Science, Technology and Development 34 (4): 265-269, 2015
ISSN 0254-6418 / DOI: 10.3923/std.2015.265.269

Forecasting Daily Precipitation Values, Using Wavelet Conjunction Models (Case Study: Tabriz and Marageh Stations, Iran)

Kazem Teimoorzadeh, Jalal Shiri and Abbas Ali Shiri
Sama Technical and Vocational Training College, Islamic Azad University,
Tabriz Branch, Tabriz, Iran

Abstract: Forecasting precipitation as a major component of the hydrological cycle is of primary importance in water resources engineering, planning and management as well as in scheduling irrigation practices. In this study, the abilities of hybrid wavelet-genetic programming [i.e., wavelet-gene-expression programming, WGEP] and wavelet-neural networks (WANN) models for daily precipitation forecasting were investigated. In the first step, the single GEP and ANN models were established to forecast daily precipitation values using the previously recorded values. The results present low accuracy. In the next step, the hybrid WGEP and WANN models were used by introducing the wavelet coefficients as GEP and ANN input vectors. The results showed that the hybrid WGEP and WANN models produced more accurate simulations than the single models.

Key words: Precipitation, gene expression programming, neural networks, discrete wavelet

INTRODUCTION

Forecasting precipitation as a major component of the hydrological cycle is of primary importance in water resources planning and management as well as in irrigation scheduling. Some numerical (Bustamante *et al*., 1999; Olson *et al*., 1995; Kisi and Shiri, 2011) and physical (Georgakakos and Bras, 1984) models for the quantitative prediction of rainfall have been developed but they are not very successful in predicting rainfall. On the other hand and in the context of emerging methodologies, the use of heuristic techniques [e.g., Gene Expression Programming (GEP) and Neural Networks (NN)] in water resources engineering has become viable.

Rabunal *et al*. (2007) applied GP and neural networks to determine the unit hydrograph of a typical urban basin. Also it has been applied in evapotranspiration modeling (Guven *et al*., 2008; Shiri *et al*., 2012, 2013, 2014), lake level forecasting (Kisi *et al*., 2012a), suspended sediment modeling (Kisi *et al*., 2012b; Kisi and Shiri, 2012a; Roushangar *et al*., 2014a), hydraulic structure modeling (Roushangar *et al*., 2014b) and rainfall-runoff simulations (Aytek *et al*., 2008; Kisi *et al*., 2013). Nevertheless, neuro-fuzzy models have been applied various hydrologic aspects, e.g., evaporation modeling (Baba *et al*., 2013; Shiri and Kisi, 2011) and air temperature forecasting (Kisi and Shiri, 2014). Also wavelet conjunction models have been applied for modeling various phenomena, including hydrological processes (Partal and Kisi, 2007; Partal and Kucuk, 2006; Kisi and Shiri, 2012b).

The present study aims at applying Wavelet-Gene Expression Programming (WGEP) and Wavelet-Neural Networks (WANN) models to forecast daily precipitation at two rain gauge stations in Iran.

MATERIALS AND METHODS

Gene expression programming: The GEP is, like genetic algorithms and genetic programming, a genetic algorithm as it uses populations of individuals, selects them according to fitness and introduces genetic variation, using one or more genetic operators (Ferreira, 2001a, b). The advantages of a system like GEP are clear from nature but the most important are: (i) The chromosomes are simple entities: linear, compact, relatively small, easy to manipulate genetically (replicate, mutate, recombine, etc), (ii) The expression trees are exclusively the expression of their respective chromosomes, they are entities upon which selection acts and according to fitness, they are selected to reproduce with modification.

The procedure to forecast precipitation is as follows:

- Selecting the fitness function
- Choosing the set of terminals, T and the set of functions, F, to create the chromosomes. In the current problem, the terminal set includes daily precipitation values. The choice of the appropriate function depends on the viewpoint of the user. In this study, different mathematical functions were utilized ({+, -, *, /},{$\sqrt{}$, $\sqrt[3]{}$, ln(x), e^x, x^2, x^3})

Corresponding Author: Kazem Teimoorzadeh, Sama Technical and Vocational Training College, Islamic Azad University, Tabriz Branch, Tabriz, Iran

- Choose the chromosomal architecture. Length of head, h = 7 and three genes per chromosomes were employed
- Choose the linking function, which was an "Addition" for this study
- Choose the genetic operators. The parameters used per run are summarized as follows:

Number of chromosomes: 30, head size: 7, number of genes: 3, linking function: addition, fitness function error type: root relative squared error, mutation rate: 0.044, inversion rate: 0.1, one point recombination rate: 0.3, two points recombination rate: 0.3, gene recombination rate: 0.1, gene transposition rate: 0.1, insertion sequence transposition rate: 0.1 and root insertion sequence transposition: 0.1.

Neural networks (ANN): An ANN has one or more hidden layers, whose computational nodes are called hidden neurons. The hidden neurons intervene between the external input and the network output in some useful manner. The detailed theoretical information about ANN can be found in (Haykin, 1998). In the present study, the ANN models were trained by using Levenberg-Marquardt technique because of the fact that this technique is more powerful and faster than the conventional gradient descent technique (Hagan and Menhaj, 1994).

Discrete Wavelet Transform (DWT): The wavelet function $\Psi(t)$, called the mother wavelet, can be defined as $\int_{-\infty}^{+\infty}\psi(t)dt = 0$. $\Psi_{a,b}(t)$ and can be obtained through compressing and expanding $\Psi(t)$ as follows:

$$\psi_{a,b}(t) = |a|^{-1/2}\psi\left(\frac{t-b}{a}\right) b\epsilon\ R, a\epsilon\ R, a \neq\ 0 \qquad (1)$$

Where:

$\Psi_{a,b}(t)$	=	The successive wavelet
a	=	The scale or frequency factor
b	=	The time factor
R	=	The domain of real numbers.

If $\Psi_{a,b}(t)$ satisfies Eq. 1 for the time series $f(t)\ \epsilon\ L^2(R)$ or finite energy signal, the successive wavelet transform of f(t) is defined as:

$$W_\psi f(a,b) = |a|^{-1/2}\int_R f(t)\bar{\psi}\left(\frac{t-b}{a}\right)dt \qquad (2)$$

where, $\overline{\Psi}(t)$ = complex conjugate functions of $\Psi(t)$. It can be seen from Equation (2) that the wavelet transform is the decomposition of f(t) at different resolution levels (scales). In other words, to filter wave for f(t) with different filters is the essence of wavelet transform.

The successive wavelet is often discrete in real applications. Let $a = a_0^j$, $b = kb_0\ a_0^j$, $a_0>1$, $b_0\ \epsilon$ R and k, j are the integer numbers. The discrete wavelet transform of f(t) can be written as:

$$W_\psi f(j,k) = a_0^{-j/2}\int_R f(t)\bar{\psi}\left(a_0^{-j}t - kb_0\right)dt \qquad (3)$$

The most common (and simplest) choice for the parameters a_0 and b_0 is 2 and 1 time steps, respectively. This power of two logarithmic scaling of time and scale is known as the dyadic grid arrangement and is the simplest and most efficient case for practical purposes (Mallat, 1989). Equation 3 becomes a binary wavelet transform when $a_0 = 2$, $b_0 = 1$.

$$W_\psi f(j,k) = 2^{-j/2}\int_R f(t)\bar{\psi}\left(2^{-j}t - k\right)dt \qquad (4)$$

The characteristics of the original time series in frequency (a or j) and time domain (b or k) are at the same time reflected by $W_\psi f(a,b)$ or $W_\psi f(j,k)$. When the frequency resolution of wavelet transform is low but the time domain resolution is high a or j becomes small. When the frequency resolution of wavelet transform is high but the time domain resolution is low a or j becomes large (Wang and Ding, 2003).

For a discrete time series f(t), when it occurs at different times t (i.e., here integer time steps are used), the DWT can be defined as:

$$W_\psi f(j,k) = 2^{-j/2}\sum_{t=0}^{N-1} f(t)\bar{\psi}\left(2^{-j}t - k\right) \qquad (5)$$

where, $W_\psi f(j,k)$ is the wavelet coefficient for the discrete wavelet of scale $a = 2^j$, $b = 2^j k$.

DWT operates two sets of function viewed as high-pass and low-pass lters. The original time series are passed through high-pass and low-pass lters and separated at different scales. The time series is decomposed into one comprising its trend (the approximation) and one comprising the high frequencies and the fast events (the detail). In the present study, the detail coefficients and approximation (A) sub-time series are obtained using Eq. 5.

Used data set: The daily rainfall values of two rain gauge stations, namely, the Tabriz and Marageh Stations located in the Northwestern of the Iran, are used in the present

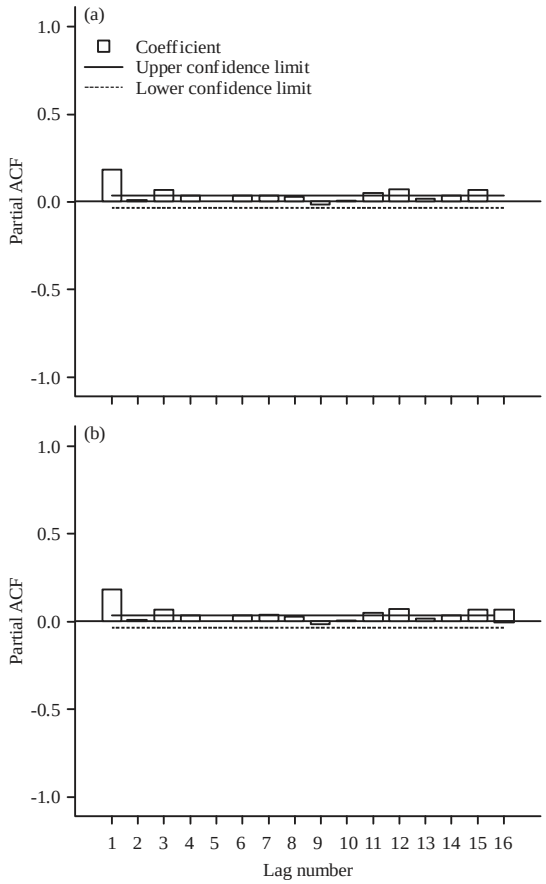

Fig. 1(a-b): PACF of rainfall data for (a) Tabriz and (b) Marageh stations

study. The observed data is 9 years (3281 patterns) long with an observation period between 2000 and 2008 for both stations.

In the applications, the first 7 years of precipitation data (2555 patterns) are used for training and the last 2 years data (726 patterns) were reserved for testing. Table 1 sums up the statistical parameters of the applied data set. In the table, the X_{mean}, X_{max}, X_{min}, S_x, C_v and C_{sx} stand for the mean, maximum, minimum, standard deviation, deviation and skewness coefficients, respectively. Figure 1 shows the Partial Auto Correlation Functions (PACF) of the rainfall data for the both studied stations. The figure clearly depicts that only introducing the precipitation values of one-immediate previous day will affect the forecast results, so the input configurations will be built accordingly.

Models performance assessment: Goodness of fit measures: Two statistical evaluation criteria were used to assess the models' performances: the root mean squared error (RMSE):

Table 1: Statistical parameters of the used data

Stations data	X_{mean}	X_{max}	X_{min}	S_x	C_V	C_{SX}
Tabriz						
Training	0.70	37	0	2.56	3.70	5.89
Testing	0.60	26	0	2.21	3.82	6.09
Whole data	0.60	37	0	2.44	3.81	6.24
Marageh						
Training	0.80	42	0	2.98	3.89	6.82
Testing	0.50	25	0	2.21	4.03	6.42
Whole data	0.70	42	0	2.88	4.11	7.39

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(P_{io} - P_{ie}\right)^2} \tag{6}$$

And the Scatter Index (SI):

$$\mathrm{SI} = \frac{\mathrm{RMSE}}{\overline{P}} \tag{7}$$

where, P_{io} and P_{ie} denote the observed and estimated precipitation values, respectively. $\overline{P}$denotes the mean observed precipitation value.

RESULTS AND DISCUSSIONS

This study aims to forecast daily precipitation values, using GEP, ANN and their hybrid models with wavelet transform. In the first step, several input combinations were introduced to the GEP and ANN models to forecast precipitation. The inputs of the applied models present the previously recorded daily precipitation values whereas the output corresponds to the precipitation value at the current time (t). As mentioned, the configurations were selected based on the partial auto correlation functions (PACF), which have been shown in Fig. 1: (i) P_t, (ii) P_{t-1}.

Table 2 represents the statistical measures of each GEP and ANN models for both stations during testing period. The results clearly show that the RMSE values are higher than the mean observed precipitation values (Table 1), explaining high values of deviation in the simulated values. Nevertheless, the SI values, describing the magnitude of RMSE which has occupied the average observed timer series, are quite high (generally greater than unity), that shows the insufficiency of GEP and ANN models capability in mapping the nonlinear precipitation phenomenon using the observed input and target values. Therefore, it is seen that the GEP and ANN models should be accompanied by some pore-processing coefficients, i.e. wavelet coefficients, to strengthen the models and improving the simulations. As mentioned in the previous section, the data were divided into training and testing periods and two new WGEP and WANN models were evaluated to forecast precipitation, as previously

Table 2: Testing statistics of the GEP and ANN models

Models	Input parameters	RMSE (mm)	SI
Tabriz			
GEP1	Pt	2.195	1.053
GEP2	Pt-1, Pt	2.184	0.992
ANN1	Pt	2.356	1.220
ANN2	Pt-1, Pt	2.684	1.280
Marageh			
GEP1	Pt	2.110	1.007
GEP2	Pt-1, Pt	2.562	1.018
ANN1	Pt	2.447	1.328
ANN2	Pt-1, Pt	2.897	1.489

Table 3: Testing statistics of the WGEP and WANN models during the test period

Models	Input parameters	RMSE (mm)	SI
Tabriz			
GEP1	Pt	0.987	0.358
GEP2	Pt-1, Pt	0.854	0.334
ANFIS1	Pt	1.025	0.407
ANFIS2	Pt-1, Pt	1.128	0.445
Marageh			
GEP1	Pt	1.002	0.403
GEP2	Pt-1, Pt	0.869	0.356
ANFIS1	Pt	1.134	0.419
ANFIS2	Pt-1, Pt	1.152	0.461

described. The WGEP and WANN models were developed by using Discrete Wavelet Transform (DWT). For the WGEP and WANN model inputs, the original time series were decomposed into a certain number of sub-time series components (Ds) by the Mallat DWT algorithm. Thus, the Ds of the original input time series were used as inputs of the WGEP and WANN and the original output time series were used as the output of the WGEP (or WANN). Three resolution levels were employed in the present study. In 3 decomposition levels, 3 details existed (2^1-day mode, 2^2-day mode and 2^3-day mode which is nearly a weekly mode) and one approximation signal.

Table 3 sums up the results of wavelet-GEP and wavelet-ANN models in the both Tabriz and Marageh stations, during the test period. From the table, it can be seen that the hybrid WGEP and WANN models can improve the modeling accuracy to a great extent. It is evident from the table that the WANN models perform relatively better than the WGEP models in the test period but the input combination in which the optimal models can be developed, are different for the hybrid models.

CONCLUSION

The accuracy of WGEP and WANN, to forecast daily precipitation was examined in the present study. In the first part of the study, single GEP and ANN models were established for forecasting daily precipitations. In the second part of the study, WGEP and WANN models were developed using the precipitation sub-time series. The sum of effective details and approximation component were used as inputs to the WGEP and WANN models. It was found that the wavelet conjunction models (WGEP and WANN) significantly increased the forecast accuracy of the single models.

REFERENCES

Aytek, A., M. Asce and M. Alp, 2008. An application of artificial intelligence for rainfall-runoff modeling. Earth Syst. Sci., 117: 145-155.

Baba, A.P.A., J. Shiri, O. Kisi, A.F. Fard, S. Kim and R. Amini, 2013. Estimating daily reference evapotranspiration using available and estimated climatic data by Adaptive Neuro-Fuzzy Inference System (ANFIS) and Artificial Neural Network (ANN). Hydrol. Res., 44: 131-146.

Bustamante, J., J. Gomes, S. Chou and J. Rozante, 1999. Evaluation of April 1999 Rainfall Forecast Over South America using the ETA Model. INPE., Cachoeria Paulista, SP., Brazil..

Ferreira, C., 2001a. Gene expression programming in problem solving. Proceedings of the 6th Online World Conference on Soft Computing in Industrial Applications, September 10-24, 2001, Lisbon, Portugal, pp: 1-22.

Ferreira, C., 2001b. Gene expression programming: A new adaptive algorithm for solving problems. Complex Syst., 13: 87-129.

Georgakakos, K.P. and L.R. Bras, 1984. A hydrologically useful station precipitation model: 1. Formulation. Water Resour. Res., 20: 1585-1596.

Guven, A., A. Aytek, M.I. Yuce and H. Aksoy, 2008. Genetic programming-based empirical model for daily reference evapotranspiration estimation. Clean-Soil Air Water, 36: 905-912.

Hagan, M.T. and M.B. Menhaj, 1994. Training feedforward networks with the Marquardt algorithm. IEEE Trans. Neural Networks, 5: 989-993.

Haykin, S., 1998. Neural Networks-A Comprehensive Foundation. 2nd Edn., Prentice-Hall, Upper Saddle River, NJ., USA, pp: 26-32.

Kisi, O. and J. Shiri, 2011. Precipitation forecasting using wavelet-genetic programming and wavelet-neuro-fuzzy conjunction models. Water Resour. Manage., 25: 3135-3152.

Kisi, O. and J. Shiri, 2012a. Wavelet and neuro-fuzzy conjunction model for predicting water table depth fluctuations. Hydrol. Res., 43: 286-300.

Kisi, O. and O. Shiri, 2012b. River suspended sediment estimation by climatic variables implication: Comparative study among soft computing techniques. Comput. Geosci., 43: 73-82.

Kisi, O. and J. Shiri, 2014. Prediction of long-term monthly air temperature using geographical inputs. Int. J. Climatol., 34: 179-186.

Kisi, O., A.H. Dailr, M. Cimen and J. Shiri, 2012a. Suspended sediment modeling using genetic programming and soft computing techniques. J. Hydrol., 450-451: 48-58.

Kisi, O., J. Shiri and B. Nikoofar, 2012b. Forecasting daily lake levels using artificial intelligence approaches. Comput. Geosci., 41: 169-180.

Kisi, O., J. Shiri and M. Tombul, 2013. Modeling rainfall-runoff process using soft computing techniques. Comput. Geosci., 51: 108-117.

Mallat, S.G., 1989. A theory for multiresolution signal decomposition: The wavelet representation. IEEE Trans. Pattern Anal. Mach. Intell., 11: 674-693.

Olson, D.A., N.W. Junker and B. Korty, 1995. Evaluation of 33 years of quantitative precipitation forecasting at the NMC. Weather Forecast, 10: 498-511.

Partal, T. and M. Kucuk, 2006. Long-term trend analysis using discrete wavelet components of annual precipitations measurements in Marmara region (Turkey). Phys. Chem. Earth Parts A/B/C, 31: 1189-1200.

Partal, T. and O. Kisi, 2007. Wavelet and neuro-fuzzy conjunction model for precipitation forecasting. J. Hydrol., 342: 199-212.

Rabunal, J.R., J. Puertus, J. Suarez and D. Rivero, 2007. Determination of the unit hydrograph of a typical urban basin using genetic programming and artificial neural networks. Hydrol. Proces., 21: 476-485.

Roushangar, K., F.V. Mehrabani and J. Shiri, 2014a. Modeling river total bed material load discharge using artificial intelligence approaches (based on conceptual inputs). J. Hydrol., 514: 114-122.

Roushangar, K., S. Akhgar, F. Salmasi and J. Shiri, 2014b. Modeling energy dissipation over stepped spillways using machine learning approaches. J. Hydrol., 508: 254-265.

Shiri, J. and O. Kisi, 2011. Application of artificial intelligence to estimate daily pan evaporation using available and estimated climatic data in the Khozestan province (South Western Iran). J. Irrig. Drainage Eng., 137: 412-425.

Shiri, J., A.A. Sadraddini, A.H. Nazemi, O. Kisi, P. Marti, A.F. Fard and G. Landeras, 2013. Evaluation of different data management scenarios for estimating daily reference evapotranspiration. Hydrol. Res., 44: 1058-1070.

Shiri, J., O. Kisi, G. Landeras, J.J. Lopez, A.H. Nazemi and L.C.P.M. Stuyt, 2012. Daily reference evapotranspiration modeling by using genetic programming approach in the Basque Country (Northern Spain). J. Hydrol., 414-415: 302-316.

Shiri, J., P. Marti and V.P. Singh, 2014. Evaluation of gene expression programming approaches for estimating daily evaporation through spatial and temporal data scanning. Hydrol. Process., 28: 1215-1225.

Wang, W. and J. Ding, 2003. Wavelet network model and its application to the prediction of the hydrology. Nat. Sci., 1: 67-71.