Jurnal Teknologi

Full paper

Minimum Input Variances for Modelling Rainfall-runoff Using ANN

Zulkarnain Hassan[a,b*], Supiah Shamsudin [c], Sobri Harun[a]

[a]*Faculty of Civil Engineering, Universiti Teknologi Malaysia, 81310 UTM Johor Bahru, Johor, Malaysia*
[b]*School of Environmental Engineering, Universiti Malaysia Perlis, 02600 Arau, Perlis, Malaysia*
[c]*Razak School of Eng. and Advan. Tech., Universiti Teknologi Malaysia, 54100 UTM-Kuala Lumpur, Kuala Lumpur, Malaysia*

*Corresponding author: zulkarnain23@live.utm.my

Article history

Received :10 March 2014
Received in revised form :
28 April 2014
Accepted :15 May 2014

Graphical abstract

Abstract

This paper presents the study of possible input variances for modeling the long-term runoff series using artificial neural network (ANN). ANN has the ability to derive the relationship between the inputs and outputs of a process without the physics being provided to it, and it is believed to be more flexible to be used compared to the conceptual models [1]. Data series from the Kurau River sub-catchment was applied to build the ANN networks and the model was calibrated using the input of rainfall, antecedent rainfall, temperature, antecedent temperature and antecedent runoff. In addition, the results were compared with the conceptual model, named IHACRES. The study reveal that ANN and IHACRES can simulate well for mean runoff but ANN gives a remarkable performance compared to IHACRES, if the model customizes with a good configuration.

Keywords: Artificial neural network; runoff; IHACRES; rainfall-runoff

1.0 INTRODUCTION

Rainfall-runoff models are the standard tools routinely designed for hydrological investigations and they are used for many purposes such as for detecting catchment response towards climatic events, calculations of design floods, management of water resources, estimation of the impact of land-use change, forecast flood and of course for stream flow prediction [2]. Simulating the real-world relationship using the rainfall-runoff models are a difficult task, since various interacting processes that involve in the transformation of rainfall into runoff are complex. Therefore, rainfall-runoff models have been classified into three types [3] in order to overcome the difficulty on simulation, which are the physically, conceptually and metric-based models.

Physically and conceptual-based models are based on physical equations that describe the real system of hydrological system of the catchment [4]. Both models are extreme data demand and composed of a large number of parameters [5]. Therefore, they are difficult to calibrate and facing over parameterization [4,6]. Metric-based models are based on extracting information that is implicitly contained in a hydrological data without directly taking into account the physical laws that underlie the rainfall-runoff processes [7]. The models are simple since no complex data are needed and easily understood, compared to the other type of models [8].

In this paper, artificial neural network (ANN), which uses the metric based-model, is applied. In recent years, ANN has been successfully used as a rainfall-runoff model [9-14]. Vos and Rientjes [7] in his paper stated that ANN has advantages over the physical and conceptual models, since it is able to simulate non-linearity in a system. It also effectively distinguishes relevant from irrelevant data characteristics. Moreover, ANN is non-parametric technique, which means that the model does not require the assumption or enforcement of constraints. Neither, it needs a priori solution structures [15].

This paper aims to demonstrate the ability of the ANN model to simulate the long range daily runoff series by only using the minimum input information such as rainfall, temperature and antecedent runoff. Hence, the conceptual model named Identification of unit Hydrographs and Component flows from Rainfall, Evaporation and Stream Flow Data (IHACRES) is applied to compare with the ANN model. The performances, abilities and shortcomings of models are discussed.

2.0 MATERIALS AND METHODS

2.1 Study Area

The simulating work is carried out using rainfall, temperature, and runoff records from the Kurau River catchment, in the state of Perak. The study area and details of the related meteorological stations are shown in Figure 1 and Table 1. The statistical indices of each meterological station are shown in Table 2. At the downstream of the catchment, a dam is located. The dam becomes a main drainage for paddy field and also acts as a source for drinking water. The area of the sub-basin covers approximately 337 km^2.

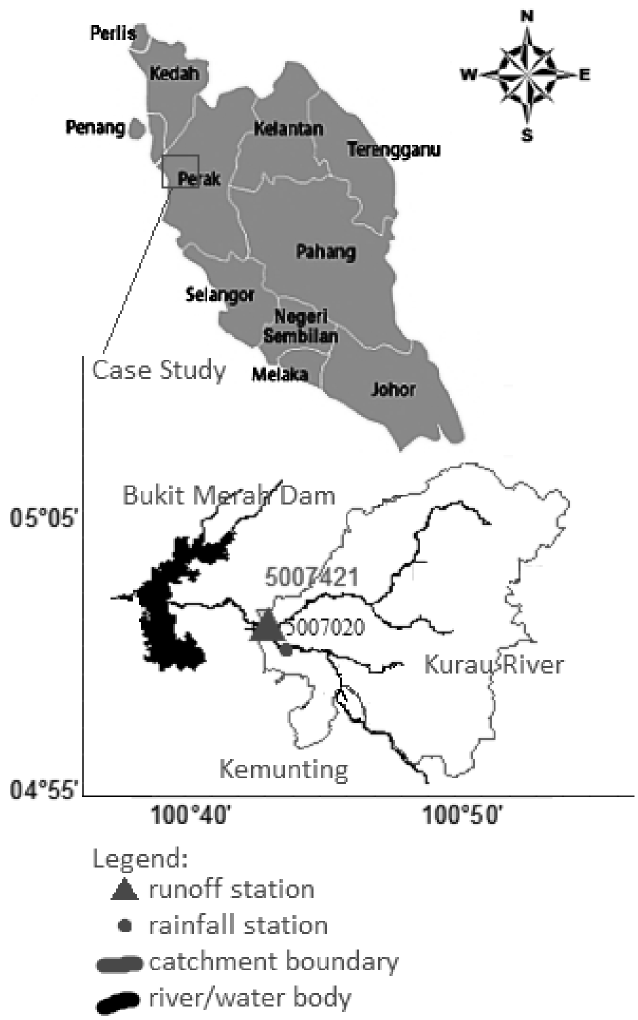

Figure 1 Location of study area

2.2 Artificial Neural Network

There are many types of ANN that have been developed, such as multilayer perceptron, radial basis, Kohonen, and Hopfield neural networks. Each type has its own strength and limitation. The study focused on the multilayer perceptron neural network (MLP) model. This model was selected because MLP shows the most promising performance compared to the other types of ANN's models. It is also widely used in the field of hydrology, particularly in the runoff analysis [16]. MLP network can be written as:

$$a = f\left(\sum_{i=1}^{n} w_i x_i + b\right) \quad (1)$$

Where a is the output of MLP, f is the transfer function, w_i is the weights, b is the bias and x_i is the input vector (i = 1, 2, …, n). In this study, two-layer feedforward network trained with backpropagation learning algorithm, as shown in Figure 2 is used.

Table 1 Detail of meteorological stations

No of Station	Type of Station	Name of Station	Location	
			Lat	Long
5007020	Rainfall	Ldg. Pondoland	05^{0} 00' 35''	100^{0} 43' 50''
48625	Temp	Ipoh, Perak	04^{0}34'01"	101^{0}06'00"
5007421	Runoff	Sg. Kurau di Pondok Tanjung	05^{0} 00' 46''	100^{0} 43' 55''

Table 2 Summary of input variables

	Mean	SD	Min	Max	Sample Variance
Training					
Rainfall (mm)	9.32	15.65	0	167.3	245.02
Temp (ºC)	26.72	1.01	22.83	31.56	1.01
Runoff (cumecs)	17.15	13.79	0.19	116.01	190.17
Validation					
Rainfall (mm)	9.53	15.46	0	118	238.95
Temp (ºC)	27.53	1.07	24.5	30.7	1.15
Runoff (cumecs)	21.56	20.47	1.22	109.19	419.07

The transfer functions used in the hidden layer are tan-sigmoid (*TANSIG*) and linear transfer function (*PURELIN*) at the output layer. The details of this MLP architecture were discussed in detail by Hassan [17].

The input data were divided into two categories, namely training (calibration) and validation periods, as shown in Table 3. In order to gain the most optimum and efficient MLP networks for daily runoff forecasting, the parameters were adjusted during the training process. The parameters were: 1) input data, 2) algorithm, 3) number of hidden neurons in hidden layer, and 4) learning rate value. Through the preliminary study, the input data for MLP model were arranged into 3 cases. The arrangement is shown in Table 4, in which $\{P(t)\}$ is rainfall of the current day, $\{T(t)\}$ is mean temperature of the current day, $\{P(t-1), P(t-2),…, P(t-n)\}$ is antecedent rainfall, $\{T(t-1), T(t-2),…, T(t-n)\}$ is antecedent temperature and $\{Q(t-1), Q(t-2),…, Q(t-n)\}$ is antecedent runoff temperature. The optimum configuration of each parameter is illustrated in Table 5.

Table 3 Period of training and validation

Condition Process	Period of Time (days)	Time Step (days)
Training	1st Feb 1968 until 31st Dec 1977; 1st Jan 1978 until 31st Dec 1979; 1st Jan 1980 until 31st Dec 1982	5448
Validation	1st Feb 1997 until 31st Dec 2000	1430

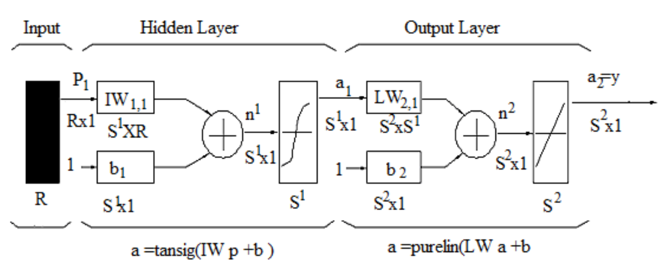

Figure 2 MLP network architecture [17]

2.3 Application of IHACRES

IHACRES is conducted using the conceptual-based model [18]. It requires between five (5) and seven (7) parameters to be calibrated and it performed well on a board with variety of catchment sizes and areas. The IHACRES model consists of two modules (Figure 3), which are: i) non-linear loss module, where rainfall (r_k) due to time step (k) is transforming into effective rainfall (u_k) and; ii) linear unit hydrograph module where u_k transforms to runoff (x_k). These modules can be written as:

$$u_k = r_k \times s_k \tag{2}$$

$$x_k = x_k^{(q)} + x_k^{(s)} \tag{3}$$

Where,

$$x_k^{(q)} = a^{(q)} x_{k-1}^{(q)} + b^{(q)} u_k \tag{4}$$

$$x_k^{(s)} = a^{(s)} x_{k-1}^{(s)} + b^{(s)} u_k \tag{5}$$

In those equations, s_k is the catchment wetness index, $x_k^{(q)}$ and $x_k^{(s)}$ are the quick and slow runoff components, $a^{(q)}$ and $a^{(s)}$ are the recession rates for quick and slow storage, and $b^{(q)}$ and $b^{(s)}$ are the fraction of effective rainfall. This transformation is similar to the concept of unit hydrograph theory, in which a configuration of linear storage acting in series and/or parallel in the catchment. In this study, the data of rainfall, temperature, and runoff series become inputs to the IHACRES model, with the similar time period as applied in the MLP model (Table 3).

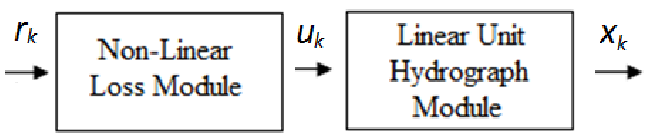

Figure 3 Concept of the IHACRES model [19]

Table 4 Input of MLP model

Models	Input
MLP1	*P(t), P(t-1), P(t-2), P(t-3), P(t-4), P(t-5), P(t-6), P(t-7), P(t-8), P(t-9)*
MLP2	*P(t), P(t-1), P(t-2), P(t-3), P(t-4), P(t-5), P(t-6), P(t-7), P(t-8), P(t-9),* *Q(t-1), Q(t-2), Q(t-3), Q(t-4), Q(t-5), Q(t-6), Q(t-7), Q(t-8), Q(t-9)*
MLP3	*P(t), P(t-1), P(t-2), P(t-3), P(t-4), P(t-5), P(t-6), P(t-7), P(t-8), P(t-9),* *T(t), T(t-1), T(t-2), T(t-3), T(t-4), T(t-5), T(t-6), T(t-7), T(t-8), T(t-9)*
MLP4	*P(t), P(t-1), P(t-2), P(t-3), P(t-4), P(t-5), P(t-6), P(t-7), P(t-8), P(t-9),* *T(t), T(t-1), T(t-2), T(t-3), T(t-4), T(t-5), T(t-6), T(t-7), T(t-8), T(t-9),* *Q(t-1), Q(t-2), Q(t-3), Q(t-4), Q(t-5), Q(t-6), Q(t-7), Q(t-8), Q(t-9)*

Table 5 Optimum configuration of the MLP model

Parameters	Values
Training Algorithm	*TRAINSCG*
No. of neurons	125
Different learning training	0.8

2.4 Model Evaluation

The evaluation of MLP and IHACRES models during training and validation was checked using the coefficient of correlation (R) and the root mean square error ($RMSE$), which are defined as:

$$R = \frac{\sum(obs-\overline{obs})(pred-\overline{pred})}{\sqrt{\sum(obs-\overline{obs})^2 \sum(pred-\overline{pred})^2}} \tag{6}$$

$$RMSE = \sqrt{\frac{\sum(obs-pred)^2}{n}} \tag{7}$$

In which, *obs* = observed streamflow value; *pred* = predicted streamflow value; $\overline{obs}$ = mean streamflow observed value, and; $\overline{pred}$ = predicted mean streamflow. The closer R value to 1 and $RMSE$ value to 0, the predictions are better.

3.0 RESULTS

The correlation analysis of time series was applied in order to evaluate the effect of antecedent rainfall, temperature and flow. The correlation results are shown in Figure 4. The auto- and partial autocorrelation statistics and the corresponding 95% confidence bands from lag 1 to 15 were simulated for rainfall (Figure 4a), temperature (Figure 4b) and runoff (Figure 4c) data series. The figures show that the partial autocorrelation function gives a significant correlation up to lag seven for rainfall, lag seven for temperature, and lag one for flow series data before dropping within the confidence limits. The decreasing trend of partial autocorrelation indicates the dominance of the autoregressive process, which is relative to the moving-average process. Hence, seven antecedent rainfalls and temperatures, and one antecedent runoff must be selected as an input to the MLP model. In order to increase the reliability in input to the MLP model, 9 antecedent rainfalls, temperatures, and runoffs were also selected.

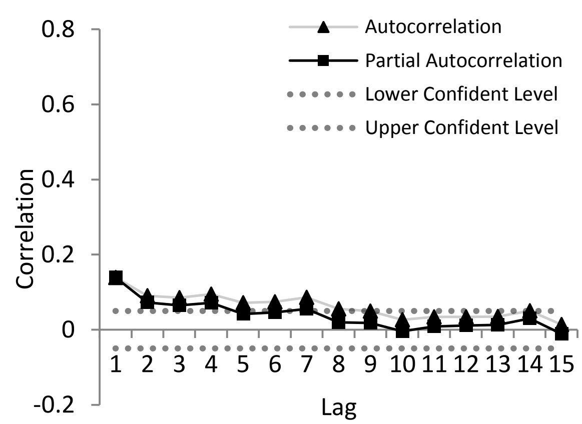

Figure 4a Auto- and partial autocorrelation functions of rainfall series (95% confidence band)

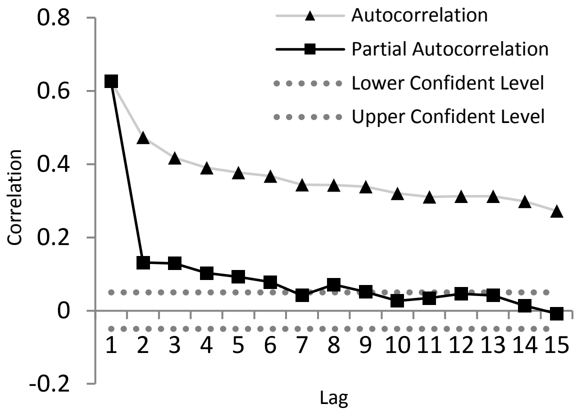

Figure 4b Auto- and partial autocorrelation functions of temperature series (95% confidence band)

The performance of the MLP and IHACRES models as compared with IHACRES by using the values of *R* and *RMSE* during the training and validation are shown in Table 6. During training, MLP4 and MLP2 show a better performance with a higher *R* value and the lowest *RMSE* value as compared to the other model. During validation, both models give a satisfied result with $R>0.5$ and $RMSE<20$cumecs. The visual representations of the inspection and comparison between the simulated and observed runoff during validation are shown in Figure 5. The IHACRES model can captures mean runoff but it is unable to capture most of the peak and low runoffs. Most of the MLP models show a good agreement between observed and simulated runoff. The summary of the performance is illustrated in Table 7.

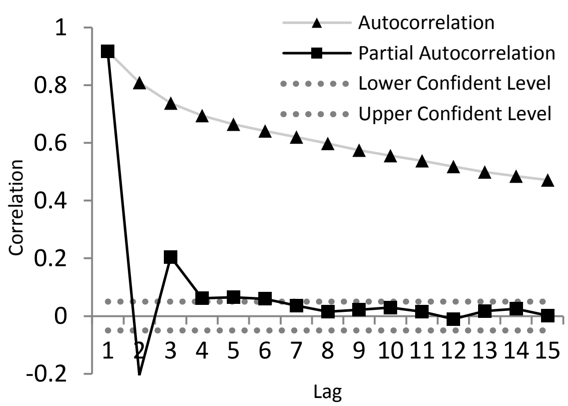

Figure 4c Auto- and partial autocorrelation functions of runoff series (95% confidence band)

Table 6 Performance during training and validation periods

Models	Training		Validation	
	R	*RMSE(cumecs)*	*R*	*RMSE(cumecs)*
MLP1	0.89	6.44	0.60	16.63
MLP2	0.99	1.89	0.85	10.93
MLP3	0.90	6.15	0.65	16.45
MLP4	0.99*	1.83*	0.85*	10.89*
IHACRES	0.71	9.70	0.60	16.53

*. The best performance

4.0 DISCUSSION

As mentioned in the introduction section, the study aims to demonstrate the ability of the ANN model, namely the MLP model to simulate the long-range daily runoff series by using the minimum input information from the Kurau River sub-catchment. The IHACRES model, which is a conceptual model, was used and become the batch mark to MLP's results. Since ANN is a metric model, identification of the input data selection is an important step. The selected data represents the characteristic of a watershed and meteorological pattern. This study used rainfall and antecedent rainfall as the main variables of the MLP model. In order to test the ability of the model, the study introduced antecedent runoff, current temperature and antecedent temperature as a combination with the main variables.

In order to select a number of antecedents of each input data (rainfall, runoff, and temperature), the correlation analysis was applied. The results (Figure 4) reveal nine antecedents of rainfall and temperature and one antecedent runoff were sufficient to be run in this study. The use additional several numbers of antecedent variables are possible to enhance performance of the model [20]. However, it is not recommended to use a many antecedent data because the study by Kuok and Bessaih [21] founds that it will reduce the performance of the ANN model. Therefore, this study only utilized the nine antecedent variables.

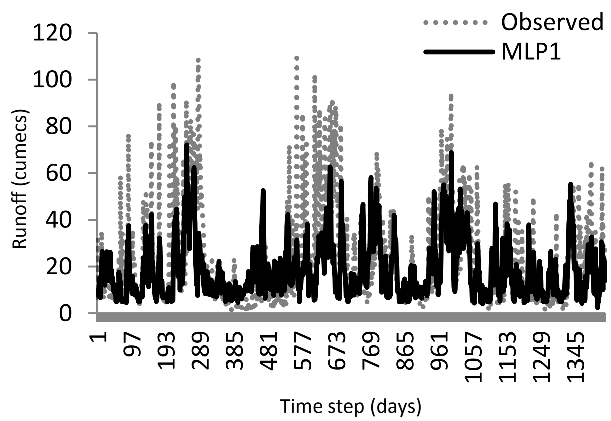

Figure 5a Performance of MLP1 during validation period

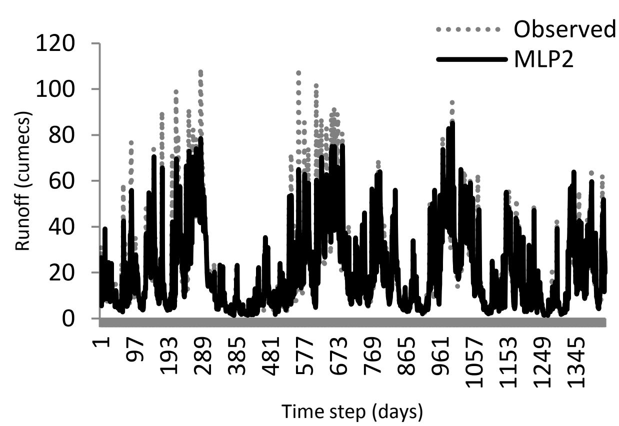

Figure 5b Performance of MLP2 during validation period

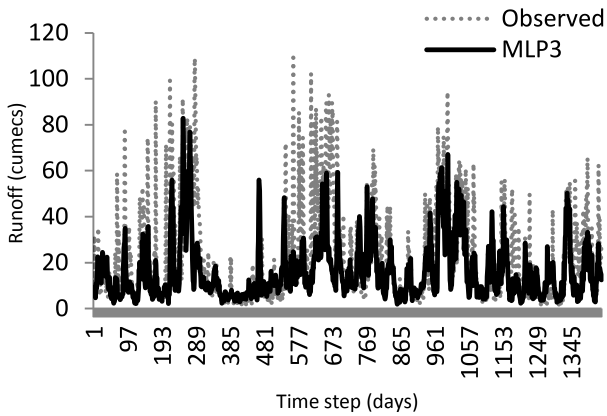

Figure 5c Performance of MLP3 during validation period

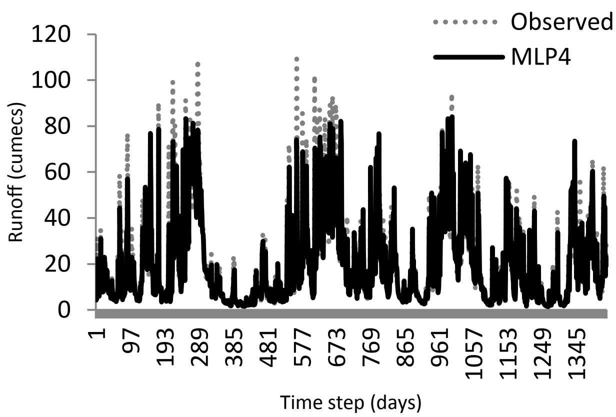

Figure 5d Performance of MLP4 during validation period

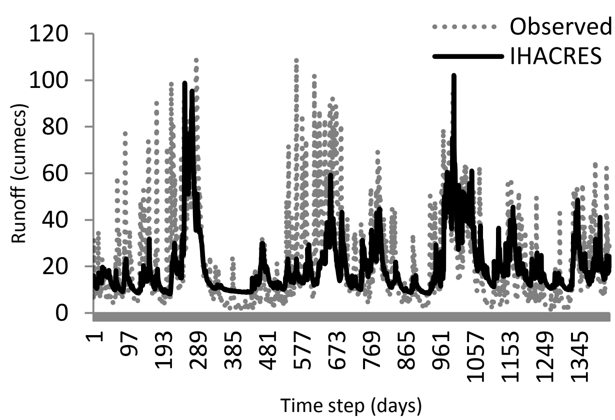

Figure 5e Performance of IHACRES during validation period

Table 7a Statistical indices of observed and simulated runoff during training period

	Mean*	SD	Min*	Max*	σ^2
Obs	17.15	13.79	0.19	116.01	190.17
MLP1	17.18 (0.03)**	11.04 (2.75)	2.23 (2.04)	101.10 (14.91)	121.84 (68.33)
MLP2	17.19 (0.04)	13.41 (0.38)	0.35 (0.16)	104.45 (11.56)	179.93 (10.24)
MLP3	17.18 (0.03)	11.26 (2.54)	1.61 (1.42)	98.86 (17.15)	126.68 (63.49)
MLP4	17.17 (0.02)	13.50 (0.29)	0.21 (0.02)	99.29 (16.72)	182.25 (7.92)
IHACRE S	16.74 (0.41)	9.95 (3.84)	0.00 (0.19)	102.71 (13.30)	98.99 (91.18)

*. Unit: cumecs
**. The different between observed and simulated

A record of 6878 days of daily rainfall, temperature and runoff series were selected in order to evaluate the performance of the MLP and IHACRES models. In order to conduct the evaluation, the data were divided into two periods, name as training and validation. During the training period, the result (Table 6) shows an application of rainfall and nine antecedent rainfalls (MLP1) was not sufficient to capture observed runoff, with the lowest R value and a high RMSE value were recorded. This situation gives effect to runoff simulation by MLP1 during the validation period, in which it gives a moderate performance and it is able to capture observed mean and low runoff (Figure 5a). Adjustment to MLP1, named as the MLP2 model, with the addition of nine antecedent runoffs resulted to a robust performance on the runoff simulation. This model is able to predict mean and based runoff very well during training and validation periods, as shown in Figure 5c. However, MLP2 is unable to detect some peak flows during the validation.

The combination of temperature and antecedent temperature is found not adequate to be used in the MLP's development. These combinations of variables were applied in MLP3 and MLP4. It is revealed that these models were slightly well performed compared to the current MLP models (named as MLP1 and MLP2).

As a conceptual model, the development of the IHACRES model is not customized like the MLP model. Each parameter of IHACRES was calibrated until it achieved the condition given. In this study, IHACRES is not seemed to perform well during the training and validation periods. The model is able to capture mean runoff but it is unable to predict well for base and peak runoff. This is as shown in Figure 5e and Table 7.

Table 7b Statistical indices of observed and simulated runoff during validation period

	Mean*	SD	Min*	Max*	σ^2
Obs	21.56	20.47	1.22	109.19	419.07
MLP1	18.37 (3.18)**	12.13 (8.34)	2.42 (1.20)	72.04 (37.15)	147.16 (271.91)
MLP2	20.49 (1.06)	17.23 (3.24)	1.29 (0.07)	85.22 (23.97)	297.02 (122.05)
MLP3	16.16 (5.39)	12.60 (7.87)	1.86 (0.64)	82.88 (26.31)	158.85 (260.22)
MLP4	19.84 (1.71)	18.08 (2.39)	1.46 (0.24)	84.06 (25.13)	327.05 (92.02)
IHACRES	19.63 (1.93)	12.93 (7.54)	8.12 (6.90)	102.07 (7.12)	167.16 (251.91)

*. Unit: cumecs
**. The different between observed and simulated

5.0 CONCLUSION

The performance of ANN using MLP's architecture is evaluated for daily runoff simulation with a long data series. As a case study, the data of Kurau River sub-catchment in Malaysia are used. This study found that the selection of input variables for MLP network can be determined by using the correlation analysis. The results show that the MLP models yield a better performance than the IHACRES model in modeling the rainfall-runoff relationship. Besides, the MLP model has the ability to simulate runoff accurately using rainfall, antecedent rainfall and antecedent runoff data as input variables. Selection of temperature and antecedent temperature data as input variables does not give a large significance toward MLP's performance.

This study used the data from one catchment and therefore, for generalization of results further studies must be performed.

Acknowledgment

The authors would like to thank the Department of Irrigation and Drainage Malaysia (DID) and Malaysian Meteorological Department (MMD) for providing the data and technical support. The project was financially supported by MOSTI, under vote 4s039 and UTM. The first author is grateful to Universiti Malaysia Perlis and the Ministry of Education Malaysia for a study leave in Universiti Teknologi Malaysia.

References

[1] Nor, N., S. Harun, and A. Kassim. 2007. Radial Basis Function Modeling of Hourly Streamflow Hydrograph. *Journal of Hydrologic Engineering*. 12(1): 113–123.

[2] Bloschl, G. 2005. *Rainfall–Runoff Modeling of Ungauged Catchments*. London: John Wiley and Sons.

[3] Zhang, G. P., and H. H. G. Savenije. 2005. Rainfall-Runoff Modelling in a Catchment with a Complex Groundwater Fow System: Application of the Representative Elementary Watershed (REW) approach. *Hydrology and Earth System Sciences*. 9: 243–261.

[4] Guéro, P. 2006. Rainfall Analysis and Flood Hydrograph Determination in the Munster Blackwater Catchment. University College Cork, Ireland.

[5] Rezaeianzadeh, M., A. Stein, H. Tabari, H. Abghari, N. Jalalkamali, E. Z. Hosseinipour, and V. P. Singh. 2013. Assessment of a Conceptual Hydrological Model and Artificial Neural Networks for Daily Outflows Forecasting. *International Journal of Environmental Science and Technology*. 10(6): 1181–1192.

[6] Beven, K. 1989. Changing Ideas in Hydrology - The Case of Physically-Based Models. *Journal of Hydrology*. 105(1–2): 157–172.

[7] Vos, N. J., and T. H. M. Rientjes. 2005. Constraints of Artificial Neural Networks for Rainfall-Runoff Modelling: Trade-Offs in Hydrological State Representation and Model Evaluation. *Hydrology and Earth System Science Discussions*. 2: 365–415.

[8] ASCE Task Committee on Application of Artificial Neural Networks in Hydrology. 2000. Artificial Neural Networks in Hydrology. I: Preliminary Concepts. *Journal of Hydrologic Engineering*. 5(2): 115–123.

[9] Hsu, K. L., H. V. Gupta, and S. Sorooshian. 1995. Artificial Neural Network Modelling of the Rainfall-Runoff Process. *Water Resources Research*. 31(10): 2517–2530.

[10] Minns, A. W., and M. J. Hall. 1996. Artificial Neural Networks as Rainfall Runoff Models. *Hydrological Science Journal*. 41(3): 399–417.

[11] Dawson, C. W., and R. L. Wilby. 2001. Hydrological Modelling Using Artificial Neural Networks. *Progress in Physical Geography*. 25(1): 80–108.

[12] Tokar, A. S., and P. A. Johnson. 1999. Rainfall-Runoff Modeling Using Artificial Neural Networks. *Journal of Hydrological Engineering*. 4(3): 232–239.

[13] Rajurkar, M. P., U. C. Kothyari, and U. C. Chaube. 2002. Artificial Neural Networks for Daily Rainfall-Runoff Modelling. *Hydrological Sciences Journal*. 47(6): 865–877.

[14] Wilby, R. L., R. J. Abrahart, and C. W. Dawson. 2003. Detection of Conceptual Model Rainfall-Runoff Processes Inside an Artificial Neural Network. *Hydrological Sciences Journal*. 48(2): 163–181.

[15] French, M. N., W. F. Krajewski and R. R. Cuykendall. 1992. Rainfall Forecasting in Space and Time Using a Neural Network. *Journal of Hydrology*. 137(1–4): 1–31.

[16] Gautama, M. R., K. Watanabea, and H. Saegusa. 2000. Runoff Analysis in Humid Forest Catchment with Artificial Neural Network. *Journal of Hydrology*. 235(1–2): 117–136.

[17] Hassan, Z. 2012. *Climate Change Impact on Precipitation on Precipitation and Streamflow in a Humid Human Tropical Watershed*. Universiti Teknologi Malaysia, Malaysia.

[18] Ye, W., B. C. Bates, N. R. Viney, M. Sivapalan, and A. J. Jakeman. 1997. Performance of Conceptual Rainfall-Runoff Models in Low-Yielding Ephemeral Catchments. *Water Resources Research*. 33(1): 153–166.

[19] Littlewood, I. G., K. Down, J. R. Parker, and D. A. Post. 1997. *IHACRES: Catchment-Scale Rainfall-Streamflow Modelling (PC version) Version 1.0*. Wallingford : Institute of Hydrology, Centre for Ecology and Hydrology.

[20] Zadeh, M. R., S. Amin, D. Khalili, and V. P. Singh. 2010. Daily Outflow Prediction by Multi Layer Perceptron with Logistic Sigmoid and Tangent Sigmoid Activation Functions. *Water Resources Management*. 24 (11): 2673–2688.

[21] Kuok, K. K., and N. Bessaih. 2007. Artificial Neural Networks (ANNs) for Daily Rainfall Runoff Modelling, *Journal - The Institution of Engineers*. 68(3): 31–42.