Modeling of Rainfall - Runoff relationship using Artificial Neural Networks

Nimiya Baby
Department of Applied Mechanics and Hydrauliics
National Institute of Technology Karnataka, Surathkal
Karnataka, India
nimiyababy110@gmail.com

Dr. Varija K.
Associate Professor, Department of Applied Mechanics and Hydrauliics
National Institute of Technology Karnataka, Surathkal
Karnataka, India
varija.nitk@gmail.com

Abstract— Relationship between rainfall and runoff plays a significant role in generation of stramflows. The objective of the study is to model the rainfall – runoff using daily, weekly and monthly data for a catchment in the coastal Karnataka region using Artificial Neural Networks. The study uses data from two rainguage stations and a riverguage station located within the catchment. Fifteen models were developed using different input combinations which included 11 daily, 2 weekly and 2 monthly models. The efficiency of the models were compared using the statistical parameters - Coefficient of Correlation (r), Index of Agreement (d), Nash-Sutcliffe Efficiency (NSE) and Root Mean Square Error (RMSE). The results indicate that the daily model with daily past one day rainfall, past 2 day rainfall, past one day maximum temperature and past one day runoff as inputs was the best. The results can be used for any future studies of the catchment.

Keywords- *ANN, empirical models, India, modeling, rainfall – runoff*

*****

I. INTRODUCTION

The need for accurate information on watershed runoff has grown rapidly during the past decades because of the acceleration of watershed management programs for conservation, development, and beneficial use of all natural resources, including soil and water. The overall objectives of all watershed management programs are: To increase infiltration into soil, to control excess runoff, to manage and utilize runoff for useful purposes, and to reduce soil erosion to protect land. Therefore, the prerequisite for any watershed development plan is to understand the hydrology of the watershed and to determine the runoff. However runoff analyses are very important for the prediction of natural calamities like floods and droughts. It also plays a vital role in the design and operation of various components of water resources projects like barrages, dams, water supply schemes, etc.

Hydrologists are often confronted with problems of prediction and estimation of hydrologic variables, such as runoff, precipitation, contaminant concentrations, water stages, groundwater recharge, and soil moisture. Hydrologic processes, like runoff are complex, nonlinear, and exhibits both temporal and spatial variability since they depend on various factors, e.g., initial soil moisture, land use, watershed geomorphology, evaporation, infiltration, distribution and duration of the rainfall.

Rainfall–runoff (R–R) models of river basins enable modellers to forecast river discharge. Such forecasts should be accurate and reliable in order to be effectively used for warning against hydrological extremes and for water resources management purposes. Over the years, several hydrological models ranging from empirical relationships to physically based models have been developed for the prediction of runoff. The use of a particular model depends on the complexity of the modelled process, the availability of data etc. In general, model is a copy of the original object. However, a model is not identical with the original. It only retains features which seem to be most crucial from some particular point of view.

A physically-based model use mathematical framework based on mass, momentum and energy conservation equations in a spatially-distributed model domain and parameter values that are directly related to catchment characteristics, while a conceptual model is based on the principle of mass conservation and simplified forms of momentum and energy conservation principles. According to the spatial description of the watershed process, hydrologic models can be classified as lumped or distributed. In a distributed model the spatial variability of vegetation, soil, topography, etc. is taken into account, while in a lumped model the spatial variability of watershed characteristics is ignored. The conceptual models are usually lumped while the physically based model in practice has to be distributed. Distributed hydrologic models require huge data and model parameter identification that may limit their strength and their adequacy for operational purposes.

Empirical (black box) models follow a data-driven modelling approach that is based on extracting and re-using information that is implicitly contained in hydrological data without directly taking into account the physical laws that underlie the hydrological transformation processes. Artificial Neural Networks (ANNs) have been pro-posed as efficient tools for and prediction in hydrology, as black-box models. ANNs are supposed to posses the capability to reproduce the unknown relationship existing between a set of input variables (e.g., rainfall) of the system and one or more output variables (e.g., runoff) [1]. Recent studies in Rainfall – Runoff modeling prove the same. Rajurkar et.al. presented an approach for modeling daily flows during flood events by coupling a simple linear (black box) model with the ANN [2]. Singh et.al. modelled Rainfall runoff using multi layer perceptron technique for Upper Kharum catchment in Chattisgarh using daily precipitation & runoff for 10 yrs of data and compared with linear regression models [3]. Sarkar & Kumar examined the applicability of ANN to model event-based rainfall-runoff process [4]. Fernando Machado et al. modelled rainfall-runoff processes using monthly data of Jangada river basin, Brazil [5]. Raghuvanshi et.al. performed Runoff & Sediment yield modeling using ANN of Upper Siwane River, India using rainfall & temperature for monsoon season of 1991-97 and developed 5 models (3 daily & 2 weekly) with 1or2 hidden layers & compared with linear regression models [6].

The objective of the study is to model the rainfall – runoff processes using daily, weekly and monthly data for the Pavanje catchment in the coastal Karnataka region.

II. MATERIALS AND METHODS

A. Study area

The study area selected is Pavanje catchment covering an area of 202.9 km^2 , located in Dakshina Kannada district of Karnataka in Southern part of India. The Pavanje river originates in the foothills of Western Ghats and joins Arabian sea. The catchment lies between 12˚58'18"N & 13˚5'19" N latitudes and 74˚46'24" E & 75˚0'34" E longitudes (Fig. 1).

The raingauge stations located within the catchment are Bajpe and Surathkal. The catchment has also got a river gauge station at Kateel. The area experiences a hot and humid type of climate. The south-west monsoon (June-Sept) is the principal rainy season for the region and the average annual rainfall is more than 2500mm. The post-monsoon season (Oct-Jan) receives occasional rainfall due to the north-east monsoon. The pre-monsoon season (Feb-May) is essentially the summer season with scanty pre-monsoon showers during April-May. The observed data used in the study includes daily rainfall, temperature and discharge for a period of 8 years, from January 1, 2001 to December 31, 2008. The required data is obtained from Indian Meteorological Department, Pune and Karnataka Water Resources Department.

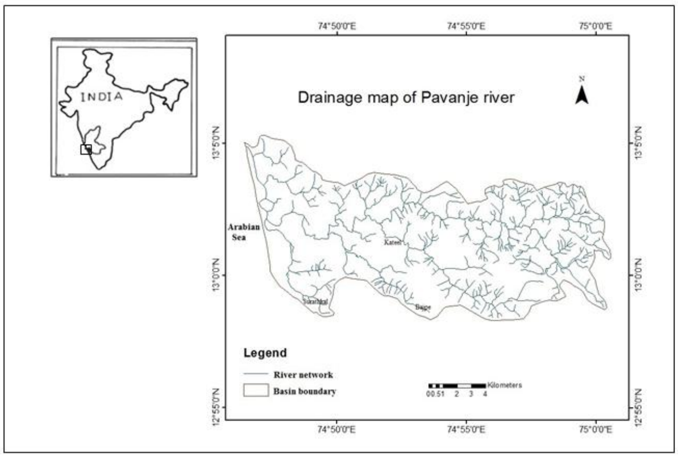

Figure. 1: Study Area

The daily data is being converted to weekly data and also to monthly data for carrying out the weekly and monthly analysis.

B. Artificial Neural Networks

ANNs mimic the functioning of a human brain by acquiring knowledge through a learning process that involves finding an optimal set of weights for the connections and threshold values for the nodes. An ANN can be viewed as a black box model in which a specific input to each node in the input layer is presented. Information passes from the input to the output side.

In an ANN architecture, the neurons are arranged in groups called layers. Each neuron in a layer operates in logical parallelism. Information is transmitted from one layer to another in serial operations. A network can have one or several layers. The basic structure of a network usually consists of three layers— the input layer, where the data are introduced to the network; the hidden layer(s), where data are processed; and the output layer, where the results for given inputs are produced [7]. The nodes in one layer are connected to those in the next, but not to those in the same layer. Thus, the output of a node in a layer is only dependent on the inputs it receives from previous layers and the corresponding weights. The connections between the input layer and the middle or hidden layer contain weights, which are usually determined through training the system. The middle layer neurons take the weighted inputs and sum them. To make a single value output from each neuron, the sum is used in an equation called a transfer function to create an output value. The basic structure of ANN is shown in Fig 2. The Levenberg-Marquardt (LM) algorithm, a standard second-order nonlinear least-squares technique, based on the backpropagation process to increase the speed and efficiency of the training was used for training the ANN models [6].

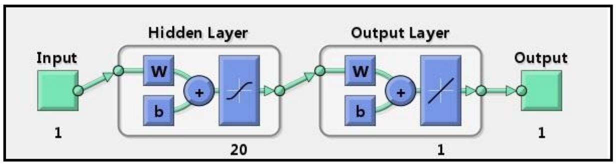

Figure 2: The basic structure of an ANN

The ASCE Task Committee on the Application of ANNs in Hydrology [8] reported that an attractive feature of ANNs is their ability to extract the relationship between the inputs and outputs of a process, without the physics being explicitly provided to them. They are able to provide a mapping from one multivariate space to another, given a set of data representing that mapping. Even if the data are noisy and contaminated with errors, ANNs have been known to identify the underlying rule. These properties suggest that ANNs may be well suited to the problems of estimation and prediction in hydrology [9].

C. Methodology

The Pavanje catchment is being delineated from toposheets (48K/16 , 48L/13, 48O/14 & 48P/1), using MapInfo Professional software. The area of the catchment is obtained as 202.9 km^2. The average rainfall and max/min temperature over the catchment is calculated using Arithmetic Average Method. The daily data is converted to weekly data for weekly analysis and also to monthly data for monthly analysis.

Most of the time in a year, the streamflow recorded is zero, indicating that rainfall is the only contributor to runoff and no considerable baseflow. Hence Pavanje catchment can be considered as intermittent catchment.

For the ANN model different input combinations including daily, weekly and monthly data, as shown in Table 1 were tried. The input data sets include present day's and antecedent days' data for analysis. Antecedent streamflow data is also used to prepare some models. The output in all the cases is discharge. 15 model combinations are prepared using different input combinations.

The models use 70% of data for training using Levenberg - Marquardt algorithm, 15% for testing and 15% for validation. The data sets with different input combinations were trained using different number of neurons ranging from 1 to 20 and the best input combination was selected based on the Regression (R) values, which measure the correlation between outputs and targets, and R=1 indicates a close relationship. Mean Square Error (MSE) is another criterion which can also be used for selecting the best model. The best model is the one with the least MSE.

TABLE 1: DESCRIPTION OF THE ARTIFICIAL NEURAL NETWORK MODELS

Model	Period	Input layers	Output layer
ANN1	Daily	P_t	Q_t
ANN2	Daily	P_t, $T_{t\,max}$	Q_t
ANN3	Daily	P_t, $T_{t\,min}$	Q_t
ANN4	Daily	P_t, $T_{t\,avg}$	Q_t
ANN5	Daily	P_t, P_{t-1}, $T_{t\,max}$	Q_t
ANN6	Daily	P_t, P_{t-1}, $T_{t\,min}$	Q_t
ANN7	Daily	P_t, P_{t-1},$T_{t\,avg}$	Q_t
ANN8	Daily	P_t, P_{t-1}, P_{t-2}, $T_{t\,max}$	Q_t
ANN9	Daily	Pt, P_{t-1}, $T_{t\,max}$, Q_{t-1}	Q_t
ANN10	Daily	P_{t-1}, $T_{t-1\,max}$, Q_{t-1}	Q_t
ANN11	Daily	P_{t-1}, P_{t-2}, $T_{t-1\,max}$, Q_{t-1}	Q_t
ANN12	Weekly	P_w	Q_w
ANN13	Weekly	P_w, $T_{w\,max}$	Q_w
ANN14	Monthly	P_m	Q_m
ANN15	Monthly	P_m, $T_{m\,max}$	Q_m

Note: P_t=t day rainfall (mm); P_{t-1}=$t-1$ day rainfall (mm); P_{t-2}=$t-2$ day rainfall (mm); Q_t=t day runoff (cumecs); $T_{t\,max}$=t day maximum temperature (°C); $T_{t\,min}$=t day minimum temperature (°C); $T_{t\,avg}$=t day average temperature (°C); $T_{t-1\,max}$=$t-1$ day maximum temperature (°C); Q_{t-1} =$t-1$ day runoff (cumecs); $T_{w\,max}$= weekly maximum temperature (°C); Pw=Weekly rainfall (mm); Q_w=Weekly runoff (cumecs); P_m=Monthly rainfall (mm); $T_{m\,max}$= monthly maximum temperature (°C); Q_m= Monthly runoff (cumecs).

Four different measures of performance were used to gauge the accuracy of the models. The coefficient of correlation (r), the Nash–Sutcliffe coefficient of efficiency (NSE) and the Index of agreement (d) were used to describe the proportion of the variance in the data that can be explained by the models.

$$r = \frac{\Sigma\{(Q_o - \bar{Q}_o)(Q_m - \bar{Q}_m)\}}{\{\Sigma((Q_o - \bar{Q}_o)^2 (Q_m - \bar{Q}_m)^2\}^{\frac{1}{2}}} \tag{1}$$

$$NSE = 1 - \frac{\Sigma (Q_m - Q_o)^2}{\Sigma (Q_o - \bar{Q}_o)^2} \tag{2}$$

$$d = 1 - \frac{\Sigma (Q_m - Q_o)^2}{\Sigma\,[abs(Q_m - \bar{Q}_o) + abs(Q_o - \bar{Q}_o)]^2} \tag{3}$$

Where Q_m and Q_o are modeled and observed values respectively.

In addition, the root mean square error (RMSE):

$$RMSE = \sqrt{\frac{\Sigma (Q_o - Q_m)^2}{N}} \tag{4}$$

Was adopted as absolute error measure.

III. Results And Discussios

Different input combinations of daily, weekly and monthly data (Table 1) have been trained using Levenberg - Marquardt algorithm and the number of neurons for the hidden layer is varied from 1 to 20. After training with different neurons (1-20) in the hidden layer, the best model is selected based on the R and MSE values. The regression values (R) and Mean Square Error (MSE) values of the best model, for training, validation and testing is shown in Table 2.

TABLE 2: R AND MSE VALUES OF THE BEST MODEL FOR TRAINING, VALIDATION AND TESTING

		Training		Validation		Testing	
Model	Layer	R	MSE	R	MSE	R	MSE
ANN1	1-8-1	0.6196	380	0.5316	400	0.5880	392
ANN2	2-11-1	0.7414	272	0.6850	250	0.6924	246
ANN3	2-17-1	0.6361	462	0.6493	372	0.6550	392
ANN4	2-3-1	0.6889	297	0.6824	301	0.7025	286
ANN5	3-18-1	0.7321	298	0.7554	274	0.7270	300
ANN6	3-17-1	0.6630	281	0.6725	260	0.6317	348
ANN7	3-17-1	0.7098	324	0.6799	340	0.6088	345
ANN8	4-12-1	0.7441	287	0.7338	294	0.7233	299
ANN9	4-16-1	0.9777	28	0.9795	25	0.9818	20
ANN10	3-12-1	0.9761	28	**0.9817**	**22**	**0.9850**	**18**
ANN11	4-12-1	**0.9798**	**22**	0.9716	24	0.9801	22
ANN12	1-18-1	0.6590	164	0.6134	178	0.6998	154
ANN13	2-19-1	0.8733	98	0.7134	103	0.8237	100
ANN14	1-18-1	0.8258	82	0.6828	103	0.7262	94
ANN15	2-13-1	0.9475	25	0.9744	24	0.8636	48

 ISSN: 2321-8169
233 – 237

TABLE 3: PERFORMANCE INDICES FOR ANN

	Training				Validation				Testing			
Model	r	d	NSE	RMSE	r	d	NSE	RMSE	r	d	NSE	RMSE
1	0.612	0.708	0.370	22.676	0.574	0.708	0.304	20.875	0.486	0.619	0.153	21.716
2	0.750	0.841	0.563	18.895	0.706	0.811	0.436	18.801	0.588	0.737	0.224	20.780
3	0.644	0.754	0.414	21.867	0.717	0.829	0.494	17.798	0.441	0.589	0.135	21.948
4	0.703	0.766	0.472	20.752	0.667	0.812	0.390	19.541	0.519	0.646	0.229	20.718
5	0.757	0.839	0.570	18.731	0.709	0.811	0.442	18.704	0.610	0.747	0.301	19.721
6	0.618	0.749	0.376	22.566	0.735	0.842	0.523	17.274	0.418	0.577	0.088	22.531
7	0.682	0.776	0.462	20.962	0.653	0.794	0.380	19.702	0.523	0.660	0.216	20.886
8	0.751	0.837	0.563	18.898	0.752	0.845	0.522	17.295	0.620	0.751	0.318	19.489
9	0.936	0.967	0.875	10.097	0.945	0.972	0.890	8.306	0.963	0.981	0.928	6.350
10	0.933	0.965	0.870	10.296	0.940	0.969	0.881	8.625	0.966	0.982	0.932	6.158
11	**0.978**	**0.989**	**0.956**	**5.964**	**0.973**	**0.986**	**0.946**	**5.832**	**0.984**	**0.991**	**0.967**	**4.290**
12	0.719	0.814	0.516	18.875	0.707	0.833	0.480	17.254	0.468	0.660	0.120	21.726
13	0.871	0.923	0.755	13.414	0.810	0.897	0.616	14.816	0.722	0.845	0.450	17.177
14	0.807	0.886	0.647	14.537	0.852	0.911	0.708	12.366	0.847	0.887	0.529	13.433
15	0.956	0.973	0.908	7.432	0.912	0.948	0.788	10.532	0.961	0.961	0.824	8.212

During training, model ANN11 (4-12-1) gives the best result, while model ANN10(3-12-1) gives the best result during validation and testing. The hydrographs generated by models 10 and 11 are almost the same. It is evident that the models are unable to simulate the peaks completely, but the low and medium flows are efficiently modelled. Weekly models (Models 12 and 13 are weak compared to daily and monthly models. Among the monthly models, model 15 is able to capture peak flows in most of the time period. The hydrographs generated by the best daily, weekly and monthly models are shown in Fig 3, Fig 4 and Fig 5 respectively.

The performance indices r, NSE, d and RMSE are calculated for different models of ANN during training, validation and testing, using equations (i), (ii), (iii) and (iv). The performance indices for the models are given in Table 3.

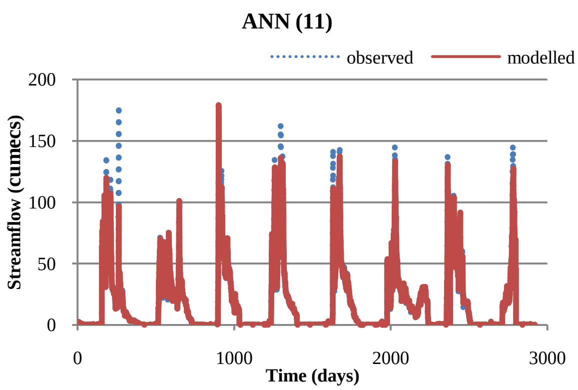

Figure 3: Hydrograph for the daily model

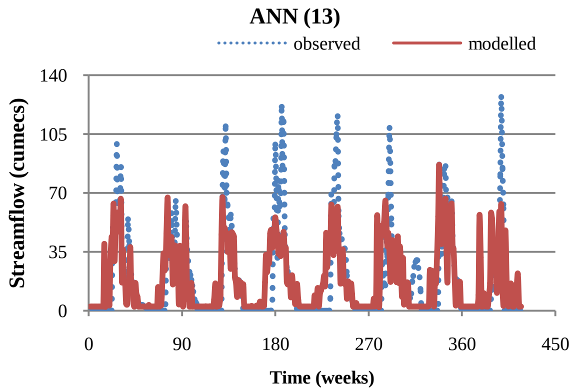

Figure 4: Hydrograph for the weekly model

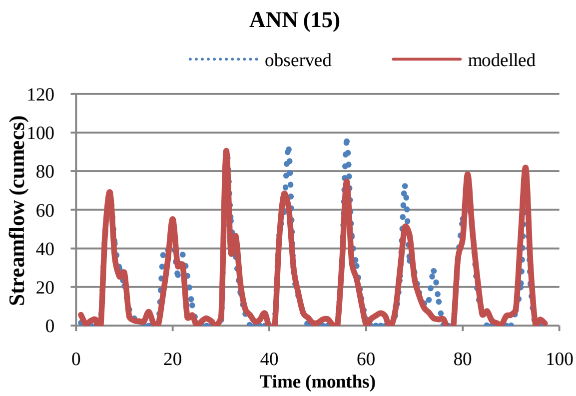

Fig 5: Hydrograph for the monthly model

ISSN: 2321-8169
233 – 237

CONCLUSIONS

In this study using daily, weekly and monthly ANN models, rainfall-runoff process of Pavanje catchment has been simulated. Four statistical parameters Coefficient of Correlation (r), Index of Agreement (d), Nash and Sutcliffe coefficient of efficiency (NSE) and Root Mean Square Error (RMSE) were used to compare the model results. The best model is the one with daily past one day rainfall, past 2 day rainfall, past one day maximum temperature and past one day runoff as inputs. The model performed well during training, validation as well as testing periods. The statistical parameters indicate that the best ANN model performed better during testing compared to training and validation time periods. The statistical parameters for training, validation and testing are r=0.978, d= 0.989, NSE=0.956, RMSE=5.964; r=0.973, d=0.986, NSE=0.946, RMSE=5.832 and r=0.984, d=0.991, NSE=0.967, RMSE=4.29 respectively. The results show that the daily ANN model has better performance in simulation of Rainfall-Runoff process than weekly and monthly models in Pavanje catchment. The model can be used for any Indian catchment with similar climatic conditions.

ACKNOWLEDGMENT

The authors acknowledge the support given by National Institute of Technology Karnataka and Karnataka Water Resources Department. Nimiya Baby expresses gratitude for the guidance given by Dr. K. Varija, Associate Professor, National Institute of Technology Karnataka, Surathkal.

REFERENCES

[1] K. Chakraborty, K. Mehrotra, C. K. Mohan and S. Ranka, “Neural Networks and Their Applications,” Review of Scientific Instruments, Vol. 65, 1992, pp. 1803-1832.
[2] M.P. Rajurkar, U.C. Kothyari, U.C. Chaube, “Modeling of the daily rainfall-runoof relationship with artificial neural network,” Journal of Hydrology 285 (2004) 96–113.
[3] Jitendra Sinha, R. K. Sahu, Avinash Agarwal, A. K. Pali and B. L. Sinha, “Rainfall -Runoff Modelling using Multi Layer Perceptron Technique - A Case Study of the Upper Kharun Catchment in Chhattisgarh”, Journal of Agricultural Engineering Vol. 50 (2): April - June, 2013.
[4] Archana Sarkar, Rakesh Kumar, “Artificial Neural Networks for Event Based Rainfall-Runoff Modeling,” Journal of Water Resource and Protection, 2012, 4, 891-897.
[5] Fernando Machado , Miriam Mine , Eloy Kaviski & Heinz Fill (2011) “Monthly rainfall–runoff modelling using artificial neural networks”, Hydrological Sciences Journal, 56:3, 349-361
[6] N. S. Raghuwanshi, R. Singh and L. S. Reddy, “Runoff and Sediment Yield Modeling Using Artificial Neural Networks: Upper Siwane River, India”, Journal of Hydrologic Engineering, Vol. 11, No. 1, January 1, 2006, ASCE.
[7] Sezin Tokar and Momcilo Markus “Precipitation-Runoff Modeling Using Artificial Neural Networks And Conceptual Models”, Journal of Hydrologic Engineering. 2000.5:156-161, ASCE
[8] ASCE Task Committee on Application of Artificial Neural Networks in Hydrology, “Artificial neural networks in hydrology. I: Pre-liminary concepts”, Journal of Hydrologic Engineering., ASCE 5(2), 115–123, 2000.
[9] ASCE Task Committee on Application of Artificial Neural Networks in Hydrology , “Artificial neural networks in hydrology II”, Journal of Hydrologic Engineering. ASCE **5**(2), 124–132, 2000.