# Forecasting monthly urban water demand using Extended Kalman Filter and Genetic Programming

Mohsen Nasseri [a,*], Ali Moeini [b], Massoud Tabesh [c]

[a] School of Civil Engineering, University of Tehran, Iran
[b] Department of Algorithms and Computations, University of Tehran, Iran
[c] Center of Excellence for Engineering and Management of Infrastructure, School of Civil Engineering, University of Tehran, Iran

## ARTICLE INFO

## ABSTRACT

In this paper, a hybrid model which combines Extended Kalman Filter (EKF) and Genetic Programming (GP) for forecasting of water demand in Tehran is developed. The initial goal of the current work is forecasting monthly water demand using GP for achieving an explicit optimum formula. In the proposed model, the EKF is applied to infer latent variables in order to make a forecasting based on GP results of water demand. The available dataset includes monthly water consumption of Tehran, the capital of Iran, from 1992 to 2002. Five best formulas based on GP results on this dataset are presented. In these models, the first five to three lags of observed water demand are used as probable and independent inputs. For each model, sensitivity of the results for each input is measured mathematically. A model with the most compatibility of the computed versus the observed water demand is used for filtering based on EKF method. Results of GP and hybrid models of EKFGP demonstrate the visible effect of observation precision on water demand prediction. These results can help decision makers of water resources to reduce their risks of online water demand forecasting and optimal operation of urban water systems.

© 2010 Elsevier Ltd. All rights reserved.

## 1. Introduction

Water is known as the most important resource in any urban development program. Most of the decisions in urban planning and sustainable development are highly dependent on forecasting of water demand. Many important decisions on environmental planning, sustainable and optimum utilization of water resources and agricultural projects depend on water demand and its prediction. Respectively, managers and engineers are interested in the water demand forecasting. Correct estimation of water demand carried by a population point is very important for water management projects, and it constitutes one of the important issues in water resources planning. Corresponding to these views, three types of temporal resolution may be encountered in water demand modeling and forecasting:

(a) Long term prediction is concerned with large scale planning and management. Most of the long term development programs in urban management are based on this type of prediction. The prediction resolution is equal or greater than one year (Willsie & Pratt, 1974; Wong, 1972; Young, 1973).

(b) Middle term prediction is applied in middle time management and its resolution is equal or grater then one month or less than one year (Maidment & Parzen, 1984a, 1984b).

(c) Short term prediction is concerned with low scale planning and management. The resolution of this type of approach varies from 1 h to some days (Billings & Agthe, 1998; Gato, Jayasuriya, & Roberts, 2007; Liu, Savenije, & Xu, 2002; Smith, Steiner, Meyer, & Erskine, 1985).

Typically in water demand simulations, a few factors such as temperature (Smith et al., 1985), rainfall (Zhou, McMahon, Walton, & Lewis, 2000, 2002), and evaporation (Willsie & Pratt, 1974; Wong, 1972) are considered.

Nowadays, artificial intelligence (AI) plays an important role in modeling and simulation of many non-convex and complex phenomena. For example, artificial neural networks (ANNs), fuzzy inference systems (FIS) and Neuro-Fuzzy have been applied in predicting of water demand (Liu et al., 2002; Tabesh & Dini, 2009; Yu, Joo, & Koo, 2002).

Also, Genetic Programming (GP) is one of the most well-known methods in AI to explain nonlinear relationships between some parameters. In GP, a user can find straight mathematical or logical relationships between some input and output (Sette & Boullart, 2001), and this can be counted as the most important property of GP. In 1997, GP has been also been used for describing some

* Corresponding author.
 *E-mail addresses:* mm_nasseri@yahoo.com, mnasseri@ut.ac.ir (M. Nasseri), moeini@ut.ac.ir (A. Moeini), mtabesh@ut.ac.ir (M. Tabesh).

nonlinear relationships in water engineering by Babovic and Abbott (1997a, 1997b). GP is a successively piece of work where,

(i) the poor information of the selected relevant variables;
(ii) understanding the shape, size and final form of the model;
(iii) traditional methods in mathematics do not provide the analytical solution.

Prior to its inherent optimized behavior and touchable resulted formula, applications of GP are being increased in simulation and prediction of some nonlinear phenomena in engineering, such as rainfall-runoff, electricity demand, etc. (Ashour, Alvarez, & Toropov, 2003; Babovic & Bojkov, 2001; Babovic & Keijzer, 2000, 2003; Drecourt, 1999; Harris, Babovic, & Falconer, 2003; Keijzer & Babovic, 2002; Khu, Liong, Babovic, Madsen, & Muttil, 2001; Lee, Lee, & Chang, 1997; Muttil & Lee, 2005; Sivapragasam, Vincent, & Vasudevan, 2007; Whigham & Crapper, 2001; Zhang, Jack, & Nandi, 2005).

In this paper, GP is applied to a water demand prediction and simulation problem. The major advantage of this approach is the possibility to achieve the shortest possible nonlinear and deterministic mathematical formula for monthly water demand prediction via evolutionary method.

The prediction is based on the previous lags of water demand observation, which helps to predict the water demand without any additional input variables.

The best created model can be accepted as an autoregressive approach in water demand simulation. Regarding to correction of the model outputs based on the water demand observation, the model results are coupled with KF for dynamic monthly water demand simulation. In the literature in question, traditionally online forecasting is based on KF and its familiar methods.

The focus of this paper is on using hybrid model which combines GP and EKF for monthly water demand forecasting. In Section 2, methods and materials included short explanation of GP and KF. In Section 3, time series modeling with GP and application of KF is presented. Results and concluding remarks are in Sections 4 and 5.

## 2. Methods and materials

### 2.1. Genetic programming

GP is a symbolic regression method based on a tree-structured approach presented by Koza (1990). This method belongs to a branch of evolutionary method, which mimics the natural process of struggle for existence (Holland, 1975). In fact, GP as a member of the genetic algorithm-based methods is applied to a population of computer programs for achieving better compatibility (Sette & Boullart, 2001).

In this method, strings and numbers (binary or real) are replaced by computational codes in the form of their tree-structured formula and most of the Darwinian operators such as selection, crossover and mutation operate on them.

Figs. 1 and 2, present the schematic process of crossover and mutation in GP. Basically, GP creates computer programs which are consisted of symbolic expression of variables, terminal and several mathematical operators (functions) in the LISP language or other computer languages. However, in GA, all chromosomes have equal lengths and genes.

The key advantage of GP is that it does not assume any priori function for the problem's solution. As a statistical criterion for compatibility of observed and computed outputs, the form of objective function in GP has not a wide range.

In other words, in GP, the building blocks (the terminal and function set) are defined initially, and the learning method subse-
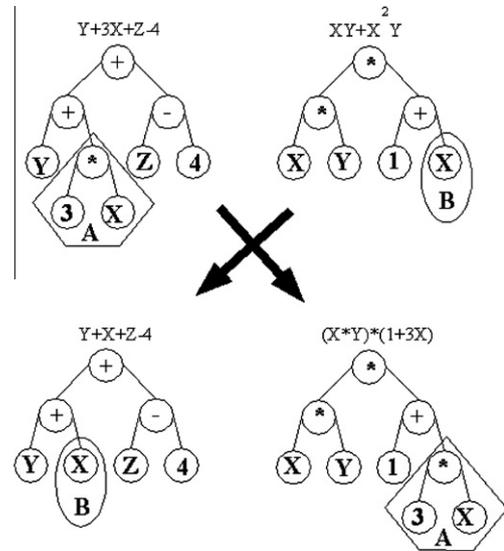


**Fig. 1.** Typical cross-over between parents, and creating new generation in GP.

quently finds its both optimal mathematical form and coefficients. Subsequent analysis of GP models may even provide additional insights into the problem at hand. Since GP evolves an equation relating the output and input variables, a major advantage of GP approach to data modeling is its automatic ability to select input variables that contribute beneficially to the model and discards the other ones. A more comprehensive presentation of GP may be found in Babovic and Abbott (1997a, 1997b). Briefly four major steps of simulation which are implemented in GP are as follows:

1. *Initialization*, creating population of the first generation randomly.
2. *Selection*, the programs which are generated randomly with the higher fitness will be selected and transferred to the next generation. There are various different paradigms in selection such as roulette-wheel, tournament, and ranking method.
3. *Preparation for the next generation*, this step includes two stages. The first stage is transferring the two best programs to the next generation in order to create new program in the next generation, and the last one is replacing the worst program(s) in the population with the transferred best programs. The best programs will be unchanged up to the end.
4. *Iteration*, going back to the second step and continuing up to the halting criteria.

In Fig. 3, flowchart of GP procedure is depicted and its similarities with GA are clearly inferable. GPLAB (version 1.1), a free downloadable toolbox of GP written in MATLAB by Silva (2002), is used for this research.

### 2.2. Extended Kalman Filter (EKF) and its application

Kalman Filter is introduced and developed by Kalman (1960), and it is known as the most prominent adaptive method of state variables and data assimilation theme. The KF is the minimum variance estimation in linear systems within the realm of stochastic dynamic mode. The method consecutively estimates the state variables of models after measuring them. There are some modifications on KF such as Extended Kalman Filter (EKF) and Unscented Kalman Filter (UKF) for nonlinear space-states. However, the KF method requires linearity and Gaussian conditions of the underlying procedures.