Environmental Research ∎ (∎∎∎∎) ∎∎∎–∎∎∎

Contents lists available at ScienceDirect

Environmental Research

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

# Mid- and long-term runoff predictions by an improved phase-space reconstruction model

Mei Hong [a], Dong Wang [b,*], Yuankun Wang [b], Xiankui Zeng [b], Shanshan Ge [a], Hengqian Yan [a], Vijay P. Singh [c]

[a] Research Center of Ocean Environment Numerical Simulation, Institute of Meteorology and oceanography, PLA University of Science and Technology, Nanjing, China
[b] Key Laboratory of Surficial Geochemistry, Ministry of Education, 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 210093, China
[c] Department of Biological and Agricultural Engineering Zachry Department of Civil Engineering, Texas A & M University, College Station, TX 77843, USA

ARTICLE INFO

*Article history:*
Received 22 September 2015
Received in revised form
14 November 2015
Accepted 19 November 2015

*Keywords:*
Mid- and long-term runoff
The improved phase-space reconstruction model
Runoff forecast
The Yellow River
Largest Lyapunov exponent

ABSTRACT

In recent years, the phase-space reconstruction method has usually been used for mid- and long-term runoff predictions. However, the traditional phase-space reconstruction method is still needs to be improved. Using the genetic algorithm to improve the phase-space reconstruction method, a new nonlinear model of monthly runoff is constructed. The new model does not rely heavily on embedding dimensions. Recognizing that the rainfall–runoff process is complex, affected by a number of factors, more variables (e.g. temperature and rainfall) are incorporated in the model. In order to detect the possible presence of chaos in the runoff dynamics, chaotic characteristics of the model are also analyzed, which shows the model can represent the nonlinear and chaotic characteristics of the runoff. The model is tested for its forecasting performance in four types of experiments using data from six hydrological stations on the Yellow River and the Yangtze River. Results show that the medium-and long-term runoff is satisfactorily forecasted at the hydrological stations. Not only is the forecasting trend accurate, but also the mean absolute percentage error is no more than 15%. Moreover, the forecast results of wet years and dry years are both good, which means that the improved model can overcome the traditional "wet years and dry years predictability barrier," to some extent. The model forecasts for different regions are all good, showing the universality of the approach. Compared with selected conceptual and empirical methods, the model exhibits greater reliability and stability in the long-term runoff prediction. Our study provides a new thinking for research on the association between the monthly runoff and other hydrological factors, and also provides a new method for the prediction of the monthly runoff.


## 1. Introduction

Floods can cause a lot of economic and social losses. Due to global warming, heavy precipitation is observed and projected to be more frequent and intense (Min et al., 2011; Im et al., 2011; Chou et al., 2012). This can substantially increase the risk of flooding. Runoff forecasting models that can provide timely and early accurate flood warning will help reduce losses and can be used for planning appropriate long-term adaptation measures. Runoff models are also necessary for appropriate water management.

However, the process of runoff has significant randomness and the runoff series is non-stationary (Zhou and Zhou, 2004). Therefore, it cannot be described by a linear system approach. Runoff models can be divided into two main categories: conceptual and empirical. Conceptual models attempt to simulate complex and nonlinear physical processes, e.g., evaporation, evapotranspiration, infiltration, surface flow, subsurface flow and groundwater flow, by employing complex mathematical formulas composed of a large number of parameters (Lidén and Harlin, 2000; Franchini and Pacciani, 1991). While the conceptual models are useful for our understanding of the physical mechanisms involved in the river flow (or any other hydrological) process, unfortunately, there are a lot of difficulties in their application (Sivakumar et al., 2002). This has caused the attention of the hydrologists to focus on another category of models, called empirical models.

Empirical models, also known as black-box models, do not attempt to formulate mathematical functions to accurately model the physical laws. They establish the relationship between input

* Corresponding author.
*E-mail address:* wangdong@nju.edu.cn (D. Wang).

http://dx.doi.org/10.1016/j.envres.2015.11.024
0013-9351/© 2015 Elsevier Inc. All rights reserved.

and output variables by using statistics from an ensemble of training data, which are good representatives of population (Liang et al., 2013). Examples of empirical models are artificial neural networks (ANNs), the generalized likelihood uncertainty estimation (GLUE) method, Non-Linear Multiple Regression (NLMR) and Adaptive Neuro-Fuzzy Inference System (ANFIS), etc.. However, empirical methods are not without flaws. For example, the number of hidden layer neurons of the ANNs model is difficult to objectively determine and the training process may easily lead to a local optimum, limiting the forecast accuracy (Jain et al., 2001). For the GLUE method, to get the combination of the optimal parameters and the selection of the optimal parameters is difficult, which will affect the long term forecast results (Beven and Binley, 1992). The NLMR method is good for non-flood situation, but is quite bad for the prediction of abnormal information, such as flood situation (Goldstein, 1986). The choice of parameters in the ANFIS model will limit the forecast accuracy (Tahmasebi, 2012). Moreover, the reliability of empirical methods is gradually reduced with increasing forecast time, so the credibility of mid- and long-term runoff forecast will become low after a long time (Kumar et al., 2005; Antar et al., 2011). To overcome these drawbacks, a number of researchers have attempted to develop more useful methods in recent years (Breaford et al., 1991; Sivakumar et al., 2001). For example, in recent years, there are a lot of real-life case studies of contemporary soft computing techniques in water resources engineering (Gholami et al., 2015; Taormina and Chau, 2015; Wu et al., 2009). These studies show that the contemporary soft computing techniques are widely used in water resources engineering and have some achievements (Wang et al., 2015; Chen et al., 2015; Chau and Wu, 2010).

The reconstruction of low-order nonlinear dynamics from the time series of a state variable has been an active area of research in the last decade. Dynamic reconstruction is the problem of approximating an unknown function that describes the state evolution of a chaotic system (Islam and Sivakumar, 2002). A variety of techniques, developed in the context of nonlinear and chaotic dynamics, have been employed to identify or predict the dynamics of observed runoff in various geographical regions (e.g. Jayawardena and Lai, 1994; Wang and Gan, 1998; Jayawardena and Gurung, 2000).

Sivakumar et al. (2001) reconstructed a single-dimensional (or variable) runoff series in a multi-dimensional phase space to represent its dynamics, and then used a local polynomial approach to predict monthly runoff in the Coaracy Nunes/Araguari River basin in northern Brazil. Sivakumar et al. (2002) compared the phase-space reconstruction method (PSR) with other empirical models, such as ANNs, GLUE etc., and they found that mid- and long-term runoff forecasts of the PSR model were better than those of the traditional empirical models. For example, Sivakumar et al. (2002a) compared the phase-space reconstruction (PSR) and artificial neural networks (ANN) for 1-day and 7-day ahead forecasts of daily river flow at the Nakhon Sawan station in the Chao Phraya River basin in Thailand. The performance of the PSR method was found to be consistently better than that of ANN. One reason for this could be that in the PSR method the flow series in the phase-space is represented step by step in the local neighborhood rather than a global approximation as is done in ANN. Another reason could be the use of multi-layer perceptron (MLP) in ANN, since MLPs may not be most appropriate for forecasting for longer lead times. But in their study, the forecast results relied heavily on embedding dimensions and the PSR runoff system considered only the runoff variable and did not consider the integrated physical process of rainfall–runoff. To overcome this problem, Sivakumar et al. (2000) analyzed monthly rainfall, runoff and runoff coefficient series using the PSR method. However, the prediction results were as accurate and the prediction accuracy did not remain constant beyond the optimal embedding dimension. They also proposed the possible extensions of the analysis as follows: (1) Use of other chaos identification methods to confirm the results obtained previously regarding the existence of chaos in the rainfall–runoff process; (2) independent analysis of other variables that influence the rainfall–runoff process, such as temperature (e.g. Jinno et al., 1995); and (3) more methods can be introduced to improve the PSR method.

Aiming at the three possible extensions, our study introduces more variables to reconstruct a new improved phase-space reconstruction model of runoff in the Yellow River basin and forecast the mid- and long-term runoff. Zhang and Hong (2008) investigated the reconstruction of a nonlinear statistical-dynamical forecast model of the El Niño Southern Oscillation (ENSO) index, and achieved satisfactory results. Since a runoff time series is also characterized by highly nonlinear, multiple time scales and generally difficult to predict effectively using traditional methods, the method due to Zhang and Hong (2008) can be introduced to reconstruct the improved phase-space reconstruction model of runoff in the Yellow River basin and forecast mid- and long-term runoff. Acknowledging that the rainfall–runoff process is complex and is affected by a number of factors, more variables (e.g. temperature, rainfall and evaporation, etc.) are incorporated in the model. To improve the conventional phase space reconstruction method, we introduced a genetic algorithm (GA) to improve the determination of root efficiency of model parameters. Compared with other optimization search methods, the genetic algorithm has three advantages: (1) The genetic algorithm operates directly on the structure of an object. The continuity of derivation and function do not need to exist (Patrascu et al., 2014). (2) The genetic algorithm has a global implicit inherent parallelism and better optimization capability (Patrascu et al., 2014). (3) The overall search strategy and search optimization methods of the genetic algorithm do not rely on the gradient information or other auxiliary knowledge in the calculation and only need the objective function affecting the search direction and the corresponding fitness function (Civicioglu, 2012). Therefore, the genetic algorithm provides a common framework to solve complex system problems. It does not depend on specific areas and can be widely used in many sciences. Hence, we chose the genetic algorithm to improve the traditional PSR model, which would make the model parameters more accurate and reasonable. Also, the model was tested for forecasting monthly runoff from the Yellow River and the Yangtze River. As Sivakumar et al. (2001) remarked, understanding runoff dynamics at the monthly scale, for example, is much more important than that at the daily scale. Finally, the forecasting results show that our new model can overcome four traditional runoff forecast problems: (1) The new model does not rely heavily on embedding dimensions and it considers the complex rainfall-runoff process by incorporating more variables (e.g. temperature and rainfall) in the model. (2) Our model can represent the nonlinear and chaotic characteristics of the runoff. (3) The medium- and long-term runoff forecast results are accurate. Moreover, the forecast results of wet years and dry years are both good, which means that the improved model can overcome the traditional "wet years and dry years predictability barrier," to some extent. (4) The forecast results of our model for different regions are all good, showing the universality of our approach.

The rest of this paper is organized as follows. In Section 2, three factors for models are chosen by the correlation analysis and physical mechanism analysis. On this basis, a new improved phase-space reconstruction model of runoff is reconstructed. The chaotic characteristics of the model are analyzed, in order to provide a preliminary evidence of the existence of chaos in the monthly rainfall-runoff process in Yellow River basin. Forecast of monthly river runoff is done at Lijin station and other five gauging