8th International Congress on Civil Engineering,
May 11-13, 2009, Shiraz University, Shiraz, Iran

Data Pre-processing Concern in Hydrological Time Series Modeling Using Artificial Neural Networks

Shahram Karimi-Googhari
Assistant Professor, Department of Water Engineering, Shahid-Bahonar University of Kerman, Kerman, Iran
E-mail: shkarimi1352@gmail.com

Abstract

Time series modeling for either data generation or forecasting of hydrologic variables is an important step in the planning and operational analysis of water resources. The capability of Artificial Neural Networks (ANN) in modeling of daily reservoir inflow forecasting was examined in a small tropical catchment. Cross-validation and pre-processing of data was considered as alternatives in modeling process. The model inputs were extracted using auto-, cross-, and partial auto-correlation functions. The results showed that the feed forward back-propagation neural networks are able to forecast extremely changeable daily reservoir inflows. Cross-validation of data improved the model performance indices. Transforming the data to normal distribution prior to training confirmed increasing significantly the model persistency and generalization in simulating an independence data set.

Keywords: Data pre-processing, Neural networks, Forecasting, Reservoir inflow

1. INTRODUCTION

Many of the techniques currently used in modeling hydrological time-series consider linear relationships among the variables. The two main groups of techniques include physically based conceptual models and time-series models. In the first group, the models usually compromise of simple forms of physical laws and are generally non-linear, time-invariant, and deterministic, with parameters that are representative of watershed characteristics [1] but ignore the spatially distributed, time-varying, and stochastic properties of the hydrological process. The practical application and calibration of conceptual models is complex and needs comprehensive field surveying and elaborate mathematical tools [2,3,4] and some degree of expertise and experience with the model [1]. In time-series modeling, traditionally, the multivariate autoregressive moving average (ARMA) models are widely used for modeling water resources time-series [5,6]. In, time-series models are more practical than conceptual models because there is no need to understand the internal structure of the physical processes that are taking place in the system being modeled. Many of the available techniques are not able to consider non-linear dynamics inherent in the transformation of rainfall to runoff.

New computing tools and black-box modeling techniques have overcame the above comments. In black box modeling, the input variables connect to output of a system with only limited knowledge about the physical behavior of the system. Techniques used for data-driven modeling can be stated as machine learning (decision tree, Bayesian methods, neural networks, reinforcement learning), soft computing (fuzzy inference systems, neuro-fuzzy), data mining (uses machine learning methods and statistics), non-linear dynamics and chaos theory. These categories often overlap each other [7].

Artificial Neural Network (ANN) is an information processing paradigm which is composed of a large number of highly interconnected processing elements working in unison to solve specific problems. The multi layer perceptron (MLP) is the most popular network architecture in use these days [8, 9, 10] due originally to [11] and discussed in most neural network textbooks (e.g. [12]).

Many researches have applied Artificial Neural Networks (ANNs) to model different complex hydrological processes. ANN methods have great generalization and usually used in practical hydrologic projects [13]. Even when there are missing data values, ANN methods can be applied for infilling missing hydrological records [14]. Some authors have compared Box–Jenkins and ANN methods [1, 15] confirming, in most cases, the superiority of ANNs. Using ANNs models for short term river flows forecasting have been investigated in the Xallas river [16], Serpis river [17], Winipeg river system [13] and Geer catchment [18] and all of them concluded the feed forward ANNs could forecast the streamflows accurately.