete wavelet transform (DWT) as follows:

\[ W_d f(j, k) = a_0^{-j/2} d_j f(t) \tilde{\psi} (a_0^{-j/2} t - k_0 b_0) \, dt \tag{3} \]

where, \( j \) denotes the decomposition level; \( k \) denotes the time translation factor. The sampling step in the time domain is adjustable towards different frequency components \( a_0^{-j/2} \). Usually, \( a_0 \) is equal to 2 and \( b_0 \) is usually equal to 2. The discrete wavelet transform can be indicated as...
the follow equation.

\[ W_{\psi}(j, k) = 2^{-j/2} \int f(t) \tilde{\psi}(2^{-j}t - k) dt \]  (4)

The high frequency component corresponding a smaller \( j \) value has a smaller sampling interval, the low frequency component corresponding a greater \( j \) value has a larger sampling interval. That is the reason why the discrete wavelet transform can realize time–frequency localization. After the time-series has been decomposed into wavelet coefficients, the time-series analysis is transformed into the study of wavelet transform coefficients.

The Mallat algorithm was put forward by Mallat on the basis of multiresolution analysis. It has been proved that it is suitable for the fast decomposition and reconstruction of signal by orthogonal wavelet. The Mallat fast algorithm includes the decomposition algorithm and the reconstruction algorithm.

(1) Mallat decomposition algorithm

\[ \begin{align*}
  c_j &= Hc_{j-1} \\
  d_j &= Gd_{j-1}
\end{align*} \]  (5)

Where, \( H \) is the decomposition low pass filter, \( G \) is the decomposition high pass filter. The original time series can be decomposed into an approximation signal \( (c_j) \) and detail signals \( (d_1, d_2, \ldots, d_j) \) respectively by using the above equation.

(2) Mallat reconstruction algorithm

The reconstruction algorithm is the inverse of decomposition algorithm.

\[ c_{j-1} = \tilde{H}c_j + \tilde{G}d_j \]  (6)

Where, \( \tilde{H} \) is the reconstruction low pass filter, \( \tilde{G} \) is the reconstruction high pass filter. Refactor \( c_j \) and \( d_j \) respectively by using Equation (5), the high frequency components \( (D_1, D_2, \ldots, D_j) \) and the low frequency component \( (C_j) \) can be obtained.

Therefore, the original time series can be calculated by follow equation:

\[ f(t) = D_1 + D_2 + \cdots + D_j + C_j \]  (7)

2.2. Fictional difference-autoregressive (FIAR) model

The key steps for establishing a FIAR \((d, p)\) model are as follows:

The first step: Analyse the long memory characteristics in hydrologic time series, and calculate the order ‘\( d \)’ of fractional difference.

The second step: Fractional order difference for time series, get a new stationary time series.

The third step: Determine the order of the AR model, which is to determine the value of ‘\( p \)’.

2.2.1. Autoregressive model

As the autoregressive prediction model method is simple and intuitive, it has been valued and widely used in hydrologic analysis. The autoregressive forecasting model can be defined as:

\[ X_t = \varphi_1(X_{t-1} - \mu) + \varphi_2(X_{t-2} - \mu) + \cdots + \varphi_p(X_{t-p} - \mu) + \mu \]  (8)

Where, \( X_t \) is the original hydrological time series; \( \varphi_1, \varphi_2, \ldots, \varphi_p \) are auto-regressive coefficients; \( \mu \) is the average value of \( X_t \); \( p \) is the order of the auto-regressive model.

Parameters in an auto-regressive model \((\varphi_1)\) can be derived by the Yule-Walker Equation:

\[ \begin{align*}
  r_1 &= \varphi_1 + \varphi_2 r_1 + \cdots + \varphi_p r_{p-1} \\
  r_2 &= \varphi_1 r_1 + \varphi_2 + \cdots + \varphi_p r_{p-2} \\
  \vdots \\
  r_p &= \varphi_1 r_{p-1} + \varphi_2 r_{p-2} + \cdots + \varphi_p
\end{align*} \]  (9)

In the above equation, \( r_i \) is the autocorrelation coefficient of hydrological time series with delay number.

Before establishing an autoregressive model, the stationarity of the hydrologic time series should be test. The autocorrelation graph and the partial correlation graph of the time series can be used to judge whether the time series is stable. If the autocorrelation function of the time series drops to zero, the time-series is stationary, otherwise it is non-stationary. Although the graphical method is simple and intuitive, it is subjective. Augment Dickey-Fuller (ADF) test method was advanced by Dickey and Fuller in 1979 (Xia, 2005). It is a type of unit root test. The ADF rule is to judge the stability of a time series by checking whether there is a unit root in it. If there is not unit root, we can infer that the time series is stationary. The ADF test can be realized directly by MATLAB software.

2.2.2. Fractional difference and fractional integral

If the time series is non-stationary, we cannot establish an autoregressive model with it. The first order difference is a frequently-used method, but it may lead to the over difference problem. Mills (2002) found the relationship between the difference order ‘\( d \)’ and the variance of the time series. The variance will increase with the increase of the difference order until it becomes a stable time series, and then it is going to increase.

Fractional order difference adopts a smaller order to overcome over-difference. The fractional difference...
The key of fractional difference method is to determine the order $d$. It can be determined by using the Equation (18).

$$d = h - 0.5$$ (18)

Where, $h$ is the Hurst exponent of the time series.

An auto-regressive model was established on the differentiated sequence to obtain the predicted data. Then fractional integral the predicted sequence, which means the inverse operation of fractional difference, the forecasting runoff sequence can be obtained.

The matrix of fractional integration is expressed as:

$$X(t) = W(t) \ast G^{-1}$$ (19)

Where, $G^{-1}$ is the inverse matrix of the matrix $G$.

3. The working framework

The main objective of this paper is to establish a combination autoregressive prediction model based on discrete wavelet decomposition. The study was divided into five steps.

1. Collect hydrologic data.
2. Decompose the hydrologic time series. Based on the multi-resolution analysis ability of wavelet, the original daily runoff time series can be decomposed into high frequency sub-series $d_j (j = 1, 2, \ldots, 6)$ and the low frequency sub-series $c_6$.
3. Process the low frequency sub-series $c_6$ by the fractional difference method to obtain a stationary sub-series.
4. Establish an autoregressive model on each stationary sub-series. Forecast the prospective values by using the autoregressive models and reconstruct them together to get the prediction runoff series.
5. Calculate the statistical error of the predictions and compare with other prediction models.

4. Case study

In this paper, hydrological data from 2002 to 2015 of Hankou Hydrologic Station in Yangtze River Basin are selected. The daily runoff data from 2002 to 2013 are used to establish the Wavelet-FIAR model. The daily runoff data of 2014 and 2015 are used to verify the prediction results. Hankou Hydrologic Station is an important hydrological station for controlling the hydrological changes of the Hanjiang River, a mainstream of the Yangtze River. The site of the project is located in the third phase of Hankou River Beach in Wuhan. The basic water gauge is located on the left side of Guanmiaojia Dock which is located in Hankou, it is 1136 km far away from the river mouth. The control basin area of Hankou Hydrologic Station is 1488036 km².
4.1. Wavelet decomposition of daily runoff time series

Generally, all mother wavelets can be divided into two types, orthogonal or nonorthogonal (Sang, Singh, Sun, Chen, and Liu, 2016). The Daubechies wavelet (dbN) is orthogonal which is usually used for the discrete wavelet transforming. The db4 wavelet is selected to decompose the daily runoff time series of Hankou hydrologic stations from 2002 to 2013, the low frequency sub-series $c_6$ and the high frequency sub-series $d_j (j = 1, 2, \ldots, 6)$ can be obtained. The result is as Figure 1.

Wavelet analysis can identify different frequency components of hydrological sequences. The graph of the low frequency sub-series $c_6$ indicates some important characteristics of the daily runoff process, it manifests as the approximate component of the daily runoff process. The principal period of the sequence is annual, and the variation trends are similar in different cycles. Focusing on the sequence in one period, the approximate component increases sostenuto until it gets to the summit, and then begin to decrease. The distinction of approximate component between flood season and non-flood season is revealed clearly in Figure 1. Besides, the high-frequency time-series include $d_1 \sim d_6$, which manifest as the detail components, indicate the detail change characteristics and the saltation of the daily runoff process with a degree of randomness.

Using the ADF test method to test the stability of the sub-series. The detail components are stationary time series and the approximate component is non-stationary time series.

4.2. Difference on the non-stationary sub-series

As the detail component $c_6$ is nonstationary, establishing an AR model with it immediately is unsuited. It is the common practice to difference the non-stationary time series. The Hurst exponent of sequence $c_6$ is calculated as 0.88, thus the order of fractional difference can be determined as 0.38 according to the Equation (18). Then fractional difference the sequence $c_6$ with order is 0.38, the time series named as $W_1$ is the result series. For comparison analyzing, the $c_6$ sequence and the original time series are differenced with first order. Two new time series named as $W_2$ or $W_3$ can be obtained respectively.

Testing all the three new time series by ADF test method, the result shows that they are all stationary. Therefore, autoregressive models can be established with them.

4.3. Model establishment and prediction

Because of the large number of the sub-sequences, choosing an autoregressive model with the same order is positive for application. For all sub-sequences in this study, the accuracy of AR (2) model is acceptable. Hence, AR (2) model is adopted in this paper, that is $p = 2$, AR (2) model is:

$$X_t = \varphi_1 (X_{t-1} - \mu) + \varphi_2 (X_{t-2} - \mu) + \mu$$

The regression coefficients can be derived as follows:

$$\varphi_1 = \frac{r_1 - r_1 r_2}{1 - r_1^2}$$
$$\varphi_2 = \frac{r_2 - r_1^2}{1 - r_1^2}$$

The result is as Figure 1.
Table 1. The parameters of different forecasting models for daily runoff series on Hankou hydrological station.

| Model          | The mean (m³) | The first order coefficient | The second order coefficient |
|----------------|---------------|-----------------------------|-----------------------------|
| DIAR (1,2)     | 17002.560     | -0.808                      | 0.017                       |
| Wavelet-DIAR (1,2) | 18050.800   | -0.426                      | -0.414                      |
| d6             | -0.658        | -1.978                      | 0.985                       |
| d5             | 3.421         | -1.940                      | 0.962                       |
| d4             | -0.055        | -1.849                      | 0.927                       |
| d3             | 0.083         | -1.604                      | 0.866                       |
| d2             | 0.007         | -0.879                      | 0.775                       |
| d1             | -0.002        | 0.730                       | 0.623                       |
| Wavelet-FIAR (0.38,2) | 21458.960  | -0.888                      | 0.985                       |
| W1             | 21458.960     | -0.888                      | 0.985                       |
| d6             | -0.658        | -1.978                      | 0.985                       |
| d5             | 3.421         | -1.940                      | 0.962                       |
| d4             | -0.055        | -1.849                      | 0.927                       |
| d3             | 0.083         | -1.604                      | 0.866                       |
| d2             | 0.007         | -0.879                      | 0.775                       |
| d1             | -0.002        | 0.730                       | 0.623                       |

In the formula, $r_1$, $r_2$ are autocorrelation coefficients of hydrological sequence delay number 1 and 2 respectively.

The forecasting model named as WFIAR (0.38,1) is established based on wavelet decomposition and fractional-order difference combination autoregressive forecasting model of daily runoff process, where 0.38 means the difference order of the approximate component and 2 means the order of autoregressive model. At the same time, a first order difference autoregressive forecasting model named as WDIAR (1,2) and a first order difference single autoregressive forecasting model named as DIAR (1,2) are also established. ‘1’ means first-order difference, ‘2’ means the order of autoregressive model. The parameters of the models are showed in Table 1.

Forecast the daily runoff series of Hankou hydrological station in 2014 and 2015 by using the above models respectively. The predicted results of different models are shown in Figures 2 and 3.

For the daily runoff in 2014 which is shown in Figure 2, the predicted value by the WFIAR (0.38,1) model is closest to the measured value, followed by the WDIAR (1,2) model, the predicted value of the DIAR (1,2) model has the biggest disparity to the measured value. Similar conclusions can be drawn from Figure 3. In allusion to the daily runoff forecast values for 2015, the WFIAR (0.38,1) model has the smallest error, followed by the WDIAR (1,2) model, the DIAR (1,2) model has the largest errors.

The Root Mean Square Error (RMSE) and the Mean Absolute Percentage Error (MAPE) are used to test the prediction accuracy. The results are shown in Table 2.

\[
RMSE = \sqrt{\frac{1}{N} \sum_{t=1}^{N} \left( f'(t) - f(t) \right)^2} \tag{22}
\]

\[
MAPE = \frac{1}{N} \sum_{t=1}^{N} \left| \frac{f'(t) - f(t)}{f(t)} \right| \times 100\% \tag{23}
\]

Where, $f'(t)$ is the forecasted series, and $f(t)$ is the measured time-series.

5. Conclusion

In this paper, a daily runoff prediction model was established, which couples the wavelet decomposition method, the fractional order difference theory and the autoregressive modelling method, named as ‘WFIAR’. The daily runoff processes of Hankou Hydrological Station in 2014 and 2015 are predicted by using the above model respectively. The WDIAR forecast model which combined with the wavelet decomposition theory, the first order difference theory and the autoregressive modelling method and the DIAR forecasting model which combined with the
first order difference theory and the autoregressive modelling method are also established for comparing. The result of the statistical errors indicates that the prediction accuracies of the three models.

The traditional difference method for dealing with non-stationary sequences is easiest and the most intuitive, but it is likely to produce over-difference problem. Some important details of the missing sequence lost that may leading to large errors in the prediction. The prediction accuracy of the WDIAR forecasting model is better than the DIAR model, but is still far from being satisfactory. Although it exploits the advantage of the wavelet decomposition in time series analysis, the over-difference problem is not solved. The WFIAR model exploits the advantages of multi-resolution wavelet analysis and decomposes hydrological sequences into high-frequency sub-sequences and low-frequency sub-sequences, and fully excavates the variation characteristics of the daily runoff series. At the same time, fractional order difference method not only transform the non-stationary time series to stationary to meet the requirements of establishing the autoregressive model but also preserves the detail variation characteristics of the daily runoff series to a great extent. It is superior to the DIAR forecasting model and the WDIAR forecasting model in prediction accuracy.

It is worth noting that the prediction model established need to determine some parameters, these parameters will directly affect the precision of the model prediction, the choice of parameters is also a cause of the hydrological analysis of uncertainty. In future studies, it could be analysed in greater detail.

This paper simply discusses the feasibility of WFIAR forecasting model and its superiority in forecasting accuracy. As regards the structure of the model, there are many areas that can be further improved.

### Acknowledgement

The authors would like to thank the editor and the reviewers for their helpful comments and constructive suggestions, which were very helpful in the revision of this correspondence. This work is supported by the Program of National Key R&D Plan 2016YFC0402106 and Science and Technology Research Project of Chongqing Education Commission under Grant KJ1600511.

### Disclosure statement

No potential conflict of interest was reported by the authors.

### Funding

This work is supported by the Program of National Key R&D Plan under Grant 2016YFC0402106 and Science and Technology Research Project of Chongqing Education Commission under Grant KJ1600511.

### References

Darain, A. B., & Azimi, S. (2016). Forecasting streamflow by combination of a genetic input selection algorithm and wavelet transforms using ANFIS models. International Association of Scientific Hydrology Bulletin, 61, 585–600.

Heng, Wang, Li, & Ding. (2002). Combined stochastic models based on wavelet transform and its application to stochastic modelling of runoff time series. Hydroelectric Energy, 1, 15–17.
Hu, S. M., Xu, W. S., & Wang, Z. J. (2006). Network traffic modelling and prediction based on fractional difference and Fuzzy-AR. *Computer Engineering & Applications*, 42(19), 104–107.

Jayawardena, A. W., Xu, P. C., & Li, W. K. (2009). Rainfall data simulation by the hidden markov model and discrete wavelet transformation. *Stochastic Environmental Research & Risk Assessment*, 23, 863–877.

Kumar, P., & Foufoula-Georgiou, E. (1993). A multicomponent decomposition of spatial rainfall fields: 1. Segregation of large- and small-scale features using wavelet transforms. *Water Resources Research*, 29(8), 2515–2532.

Kumar, P., & Foufoula-Georgiou, E. (1997). Wavelet analysis for geophysical applications. *Reviews of Geophysics*, 35(4), 385–412.

Labat & Dlabatd. (2005). Recent advances in wavelet analysis: Part 1: A review of concepts. *Journal of Hydrology*, 314, 285–288.

Lin, B. D., Xia, L. Z., & Jie, H. H. (2013). Application of autoregressive model for Low flow prediction. *Journal of Water Resources Research*, 02(3), 222–227.

Mills. (2002). *Econometric models of financial time series*. Beijing: Economic sciences press.

Nowak, K. C., Rajagopalan, B., & Zagona, E. (2011). Wavelet Auto-Regressive Method (WARM) for multi-site streamflow simulation of data with non-stationary spectra. *Journal of Hydrology*, 410, 1–12.

Pereira, G. A. D. A., & Souza, R. C. (2014). Long memory models to generate synthetic hydrological series. *Mathematical Problems in Engineering*, 2014, (2014-7-16), 2014(1), 1–8.

Sang, Y. F. (2012). A practical guide to discrete wavelet decomposition of hydrologic time series. *Water Resources Management*, 26(11), 3345–3365.

Sang, Y. F., Dong, W., & Wu, J. C. (2009). The relation between periods’ identification and noises in hydrologic series data. *Journal of Hydrology*, 368(1), 165–177.

Sang, Y. F., Singh, Vijay P., Sun, F. B., Chen, Y. N., Liu, Y., & Yang, M. Y. (2016). Wavelet-based hydrological time series forecasting. *Journal of Hydrologic Engineering*, 21, 06016001.

Wang, W. S., Ding, J., & Heng, T. (2003). Study on the periodicity and long-term forecast of the annual maximum peak discharge at Yichang Station of Yangtze River. *Journal of Sichuan University*, 35, 20–23.

Wang, W. S., Ding, J., & Li, Y. Q. (2005). *Hydrological wavelet analysis*. Beijing: Chemical industry press.

Wang, W. S., Liang, J. J., & Ding, J. (2016). *Random hydrography*. Beijing: China Water Power Press.

Wang, H. R., & Liu, X. H. (2010). Prediction of hydrological series based on wavelet transform and support vector machine. *Journal of Tsinghua University*, 50, 1378–1382.

Wang, W., Yuan, P., & Ding, J. (2000). Wavelet analysis and its application to stochastic simulation of daily flow. *Journal of Hydraulic Engineering*, 31, 43–48.

Xia, N. X. (2005). A comparative study of the DF, ADF test and PP test of unit roots. *Quantitative Economic and Technological Economic Research*, 22(9), 129–135.

Yao, J. Q. (2001). Application of continuous wavelet transform in the analysis of precipitation in nearly 100 years in Shanghai. *Meteorology*, 27(2), 20–24.

Zhao, Y. L., Ding, J., & Deng, Y. R. (1998). Wavelet network model of phase space and its application in hydrologic prediction. *Advance in Waterence*, 9, 252–257.

Zuo, Y. B., Wang, X. B., & Xu, X. L. (2009). Application of DFAR model in fault prediction of rotating machinery. *China Mechanical Engineering*, 20(12), 1460–1463.