Computers & Geosciences 37 (2011) 1692–1701

Contents lists available at ScienceDirect

Computers & Geosciences

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

Comparison of genetic programming with neuro-fuzzy systems for predicting short-term water table depth fluctuations

Jalal Shiri [a,*], Özgur Kişi [b]

[a] Water Engineering Department, Faculty of Agriculture, University of Tabriz, Tabriz, Iran
[b] Civil Engineering Department, Hydraulics Divisions, Erciyes University, Kayseri, Turkey

ARTICLE INFO

Article history:
Received 13 June 2010
Received in revised form
31 October 2010
Accepted 5 November 2010
Available online 10 December 2010

Keywords:
Groundwater depth fluctuation
Genetic programming
Neuro-fuzzy
Forecast

ABSTRACT

This paper investigates the ability of genetic programming (GP) and adaptive neuro-fuzzy inference system (ANFIS) techniques for groundwater depth forecasting. Five different GP and ANFIS models comprising various combinations of water table depth values from two stations, Bondville and Perry, are developed to forecast one-, two- and three-day ahead water table depths. The root mean square errors (*RMSE*), scatter index (*SI*), Variance account for (*VAF*) and coefficient of determination (R^2) statistics are used for evaluating the accuracy of models. Based on the comparisons, it was found that the GP and ANFIS models could be employed successfully in forecasting water table depth fluctuations. However, GP is superior to ANFIS in giving explicit expressions for the problem.

1. Introduction

Physical-based numerical groundwater flow models are powerful tools for representing high spatial and temporal variations of aquifers. However, this capability renders the models data intensive, and to achieve acceptable simulations and prediction performance, the properties and conditions of the groundwater system must be accurately presented within the model's space and time domains (Coppola et al., 2003; Feng et al., 2008). Because the properties and conditions of groundwater can never be ascertained with absolute accuracy, unavoidable discrepancies between the model and the real-world system reduce simulation accuracy hinders efforts to appropriately manage the groundwater resources (Coppola et al., 2005). Therefore, empirical models may be considered as alternative methods and can provide useful results without costly calibration time (Daliakopoulos et al., 2005; Box and Jenkins, 1976; Hipel and Mc Leod, 1994). However, these models have their own limitations, because they are data demanding models and they are not adequate when the dynamical behavior of the hydrological system changes in time (Bierkens, 1998).

In the recent past, the use of Artificial Intelligence techniques, such as Genetic Programming (GP), Adaptive Neuro-Fuzzy Inference System (ANFIS) and Artificial Neural Networks (ANNs) have become viable: Coulibaly et al. (2001) applied ANNs for modeling of monthly groundwater level fluctuations; Coppola et al. (2005) developed ANNs for accurately predicting potentiometric surface elevations; Daliakopoulos et al. (2005) applied ANN for forecasting groundwater level; Szidarovszky et al. (2007) introduced a hybrid ANNs-numerical model for groundwater problems; Coppola et al. (2007) applied a combination of ANN modeling with multi-objective optimization for a complicated real-world groundwater management problem in New Jersey; and Feng et al. (2008) applied ANNs to investigate the effects of human activities on regional groundwater levels; Yang et al. (2009) applied ANN for forecasting groundwater levels in Western Jilin Province, China.

The focus of the current paper is on the application of GP and ANFIS data driven models to forecast groundwater table depth time series. The methodology of GP was first proposed by Koza (1992), as a generalization of Genetic Algorithms (GA) (Goldberg, 1989). The fundamental difference between GP and GAs lie in the nature of individuals, where in GAs individuals are linear strings of fixed length (as chromosomes), while in GP individuals are nonlinear entities of different sizes and shapes (as parse trees). Major advantages of GP are that it can be applied to areas where (a) the interrelationships among the relevant variables are poorly understood (or where it is suspected that the current understanding may well be less than satisfactory), (b) finding the ultimate solution is hard, (c) conventional mathematical analysis does not, or cannot, provide analytical solutions, (d) an approximate solution is acceptable (or is the only result that is ever likely to be obtained), (e) small improvements in the performance are routinely measured (or easily measurable) and highly valued, and (f) there is a large amount of data, in computer readable form, that requires examination, classification, and integration (such as satellite observations) (Banzhaf et al., 1998). Also effective data driven neuro-fuzzy models have received more attention in the

* Corresponding author. Tel.: +984113392775; fax: +984113345332.
E-mail address: j_shiri2005@yahoo.com (J. Shiri).

0098-3004/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.cageo.2010.11.010

recent past. ANFIS was firstly introduced by Jang (1993), Jang and Sun (1995) and Jang et al. (1997), and later on widely applied in engineering problems. Jang (1993) introduced architecture and a learning procedure for the Fuzzy-Inference Systems (FIS) that uses a neural network learning algorithm for constructing a set of fuzzy if-then rules with appropriate membership functions (MFs) from the specified input–output pairs. This procedure is called an adaptive network-based-fuzzy inference system (ANFIS). There are largely two approaches for fuzzy inference systems, namely the approaches of Mamdani (Mamdani and Assilian, 1975) and Sugeno (Takagi and Sugeno, 1985). The differences between the two approaches arise from the consequent part. Mamdani's approach uses fuzzy membership functions, whereas Sugeno's approach uses linear or constant functions. The neuro-fuzzy model used in this study implements Sugeno's fuzzy approach (Takagi and Sugeno, 1985) to obtain the values for the output variable from those of input variables. For a given input–output data set, various Sugeno models may be developed by using different identification methods (i.e., grid partitioning, subtractive clustering and Gustafson–Kessel clustering methods). However, the recent researches demonstrated that the type of identification method does not affect the results rigorously (Vernieuwe et al., 2005). Therefore, the commonly used grid partitioning identification method was applied for constructing the neuro-fuzzy models in this paper. The grid partitioning method proposes independent partitions of each antecedent variable through defining the membership functions of all antecedent variables. A major problem with application of this method is that the construction of the membership functions of each variable is not dependent on each other, hence the relationship between the variables is omitted.

One of the strong points of using GP over other data driven techniques (e.g., ANFIS) is that it can produce explicit formulations (model expression) of the relationship that rules the physical phenomenon. Such expressions may be subject to some physical interpretations. Actually, the comprehensibility of GP models is also a way to reduce the risk of over-fitting to training data and improve generalization of resulting models. In this way, one may perform knowledge discovery using GP, finding some confirmation of well-known physical relationships and evolving interesting new formulae, as an upgrading of particular cases of study.

Review of all of the applications of GP and ANFIS in hydrology and water resources engineering is beyond the scope of this paper and only some limited studies are discussed here. Babovic et al. (2002) applied GP for modeling of risks in water supply. Aytek and Alp (2008) applied GP to rainfall-runoff modeling. Aytek and Kisi (2008) applied GP to suspended sediment transport streams. Ghorbani et al. (2010) applied GP to forecast averaged sea water level values. Kisi and Shiri (2010) applied GP and ANFIS techniques for predicting short-term and long term river flow.

Kisi (2005) estimated suspended sediment using neuro-fuzzy and neural network approaches. Kisi (2006) proposed a neuro-fuzzy computing technique for daily pan evaporation modeling. Partal and Kisi (2007) proposed a new wavelet-neuro-fuzzy conjunction model for precipitation forecast. Kisi (2009) applied evolutionary fuzzy models for river suspended sediment concentration estimation.

To the best knowledge of the authors, no study has been carried out to predict groundwater table fluctuations using GP and ANFIS. This provides an impetus for the current work. The aim of this study is the application and comparison of GP and ANFIS for forecasting short-term daily groundwater table depths. It is relevant to remarked that the models investigated here are normally applied within deterministic frameworks in professional practices, which has encouraged the practice of comparing the actual with predicted values. However, this is a black-and-white approach for selecting the merits of a method and does not necessarily measure the impact on the decision.

2. Specification of the study

2.1. Used data

The data set used in this study was obtained from Illinois State Water Survey, U.S (www.isws.illinois.edu/data.asp). The time series of daily depth to water table records from two wells were used: Bondville (station no: 421832; FIPS code: 019; Latitude: 40°05′N; Longitude: 88°37′W; Altitude: 213 m) and Perry (station no: 421843; FIPS Code: 149; Latitude: 39°80′N; Longitude: 90°83′W; Altitude: 213 m). Groundwater levels are monitored continuously with Stevens Type-F paper chart recorders. The initial time series data of water table depth were obtained at daily intervals, but different disciplines require the processing and applying of data at various time intervals, according to the degree of desirability and necessity for different applications. For instance, daily groundwater depth data are important in irrigation scheduling in arid and semi-arid region, where the water is scarce, especially in the period when the water consumptive use of plants are high. Obviously, for making use of groundwater as subsurface-irrigation modeling input, one should have some meteorological data to get a more capable real-time and trustworthy forecasting model, but these data were not available to the authors. It was therefore decided to compare the GP with ANFIS for short-term (i.e., one-, two- and three-day ahead) water table depth fluctuations at a number of designated time intervals for the time series, using some input combinations. The water table data of September 01, 2001–August 30, 2008 were employed for training and testing of GEP and ANFIS models. For each well, the first five years data were used to train the models and the remaining data were used for testing. The periods from which training and testing data were chosen span the same temporal seasons (September–August). The daily statistical parameters of the water table data are given in Table 1. The data of Perry Well show more scattered distribution than those of the Bondville Well (see the C_{sx} values in Table 1). In the training data of the Bondville Well, minimum and maximum values fall in the range 0.37–9.89 m. However, the minimum of the testing data of the Bondville Well is 0.06 m, which is lower than the corresponding training set's value. This may cause some extrapolation difficulties in prediction of minimum values (Kisi, 2007). Fig. 1 represents the observed depth-to-water table data for both Bondville and Perry wells. From the figure it is understood that the water table suffers more fluctuations in Perry well. The high depth-to-water table values in the figure are corresponded to dry season, while the low values show the wet periods. It can be also concluded from the figure that there are some memory in the Bondville system, while finding a general trend and memory for Perry well data is somewhat difficult. However, there can be seen some memory in test period for Perry well. The auto-correlation functions of the water table depths for the (a) Bondville Well and (b) Perry Well are shown in Fig. 2. It is clear from this figure that the auto-correlation values of the water table depths are significantly high for both wells.

Table 1
Daily statistical parameters of each well data set (The statistical parameters presented are X_{mean}, X_{max}, X_{min}, S_d, C_v and C_{sx} denote the mean, maximum, minimum, standard deviation, coefficient of variation and skewness, respectively).

Data set	Well	Statistical parameter					
		X_{mean}	X_{max}	X_{min}	S_d	C_v	C_{sx}
Training	Bondville (421832)	5.04	9.89	0.37	2.28	0.45	0.16
	Perry (421843)	9.02	21.38	0.04	6.01	0.66	0.49
Testing	Bondville (421832)	4.75	9.81	0.06	2.56	0.54	0.34
	Perry (421843)	8.63	19.52	0.24	6.30	0.73	0.65