Environmental Research 160 (2018) 269–281

Contents lists available at ScienceDirect

Environmental Research

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

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

Dong Wang[a,*], Alistair G. Borthwick[b,c], Handan He[a], Yuankun Wang[a,*], Jieyu Zhu[a], Yuan Lu[a], Pengcheng Xu[a], Xiankui Zeng[a], Jichun Wu[a], Lachun Wang[d], Xinqing Zou[d], Jiufu Liu[e,f], Ying Zou[e,f], Ruimin He[e,f]

[a] 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, PR China
[b] School of Engineering, The University of Edinburgh, Edinburgh EH9 3JL, UK
[c] School of Engineering, The University of Edinburgh, St Edmund Hall, Queen's Lane, Oxford OX1 4AR, UK
[d] School of Geographic and Oceanographic science, Collaborative Innovation Center of South China Sea Studies, Nanjing University, Nanjing, PR China
[e] Nanjing Hydraulic Research Institute, Nanjing, PR China
[f] State Key Laboratory of Hydrology, Water Resources and Hydraulic Engineering, Nanjing Hydraulic Research Institute, Nanjing, PR China

A R T I C L E I N F O

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

A B S T R A C T

Accurate, fast forecasting of hydro-meteorological time series is presently a major challenge in drought and flood mitigation. This paper proposes a hybrid approach, wavelet de-noising (WD) and Rank-Set Pair Analysis (RSPA), that takes full advantage of a combination of the two approaches to improve forecasts of hydro-meteorological time series. WD allows decomposition and reconstruction of a time series by the wavelet transform, and hence separation of the noise from the original series. RSPA, a more reliable and efficient version of Set Pair Analysis, is integrated with WD to form the hybrid WD-RSPA approach. Two types of hydro-meteorological data sets with different characteristics and different levels of human influences at some representative stations are used to illustrate the WD-RSPA approach. The approach is also compared to three other generic methods: the conventional Auto Regressive Integrated Moving Average (ARIMA) method, Artificial Neural Networks (ANNs) (BP-error Back Propagation, MLP-Multilayer Perceptron and RBF-Radial Basis Function), and RSPA alone. Nine error metrics are used to evaluate the model performance. Compared to three other generic methods, the results generated by WD-REPA model presented invariably smaller error measures which means the forecasting capability of the WD-REPA model is better than other models. The results show that WD-RSPA is accurate, feasible, and effective. In particular, WD-RSPA is found to be the best among the various generic methods compared in this paper, even when the extreme events are included within a time series.

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 (Bardossy and Plate, 1992; Bardossy and Li, 2008; Cai et al., 2009; Whitworth et al., 2012; Chen et al., 2015; Mehran et al., 2015). A major challenge is presently faced in how to ensure the sustainability of water resources, and this is made harder by the insufficiency of hydrologic data in developing countries (Qian and Leung, 2007; Leung et al., 2013; Hong et al., 2016). 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 hydro-meteorological process is particularly complicated because of climate and anthropogenic drivers (Gao and Sorooshian, 1994). Furthermore, this process involves several uncertain factors such as randomness, fuzziness and chaos (Sivakumar et al., 1999; Han and Bray, 2006; Han et al., 2007; Kavvas et al., 2013; Wang et al., 2015a, 2015b). 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

* Corresponding authors.
E-mail addresses: wangdong@nju.edu.cn (D. Wang), yuankunw@nju.edu.cn (Y. Wang).

http://dx.doi.org/10.1016/j.envres.2017.09.033
Received 21 June 2017; Received in revised form 1 September 2017; Accepted 29 September 2017

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; Adamowski and Sun, 2010; Nalley et al., 2012; Belayneh and Adamowski, 2013; Liu et al., 2013; Belayneh et al., 2014, 2016; Hong et al., 2016).

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., 2013a, 2013b), 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; Li et al., 2016; Pan et al., 2017), waterlog disaster risk evaluation (Jin et al., 2015), sustainability assessment of a water resources system (Du et al., 2015) and safety analysis (Chong et al., 2017).

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. (2013a, 2013b) 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 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. Wang et al. (2012) used weighted rank set pair method to establish Annual runoff forecasting model. Hou et al. (2017) presented rank set pair analysis (SPA) as a new method to build ensemble surrogate (ES) model, and conducted a comparative research to select a better ES modeling pattern for the SEAR strategy optimization problems.

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 middle and 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 Adamowski et al. (2012). 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:

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

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