Runoff Generation Using Soil Moisture and Neural Network Models

U. Yadav
Department of Natural Resources and Mines, PO Box 318 Toowoomba, QLD 4350, Australia
(YadavU@nrm.qld.gov.au)

Abstract: Rainfall-runoff modeling is an integral part of a hydrologic model such as the Integrated Quantity and Quality Model (IQQM) being developed for the Lombok catchment in Indonesia. The main objective is to develop a tool for efficient water distribution and management. One of the problems encountered for runoff generation was a shortage of daily data for the purposes of model calibration and validation. Given this limitation, a comparative analysis is being undertaken to select the most appropriate rainfall-runoff model for the Lombok catchment. The paper examines rainfall-runoff generation processes using traditional soil moisture accounting models such as the Sacramento model as well as an artificial neural network back-propagation model (NNM). Simulated streamflows resulted in higher level of fitness in the case of the NNM.

Keywords: Rainfall-runoff modeling; IQQM; Artificial neural network; Sacramento; Back-propagation

1. INTRODUCTION

A research project was undertaken by the Queensland Centre for Climate Applications, Toowoomba, to assess the impact of climate variability in water and crop management in Lombok, Indonesia. The Australian Centre for International Agricultural Research funded the project entitled “Capturing the Benefits of Seasonal Climate Forecasts in Agricultural Management”. One objective of this project was to develop decision support systems to make informed decisions on water allocation, cropping systems, and irrigation strategies based on seasonal climate forecasts (Abawi et. al, 2002).

To fulfill the objectives, a generalized river basin hydrologic model named Integrated Quantity and Quality Model (IQQM) is being developed for the study area to be used as a decision support tool in the next stage of the project scheduled to start in July 2003. Rainfall-runoff modelling is an integral part of the development of the IQQM model. This paper compares rainfall runoff modeling in one river of Lombok island using Sacramento and Back-propagation Neural Network models.

2. DESCRIPTION OF THE STUDY AREA

Lombok island lies in the eastern part of the Indonesian Archipelago and covers an area of approximately 4,800 square kilometers. The catchment area of Sesaot river (Figure 1) at the Keling streamflow gauging station is 35 sq km. The land cover in the catchment is mainly tropical forest. The geological features comprise of tertiary volcanic rocks and sedimentary deposits (Beture Setame, 1992). The river is fed by numerous springs emerging from volcanic ash deposits and therefore, groundwater contribution to the streamflow is significant.

Figure 1. Location Map of Sesaot River

3. OBJECTIVE AND SCOPE OF THE STUDY

The major difficulty in the development of the IQQM model for Lombok has been a lack of the streamflow data needed for model calibration and validation. Therefore, it has been decided to use a rainfall-runoff model to generate streamflow data. Rainfall-runoff models can be classified into two groups – black box models and process models. NNM is a black box model that relates input and
output through complex mapping functions while the Sacramento model relates rainfall-runoff relationships through conceptualization of the hydrological processes in the catchment. A comparison of black box and process models is given by Chiew et. al. (1993), however, this does not include neural network models.

Hsu et al. (1993) found that an artificial neural network model matched the hydrograph most closely for a medium size river in Collins, Mississippi, USA, while the recession and low flow performance was better than that of the Sacramento model. Sajikumar and Thandaveswara (1999) found that an artificial neural network (ANN) was the most efficient of the black-box models tested for calibration periods as short as 6 years. These results prompted the researcher to conduct a comparison test of both the Sacramento and neural network models for the Sesaot river. It is expected that the research would contribute in the selection and the application of an appropriate model, as no comparison of rainfall-runoff models for this river has been carried out before.

Usually 20-30 years of daily streamflow and rainfall data is desirable for calibration and validation of rainfall-runoff models to cover different climate cycles. However, for the Sesaot river, only 8 years of daily rainfall and streamflow data were available. This limitation applies to both the modeling approaches. In this study, both models were run using the same data and period of analysis. The criteria used for comparison are correlation coefficient, coefficient of efficiency and volume ratio. Although qualitative in nature, there is no substitute for visual inspection of observed and simulated hydrographs, therefore these graphs are also included.

4. SACRAMENTO MODEL

The Sacramento model is a spatially lumped continuous soil moisture accounting model developed by the United States National Weather Service and the California Department of Water Resources (Burnash, 1973). The schematic diagram of the model is shown in Figure 2.

Although the soil mantle can be divided theoretically into infinite zones, in practice, it consists of mainly two zones: the upper zone and lower zone as shown in Figure 2. The upper zone represents short-term storage capacity and contributes primarily to surface run-off and interflow. The lower zone characterizes long-term storage capacity and contributes supplemental and primary groundwater flow. The input data for the Sacramento model consists of catchment area, spatially averaged precipitation, evaporation, pan coefficient and observed flows for calibration and validation of the model. In addition, it requires initial values of the soil water storage contents and ordinates of the unit hydrograph for surface flow. The soil-water characteristics and conceptual linkages among themselves are defined by a set of parameters estimated during calibration. One of the important criteria for parameter estimation is the sufficiency of system excitation given the input during the validation period (Wood and O’Connell, 1985).

5. BACK PROPAGATION NEURAL NETWORK MODEL

A standard back propagation neural network model (NNM) is a multi-layer perceptron trained through the back propagation of errors algorithm developed by Werbos (1974), and Rumelhart et. al. (1986). A multi-layer perceptron is a feed

Figure 2. Conceptualisation of Sacramento Model
forward net with one or more hidden layers of nodes or neurons between input and output layers. The perceptron is a neural network model with learning and adaptation features developed by Rosenblatt (1958).

A feedforward multi-layer neural network model is shown in Figure 3. The circles in the diagram represent neurons and $x_1, x_2, ..., x_n$ are the inputs applied to them. An arrow connects the neurons in two adjacent layers and holds the synaptic weight reflecting the strength of the linkage. These weights are adjusted by a feedback mechanism during the learning process.

At each neuron $i$, the incoming signals are multiplied by the synaptic weights and added together to get an effective input signal as given by (1).

$$Y_i = \sum_{j=0}^{n} w_{ij} x_j$$

where $Y_i$ is the effective signal at neuron $i$, $w_{ij}$ is the synaptic weight associated with the arrow linking neuron $i$ and neuron $j$, $x_j$ is the incoming signal and $n$ is the total number of signals.

Then, a threshold function is applied to compute the activation of the neuron or output. The output or activated value $O_i$ at neuron $i$ is given by

$$O_i = f(Y_i)$$

where $Y_i$ is the effective input signal, $\theta$ is a bias term similar to a threshold and $k$ is the gain of the sigmoid whose value may vary from $-\infty$ to $+\infty$. In this study, $k$ has been fixed to 1. Under the backpropagation algorithm, the weights are adjusted recursively working from the output nodes towards the first hidden layer using equation (4) below:

$$w_{ij}^{t+1} = \alpha w_{ij}^t + \eta \delta_j O_i$$

where $w_{ij}^{t+1}$ and $w_{ij}^t$ are the synaptic weights at time step $t+1$ and $t$, respectively, $\alpha$ is the momentum factor, $\eta$ is the learning rate, $\delta_j$ is the error signal term and $O_i$ is the output of neuron $i$ or an input. The error signal $\delta_j$ for an output node $j$ is given by

$$\delta_j = O_j (T_j - O_j)(1 - O_j)$$

where $T_j$ is the target output of node $j$ and $O_j$ is the actual output. For an internal hidden node, the error signal is

$$\delta_j = O'_j(1 - O'_j) \sum_p \delta_p w_{pj}$$

where $O'_j$ is either output of node $j$ or an input.

The network architecture of the NNM model used in this study consists of three layers of feedforward net — input, output and a hidden layer. The determination of the correct number of hidden layers and the number of neurons in an ANN needed to solve a specific task is still an open problem (Birikundavyi et al., 2002). However, a single hidden layer network can approximate any continuous and bounded multivariate function provided that sufficient numbers of neurons exist (Cybenko, 1989). Further, De Villars and Barnard (1993) found that a two hidden layer network converges with less accuracy than its single hidden layer counterpart. Therefore, a one hidden layer network was chosen.
in this study.

A trial and error experiment was conducted with 5, 10, 15, 30, 40, 50 and 90 neurons and it was found that 40 neurons in the hidden layer with 90 neurons in the input layer and one neuron in the output layer provided the best performance for this study. The computer execution time increased from 32 seconds to 596 seconds, when the neurons were increased from 40 to 90 with very little improvement on the accuracy.

6. METHODOLOGY

Three modeling scenarios were tested with the NNM. In the first scenario NNM-1, the input data to nodes were streamflow data for the previous 30 days and evaporation and rainfall data from the present day to the last 29 days. The second scenario NNM-2 involved only evaporation and rainfall data, while in the last scenario NNM-3 only rainfall data were used for the same corresponding period of NNM-1.

The recorded streamflows of Sesaot river were divided into three periods: (1) Training period (2) Testing period and (3) Validation period for the NNM modeling. The training data were used to find the optimal set of connection weights and the testing data were used to avoid over-fitting and to find the effects of other parameters such as learning rate, momentum factor and number of neurons in hidden layers. Over-fitting occurs when the network performs well on the training data but poorly on the test data (Bowden et. al., 2002). To avoid over-fitting, the training of the net was stopped when the improvement in the coefficient of efficiency was less than 0.0001 for NNM-1 and 0.001 for NNM-2 and NNM-3.

Maier and Dandy (2000) stated that the data sets for validation period must be kept separate from the model development process. Accordingly, these data sets were used only to test the generalization capability of the model and to compare the results with the Sacramento model.

As the available records were very short, the training period was limited to the first year. The other two periods were selected such that the mean and standard deviation of the samples do not change significantly from one period to another. The statistics of the observed data in these periods are shown in Table 1.

The observed data were transformed such that their values lie between 0.9 and 0.1 in order to avoid flat regions of the sigmoid function. Random values lying in the range of -0.5 to +0.5 were assigned to the initial weights to overcome the symmetry problem. Usually, the learning rate is selected between 0.05-0.25 to ensure that the network will find an optimum solution (Burian et. al., 2001, Kutza, 1996). In this study, a learning rate of 0.2 was used for all three scenarios and a momentum factor of 0.6 for NNM-1 and 0.9 for NNM-2 and NNM-3 produced optimum results after experimenting with various combinations of these values.

The Sacramento model was run using the procedure defined in the IQQM User Manual (DLWC 1999). Hsu et al. (1993) used six months of buffer period in their Sacramento model to minimize the effects of errors in the initial state variables on calibration results. In this study, the first year was used as the buffer period. The observed and simulated hydrograph were matched visually paying special attention to the end of this period. The outputs for soil storage contents at the end of the buffer period were input as initial storage contents for the next calibration period of two years. In this stage the Sacramento model parameters were optimized based on the coefficient of efficiency with a target value of unity. The model was then run with the calibrated parameters for the validation period.

7. RESULTS

The observed and simulated hydrographs from the Sacramento model for the validation period are shown in Figure 4 and the statistics of the observed and simulated values are shown in Table 2.

While it is common practice to get a correlation coefficient of 0.80 or more in rainfall-runoff modeling for broader acceptance, it was not possible to achieve this for the Sacramento model. There may be numerous reasons including unmeasured abstraction of the flows, uneven distribution of rainfall in the catchment and quality of recorded data. The statistics for the three NNM scenarios are shown in Table 2. The observed and NNM simulated hydrographs for the validation period are plotted in Figure 5.

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Training Period</th>
<th>Testing Period</th>
<th>Validation Period</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Jan 92-Dec 92</td>
<td>Jan 93-Dec 94</td>
<td>Jan 95-Dec 96</td>
</tr>
<tr>
<td>Mean Flow ML/day</td>
<td>232</td>
<td>289</td>
<td>248</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>153</td>
<td>220</td>
<td>163</td>
</tr>
<tr>
<td>Skewness</td>
<td>0.86</td>
<td>1.04</td>
<td>1.04</td>
</tr>
</tbody>
</table>
### Table 2. Statistics of the Observed and Simulated Flows during Validation Period

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Observed Flows ML/day</th>
<th>Simulated Flows ML/day</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Sacramento</td>
</tr>
<tr>
<td>Mean</td>
<td>248</td>
<td>251</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>1.04</td>
<td>0.20</td>
</tr>
<tr>
<td>Skewness</td>
<td>0.04</td>
<td>0.999</td>
</tr>
<tr>
<td>Volume ratio (simulated/observed)</td>
<td>0.748</td>
<td>0.963</td>
</tr>
<tr>
<td>Correlation coefficient</td>
<td>0.491</td>
<td>0.926</td>
</tr>
</tbody>
</table>

NNM-1 gives the best result, but it uses previous day observed flow as an input and so this method is only useful for real time flow forecasting. With the second scenario of NNM similar input data to the Sacramento model has been used. The correlation coefficient and coefficient of efficiency for NNM-2 are found to be higher than for the Sacramento model.

However, for NNM-3 with only rainfall data as input, the correlation coefficient and coefficient of efficiency deteriorate to the same level as for the Sacramento model. This result shows that evaporation plays an important role in synthesizing streamflow data with NNM.

![Figure 5. Observed and NNM simulated flows during validation period.](image)

8. **CONCLUSIONS AND RECOMMENDATIONS**

There is no substitute for sufficient high quality data in rainfall runoff modeling, however, a rigorous analysis with both conceptual and black box modeling techniques such as NNM provides more confidence in the outcomes of the modeling.

- NNM-1 produced excellent results with a correlation coefficient and coefficient of efficiency of 0.963 and 0.926, respectively. The simulated and observed peaks of hydrographs for this scenario also match very well (Figure 4).
- Consequently, there seems significant scope for using NNM for real time flow forecasting and further research will enhance the acceptability of NNM concepts.

![Figure 4. Observed and Sacramento simulated flows during validation period.](image)

Given the linkages between temperature and evaporation, it is interesting to note that temperature did not play a significant role in a study carried out by Cannas et al., (2001). As NNM is a black box model, it is recommended to perform significance tests on the improvement in the accuracy of flow prediction if other climate variables such as temperature and relative humidity are used in the NNM model.

The Sacramento model component of IQQM does not have a feature to provide previous day observed flow as input for real time flow forecasting and NNM has this capability. It showed excellent results in this study and therefore, it is recommended to explore this area further.

9. **ACKNOWLEDGMENTS**

The author is grateful to David Freebairn, Geoff Titmarsh, Sunil Dutta, John Ritchie and the reviewers for commenting on the manuscript.
9. REFERENCES


