Jurnal
Teknologi

Full Paper

STREAMFLOW PREDICTION IN UNGAUGED CATCHMENTS IN THE EAST COAST OF PENINSULAR MALAYSIA USING MULTIVARIATE STATISTICAL TECHNIQUES

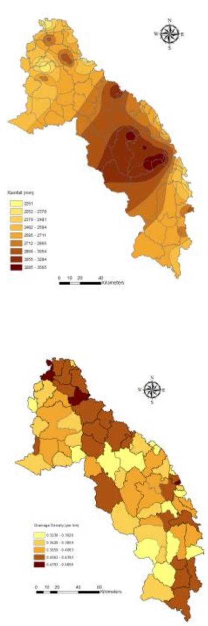
Article history
Received
23 May 2015
Received in revised form
10 January 2015
Accepted
5 April 2015

Salaudeen Abdul Razaq[a*], Tarmizi Ismail[b], Arien Heryansyah[b], Umar Faruk Lawan[b], Mahiuddin Alamgir[b], Sahar Hadi Pour[b]

[a]Department of Civil Engineering, Faculty of Engineering, Abubakar Tafawa Balewa University, (ATBU), Bauchi, Nigeria
[b]Department of Hydraulics and Hydrology, Faculty of Civil Engineering, Universiti Teknologi Malaysia, 81310 UTM Johor Bahru, Johor, Malaysia

*Corresponding author
aburuqayyatiyara@gmail.com

Graphical abstract



Abstract

The east coast of Peninsular Malaysia is one of the most vulnerable regions of Malaysia to hydrological disasters, which is believed to become more vulnerable due to climate change. Studies to have better understandings of the hydrological processes in the region are therefore, of paramount importance for disaster risk mitigation. However, unavailability of long-term river discharge data is one of the major constraints of hydrologic studies in the area. The major objective of this study is to predict river discharge in ungauged river basins in the study area. For this purpose, a set of multiple linear regression equations and exponential functions have been developed, which are expressed in the forms of multivariate equations. Available streamflow data along with other catchment characteristics from gauged catchments were used to develop the equations and were subsequently applied to the poorly gauged or ungauged catchments within the study area for prediction of streamflow. In this present study, 4 to 7 explanatory variables were selected as the input variables, which comprise of climatic, geomorphologic, geographic characteristics, soil properties, land use pattern and land cover of the area. Ten flow metrics as maximum, 0.05, 0.10, 0.25, 0.50, 0.75, 0.90, and 0.95, mean and minimum were therefore predicted. Thus, the results of the developed multivariate equations revealed the model to be capable of predicting the desired flow metrics at ungauged catchments in the area under consideration with reasonable accuracy.

Keywords: Flow duration curve, streamflow prediction, ungauged catchments, east coast of Peninsular Malaysia, multivariate equations

1.0 INTRODUCTION

Predictions of runoff hydrographs are key requirements for designing hydraulics structures, water resources planning and management, hydropower operation, hydrological disaster risk management as well as in assessing the effects of environmental changes [1, 2]. Streamflow predictions are approaches used to extrapolate by hydrological information transfer of contiguous ungauged catchment from gauged data through, hydrological model simulation, and other relevant methods [3]. However, runoff data are not available in many catchments of interest. Therefore, it is often required to predict runoff hydrographs of ungauged catchments from other information within that catchment or from other catchments [4]. Many methods have been developed and applied in different parts of the world for this purpose [1, 5, 6]; however, prediction in ungauged basins remains a major challenge in hydrology.

It is much more challenging in tropical regions where most of the catchments are ungauged, as such; the need for improved knowledge of flow variability in such regions become very urgent, especially in the context of changing hydrological processes and growing hydrological disasters due to climate change. Increasing severity and frequency of floods due to changing rainfall pattern is a growing concern in the east coast of peninsular Malaysia [7-9]. Major challenge in hydrological studies in the area is the unavailability of reliable and long-term streamflow data in most of the catchments of interest.

Flow duration curve (FDC) presents the probability of flood of a particular magnitude to be equaled or exceeded over a historical period. With FDC, a comprehensive graphical view of the historical change in the overall flow event is possible, for the catchment of interest. Because of highly skewed nature of daily streamflow data, nonparametric approach will be used for the FDC framework rather than the parametric to avoid high tendencies to biasness. Median annual FDC being most appropriate method for this type of study have been chosen [10].
Several models have been developed by different researcher for flow predictions in ungauged catchments across the globe. For example, Yaşar and Baykan [1] developed a model for predictions of FDCs in ungauged basins in US with Quasi-Newton method, known as Estimation of Regionalized Flow Duration Curve (EREFDC). Lacombe *et al*. [5] suggested predictions of streamflow in Makong basin, where a set of power law equations were developed and were reported to perform well. A methodology for estimating the flood frequency curve at sites for which there is no sufficient information to allow the direct calibration of a rainfall-runoff model relevant for that site was developed by Jones and Kay [2]. The present research is however, targeted at the east coast of Peninsular Malaysia by making comparison accordingly between linear regression equations and exponential functions for improved knowledge of flow variability of the area concerned.

2.0 METHODOLOGY

2.1 Description of the Study Area

The present study area comprises of some catchments in the states of Kelantan and Terengganu in the east coast of Peninsular Malaysia. It is inscribed by Longitude 101^0 and 103^{0}E, and Latitudes 4^0 and 5^0 N. It commands a total watershed area of approximately 17,000km^2, consisting of about 58 sub - catchment areas. The minimum and maximum elevations within the project area are approximately 3m and 2600m respectively above mean sea level. The River Kelantan catchment for instance is the longest river in catchment recording 180 to 300 m width covering a distance of 248 km and drains an area of 13,100 km^2. Approximately 68.5% of the population was reported to inhabit the river basin and they are as such; predisposed to the detrimental effects of frequent flooding episodes [11]. Terengganu on the other hand, is bordered in the northwest, southwest and in the east by the state of Kelantan, Pahang and South China Sea respectively. The total population of the state was put at 1,015,776 according to 2010 census with a density of 78 km^2. The region has tropical monsoon climate known to be fairly hot and humid throughout the year with temperature ranging between 21°C and 32°C all year round [12]. Figure 1 Shows the location map and the streamflow gauging stations.

2.2 The Multivariate Equations

In order to predict the various flow metrics in the catchments, multiple linear regressions (Eq.(1)) from the available streamflow data from some stations were developed to established functional relationship among the variables as already being used in various catchments of the world [5].

$$Q_i = \beta_0 + \beta_1.X_1 + \beta_2.X_2 + \cdots + \beta_n.X_n + \varepsilon \qquad (1)$$

A logarithmic transformation of the variable prior to multiple linear regression, results in the exponential equation as given below:

$$Q_i = \exp\{[\beta_0 + (\beta_1 X_1) + (\beta_2 X_2 + \cdots + (\beta_n X_n)] - 1\} + v \qquad (2)$$

Where Q is the streamflow from n catchments with characteristics X (i = 1,2,…n) whose coefficients are β (i = 1,2,…n).

β_0 is the y intercept and, ε, v are the normally and log-normally errors of the models.

The transformed data for the catchment characteristics, X_i and flow Q_i will be incremented by one prior to being used in the regression analysis. As such; the variables, Q_i and X_i were replaced by Q_i+1 and X_i+1 respectively.

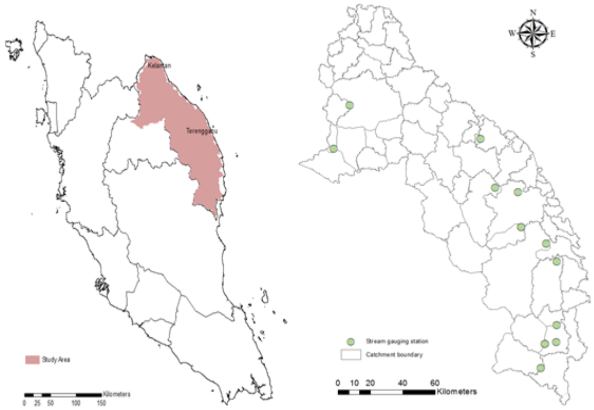


Figure 1 Shows the location map and the streamflow gauging stations of the study area

2.3 Streamflow

Available records of daily streamflow data ranging from 6 – 51years for minimum of twelve gauging stations obtained from the Department of Irrigation and Drainage (DID), Malaysia were processed for this study. All - year of records median annual FDCs for all the gauged catchments were computed rather than period of records FDCs as suggested by Vogel and Fennessey [10]. Seven flow percentiles namely 0.05, 0.10, 0.25, 0.50, 0.75, 0.90, and 0.95 were estimated from the FDCs. The maximum, minimum and mean flows were equally selected. The above ten metric flows were used as the candidate explanatory variables for the development of the regression equations.

2.4 Rainfall

The east coast of Peninsular Malaysia is one of the zones most vulnerable to natural disasters and climate change [13]. There has been increase in the cases of extreme rainfall events in the region in recent years, particularly during the northeast monsoon season [14]. Long-term rainfall data in the gauged catchments were equally collected from DID, Malaysia for the same hydrological year as the streamflow data, and was subsequently used to compute the areal precipitation time series for each gauged catchment, using inverse weighted distance method. The median annual rainfalls [5] were extracted rather than the mean annual rainfall because of the positively skewed nature that are associated with hydrologic data as mentioned earlier, thus, avoiding bias data possibilities. Correlations of rainfall with each of the explanatory variables for prediction were tested and were found to record the highest positive correlation coefficient.

2.5 Geomorphologic and Geographic Characteristics

Table 2 presents the summary of explanatory variables used in the multiple regression analysis. The geomorphologic characteristics are essential components of any hydrologic studies of a catchment as this presents properties such as; catchment area, perimeter, mean slope, drainage density etc. For example, the mean slope, drainage area, perimeter and drainage density were extracted from ASTER DEM data available at USGS website. Land-use, and land-cover, soil properties, etc. were prepared using geographical information system (ArcGIS). The drainage area reflects the volume of runoff that can be generated from rainfall, and hence, it serves as one of the most important input data for the hydrologic model. The watershed slope is an important factor in the runoff momentum of a particular flood magnitude. It reflects the rate of change of elevation with respect to distance along the principal flow path; while the drainage density is the total length of the stream in the catchment normalized by the catchment area. The catchment slope plays an important role on the drainage density [15]. Higher per cent of the streams are found to flow towards north-east and east directions.

Temperature time series data of the study area were collected from Malaysian Meteorological Department (MMD) and were used to compute the reference evapotranspiration (ET_0) at each catchment using Hargreaves method [15]. This can be used to determine the reference evapotranspiration of the concerned area as suggested by [Droogers and Allen [16], Jabloun and Sahli [18], Tabari *et al*. [19]]. However, the annual variations in ET_0 across the study area, were found not to be significant, and were therefore excluded from the analysis.

2.6 Soil Characteristics, Land Use/Land Cover

The soil properties of the catchments were derived from the available generalized soil map of Peninsular Malaysia, 1970. There are seventeen different soil groups as detailed on the map. Appraisal of the study "keys to the identification of the Malaysian soils using parent materials" reported by Paramananthan [20] gave better insight to the soil groups. Two major properties of interest which are likely to control the hydrological characteristics of the catchments were obtained from the map. These are: soil texture and soil depths [5]. The soil texture groups were categorized into three major groups as shown in Table 1 below:

Table 1 Soil groups

Top soil texture	Soil depth	Remark
Coarse	<50cm	Shallow
Medium	50 - 100cm	Moderately deep
Fine	>100cm	Deep

The Land use and land cover are highly dynamic and rarely in a stable equilibrium [21] and has been reported to have great effects on the runoff generation in a catchment. The land use map of Peninsular was as well studied. This provides useful information about the land use practice in the catchment, because of its great influence on the catchment hydrologic response. It is a common practice in recent years to convert large tracks of forest to agricultural land with consequence of higher flood frequency and drought [22]. The soil characteristics and land-use of the area formed the rationale for curve number derivation for the catchments according to US Soil Conservation Service (SCS). In SCS, these are referred to the soil groups, hydrologic conditions and antecedent moisture condition [16, 23-25]. The hydrologic condition of the catchments defines the vegetation or ground cover and are expressed as poor, fair and good represented as < 50%, 50 - 75% and > 75%. In deriving the curve number, antecedent soil moisture condition II was selected for the study area depicting average condition.

Table 2 Explanatory variables used in the multiple regression analysis

Variable	Definition	Unit	Minimum value	Median value	Maximum value
		Climatic characteristics			
Rain	Median annual rainfall	mm/year	2586	2735	3369
		Geomorphologic characteristics			
Area	Drainage area	km^2	18.40	494.03	2788.00
Peri	Periphery	km	19.76	136.23	265.00
Strm	Longest stream	km	4.10	33.80	223.63
Slop	Mean slope	%	0.39	1.16	2.41
Drai	Drainage density	km^{-1}	0.36	0.38	0.41
Curv	Curve number		75	78.5	89
		Soil characteristics			
Depth	Soil depth	cm	25	75	150

3.0 RESULTS AND DISCUSSION

The explanatory variables for streamflow predictions in the east coast of Peninsular Malaysia are presented in Table 3 below. In the table, there are two rows for each flow metrics, the upper rows presents the coefficients of the multiple linear regression equations, while the lower row shows the logarithmic transformed data incremented by unity. The two sets of equations are presented for comparison of the models, by comparing the performance indicators as shown in columns 8 and 9 of Table 3 respectively. The illustrative equation of the multiple linear regression equations are as presented in (Eq. (3) and (4)).

Table 3 Parameters of the multiple regression models; Column 1: Flow metrics for which predictions are made. Column 2 - 7: Coefficients of the explanatory variables. Column 8 and 9: Performances of the models (%)

Q (m^3s^{-1})	β_0	Explanatory Variables (β_i, i >0)					Performance		
		Rain	Area	Peri	Slop	Curv	NSE	RMSNE	R^2_{adj}
Max	-2112.9	13.38	0.875		169.36	19.65	93.3	0.75	96.0
	-6.2	0.71		1.483	0.40	0.18	98.8	1.37	89.5
0.05	-778.6	1.97	0.247		50.20	8.20	97.5	0.00	88.7
	-24.6	0.70		1.585	0.62	3.17	96.6	1.94	93.1
0.10	-592.9	1.40	0.174		31.79	6.33	88.3	0.00	81.6
	-24.6	0.70		1.585	0.62	3.17	95.6	7.69	90.6
0.25	-374.2	1.12	0.103		18.32	3.95	86.7	0.00	79.1
	-28.9	0.85		1.542	0.67	3.71	93.8	3.97	86.2

Table 3 Continued

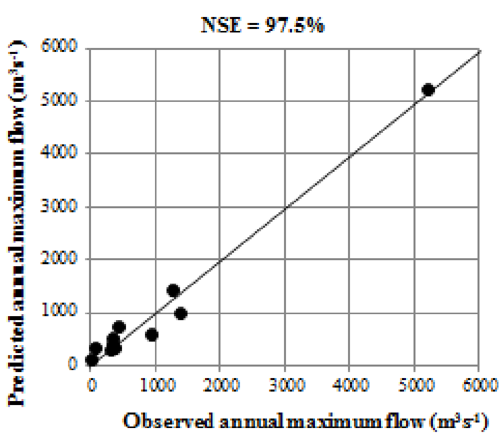
0.50	-270.8	0.89	0.072		12.46	2.85	84.1	0.00	75.0
	-27.0	0.91		1.48	0.58	3.30	89.0	2.29	81.1
0.75	-194.8	0.66	0.054		8.31	2.03	82.7	0.00	72.8
	17.9	0.99		1.37	0.43	1.65	96.1	5.95	72.0
0.90	-150.3	0.53	0.044		6.13	1.55	83.2	0.00	73.6
	-12.6	1.01		1.366	0.31	0.60	97.3	6.04	67.4
0.95	-135.3	0.53	0.04		5.46	1.38	82.7	2.38	75.1
	-7.6	1.07		1.344	0.28	0.35	97.6	5.80	64.0
Min	-77.8	0.60	0.026		3.25	0.67	94.6	0.00	91.5
	-1.5	1.3		1.184	0.41	0.20	91.6	4.98	48.2
Mean	-360.6	0.99	0.104		20.27	3.80	89.2	0.69	82.5
	-26.9	0.78		1.546	0.64	3.41	95.6	3.20	88.9

$$Q_{max} = -2112.957 + 13.382Rain + 0.875Area + 169.36Slop + 19.656Curv \quad (3)$$

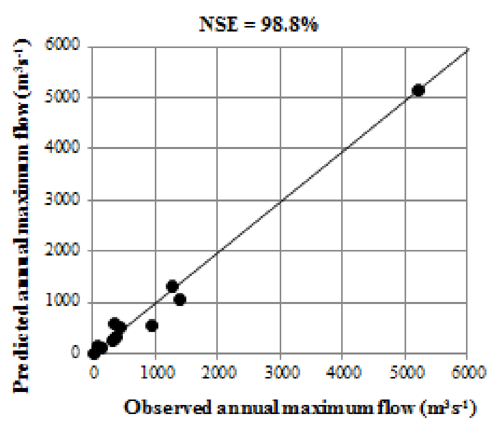
The logarithmic transformed dataset prior to multiple linear regression, results in the exponential equation as given below:

$$Q_{0.50} = \exp\{[-27.033 + 0.917Rain + 1.480Peri + 0.588Slop + 3.305Curv] - 1\} \quad (4)$$

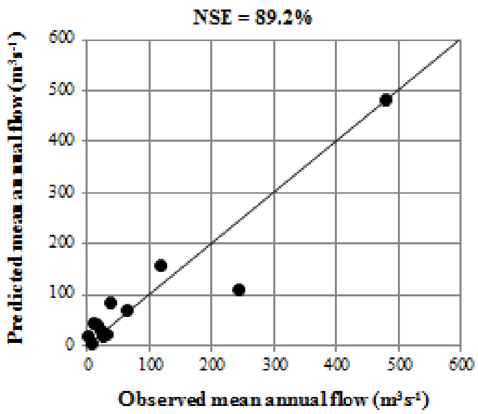
The performance indicators in Table 3 above shows that logarithmic transformed data, prior to multiple linear regression and replacing the area with periphery generally have better perform than those that did not undergo transformation except at low flow metrics where the later proved to perform better. The performance of the two models can be said to record high performances and hence can be confidently applied to any catchment within the study area, once the input variables have been established. Figure 2 presents excellent comparison between observed and predicted flows for the models.



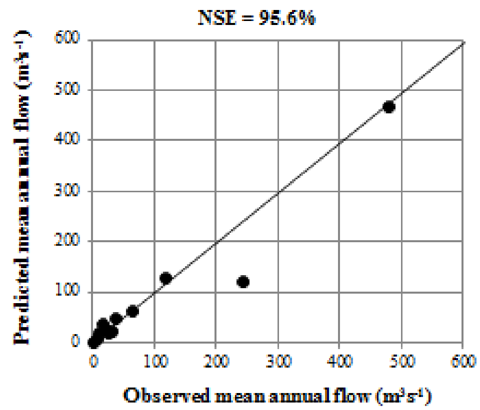
a) Measured vs pred. Q_{max} (Linear Model)



(b) Measured vs pred. Q_{max} (Log model)



c) Measured vs pred. Q_{mean} (Linear Model)



(d) Measured vs pred. Q_{mean} (Log model)

Figure 2 Comparison of measured and predicted flows for the linear and log models (a-b): annual maximum and (c-d) mean annual

4.0 CONCLUSION

A set of multivariate linear and exponential models have been developed and validated in the present study, to predict streamflow in ungauged catchments in the east coast of peninsular Malaysia, which is considered the most vulnerable region of Malaysia to hydrological disasters. The models are specifically developed for the catchments of east coast and may not perform well in other catchments other than these, considering the fact that the input parameters were developed from the established catchment characteristics of the area under study. The model is found to predict streamflow in ungauged catchments with reasonable accuracy. The R^2_{adj} for the best models are of the order of 82.5 to 96%. The best of the model may be chosen to suit a particular flow metric applicable. However, the logarithmic transformed data with resulting exponential function suggested a higher predictive power except for very low flow where the multiple linear equations proved to be better. It is therefore expected that the methodology used in the present study can be used in other tropical regions for predictions of streamflow in ungauged river basins.

Acknowledgement

We are grateful to Ministry of Higher Education (MOHE)-Malaysia and Universiti Teknologi Malaysia (UTM) for financial support of this research through FRGS research project (Vote No. 4F541).

References

[1] Yaşar, M. and Baykan, N. O. 2013. Prediction of Flow Duration Curves for Ungauged Basins with Quasi-Newton Method. *Journal of Water Resource and Protection*. 5: 97-110.

[2] Jones, D. A. and Kay, A. L. 2007. Uncertainty Analysis For Estimating Flood Frequencies For Ungauged Catchments Using Rainfall-Runoff Models. *Advances in Water Resources*. 30(5): 1190-1204.

[3] Goswami, M., O'Connor, K. M. and Bhattarai, K. P. 2007. Development Of Regionalisation Procedures Using A Multi-Model Approach For Flow Simulation In An Ungauged Catchment. *Journal of Hydrology*. 333(2-4): 517-531.

[4] Salinas, J. L., Laaha, G., Rogger, M., Parajka, J., Viglione, A., Sivapalan, M. and Blöschl, G. 2013. Comparative Assessment Of Predictions In Ungauged Basins. Part 2: Flood And Low Flow Studies. *Hydrology and Earth System Sciences*. 17(7): 2637-2652.

[5] Lacombe, G., Douangsavanh, S., Vogel, R. M., McCartney, M., Chemin, Y., Rebelo, L. M. and Sotouke, T. 2014. Multivariate Power-Law Models For Streamflow Prediction In The Mekong Basin. *Journal of Hydrology*. 2: 35-48.

[6] Kovacs, A., Honti, M., Zessner, M., Eder, A., Clement, A. and Bloschl, G. 2012. Identification Of Phosphorus Emission Hotspots In Agricultural Catchments. *Sci Total Environ*. 433: 74-88.

[7] Awadalla, S. and Noor, I. M. 1991. Induced Climate Change On Surface Runoff In Kelantan Malaysia. *International Journal of Water Resources Development*. 7(1): 53-59.

[8] Basarudin, Z., Adnan, N. A., Latif, A. R. A., Tahir, W. and Syafiqah, N. 2014. Event-based rainfall-runoff modelling of the Kelantan River Basin. *IOP Conference Series: Earth and Environmental Science*. 18: 012084.

[9] Pradhan, B. 2009. Flood Susceptible Mapping And Risk Area Delineation Using Logistic Regression, GIS And Remote Sensing. *Journal of Spatial Hydrology*. 9: 1-18.

[10] Vogel, R. M. and Fennessey, N. M. 1994. Flow Duration Curve: New Interpretation And Confidence Intervals. 485-503.

[11] Adnan, N. A., Atkinson, P. M., Yusoff, Z. M. and Rasam, A. R. A. 2014. Climate Variability and Anthropogenic Impacts on a Semi-Distributed Monsoon Catchment Runoff Simulation.

[12] Abdullahi, M. G., Gasim, M. B. and Juahir, H. 2015. Determination of Groundwater Level Based on Rainfall Distribution: Using Integrated Modeling Techniques in Terengganu, Malaysia. *Journal of Geology & Geosciences*. 4(1): 1-4.

[13] Pour, S., Harun, S. and Shahid, S. 2014. Genetic Programming for the Downscaling of Extreme Rainfall Events on the East Coast of Peninsular Malaysia *Atmosphere*. 5(4): 914-936.

[14] Mayowa, O. O., Pour, S. H., Shahid, S., Mohsenipour, M., Harun, S. B., Heryansyah, A. and Ismail, T. 2015. Trends In Rainfall And Rainfall-Related Extremes In The East Coast Of Peninsular Malaysia. *Journal of Agriculture and Rural Development in the Tropics and Subtropics.* 124(8): 1609-1622.

[15] Tukimat, N. N. A., Harun, S. and Shahid, S. 2012. Comparison Of Different Methods In Estimating Potential Evapotranspiration At Muda Irrigation Scheme Of Malaysia. *Journal of Agriculture and Rural Development in the Tropics and Subtropics.* 113(1).

[16] McCuen, R. H. 1989. Hydrologic Analysis and Design. B. Stenquist, ed. A Simon and Schuster/Aviacom upper saddle river, New Jersey, 07458. 833-853.

[17] Droogers, P. and Allen, R. G. 2002. Estimating Reference Evapotranspiration Under Inaccurate Data Conditions. *Irrigation and Drainage Systems.* 16: 33-45.

[18] Jabloun, M. and Sahli, A. 2008. Evaluation Of FAO-56 Methodology For Estimating Reference Evapotranspiration Using Limited Data Available To Tunisia Evapotranspiration Using Limited Climatic Data Application To Tunisia. *Agricultural Water Management.* 95: 707-715.

[19] Tabari, H., Grismer, M. E. and Trajkovic, S. 2011. Comparative Analysis Of 31 Reference Evapotranspiration Methods Under Humid Conditions. *Irrigation Science.* 31(2): 107-117.

[20] Paramananthan, S. 2012. Keys To The Identification Of The Malaysian Soils Using Parent Materials. P. A. S. S. M. S. BHD, ed. 1-37.

[21] Niehoff, D., Fritsch, U. and Bronstert, A. 2002. Land-Use Impacts On Storm-Runoff Generation: Scenarios Of Land-Use Change And Simulation Of Hydrological Response In A Meso-Scale Catchment In SW-Germany. *Journal of Hydrology.* 267: 80-93.

[22] Valentin, C., Agus, F., Alamban, R., Boosaner, A., Bricquet, J. P., Chaplot, V., de Guzman, T., de Rouw, A., Janeau, J. L., Orange, D., Phachomphonh, K., Do Duy, P., Podwojewski, P., Ribolzi, O., Silvera, N., Subagyono, K., Thiébaux, J. P., Tran Duc, T. and Vadari, T. 2008. Runoff And Sediment Losses From 27 Upland Catchments In Southeast Asia: Impact Of Rapid Land Use Changes And Conservation Practices. *Agriculture, Ecosystems & Environment.* 128(4): 225-238.

[23] Kabiri, R., Chan, A. and Bai, R. 2013. Comparison of SCS and Green-Ampt Methods in Surface Runoff-Flooding Simulation for Klang Watershed in Malaysia. *Open Journal of Modern Hydrology.* 3(3): 102-114.

[24] Schneider, L. E. and McCuen, R. H. 2005. Statistical Guidelines for Curve Number Generation. *Journal of Irrigation and Drainage Engineering.* 131: 282-290.

[25] Soulis, K. X. and Valiantzas, J. D. 2012). SCS-CN Parameter Determination Using Rainfall-Runoff Data In Heterogeneous Watersheds: The Two-CN System Approach. *Journal of Hydrology and Earth System Sciences.* 16(3): 1001-1015.