Jurnal
Teknologi

Full Paper

FLOOD ESTIMATION AT UNGAUGED SITES USING GROUP METHOD OF DATA HANDLING IN PENINSULAR MALAYSIA

Basri Badyalina, Ani Shabri*

Department of Mathematics, Universiti Teknologi Malaysia

Article history
Received
10 February 2014
Received in revised form
5 April 2015
Accepted
1 August 2015

*Corresponding author
ani@utm.my

Graphical abstract

Abstract

Group Method of Data Handling (GMDH) have been successful in many fields such as economy, ecology, medical diagnostics, signal processing, and control systems but given a little attention in hydrology field especially for flood estimation at ungauged sites. Ungauged site basically mean the site of interest is no flood peak data available. This paper presented application of GMDH model at ungauged site to predict flood quantile for T=10 year and T=100 year. There five catchment characteristics implement in this study that are catchment area, elevation, longest drainage path, slope of the catchment and mean maximum annual rainfall. The total number of catchment used for this study is 70 catchments in Peninsular Malaysia. Four quantitative standard statistical indices such as mean absolute error (MAE), root mean square error (RMSE) and Nash-Sutcliffe coefficient of efficiency (CE) are employed. Based on these results, it was found that the GMDH model outperforms the prediction ability of the traditional LR model.

Keywords: Linear regression, group method of data handling, ungauged

Abstrak

Kaedah Kumpulan Pengendalian Data (GMDH) telah berjaya dalam pelbagai bidang seperti ekonomi, ekologi, diagnostik perubatan, pemprosesan isyarat, dan sistem kawalan tetapi diberi sedikit perhatian dalam bidang hidrologi terutamanya dalam anggaran banjir di tapak ungauged. Tapak Ungauged pada dasarnya bermakna tapak tersebut tidak wujud data puncak banjir. Kajian ini membentangkan kertas kerja berkenaan applikasi model GMDH di tapak ungauged untuk meramalkan kuartil banjir untuk T = 10 tahun dan T = 100 tahun. Terdapat lima ciri-ciri tadahan diperlukan untuk melaksanakan dalam kajian ini iaitu kawasan tadahan, ketinggian, jalan perparitan paling panjang, cerun tadahan dan purata maksimum hujan tahunan. Jumlah tadahan yang digunakan untuk kajian ini adalah 70 kawasan tadahan di Semenanjung Malaysia. Empat indeks statistik standard kuantitatif seperti min ralat mutlak (MAE), punca min ralat kuasa dua (RMSE) dan pekali Nash-Sutcliffe kecekapan (CE) bekerja. Berdasarkan keputusan ini, didapati bahawa model GMDH yang melebihi performa keupayaan ramalan model LR tradisional.

Kata kunci: Regresi linear, kaedah kumpulan pengendalian data, ungauged

1.0 INTRODUCTION

Flood are one the most frequent natural disaster occur in Peninsular Malaysia which happen almost every year. Flood causes a lot of damages to properties, infrastructures and even loss of people lives. The rising floodwaters cut off water, food, electricity supplies, forcing evacuees to seek shelter in relief centers. Flood surely cannot be prevented from occurring but human beings can prepare for it. This make a reliable estimation of flood quantiles is important for flood risk assessment project (e.g., dams, spillways, road, and culverts), the safe design of the river system (Besaw *et al*., 2010; Seckin, 2011). However, it often happen the historical data at site of interest not always available. Although at-site of interest may have some available data but the data is not enough to describe the catchment flow because of the changes in watershed characteristics such as urbanization (Pandey and Nguyen, 1999). Robson and Reed (1999) stated that flood estimation become a problem when the estimation is at ungauged site where no flood peak data available. Mamun *et al*. (2012) stated that river located in Malaysia is gauged only at a strategic location and other river is usually ungauged. This could be a problem for a developing country like Malaysia when the development projects are located at ungauged site. Typically some site characteristics for the ungauged sites are known. Thus, regionalization is carried out to make the estimation of flow statistics at ungauged sites using physiographic characteristics. Regionalization technique includes ng a probability distribution to series of flow and then linking the relationship to catchment characteristics (Dawson *et al*., 2006). The variables affecting the flood quantile estimation include catchment characteristics (size, slope, shape and storage characteristics of the catchment), storm characteristics (intensity and duration of rainfall events), geomorphologic characteristics (topology, land use patterns, vegetation and soil types that affect the infiltration) and climatic characteristics (temperature, humidity and wind characteristics) (Hosking and Wallis 1997; Jain and Kumar 2007). In relating flood quantile at site of interest to catchment characteristics a power form equations are mostly used (e.g., Thomas and Benson 1970; Fennessey and Vogel 1990; Mosley and Mckerchar 1993; Pandey and Nguyen, 1999; Seckin, 2011; Mamun, 2012). At ungauged sites linear regression (LR) model is always reliable to make estimates of flow statistics or flood quantiles (see e.g. Vogel and Kroll, 1990; Shu and Ouarda, 2008; Pandey and Nguyen, 1999). Mohamoud (2008) used step-wise linear regression to identify dominant landscape and climate descriptor from 29 catchments and then developed flow duration curves that managed to forecast flow in nearby ungauged catchments. Mamun *et al*. (2012) used linear regression of various return periods in Peninsular Malaysia.

Recently the group method of data handling (GMDH) algorithm has been successfully used to deal with uncertainty, linear or nonlinearity of systems in a wide range of disciplines such as economy, ecology, medical diagnostics, signal processing, and control systems (Oh and Pedrycz, 2002; Nariman-Zadeh *et al*., 2002; Kondo, *et al.,* 2005, 2006; Onwubolu, 2009). Although GMDH was a useful statistic tool used in many areas but in hydrology it only just a few studies involving application of GMDH and especially in ungauged problem there are none. The prediction accuracy was surprising successful for researchers who used the GMDH in modeling. The GMDH algorithm can be devoted to developing polynomial structure for modeling highly nonlinear systems with large number of inputs. The GMDH models are layered structures that exhibit a number of significant advantages as contrasted to other nonlinear modeling techniques. Tamura and Halfon (1980) used GMDH to model and identify water quality dynamics in Lake Ontario and the results showed that GMDH can be usefully employed to develop lake models with very low expenses of manpower and computer time. Huang and Shin (2002) applied the GMDH for short-term load forecast of a power system. In addition, Sforna (1995) introduced the GMDH in underline the link between the variables, temperature and electric load which examines the entire national electric network and also considers more limited areas such as regional and departmental networks. The aims of the present investigation are: 1) to explore the potential application of group method of data handling (GMDH) solutions to the problem of flood estimation in ungauged catchments; 2) to compare GMDH model estimation performance with conventional method linear regression.

2.0 METHODOLOGY

2.1 Group Method of Data Handling

Group Method of Data Handling model was introduced by Ivakhnenko on 1970 to solve complex non-linear multidimensional that has short data series (Ivakhnenko, 1970). GMDH model is based on principal of heuristic of self-organization to identify mathematical model between input and output signal (Ivakhnenko, 1970; Onwubulu *et al.,* 2007; Najafzadeh & Barani, 2011). GMDH also can solve the modeling problem that has multi-input to single output data (Sharma & Onwubolu, 2009). The GMDH algorithm that describes the relationship between input and output signal can be represented by Volterra series (Ivakhnenko 1970; Farlow 1981) in form of:

$$y = v_0 + \sum_{i=1}^{n} c_i x_i + \sum_{i=1}^{n}\sum_{j=1}^{n} v_{ij} x_i x_j + \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n} v_{ijk} x_i x_j x_k + \ldots \quad (1)$$

which also be known as Kolmogorov-Gabor polynomial. From Eqn. 1, is referring to input variable vector, is the number of input and is vector of coefficient weight). GMDH are self-organizing networks, developed in a layer by layer basis, following a systematic expansion exposure. The original GMDH method is called Multilayered Iterative Algorithm (MIA GMDH). There are four advantages to this algorithm: A small training set is required, the multiple layer structure of the designed system results in a feasible way of implementing high degree multinomial, the computation burden is reduced, and inputs/functions of inputs that have little impact on the output are automatically filtered out (Chang *et al*., 1999). Networks developed using methods based on GMDH concepts tend to have fewer, but far more flexible, nodes than a typical artificial neural network (Tamura and Halfon, 1980). The GMDH algorithm only used second order polynomial (Ivakhnenko 1970; Farlow 1981; Srinivasan 2008; Najafzadeh & Barani 2011) in form of:

$$\hat{y} = v_0 + v_1 x_i + v_2 x_j + v_3 x_i x_j + v_4 x_i^2 + v_5 x_j^2 \tag{2}$$

Eq. 2 as partial description (PD) provide the mathematical relation between the input and output variable. Least square method mostly applied in GMDH to obtain the weight coefficients for the models (Ivakhnenko 1971; Farlow 1984; Zadeh *et al.* 2002). The data set that consist of input and output is divided into two subset data that is training and forecasting. The input variables are pair using partial description in Eq. 2 in training data set. Then least square method applied in Eq. 2 to obtain the vector of coefficient.

$$Gv = Q_T \tag{3}$$

where v is the vector of coefficient of the partial description in Eq. 2.

$$\boldsymbol{v} = \{v_0, v_1, v_2, v_3, v_4, v_5\} \tag{4}$$

and

$$\boldsymbol{Q}_T = \begin{pmatrix} Q_{T,1} & Q_{T,2} & \cdots & Q_{T,p} \end{pmatrix}^T \tag{5}$$

$$\boldsymbol{G} = \begin{bmatrix} 1 & x_{1i} & x_{1j} & x_{1i}x_{ij} & x_{1i}^2 & x_{1j}^2 \\ 1 & x_{2i} & x_j & x_{2i}x_{2j} & x_{2i}^2 & x_{2j}^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{pi} & x_{pj} & x_{pi}x_{pj} & x_{pi}^2 & x_{pj}^2 \end{bmatrix} \tag{6}$$

Then, the best-estimated coefficients of partial description in Eq. 4.8 were obtained in the form of:

$$\boldsymbol{v = (G^T G)^{-1} G^T Y} \tag{7}$$

Therefore in each layer the total number of PD generated in the form of:

$$U = {}^nC_2 = \frac{n!}{(n-2)!2!} \tag{8}$$

where n is the number of input in each layer. The vector coefficient of each PD is determined using regression analysis then forming the quadratic equation which approximates the output yˆ in Eq. 2. After completion the process, the algorithm has constructed U number of new input variable but only one from U is chosen for the next layer based on regularity criterion r_j. This approach for identification of GMDH-type networks are called as error-driven approach (Zadeh *et al*. 2002).

$$r_k^2 = \frac{\sum_{i=1}^{p}\left(Q_{T,i} - \hat{Q}_{T,i}\right)^2}{\sum_{i=1}^{p} Q_{T,i}^2} \qquad j = 1, 2, 3, \ldots, U$$

where n is the number of input in each layer. The vector coefficient of each PD is determined using linear regression then forming the quadratic equation which approximates the output $\hat{y}$ in Eq. 4.8. After completing the process, the algorithm has constructed U number of new input variable but only one from U is chosen for the new input of GMDH based on RMSE value. After determining the new input, the whole GMDH process is repeated again. If $r_k \le r_{k-1}$, set new input variables and repeat the GMDH process, otherwise if r_k show an improvement the process is stopped and use the results from the previous minimum value of r_k.

2.2 Linear Regression Based on Regionalization

The variation in streamflow characteristics such as mean annual flow and flood quantiles are much related to the variations of physiographic and climatic factors. Using this fact empirical equations develop to relate streamflow characteristics with the metrological and physiographic variables. Shu and Ouarda (2008) pointed out that linear regression model is frequently used to estimate flood quantile as a function of site physiographical and other characteristics can be expressed. However, in practice the most commonly used relationship between the flood quantiles (Q_T) and catchment characteristics is the power form function (Mosley and Mckerchar, 1993). The power function has the following form:

$$Q_T = \alpha_0 A_1^{\alpha_1} A_1^{\alpha_2} \cdots A_2^{\alpha_n} \varepsilon_0 \tag{10}$$

where $\alpha_1, \alpha_2, \ldots, \alpha_n$ are the model parameters, $A_1, A_2, \ldots, A_n$ are the site characteristics, ε_0 is the multiplicative error term, n is the number of sites characteristics and Q_T is the flood quantile of T-year return period. The power form model on Eq. 4.1 can be linearized by a logarithmic transformation whereas the parameters of the linearized model can be estimated by a linear regression model. In other word, taking logs on both sides, Eq. (10) can be express as;

$$\log(Q_t) = \log(\alpha_0) + \alpha_1 \log(A_1) \ldots + \alpha_n \log(A_n) + \log(\varepsilon_0)$$

or

$$Y = X\beta + e \tag{12}$$

where

$Y = \log(Q_T)$ for $i = 1, 2, \ldots, m$: vector of flood quantiles from m sites.

$\beta = [\log(\alpha_0), \alpha_1, \ldots, \alpha_n]$: vector of coefficients;

$X = [(1, \quad \log A_i)]$: matrix of the logarithm of the physiographic and meteorological characteristics with the first column being equal to one.

$e = [\log(\varepsilon_0)]$: matrix of the logarithm of the error terms ε_0, which are assumed to be independent,

$m =$ total number sites

$n =$ number of independent variables excluding the constant term

Linear Regression builds relationship between the explanatory variables and response variables. One of the purposes of linear regression is to predict or estimate, the value of one variable from known or assumed values of other variables related to it. Linear regression generates model that can be used to forecast or estimate future values of the response variable given specified values of the explanatory variables. The goal in linear regression analysis is to identify variables that carry information about another variable and not to extrapolate from present conditions to future conditions. Linear regression can also be used for the related purposes of estimation and description. After applying the logarithmic to the power form model the parameters can be estimated using linear regression model.

2.3 Jackknife Procedure (Abdi &Williams, 2010)

The jackknife or "leave one out" procedure is a cross-validation technique first developed by Quenouille to estimate the bias of an estimator. John Tukey then expanded the use of the jackknife to include variance estimation and tailored the name of jackknife. The jackknife estimation of a parameter is an iterative process. First the parameter is estimated from the whole sample. Then each element is, in turn, dropped from the sample and the parameter of interest is estimated from this smaller sample. This estimation is called a partial estimate (or also a jackknife replication). A pseudo-value is then computed as the difference between the whole sample estimate and the partial estimate. These pseudo-values reduce the (linear) bias of the partial estimate (because the bias is eliminated by the subtraction between the two estimates). Although the jackknife makes no assumptions about the shape of the underlying probability distribution, it requires that the observations are independent of each other. Technically, the observations are assumed to be independent and identically distributed. This means that the jackknife is not, in general, an appropriate tool for time series data. When the independence assumption is violated, the jackknife underestimates the variance in the data-set which makes the data look more reliable than they actually are.

2.4 Evaluation Criteria

To assess the performance of each regional flood frequency analysis model, the following numerical indices are used: mean absolute error (MAE), root mean square error (RMSE) and Nash-Sutcliffe coefficient of efficiency (CE). MAE, RMSE and CE are provided in Eq. 10 – Eq. 12, respectively.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|Q_{T,i} - \hat{Q}_{T,i}\right| \tag{13}$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(Q_{T,i} - \hat{Q}_{T,i}\right)^2} \tag{14}$$

$$CE = 1 - \frac{\sum_{i=1}^{n}\left(Q_{T,i} - \hat{Q}_{T,i}\right)^2}{\sum_{i=1}^{n}\left(Q_{T,i} - \overline{Q}_{T,i}\right)^2} \tag{15}$$

where $Q_{T,i}$ is the observed flows, $\hat{Q}_{T,i}$ is the predicted flows, $\overline{Q}_{T,i}$ is the mean of the observed flows, $\overline{\hat{Q}}_{T,i}$ is the mean of the predicted flows and n is the number of flow series that have been modeled. The MAE is related with the prediction bias whereas the RMSE is associated with the model error variance. Both MAE and RMSE evaluate how closely the predictions match the observations by judging the best model based on the relatively small MAE and RMSE values. The coefficient of efficiency (CE) provides an indication of how good a model is at predicting values away from the mean. CE ranges from $-\infty$ in the worst case to 1 (perfect fit). The efficiency of lower than zero indicates that the mean value of the observed flow would have been a better predictor than the model.

3.0 CATCHMENT DATA SET

3.1 Introduction

The data were obtained from Department of Irrigation and Drainage, Ministry of Natural Resources and Environment, Malaysia. There were seventy gauged stations selected including all the stations located at Peninsular Malaysia. They are located within latitude 1° N-5° N and longitudes of 100° N-104° N. The stations include wide variety of basins region ranging between 16.3 km2 to 19,000 km^2. The period of the flow series for different sites varies from 11 -50 years starting from 1959 – 2009. These data were processed in two stages. First, catchment descriptors were extracted for each site.

Second, the annual peak flow was used to estimate selected T-year flood events for each catchment.

3.2 Catchment Descriptors

The variables selected in this study on the basis of previous study by Seckin (2011) and by Shu and Ouarda (2008). The four physiographical variables are catchment area, elevation, mean river slope and longest drainage path. The meteorological variable is mean annual total rainfall. The summary statistics of these variables are presented in Table 1. The descriptive statistics include minimum, maximum, mean and standard deviation for each variable. The variables shown in the table are catchment area (AREA), mean elevation (ELV), longest drainage path (LDP), mean river slope (SLP) and annual mean total rainfall (AMR).

3.2 GMDH and LR implementation

GMDH and LR is simulated using MATLAB software. Flood quantile for T=10 year and T=100 and catchment characteristics are converted into the natural logarithm form. The must be converted into natural logarithm form before implement GMDH model and LR model. GMDH model used second order polynomial as partial description to construct mathematical relation between input (AREA, ELE, MCS, LDP and AMR) and output variables. The process is repeated until GMDH achieve the most optimize solution.

Table 1 Descriptive statistics of hydrologic, physiographical and meteorological variables

Variables	Min	Mean	Max	STD
AREA [km^2]	30	1787.05	19000	3676.28
ELV [m]	4	99.49	1450	249.99
LDP [m]	3800	38457.97	280000	59553.88
SLP [%]	0.01	0.40	2.56	0.50
AMR [mm]	314.30	2099.75	4678.70	717.26
Q_{10} [m^3/s]	12.87	716.15	7256.76	1451.10
Q_{100} [m^3/s]	43.82	1194.17	11218.89	2270.77

4.0 RESULTS AND DISCUSSION

The objective of this paper is to assess the performance of the GMDH model in estimating flood quantile at ungauged sites in Peninsular Malaysia. there are five variables using in this study. The five variables are area, elevation, longest drainage path, mean catchment slope and annual mean total rainfall. The performance of each model depend on it prediction quantiles. The prediction quantiles compared in the real domain and not the logarithm transformation (Pandey & Nguyen 1999). To simulate the ungauged site, a jackknife procedure is implemented. In jackknife procedure, one site is removed from data and model parameters are estimated using the data from remaining site. The process is repeated until all stations are removed at least one (Pandey and Nguyen, 1999).

Table 2 Comparative performance between models obtained from the jackknife procedure

	T = 10 year		
Model	RMSE	MAE	CE
LR	820.9721	402.3754	0.7147
GMDH	427.3743	145.4721	0.9124
	T = 100 year		
Model	RMSE	MAE	CE
LR	1396.4232	706.0263	0.5983
GMDH	807.1215	309.5258	0.8866

Thus, the total number of developed model becomes equal to the number of sites in the region. Separate models are develop for 10- and 100-year flood quantiles.In order to assess the performance of proposed model, mean absolute error (MAE), root mean square error (RMSE) and Nash-Sutcliffe coefficient of efficiency (CE) were determined on Table 2. Both LR model and GMDH model used five catchment characteristics (AREA, ELV, LDP, SLP and AMR) as input. Table 2 show the RMSE, MAE and CE statistics for LR model and GMDH model. In estimating for flood quantile T=10 years, RMSE, MAE and CE for LR model are 820.9721, 402.3754 and 0.7147 meanwhile for GMDH model are 427.3743, 145.4721 and 0.9124. Then estimating for flood quantile T=100 years, RMSE, MAE and CE for LR model are 1396.4232, 706.0263 and 0.5983 meanwhile for GMDH model are 807.1215, 309.5258 and 0.8866. The RMSE and MAE indices provide an assessment of the prediction relative accuracy on square and absolute scale of error, respectively. The lower value of RMSE and MAE statistic indicate that the model prediction was closed to the observation value. GMDH model has lower RMSE and MAE value compare to LR model which indicated the prediction of GMDH model is better and more closed to observation flows. The NASH statistic provides overall assessment estimation. Model with NASH values that close to 1 mean the model produce near perfect estimation. According to Shu and Ouarda (2008) the NASH values that are closed and greater than 0.8 are generally acceptable. The NASH value for GMDH are over than 0.8 for both estimation of flood quantile which indicates that GMDH model achieved acceptable result. RMSE and MAE used to measure the performance accuracy of models implement in this study. The plot between observed and model-predicted quantile in the form of quantile-quantile ("Q-Q") plots are given in Figure 2 and Figure 3. The "Q-Q" plot is a subjective means of assessing closeness of the predicted are closed to fitted ones. If predicted quantiles are closed to fitted ones, then the points in the "Q-Q" plots should fall closed to the 45° line. It can be seen that GMDH model prediction of flood quantile are more closed to the 45° line compare to LR model for T=10 and T=100 year.

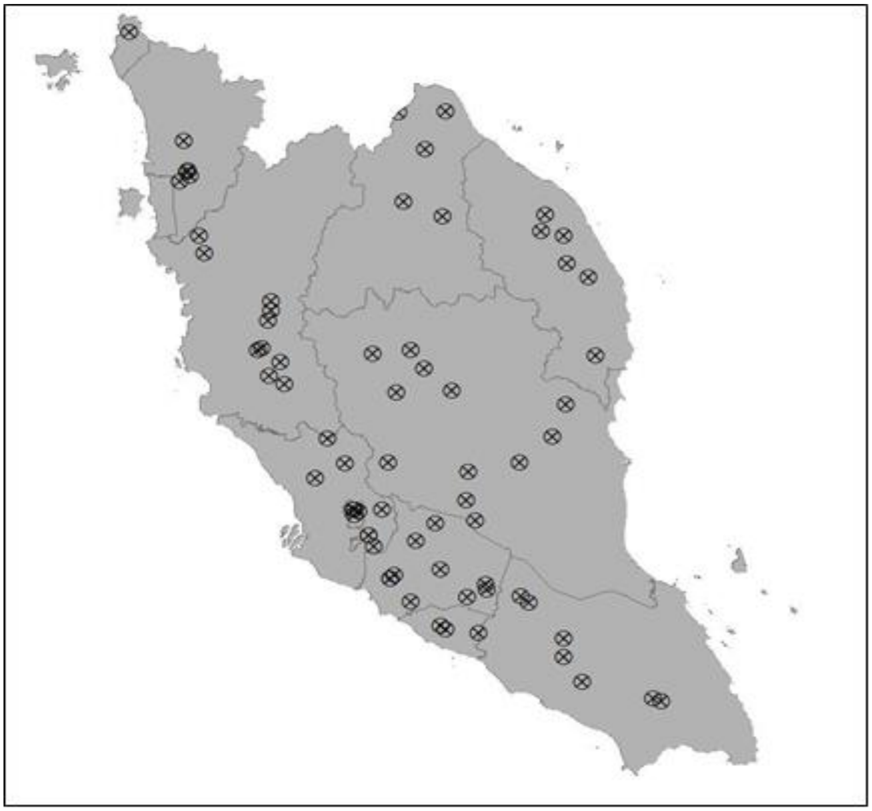

Figure 1 Map showing location of stream flow stations used in the study

4.0 CONCLUSION

The results obtained in this study show that GMDH model can be used to estimate flood quantile for ungauged site. The GMDH model was compared with LR model because LR model is the most common model used in estimating flood quantile at ungauged site. For modeling study, hydrologic and physiographic data from 70 catchments in the province of Peninsular Malaysia were used. The jackknife procedure is needed to simulate ungauged site. In this study , five input variables were implement that are catchment area, elevation, longest drainage path, slope of the catchment and annual mean rainfall to applied in GMDH model and LR model. The performance of each model is examine using RMSE, MAE and CE. To cover both the high and low sides of the flood distribution, the flood quantiles associated with 10- and 100-year return periods were considered. The comparison between GMDH model and LR model shows that GMDH model performance of GMDH model is better than LR model in estimating flood quantile. Finally, an investigation of GMDH model could yield further insights into the relationships between catchment properties and flood estimation in ungauged catchments.

Acknowledgement

The authors thankfully acknowledge the financial support by Kementerian Pendidikan Malaysia (MyBrain15) and also would like to thank the Department of Irrigation and Drainage, Malaysia for providing data.

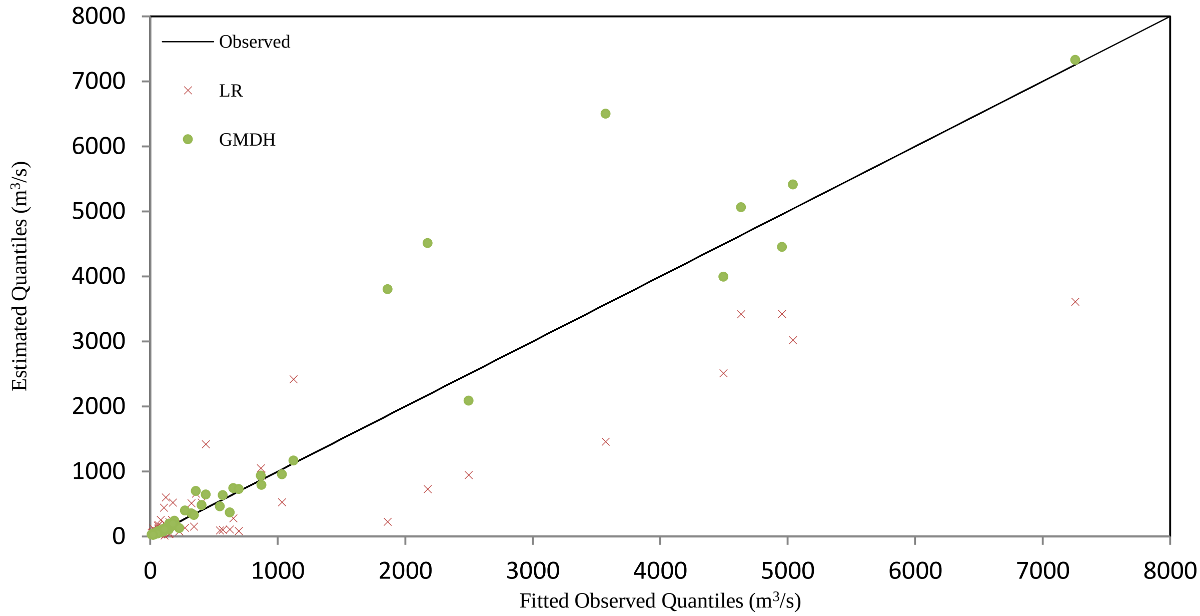

Figure 2 Comparison of the observed and estimated flood quantiles values using LR model and GMDH model for *T*=10 years

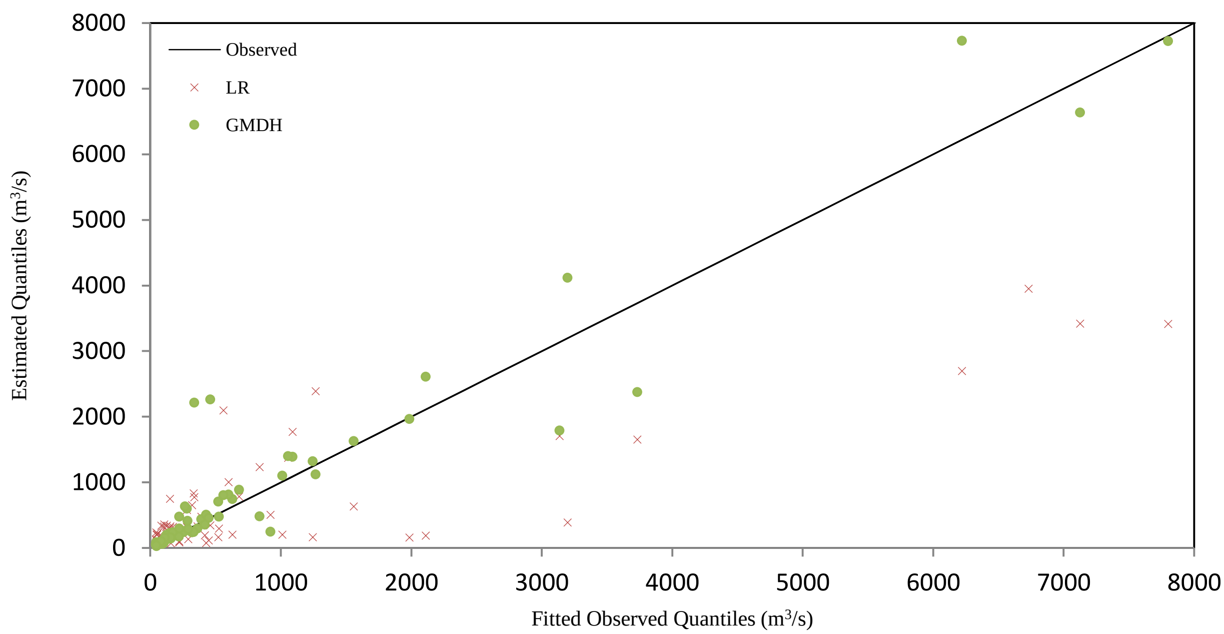

Figure 3 Comparison of the observed and estimated flood quantiles values using LR model and GMDH model for *T*=100 years

References

[1] Besaw, L. E., Rizzo, D. M., Bierman, P. R. and Hackett W. R. 2010. Advances in Ungauged Streamflow Prediction Using Artificial Neural Networks. *Journal of Hydrology*. 386: 27-37.

[2] Seckin, N. 2011. Modeling Flood Discharge at Ungauged Sites Across Turkey Using Neuro-Fuzzy and Neural Network. *Journal of Hydroinformatics.* 13(4):842-849.

[3] Robson, A. J. & Reed, D. W. 1999. *Flood Estimation Handbook*. *Statistical Procedures for Flood Frequency Estimation*. Institute of Hydrology, Wallingford, UK. 3.

[4] Mamun, A. A., Hashim, A. and Amir, Z. 2012. Regional Statistical Models for the Estimations of Flood Peak Values at Ungauged Catchments: Peninsular Malaysia. *Journal of Hydrologic Engineering*. 17(4): 547-553.

[5] Dawson, C. W., Abrahart, R. J., Shamseldin, A. Y. and Wilby, R. L. 2006. Flood Estimation at Ungauged Sites Using Artificial Neural Network. *Journal of Hydrology*. 319: 391-409.

[6] Hosking, J. R. M. and Wallis, J. R. 1997. *Regional Frequency Analysis: An Approach Based on L-Moments*. UK: Cambridge, University Press.

[7] Jain, A. and Kumar, A. M. 2007. Hybrid Neural Network Models for Hydrologic Time Series Forecasting. *Applied Software Computing. 7*: 585-592.

[8] Thomas, D. M. and Benson, M. A. 1970. Generalization of Streamflow Characteristics from Drainage-Basin Characteristics. *US Geological Survey, Water Supply Paper.*

[9] Fennessey, N. and Vogel., R. M. 1990. Regional flow-Duration Curves for Ungauged Sites in Massachusetts. *Journal Water Resource Planning Management.* 116(4).

[10] Mosley, M. P. and McKerchar, A. 1993. *"Streamflow''Handbook of Hydrology*. New York: McGraw-Hill.

[11] Vogel, R. M. and Kroll, C. N. 1990. Generalized Low-Flow Frequency Relationships for Ungauged Sites in Massachusetts. *American Water Resources Association.* 26(2): 241-253.

[12] Shu, C. and Ourda, T. B. M. J. 2008. Regional Flood Frequency Analysis at Ungauged Sites Using the Adaptive Neuro-Fuzzy Inference System. *Journal of Hydrology*. 31-43.

[13] Mohamoud, Y. M. 2008. Prediction of Daily Flow Duration Curves and Streamflows for Ungauged Catchments Using Regional Flow Duration Curves. *Hydrological Sciences.* 53(4): 706-724.

[14] Oh, S. K., and Pedrycz, W. 2002. The Design of Self-Organizing Polynomial Neural Networks. *Inf. Science*. 141: 237-258.

[15] Zadeh, N., Darvideh, A., Felezi, M. A. and Gharababaei, H. 2002. Polynomial Modeling of Explosive Compaction Process of Metallic Powders using GMDH-type Neural Networks and Singular Value Decomposition. *Modelling Simulation Material Science Engineering.* 10: 727-744.

[16] Kondo, T. and Pandya, A. S. 2004. *Identification of the Multi-Layered Neural Networks by Revised GMDH-Type Neural Network Algorithm with PSS Criterion*. Springer-Verlag Berlin Heidelberg. 1051-1059.

[17] Onwubolu, G. C. 2009. Prediction of Burr Formation during Face Miling using a Hybrid GMDH Network Model with Optimized Cutting Conditions. *Journal Advances Manufactured Technology.* 44: 1083-1093.

[18] Tamura, H. and Halfon, E. 1980. Identification of a Dynamic Lake Model by the Group Method of Data Handling: an Application to Lake Ontario. *Ecological Modeling.* 11: 81-100.

[19] Ivakhnenko, A. G. 1970. Polynomial Theory of Complex Systems. *IEEE Transactions System Manufactured Cybern.* 1: 364-378.

[20] Najafzadeh, M. and Barani, G. A. 2011. Comparison of Group Method of Data Handling Based Genetic Programming and Back Propagation Systems to Predict Scour Depth Around Bridge Piers. *Scientia Iranica.* 18(6): 1207-1213.

[21] Sharma, A. and Onwubolu, G. 2009. Hybrid Particle Swarm Optimization and GMDH System. *Hybrid Self-Organizing Modeling Systems.* 211: 193-231.

[22] Farlow, S. J. 1984. *Self-Organizing Methods in Modeling: GMDH Type Algorithms.* 270 Madison Avenue, New York: Marcel Dekker, Inc. 54.