Journal of Hydro-environment Research 20 (2018) 77–92

Contents lists available at ScienceDirect

Journal of Hydro-environment Research

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

A new hybrid model for nonlinear and non-stationary runoff prediction at annual and monthly time scales

Xu Chen, Fa-wen Li*, Ping Feng

State Key Laboratory of Hydraulic Engineering Simulation and Safety, Tianjin University, Tianjin 300072, China

ARTICLE INFO

Keywords:
BDS test
ADF test
MM-K test
Improved ensemble intrinsic time-scale decomposition
Improved nearest neighbor bootstrapping regressive model
Long-term runoff prediction

ABSTRACT

Due to the effects of frequent anthropogenic activities and climate change, the natural annual runoff series presents typical nonlinear, non-stationary and multiple-scale characteristics, which triggers the common problem of low accuracy for long-term runoff predictions. Therefore, the main goal of this study was to improve the long-term prediction accuracy for the nonlinear and non-stationary runoff series by introducing a new hybrid approach based on improved ensemble intrinsic time-scale decomposition (IEITD) and improved nearest neighbor bootstrapping regressive (INNBR) methods. First, the Brock-Dechert-Scheinkman (BDS) test and Augmented Dickey-Fuller (ADF) test were used to identify the nonlinear and non-stationary characteristics of annual runoff series, respectively, and the Modified Mann–Kendall (MM-K) test and Mann–Kendall (M-K) abrupt change test methods were applied to explore the drivers of non-stationarity. On this basis, the IEITD was used to decompose the original annual runoff series into several proper rotation components (PRCs) and a monotonic residual trend term to make the non-stationary runoff time series stationary. Then, the INNBR model was applied to predict the respective PRCs. The residual series was predicted by the polynomial fitting method. Finally, the predictive results of each PRC and residual series were summed to obtain an ensemble forecast for the runoff series. The performance of the new hybrid approach was tested by the annual (nine hydrological stations) and monthly (three hydrological stations) runoff data covering 1956–2011 in the Luanhe River basin in China. Results suggested: (1) the annual runoff time series of nine hydrological stations presented obviously dependent nonlinear structure; (2) the annual runoff series of nine hydrological stations were all non-stationary; (3) human activity, rather than change in precipitation, was the major driving factor of runoff decline in the Luanhe River basin; (4) For the performance evaluation criteria of Nash-Sutcliffe efficiency coefficient (NSEC), compared with Nearest Neighbor Bootstrapping Regressive (NNBR) and INNBR, the precision of IEITD-INNBR model almost increase by 227% and 37% on the average of nine hydrological stations; (5) the new hybrid approach of combining the IEITD and INNBR models outperformed the other two models (NNBR and INNBR) tested, and it is capable of capturing the nonlinear, non-stationary and multiple-scale characteristics of complex runoff time series and obtaining higher predictive precision.

1. Introduction

Surface runoff occurs when there is more water than land can absorb, and this can cause soil erosion and transport and the diffusion of fertilizers and other pollutants from soil towards nearby waterbodies. Therefore, from a managerial perspective, the capability of predicting the magnitude of surface runoff is crucial, because predictions make it possible to prevent or mitigate the above reported adverse events. Runoff predictions, especially long-term (annual and monthly) runoff predictions, are of great significance for environmental protection, water quality management, hydropower generation, flood mitigation, disaster alleviation and optimal allocation of water resources (Toro et al., 2013). However, to meet the constant growth of water consumption due to population growth and economic development, modern water resources management requires more efficient techniques or methods to improve the accuracy of runoff predictions. Therefore, in recent decades, long-term runoff predictions have captured increasing attention from academic researchers in operational hydrology.

Due to the comprehensive influence of numerous factors, such as anthropogenic activities and underlying surface conditions, the runoff time series generally displays uncertain, weakly dependent, complicatedly multi-dimensional, highly nonlinear and non-stationary characteristics (Zhao and Chen, 2015). In recent years, with the aggravating

* Corresponding author.
E-mail address: lifawen@tju.edu.cn (F.-w. Li).

https://doi.org/10.1016/j.jher.2018.05.004
Received 9 September 2017; Received in revised form 23 May 2018; Accepted 28 May 2018
Available online 30 May 2018

influence of climatic change and anthropogenic activities, these characteristics appear particularly outstanding. Therefore, the way in which to process the characteristics of multi-dimension, nonlinearity and non-stationarity and, further, to improve the prediction accuracy has become the key of long-term runoff predictions. However, many previous prediction models have been constructed using the original series as the input variables, which inevitably misses some features of different resolutions. If the prediction model is established only based on one resolution component, the changes in the internal mechanism of runoff cannot be completely captured (Chou and Wang, 2004). Currently, aiming at the multi-dimensional, nonlinear and non-stationary characteristics of time series, several self-adaptive time-frequency analysis methods, including empirical mode decomposition (EMD) (Huang et al., 1998), local mean decomposition (LMD) (Sun et al., 2016), and intrinsic time-scale decomposition (ITD) (Feng et al., 2016), have been developed. Nevertheless, the EMD method suffers from disadvantages, such as mode mixing, envelope overshoot and envelope undershoots of cubic spline interpolation (Lin, 2012). In addition, the LMD method is greatly influenced by noise, which produce many extra and useless frequency components; accordingly, it is difficult for LMD to be applied practically (Yang et al., 2012). In contrast, the ITD method is specifically proposed for application to nonlinear and non-stationary signals in nature, and it can extract the underlying dynamics with changes on multiple time-scales from complex systems. Compared with the decomposition algorithms of EMD and LMD, the ITD algorithm has the advantages of low computational complexity, small edge effects, avoiding wavelet base selection, real-time processing of mass data and obtaining proper rotation component with physical significance (Zen et al., 2011). Unfortunately, the ITD method extracts the baseline signal based on the linear transformation, and this may result in the distortion of signal and produce burr. Another drawback of ITD is the mode mixing, which means one proper rotation components (PRC) contains different scale components, or a signal of similar scale is located at different PRC components. The mode mixing problem not only leads to serious aliasing in the time-frequency distribution but also causes the physical meaning of individual PRC obscurity. Therefore, in this study, to effectively overcome those deficiencies of the ITD method, a new time-frequency analysis method, called the improved ensemble intrinsic time-scale decomposition (IEITD) method, is proposed to make the non-stationary series stationary. In the new IEITD method, a cubic spline interpolation is employed to fit the baseline control points, the mirror extension is adopted to address the endpoint of the given signal, and the assistant white noise is introduced to eliminate the mode mixing problem, which significantly improve the performance of the ITD method.

Many previous researchers have used the time-series models, including the AR model, MA model and ARIMA model, to predict the runoff (Boukharouba, 2012; Mohammad et al., 2006; Zhang et al., 2011). These traditional methods tend to be based on the assumption of stationarity and specific distribution (Jain and Kumar, 2007). However, with the influence of climate change and anthropogenic activities, the natural runoff series presents certain multi-dimensional, nonlinear, non-stationary and non-normal characteristics. These traditional methods neglect the nonlinear, non-stationary and non-normal information hidden in the runoff series, which leads to the poor performance in long-term runoff predictions. With a strong ability of nonlinear mapping, artificial neural networks (ANN), support vector machines (SVM), chaotic theory models and fuzzy predictive models have emerged as the times requirement. These new methods have obtained better effects in some specific predictions (Ding and Dong, 2015; Nayak et al., 2007; Sharma et al., 2015; Young et al., 2015). However, they inevitably contain low convergence and local optimum problems. Due to the superiority of simplicity and the avoidance of distributional and linear hypotheses, the NNBR model has been applied widely in various fields (Breidenbach et al. (2012); Liu et al., 2016; Qin et al., 2015). Currently, the study and application of the NNBR model generally chooses the Euclidean distance as the similarity function (Nguyen et al., 2015; Zolfaghari et al., 2016; Shen et al., 2016). The Euclidean distance has the advantages of simple calculations and giving a clear geometric interpretation. Nevertheless, the Euclidean distance only focuses on the "quantity" and neglects the "vector". To offset this deficiency, this paper proposes that the Euclidean distance and cosine distance are simultaneously used to describe the similarity between vectors. The NNBR model is developed by introducing a new hybrid technical index, which is a combination of the Euclidean distance and the cosine distance, to measure the similarity between vectors in this research. Therefore, in view of the above disadvantages of traditional methods and new machine learning models, the IEITD-INNBR hybrid model, which can not only sufficiently consider the multidimensional, nonlinear and non-stationary characteristics but also effectively avoid the assumption about the dependent form and probability distribution form of research objectives, is proposed for long-term runoff predictions.

In this paper, based on the multidimensional, nonlinear and non-stationary characteristics of runoff series, the IEITD is used to decomposed the nonlinear and non-stationary runoff series into several PRCs and a monotonic trend term; based on the ubiquitous pattern of assumption-calibration-verification in the traditional time-series models and new machine learning techniques for runoff prediction and the drawback of the NNBR model in measuring the vector similarity only focusing on the "quantity", the improved NNBR (INNBR) model is proposed. On this basis, the IEITD technique and the INNBR model are coupled to obtain the new IEITD-INNBE hybrid model.

The objective of this study is to develop a new hybrid model to improve the long-term prediction accuracy with focus on the nonlinear and non-stationary characteristics of runoff time series. The process of the study presented are as follows: 1) investigate the nonlinear and non-stationary characteristics of complex hydrologic time series; 2) explore the drivers of non-stationary runoff time series using the trend and abrupt test; 3) improve the ITD method and NNBR model to obtain the new IEITD-INNBE hybrid model; 4) predict the annual runoff by using the proposed IEITD-INNBR hybrid model and check the application of hybrid model for different temporal resolutions and extreme climate conditions by using the month runoff series; and 5) evaluate the performance of the proposed hybrid model by comparing it with the single NNBR and INNBR models.

2. Study area and data

The Luanhe River basin, located in northeast China, was selected as the research object in this study (Fig. 1). The Luanhe River lies between 115°30′–119°45′E and 39°10′–42°40′N and has a drainage area of 33,700 km^2, with an altitude ranging from 2 m to 2205 m, which decreases from the highest point in the northwest to the lowest point in the southeast.

In this study, the annual runoff series of nine typical hydrological stations are selected as the research objects. From the upstream to the downstream, the stations are in the following order: Goutaizi, Boluonuo, Sandaohezi, Xiahenan, Hanjiaying, Chengde, Xiabancheng, Liying and Pingquan, which are located in the Xiaoluanhe River sub-basin, Xingzhouhe River sub-basin, Luanhe River sub-basin, Yimatuhe River sub-basin, Yixunhe River sub-basin, Wuliehe River sub-basin, Laoniuhe River sub-basin, Liuhe River sub-basin and Baohe River sub-basin, respectively. The locations of the stations and their elevations are presented in Fig. 1. The annual runoff data obtained from nine stations in the Luanhe River basin, from 1956 to 2011, were used in this study. Furthermore, to examine the proficiency of the hybrid model in different temporal resolutions, the month runoff of Sandaohezi, Hanjiaying and Liying from 2006 to 2011 was also used to test the performance of the model.