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

Statistical Functions as an Auxiliary Means in Hydrological Time Series Modeling Using New Data Driven Techniques

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

Abstract

Interesting in using of relatively new data driven modeling techniques (eg. neural networks or neuro-fuzzy) is being increased among water resources engineering researchers in recent years. One of difficult task in hydrological modeling using these new computational intelligence techniques is selecting appropriate number of model inputs. In this study, the ability of using auto-, cross-, and partial auto-correlation functions were evaluated for model input selection. The method was applied for Sembrong dam's reservoir inflow data located in Malaysia. The selected inputs were used to construct simple neural network and neuro-fuzzy models. Results show using these statistical functions can reduce attempts in modeling hydrological time series using computational intelligence techniques.

Keywords: Input Selection, Statistical functions, Artificial neural networks, Neuro-fuzzy

1. INTRODUCTION

Time series modeling for forecasting of hydrologic variables is an important step in the operational analysis of water resources. Recently a growing interest in the modeling of complex real-world problems has developed a variety of test procedures for detecting the nonlinearity relationships have evolved. The artificial neural networks (ANNs) and neuro-fuzzy (NF) are two relatively new techniques which have applied successfully in hydrological time series modelling. Unfortunately, for many applications, the theory does not guide the model building process by suggesting the relevant input variables. This particular difficulty makes it attractive to consider an 'atheoretical' but flexible class of statistical models.

For a hydrological modeling problem, it is common to have large number of potential inputs to the model under construction. Using the large number of inputs for a model not only spoil the transparency of the underlying model, but also increase the complexity of necessary computation. Therefore, it is worthwhile to do input selection. Specifically, the purposes of input selection include:

- Remove irrelevant inputs
- Remove inputs that depends on other inputs
- Make the underlying model more concise and transparent
- Reduce the time for model construction

In the hydrological time series modeling using the variable values in the past times (lag times) as input is very common. For instance, in developing artificial neural network for modeling rainfall-runoff process, the time of concentration in the catchment can give very useful information about the number of time steps in the past for which rainfall needs to be considered as input neurons [1]. Based on guidelines, the input vectors to the model will be identified using either an Auto-correlation function (ACF), partial auto-correlation (PACF) or a cross-correlation function (CCF) among the variables in question [2].

In this study, we explain how the reservoir inflow data extracted for a dam catchment located in a tropical area and examine the ACF, PACF for these data to select the appropriate lag time as model inputs. The CCF applied to average rainfall data across the dam catchment to choose the best lag numbers of rainfall data as model inputs. The ability of statistical function in ANN and NF models developing was investigated.