

Flood Frequency Analysis at Ungauged Site Using Group Method of Data Handling and Canonical Correlation Analysis

Basri Badyalina¹ & Ani Shabri¹

¹ Department of Mathematics, Universiti Teknologi Malaysia, Malaysia

Correspondence: Ani Shabri, Department of Mathematics, Universiti Teknologi Malaysia, Johor Bahru, Malaysia. E-mail: ani@utm.my

Received: November 17, 2014

Accepted: November 29, 2015

Online Published: April 16, 2015

doi:10.5539/mas.v9n6p48

URL: <http://dx.doi.org/10.5539/mas.v9n6p48>

Abstract

Model based on canonical correlation analysis (CCA) and group method of data handling (GMDH) are explicate to obtain a better flood quantile estimation at ungauged sites. CCA is used to build a canonical physiographical space by applying the site characteristics from gauged station. Then GMDH model is used to distinguish the functional relationship between flood quantiles and the physiographic variables in the CCA space. The proposed model is applied to 70 catchments in Peninsular Malaysia. The jackknife procedure is used to evaluate the performance of proposed model. Result of proposed model compared with Traditional CCA model, linear regression (LR) model and GMDH model. The results indicated that the proposed model CCA-GMDH deliver the best performance among all models in terms of prediction accuracy.

Keywords: ungauged, group method of data handling, canonical correlation analysis

1. Introduction

Flood event is one of the most life-threatening and repeated type of natural disasters that take place in Peninsular Malaysia. Flood event contributes to a lot of damages to properties, infrastructures and even loss of people lives. Flood undoubtedly cannot be prevented from occurring, but human being can prepare for it. This problem makes a reliable estimation of flood quantiles is necessary for planning flood risk project (e.g., roads, culverts and dams), the safe design of the river system, and it give a closed valuation budget of flood protection project. In order to acquire accurate estimation of flood quantiles, recorded historical time series data of the stream flows required. Usually, long term historical data needed for estimation to produce a more reliable outcome compared to short term data and may also reduce the risk. However, it usually occurs that the historical data at target site not always available. Although at-site of interest may have some available data but the data are not enough to describe the catchment flow because of the changes in watershed characteristics such as urbanization (Pandey and Nguyen, 1999). The UK Flood Estimation Handbook (FEH) notes “many flood estimation problems arise at ungauged sites that there are no flood peak data” (Reed and Robson, 1999).

Mamun et al. (2012) stated that the river located in Malaysia is gauged only at a strategic location, and another river is usually ungauged. The ungauged river could become a problem to the developer when development projects located at ungauged catchments. Typically some site characteristics at the ungauged sites are exist. Thus, regionalization is conducted to make the estimation of flood quantile at ungauged sites using physiographic characteristics. Regionalization technique consists of fitting a probability distribution to series of flow and then relating the relationship to catchment characteristics (Dawson et al., 2005). The variables t the flood quantile estimation includes storm characteristics (duration and intensity events), basin characteristics (slope, size, storage and shape characteristics of the catchment), climatic characteristics (humidity, wind and temperature characteristics) and geomorphologic characteristics (topology, land use patterns, vegetation and soil types that affect the infiltration) (Jain and Kumar 2007). In relating flood quantile at site of interest to catchment characteristics a power form equations are mostly used (Pandey and Nguyen, 1999; Seckin, 2011; Mamun, 2012).

Canonical correlation analysis (CCA) is also frequently used approach to defining hydrological neighborhoods (Ouarda et al., 2000, 2001). Cavadias (1990) introduced CCA to flood quantile estimation where region formed on the basis of visual judgments of clustering patterns. Shu and Ourda (2007) using artificial neural network in canonical correlation analysis produce a better estimation of flood quantile compare used only artificial neural

network to estimate flood quantile in Quebec, Canada. Ouarda et al. (2000) applied the CCA approach to estimating extreme flood quantiles in Quebec, Canada. On the following year Ouarda et al. (2001) proposed the additional improvement to the method and detailed algorithm to delineate homogenous regions for gauged and ungauged sites using CCA.

At ungauged sites linear regression (LR) model is always worthy of depended on estimates of flow statistics or flood quantiles (Shu & Ouarda, 2008; Pandey & Nguyen, 1999). Mamun et al. (2012) used linear regression of various return periods in ten flood regions in Peninsular Malaysia. Linear regression usually integrated with CCA to provide quintile estimation especially at ungauged sites (Shu & Ouarda, 2007). There are many models whereby the relationship between catchments streamflow and catchments characteristics can be expressed. However, in practice the most commonly used relationship between the flood quantiles and catchment characteristics is the power form function (Thomas and Benson, 1970). The power function has the following form:

$$Q_T = \alpha_0 A_1^{\alpha_1} A_1^{\alpha_2} \dots A_n^{\alpha_n} \varepsilon_0 \quad (1)$$

where $\alpha_1, \alpha_2, \dots, \alpha_n$ are the model parameters, A_1, A_2, \dots, A_n are the site characteristics, ε_0 is the multiplicative error term, n is the number of sites characteristics and O_T is the flood quantile of T-year return period. The power form model on Eq. 1 linearized by a logarithmic transformation whereas the parameters of the linearized model can be estimated by a linear regression model. McCuen et al. (1990) stated that the solution obtained by linear regression methods theoretically unbiased in the logarithmic domain but biased in real flood flow area.

As an alternative to standard nonlinear regression methods, group method of data handling (GMDH) by Badyalina et al. (2014) for flood quantile estimation. The GMDH algorithm was first presented by a Ukrainian scientist Ivakhnenko and his colleagues in 1968 to produce mathematical models of complex systems by handling data samples of observations (Ivakhnenko and Ivakhnenko, 1974). The GMDH method was initially formulated to resolve higher order regression polynomials specifically for solving modeling and classification problems. The GMDH model was developed by Ivakhnenko to identify nonlinear relationship between inputs and output variables. Oh and Pedryz (2002) stated that GMDH performs the self-organizing control of data mining process, is a high efficiency and intelligible algorithm for constructing optimization model by objective approach from the original input variables.

In the present paper, regional flood quantile estimation methods based on CCA and GMDH model are proposed. CCA is used to define a transformed physiographical space. A GMDH then used to establish the nonlinear relationships between the site physiographical space and hydrological variables is estimated. A comparison study is carried out between the proposed model and several other model using data from the province of Peninsular Malaysia.

2. Methodology

2.1 Group Method of Data Handling

Group Method of Data Handling (GMDH) model was introduced by Ivakhnenko on 1970 to solve complex non-linear multidimensional that has short data series (Ivakhnenko 1970). The GMDH algorithm that describes the relationship between input and output signal can be represented by Volterra series (Ivakhnenko 1970) in the form of:

$$y = v_0 + \sum_{i=1}^n c_i x_i + \sum_{i=1}^n \sum_{j=1}^n v_{ij} x_i x_j + \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n v_{ijk} x_i x_j x_k + \dots \quad (2)$$

Eq. 2 known as Kolmogorov-Gabor polynomial. From Eq. 2, $\vec{X}(x_1, x_2, \dots, x_n)$ is referring to input variable vector, n is the number of inputs and $V(v_1, v_2, \dots, v_n)$ is a vector of the coefficient weight. In the GMDH algorithm, Eq. 2 is called the complete description of the nonlinear system. However, most application only used second order polynomial called partial descriptions (PD) of the nonlinear system that can be expressed by a system of transfer function consisting of only two variables (Srinivasan 2008; Najafzadeh & Barani 2011). The PD is in the form of:

$$\hat{y} = v_0 + v_1 x_i + v_2 x_j + v_3 x_i x_j + v_4 x_i^2 + v_5 x_j^2 \quad (3)$$

Eq. 3 as partial description (PD) provides the mathematical relation between the input and output variable. Linear regressions mostly used in GMDH to obtain the weight coefficients for the models (Ivakhnenko 1970;

Zadeh et. al 2002). The data set that consists of input and output are divided into two subsets that are the modeling and forecasting based on jackknife procedure. The input variable $X = \{x_{ij}\}$ are paired using partial description in Eq. 3 in modeling data set. Then a linear regression used in Eq. 3 is to obtain the vector of coefficients.

$$Gv = Y \quad (4)$$

Where v is the vector of coefficient of the partial description in Eq. 3.

$$v = \{v_0, v_1, v_2, v_3, v_4, v_5\} \quad (5)$$

and

$$Y = (y_1 \quad y_2 \quad \dots \quad y_p)^T \quad (6)$$

is a vector of output from training data set.

$$G = \begin{bmatrix} 1 & x_{1i} & x_{1j} & x_{1i}x_{1j} & x_{1i}^2 & x_{1j}^2 \\ 1 & x_{2i} & x_{2j} & x_{2i}x_{2j} & x_{2i}^2 & x_{2j}^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{pi} & x_{pj} & x_{pi}x_{pj} & x_{pi}^2 & x_{pj}^2 \end{bmatrix} \quad (7)$$

Then, the best-estimated coefficients of partial description in Eq. 4 were obtained in the form of:

$$v = (G^T G)^{-1} G^T Y \quad (8)$$

Therefore in each layer the total number of PD generated, and the RMSE are as follow:

$$U = {}^n C_2 = \frac{n!}{(n-2)!2!} \quad (9)$$

$$RMSE_k = \sqrt{\frac{1}{p} \sum_{i=1}^p (y_i - \hat{y}_{i,k})^2} \quad k \in [1, 2, \dots, U] \quad (10)$$

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} . After completing the previous process, the algorithm has constructed U number of new inputs variable, but only one from U is chosen for the new input of GMDH based on RMSE value. This method for identification of GMDH-type networks is called as error-driven approach (Zadeh et. al. 2002). After determining the new input, the whole GMDH process is repeated again. If $RMSE_k \leq RMSE_{k-1}$, set new input variables and repeat the GMDH process, otherwise if RMSE shows an improvement the process is stopped and use the results from the previous minimum value of RMSE.

2.2 Canonical Correlation Analysis (Shu & Ouarda, 2007)

Canonical correlation analysis (CCA) is a path explaining the linear relationship between two sets of variables. Consider X and Y are two random variables, CCA computes two sets of basis vectors (canonical variables), one for X and the other for Y , such that the correlations between the projections of the variables onto these basis vectors are mutually maximized (Muirhead, 1982). The maximum number of canonical variable pairs is equal to or less than the smallest dimensionality of the two variables. Let W and V be linear combinations of X and Y , respectively.

$$W = \alpha' X \quad (11)$$

$$V = \beta' Y \quad (12)$$

Let Σ be a covariance matrix of variable X and Y , defined as

$$\Sigma = \text{cov} \begin{pmatrix} X \\ Y \end{pmatrix} = \begin{bmatrix} \sum X & \sum XY \\ \sum YX & \sum Y \end{bmatrix} \quad (13)$$

The correlation between W and V can be calculated as

$$\rho = \frac{\alpha' \sum XY^\beta}{\alpha \sum X^{\alpha\beta} \sum Y^\beta} \quad (14)$$

The goal of CCA is to find the vectors of α and β maximizing ρ subject to the constraint that W and V must have unit variances. Once the first pair of canonical variables is obtained, other pairs of canonical variables can be obtained in the uncorrelated directions to the previous ones by maximizing Eq. 14 subject to the constraint of unit variance. CCA was recently used by Chokmani and Ouarda (2004) to construct a transformed space defined by the physiographical and meteorological characteristics. The hydrological variables (flood quantiles in our case) are usually not continuous in the geographical space. However, they are constant in the canonical physiographical space (Chokmani and Ouarda, 2004). This characteristic is crucial for flood estimation at ungauged sites. Because the physiographic variables and the meteorological variables are available at the ungauged sites, one can easily locate an ungauged site in the physiographical space constructed by these variables. For more detailed information regarding CCA, the readers are referred to Ouarda et al. (2001).

2.3 Integrating CCA and GMDH for Regional Flood Frequency Analysis at Ungauged Sites

Usually at ungauged sites, historical flood data are not available and became a problem if directly used to estimate the hydrological variables such as flood quantiles. Contrary, building a functional relationship between the hydrological variables and the physiographical variables make the estimation of flood quantile is possible at ungauged site. Estimation of various flood quantile required model to used data from gauged station that around ungauged site. In the proposed model shown in this paper, the original physiographical variables are converted into canonical space. Then the expected variables are used as input variables for GMDH model to estimate the particular flood quantile. Suppose a set of catchments characteristics, X and hydrological variables, Y are related with each gauged station. Applying CCA, canonical variables W and V can get as a linear combination of W and V , respectively. The coefficients used for the combination computed so that the correlation between the variables W and V is maximized. Knowing the combination coefficients, the physiographical variable X_u for an ungauged site can be easily projected into the CCA space to obtain the physiographical variable in the CCA space. The goal of the GMDH model is to approximate the functional relationship between the canonical variables and the hydrologic variables Y which act as an input and output of a GMDH, respectively. The canonical variables V not used in the GMDH training and estimation phase. To achieve this goal, the GMDH must be trained using the samples from the gauged sites in the study area.

2.4 Evaluation Criteria

The performance of each model is evaluated with the following error indices which are the mean absolute error (MAE), root mean square error (RMSE) and Nash-Sutcliffe coefficient of efficiency (CE) and correlation coefficient. The definitions of MAE, RMSE, CE and BIASr, are provided in Eq. 15- Eq. 18, respectively.

$$MAE = \frac{1}{n} \sum_{i=1}^n |Q_{T,i} - \hat{Q}_{T,i}| \quad (15)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (Q_{T,i} - \hat{Q}_{T,i})^2} \quad (16)$$

$$CE = 1 - \frac{\sum_{i=1}^n (Q_{T,i} - \hat{Q}_{T,i})^2}{\sum_{i=1}^n (Q_{T,i} - \bar{Q}_{T,i})^2} \quad (17)$$

$$BIASr = \frac{1}{n} \sum_{i=1}^n \left(\frac{Q_{T,i} - \hat{Q}_{T,i}}{Q_{T,i}} \right) \quad (18)$$

where $Q_{T,i}$ is the observed flows, $\hat{Q}_{T,i}$ is the predicted flows, $\bar{Q}_{T,i}$ is the mean of the observed flows, $\bar{\hat{Q}}_{T,i}$ is the mean of the predicted flows and n is the number of flow series that have been modeled. 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. Variance and the multiplied standard deviations of observed and predicted values.

3. Case Study

The hydrometric station network of Peninsular Malaysia is chosen as the case study of this work. According to the following criteria, 70 hydrometric stations located in Peninsular Malaysia are selected.

1. To get reliable at-site estimation, a historical flood record of 15 years or longer are needed.
2. The gauged river should present natural flow regime.
3. The selected stations are located within latitude $1^{\circ}N - 5^{\circ}N$ and longitude of $100^{\circ}E - 104^{\circ}E$. The areas of these catchments are ranging between 16.3 km² to 19,000 km².

The historical data of 70 catchments in the province of Peninsular Malaysia were implemented in this study. They are located within latitude $1^{\circ}N - 5^{\circ}N$ and longitude of $100^{\circ}E - 104^{\circ}E$. The areas of these catchments are ranging between 16.3 km² to 19,000 km². The locations of these catchments are shown in Fig. 1. Three types of data, physiographical, meteorological and hydrological are used in this study. The variables selected in this study on the basis of previous study by Seckin (2011) and by Shu and Ouarda (2007). Four physiographical variables are the 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), elevation (ELV), longest drainage path (LDP), mean river slope CCA-GMDH(SLP), annual mean total rainfall (AMR), magnitude of flood for return period T=10 year (Q_{10}), magnitude of flood for return period T=50 year (Q_{50}) and magnitude of flood for return period T=100 year (Q_{100}). The return period were estimated using selected distribution at each station. There are five distribution used in this study that are generalized extreme value distribution (GEV), generalized logistic distribution (GLO), generalized pareto distribution (GPA), pearson 3 distribution (P3) and three parameter lognormal (LN3). The flow data at each station were fitted using these five distributions. Then the best fitted distribution represent the station flow pattern used to estimate the flood quantile.

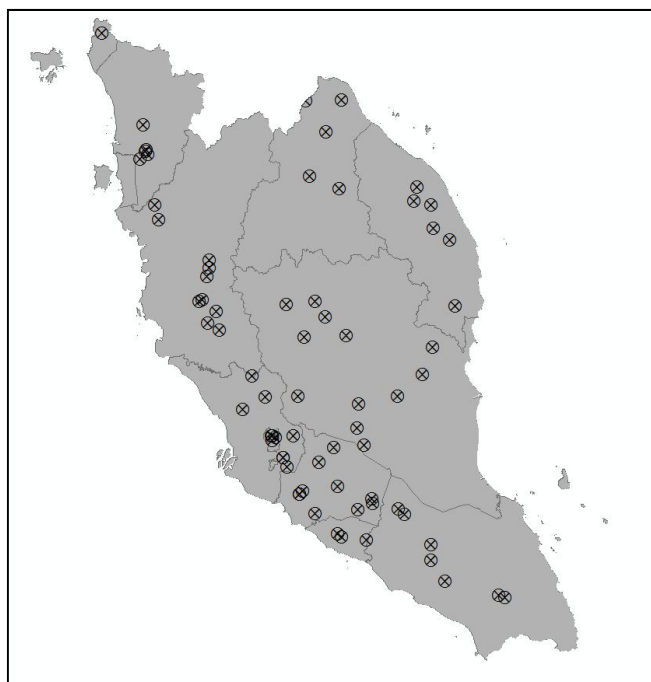


Figure 1. Map showing location of stream flow stations used in the study

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_{50} [m ³ /s]	29.54	1043.45	10089.80	2029.01
Q_{100} [m ³ /s]	43.82	1194.17	11218.89	2270.77

4. Discussion

There is one model proposed in this paper and three models used for comparison purpose are applied to the study area database. The proposed model is a combination between group method of data handling and canonical correlation analysis (CCA-GMDH). 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 estimated parameters are in turn used to predict quantile for the site not used in the model development. The process is repeated until all stations are removed at least once. The input variables for all models are catchment area, elevation, longest drainage path, river slope and annual maximum mean total rainfall. The results obtained using jackknife validation procedures are presented in Table 2. For each cell of Table 2, bold font denotes the best performing approach. A model can be claimed to produce a perfect estimation if the *NASH* criterion equal to 1. The model can be considered acceptable if the *NASH* criterion is greater than 0.8. The four models, ranked according to their performance in the *NASH* criterion from highest to lowest in estimating the 10, 50 and 100-year flood quantiles are listed as follows: GMDH-CCA, GMDH, Tradition-CCA and LR. The *NASH* value obtained from both CCA-GMDH and GMDH model in estimating the three particular quantiles are all above 0.8. This indicates that the GMDH models in the CCA space can provide satisfactory estimates.

Table 2. Jackknife validation results

	Hydrological Variables	Traditional			
		LR	CCA	GMDH	CCA-GMDH
CE	q10	0.6610	0.6733	0.7832	0.9210
	q50	0.6332	0.6455	0.8839	0.9406
	q100	0.5884	0.6004	0.8821	0.9565
RMSE	q10	841.8803	826.4962	673.3238	406.4181
	q50	1010.2217	993.1600	568.4410	406.4128
	q100	1242.7308	1224.5268	665.0904	403.8808
MAE	q10	406.8528	399.7822	235.8594	174.9786
	q50	496.8859	488.5106	202.7040	165.4816
	q100	618.1738	608.1246	289.2933	157.0950
BIASr	q10	-0.3383	-0.3331	-0.7849	-0.0973
	q50	-0.3425	-0.3366	-0.1159	-0.0472
	q100	-0.3549	-0.3477	0.1672	-0.0393

RMSE indices provide assessment of prediction accuracy in absolute and relative scale, respectively. The CCA-GMDH model has the best performance among all the models according to *RMSE* indices. Meanwhile for *MAE* indices, CCA-GMDH outperformed other models when estimating the 10, 50 and years flood quantile. The result obtained showed that CCA-based GMDH model showed significant improvement compared to GMDH model applied in original physiographical space. The proposed model lead to a better performance in the estimation because the combination both linear and nonlinear method. Other than that, after applying GMDH on CCA physiographical space, GMDH model chooses the best input to obtain a better estimation. The *BIASr* indices provide indication whether model tends to overestimate or underestimate. An analysis based on the *BIASr* index is used, both CCA-GMDH model and GMDH model underestimates flood quantiles. Estimation obtained from CCA-GMDH has the lowest bias. Overall, the CCA-GMDH model leads to a much better performance with CE, *RMSE* and *MAE* indices compare than GMDH, Traditional CCA and LR model. These indicate that applying GMDH model in the physiographical space can significantly improve the performance of

GMDH models than in original physiographical space. Chokmani and Ouarda (2004) concluded that the CCA technique is more capable of characterizing the physiographical space for conducting flood quantile estimation. The research result of this paper is consistent with their conclusions. The GMDH model outperforms both in the original space and the CCA physiographical space according to most performance indices. Thus, the CCA-GMDH model is better than traditional LR model.

5. Conclusions

The methodology of integrating the CCA technique and GMDH for flood quantile estimations at ungauged sites presented in this paper. CCA is used to project the site characteristics into the canonical physiographical space. GMDH model then used to approximate the functional relationship between flood quantiles and the projected physiographic variables. Three various return period in this study were used to see the capability of the model to estimate for short term and long term. CCA-GMDH model was compared to three other models that are LR, Traditional CCA and GMDH models. The result shows CCA-GMDH outperformed the comparison model in relative accuracy in estimation of flood quantile.

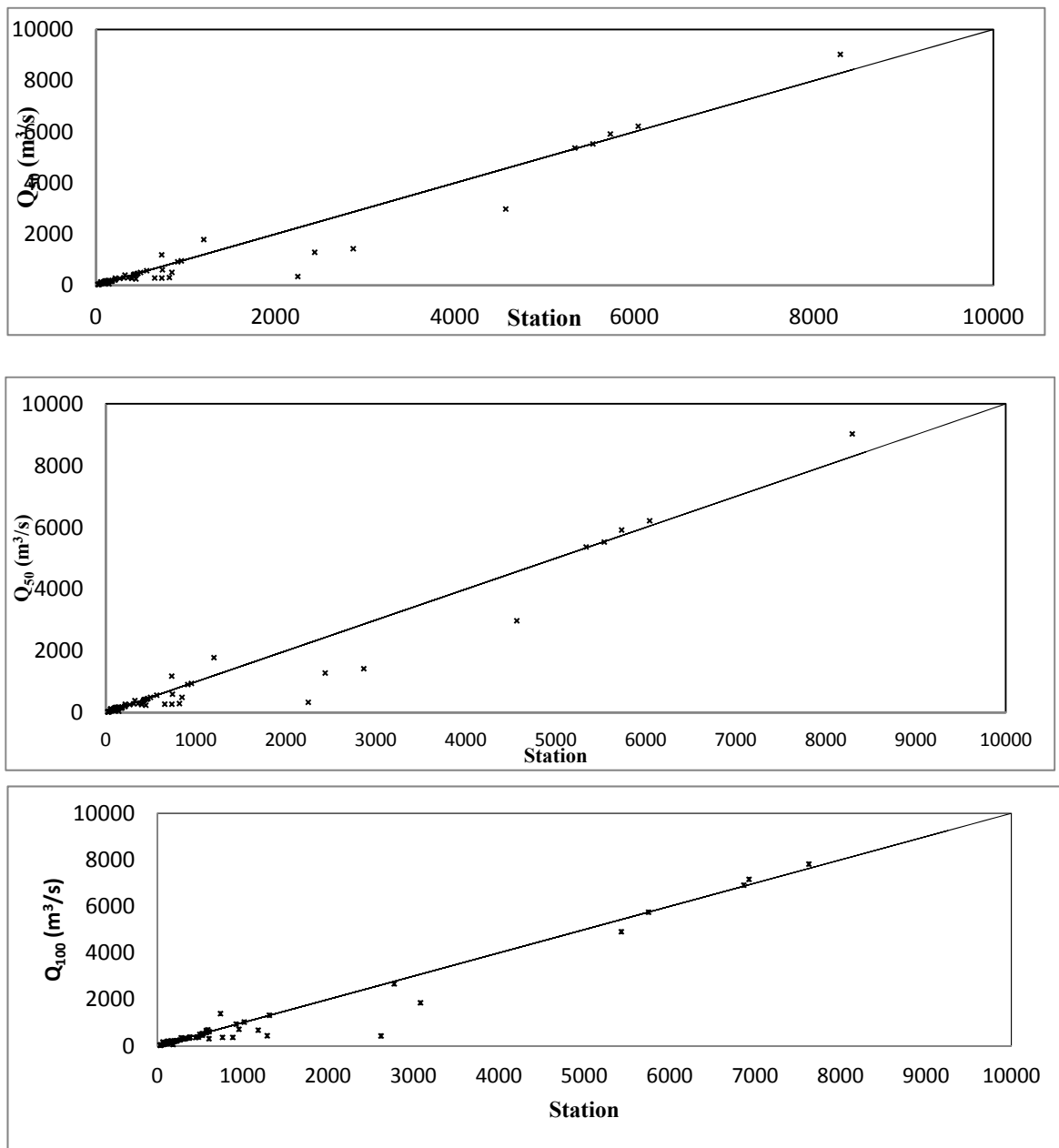


Figure 2. Jackknife estimation using the CCA-GMDH model

References

- Badyalina, B., & Shabri, A. (2014). Streamflow Estimation at Ungauged Site Using Wavelet Group Method of Data Handling in Peninsular Malaysia. *Int. Journal of Math. Analysis*, 8(11), 513-524. <http://dx.doi.org/10.12988/ijma.2014.4251>
- Cadavias, G. S. (1990). The Canonical Correlation Approach to Regional Flood Estimation. *IAHS Publication*, 191, 171-178.
- Dawson, C. W., Abrahart R. J., Shamseldin, A. Y., & Wilby R. L. (2006). Flood estimation at ungauged sites using artificial neural network. *Journal of Hydrology*, 319, 391-409. <http://dx.doi.org/10.1016/j.jhydrol.2005.07.032>
- Ivakhenko, A. G. (1970). Polynomial Theory of Complex Systems. *IEEE Transactions System Manufactured Cybern*, 1, 364-378. <http://dx.doi.org/10.1109/TSMC.1971.4308320>
- Ivakhnenko, A. G., & Ivakhnenko, N. A. (1974). Long Rand Prediction of Random Processes by GMDH algorithms using the unbiasedness Criterion and Balance of Variables Criterion. *Automatyka Soviet Automatic*, 4, 52-58.
- Jain, A., & Kumar A. M. (2007). Hybrid neural network models for hydrologic time series forecasting. *Applied Software Computing*, 7, 585-592. <http://dx.doi.org/10.1016/j.asoc.2006.03.002>
- Mamun, A. A., Hashim, A., & Amir, Z. (2012). Regional Statistical Models for the Estimations of Flood Peak Values at Ungauged Cathments: Peninsular Malaysia. *Journal of Hydrologic Engineering*, 17, 547-553. [http://dx.doi.org/10.1061/\(ASCE\)HE.1943-5584.0000464](http://dx.doi.org/10.1061/(ASCE)HE.1943-5584.0000464)
- McCuen, R. H., Leahy R. B., & Wallis J. R. (1975). Problems with Logarithmic Transformation In Regression. *Journal of Hydraulic Engineering*, 116, 414-428. [http://dx.doi.org/10.1061/\(ASCE\)0733-9429\(1990\)116:3\(414\)](http://dx.doi.org/10.1061/(ASCE)0733-9429(1990)116:3(414))
- Najafzadeh, M., & 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, 1207-1213. <http://dx.doi.org/10.1016/j.scient.2011.11.017>
- Oh, S. K., & Pedrycz, W. (2002). The Design of Self Organizing Polynomial Neural Networks. *Information Science*, 141, 237-258. [http://dx.doi.org/10.1016/S0020-0255\(02\)00175-5](http://dx.doi.org/10.1016/S0020-0255(02)00175-5)
- Ouarda, T. B. M. J., Girard, C., Cadavias, G. S., & Bobee, B. (2001). Regional flood frequency estimation with canonical correlation analysis. *Journal Hydrology*, 254, 157-173. [http://dx.doi.org/10.1016/S0022-1694\(01\)00488-7](http://dx.doi.org/10.1016/S0022-1694(01)00488-7)
- Ouarda, T. B. M. J., Hache, M., Breneau, P., & Bobee, B. (2000). Regional flood peak and volume estimation in a Northern Canadian basin. *Journal Cold Region Engineering*, 14, 176-191. [http://dx.doi.org/10.1061/\(ASCE\)0887-381X\(2000\)14:4\(176\)](http://dx.doi.org/10.1061/(ASCE)0887-381X(2000)14:4(176))
- Pandey, G. R., & Nguyen, V. T. V. (1999). A comparative study of regression based methods in regional flood frequency analysis. *Journal of Hydrology*, 225, 92-101. [http://dx.doi.org/10.1016/S0022-1694\(99\)00135-3](http://dx.doi.org/10.1016/S0022-1694(99)00135-3)
- Reed, D. W., & Robson, A. J. (1999) Flood Estimations Handbook. *Centre for Ecology and Hydrology*, UK.
- Seckin, N. (2011). Modelling flood discharge at ungauged sites across Turkey using neuro fuzzy and neural network. *Journal of Hydroinformatics*, 13, 842-849. <http://dx.doi.org/10.2166/hydro.2010.046>
- Shu, C., & Ouarda, T. B. M. J. (2007). Flood Frequency Analysis at Ungauged sites using artificial neural networks in canonical correlation analysis physiographic space. *Water Resources Research*, 43, 1-12. <http://dx.doi.org/10.1029/2006WR005142>
- Shu, C., & Ouarda, T. B. M. J. (2008). Regional flood frequency analysis at ungauged sites using the adaptive neuro-fuzzy inference system. *Journal of Hydrology*, 31-43. <http://dx.doi.org/10.1016/j.jhydrol.2007.10.050>

Copyrights

Copyright for this article is retained by the author(s), with first publication rights granted to the journal.

This is an open-access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).