International Journal of Water Resources and Environmental Engineering Vol. 3(13), pp. 303 - 307, 8 December, 2011
Available online at http://www.academicjournals.org/ijwree
DOI: 10.5897/IJWREE10.050
ISSN 1991-637X ©2011 Academic Journals

Full Length Research Paper

Artificial neural network model to assess the impacts of land development on river flow

M. Y. Mohamed

Faculty of Agriculture, Al-Zaiem Al Azhary University, Sudan. E-mail: Mustafayousif70@gmail.com.

Accepted 20 June, 2011

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, groundwater recharge decreases due to increase impervious surfaces and decrease 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 physical 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, artificial neural network (ANN) model was developed. Landsat data was used in this study in view of its ability to provide useful information on landuse dynamics. The model's performance in both training and testing phases were evaluated 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 phases 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.

Key words: Artificial neural network, river flow, landsat data, land use changes.

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 of the amount and rate of runoff. A lot of rain water makes its way to the sea during rainy seasons due to high rate of runoff without being used for the human needs. Deforestations, urbanization and other land use activities can significantly alter the seasonal and annual distribution of stream and base flows within a watershed. 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 nonlinearity of physical processes, conflicting spatial and temporal scales, and uncertainties 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, in particular in design and management of water resources.

While conceptual or physical 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 situations, it is preferable 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. 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 land use could be based on the application of artificial neural networks (ANN). The use of ANN models in water resources applications has grown

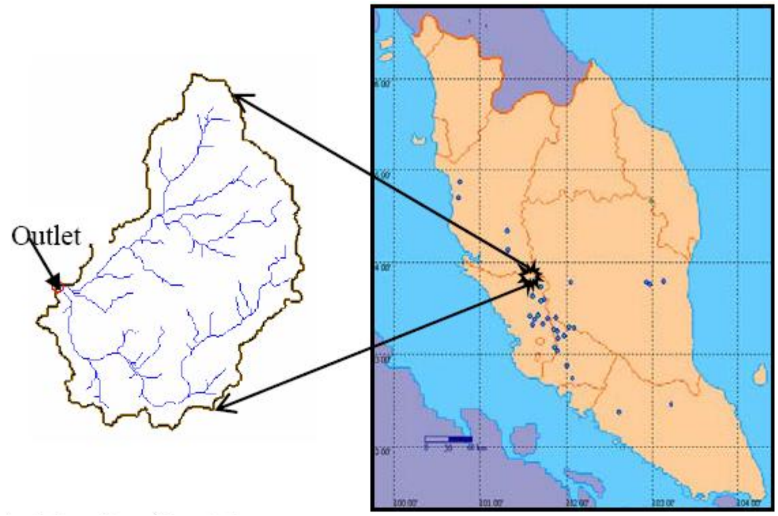

Figure 1. Location of the study area.

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. The major issues that need to be addressed in constructive algorithms include the way additional nodes are connected to the existing network, how to determine optimal connection weights (new and existing) once a node has been added and, when to stop the addition of hidden nodes (Kwok and Yeung, 1997). 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 land use. The performance of the network in training and validation using a feed forward back propagation network model to predict the runoff from the land use, soil moisture and rainfall was found to be quite satisfactory and the model can be used for estimation of flows for ungauged periods.

METHODOLOGY

Data set

This study was conducted in a 200 km watershed located in the southeastern part of the state of Perak and northeastern part of the state of Selangor, Malaysia. The area lies between 3° 36' 23" to 3° 47' 55" N and 101° 30' 53" to 101° 39' 33" E. The area is characterized by high temperature and humidity with relatively small seasonal variation. The mean relative humidity is 77%, while the minimum and maximum temperatures are 26 and 32°C respectively.

The average rainfall ranges from 2,000 to 3,500 mm. The mean annual evaporation ranges from 1,200 to 1,650 mm, and the average daily sunshine is 6.2 h. The wind is calm for most of the year, with the average daily wind speed being 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 a few small or medium sized urbanized areas especially along the river banks and roadsides. The main tributaries of the river are the Bernam and Inki Rivers (Figure 1).

LANDSAT 5 satellite images (path 127 and row 57) of 30 m resolution for the years 1989, 1993, 1995, 1998 and 2001 were processed using ERDAS IMAGINE 8.4 (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 (FCCs) were used for the visual examination and interpretation. The training signatures to perform the supervised classification were collected from hardcopy and topographic maps. In areas where there were no distinct spectral signatures within the land cover types as a result of mixed pixels, ground truth data was collected using Global Positioning System (GPS) facilities, and onscreen digitizing techniques were applied to clearly demarcate the classes.

The classified thematic raster maps were vectorized and converted to landuse maps. ARCGIS 8.3 was used to generate the map's databases and to perform the computations. Four major types of land use were identified in the study area, which are forest, rubber, oil palm and built-up areas. The percentages of land use areas for different years were calculated and hence, changes in the land use can be detected through the years.

Artificial neural network model

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

(1) Selection of inputs and outputs data those are suitable for calibration and validation.
(2) Selection of the model structure and estimation of the model parameters.
(3) Model generalization.
(4) Validation of the identified model.

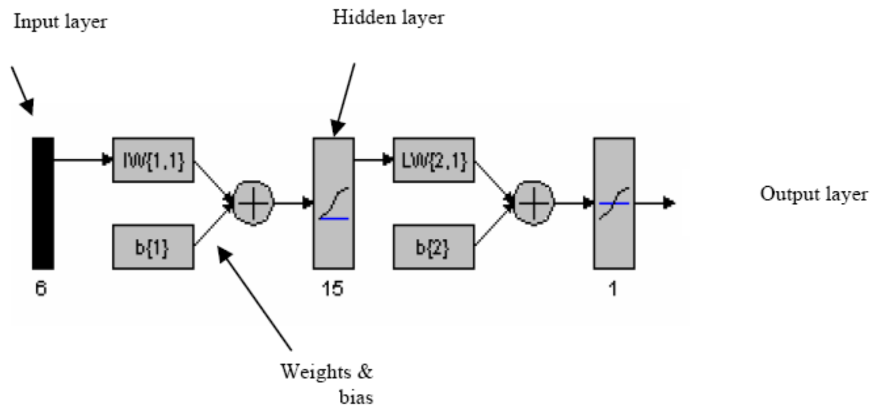

Figure 2. ANN structure for UBRB.

(5) Running the model for land use scenarios.

It has also been suggested that it might be best to fix the number of nodes, rather than the number of hidden layers, and to optimize the connections between the nodes, as well as the connection weights associated with each connection, during training (Kumar, 1993).

For the purpose of this study, six parameters were selected to represent the input layer, which were monthly rainfall (mm), antecedent soil moisture index, and percentages areas of the four major land uses found in the study area, which are 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 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 algorithm.

There is no standard rule to define network structures. In this study, the selection of the optimum network structure was performed by trial and error. Multilayer networks using the back propagation algorithm were selected to construct the network. The Levenberg-Marquardt (LM) training algorithm was used. 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 at mean square error (MSE) of 0.005.

The input layer composed of six neurons, while 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 under fitting and difficulty in mapping, while too many neurons lead to over fitting and increase of training time. Using the 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 over fitting. 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. However, it is often useful to investigate the network's response in more detail. To perform this, a regression analysis between the network output and the corresponding targets were conducted to determine the slope and correlation coefficient R^2. The statistical criteria that were used to evaluate the model performance were mean absolute error (MAE), MSE and U Theil's coefficient. A T-test with 95% confidence was carried out to compare the means of the observed and simulated data. In order to validate the scatter of the output values, 15 and 20% deviation bands were used for the training and testing phases respectively. These criteria were employed to measure the goodness of fit of the model and to test the model efficiency of both the 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.

Model application

Rainfall pattern from the year 1989 was superimposed to determine the runoff amount based on the land use of the years of 1989,1993, 1995, 1998 and 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 into built-up areas. Changes in the runoff amount will be due to the changes in land use.

RESULTS AND DISCUSSION

The optimum ANN 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. Table 1 shows the percentage of different land use for the different five years under investigation. The simulated values were compared with the observed data (output target) for the training and testing data sets. Both graphical and statistical analyses was conducted to validate the ANN model's performance. Figures 3 to 6 show the relationships between the

Table 1. Landuse area percentage in the study area.

Landuse type	% of landuse 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

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

observed and simulated flow for both the training and testing phases. Strong correlations and linear fits were observed for both cases (Figures 4 and 6).

In order to evaluate the model's performance in predicting the runoff amount, MAE, MSE, U Theil's coefficient, R^2 and T-test with 95% confidence level analyses were conducted, and the results are as shown in Table 2. It is observed that the model's outputs are within the confidence level. Plots of the scattered points for the training and testing data show (Figures 7 and 8 respectively) that most of the data points are within the 15 and 20% deviation lines for the training and testing phases respectively.

M and *B* correspond to the slope 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. For this model *M* and *B* were found to be 0.88 and 0.017 respectively for the training phase. The correlation coefficient R^2 between the outputs and targets is a measure of how well the variation in the output is explained by the targets. If R^2 is equal to 1, there is a perfect correlation between targets and outputs. R^2 was found to be 0.94 for the training phase. The model's performance was evaluated for the testing phase using a data set that had not been used during the model's training, and it gave *M*, *B*, and R^2 values of 0.85, 0.016 and 0.89 respectively. Table 3 shows the regression results from the model's performance test.

From the above analyses, the model shows good performance in simulating runoff based on land use in the both training and testing phases. Hence, the model can be applied to simulate river flow for different land use scenarios. This will enable the model's users to predict the

Table 3. Regression analysis

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

impacts of the land use changes on the basin's runoff. From previous studies (Mohan and Shrestha, 2000; Mustafa et al., 2004), there is an evidence that change in runoff amount due to land use change is constant regardless of rainfall pattern. Hence, rainfall pattern can be used to investigate the impacts of land use 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 land use patterns for the years of 1989, 1993, 1995, 1998 and 2001, and the proposed land use plan for 2020. Figure 9 shows the results obtained. It is observed that there was no significant change in the runoff amount through the years of 1989 to 2001. This was due to the lack of land development during that period, while the proposed plan for 2020 will increase the monthly runoff rate by 20% compared to the year 2001. The model was tested with deforestation / urbanization percentages of 10, 50 and 80% using the same rainfall amount. The percentage change in the runoff amount due to the change in the land use is shown in Figure 9.

Conclusions

The following can be drawn from this study:

(i) The ANN model shows very good performance in runoff prediction. The model outputs are within the 95% confidence level, and ± 15 and 20% deviation lines for the training and testing phases respectively. The correlation coefficient between the observed and simulated outputs was found to be very high for both the training and testing phases.
(ii) 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 ungauged periods. Development plans of 2020 will lead to increase of monthly runoff amount by 20% compared to the year 2001.
(iii) 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

Atlanta G (1990). ERDAS V.8.4 System Guides, ERDAS, Inc., Kwok TY, Yeung DY (1997). Objective functions for training new hidden units constructive neural networks. IEEE Trans. Neural Netw., 8(5): 1131–1148.

Kumar K (1993). Optimization of the neural net connectivity pattern using a backpropagation algorithm. Neuron-computing, 5: 273–286.

Mohan S, Shrestha MN (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, Bangalore, India.

Mustafa YM, M Amin MS, Lee TS, Shariff ARM (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 DE, Hinton GE, Williams RJ (1986). Learning internal representation by error propagation. Parallel Distributed Processing (Foundations, D. E. Rumelhart and J. L. McClelland (Eds) MIT Press), Cambridge, Mass.

Shrestha MN (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.