

Spatial rainfall prediction using optimal features selection approaches

Keyvan Asghari and Mohsen Nasseri

ABSTRACT

Rainfall as a semi-random hydrological event is difficult to forecast due to some very complicated and unforeseen physical factors and their chaotic behavior. Artificial neural networks (ANN), which perform a nonlinear mapping between inputs and outputs, have played a crucial role in rainfall forecasting. In this paper, some feature selection approaches have been implemented to simulate the regional scale rainfall field in order to address a few deficiencies of ANN, such as selection of informative features of input data encountered in hydrological processes. The main simulator is a multi-layer perceptron neural network optimized by simple genetic algorithm (GA) to determine optimal input vectors in order to compare with other statistical approaches. Current rainfall from a limited number of neighboring stations is shown to be valuable to forecast current rainfall of certain target stations in the province of Fars in Iran for 30 min leading time. Among the studied features selection approaches such as chi-squared, linear correlation coefficient and mutual information (MI), the results by MI have considerable competency with regard to computational efficiency using the optimized scenario by GA.

Key words | artificial neural networks, features selection, genetic algorithm, rainfall forecasting

Keyvan Asghari (corresponding author)
Department of Civil Engineering,
Isfahan University of Technology,
Isfahan,
Iran
E-mail: kasghari@cc.iut.ac.ir

Mohsen Nasseri
Department of Civil Engineering,
College of Engineering,
University of Tehran,
Tehran,
Iran

INTRODUCTION

Rainfall is one of the most important processes of the hydrologic cycle. Being able to forecast rainfall helps in making valuable decisions and performing strategic planning; yet, this is a complicated process largely due to the variability it displays over a wide range of scales both in time and space. Flash flooding, being a product of intense rainfall, is a life-threatening phenomenon and developing a rainfall forecasting and flood warning system can reduce the catastrophic consequences. Both internal and external characteristics of rainfall field depend on many factors including: pressure, temperature, wind speed and its direction, meteorological characteristics of the catchments, and so on. A physically based approach for rainfall forecasting has several advantages. However, given the short time scale, the small catchments area, and the massive costs associated with collecting the required data, it is not a feasible alternative in most cases as it involves many variables which are interconnected in a very complicated way. The

complexity and nonlinearity inherent in rainfall pattern makes it attractive to try data-driven models for simulation and forecasting purposes. Govindaraju (2000a, b) reported a number of studies which have used artificial neural networks (ANN) to forecast rainfall over a short time interval.

As a pioneering study of ANN application in rainfall forecasting, French *et al.* (1992) developed the first simulation scheme to forecast 1 h ahead, two-dimensional rainfall fields on a regular grid. The results of their investigation indicate that an ANN is quite capable of capturing the complex relationship associated with spatio-temporal evolution of rainfall inherent in a complex rainfall simulation model. Application of ANN with different network architectures for different data arrangements in hydrological and hydraulic modeling were continued in studies by Minns (2000), Toth *et al.* (2000), Luk *et al.* (2000a, b) and Ramirez *et al.* (2005), among others.

doi: 10.2166/nh.2014.178

The process of selecting an appropriate and relevant input vector in ANN modeling is an important step. The importance of input determination for ANN models were studied by Bowden *et al.* (2005a, b). Two methodologies, namely mutual information (MI) and self-organizing map (SOM), integrated with a genetic algorithm (GA) and general regression neural network, tested with synthetic data, were presented. The results indicated improvement in input selection by MI as it was able to exclude all insignificant inputs. Compared to MI, SOM determines the inputs in two steps: first it reduces the input dimension and in the second step it selects the subset of important model inputs using GA and regression network. Recently, various approaches were introduced for selecting the input variables in soft computing algorithms. Noori *et al.* (2011) evaluated the performances of three input selection techniques in order to reduce the number of input variables in support vector machines (SVM) modeling for predicting the monthly streamflow. Principal component analysis demonstrated to be superior over the other two techniques. Two new methodologies related to decision tree algorithms, namely M5 Model Tree and REPTree, studied by Senthil Kumar *et al.* (2012) for modeling sediment concentration. They proved that the M5 model outperforms the trial-and-error procedures such as ANN. In a rainfall–runoff modeling, Nourani & Parhizkar (2013) developed a wavelet-ANN model to decompose the rainfall and runoff time series into several sub-series and subsequently, the selected sub-series by a SOM method were introduced as input data to forecast the runoff.

Hong (2008) developed coupled recurrent ANN and SVM to solve nonlinear regression and rainfall time series prediction during typhoon periods from Northern Taiwan. Moreover, the chaotic particle swarm optimization algorithm has been employed to choose the SVM model parameters. SOM and the multilayer perceptron network (MLPN) have been combined by Lin & Wu (2009) to develop a new prediction model. Clustering, based on SOM, has been implemented on the dataset and MLPN has been used as the nonlinear regression technique to construct the relationship between the input and output data in each cluster. The proposed model precisely forecasts a typhoon rainfall in the Tanshui River Basin and was compared with the conventional neural network model.

In a previous study by Nasser *et al.* (2008), a model was developed to analyze small scale behavior of rainfall in a target station using surrounding stations by optimizing input parameters (stations and their lags) and the ANN architecture. They showed that the number of effective rain stations as input parameters decreases when combined use of ANN and GA is implemented. The effectiveness of cumulative data type simulation versus discrete modeling was also reported. The effective number of time lags and number of input stations were identified to reduce the training time significantly, compared to standard ANN. The results also showed that instead of using all available rain gauges, utilizing certain combinations will result in better forecasts. In a recent study by Kisi & Shiri (2012), a coupled wavelet and neuro-fuzzy approach were proposed as a new conjunction model to forecast short-term groundwater level. The combined model showed better performance compared to a simple neuro-fuzzy model.

In continuation of previous works, this paper investigates the effectiveness of some statistical data mining approaches in the realm of feature selection for reducing the computational cost of modeling. A strategy is developed to forecast a rainfall field at Zarghan and Shiraz recording rain gauges using surrounding gauges in a catchment located in Fars province (Figure 1). Since GA has been implemented in an enhanced-ANN model as a tool to optimize the parameters, it is also used here to have a benchmark for new proposed feature selection performances. The fitness function of GA is the error function of training dataset in ANN model (mean square error (MSE)), and the backward featured selection method is evaluated with results of mathematical sensitivity analysis. In the following sections, methodology of the modeling, description of study area including the data, results and discussion of modeling application, and concluding remarks are presented.

METHODOLOGY

The methodology of the current study is presented in four subsections as follows: brief description of ANN and GA coupling, the feature selection algorithms, sensitivity analysis procedures, and finally, the modeling procedure.

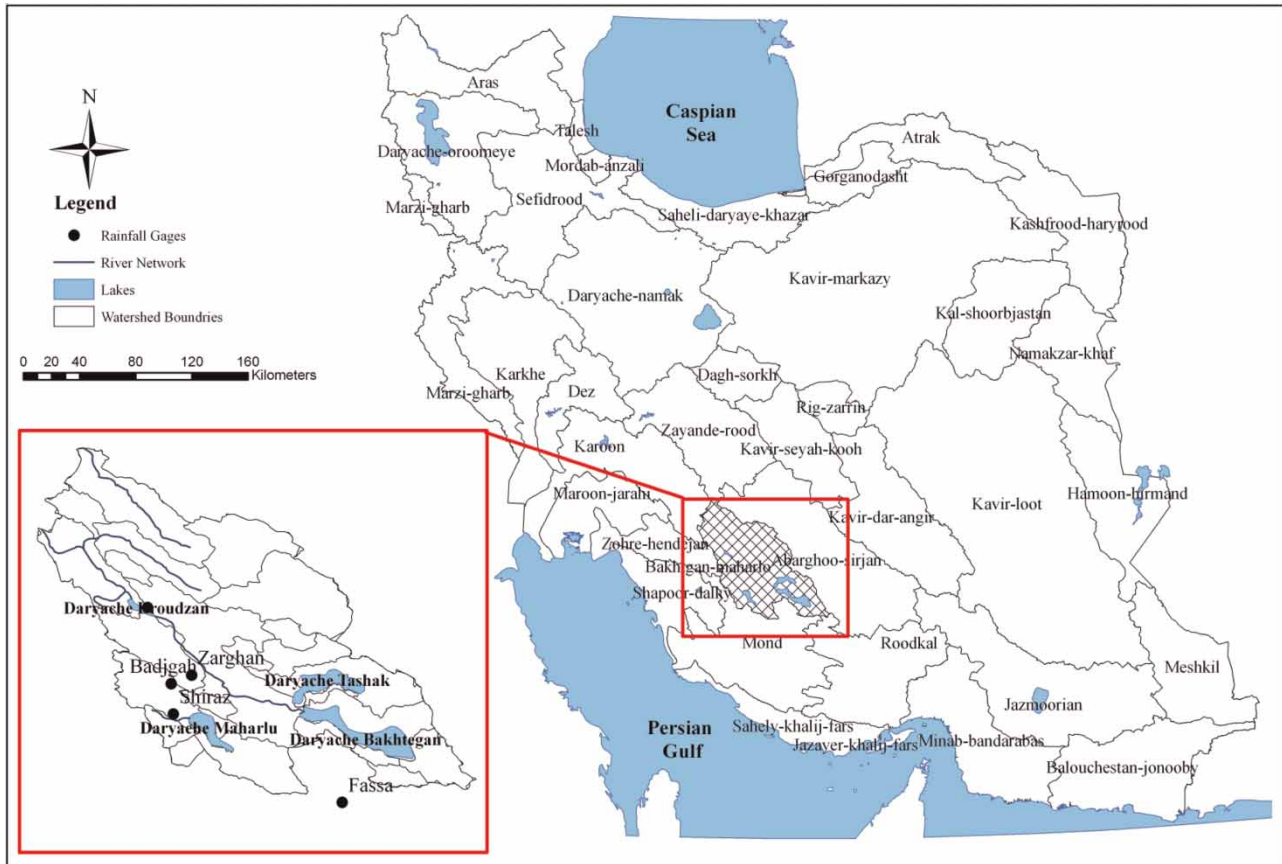


Figure 1 | Location of the study area and positions of the five recording rain gauges.

ANN and GA

A neural network is a parallel, distributed information processing structure consisting of interconnected processing elements called neurons and unidirectional signal channels called connections. Each processing element branches into as many output connections as desired and carries signals known as neuron output signals. The neuron output signal can be of any mathematical type desired. In other words, it is a mathematical representation of a biological neural network. In terms of implementation, it is basically a coupled input-output map constructed through an iterative procedure.

A neural network learns by way of training. Training may be supervised or unsupervised. In this study supervised training has been used, which provides the network with the desired response. In the training process, first, the input data are presented to the network, and then the network modifies

the weights of the neurons and adjusts them to predict the next point in the input data with desired accuracy. When this procedure is carried out with a large sample of the input data (comprising the training set), the neural network 'learns' the relationship between input and output data and then the trained network can be used to make a prediction for a point outside the training set. This process of predicting a point outside the training set is called testing. The input-hidden layer transformation performs a continuous nonlinear mapping of the n_0 input values x_i to the intermediate variables y_j ; the parameters of this transformation are the weights w_{ij} . In a similar way, the hidden-output layer transformation performs a (linear or nonlinear) mapping of the n_1 intermediate variables y_j to the output variables z_k ; the parameters of this mapping are the weights v_{jk} .

The most widely used neural network model is the multi-layer perceptron (MLP), in which the connection weight training is normally completed by a back-propagation (BP)

learning algorithm (Rumelhart *et al.* 1986; Zhang *et al.* 1998). The idea of weight training in MLPs is usually formulated as minimization of an error function, such as MSE between the network output and the target output, by iteratively adjusting connection weights.

ANN training is performed to determine the weights associated with the network in a near optimal way using an appropriate algorithm. In this regard, researchers have reported limitations such as uncertainty in convergence to global optimum in training the network parameters (Sexton *et al.* 1998). In a more recent paper, the topology and structure of ANN has been developed dynamically with sequential iterative learning process (Ghiassi *et al.* 2005).

GA is considered to be a heuristic, probabilistic, combinatorial, search-based optimization technique based on the biological process of natural evolution developed by Holland (1975). Goldberg (1989) discussed the mechanism and robustness of GA in solving nonlinear optimization problems. The combination of GA and ANN are found especially appropriate to undertake the forecasting problems with their advantages in using the ANN to learn complex nonlinear mapping and that of GA in searching for optimum parameters. Maniezzo (1994) applied GA in training a BP neural network.

In the realm of feature selection methods and wrapper technique, GA has been applied in different contexts as to optimize the number of hidden neuron nodes, their bias, and selecting the most pertinent input parameters. The structure of the coupling of GA and ANN is shown in Figure 2.

Feature selection

Feature selection is a general procedure of selecting a suitable subset of the pool of original feature spaces according to discrimination capability to improve the quality of data, and performance of a simulation technique. The process of feature selection is a challenge in a number of different tasks such as classification and data mining. Recently, the growing importance of knowledge discovery and data mining in practical applications has made the feature selection problem a motivating topic, especially when mining of knowledge from a real-world database is the target. Feature selection techniques can be categorized into three main branches (Tan *et al.* 2006), namely, embedded approaches, wrapper approaches, and filter approaches.

Embedded approaches are preventive and they have been developed to suit particular classification algorithms. Most of the known general and applicable approaches in feature selection can be categorized in the two broad classes of wrapper and filter methods (Guyon & Elisseeff 2003). In the wrapper methods, the objective function is usually a pattern classifier or a mathematical regression model which evaluates feature subsets by their predictive accuracy (recognition or regression evaluation on test dataset) using statistical re-sampling or cross-validation approaches. These methods measure the model performance and most possible subsets of input variables to find the appropriate input sets based on their calibration results. The suitable input arrangement can be set up through several ways. Some of the most well-known methods are presented as follows (Liu & Yu 2005):

- Forward selection, where the input set increases from a single input until model performance is no longer improved.
- Backward elimination, where the input set initially includes all candidates and candidates are removed one at a time.
- Optimization approach, where the decision to include each input is encoded as a variable within the overall model optimization.

The most important weakness of wrapper methods is their computational cost which makes them a poor choice for large feature sets. Achieving the best performance subset of input dataset is a very complicated procedure and reduces the performance reliability when considering the interaction of model parameters and its structure. In ANN modeling, components such as number of layers, training paradigm, and model parameters are samples of the effective parameters in model performance and directly influence the feature selection process. Genetic programming as another method could be a good example of a wrapper optimization approach in modeling.

Another method in the field of feature selection is filter approach. The method is a model-free technique utilizing a statistical criterion to find the dependence condition between the input candidates and output variable(s). This criterion acts as a statistical benchmark for reaching the suitable input variable dataset. The method makes feature selection more efficient which is the known advantage of filter-based selection over costly wrapper methods. It should be noted that the performance of the selection depends on the statistical dependency

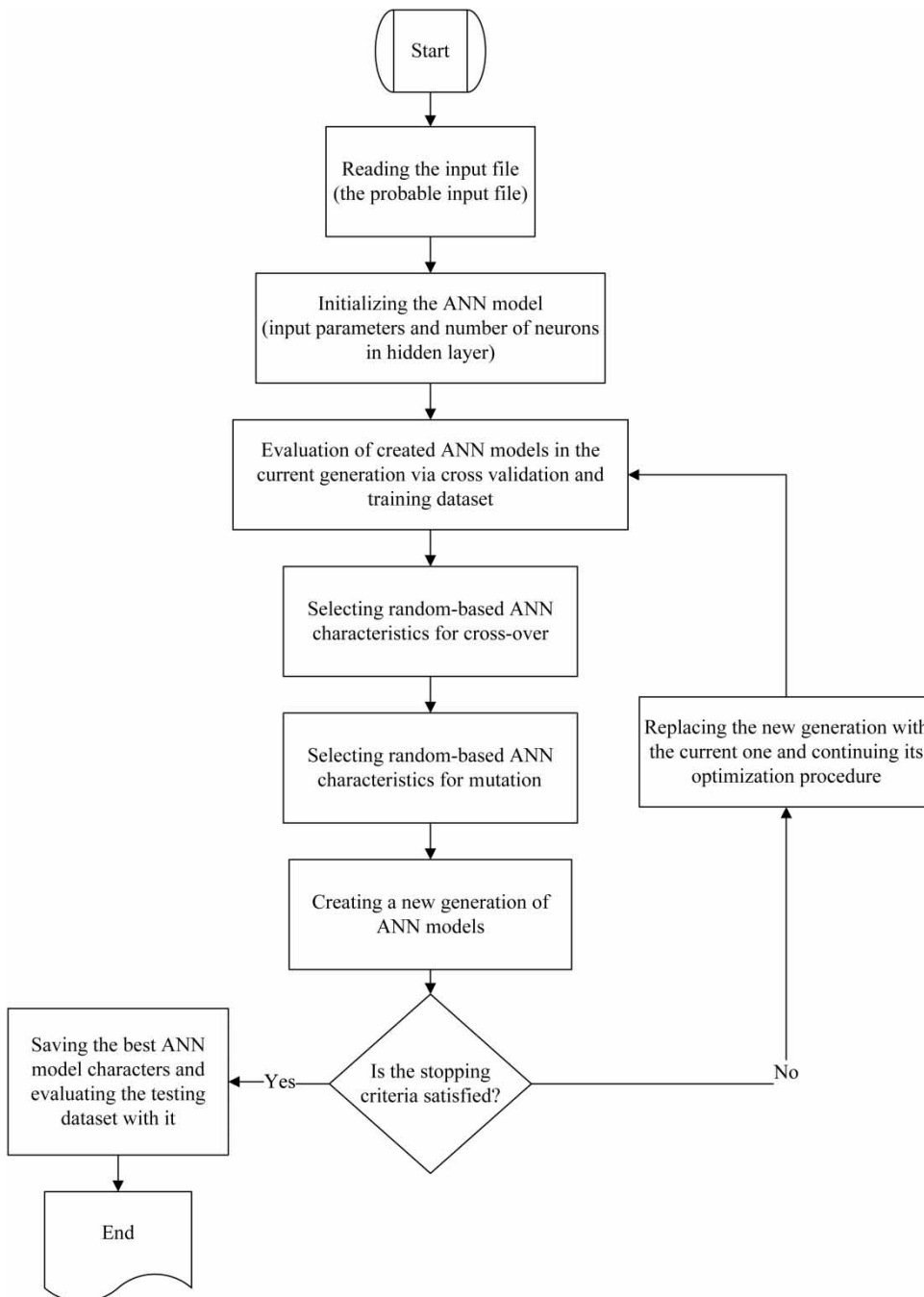


Figure 2 | The flowchart of the coupled GA and ANN.

measures in which the intrinsic characteristics of the training data are exploited. Three filter-based feature selection approaches used in this study are briefly explained.

Linear correlation coefficient (LCC) is a usual adopted criterion of dependency between input and output variables.

Battiti (1994) discusses the efficiency of LCC related to effects of noise and data transformation during data preprocessing and feature selection. Despite the popularity and simplicity of LCC in exploring the dependency of variables, this approach was shown to be inappropriate for real

nonlinear systems. Chi-squared criterion is considered for evaluating the goodness of fit and is based on nonlinearity of data distribution and known as a classical nonlinear data dependency criterion (Manning *et al.* 2008).

MI, as another filtering method, describes the reduction amount of uncertainty in estimation of one parameter when another is available. This method as a non-linear filtering method with its statistical criterion has recently been found to be more suitable in feature selection. It has also been found to be robust due to its insensitivity to noise and data transformations and also has no pre-assumption in correlation of input and output variables (May *et al.* 2008).

In the section on modeling applications, implementation of these feature selection approaches in rainfall forecasting is described. An optimized method based on backward-wrapper methodology is applied on the dataset and the results are compared with three filter-based feature selection. In this paper, effectiveness of partial autocorrelation function (PACF) in autonomous modeling will also be evaluated.

Sensitivity analysis

Sensitivity analysis is considered an essential step in all mathematical-based model applications. The purpose of conducting a sensitivity analysis is two-fold: that is first to evaluate the model's response to changes in input parameters, and second to quantify the likely uncertainty of the calibrated model resulting from uncertainties associated with the input parameters, stresses, and boundary conditions (Skaggs & Barry 1996; Nasserli *et al.* 2008). The main advantage of performing sensitivity analysis is to identify sensitive parameters or processes associated with model output. In neural networks like any mathematical-based model, sensitivity analysis provides feedback as to which input parameters are the most significant. Considering S as a dimensionless measure of the impact of change in one parameter on the output result, the sensitivity, S_{ij} of the y_i model dependent variable due to the change in the input parameter, x_j can be expressed as:

$$S_{ij} = \frac{\partial y_i}{\partial x_j} \quad (1)$$

where i is the index for i th model dependent variable (i.e., model output), j is the index for j th model input parameter.

For a model with n output parameters, the number of sensitivity coefficients that could be generated is given by $n \times m$.

It has to be emphasized that computation of sensitivity coefficients in a typical input-output model is not a trivial task as the author are not faced with a close form function but a complex procedure to convert input to output. As a result, some numerical scheme should be employed to compute those coefficients. Sensitivity analysis should be done for a range of input parameters. For the networks which are developed in this research, sensitivity coefficient of each input is computed using their mean value and variances.

Modeling procedure

In this section, implementation of two general feature selection methods is presented. These methods are backward wrapper improved with stochastic optimization and filter-based feature selection methods. First, in the wrapper method, the integration of GA within neural network architecture (GAANN) is shown to speed up the optimal parameters selection of the network structure. An integrated GAANN model combines stochastic optimization, as a backward feature selection, with an intelligent neural network simulation scheme (Figure 3). Also, the outline of this procedure is given in the Appendix (available online at <http://www.iwaponline.com/nh/046/178.pdf>). In this method, a pool of probable effective input variables is optimized by GA in creating the best ANN input dataset. After convergence of the optimization process, similar mathematical sensitivity analysis used by Nasserli *et al.* (2008) is applied on the selected input variables to prune probable non-effective ones. With the new pool of selected input parameters, the next optimization and backward feature selection run is performed.

This procedure is continued to achieve a stable input dataset and model structure. In this regard, the topology of created ANN is adjusted with GA to achieve the best model with assumed input parameters. It is clear that ANN topology and the type of dataset fully affect model efficiency and performance. Through the use of ANN modeling usually a suboptimal learning process is achieved. By introducing the integrated GAANN model, an optimization stage completes the global search and results for a better selection of parameters and input features. The main drawback of the latter model is the major increase in computational cost.

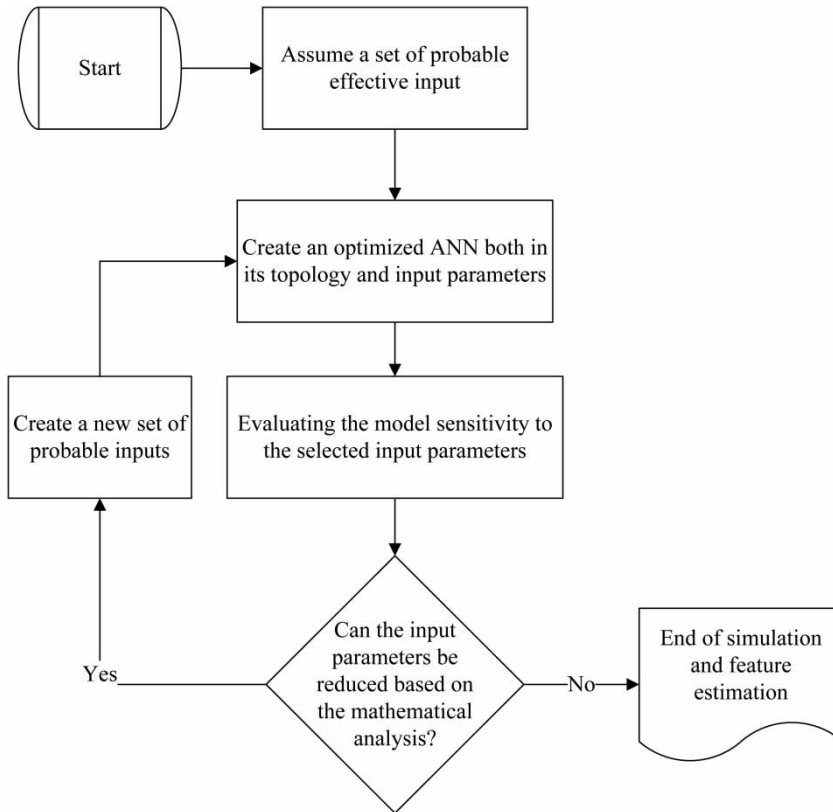


Figure 3 | The flowchart of the proposed feature selection algorithm.

In the second part of modeling, different methods of feature selections such as LCC, chi-squared and MI are implemented. According to Sharma (2000), the MI, a dimensionless index, can be estimated using an approximation function such as

$$MI = \frac{1}{N} \sum_{i=1}^N \log_e \left[\frac{f_{x,y}(x_i, y_i)}{f_x(x_i) f_y(y_i)} \right] \quad (2)$$

where $f_x(x_i)$, $f_y(y_i)$ and $f_{x,y}(x_i, y_i)$ are the respective marginal and joint probability density functions estimated at the sample data points.

These methods act as a preprocessing stage for ANN and can be substituted for the GA part of GAANN model hence reducing the computational cost. A brief explanation of filtering and selection of informative subset of input data is presented in Sotoca & Pla (2010).

To evaluate the model's performance, three criteria are used. In addition to coefficient of determination (R^2), the following goodness of fit criteria was used to measure the performance of the model

$$MSE = \frac{\sum_{j=0}^P \sum_{i=0}^N (d_{ij} - y_{ij})^2}{N \times P} \quad (3)$$

and

$$NMSE = \frac{\sum_{j=0}^P \sum_{i=0}^N (d_{ij} - y_{ij})^2}{N \times P \times \sigma_d} \quad (4)$$

where MSE is the mean squared error (mm^2); P is the number of output processing elements; N , number of exemplars in the dataset; y_{ij} , network output for exemplar i at processing element j ; and d_{ij} target output for exemplar i at processing element j . Normalized mean square error (NMSE) is a dimensionless dissimilarity coefficient

obtained by dividing MSE by variance of desired outputs (σ_d).

STUDY AREA AND DATA

The study area selected for this research is the Bakhtegan catchments, located in south-west Iran. It is located between 28°58' and 30°47' north and 52°32' and 53°41' east with a size of 31,492 km² and a semi-arid to arid climate (Figure 1). Nearly half of the area is located in the Zagroos mountain range and the remaining southern part is hillside. Average annual rainfall over the catchment varies from 200 to 750 mm and it is identified as a Mediterranean climate.

The selected five recording rain gauges have recorded at least 17 years of rainfall (1986–2002) with a temporal resolution of 30 min intervals. The rainfall data were collected from recording gauges operated by the Iranian National Meteorological Organization. Zarghan and Shiraz stations were selected as the target rain gauges of interest for rainfall prediction. The characteristics of the recording rain gauges are summarized in Table 1. After some preliminary data manipulation, 19 events were selected keeping zeros among rain gauges for synchronization purposes. These 19 events result in 944 input–output pairs for which 66% of the dataset were used for training (626 pairs), 23% for testing (217 pairs), and the remaining 11% for validation (150 pairs). Due to chaotic behavior of the rainfall field, the data transformation has been implemented using Equation (1) to reduce the variance of variation

$$y = a \log_{10}(x + b) \quad (5)$$

where a and b are equal to 0.5 and 1, respectively, and all original zeros were transformed to zero in the normalized dataset.

RESULTS AND DISCUSSION

In this section the efficiency of the filter-based feature selection versus optimized backward wrapper method for rainfall forecasting has been evaluated. This methodology has been examined for two synoptic rain gauges, Zarghan and Shiraz, within the selected catchment.

In the first step, comprehensive simulations to achieve the best set of appropriate features (inputs) for Zarghan station have been executed. GAANN has been executed 13 consecutive times and the results are presented in Tables 2 and 3. In Table 2, statistical performance of the best model based on different GA parameters, within a standard acceptable range, including ANN model topology are given. The range of mutation and crossover rates (P_m and P_c) are 1.2%–1.4% and 92%–98%, respectively. In a successive analysis presented in Table 3, mathematical sensitivity derived from a method proposed by Nasserli *et al.* (2008) is carried out and results of each sensitive input variable relative to target value are identified. It is observed that different GA parameters lead to different optimum numbers of inputs and different ANN topology. For updating and improving the performance, the ineffective inputs and ones with low effectiveness are omitted from the list of input parameters. The most prominent and effective rainfall value from prior lags happened to be from the first lag of Zarghan station itself. Other effective rainfall values with less degree of sensitivities are the third lag rainfalls of Shiraz and Doroodzan stations which are both

Table 1 | Location of gauges in the catchment

Station	Longitude (East)	Latitude (North)	Elevation above mean sea level (m)	Mean annual precipitation (mm)	Available data
Shiraz	52:32	29:36	1,488	324	1964–now
Badjgah	52:32	29:36	1,810	315	1968–now
Zarghan	52:43	30:47	1,596	328	1986–now
Doroodzan	53:26	30:13	1,462	474	1986–now
Fassa	53:41	28:58	1,288	295	1974–now

Table 2 | Statistic and genetic properties of models for Zarghan station

Model No.	1	2	3	4	5	6	7	8	9	10	11	12	13
Optimum hidden layer nodes	10	7	7	9	6	8	8	10	4	6	6	9	10
$P_m(\%)$	1.2	1.2	1.2	1.2	1.3	1.3	1.3	1.3	1.4	1.4	1.4	1.4	1.2
$P_C(\%)$	92	94	96	98	92	94	96	98	92	94	96	98	96
R^{2a}	0.01	0.42	0.59	0.58	0.6	0.44	0.43	0.425	0.49	0.59	0.56	0.59	0.55
MSE (mm ²)	0.43	0.13	0.09	0.07	0.08	0.12	0.12	0.123	0.111	0.08	0.09	0.08	0.088
NMSE ^a	1.72	0.66	0.42	0.41	0.39	0.59	0.69	0.6	0.513	0.41	0.44	0.41	0.457

^a R^2 and NMSE are dimensionless quantities.

Table 3 | Sensitivity coefficients of input variables [symbol (-) is ignored input parameter in the next model] for Zarghan station as target

Model No.		1	2	3	4	5	6	7	8	9	10	11	12	13	Average of SC (dimensionless)
First lag	Shiraz	0	-	-	-	-	-	-	-	-	-	-	-	-	0
	Badjgah	0.02	-	-	-	-	-	-	-	-	-	-	-	-	0
	Zarghan	0.4	0.65	0.7	0.73	0.67	0.74	0.61	0.68	0.61	0.72	0.75	0.71	0.75	0.66
	Doroodzan	0.043	0.02	0.03	-	-	-	-	-	-	-	-	-	-	0
	Fassa	0.054	0.073	0.066	0.065	0.11	0.2	0.06	0.07	0.15	0.08	0.09	0.06	0.07	0.09
Second lag	Shiraz	0	-	-	-	-	-	-	-	-	-	-	-	-	0
	Badjgah	0.03	-	-	-	-	-	-	-	-	-	-	-	-	0
	Zarghan	0	-	-	-	-	-	-	-	-	-	-	-	-	0
	Doroodzan	0	-	-	-	-	-	-	-	-	-	-	-	-	0
	Fassa	0.01	-	-	-	-	-	-	-	-	-	-	-	-	0
Third lag	Shiraz	0.08	0.18	0.14	0.15	0.15	0.15	0.13	0.18	0.1	0.17	0.17	0.11	0.17	0.15
	Badjgah	0.01	-	-	-	-	-	-	-	-	-	-	-	-	0
	Zarghan	0.03	-	-	-	-	-	-	-	-	-	-	-	-	0
	Doroodzan	0.06	0.03	0.09	0.13	0.11	0.21	0.07	0.13	0.14	0.16	0.16	0.06	0.11	0.1
	Fassa	0.05	0	-	-	-	-	-	-	-	-	-	-	-	0

located east of the target station. The least effective input data happened to be the Fassa station.

It is clear that after four simulation runs, an optimal set of input parameters are obtained and further simulations carried out to test any inconsistency of the model associated with recognizing the best ANN inputs and their related parameters. Figure 4 presents the scattered plots of transformed observed and computed testing rainfall dataset in Zarghan station based on results of model numbers 5 and 13. Linear trend lines of these two models are under the line 1:1, indicating that the computed values of rainfall are generally underestimated.

For evaluation of direct filtering methods discussed in the feature selection section, correlations of all inputs (15 variables) with the rainfall observation in Zarghan station are computed using LCC, chi-squared correlation, and MI.

The results are presented in Table 4. In this table, the effectiveness ranks of the stations are presented based on their governing relationships for three selected methods. All three lags of the Zarghan station have better correlation with the target station and their differences are not significant. In the chi-squared method, Zarghan's first two lags showed the most differentiable relevancy to the current rainfall followed by Shiraz and Fassa stations; however, the result does not support the GAANN outcomes.

As the third differentiator, results of MI clearly indicate nearly the same performance in selecting the most relevant inputs as compared with the GAANN algorithm. The justification for using MI index as a simple and fast procedure was evident when it is judged against the GA as a method with very low convergence speed.

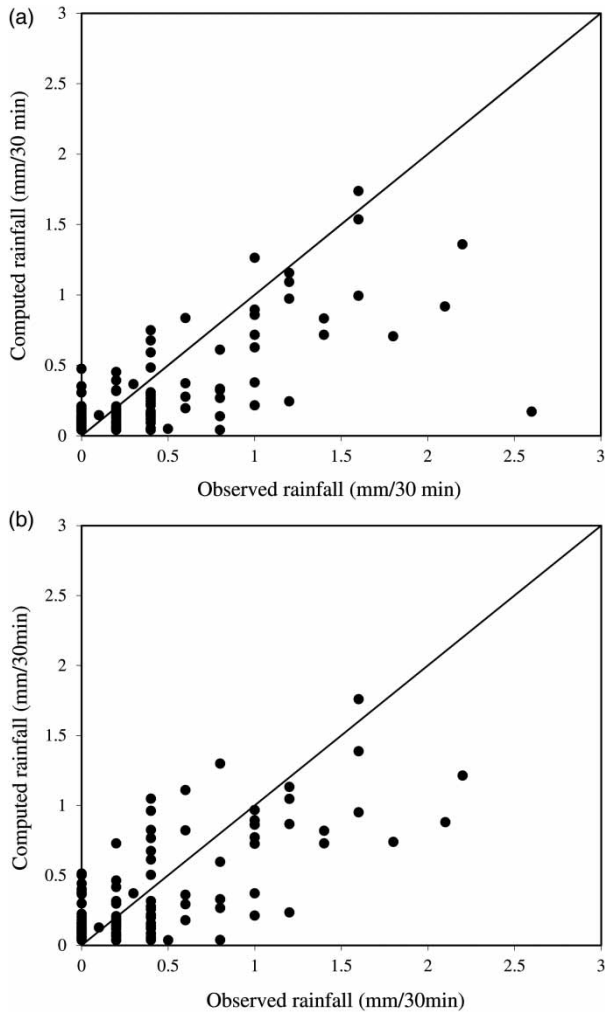


Figure 4 | Observed vs. computed rainfall (mm/30 min) in models (a) 5 and (b) 13.

Table 4 | Contribution of the network stations based on three filter methods (LCC, chi-squared and MI index) for Zarghan station as target

Filter method	Lags	Stations				
		Shiraz	Badjgah	Zarghan	Doroodzan	Fassa
Linear correlation coefficient	First lag	0.459	0.157	0.674	0.209	0.130
	Second lag	0.443	0.125	0.545	0.167	0.133
	Third lag	0.428	0.100	0.426	0.122	0.117
	Ranked based on LCC	2	5	1	3	4
Chi-squared	First lag	1.954	608	2.371	874	1.336
	Second lag	2.351	581	3.175	699	929
	Third lag	1.541	1.069	1.248	508	862
	Ranked based on chi-squared	2	4	1	5	3
MI index	First lag	0.052	0.016	0.108	0.028	0.017
	Second lag	0.048	0.018	0.069	0.022	0.017
	Third lag	0.050	0.018	0.048	0.018	0.018
	Ranked based on MI	2	5	1	3	4

All values are dimensionless.

For further justification of previous results, a similar analysis was performed using Shiraz rain gauge as the target station. First, a comprehensive simulation for Shiraz station was performed. The statistical performance, optimum number of hidden layer nodes and SC results of selected features are reported in Tables 5 and 6, respectively. Based on the results of the first run (the first model), Shiraz as the target and Zarghan as the next effective station are the most sensitive stations in estimating the output. Their first, second, and third 30 min rainfall lags are considered as the effective inputs. The results of the next two runs (second and third models) were consistent with the results of the first model. Omitting the stations with low SC values, again, Shiraz and Zarghan were the two effective stations in the second model. Finally, in the third model, Shiraz remained the only effective station in predicting the lead time rainfall of the target station with the first to third lags having the highest to lowest effectiveness, respectively. To evaluate the performance of filter-based methods using LCC, chi-squared, and MI index, the same procedures were performed and the results reported in Table 7. Each method was ranked based on the most to least effective station in predicting the rainfall and the results were in harmony compared with the GAANN approach. The CPU runtimes for execution of GAANN algorithm on a personal computer with 3 GHz processor for the first, second, and third model of Shiraz station were 107.5, 76.4, and 63.2 hours, respectively. While use of GAANN is

Table 5 | Statistic and genetic properties of models for Shiraz station

Model No.	1	2	3
Optimum hidden layer nodes	10	8	8
$P_m(\%)$	1.1	1.0	1.2
$P_C(\%)$	90	96	94
R^{2a}	0.58	0.61	0.64
MSE (mm ²)	0.26	0.18	0.14
NMSE ^a	0.42	0.40	0.39

^a R^2 and NMSE are dimensionless quantities.

Table 6 | Sensitivity coefficients of input variables (symbol (-) is ignored input parameter in the next model) for Shiraz station as target

Model No.		1	2	3	Average of SC
First lag	Shiraz	0.31	0.68	0.71	0.57
	Badjgah	0.10	-	-	0.10
	Zarghan	0.18	0.15	-	0.16
	Doroodzan	0.09	-	-	0.09
	Fassa	0.05	-	-	0.05
Second lag	Shiraz	0.18	0.12	0.16	0.15
	Badjgah	0.07	-	-	0.07
	Zarghan	0.15	0.08	-	0.12
	Doroodzan	0.06	-	-	0.06
	Fassa	0	-	-	-
Third lag	Shiraz	0.11	0.11	0.11	0.11
	Badjgah	0.05	-	-	0.05
	Zarghan	0.09	-	-	0.09
	Doroodzan	0	-	-	-
	Fassa	0	-	-	-

computationally expensive, the advantage of the three proposed methods to the GAANN is that it is not iterative in nature and its efficiency overcomes the GAANN approach.

To investigate the effect of cumulative data on forecasting precision, a simulation result, based on the previous five lags of Zarghan station itself, was analyzed. Based on Equation (1), results of the mathematical sensitivity analysis, presented in Table 8, indicate that the first lag should be selected as an effective input, implying very short memory of rainfall for the Zarghan station. Scatter plot of the observed vs. forecasted non-transformed rainfall of the cumulative dataset is shown in Figure 5. Also, results from PACF analysis verifies that the most important feature (first lag) is similar to the optimized feature selected procedure proposed in this paper (Figure 6).

CONCLUDING REMARKS

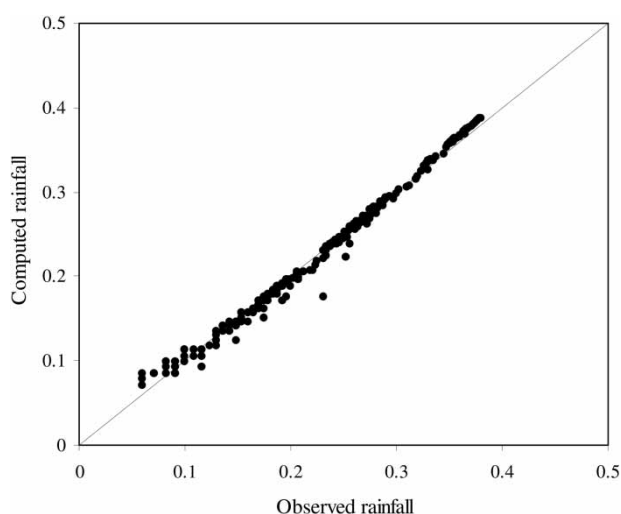
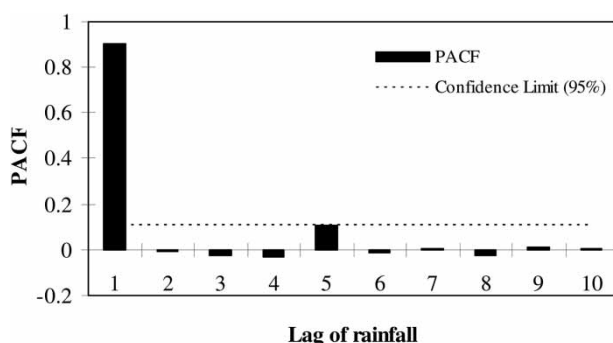
Various researchers have investigated the short-time forecasting of hydrological processes such as rainfall fields using ANN or a combination of ANN and GA. Although they have obtained practical findings, most of these efforts have focused on identifying an efficient method of selecting appropriate and relevant input data for any data-driven models. The process of selecting input data using trial and error methods for ANN has already been improved through

Table 7 | Contribution of the network stations and their first to third lags using three filter methods (LCC, chi-squared, and MI index) for Shiraz station as target

Filter method	Lags	Stations				
		Shiraz	Badjgah	Zarghan	Doroodzan	Fassa
Linear correlation coefficient	First lag	0.744	0.126	0.401	0.040	0.053
	Second lag	0.609	0.124	0.345	0.028	0.025
	Third lag	0.521	0.109	0.300	0.011	0.021
	Ranked based on LCC	1	3	2	5	4
Chi-squared	First lag	2,895	710	1,475	785	702
	Second lag	3,356	681	2,785	587	512
	Third lag	2,531	589	1,152	432	389
	Ranked based on chi-squared	1	3	2	4	5
MI index	First lag	0.151	0.033	0.048	0.023	0.022
	Second lag	0.098	0.027	0.039	0.022	0.023
	Third lag	0.077	0.027	0.039	0.021	0.032
	Ranked based on MI	1	3	2	5	4

Table 8 | Sensitivity coefficients for different lags of the PACF model with cumulative dataset in Zarghan station. (Results are rounded to the nearest hundredths)

Model inputs	Sensitivity coefficients
First lag	1.03
Second lag	0
Third lag	0
Fourth lag	0
Fifth lag	0

**Figure 5** | Observed vs. computed of non-transformed dimensionless cumulative rainfall.**Figure 6** | Partial auto-correlation results for Zarghan rain gauge with cumulative dataset.

employing GA for an optimized selection of featured inputs. In this paper, we have examined three feature selection methods: LCC, chi-squared, and MI, and demonstrated that they are more efficient than the combination methods of GA and ANN.

In order to evaluate the performance of the three proposed methods against the combined GA and ANN (GAANN) algorithms as a backward wrapper method, first we applied the GAANN to forecast 30 min lead rainfall of a target station using the surrounding gauge stations for two different target stations. The GAANN method was effective in generating an optimized configuration of an ANN model with MSE of 0.08 (mm^2) for Zarghan station. However, the approach is computationally inefficient due to the searching method built into the algorithm. Use of the cumulative approach of arranging the input data was found to be favorable in comparison with those developed using discrete data. Based on sensitivity analysis performed on 15 input variables, only four of the variables proved to be effective in forecasting the target rainfall.

The performance of the ANN model combined with the GA approach was compared with LCC, chi-squared, and MI as three filter-based feature selection methods. As a result, the MI indices, as an indicator to identify the most appropriate input to ANN approach, practically correspond to the selected parameters using GAANN. In terms of ease of development and reducing the computational effort, the selection of input variables based on MI analysis of the data, rather than application of specific heuristics approach, has considerable computational advantage.

REFERENCES

- Battiti, R. 1994 Using mutual information for selecting features in supervised neural net learning. *IEEE Trans. Neural Netw.* **5** (4), 537–550.
- Bowden, G. J., Dandy, G. C. & Maier, H. R. 2005a Input determination for neural network models in water resources applications, Part 1, background and methodology. *J. Hydrol.* **301** (1–4), 75–92.
- Bowden, G. J., Dandy, G. C. & Maier, H. R. 2005b Input determination for neural network models in water resources applications, Part 2, Case study: forecasting salinity in a river. *J. Hydrol.* **301** (1–4), 93–107.
- French, M. N., Krajewski, W. F. & Cuykendal, R. R. 1992 Rainfall forecasting in space and time using a neural network. *J. Hydrol.* **137** (1–4), 1–37.
- Ghiassi, M., Saidane, H. & Zimbra, D. K. 2005 A dynamic artificial neural network model for forecasting time series events. *Int. J. Forecast.* **21** (2), 341–362.
- Goldberg, D. E. 1989 *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison Wesley, Boston, MA.

- Govindaraju, R. S. 2000a Artificial neural network in hydrology, I: preliminary concepts. *J. Hydrol. Eng.* **5** (2), 115–123.
- Govindaraju, R. S. 2000b Artificial neural network in hydrology, II: hydrological applications. *J. Hydrol. Eng.* **5** (2), 124–137.
- Guyon, I. & Elisseeff, A. 2003 An introduction to variable and feature selection. *J. Mach. Learn. Res.* **3**, 1157–1182.
- Holland, J. 1975 *Adaptation in Natural and Artificial Systems*. University of Michigan Press, Ann Arbor, MI.
- Hong, W. Ch. 2008 Rainfall forecasting by technological machine learning models. *Appl. Math. Comput.* **200** (1), 41–57.
- Kisi, O. & Shiri, J. 2012 Wavelet and neuro-fuzzy conjunction model for predicting water table depth fluctuations. *Hydrol. Res.* **43** (3), 286–300.
- Lin, G. F. & Wu, M. C. 2009 A hybrid neural network for typhoon-rainfall forecasting. *J. Hydrol.* **375** (3–4), 450–458.
- Liu, H. & Yu, L. 2005 Toward integrating feature selection algorithms for classification and clustering. *IEEE Trans. Knowl. Data Eng.* **17** (4), 491–502.
- Luk, K. G., Ball, J. E. & Sharma, A. 2000a A study of optimal lag and statistical inputs to artificial neural network for rainfall forecasting. *J. Hydrol.* **227**, 56–65.
- Luk, K. G., Ball, J. E. & Sharma, A. 2000b An application of artificial neural network for rainfall forecasting. *Math. Comput. Model.* **33**, 683–693.
- Maniezzo, V. 1994 Genetic evaluation of the topology and weight distribution of Neural Network. *IEEE Trans. Neural Netw.* **5** (1), 39–53.
- Manning, C. D., Raghavan, P. & Schütze, H. 2008 *Introduction to Information Retrieval*. Cambridge University Press, Cambridge, MA.
- May, R. J., Maier, H. R., Dandy, G. C. & Fernando, T. G. 2008 Non-linear variable selection for artificial neural networks using partial mutual information. *Environ. Model. Softw.* **23**, 1312–1326.
- Minns, A. W. 2000 Subsymbolic methods for data mining in hydraulic engineering. *J. Hydroinform.* **2** (1), 3–13.
- Nasserri, M., Asghari, K. & Abedini, M. J. 2008 Optimized scenario for rainfall forecasting using genetic algorithm coupled with artificial neural network. *Expert Syst. Appl.* **35**, 1415–1421.
- Noori, R., Karbassi, A. R., Moghaddamnia, A., Han, D., Zokaei-Ashtiani, M. H., Farokhnia, A. & Ghafari Gousheh, M. 2011 Assessment of input variables determination on the SVM model performance using PCA, Gamma test, and forward selection techniques for monthly stream flow prediction. *J. Hydrol.* **401**, 177–189.
- Nourani, V. & Parhizkar, M. 2013 Conjunction of SOM-based feature extraction method and hybrid wavelet-ANN approach for rainfall–runoff modeling. *J. Hydroinform.* **15** (3), 829–848.
- Ramirez, M. C. V., Velho, H. F. C. & Ferreira, N. J. 2005 Artificial neural network technique for rainfall forecasting applied to the Sao Paulo region. *J. Hydrol.* **301**, 146–162.
- Rumelhart, D. E., Hinton, G. E. & Williams, R. J. 1986 Learning internal representation by error propagation. In: *Parallel Distributed Processing: Explorations in the Microstructure of Cognition* (D. E. Rumelhart & J. L. McClelland, eds). MIT Press, Cambridge, MA, pp. 318–362.
- Senthil Kumar, A. R., Ojha, C. S. P., Goyal, M., Singh, R. D. & Swamee, P. K. 2012 Modelling of suspended sediment concentration at Kasol in India using ANN, fuzzy logic and decision tree algorithms. *J. Hydrol. Eng.* **17** (3), 394–404.
- Sexton, R. S., Dorsey, R. E. & Johnson, J. D. 1998 Toward global optimization of neural networks: a comparison of the genetic algorithm and back propagation. *Decis. Support Syst.* **22**, 171–185.
- Sharma, A. 2000 Seasonal to interannual rainfall probabilistic forecasts for improved water supply management: Part 1 – A strategy for system predictor identification. *J. Hydrol.* **239**, 232–239.
- Skaggs, T. H. & Barry, D. A. 1996 Sensitivity methods for time continuous, spatially discrete groundwater contaminant transport models. *Water Resour. Res.* **32**, 2409–2420.
- Sotoca, J. M. & Pla, P. 2010 Supervised feature selection by clustering using conditional mutual information-based distances. *Pattern Recog.* **43** (6), 2068–2081.
- Tan, P. N., Steinbach, M. & Kumar, V. 2006 *Introduction to Data Mining*. Addison Wesley, Boston, MA.
- Toth, E., Brath, A. & Montanari, A. 2000 Comparison of short-term rainfall prediction models for real-time flood forecasting. *J. Hydrol.* **239**, 132–147.
- Zhang, G., Patuwo, B. E. & Hu, M. Y. 1998 Forecasting with artificial neural networks: the state of the art. *Int. J. Forecast.* **14** (1), 35–62.

First received 3 November 2013; accepted in revised form 22 April 2014. Available online 30 June 2014