

A genetic programming technique for lake level modeling

Ali Aytek, Ozgur Kisi and Aytac Guven

ABSTRACT

The potential of gene expression programming (GEP) approach for modeling monthly lake levels is investigated. The application of the methodology is presented for the monthly water level data of Van Lake, which is the biggest lake in Turkey. The root mean square errors, mean absolute relative errors, determination coefficient, and modified coefficient of efficiency (E_M) are used for evaluating the accuracy of the genetic programming-based models. The results of the proposed models are compared with those of the neuro-fuzzy models. The comparison results indicate that the suggested GEP-based models perform better than the neuro-fuzzy models in forecasting monthly lake levels.

Key words | gene expression programming, lake level, neuro-fuzzy, soft computing

Ali Aytek
Sahinbey Municipality,
27310 Gaziantep,
Turkey

Ozgur Kisi (corresponding author)
Civil Engineering Department,
Canik Basari University,
Hydraulics Division, 55000 Samsun,
Turkey
E-mail: okisi@basari.edu.tr

Aytac Guven
Civil Engineering Department,
Gaziantep University,
Hydraulics Division, 27310 Gaziantep,
Turkey

INTRODUCTION

Lake level fluctuations are important for planning, design, construction, and operation of lakeshore structures and also in the management of fresh water lakes for water supply purposes. It is necessary to develop models for simulation of the extreme or abnormal level variations in order to control future lake level changes. Level measurements or their future equally likely replicas obtained through a simulation model are a direct way of obtaining lake management decision variables. Although it is possible to establish sophisticated models taking into consideration the hydrological and hydro-meteorological variables, such as the rainfall, runoff, temperature, and evaporation, it is economically preferable if a model that simulates the level variations on the basis of past level records is available to the decision-maker whether he/she be administrator, local authority, or a technical operator (Sen *et al.* 2000).

Genetic programming (GP) can be successively applied to areas where: (i) the interrelationships among the relevant variables are poorly understood (or where it is suspected that the current understanding may well be wrong); (ii) finding the size and shape of the ultimate solution is hard and a major part of the problem; (iii) conventional mathematical analysis does not, or cannot, provide analytical solutions; (iv) an approximate solution is acceptable (or is the only

result that is ever likely to be obtained); (v) small improvements in performance are routinely measured (or easily measurable) and highly prized; (vi) there is a large number of data, in computer readable form, that requires examination, classification, and integration (such as molecular biology for protein and DNA sequences, astronomical data, satellite observation data, financial data, marketing transaction data, or data on the World Wide Web) (Banzhaf *et al.* 1998).

In the last decades, it was observed that the genetic expression programming (GEP) was successfully applied in the field of water resources engineering (Cousin & Savic 1997; Savic *et al.* 1999; Babovic & Keijzer 2002; Harris *et al.* 2003; Rabunal *et al.* 2007; Aytek & Kisi 2008; Guven & Aytek 2009; Rodríguez-Vazquez *et al.* 2012; Fallah-Mehdipour *et al.* 2013; Nourani *et al.* 2013; Shiri *et al.* 2013). Cousin & Savic (1997), Savic *et al.* (1999), Drecourt (1999), Whigham & Crapper (1999, 2001), Babovic & Keijzer (2002) applied GEP to rainfall-runoff modeling. Babovic *et al.* (2001) applied GEP to sedimentary particle settling velocity equations. Harris *et al.* (2003) studied velocity predictions in compound channels with vegetated floodplains using GEP. Dorado *et al.* (2003) studied prediction and modeling of the rainfall-runoff transformation of a typical urban

basin using artificial neural networks (ANNs) and GEP. Giustolisi (2004) determined Chezy resistance coefficient in corrugated channels using GEP. Rabunal *et al.* (2007) determined the unit hydrograph of a typical urban basin using GEP and ANNs. Babovic (2000) used experimental flume data utilized by Zyserman & Fredsoe (1994) and expressed a new formulation for bed concentration of suspended sediment. Kizhisseri *et al.* (2005) used GEP methodology to explore a better correlation between the temporal pattern of fluid field and sediment transport by utilizing two data sets; one from numerical model results and other from Sandy Duck field data. Aytek & Kisi (2008) used the GEP approach for daily suspended sediment modeling. Recently, Guven *et al.* (2008) applied GEP for estimation of reference evapotranspiration and Guven & Aytek (2009) applied GEP for stage–discharge relationship. Shiri & Kisi (2010) compared the GEP to neuro-fuzzy models in predicting short-term groundwater table depth fluctuations. Azamathulla & Ahmad (2012) used GEP for computing transverse mixing coefficients. Shiri *et al.* (2013) used GEP for estimating daily evapotranspiration values in northern Spain. The present study is the first study indicating the input–output mapping capability of GEP technique in modeling of lake levels.

Recently, fuzzy logic was successfully used in modeling lake levels. Altunkaynak & Sen (2007) used fuzzy logic for forecasting lake levels. Sener & Morova (2011) compared fuzzy and linear regression models in modeling lake level fluctuations. Yazar *et al.* (2009) compared neuro-fuzzy (NF) and ANN models for modeling lake level changes and found that the NF model performed better than the ANN model. Guldal & Tongal (2012) compared recurrent ANN (RANN), NF, and stochastic models in lake level forecasting and found that the RANN and NF models showed similar accuracy and they performed better than the autoregressive (AR) and autoregressive moving average (ARMA) models.

The purpose of this study is to develop a mathematical model for estimation of monthly lake levels based on GEP. The GEP technique is used to forecast the surface water level of Van Lake, which is situated in eastern Turkey. The results obtained using the GEP models are compared with those of the NF models.

METHODOLOGY

Gene expression programming

GEP is an extension to GP that evolves computer programs of different sizes and shapes encoded in linear chromosomes of fixed length (Ferreira 2001a, b). There are only two main players in GEP: the chromosomes and the expression trees (ETs), the latter consisting of the expression of the genetic information encoded in the former. The process of information decoding (from the chromosomes to the expression trees) is called translation; and this translation implies obviously a kind of code and a set of rules. The genetic code is very simple: a one-to-one relationship between the symbols of the chromosome and the functions and terminals they represent (Figure 1). The rules are also quite simple: they determine the spatial organization of the functions and terminals in the expression trees and the type of interaction between sub-expression trees in multi-genic systems (Ferreira 2006).

The fundamental step of the GEP process begins with the random generation of the chromosomes of a certain number of individuals (the initial population). These chromosomes are then expressed and the fitness of each individual is evaluated against a set of fitness cases (also called selection environment which, in fact, is the input to a problem). The individuals are then selected according to their fitness (their performance in that particular environment) to reproduce with modification, leaving progeny with new traits. These new individuals are, in their turn, subjected to the same developmental process: expression of the genomes, confrontation of the selection environment, selection, and reproduction with modification. The process is repeated for a certain number of generations or until a good solution has been found (Ferreira 2006). The reproduction, sequentially, consists of replication, mutation, inversion, insertion sequence (IS) transposition, root IS (RIS) transposition, gene transposition, one-point recombination, two-point recombination, and gene recombination.

There are five major steps in preparing to use GEP, of which the first is to choose the fitness function. The fitness of an individual program i for fitness case j is evaluated in

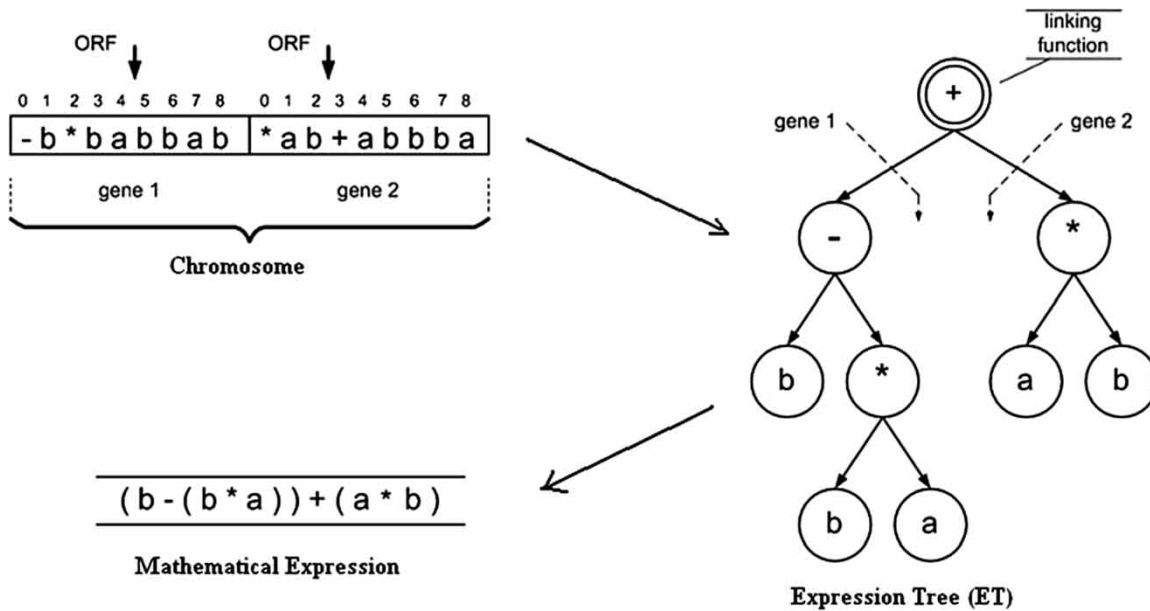


Figure 1 | Individual representations in GEP.

Ferreira (2006) by:

$$\text{If } E(ij) \leq p, \text{ then } f_{(ij)} = 1; \text{ else } f_{(ij)} = 0 \quad (1)$$

where p is the precision and $E(ij)$ is the error of an individual program i for fitness case j . For the absolute error it is expressed by:

$$E_{(ij)} = |P_{(ij)} - T_j| \quad (2)$$

Again for the absolute error, the fitness f_i of an individual program i is expressed by

$$f_i = \sum_{j=1}^n (R - |P_{(ij)} - T_j|) \quad (3)$$

where R is the selection range, $P_{(ij)}$ is the value predicted by the individual program i for fitness case j (out of n fitness cases), and T_j is the target value for fitness case j . So, for a perfect fit, $P_{(ij)} = T_j$ for all fitness cases and maximum fitness $f_{\max} = R_n$, where n is the number of fitness cases.

The second major step consists of choosing the set of terminals T and the set of functions F to create the chromosomes. In this problem, the terminal set consists obviously of

the independent variables, i.e., $T = \{L_{t-1}, L_{t-2} \dots\}$. Here, L_{t-1} , and L_{t-2} refer to the lake levels for months $t-1$ and $t-2$. The choice of the appropriate function set is not so obvious; however, a good guess can always be helpful in order to include all the necessary functions. In this study, four basic arithmetic operators (+, -, *, /) were utilized for model development.

The third major step is to choose the chromosomal architecture, i.e., the length of the head and the number of genes. Length of the head, $h = 8$, and two genes per chromosome were employed. The fourth major step is to choose the linking function. In this study, the sub-ETs were linked by addition. Finally, the set of genetic operators that cause variation and their rates are chosen.

CASE STUDY

The monthly lake level data of Van Lake, the largest lake in Turkey, are used in the study (see Figure 2). Van Lake is also the biggest soda lake in the world and the world's fourth closed basin lake with a volume of about 60 km^3 . The lake is located on the Anatolian high plateau in eastern Turkey near the border of Iran. It has a surface area of



Figure 2 | The location of Van Lake in Turkey (Sen et al. 2000).

3,713 km² and is more than 119 km across at its widest point. Its surface elevation from the mean sea level is 1,650 m and the deepest point is 457 m. The lake is surrounded by hills and mountains whose elevations reach 4,000 m above mean sea level. The Suphan Mountain, a volcanic mass, rises to about 4,434 m. Only limited species of fresh water fish can live in the lake and the water cannot be used for drinking or irrigation because of its high salinity. The lake is fed by rainfall, small rivers, and snowmelts. During winter months most precipitation falls as snow and the lake is at its lowest level. Towards the end of spring, heavy rainfalls occur and high runoff rates with the snowmelt from surrounding mountains result in more than 80% of the annual discharge from the catchment reaching the lake. The winter is very cold with frequent temperatures below 0 °C in this area for almost 3 months (December to February) of the year. The summer (July to September) is warm and dry and the average temperature is 20 °C. The variation of diurnal

temperature is about 20 °C. The lake has a drainage area of 12,500 km² and no natural outlets. On average, 4.2 km³ of water is lost annually from the lake surface to the atmosphere by evaporation. This loss is balanced by the long-term averages of annual precipitation and surface runoff amounts of 1.7 and 2.5 km³, respectively. Kadioglu et al. (1997) have shown that the fluctuations in lake level are entirely dependent on the natural variability of the hydrological cycle and the drainage basin is not affected by any longer-term climatic changes (Sen et al. 2000; Altunkaynak et al. 2003).

For Van Lake, 59 years (708 monthly levels) of lake level data between 1944 and 2002 are available. The observed data are for hydrologic years, i.e., the first month of the year is October and the last month of the year is September. After examining the data and noting the periods in which there were gaps in the lake level variable, the periods for calibration and test were chosen. The first 47 years of level data (560 months, 80% of the whole data set) were used for calibration and the remaining

Table 1 | The monthly statistical parameters of lake level data sets

Data set	x_{mean} (m)	S_x (m)	C_{sx} (m)	x_{min} (m)	x_{max} (m)	r_1	r_2	r_3	r_4
Training data	1,648	0.911	0.654	1,647	1,651	0.988	0.960	0.925	0.894
Test data	1,650	0.587	-0.228	1,648	1,651	0.987	0.957	0.921	0.888
Entire data	1,648	0.911	0.654	1,647	1,651	0.995	0.982	0.967	0.953

12 years (140 months, 20% of the whole data set) were used for testing. The monthly statistics of the data sets are given in Table 1. In this table x_{mean} , S_x , C_{sx} , x_{min} , x_{max} , r_1 , r_2 , r_3 , and r_4 denote the overall mean, standard deviation, skewness, lag-1, lag-2, lag-3, and lag-4 auto-correlation coefficients, respectively. The statistical properties of the training data set are different from those of the test data set. The lake level data have significantly high auto-correlations.

APPLICATION AND RESULTS

The root mean square errors (RMSE), mean absolute relative errors (MARE), determination coefficient (R^2), and modified coefficient of efficiency (E_M) statistics are used to evaluate the accuracy of GEP and NF models. The R^2 measures the degree to which two variables are linearly related. RMSE and MARE provide different types of information about the predictive capabilities of the model. The RMSE measures the goodness-of-fit relevant to high lake level values whereas the MARE yields a more balanced perspective of the goodness-of-fit at moderate lake levels (Karunanithi et al. 1994). Physically, the coefficient of efficiency, E , measures the differences between the observations and predictions relative to the variability in the observed data themselves. A value of 90% and above indicates a very satisfactory performance whereas a value below 80% indicates an unsatisfactory performance (Legates & McCabe 1999). The issue of oversensitivity of E

to extreme values (caused by squaring the difference terms), therefore, introduced a modified coefficient of efficiency, E_M , which uses the absolute differences, rather than their squares. The RMSE, MARE, and E_M are defined as:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n ((Y_{\text{observed}})_i - (Y_{\text{estimate}})_i)^2} \quad (4)$$

$$\text{MARE} = \frac{1}{n} \sum_{i=1}^n \frac{|(Y_{\text{observed}})_i - (Y_{\text{estimate}})_i|}{(Y_{\text{observed}})_i} 100 \quad (5)$$

$$E_M(\%) = 1 - \frac{\sum_{i=1}^n |((Y_{\text{estimate}})_i - (Y_{\text{observed}})_i)|}{\sum_{i=1}^n |((Y_{\text{estimate}})_i - Y_{\text{mean}})|} 100 \quad (6)$$

in which n is the number of data set, Y is the monthly lake level.

The correlation analyses of the data are employed for selecting appropriate input vectors in GEP- and NF-based lake level prediction models. The auto-correlation statistics and the corresponding 95% confidence bands from lag 0 to lag 10 are estimated for the lake level time series (Figure 3). It can be obviously seen from the figure that the lake level data have significant auto-correlations. The estimated partial auto-correlation statistics and corresponding 95% confidence limits between lag 0 and lag 10 are

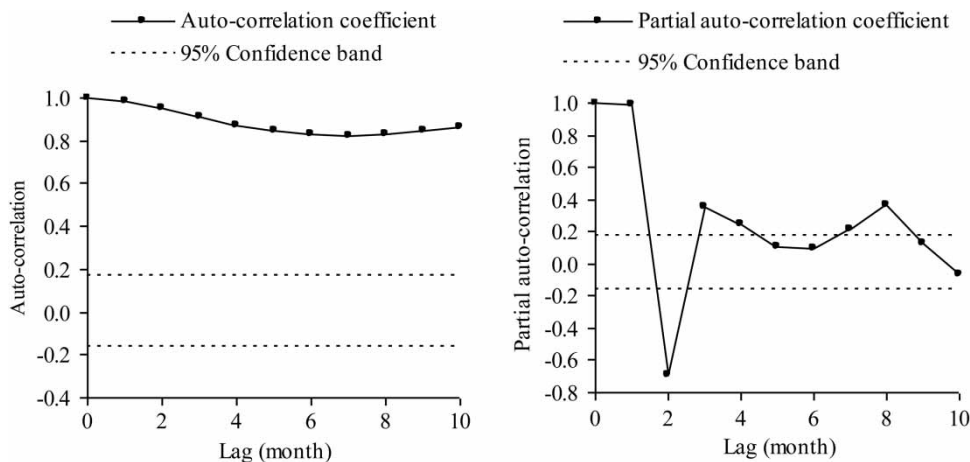


Figure 3 | The auto- and PACFs of the lake level time series.

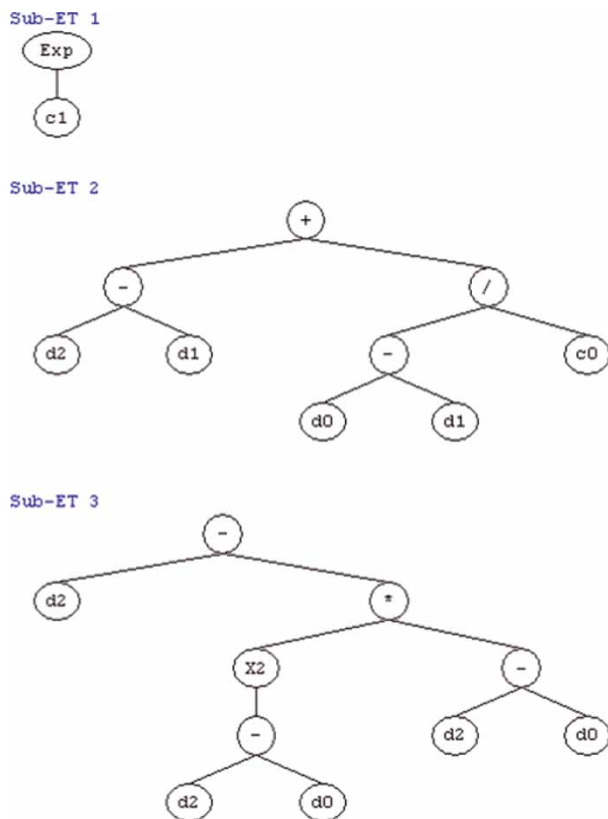


Figure 4 | Expression tree (ET) for GEP3 model (ET = Sub ET1 + Sub ET2 + Sub ET3).

Table 2 | The statistics of the NF models in test period

	Observed	NF1	NF2	NF3	NF4
Avr. (m)	1,649.57	1,649.4	1,649.4	1,649.4	1,649.3
Min. (m)	1,648.41	1,648.4	1,648.4	1,648.4	1,648.4
Max. (m)	1,650.53	1,650.1	1,649.9	1,650.0	1,649.8
St. Dev. (m)	0.585	0.479	0.422	0.454	0.374
RMSE (m)		0.189	0.279	0.228	0.342
MARE		0.0093	0.0135	0.0108	0.0159
R^2		0.972	0.976	0.972	0.952
E		0.976	0.988	0.990	0.991

also presented in Figure 4. The partial auto-correlation function (PACF) indicates significant correlation up to lag 4 and, thereafter, falls within the confidence limits. The rapid decaying pattern of the PACF confirms the dominance of the autoregressive process, relative to the moving average process. The auto- and partial auto-correlation coefficients suggest that incorporating lake level values of up to 4

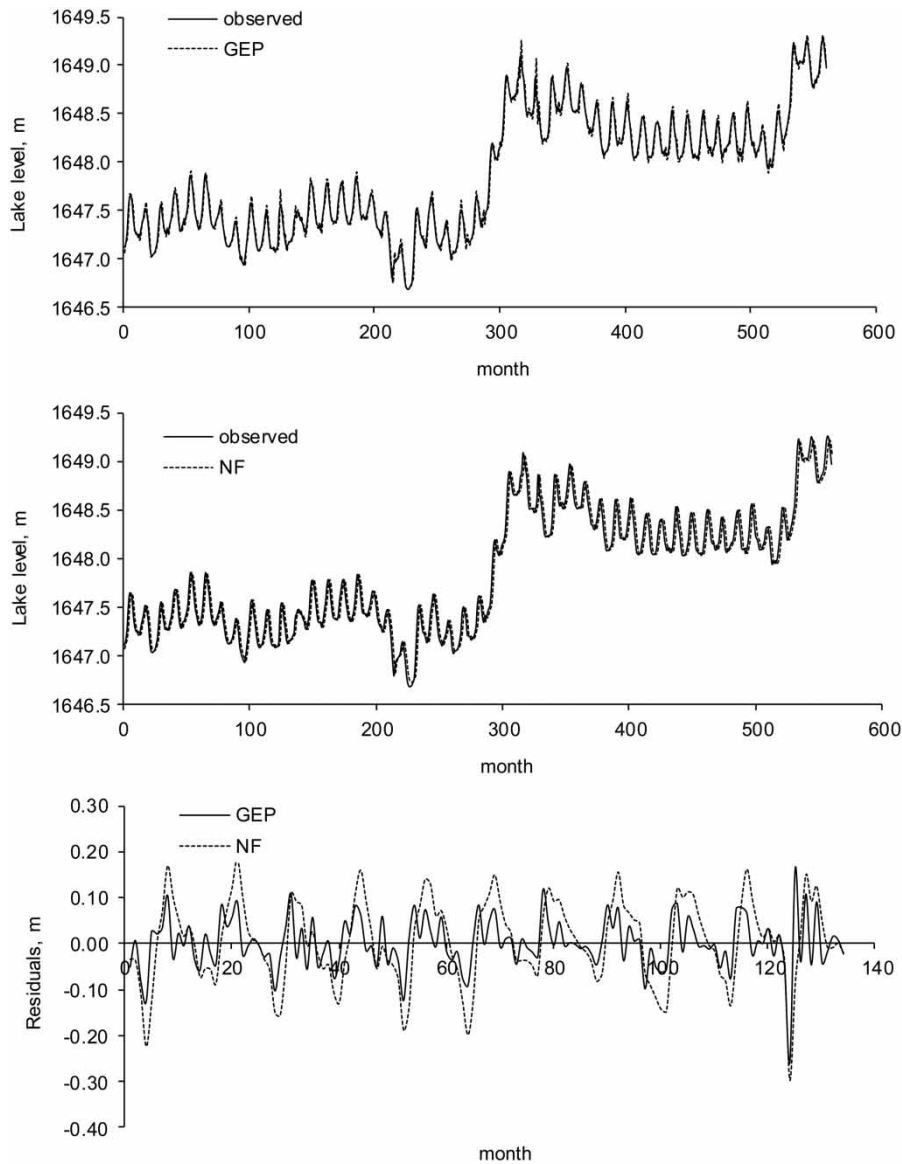
months' lag are sufficient as input vectors to the GEP- and NF-based models.

Considering the correlation analyses, four input combinations based on monthly lake levels of previous periods are evaluated to estimate current level value. The input combinations evaluated in the study are: (i) L_{t-1} , (ii) L_{t-1} and L_{t-2} , (iii) L_{t-1} , L_{t-2} , and L_{t-3} , and (iv) L_{t-1} , L_{t-2} , L_{t-3} , and L_{t-4} . ANFIS toolbox is used for the application of NF models. Different numbers of membership functions were tried for the NF models. Hybrid learning algorithm combining gradient descent and nonlinear least square algorithms is used for training NF models and 100 iterations are found to be enough for calibration of the models. The optimal models are found to have two Gaussian membership functions for each input (L_{t-1} , L_{t-2} , L_{t-3} , and L_{t-4}). The comparison statistics of the optimal NF models are provided in Table 2 for the test period. It is clear from the table that the NF1 model comprising L_{t-1} input performs better than the other models with respect to various comparison statistics used in the present study. The comparison statistics of the optimal GEP models are given in Table 3. The table indicates that the GEP3 and GEP4 models have almost the same accuracy and they perform better than the other GEP models. Among the GEP models, the GEP1 model whose input is the lake level of one previous month has the highest RMSE (0.093 m), MARE (0.0045%), and the lowest R^2 (0.975), E_M (0.991). The GEP3 model whose inputs are less than those of the GEP4 is selected for the comparison with the optimal NF1 model. Comparison of Tables 2 and 3 clearly reveals that the optimal GEP3 model performs better than the optimal NF1 model from the viewpoints of RMSE, MARE, R^2 , E_M , and other statistics.

The best generation individuals for GEP3 have been obtained as chromosomes: 30; genes: 3; head size: 6; gene size: 23; and the linking function: addition. The best fitness is 0.942. The GEP technique solves the problem of constant creation by using a special terminal named ephemeral random constant. For each ephemeral random constant used in the trees of the initial population, a random number of a special data type in a specified range is generated. Then, these random constants are moved around from tree to tree by the crossover operator (Figure 4). After putting the corresponding values, the final explicit

Table 3 | The comparison statistics of the GP and NF models in test period

	Observed	GEP1	GEP2	GEP3	GEP4
Avr. (m)	1,649.57	1,649.57	1,649.56	1,649.57	1,649.57
Min. (m)	1,648.41	1,648.41	1,648.38	1,648.40	1,648.40
Max. (m)	1,650.53	1,650.53	1,650.61	1,650.55	1,650.54
St. Dev. (m)	0.585	0.585	0.594	0.590	0.590
RMSE (m)		0.093	0.066	0.059	0.058
MARE		0.0045	0.0033	0.0028	0.0027
R^2		0.975	0.987	0.990	0.990
E_M		0.991	0.994	0.995	0.995

**Figure 5** | Time variation of lake level estimates by GEP and NF models in training period.

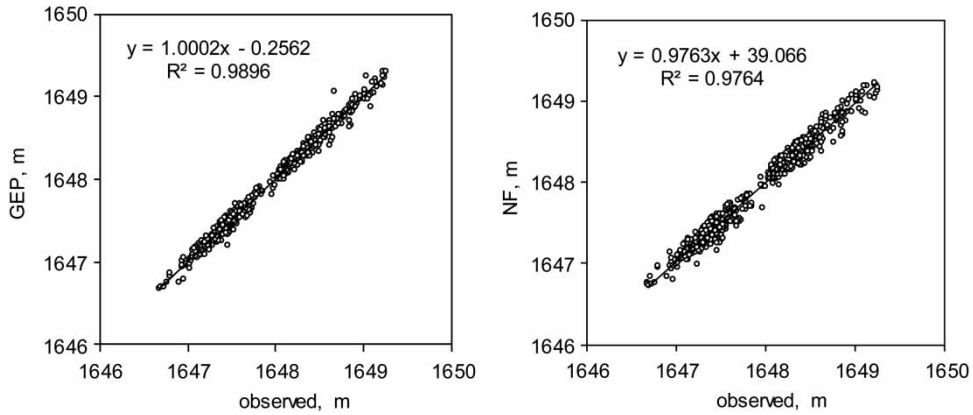


Figure 6 | Scatterplots of the lake level estimates by GEP and NF models in training period.

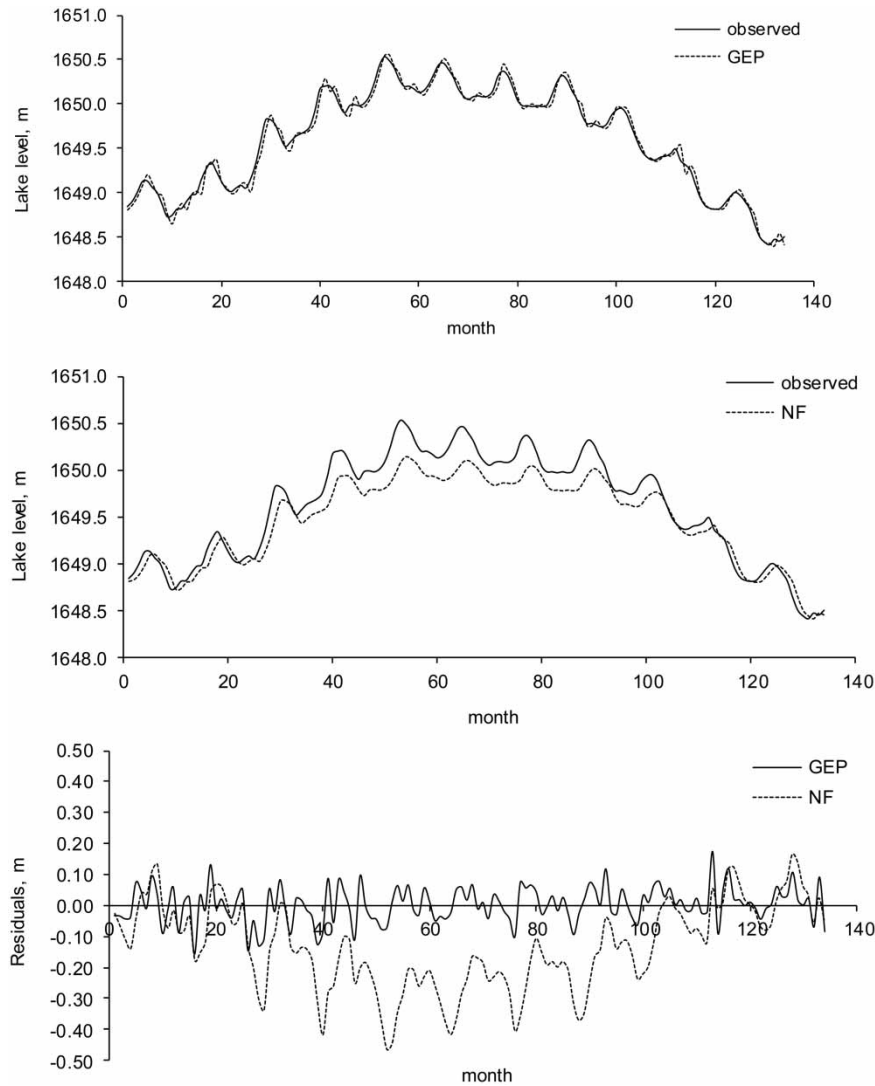


Figure 7 | Time variation of lake level estimates by GEP and NF models in test period.

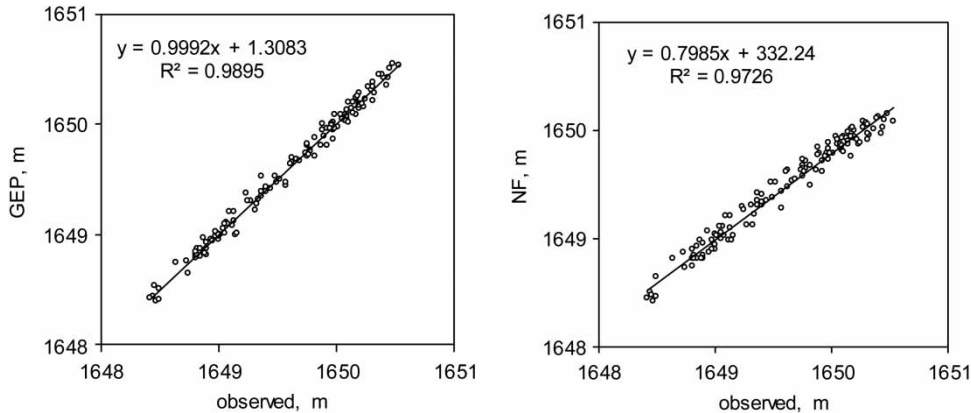


Figure 8 | Scatterplots of the lake level estimates by GEP and NF models in test period.

formulation of GEP3 based on the input variables (L_{t-1} , L_{t-2} , and L_{t-3}) and the models constants becomes:

$$L_t = (L_{t-1} - L_{t-3})^3 + L_{t-1} + 1.287(L_{t-1} - L_{t-1}) + 0.003 \quad (7)$$

The training performances of the GEP3 and NF1 models are illustrated in Figure 5. The residuals of the models are also provided in this figure. It can be seen especially from the residual graph that the GEP model approximates corresponding lake levels better than the NF. Figure 6 shows the scatterplots of the model estimates in the training period. It is obviously seen from the scatterplots that the GEP performs better than the NF model. The GEP and NF estimates in the test period are compared with the observed monthly lake levels in the form of a hydrograph in Figure 7. It can be seen from the hydrograph and residuals that the GEP model performs better than the NF especially for the extreme values. The over- and under-estimations are obviously seen for the NF model. Figure 8 denotes the estimates of GEP and NF models in the form of scatterplots. As can be seen from the fit line equations

(assume that the equation is $y = a_0x + a_1$) in the scatterplots, the a_0 and a_1 coefficients for the GEP model are respectively closer to the 1 and 0 with a higher R^2 value of 0.989 than those of the NF model. It can be obviously seen from the scatterplots that the estimates of GEP model are less scattered and closer to the exact line than those of the NF model.

The results are also tested by using one-way analysis of variance (ANOVA) and t -test for verifying the robustness (the significance of differences between the model estimates and observed values) of the models. Both tests are set at a 95% significant level. Namely, differences between observed and estimated values are considered significant when the resultant significance level (p) is lower than 0.05 by use of two-tailed significance levels. The statistics of the tests are given in Table 4. The GEP model respectively yields smaller testing values (0.0011 and 0.457) with a higher significance level (0.973) for the ANOVA and t -test than the NF model. According to the test results, the GEP model seems to be more robust (the similarity between the observed lake levels and GEP estimates are significantly high) in lake level modeling than the NF.

Table 4 | ANOVA and t -test for lake level prediction

	ANOVA F-statistic	t-test	
		t-statistic	Resultant significance level
GEP	0.0011	0.457	0.973
NF	4.234	10.94	0.041

CONCLUSIONS

This study indicates the ability of the GEP technique for modeling monthly lake levels. The GEP model is explicit and simple and can be used by anyone not necessarily being familiar with GEP. The lake level estimates obtained by using

GEP are compared with NF models. The GEP results are found to be better than those obtained using the NF models and confirm the ability of this approach to provide a useful tool in solving specific problems in hydrology, such as lake level modeling. The results suggest that the GEP approach may provide a superior alternative to the NF models.

ACKNOWLEDGEMENT

The third author would like to thank the Research Foundation of Gaziantep University for the support provided during this study.

REFERENCES

- Altunkaynak, A. & Sen, Z. 2007 Fuzzy logic model of lake water level fluctuations in Lake Van, Turkey. *Theor. Appl. Climatol.* **90** (3/4), 227–233.
- Altunkaynak, A., Ozger, M. & Sen, Z. 2003 Triple diagram model of level fluctuations in Lake Van, Turkey. *Hydrol. Earth Syst. Sci.* **7** (2), 235–244.
- Aytek, A. & Kisi, O. 2008 A genetic programming approach to suspended sediment modeling. *J. Hydrol.* **351** (3–4), 288–298.
- Azamathulla, H. Md. & Ahmad, Z. 2012 Gene-expression programming for transverse mixing coefficient. *J. Hydrol.* **434–435**, 142–148.
- Babovic, V. 2000 Data mining and knowledge discovery in sediment transport. *Comput.-Aided Civ. Infrastruct. Eng.* **15** (5), 383–389.
- Babovic, V. & Keijzer, M. 2002 Declarative and preferential bias in GP-based scientific discovery. *Genet. Program. Evol. Mach.* **3** (1), 41–79.
- Babovic, V., Keijzer, M., Aguilera, D. R. & Harrington, J. 2001 Automatic Discovery of Settling Velocity Equations, D2K Technical Report, D2K-0201-1.
- Banzhaf, W., Nordin, P., Keller, R. E. & Francone, F. D. 1998 *Genetic Programming*. Morgan Kaufmann, San Francisco, CA.
- Cousin, N. & Savic, D. A. 1997 A rainfall-runoff model using genetic programming, Centre for Systems and Control Engineering, Report No. 97/03, School of Engineering, University of Exeter, Exeter, UK, p. 70.
- Dorado, J., Rabunal, J. R., Pazos, A., Rivero, D., Santos, A. & Puertas, J. 2003 Prediction and modelling of the rainfall-runoff transformation of a typical urban basin using ANN and GP. *Appl. Artif. Intell.* **17**, 329–343.
- Drecourt, J. P. 1999 Application of Neural Networks and Genetic Programming to Rainfall-Runoff Modeling, D2K Technical Report 0699-1-1, Danish Hydraulic Institute, Denmark.
- Fallah-Mehdipour, E., Haddad, O. B. & Marino, M. A. 2013 Developing reservoir operational decision rule by genetic programming. *J. Hydroinf.* **15** (1), 103–119.
- Ferreira, C. 2001a Gene expression programming in problem solving. 6th Online World Conference on Soft Computing in Industrial Applications (invited tutorial).
- Ferreira, C. 2001b Gene expression programming: A new adaptive algorithm for solving problems. *Complex Syst.* **13** (2), 87–129.
- Ferreira, C. 2006 *Gene-Expression Programming; Mathematical Modeling by an Artificial Intelligence*. Springer, Berlin, Heidelberg, New York.
- Giustolisi, O. 2004 Using genetic programming to determine Chezy resistance coefficient in corrugated channels. *J. Hydroinf.* **6**, 157–173.
- Guldal, V. & Tongal, H. 2012 Comparison of recurrent neural network, adaptive neuro-fuzzy inference system and stochastic models in Eğirdir Lake level forecasting. *Water Resour. Manage.* **24** (1), 105–128.
- Güven, A. & Aytek, A. 2009 A new approach for stage-discharge relationship: Gene-expression programming. *J. Hydrol. Eng.* **14** (8), 812–820.
- Güven, A., Aytek, A., Yuce, M. I. & Aksoy, H. 2008 Genetic programming-based empirical model for daily reference evapotranspiration estimation. *CLEAN Water Soil Air* **36**, 905–911.
- Harris, E. L., Babovic, V. & Falconer, R. A. 2003 Velocity predictions in compound channels with vegetated floodplains using genetic programming. *Int. J. River Basin Manage.* **1** (2), 117–123.
- Kadioglu, M., Sen, Z. & Batur, F. 1997 The greatest soda-water lake in the world and how it is influenced by climatic change. *Ann. Geophys.* **15**, 1489–1497.
- Karunanithi, N., Grenney, W. J., Whitley, D. & Bovee, K. 1994 Neural networks for river flow prediction. *J. Comp. Civil Eng. ASCE* **8** (2), 201–220.
- Kizhisseri, A. S., Simmonds, D., Rafiq, Y. & Borthwick, M. 2005 An evolutionary computation approach to sediment transport modeling. *Fifth International Conference on Coastal Dynamics*, April 4–8, 2005, Barcelona, Spain.
- Legates, D. R. & McCabe, G. J. 1999 Evaluating the use of ‘goodness-of-fit’ measures in hydrologic and hydroclimatic model validation. *Water Resour. Res.* **35**, 233–241.
- Nourani, V., Komasi, M. & Alami, M. T. 2013 Geomorphology-based genetic programming approach for rainfall-runoff modeling. *J. Hydroinf.* **15** (2), 427–445.
- Rabunal, J. R., Puertas, J., Suarez, J. & Rivero, D. 2007 Determination of the unit hydrograph of a typical urban basin using genetic programming and artificial neural networks. *Hydrol. Process.* **21**, 476–485.
- Rodríguez-Vázquez, K., Arganis-Juárez, M. L., Cruickshank-Villanueva, C. & Domínguez-Mora, R. 2012 Rainfall-runoff modelling using genetic programming. *J. Hydroinf.* **14** (1), 108–121.
- Savic, A. D., Walters, A. G. & Davidson, J. W. 1999 A genetic programming approach to rainfall-runoff modeling. *Water Resour. Manage.* **13**, 219–231.

- Sen, Z., Kadioglu, M. & Batur, E. 2000 [Stochastic modelling of the Van Lake monthly level fluctuations in Turkey](#). *Theor. Appl. Climatol.* **65**, 99–110.
- Sener, E. & Morova, N. 2011 Determination of water level changing of Burdur Lake with fuzzy logic and linear regression analysis. *J. Natural Appl. Sci.* **15** (1), 60–66.
- Shiri, J. & Kisi, O. 2010 Comparison of genetic programming with neuro-fuzzy systems for predicting short-term water table depth fluctuations. *Comput. Geosci.* **37** (10), 1692–1701.
- Shiri, J., Nazemi, A. H., Sadraddini, A. A., Kisi, O., Landaras, G., Fard, A. F. & Marti, P. 2013 Evaluation of different data management scenarios for estimating daily reference evapotranspiration. *Hydrol. Res.* **44** (6), 1058–1070.
- Whigham, P. A. & Crapper, P. F. 1999 Time series modelling using genetic programming: an application to rainfall–runoff models. In *Advances in Genetic Programming 3*. (L. Spector, W. B. Langdon, U. O'Reilly & P. J. Angeline, eds). MIT Press, Cambridge, MA, pp. 89–104.
- Whigham, P. A. & Crapper, P. F. 2001 [Modeling rainfall-runoff using genetic programming](#). *Math. Comput. Model.* **33**, 707–721.
- Yarar, A., Onucyildiz, M. & Coptu, N. K. 2009 [Modelling level change in lakes using neuro-fuzzy and artificial neural networks](#). *J. Hydrol.* **365** (3–4), 329–334.
- Zyserman, J. A. & Fredsoe, J. 1994 [Data analysis of bed concentration of suspended sediment](#). *J. Hydraul. Eng. ASCE* **120** (9), 1021–1042.

First received 17 April 2013; accepted in revised form 13 August 2013. Available online 20 September 2013