XVIII International Conference on Water Distribution Systems Analysis, WDSA2016

Gene expression programing in long term water demand forecasts using wavelet decomposition

Peyman Yousefi[a], Sina Shabani[a], Hadi Mohammadi[a], and Gholamreza Naser[a]*

[a]Okanagan School of Engineering, The University of British Columbia, Canada

Abstract

Increasing draught seasons and lack of access to potable water reserves have been the major risks threatening water authorities and governments over the recent years. Therefore, long term water forecasts are receiving much more attention nowadays. Unlike the conventional projection of historical water demand, researchers have tried to implement sophisticated mathematical models to predict demand of water. Gene expression programming (GEP), as a relatively new forecasting technique, remains to be explored in this endeavor. The main purpose of this research was to assess the performance of GEP models using wavelet decomposition with 2 transfer functions (*db2* and *haar*) and 3 levels. Results of this study showed GEP models can be highly sensitive to wavelet decomposition if all combinations of proper lag times are used as inputs feeding these models.

Keywords: Water Demand Forecasting; Gene expression Programming; Wavelet Decomposition, Lag Time.

1. Introduction

Highly uncertain governing factors are coupled to water demand, making water distribution systems (WDSs) one of the most complex infrastructures in terms of resource management. Climate change has significantly shifted water availability patterns over the recent years [1]. On the other hand, sudden economic expansions and increasing urbanization escalated water demand worldwide [2]. Consequently, governments are expected to be prepared for

---

* Corresponding author. Tel.: +1-250-807-8464; fax: +1-250-807-9850.
  E-mail address: bahman.naser@ubc.ca

highly uncertain anticipated future of water demands. Projection of historical water usage patterns is no longer applicable due to high complexity of the mentioned global drivers of demand. Therefore, an educated/engineering prediction of demand seems necessary for efficient management of this valuable resource. Water demand forecasting has attracted many researchers over the last decade. A wide range of modeling techniques has been employed to assess the accuracy of such forecasts. Multiple linear regression [3-5], time-series analysis [6-7], artificial neural networks [8-9], hybrid models [10-11], and support vector machines [12-13] have been compared by scholars working in this field. Recently, gene expression programming has been employed by researchers in forecasting models [14]. Unlike most of the emerging complicated techniques, GEP is not a black-box model failing to deliver models formulation. Therefore, it can have a certain edge over conventional models. Coupled wavelet-GEP models gave promising results in other fields such as precipitation prediction [15], runoff forecasting [16], and stream-flow prediction [17]. Therefore, the prime objective of this research is to investigate the performance of this approach in long term forecasting of water demand in City of Kelowna district (CKD), Canada.

2. Methodology

2.1. Gene Expression Programming

Inspired by Darwin's theory of evolution, gene expression programming (GEP) was proposed by Ferreira [18], as an alternative or complement to other genetic based computer programming techniques like genetic programing (GP) and genetic algorithms (GA). This model works based on two simple entities: 1) chromosomes 2) expression trees. It starts with random generation of chromosomes which are linear fixed string of numbers defined by the genes. Moreover, unconstrained applications of genetic operators (e.g. Replication, recombination, mutation, and etc.) are allowed on these linear chromosomes. Figure 1 shows a simple structure or expression tree (ET) diagram of a sample candidate solution which shows how the encoding differs from GP and GA. Such diagrams should be read from left to right. These models are based on a training which enhances the algorithms to look for the optimum candidate solution or "offspring/children" among the generated population subjected to a selection environment. In this paper, the maximum fitness was used as stopping condition of the developed GEP models. Following previous researchers [15] and suggested values by Ferreira [19], 30 chromosomes, 8 head sizes, and 3 genes were used for model structures. Root mean square of error (RMSE) was used as a fitness function to fit a curve to target values.
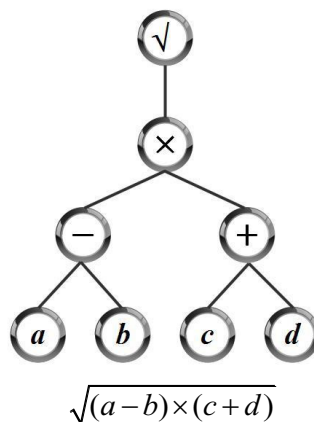

$$\sqrt{(a-b) \times (c+d)}$$

Fig. 1. Structure of same GEP candidate solution