Performance evaluation of ANN and geomorphology-based models for runoff and sediment yield prediction for a Canadian watershed

A. Sarangi[1,*], C. A. Madramootoo[2], P. Enright[2], S. O. Prasher[2] and R. M. Patel[2]

[1]Water Technology Centre, Indian Agricultural Research Institute, New Delhi 110 012, India
[2]Department of Bioresources Engineering, Macdonald Campus, McGill University, Montreal H9X3V9, Canada

Artificial Neural Network (ANN) and regression models were developed using watershed-scale geomorphologic parameters to predict surface runoff and sediment losses of the St. Esprit watershed, Quebec, Canada. Geomorphological parameters describing the land surface drainage characteristics and surface water flow behaviour were empirically associated with measured rainfall and runoff data and used as input to a three-layered back-propagation feed-forward neural network model. Morphological parameters such as bifurcation ratio, area ratio, channel length ratio, drainage factor and relief ratio were selected using the Multivariate Adaptive Regression Splines tool, based on their relative importance in prediction of runoff and sediment yield. Regression models were developed using the curve-fitting toolbox of MATLAB software and compared with the results obtained from ANN models. The coefficient of determination (R^2) and model efficiency factor (E) were estimated to ascertain the model performance. Geomorphology-based ANN model validation statistics resulted in R^2 values ranging from 0.85 to 0.95 and E values from 0.74 to 0.82 for peak runoff rate and R^2 values from 0.78 to 0.93 and E values from 0.71 to 0.76 for sediment loss. Using geomorphology-based regression models, R^2 values for the same dataset varied from 0.78 to 0.88 ($0.74 > E > 0.69$) for peak runoff rate prediction and 0.39 to 0.54 ($0.53 > E > 0.46$) for sediment prediction. When morphological parameters were not associated with rainfall depth and peak runoff rate, prediction error statistical parameter values (R^2 and E) were less for both neural network and regression models. Thus, associating selected geomorphological parameters with rainfall depth and peak runoff rate enhances the accuracy of runoff rate and sediment loss predictions from the watershed. Furthermore, ANN models performed better than the regression equations.

Keywords: Artificial Neural Network, geomorphology, regression splines, runoff, sediment yield.

KNOWLEDGE of landscape morphology along with the hydrologic processes is required to conceptualize the generation of runoff and sediment loss from precipitation events. In the past few decades, great strides have been made in conceptualizing the process of runoff generation and sediment yield from watersheds through modelling approaches. Models are classified based on their degree of representation of the physical processes involved in abstraction of the real world phenomenon. With the increasing degree of representation, models are classified as black-box models, conceptual models and physically based distributed models. The physically based distributed model can be considered a better choice in a rigorous theoretical sense. However, the significant data requirements of such models, coupled with the time involved in model development, calibration and validation compared to other model categories, make them an unfavourable choice in operational hydrology[1]. Lumped conceptual models are favoured in terms of their limited data requirements and inclusion of a conceptual framework, but require a lengthy calibration and parameterization process. In this context, use of soft computing and data mining tools offers an alternative to the distributed and physics-based modelling approaches. The Artificial Neural Network (ANN), a soft computing tool, belongs to black-box modelling category and has its own limitations[2]. The main advantage of the ANN approach over traditional methods is that it does not require the complex nature of the underlying process under consideration to be explicitly described in a mathematical form[3]. Other advantages of ANN over conventional models are discussed in detail by French *et al.*[4]. In recent years, ANN models have attracted researchers in many scientific and engineering disciplines, since they are capable of correlating large and complex multi-parameter datasets without any prior knowledge of the relationships between the parameters. However, several studies[5–8] have argued for more cautious approaches that include considerations of relevant physics and statistical principles in an effort to make ANN more useful as a practical as well as research

*For correspondence. (e-mail: arjamadutta.sarangi@elf.mcgill.ca)

tool. Therefore, determination of appropriate model inputs, development of a suitable network architecture, and parameter estimation have been identified as aspects that need further attention[9]. The objective of this study is to relate judiciously selected dimensionless geomorphological parameters reflecting watershed hydrology with rainfall depth and duration to predict runoff and again relate the parameters with runoff rates to predict sediment loss, using ANN and regression models.

Background information

The ANN technique was formulated based on the cognitive response of the human brain. ANNs were first developed in the 1940s, and in recent decades, considerable interest has been raised over their applications in hydrological modelling, as the current algorithms overcome the limitations of previous network algorithms[10]. The increasing use of ANN in estimating and predicting water resources variables has been documented in recent studies[5,6,9,11]. Among the plethora of applications of ANN, the systems approach, in general, may be categorized into a pattern-mapping problem, i.e. input–output mapping as in the case of black-box modelling. For this purpose, one of the best suitable architectures is a feed-forward network[12]. It was also established by Hornik *et al.*[13] that a feed-forward network could be considered as a general nonlinear approximator. This property of generality prompted researchers to use ANN models for predicting the complex hydrologic responses such as estimation of runoff and sediment loss from watersheds.

Smith and Eli[14] trained a three-layer Back Propagation (BP) ANN model to predict runoff from stochastically generated rainfall patterns obtained by analysis of weather radar data of Advanced Very High Resolution Radiometer (AVHRR) and cloud cover images. They revealed that the ANN model predicted peak runoff rate and times to peak values of a small synthetic watershed (0.1 ha) for 76 rainfall events were in line with the observed values (Root Mean Square Error (RMSE) ranging from 0.1 to 0.3). However, this study raised a couple of questions related to the suitability of ANN approaches in explaining the physics of the rainfall-runoff process and accounting for the watershed drainage patterns. In this study, one of the apprehensions on the suitability of ANN to recognize the differences in watershed drainage network and predict runoff and sediment losses more accurately, is addressed by including Horton's geomorphological parameters.

Tokar and Johnson[10] used a three-layer BP ANN to forecast daily runoff as a function of daily precipitation, temperature and snowmelt for the Little Patuxent River watershed in Maryland. The ANN model proved to be a promising alternative to the existing statistical regression and simple conceptual rainfall-runoff-based models. ANN models often represent an improvement upon the prediction accuracy and flexibility of current methods. Anmala *et al.*[15] developed a three-layer feed-forward ANN with a BP learning algorithm with five input nodes, monthly precipitation from four different stations and mean monthly temperature of a Kansas watershed. The monthly mean runoff was chosen as output. It was reported that the feed-forward ANN, without time-delayed input, did not provide a significant improvement over other regression techniques. However, inclusion of recurrent ANNs resulted in better performance. Cannon and Whitfield[16] concluded ANNs to be superior to stepwise linear regression procedures while conducting a study on predicting runoff from five-day mean stream flow data collected from 21 watersheds of British Columbia, Canada. Sarangi and Bhattacharya[17] developed an empirical model using regression techniques to predict sediment concentration from runoff rate associated with geomorphological parameters. The relationship between runoff rate and dimensionless geomorphological parameters was obtained by trial and error techniques. The developed model performed better ($R^2 = 0.92$ and $E = 0.87$) when validated for watersheds of Damoder Valley Corporation, India. Nagy *et al.*[18] used a feed-forward three-layer BP ANN model to predict the sediment concentration in rivers using eight input parameters, reflecting sediment and river-bed information. The ANN approach provided better results than other formulae used for estimation of sediment concentration. Sudheer *et al.*[3] developed a new approach for designing the network structure in an ANN-based rainfall-runoff model. Their method used statistical properties such as an autocorrelation function and a partial autocorrelation function of the data series, in identifying a unique input vector that best represented the process for the basin, and a standard algorithm for training. The methodology was validated using data from a river basin in India. The results of the study were highly promising.

Yitian and Gu[19] developed a mass-conservation transfer function for flow and sediment yield in rivers and incorporated the models into an ANN architecture, based on an actual river network architecture. They also expanded hydrological applications of the ANN modelling technique to sediment yield predictions. The ANN river system model was applied to model daily discharges and annual sediment discharges in the Jingjiang reach of the Yangtze River and Dongting Lake, China. An assessment of model accuracy demonstrated that the ANN technique is a powerful tool for real-time prediction of flow and sediment transport in a complex network of rivers. Zhang and Govindaraju[11] developed a Geomorphology-based ANN (GANN) for prediction of runoff over watersheds. Watershed morphological parameters such as bifurcation ratio, area ratio, channel length and slope ratio, required for the development of a Geomorphologic Instantaneous Unit Hydrograph (GIUH), were used for development of flow path probabilities. The path probabilities were used as connection weights between the hidden and output layers. In the three-layered ANN architecture, the input layer consisted of the rainfall excesses from the current time-step and previous time-steps. The number of previous time-steps, included as inputs, was obtained by trial and error. The number of hidden layers was equal to the number of paths that the flow could take in a fourth-order channel network. It was concluded that GANN offered a promising step towards elevating ANN from purely empirical models to those based on geomorphology. With the added flexibility provided through the connection weights between input and hidden-layer nodes, they were shown to perform better than the GIUH model.

A few ANN models for the prediction of runoff and sediment loads consider the geomorphological parameters as inputs or as network weights to predict the runoff and sediment loss over watersheds. However, it was revealed from the review of ANN-based hydrological models that no research investigation has mathematically associated the geomorphological parameters with rainfall or runoff and used these as input to the neural network model to predict the runoff and sediment yield on a watershed scale.

Study watershed and data

The St. Esprit watershed (26.1 km^2) is located in the province of Quebec, Canada, approximately 50 km north of Montreal (Figure 1). It is located in the St. Esprit river basin (210 km^2), which is a tributary of the L'Assomption watershed (4220 km^2). The watershed is located between 45°55′00″ and 46°00′00″N lat, and 73°41′32″ and 73°36′00″W long. The maximum difference in elevation from the outlet to the highest point of the watershed is about 50 m. The principal watercourse of the delineated St. Esprit watershed (Figure 2) is 9 km long and there are a total of 60.3 km of watercourses within the watershed[20]. The climate of the watershed is temperate. The frost-free growing season varies from 122 to 138 days with a mean annual precipitation of 998 mm, of which roughly 20% occurs as snow. The mean annual temperature is 5.2°C and the daily mean temperature in the month of July, which is the peak summer month in Quebec, varies between 18 and 21°C. Soils formed from glacial tills (sandy loams and loams) are located in the upland areas that occupy approximately 37% of the watershed. Soils formed from marine sediments (clay, clay loam) occupy 38% of the watershed, and the balance of the soils (sand to loamy clay) is formed mostly from alluvial deposits. About 64% of the watershed is planted with corn, cereal, soybean, vegetable, hay and pasture. The rainfall, soil temperature, runoff and sediment concentration values for selected events during the snow-less periods (April to September) for four years (1994–97) were extracted from the telemetric sensor data of the watershed gauging station. These periods were so chosen to ensure proper flow of water over the land surface, without the effect of snowfall and subsequent freezing and thawing on land surfaces. The freezing of water on land surfaces during snowfall restricts the

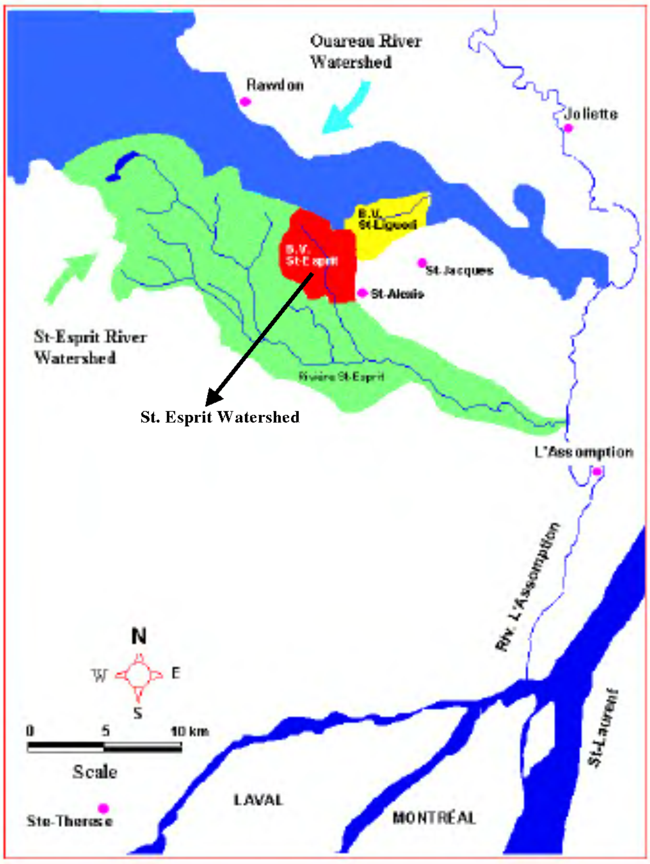

Figure 1. Location map of St. Esprit watershed.

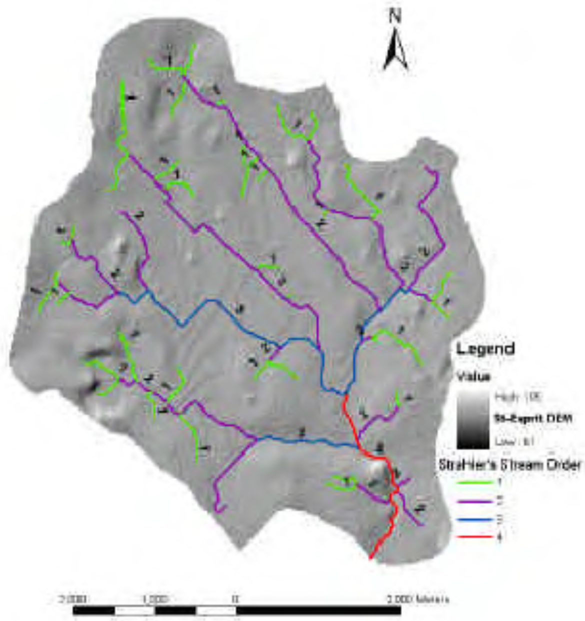

Figure 2. Stream order network of St. Esprit watershed generated using an interface with ArcGIS tool.

Table 1. Geomorphological parameters estimated using WMET interface for St. Esprit watershed, Quebec, Canada

Geomorphological parameter	St. Esprit watershed				
Area (km^2)	26.093				
Perimeter (km)	23.679				
Maximum length (km)	7.35				
Maximum elevation (m)	105				
Minimum elevation (m)	65				
Watershed relief (km)	0.04				
Relief ratio	0.006				
Relative relief	0.002				
Elongation ratio	0.823				
Mean slope (km/km)	0.012				
Stream characteristics (Strahler's stream ordering system)[23]	Number of streams	Length (km)	Area (km^2)	Mean length (km)	Mean area (km^2)
1st order streams	40	15.76	10.08	0.39	0.252
2nd order streams	17	25.89	17.6	1.523	1.035
3rd order streams	3	7.185	17.83	2.39	5.943
4th order streams	1	2.663	25.21	2.663	25.21
Horton's parameters		$R_L = 1.856$	$R_B = 3.597$	$R_A = 4.742$	
Total length of streams of all orders (km)	51.5				
Stream frequency of the watershed (km^{-2})	2.338				
Drainage factor of the watershed	0.60				
Shape factor	1.878				
Form factor	0.533				
Circulatory ratio	0.765				
Drainage density (km^{-1})	1.974				
Ruggedness number	0.079				
Hypsometric integral (H_{si})	0.52				

flow and the effect of geomorphology is not properly reflected in hydrologic response of watersheds. From the recorded dataset of the St. Esprit watershed, 64 rainfall events were selected for analysis and the corresponding Direct Runoff Hydrographs (DRHs) were derived from the total storm hydrograph using the straight-line baseflow separation technique[21]. Further, the peak and mean runoff rates for all the selected events were estimated from the generated DRHs. The collected sediment samples were analysed in the laboratory to estimate the sediment concentrations in the runoff and the average sediment yield rate of individual events was calculated.

Estimation of geomorphological parameters

The ArcGIS® tool of ESRI was used to generate geomorphological parameters by developing an interface, WMET (Watershed Morphology Estimation Tool) within the ArcGIS® environment using Visual Basic for Applications programming language[22]. The stream network of the St. Esprit watershed is shown in Figure 2 and estimated morphological parameters are displayed in Table 1. The stream order and network is based on Strahler's ordering scheme[23], according to which Horton's bifurcation law of stream numbers can be expressed quantitatively as[24]:

$$N_i/N_{i+1} = R_B, \tag{1}$$

where N_i and N_{i+1} are the number of streams in order i and $i+1$ respectively, $i = 1, 2, \ldots \Omega$ and Ω is the highest stream order of the watershed, and R_B is the bifurcation ratio.

$$\bar{L}_{i+1} / \bar{L}_i = R_L, \tag{2}$$

where R_L is the stream length ratio, and the average length of channel of order i is given by

$$\bar{L}_i = \frac{1}{N_i}\sum_{j=1}^{N_i} L_{j,i}, \tag{3}$$

For the drainage area,

$$\bar{A}_{i+1} / \bar{A}_i = R_A, \tag{4}$$

where R_A is the area ratio and average area of order i is given by

$$\bar{A}_i = \frac{1}{N_i}\sum_{j=1}^{N_i} A_{j,i}, \tag{5}$$

where $A_{j,i}$ is the total area that drains into the jth stream of order i.

These empirical laws state that the bifurcation ratio, length ratio and area ratio are unique representative parameters for a given watershed and are fixed values for a given watershed system. These parameters are estimated by graphically plotting stream order against the number of streams, stream length and contributing stream area in a semi-logarithmic graph and calculating the anti-log of the slopes of the best fit lines to obtain R_B, R_L and R_A (Figure 3). The estimated parameters ($R_B = 3.597$, $R_L = 1.856$ and $R_A = 4.742$) of the St. Esprit watershed were within the ranges (i.e. $3 < R_B < 5$, $1.5 < R_L < 3.5$ and $3 < R_A < 6$) generally found for natural watersheds[25,26].

Modelling tools and methodology

ANN

ANNs, which emulate the parallel distributed processing of the human nervous system, have proven to be successful in dealing with complicated problems such as function approximation and pattern recognition. The stored information-processing elements are interconnected and organized in layers. The connection strengths, also called network weights, can be adapted such that the output of the network matches a desired response. In hydrology, prediction of runoff and sediment loss from watershed systems has been a difficult subject due to complexity of the physical processes involved and variability of rainfall in space and time. The most commonly used ANN for hydrological modelling is a feed-forward network with the BP training algorithm[11], which is also capable of nonlinear pattern recognition and memory association. Standard multi-layer feed-forward networks are capable of approximating any measurable function to any desirable degree of accuracy. In that sense, the multi-layer feed-forward architecture gives neural networks the potential of being universal approximators rather than the specific choice of an activation function. In general, application of ANN in modelling, design or problem-solving is preferred in situations where the system response parameters of a real-world phenomenon are either poorly defined or misunderstood, and where observations of the process may be difficult or impossible to perform, and also when it is difficult to recognize the complex relationships between aspects of the process under investigation[27].

Neural network architecture: According to Sudheer *et al.*[3], one of the most critical questions when applying ANN to modelling of the rainfall-runoff process, is what architecture should be used to map the processes effectively. The input vectors to the selected ANN model, the number of hidden layers, the learning rule and the number of output vectors greatly influence the performance of the model. Moreover, there are no fixed rules for developing an ANN, even though a general framework can be followed based on previous successful applications in engineering[3]. Using available data of the study watershed, trial and error approach was employed in the present analysis to select the optimal neural network architecture[15]. Different combinations of input parameters and number of hidden layers

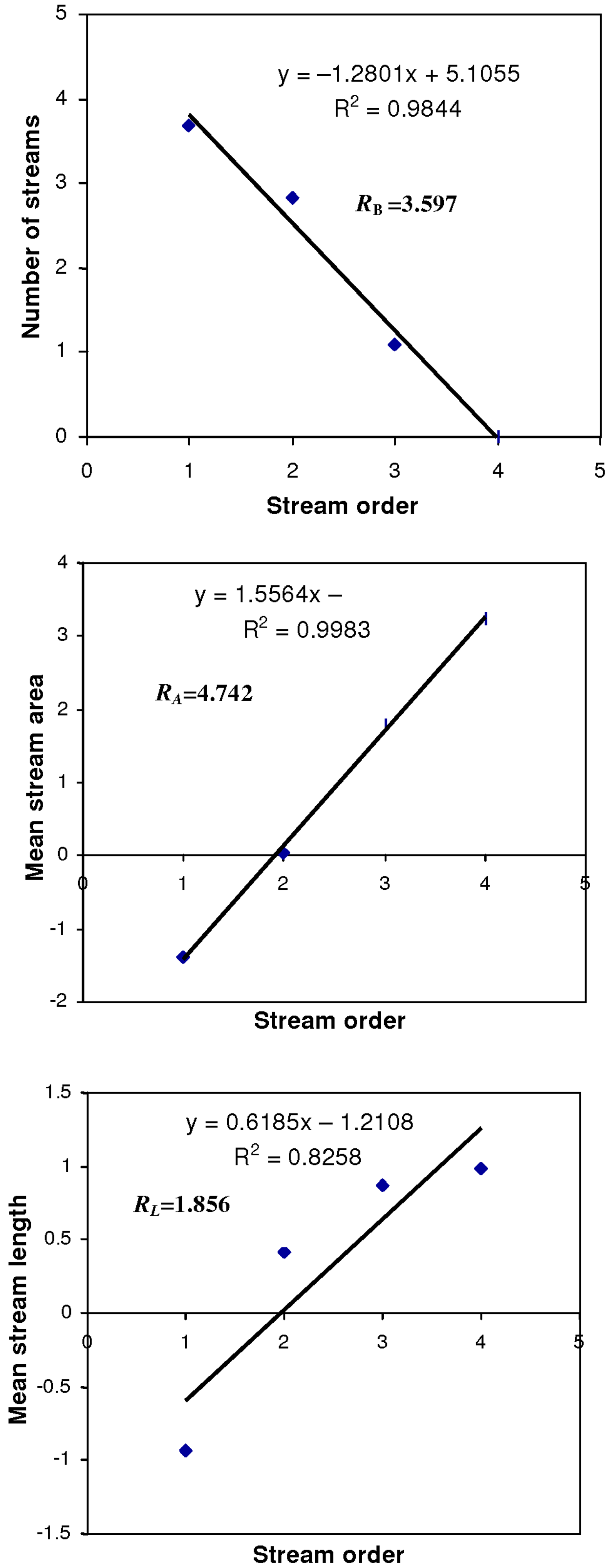

Figure 3. Estimation of R_B, R_A and R_L values of St. Esprit watershed.

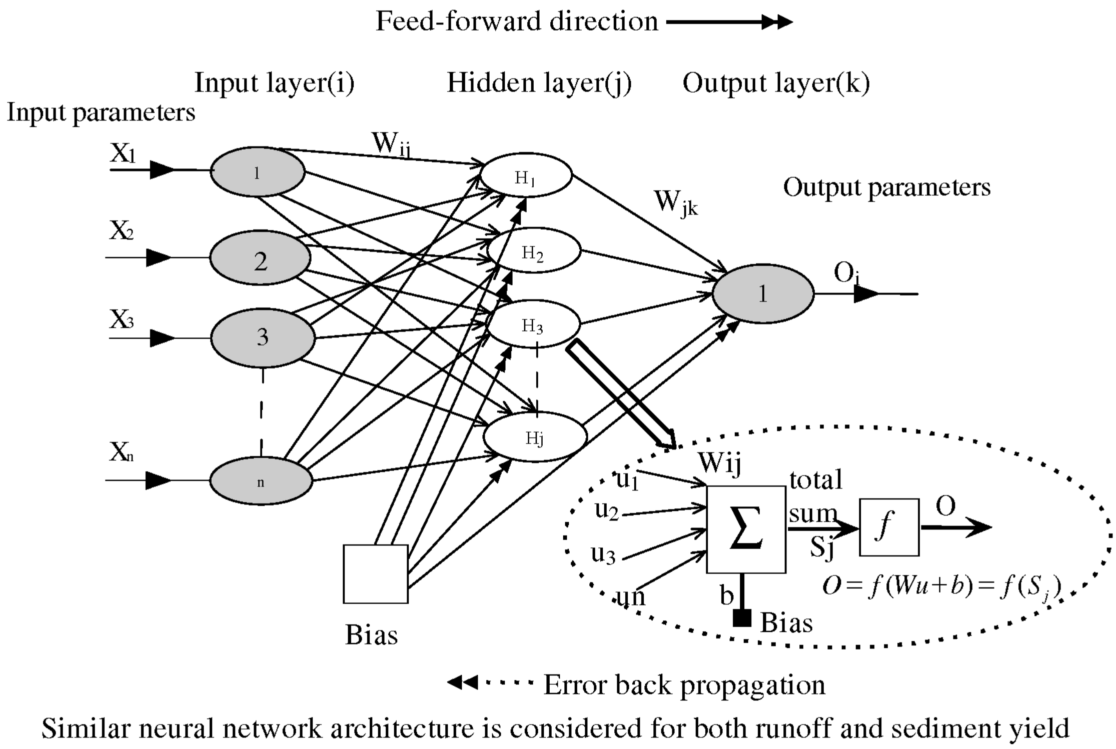

Figure 4. Neural network architecture used for runoff and sediment loss prediction.

corresponding to a single output were tried to select the ANN model architecture. The RMSE, which indicates the performance of an ANN model, was considered as the criterion for selection of optimal architecture. While the inclusion of two hidden layers increased the model running time significantly, it had little effect on RMSE. The number of input parameters in the ANN was determined using the Multvariate Adaptive Regression Spline (MARS)[28], in which the relative importance of each variable was obtained. MARS estimates the relative importance of a variable by comparing the effect of an individual variable on the goodness-of-fit of the regression model and lists variables in the order of their sensitivity with respect to model accuracy.

Therefore, a three-layer feed-forward neural network with BP learning algorithm was selected for rainfall-runoff and runoff-sediment yield modelling (Figure 4).

Each neuron has a number of input arcs u_i (i = 1 to n) connected (Figure 4), and associated with each i, there is a weight W_{ij} which represents a factor by which a value passing to the neuron is multiplied. A neuron sums the values of all inputs.

$$S_j = \sum_{i=1}^{n} W_{ij} u_i + b. \qquad (6)$$

In Figure 4, Wu corresponds to the summation term used in eq. (6). The term b is called a bias. Finally, an activation function is applied to the value S_j to provide a final output O from the neuron. When a BP training algorithm is used for training a network, the sigmoid activation function is most often used[29]. The sigmoid function is bounded above and below (0 and 1), continuous and differentiable everywhere[30]. The sigmoid function (φ) is given by

$$\varphi(S_j) = \frac{1}{1+e^{S_j}}. \qquad (7)$$

Network hidden-layer nodes: The number of nodes in the hidden layer plays a significant role in ANN model performance. Figure 5 shows the RMSE for datasets simulated by ANN with varying number of nodes in the hidden layer. It was observed that the RMSE was minimum for 16 nodes in the ANN model of rainfall and runoff, whereas for runoff and sediment yield ANN model, RMSE was minimum for eight nodes. Also, with further increase in the number of hidden nodes for runoff-sediment ANN model, there was marginal increase in RMSE. However, for eight nodes in the hidden layer, both the ANN models performed optimally with respect to processing time and RMSE estimates. Zhang and Govindaraju[11] used the number of flow paths as the number of nodes in the hidden layer of a GANN. The maximum number of possible flow paths in a watershed drainage network is given by $2^{\Omega-1}$, where Ω is the highest stream order of the watershed. Each flow path is made up of an overland plane and one

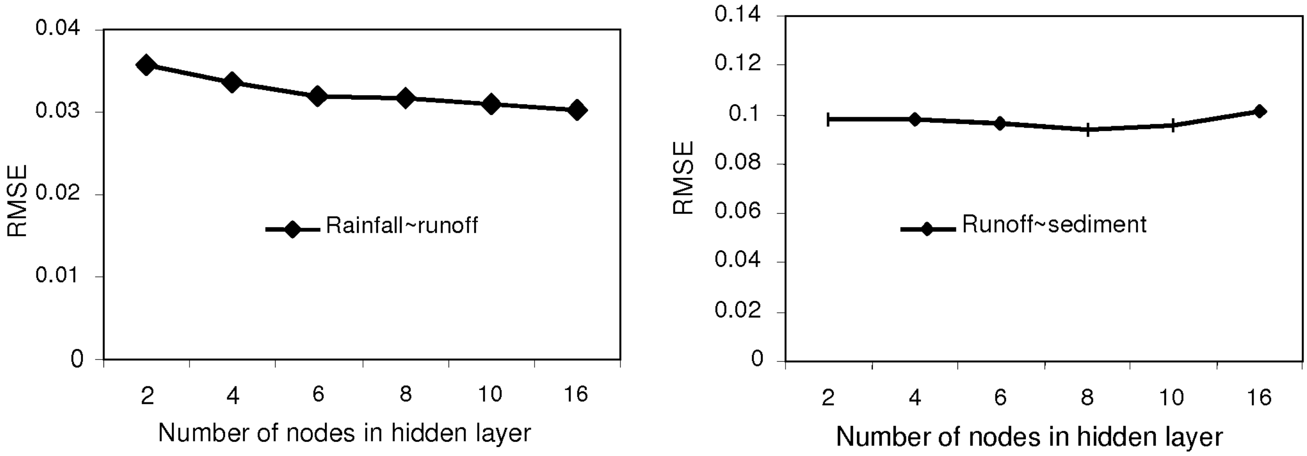

Figure 5. Variation of RMSE with number of nodes in hidden layer.

or more channels. The flow path and path probabilities are responsible for translating runoff water over the watershed towards the outlet[11]. In the present study, the St. Esprit watershed was a fourth order watershed, so the possible flow paths are $2^{(4-1)} = 2^3 = 8$, which conforms to the low RMSE values for eight nodes, as shown in Figure 5. Hence, it is evident that the geomorphological features of the watershed govern the neural network architecture for runoff and sediment yield modelling.

Data preparation and standardization: In the present study, 50% of the data (32 sets) was used for training, 25% (16 sets) for testing and the remaining 25% (16 sets) was used for validation of the ANN models. For ANN model development, the Neural Works Professional II+, version 5.23 was used. For development of regression models, 75% of the data was used for model calibration and 25% for model validation. The MARS software (version 2.0) was used for carrying out sensitivity analysis and prioritizing the parameters which are more sensitive to generation of runoff and sediment yield. The regression models were developed using the curve-fit toolbox of MATLAB 6.5 tool. The dimensionless geomorphological parameters (Table 1) were mathematically associated with the event-based rainfall depth values. A trial and error technique was employed to find the best association in terms of R^2 value of the regression equation generated by association of the input and output parameters. Logarithmic, exponential and trigonometric transformations, as well as different roots of the geomorphological parameters were tried and associated with the rainfall depth in peak runoff rate prediction and associated with mean runoff rate in sediment prediction. The selected composite parameters were then used as independent variables (X_m) for developing a relationship between dependent variables (Y_n), i.e. between peak runoff rate and sediment concentration.

$$Y_n = f(X_m), \tag{8}$$

where

$$X_m = X_1 f(GP_a) + X_2 f(GP_b) + X_3 f(GP_c) + \ldots + X_m f(GP_m). \tag{9}$$

GP_a to GP_m are different geomorphological parameters and $f(GP_a)$ through $f(GP_m)$ are the logarithmic, exponential and root transformations associated as mathematical relationships (e.g. power, multiplication and division) with variables X_1 to X_m respectively.

Due to the nature of the sigmoid function, it is necessary to standardize the data, i.e. to convert it to a range 0 and 1. Without this, for the large values of input variables obtained through eq. (9), the ANN would require extremely small weighting factors and this could cause computational inaccuracies due to floating point calculations and sluggish training, and the gradient of sigmoid function at extreme values would be approximately zero[29]. Therefore, in the present study, the input values were standardized with respect to the range of the dataset (eq. (10)), for better model predictions compared to other approaches of standardization[29].

$$N_i = \frac{R_i - \mathrm{Min}_i}{\mathrm{Max}_i - \mathrm{Min}_i}, \tag{10}$$

where R_i is the real value applied to node i, N_i is the respective standardized value for the node, Max_i and Min_i are respectively, the maximum and minimum of all values applied to the node. The ANN model-predicted output values were destandardized to generate the predicted values of runoff rate and sediment yield for comparison with the observed values.

The datasets were randomized using the Excel™ spreadsheet data-sorting capability to nullify the presence of any existing trend and inherent properties within the data. The presence of specific trends within the data may provide improper training and testing during the ANN model development[29,30]. Therefore, in this study, twenty such shuffled sets were prepared for input to the ANN architecture, and the RMSE of model training and R^2 of the observed and model predicted values were noted. The best ANN model was developed using one of the shuffled datasets having minimum RMSE and R^2 close to one.

Training neural networks: The neural network learns by adjusting the biases and weights linking the neurons. Before training, the initial network biases and weights were assigned small random values[29]. The learning process is similar to the calibration of conceptual models. ANNs are trained with a set of known input and output data. The training process was time-consuming, repeated with a number of different sets of shuffled data. RMSE was noted for each analysis and cross validation was also performed to estimate R^2 values. The learning process is terminated when an optimum prediction statistics is obtained in relation to epoch size and cross-validation results. Epoch is the number of sets of training data presented to the learning cycles between weight updates. It is recommended that the number of epochs should be less than the number of input datasets fed to the ANN model for training and testing. In the present study, the Normalized Cumulative Delta Rule (NCDR) was implemented with the sigmoid transfer function. NCDR was independent of the epoch size due to its normalized function; however, in an attempt to substantiate this concept, epochs ranging from 16 to 30 were implemented. Changes in epoch size had no significant effect on ANN performance using validation datasets. Once the training process was satisfactorily completed, the network was saved, the test and validation datasets recalled, and values predicted by the model were compared with the observed values for the particular events. If the prediction error statistics for these datasets are good, then the neural network model can be considered to perform well for forecasting the runoff with different sets of rainfall data of the watershed.

Regression model development

The MARS software was used to estimate the relative significance of variables through sensitivity analysis. These parameters obtained by operating the MARS tool were further used in MATLAB for development of regression equations. The concept of spline used in MARS is that of the knot, marking the end of a region of data and the beginning of another. The behaviour of functions changes at the knot. MARS finds the location and number of knots needed in a forward–backward stepwise fashion. Basic Functions (BFs) are the machinery used in searching for knots. Hockey stick BFs are used in the final prediction model and serve as the core building block of MARS[28]. The curve-fit toolbox of MATLAB was used extensively to study the data trends and explore different functions to develop a multiple regression model. The equation builder was used to develop an equation relating the morphological parameters. Different possible mathematical associations were tried with the independent variables (eq. (9)). The developed model was applied to a validation dataset and the predicted values were compared with observed values. The model efficiency factor E and R^2 of observed and predicted values were estimated for different predictions on validation datasets. The best model was selected based on the E value approaching one[22]. The model efficiency factor was estimated for all the 20 validation sets using the relation:

$$E = 1 - \frac{\sum_{i=1}^{n} (p_i - o_i)^2}{\sum_{i=1}^{n} (o_i - \overline{o})^2}, \qquad (11)$$

where n is the total number of observations, o_i the ith observed value, $\hat{o}$ the mean of observed values, and p_i the ith predicted value.

Results and discussion

Results of rainfall runoff models

Regression models for peak runoff rate: The best regression equation obtained with the calibration dataset of the total rainfall depth and corresponding peak runoff rate of rainfall events selected for the periods from 1994 to 1997, excluding any geomorphological parameters using the MATLAB tool was:

$$R = 0.047 * P^{1.208} \qquad (R^2 = 0.78), \qquad (12)$$

where R is the peak runoff rate in m^3/s and P is the event-based rainfall depth in mm.

The MARS tool was used to extract the relative importance of geomorphological parameters as associated with rainfall and accordingly, R_A, R_B, R_L, D_f (drainage factor) and R_R (relief ratio) were selected as sensitive parameters for generation of runoff from rainfall. These parameters were included in the development of the regression model and different combinations of parameter values in the curve-fit tool box of MATLAB resulted in a multiple regression model,

$$R = 0.13P^{\sqrt{R_B}} - 0.87P^{\sqrt{R_L}} - 0.02P^{\sqrt{R_A}} + 4.631P^{\sqrt{D_f}} - 48.15P^{\sqrt{R_R}} + 45.5 \qquad (R^2 = 0.88). \qquad (13)$$

Neural network models: ANN models obtained using the 20 shuffled datasets, when subjected to validation, resulted in RMSE ranging from 0.0261 to 0.0573 and R^2 values

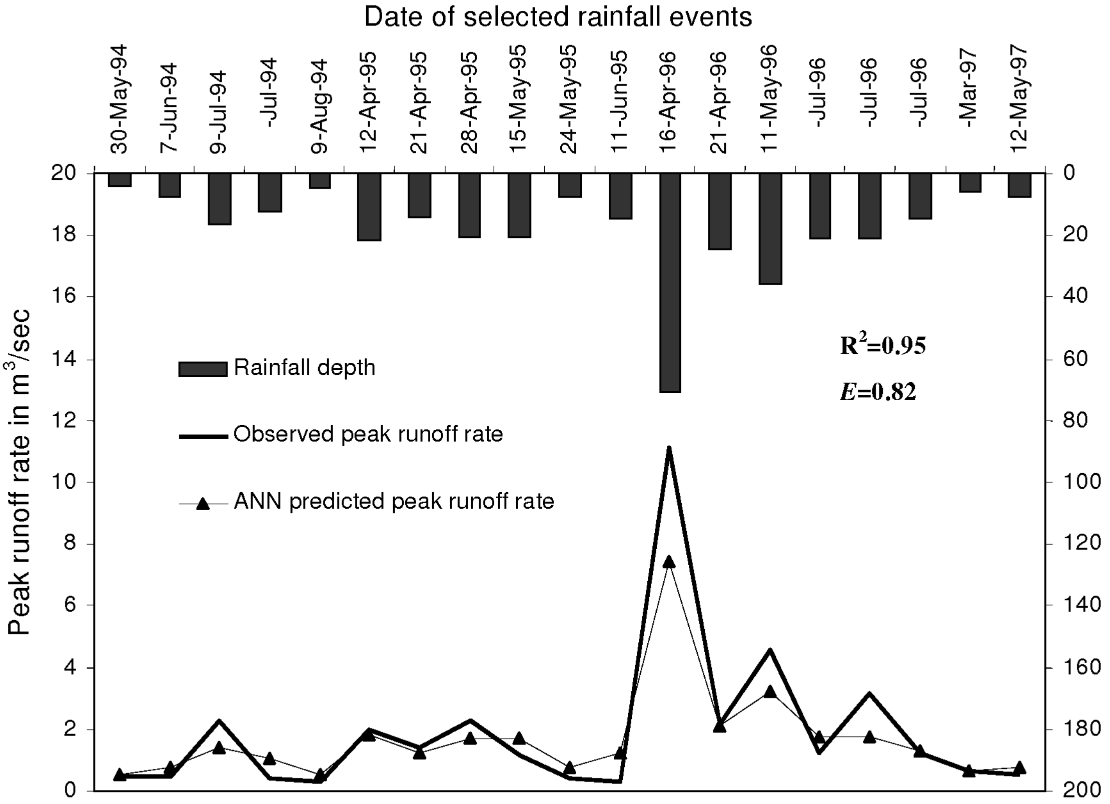

Figure 6. ANN predicted peak runoff rate from geomorphology-based rainfall input parameters.

from 0.69 to 0.76. The ANN models were for the network architecture having single input layer with two nodes, one for event-based rainfall depth and the other for rainfall duration. There were eight nodes contained in a single hidden layer with one node in the output layer as peak runoff rate. The ANN model was first developed without associating any geomorphological parameters. It was observed that the standardization of input and output parameters (eq. (10)) for training and testing the ANN model resulted in slightly better RMSE during model training, but without any significant change in R^2 value during model validation.

Further, GANN models were developed by feeding the data to the network architecture with one input layer having six nodes, where five nodes represented the values of $R^{\sqrt{R_B}}$, $R^{\sqrt{R_A}}$, $R^{\sqrt{R_L}}$, $R^{\sqrt{R_R}}$ and $R^{\sqrt{D_f}}$, and one node was fed with rainfall durations of the selected events. The single hidden layer consisted of eight nodes and the output layer was for one node representing the peak runoff rate values. The input and output values were also standardized using eq. (10). The standardization also resulted in a better RMSE and less significant change in R^2 values. This ANN model architecture was tested, trained and then validated for all the 20 shuffled datasets. The model performance for validation datasets resulted in R^2 values ranging from 0.78 to 0.95 and E values from 0.71 to 0.82 for the shuffled datasets. Results of the model performance with the highest E values of 0.82 and R^2 of 0.95 are presented in Figure 6.

The GANN model predicted peak runoff rates were seen to be in line with the observed values for the selected rainfall events (Table 2). It was observed during the model training that variation in the number of epochs did not improve the model prediction. Finally, the developed ANN model was applied on all the 64 datasets to predict the peak runoff rate values, and the 1 : 1 line of observed and model-predicted peak runoff rates resulted in the coefficient of determination (R^2) of 0.83 (Figure 7). The model efficiency factor E was reduced to 0.61. Moreover, performance of the best-fit ANN model with all the selected datasets revealed that the model can be used to predict peak runoff rate from event-based rainfall depths over St. Esprit watershed with acceptable accuracy.

Results of runoff and sediment yield models

Regression models for sediment yield: Regression models to predict sediment yield were developed from the dataset of mean runoff rates (m^3/s) and mean sediment yield rates (mg/l) for the selected rainfall events over the St. Esprit watershed. The average value of the runoff and sediment flow rates indicates a lumped parameter estimation approach in which the total sediment yield resulting from a DRH can be approximated by multiplying the runoff volume with mean sediment yield rates. Moreover, aver-

Table 2. Observed and predicted peak runoff rates for different rainfall events on St. Esprit watershed using ANN model with validation dataset, $R^2 = 0.95$ and model efficiency (E) = 0.82

Date of rainfall event	Rainfall depth (mm)	Rainfall duration (h)	Peak runoff rate (observed) (m^3/s)	Peak runoff rate (predicted) (m^3/s)
30-May-94	4.2	5.5	0.44	0.528
7-Jun-94	7.6	6.25	0.48	0.746
9-Jul-94	16.4	12.5	2.27	1.389
16-Jul-94	12	11.0	0.41	1.053
9-Aug-94	4.4	3.5	0.32	0.540
12-Apr-95	21.4	14.25	1.98	1.808
21-Apr-95	14	10.0	1.38	1.202
28-Apr-95	20.2	13.5	2.29	1.704
15-May-95	20.2	14.5	1.19	1.704
24-May-95	7.6	5.03	0.43	0.746
11-Jun-95	14.6	11.25	0.30	1.248
16-Apr-96	70.6	14.5	11.12	7.417
21-Apr-96	24.4	11.25	2.17	2.078
11-May-96	35.9	17.56	4.56	3.239
15-Jul-96	20.8	12.5	1.21	1.756
19-Jul-96	14.9	11.5	1.21	1.271
29-Mar-97	5.6	2.5	0.64	0.616
12-May-97	7.6	4.5	0.51	0.746

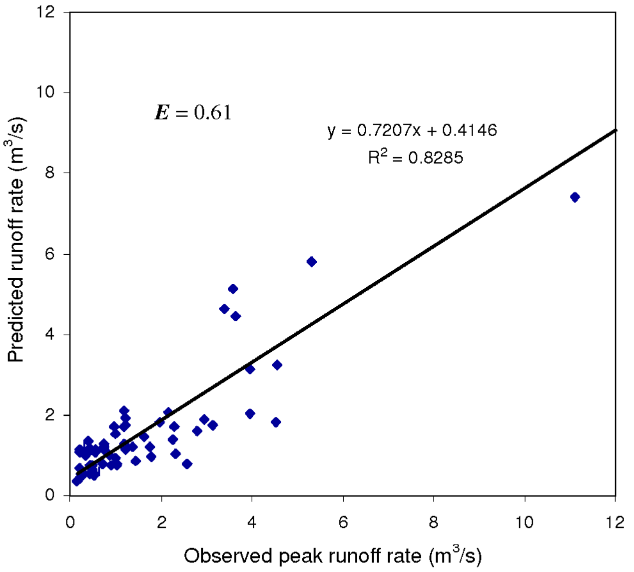

Figure 7. Scatter-plot of observed and computed peak runoff rates using the developed ANN model on complete set of randomized data.

aging of the sedimentation process from DRHs is more useful than the time-paced sediment yield rates, in watershed management for selection of appropriate soil and water conservation structures. While developing the regression models, the mean runoff and sediment loss rates of the selected events from the year 1994 to 1997 were fitted using the curve-fit tool of MATLAB software. The best model derived from these data, without geomorphological parameters, is

$$S_y = 0.1 * R_a^{0.83}, \quad R^2 = 0.39, \tag{14}$$

where S_y is the mean sediment yield (mg/l) and R_a is the mean runoff rate (m^3/s). Using the morphological parameters R_A, R_B, R_L, D_f and R_R, the best-fit equation obtained by trial and error approach was

$$S_y = 2.4R_a^{\sqrt{R_B}} - 6.308R_a^{\sqrt{R_R}} - 0.82R_a^{\sqrt{R_A}} \\ -3.77R_a^{\sqrt{R_L}} + 3.838R_a^{\sqrt{D_f}} - 57.44, \quad R^2 = 0.51. \tag{15}$$

The geomorphological parameters used in eq. (15) were in the order of relative importance obtained from MARS, starting with R_A being the most sensitive parameter having 42% relative importance followed by R_B, R_L, D_f and R_R, with decreasing trend of relative importance of 20, 15, 10 and 8% respectively. It was observed that by including another sensitive geomorphologic parameter (hypsometric integral)[22,25], as indicated by MARS with relative importance of 5%, R^2 was increased by only 0.03 units, i.e. to 0.54, but subsequent addition or deletion of these morphological parameters in the regression equation did not yield better R^2.

Neural network model for sediment yield: Using the approach detailed in previous sections, the R_a and S_y values were shuffled and 20 randomized datasets were prepared for 20 validations. ANN models were developed without associating the morphological parameters having network architecture of a single input layer with one node representing the mean runoff rates of the selected rainfall events, a single hidden layer with eight nodes, and a single output layer with one node of S_y values. The ANN models were trained, tested and validated for all the shuffled datasets obtained, as discussed earlier. RMSE values were different for different datasets and ranged from 0.047 to

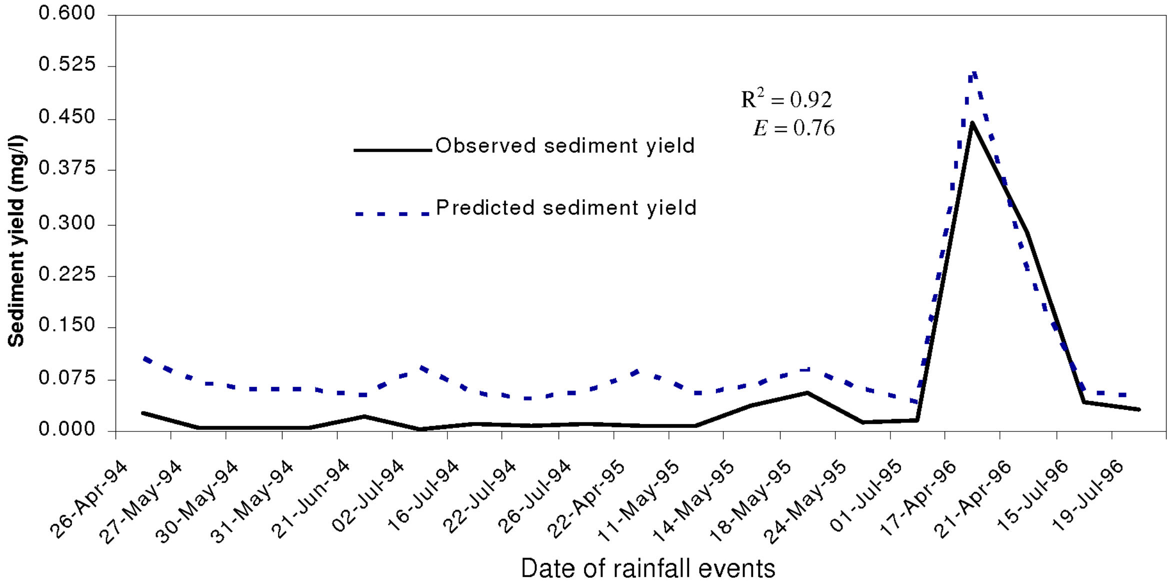

Figure 8. ANN model-predicted sediment yield from geomorphology-based mean runoff rate for different events.

0.069. The validation statistics corresponding to these RMSE values resulted in R^2 values ranging from 0.68 to 0.76 and E values varied from 0.59 to 0.68. The highest R^2 value was for the dataset with the lowest RMSE, whereas the E value for the lowest RMSE was not the highest.

Geomorphological parameters were then associated with rainfall values (eq. (15)) and the ANN architecture with a single input layer with five input nodes representing the values of $R_a^{\sqrt{R_B}}$, $R_a^{\sqrt{R_A}}$, $R_a^{\sqrt{R_L}}$, $R_a^{\sqrt{R_R}}$ and $R_a^{\sqrt{D_f}}$, one hidden layer with eight nodes and a single output layer with one node was trained, tested and validated for all the datasets, as discussed earlier. The resulting RMSE were from 0.0364 to 0.096, R^2 values varied from 0.75 to 0.93 and E values from 0.65 to 0.76. The highest R^2 (0.93) occurred for the lowest RMSE (0.0364), while the highest E value (0.76) occurred for RMSE value of 0.0385. Finally, the ANN model with highest E value (0.76) was selected and the corresponding R^2 was estimated to be 0.92. The observed and predicted values are presented in Figure 8. It can been seen from Figure 8 that the ANN model-predicted sediment yields for a couple of events showed comparatively larger deviation from the observed sediment yield. The recorded data of the watershed revealed that the events of 26 April 1994, 27 May 1994, 2 July 1994, 22 April 1995 and 21 April 1996, which showed more deviation of sediment yield rate, were for longer duration (>10 h) and very low intensities. Such type of events produce less sediment outflow[21], which is also reflected in the recorded sediment yield data of the watershed. Therefore, it was shown that the performance of the developed ANN model in prediction of sediment yield due to short duration and high intensity events was better than the long duration, low intensity events. This may be attributed to the black-box nature of the ANN model, which does not consider the detailed water-budgeting, surface-water routing and accounting of the spatial and temporal watershed hydrologic responses for generation of runoff and sediment losses.

Conclusion

In the present study, an effort was made to compare the ANN and regression models for prediction of peak runoff rate from runoff depth and predication of mean sediment yield rate from the mean runoff rate resulting from the rainfall events of different intensities and durations over the St. Esprit watershed. The Horton's geomorphological parameters were associated with the hydrologic parameters for development of the ANN and regression models. The ANN and regression models developed through combination of soft computing techniques (i.e. ANN and MARS) and mathematical association of the sensitive geomorphological parameters with the rainfall and runoff were standardized for modelling the watershed hydrologic responses. However, neural network and regression models developed for one watershed cannot be applied to other watersheds, and also the functional relationship of geomorphological parameters with rainfall and runoff will differ from one watershed to another. However, the ANN modelling techniques and methods developed through this research can be replicated over other watershed systems to account for hydrological responses. Efforts should also be made to associate hydrological parameters with the watershed morphological parameters through different mathematical functions, to develop geomorphologic association functions leading to more accurate prediction of runoff and sediment losses.

Nonetheless, keeping in view the objectives of the study, it was confirmed that the inclusion of morphological parameters in ANN and regression models enhanced model prediction. In general, GANN models performed better than regression equations. However, performance of the ANN model was more accurate for short duration and high intensity rainfall events of the St. Esprit watershed. These research findings necessitate application to other watershed systems and large datasets in India and abroad under different agro-ecological regions, to strengthen the methodology of ANN-based approaches for runoff and sediment yield prediction.

1. Gautam, M. R., Watanabe, K. and Saegusa, H., Runoff analysis in humid forest catchment with artificial neural network. *J. Hydrol.*, 2000, **235**, 117–136.
2. Hsu, K. L., Gupta, H. V. and Sorooshian, S., Artificial neural network modeling of the rainfall-runoff process. *Water Resour. Res.*, 1995, **31**, 2517–2530.
3. Sudheer, K. P., Gosain, A. K. and Ramasastri, K. S., A data-driven algorithm for constructing artificial neural network rainfall-runoff models. *Hydrol. Process.*, 2000, **16**, 1325–1330.
4. French, M. N., Krajewski, W. F. and Cuykendall, R. R., Rainfall forecasting in space and time using a neural network. *J. Hydrol.*, 1992, **137**, 1–31.
5. ASCE, Artificial neural networks in hydrology–I: preliminary concepts. *J. Hydrol. Eng.*, ASCE task committee on application of ANNS in hydrology, 2000, **5**, 115–123.
6. ASCE, Artificial neural networks in hydrology–II: hydrologic applications. *J. Hydrol. Eng.,* 2000, **5**, 124–137.
7. Gupta, H. V., Hsu, K. and Sorooshian, S., Effective and efficient modeling for streamflow forecasting. In *Artificial Neural Networks in Hydrology* (eds Govindaraju, R. S. and Rao, A. R.), Kluwer, Dordrecht, 2000, pp.7–22.
8. Hsu, K., Gao, X. and Sorooshain, S., Precipitation estimation from remotely sensed information using artificial neural network. *J. Appl. Meteorol.*, 1997, **36**, 1176–1190.
9. Maier, H. and Dandy, G. C., Neural networks for the prediction and forecasting of water resources variables: a review of modeling issues and applications. *Environ. Model. Software*, 2000, **15**, 101–124.
10. Tokar, A. S. and Johnson, P. A., Rainfall runoff modeling using artificial neural networks. *ASCE J. Hydrol. Eng.*, 1999, **4**, 232–239.
11. Zhang, B. and Govindaraju, R., Geomorphology-based artificial neural networks (GANNs) for estimation of direct runoff over watersheds. *J. Hydrol.*, 2003, **273**, 18–34.
12. Sajikumar, N. and Thandaveswara, B. S., A non-linear rainfall-runoff model using an artificial neural network. *J. Hydrol.*, 1999, **216**, 32–55.
13. Hornik, K., Stinchcombe, M. and White, M., Multi-layer feed forward networks are universal approximators. *Neural Networks*, 1989, **2**, 359–366.
14. Smith, J. and Eli, R.N., Neural network models of rainfall-runoff process. *ASCE J. Water Resour. Plann. Manage.*, 1995, **121**, 499–508.
15. Anmala, J., Zhang, B. and Govindaraju, R., Comparison of ANNs and empirical approaches for predicting watershed runoff. *ASCE J. Water Resour. Plann. Manage.*, 2000, **126**, 156–166.
16. Cannon, A. J. and Whitfield, P. H., Downscaling recent streamflow conditions in British Columbia, Canada using ensemble neural networks. *J. Hydrol.*, 2002, **259**, 136–151.
17. Sarangi, A. and Bhattacharya, A. K., Use of geomorphological parameters for sediment yield prediction from watersheds. *J. Soil Water Conserv.,* 2002, **44**, 99–106.
18. Nagy, H. M., Watanabe, K. and Hirano, M., Prediction of sediment load concentration in rivers using Artificial Neural Network Model. *J. Hydraul. Eng.*, 2002, **128**, 588–595.
19. Yitian, L. and Gu, R. R., Modeling flow and sediment transport in a river system using an artificial neural network. *Environ. Manage.*, 2003, **31**, 122–134.
20. Romero, D., Madramootoo, C. A. and Enright, P., Modelling the hydrology of an agricultural watershed in Quebec using SLURP. *Can. Biosyst. Eng.*, 2002, **44**, 1.11–1.20.
21. Chow, V. T., Maidment, D. R. and Mays, L. W., *Applied Hydrology*, McGraw-Hill, NY, 1988.
22. Sarangi, A., Madramootoo, C. A. and Singh, D. K., Development of ArcGIS assisted user interfaces for estimation of watershed morphologic parameters. *J. Soil Water Conserv.*, 2004**, 3**, 139–149.
23. Strahler, A., Quantitative analysis of watershed geomorphology. *EOS Trans. AGU*, 1957, **38**, 913–920.
24. James, I. D. and Burgess, S. J., Selection, calibration and testing of hydrologic models. In *Hydrological Modeling of Small Watersheds* (eds Haan, C. T., Johnson, H. P. and Brakensiek, D. L.), American Society of Agricultural Engineers, St. Joseph, MI, 1982, pp. 215–257.
25. Ritter, D. F., Kochel, R. C. and Miller, J. R., *Process Geomorphology*, McGraw Hill, Boston, 2002.
26. Jain, V. and Sinha, R., Evaluation of geomorphic control on flood hazards through Geomorphic Instantaneous Unit Hydrograph. *Curr. Sci.*, 2003, **85**, 1596–1600.
27. Haykin, S., *Neural Networks*, Prentice Hall, NJ, 1999, 2nd edn.
28. Friedman, J. H., Multivariate adaptive regression splines. *Ann. Stat.*, 1991, **19**, 1–141.
29. Sivakumar, B., Jayawardena, A. W. and Fernando, T. M. K. G., River flow forecasting: use of phase-space reconstruction and artificial neural networks approaches. *J. Hydrol.*, 2002, **265**, 225–245.
30. Dawson, C. W. and Wilby, R., An artificial neural network approach to rainfall-runoff modeling. *Hydrol. Sci. J.*, 1998, **43**, 47–66.

Received 9 May 2005; revised accepted 5 September 2005