Artificial Neural Network Model to Assess the Impacts of Land Development on The River Flow

M.Y. Mohamed

Fac. of Agric., Al-Zaiem Al Azhary Univ., Sudan

Abstract

With rapid land development and limited water resources, good quality water becomes an important commodity. Some types of land development can be associated with increased impervious area that causes increase in surface runoff and decrease in ground water recharge. Both of these processes can have large-scale ramifications through time. Increased runoff results in higher flows during rainfall events. On the other hand the groundwater recharge decreased due to the increase in the impervious surfaces and decrease in the soil infiltration rate. Hence there is a need to quantify the impacts of landuse changes from the point of minimizing potential environmental degradation. The objective of this study is to develop a model for assessing the impacts on the watershed runoff due to changes in landscape patterns. While conceptual or physically based models are of importance in the understanding of hydrologic processes, there are many practical situations where the main concern is with making accurate predictions at specific locations. For this purpose an Artificial Neural Network model (ANN) was developed. Remote sensing was used in this study in view of its providing useful information on landuse dynamics. The model performance in both training and testing phases were checked using mean absolute error (MAE), mean square error (MSE), U Theil's coefficient and regression analysis. The correlation coefficients between simulated and real data were found to be 0.94 and 0.89 for the training and testing, respectively. Most of the data points were within the confidence level of 95 %. The model can be used as a decision making tool when formulating landuse policies. It can be a practical tool for hydrologists, engineers and town and country planners. The model can be used in different areas with different conditions, like dry areas, taking in consideration that the model should be calibrated for the use in new areas.

Key words: Land development, Artificial Neural Network, river flow, remote sensing.

Introduction

Due to land cover changes, many watersheds and river basins soils are converted to impervious surfaces which lead to decrease in the soil infiltration rate and consequently increase the amount and rate of runoff. A lot of rain water makes its way to the sea during the rainy season due to the higher runoff without being used for the human needs. Deforestations, urbanization, and other landuse activities can significantly alter the seasonal and annual distribution of stream flow within a watershed. It is likely that such changes can also affect the seasonal and annual distribution of base flow. Understanding how these activities have influenced stream flow pattern may enable planners to formulate policies to minimize the undesirable effects of future land development.

Most hydrologic processes have a high degree of temporal and spatial variability and are further plagued by issues of non linearity of physical processes, conflicting spatial and temporal scales, and uncertainty in parameter estimation. Determining the relationship between rainfall and runoff for a watershed is one of the most important problems faced by hydrologists and engineers. Hydrologists attempt to provide rational answers to problems that arise in design and management of water resources.

While conceptual or physically based models are of importance in understanding of hydrologic processes, there are many practical situations where the main concern is with making accurate predictions at specific locations. In such situation it is preferred to implement a simple "black box" model to identify a direct mapping between the inputs and outputs without detailed consideration of the internal structure of the physical process. In the class of black box models, a method to predict the runoff response of the watershed on the basis of known meteorological data, hydrologic time series, soil condition and spatial distribution of the landuse could be based on the application of artificial neural networks (ANN). The use of ANN models in water resources applications has grown considerably over the last decade. An attractive feature of ANN is its ability to extract the relation between the inputs and outputs of a process, without the physics being explicitly provided to them.

Shrestha (2002), Stated that the relationship between the changes of the runoff values for the change in rainfall was found to be nonlinear for different landuse. The performance of the network in training and validation using feed forward back propagation network model to predict the runoff from the landuse, soil moisture and rainfall was found to be quite satisfactory and the model can be used for estimation of flows for un-gauged periods.

Remote sensing and geographic information system are increasingly becoming an important tools in hydrology and water resources development; this is due to the fact that most of the data required for hydrological analysis can easily be obtained from remotely sensed images. The conventional methods of detecting landuse changes are costly and low in accuracy. Remote sensing technique, because of its capability of synoptic viewing and repetitive coverage provides useful information on landuse dynamics. The greatest advantage of using remotely sensed data for hydrological modeling is its ability to generate information in spatial and temporal domain which is very crucial for successful model analysis, prediction and validation.

Methodology

For efficient mapping of nonlinear rainfall-runoff process, an Artificial Neural Network (ANN) model was developed to to assess the change in runoff due to landuse changes as the landuse is one of the main model's inputs. This study was conducted in a 200 km^2 watershed located in the Southeastern part of Perak state and Northeastern part of Selangor state, Malaysia. The area lies between 3^0 36' 23'' to 3^0 47' 55'' North and 101^0 30' 53'' to 101^0 39' 33'' East. The area is characterized by high temperature and high humidity with relatively small seasonal variation. The mean relative humidity is 77%, while the minimum and maximum temperatures are 26^0C and 32^0C respectively.

The average rainfall ranges from 2000 to 3500 mm. The mean annual evaporation ranges from 1200 to 1650 mm, and the average daily sunshine hour is 6.2 hours. The wind is calm for most of the year; the average daily wind speed is 89 km/day. Six soil series are found within the study area. The dominant vegetation cover in the river basin consists of tropical hill rainforests, oil palm and rubber. Other land covers that can be found are few small or medium sized urbanized built up areas especially along the river banks and roadsides. The main tributaries of the river are Bernam and Inki Rivers (Figure 1).

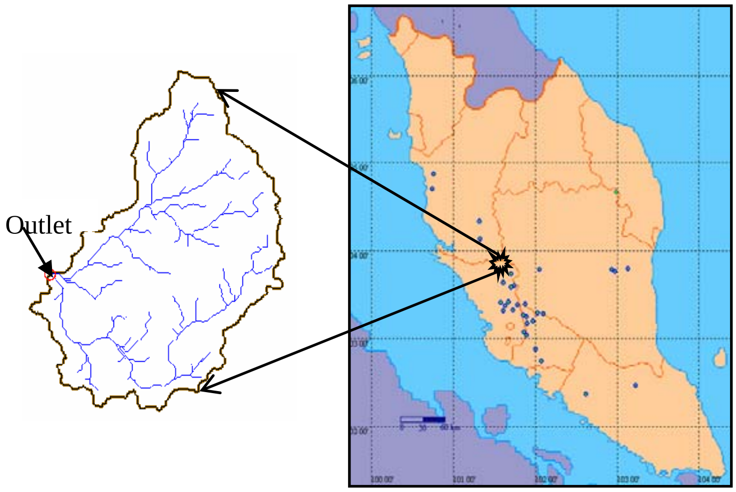

Fig. 1. Location of the study area

1. Landuse change detection

LANDSAT satellite images path/row 127/57 of 30 meter resolution for the years 1989, 1993, 1995, 1998 and 2001 were processed using ERDAS IMAGINE 8.4 software (ERDAS, 1990). The images were enhanced, registered, and classified into different land use types using supervised classification with average classification accuracy of 90%. The false composite colors were used for the visual examination and interpretation. The training signatures to perform the supervised classification were collected from hard copy maps and topographic maps. In areas where there was no distinct spectral signature within the land cover types as a result of mixed pixels the ground truth data was conducted using the global poisoning system (GPS) facilities, and on screen digitizing technique applied to clearly demarcate the classes.

The classified thematic raster maps were vectorized and converted to landuse maps, ARCGIS 8.3, GIS software by ESRI, was used to generate the map's data bases and to perform the needed calculation. Four major types of landuse identified in the study area were forest, rubber, oil palm and built-up areas. The percentage of the landuse areas for different years was calculated and hence the change in the landuse can be detected through the years.

2. Artificial Neural Network model (ANN)

The feed-forward neural network was used in this study, The error back propagation learning algorithm (Rumelhart, 1986), which is the most popular and most used in the field of water resources management was used as training algorithm to train

ANN. Model training was accomplished by providing suitable inputs for the model, computing the output and adjusting the interconnection weights until the desired outputs were obtained. The network architecture that resulted in the minimum error over the training epochs was considered as the optimal architecture, which was obtained by trial and error. The general steps followed to identify and validate the ANN model for this study can be expressed as follow:

(1) Selection of inputs and outputs data those are suitable for calibration and validation.
(2) Selection of the model structure and estimating the model parameters.
(3) Model generalization.
(4) Validation) of the identified model.
(5) Running the model for land use scenarios.

For the purpose of this study six parameters were selected to represent the input layer. There were the monthly rainfall (mm), antecedent soil moisture index and percentage areas of four major landuse found in the study area, namely, Forest, Rubber, oil palm and developed areas. There was only one output from the model which is the monthly runoff (mm). The observed flow data in (mm) for the years 1989, 1993, 1995, 1998, and 2001 were used as target for the purpose of training and testing the model. The average soil moisture index was determined by taking AMC II and III to represent the dry and wet seasons respectively. The index values were taken as integer values that 2 and 3 was used instead of AMC II and III respectively. The data sets were divided into two segments, 85% of the data was used in the training phase, and 15% was used to test the model. The data sets were normalized and scaled to be within the standard range of (0 to 1) which is required by the model's algorism.

There is no standard rule to define the network structure. In this study the selection of the optimum network structure was performed by trial and error. Multilayer networks using the backpropagation algorithm were selected to construct the network. Levenberg-Marquardt (LM) training algorithm was used. Sigmoid transfer function is commonly used in multilayer networks that are trained using the backpropagation algorithm. The Log-sigmoid transfer function was used in the hidden layer while Hard-limit transfer function was used in the output layer. The error goal was set as MSE equal to 0.005 for this study.

The input layer was composed of 6 neurons and the output layer has only one neuron. The hidden layer started with small number of neurons and increased progressively until the optimum structure was reached. Too few neurons lead to underfitting and difficulty in mapping and too many neurons lead to overfitting and increase the training time. Using optimum network architecture, the ANN model was trained for given inputs and output sets.

One of the problems that occur during neural network training is overfitting. The error in the training set is driven to a very small value, but when new data is presented to the network the error becomes large. The network has memorized the training examples, but it has not learned to generalize to new situations. Early stopping technique was used to improve the model generalization. Since the LM training algorithm, which converges too rapidly, was used in this model the training parameters need to be adjusted so that the convergence is relatively slow.

The performance of a trained network can be measured to some extent by the errors on the training, validation and test sets. But it is often useful to investigate the network response in more details. To perform this, a regression analysis between the network output and the corresponding targets were conducted to determine the slope and correlation coefficient (R). The statistical criteria that were used to evaluate the model performance were namely, mean absolute error (MAE), mean square error (MSE) and U Theil's. T-test with 95% confidence for comparing the means of observed and simulated data was carried out. To check the scatter of the output values, 15 % and 20% deviation bands were used for training and testing, respectively. These criteria were employed to measure the goodness of fit of the model and used to test the model efficiency in both training and testing phases.

Final weights and bias values calculated during training phase for the network were used in the testing phase. This phase involves evaluating the network performance on a set of test problems that were not used for training. The model output from the testing process was compared to the observed data and examined using the same statistical criteria that was used during the training phase.

3. Model application

Rainfall pattern from the year 1989 was superimposed to determine the runoff amount based on the landuse of the years 1989,1993, 1995, 1998, 2001 and for the development plans of the year 2020 proposed by the Department of Town and Country Planning, Malaysia. In this plan all the rubber and oil palm areas will be converted to built-up areas. Change in the runoff amount will be due to the change in the landuse.

Results and Discussion

The optimum model structure was accomplished after several trial and error operations to define the number of hidden layers and the number of neurons in each layer. It was found that a network of six neurons in the input layer, one hidden layer with 15 neurons and only one neuron in the output layer (6-15-1),is the optimum structure to model the basin runoff in this study. Figure 2 illustrates the network structure.

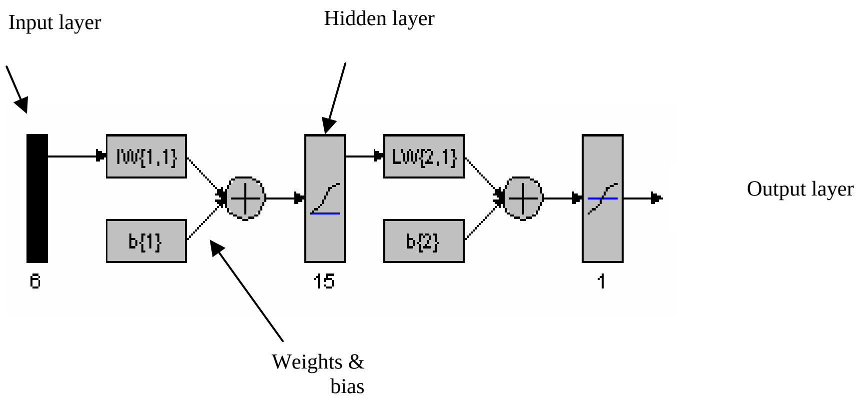

Fig. 2. ANN Structure for UBRB

Table 1 shows the percentage of different landuse for the different five years under investigation.

Table 1: Landuse area percentage in the study area

Landuse type	% of land use area				
	1989	1993	1995	1998	2001
Developed area	4	4	5	6	6
Oil Palm	5	9	10	11	15
Rubber	21	19	17	14	12
Forest	70	68	68	69	67

The simulated values were compared with the observed data (output target) for training and testing data sets. Both graphical and statistical analysis was conducted to check the model performance. Figures 3 to 6 show the relationships between the observed and simulated flow and the best linear fit. the high correlation and linear fit between the simulated and actual data can be observed.

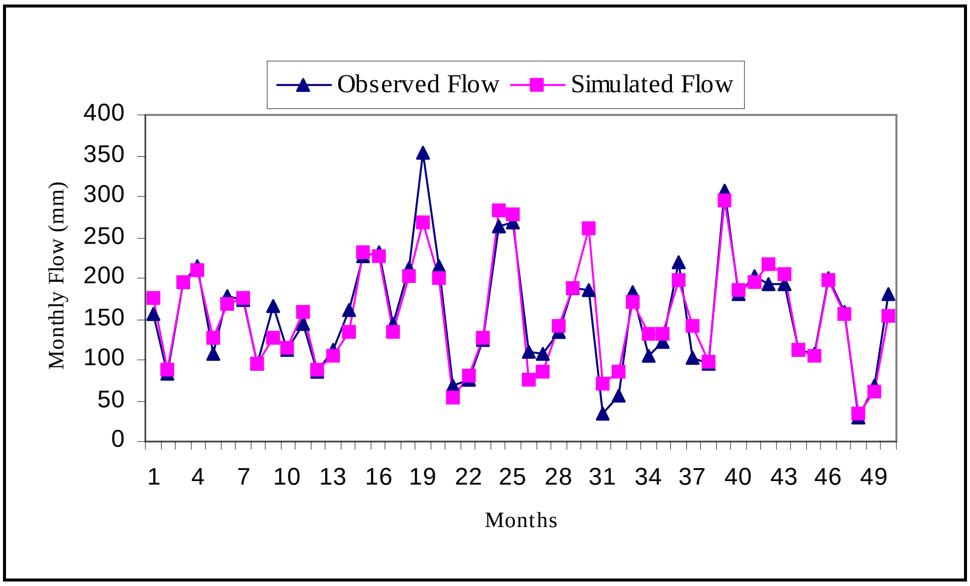

Fig.3. Model performance during the training process

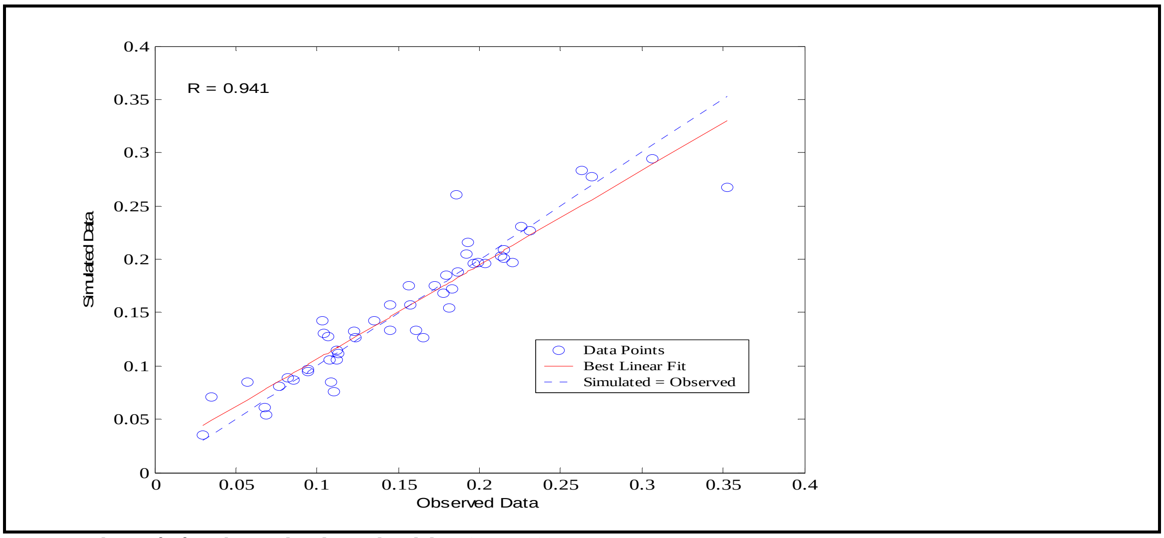

Fig. 4 Best linear fit for Observed and Simulated data (training)

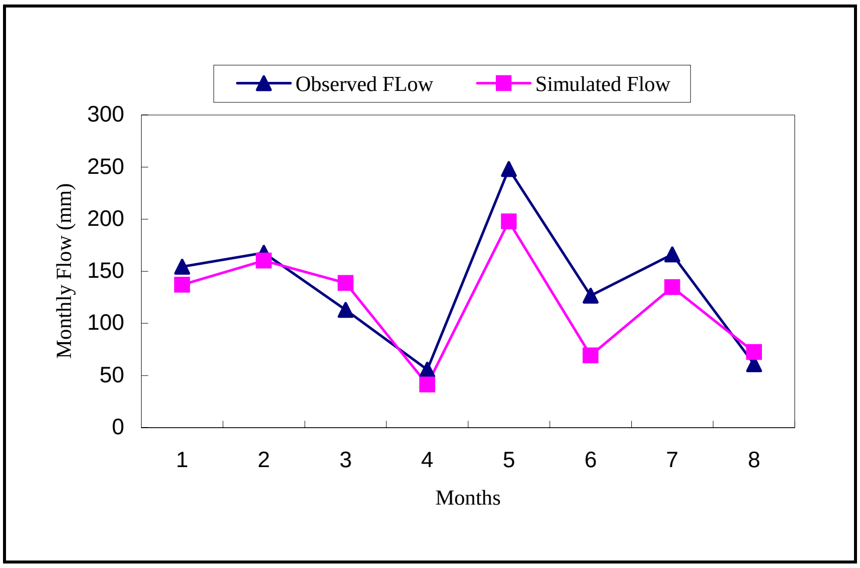

Fig.5. Model performance during the testing process

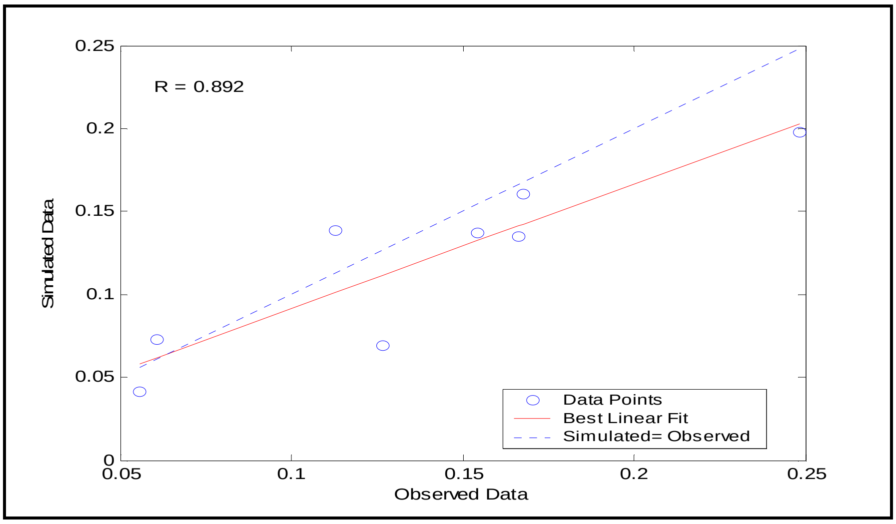

Fig. 6 Best linear fit for Observed and Simulated data (testing)

To evaluate the model performance in predicting the runoff amount, MAE, MSE, U' Thiel test, determination coefficient (R^2) and T-test with 95% confidence level analysis were conducted and the results are shown in Table 2. The model outputs are within the confidence level.

Table 2 Model performance

Statistical Criteria	Model performance	
	Training	testing
MAE	0.001	17.6
MSE	4.77	5.6
R^2	0.88	0.79
U Theil's	0.06	0.11
T-test (95% con.)	0.99	0.13

A plot of the scattered points for the training and testing data show that most of the data points are within the 15 % and 20% deviation lines for the training and testing phases, respectively, (Figures 7 and 8).

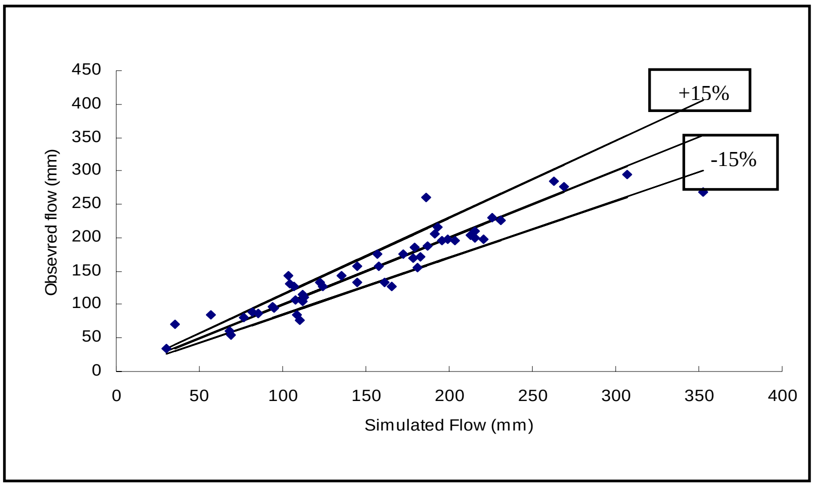

Fig. 7 Scatter plot of the training outputs values

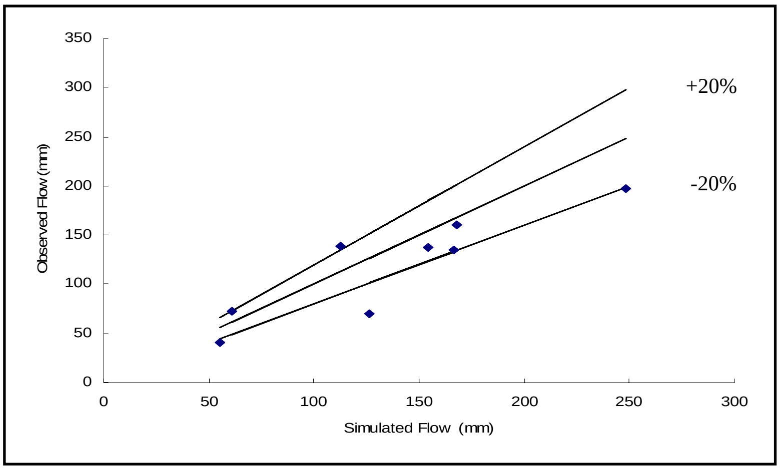

Fig. 8 Scatter plot of the testing outputs values

M and B, correspond to the slope parameters and the y-intercept of the best linear regression relating targets to network outputs. If there is a perfect fit (outputs exactly equal to targets), the slope would be 1, and the y-intercept would be 0. In this model M and B were found to be 0.88, 0.017, respectively for the training phase.

The correlation coefficient factor (R) between the outputs and targets is a measure of how well the variation in the output is explained by the targets. If this number is equal to 1, then there is a perfect correlation between targets and outputs. R value was found to be 0.94 for the training phase. The model performance was check for the testing phase using a set of data that never seen by the model during the model training and it gave 0.85, 0.016 and 0.89 for the M, B, and R respectively. Table 3 shows the regression results from the model performance test.

Table 3: Regression analysis

Regression parameters	Training	Testing
M	0.88	0.85
B	0.017	0.016
R	0.94	0.89

From the above analysis, the model shows high performance in simulating the runoff based on the landuse in both training and testing processes. Hence the model can be applied to simulate the river flow from different landuse scenarios. This will enable the model users to predict the impacts of the landuse changes on the basin's runoff.

From previous studies (Mohan, 2000), there is an evidence that the change in the runoff amount due to the landuse change is constant regardless the rainfall pattern. The same result was obtained by (Mustafa, 2004). Hence a certain rainfall pattern can be used to investigate the impacts of landuse change on the runoff amount by imposing the same rainfall pattern to different landuse combinations.

A rainfall amount of 300 mm from the year 1989 was used to run the model with different landuse pattern for the years 1989, 1993, 1995, 1998, 2001 and from the proposed landuse plan for the year 2020. Figure 9 demonstrates the results of these applications.

From Figure 9 it can be recognized that no significant change in the runoff amount through the years 1989 to 2001. This was due to the lack of land development during that period, while the proposed plan of the year 2020 will increase the monthly runoff rate by 20% compared to the year 2001.

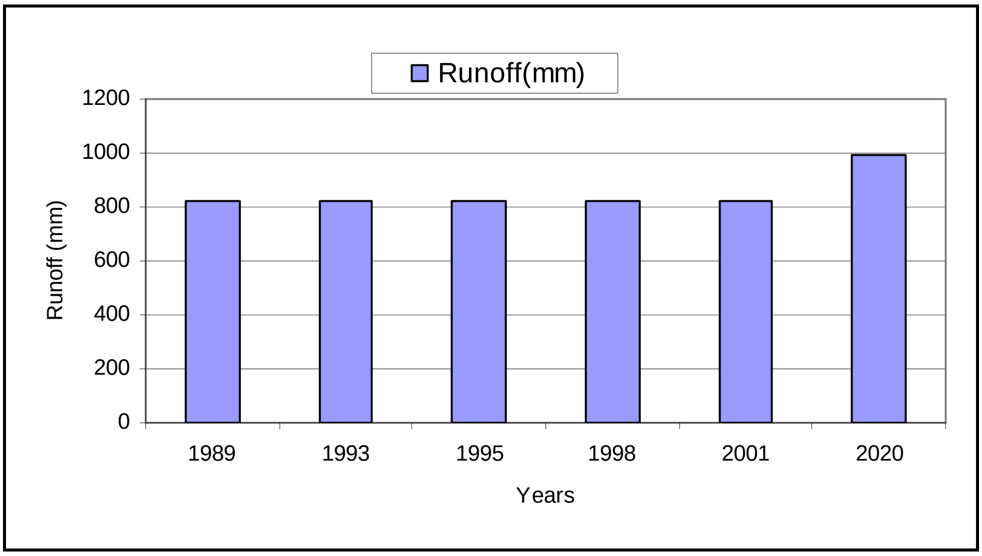

Fig. 9 Change in runoff amount due to change in landuse

The model was test with deforestation/urbanization percentage of 10%, 50% and 80% using the same rainfall amount. The percentage change in the runoff amount due to the change in the landuse is shown in Figure 10 and Table 4.

Conclusion

The following conclusions can be drawn from this study:

- The ANN model shows very good performance in runoff prediction. The model outputs are within 95% confidence level and $\pm$ 15 % and 20% deviation lines for training and testing process, respectively. The correlation coefficient between the observed and simulated outputs was found to be very high for both training and testing phases.
- The determination of optimal network architecture is found to be critical for efficient mapping of rainfall runoff relationship.
- The model can be used for flow estimation during un-gauged periods.
- Development plans of 2020 will lead to increase of monthly runoff amount by 20% compared to the year 2001.
- ANN model with optimal architecture is very useful tool to assess the hydrological effects for a given landuse condition. The model can be used as decision making tool to formulate the landuse policies.

References

ERDAS V.8.4 System Guides, ERDAS, Inc., Atlanta, Georgia. 1990.

Mohan S. and M.N. Shrestha. 2000. A GIS based Integrated Model for Assessment of Hydrological change due to Land use modifications. *Proceeding of symposium on Restoration of Lakes and Wetlands. Indian Institute of Science,* Banglore, India.

Mustafa Y. M., M.S.M Amin, T.S. Lee and A.R.M Shariff. 2004. Assessment methodology for impacts of land development on surface runoff using remote sensing and GIS. *International Conference on Water Resources and Arid Areas, Riyadh, Saudi- Arabia* .

Rumelhart, D.E., G. E. Hinton and R. J. Williams. 1986. Learning internal representation by error propagation. *Parallel Distributed Processing* (Foundations, D. E. Rumelhart and J. L. McClelland (Eds) MIT Press), Cambridge, Mass.

Shrestha M. N. 2002. Integrated ANN Modeling for Assessment of Runoff due to Landuse change using Remote Sensing and GIS. Proceedings of 23rd Asian Conference on Remote Sensing, Kathmandu, Nepal.