

## Prediction in Ungauged River Basin in the West Coast of Peninsular Malaysia using Linear Regression Model

Tarmizi Ismail<sup>1\*</sup>, Sobri bin Harun<sup>1</sup>, Shamsuddin Shahid<sup>1</sup>, Zaitul Marlizawati Zainudin<sup>2</sup>,  
Mahiuddin Alamgir<sup>1</sup>, and Usman Ullah Sheikh<sup>3</sup>

<sup>1</sup>Department of Hydraulics & Hydrology, Faculty of Civil Engineering, UTM

<sup>2</sup>Department of Mathematics, Faculty of Science, UTM

<sup>3</sup>Faculty of Electrical Engineering, UTM

### Abstract

A linear multiple regression based regionalization method has been proposed in this study to simulate streamflow in ungauged catchment in the east coast of peninsular Malaysia. Identification of unit Hydrographs And Component flows from Rainfall, Evapotranspiration and Streamflow (IHACRES) rainfall-runoff model was used to develop the relationship between model parameters and physical catchment descriptors of eight gauged catchments distributed over the west coast of peninsular Malaysia. The IHACRES model was calibrated and validated individually for each catchment with the available data for the time periods varying between three to sixteen years. The Nash-Sutcliffe efficiency index was used as criteria to evaluate the model performance. As the relationships between the physical catchment descriptors and hydrologic response characteristics are not necessarily linear, different forms of transformations were performed in order to find the most appropriate relationship. Finally, the obtained regression equations were used for simulating stream discharge in Sg Layang catchment located in the south of peninsular Malaysia. The result of the regional model was compared with observed monthly stream flow data of the catchment to assess the ability of regional model. The obtained results revealed that the regional model was able to replicate the historical monthly average flow. However, the relationship between the catchment area and hydrologic response characteristics were not fully understood by regional model which emphasize the need of consideration of other dominant catchment factors for prediction in ungauged basin.

**Key words:** Regionalization, prediction in ungauged basin (PUB), IHACRES, linear regression, peninsular Malaysia

---

\* Corresponding Author. E-mail: [tarmiziismail@utm.my](mailto:tarmiziismail@utm.my)

## 1. Introduction

Predictions of runoff are key requirement 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 (Kovacs et al., 2012). However, runoff data are not available in many catchments of interest. Therefore, it is often required to predict runoff of ungauged catchments from other information within that catchment or from other catchments (Parajka et al., 2013). Many methods have been developed and applied in different parts of the world for this purpose (Kovacs et al., 2012; Parajka et al., 2013; Wang et al., 2014). However, prediction in ungauged basins remains a major challenge in hydrology.

It is much more challenging in tropical region where most of the catchments are ungauged. However, the need for improved knowledge of flow variability in such region become very urgent, especially in the context of changing hydrological processes due to changes in landuse 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 in recent years. However, the major challenge in hydrological studies in the area is the unavailability of reliable and long-term streamflow data.

A possible solution to this challenge is the regionalization of a hydrologic model. A regionalization method is applied to simulate streamflow from an ungauged catchment as an alternative approach. Several regionalization methods have been proposed in the literature for the development of catchment model parameters (Wang et al., 2014). Among those the most widely used techniques are regressions between the model parameters and physical catchment descriptors. Typically, linear multiple regressions are the most widely used as regionalization technique where each model parameter is estimated independently from the others (Post and Jakeman, 1996).

Regionalization approach has been previously reported in various studies. Kokkonen *et al.*, (2003) used 13 catchments with areas ranging below 16km<sup>2</sup> to developed regionalization model of the Coweeta watershed. Paola *et al.*, (2006) extended the study for more and bigger watersheds ranging from 2.88 to 9500km<sup>2</sup>. In a comparative study of 320 catchments in Austria, Parajka *et al.*, (2005) found similarity and kriging approaches performs best among the four methods used for transposing model parameters to the ungauged catchments.

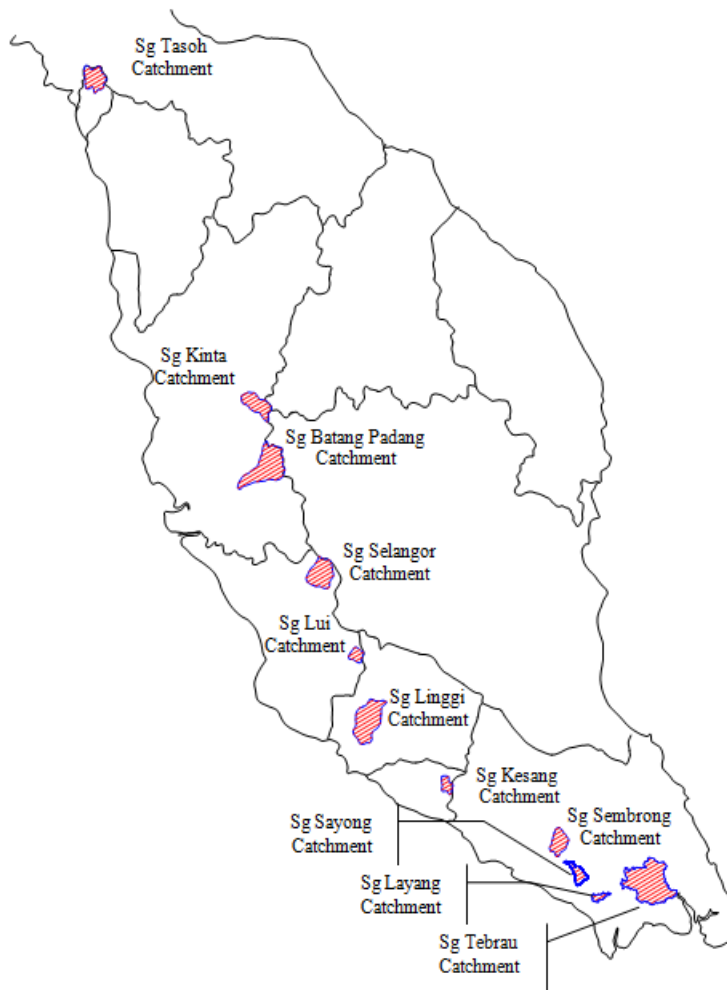
A linear multiple regression based regionalization method has been proposed in this study to simulate streamflow in ungauged catchment in the east coast of peninsular Malaysia. Identification of unit Hydrographs And Component flows from Rainfall, Evapotranspiration and Streamflow (IHACRES) rainfall-runoff model was used to develop the relationship between model parameters and physical catchment descriptors of eight gauged catchments distributed over the west coast of peninsular Malaysia. IHACRES has been used to address regionalization issues that require methods for estimating the model parameters from independent means rather than direct correlation from rainfall-discharge time series (Post and Jakeman, 1996; Sefton and Howarth, 1998; Kokkonen *et al.*, 2013). The parametric efficiency of the model also provides an advantage to relate its parameters to landscape attributes (Croke *et al.*, 2005) and is therefore potentially suited for streamflow prediction in ungauged catchment through regionalization method.

## **2. Methodology**

The selection of the catchments for the study was constrained by data series lengths. Therefore, only eight catchments on the west coast of Peninsula Malaysia that experienced a similar climatic influence were selected. The relative location of the selected catchments in west coast of peninsular Malaysia is presented in Figure 1. The catchment areas range in size from 185.5km<sup>2</sup> to 1227.2km<sup>2</sup>. Rainfall time series data for eight selected catchments that behave similar climatic conditions in the west coast of Peninsula Malaysia are compiled and analyzed. Description of the methods used in the present study is given below in details.

### *2.1 Simulation of streamflow using IHACRES*

IHACRES was used in this study to simulate flows. The IHACRES is a lumped conceptual rainfall runoff model initially developed for application in temperate catchments (Jakeman *et al.*, 1990). It has undergone several considerable enhancements to suit various applications (Jakeman *et al.*, 1990; Jakeman and Hornberger, 1993; Schreider *et al.*, 1997; Evans and Jakeman, 1998). The IHACRES Classic Plus software (Croke *et al.*, 2006) used in this study is the modification of the IHACRES\_PC software (Littlewood *et al.*, 1997) which has been developed to increase the functionality of the software with the inclusion to ephemeral streams (Ye *et al.*, 1997; Schreider *et al.*, 1996; Croke *et al.*, 2006). Details of IHACRES can be found in Jakeman *et al.* (1990) and Jakeman and Hornberger (1993).



**Figure 1:** Location of selected catchments used in the present study

## 2.2 Linear regionalization

IHACRES was used to develop the relationship for gaged catchments between model parameters and physical catchment descriptors. Catchment area and aspect ratio are the two descriptors that were used to derive the relationship. The hydro-meteorological data for the selected catchments varies from three to 16 years period where the streamflow data often becomes the limiting factor. These observed data from each catchment was thoroughly screened to represent a good quality concurrent period with no missing data. The data are divided into almost two equal portions for calibration and validation. Detail catchment characteristics and the hydro-meteorological data for all catchments used in calibration and validation are given in Table 1. The model was calibrated and validated individually for every catchment. The Nash and Sutcliffe efficiency index  $R^2$  was used as criteria in the selection of grid analysis to evaluate the calibration.

**Table 1:** Calibration and validation periods

<b>Catchment</b>	<b>Calibration</b>	<b>Validation</b>	<b>Data Period</b>
Kesang	Jan 26, 1978 - Jan 12, 1986	Jan 1, 1977 - Jan 25, 1978 and Jan 13, 1986 - Dec 31, 1992	1977 - 1992
Linggi	Jan 1, 1985 - Jul 4, 1986	Jul 5, 1986 - Dec 31, 1987	1985 - 1987
Lui	Nov 9, 1977 - July 5, 1983	Jan 1, 1977 - Nov 8, 1977 and July 6, 1983 to Dec 31, 1985	1977 - 1985
Sayong	Dec 30, 1989 - Nov 12, 1991	Jan 1, 1989 - Dec 29, 1989 and Nov 13, 1991 - Dec 31, 1992	1989 - 1992
Selangor	Apr 21, 1982 - Apr 19, 1987	Jan 1, 1980 - Apr 20, 1982 and Apr 20, 1987 - Dec 31, 1989	1980 - 1989
Sembrong	Oct 30, 1983 - Mar 19, 1985	Jan 1, 1983 - Oct 29, 1983 and Mar 20, 1985 - Mar 31, 1985	1983 - 1985
Tasoh	Jan 1, 1980 - Jan 20, 1984	Jan 21, 1984 - Dec 31, 1987	1980 - 1987
Tebrau	Jan 1, 1981 - Dec 31, 1982	Jan 1, 1983 - Dec 31, 1983	1981 - 1983
Layang	-	Jan 1, 2004 - Dec 31, 2008	2004 - 2008

After the calibration, the relationships between model parameters and physical descriptor of catchments were investigated using linear regressions. The variables considered are the catchment area and the aspect ratio that reveal the most significant impacts on catchment response to rainfall. The model parameters were derived from these relationships on the basis of the area of the study catchment. These parameters were set in the linear module and identically specified the start and end parameter values in the non-linear module to generate the streamflow using simple morphological information and records of rainfall and temperature.

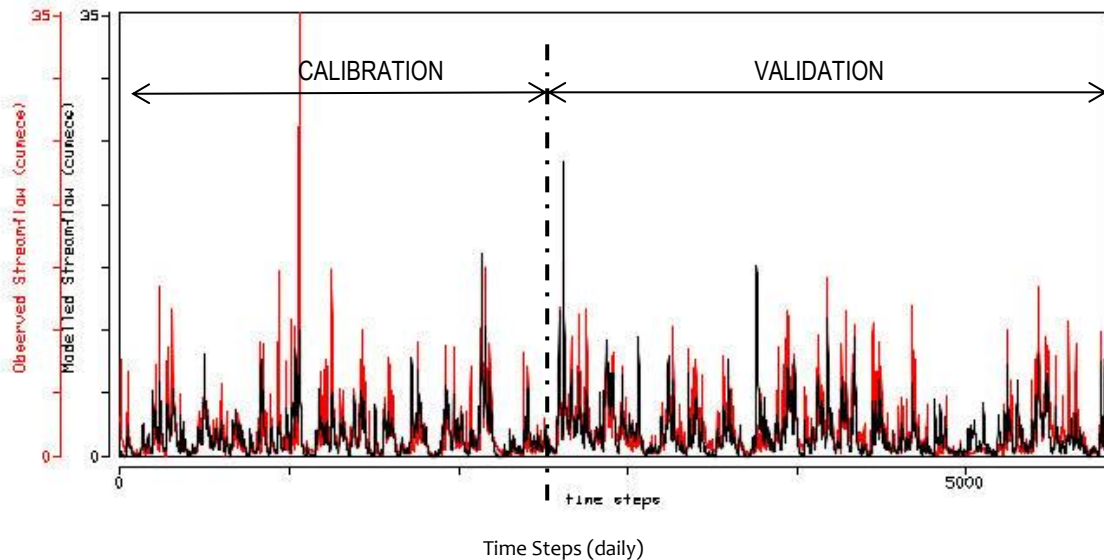
The results of the regional model were compared to the model obtained from the calibration of gauged catchment for selection in simulation of daily flow under future climate scenarios. Four different statistical performance criteria that assess the agreement between the observed and estimated flow in calibrations and validations were computed, namely, the Nash

and Nash-Sutcliffe Efficiency ( $R^2$ ), the mean error or bias ( $ME$ ), the dimensionless mean error or relative bias ( $ME/\bar{Q}_o$ ), and the root mean square error ( $RMSE$ ) have been considered.

### 3. Results and Discussion

#### 3.1 Calibration of IHACRES models

IHACRES models were calibrated and validated for each of the eight catchments under study. Calibration and validation of IHACRES in Sg Kesang catchment is shown in Figure 2. The figures shows good match between observed and simulated streamflow during both model calibration and validation. The statistical performance of IHACRES model during calibration and validation in different catchments are given in Table 2. The values are found above satisfactory in all catchments during both model calibration and validation.



**Figure 2:** Results of simulation for Sg Kesang Catchment

#### 3.2 Regionalization

For the gauged catchments, the correlation between IHACRES model parameters and catchment area was determined by regression. The selected gauged catchments were considered to have similar hydrologic response characteristics. Therefore, this correlation is assumed to reveal a significant relationship on catchment response to rainfall.

Since all the eight catchments under study were gauged, the simulated flows were compared to the observed flows in order to assess the accuracy of the predictions. Table 3 shows the statistical performance of regionalized model during calibration and validation.

The values of  $R^2$  were found in the range from 0.55 to 0.78 and 0.44 to 0.83 during model calibration and validation, respectively. As the relationships between the physical catchment descriptors and hydrologic response characteristics are not necessarily linear, other forms of transformations were performed in order to find the most appropriate relationship. The calibrated model parameters that represent the non-linear and linear modules are regressed to derive the best fitted equation. Equations 1 to 6 represent the optimum predictive equations obtained using regionalization approach for each model parameters.

*Parameter for non-linear module*

$$c = -3 \times 10^{-7} A + 0.0012 \quad (1)$$

$$t_w = -0.0113A + 16.585 \quad (2)$$

$$f = 0.0001A + 2.32 \quad (3)$$

*Parameter for linear module*

$$\alpha^{(s)} = -0.1152[\ln(A)] - 0.2368 \quad (4)$$

$$\beta^{(s)} = 8 \times 10^{-5} A + 0.1149 \quad (5)$$

$$\tau^{(s)} = -0.0041A + 11.251 \quad (6)$$

These correlations were used for deriving the model parameters representing the Sg Layang study catchment with an area of  $36.8 \text{ km}^2$ . The obtained results are discussed in next section.

### 3.3 Model Performance Analysis

Model calibration from eight selected catchments has shown that regional relationship between hydrologic response characteristics and catchment area can be used in order to make a prediction of daily streamflow of an ungauged catchment from an appropriate rainfall and temperature time series data. However, the streamflow resulting from these predictions are poorer than the approach obtained from direct model calibration of gauged catchment. This is indicated by the statistical performance criteria as shown in Table 3. It shows that the relationship between the catchment area and hydrologic response characteristics are not fully understood. The fact that reasonable model results were obtained for most of the catchments is probably due to homogeneity of these catchment in terms of their hydrologic response and

meteorological factors. Other dominant factors may be considered to provide a better understanding to improve the results.

**Table 2:** Statistical performance IHACRES during calibration and validation in different catchments used for regionalization

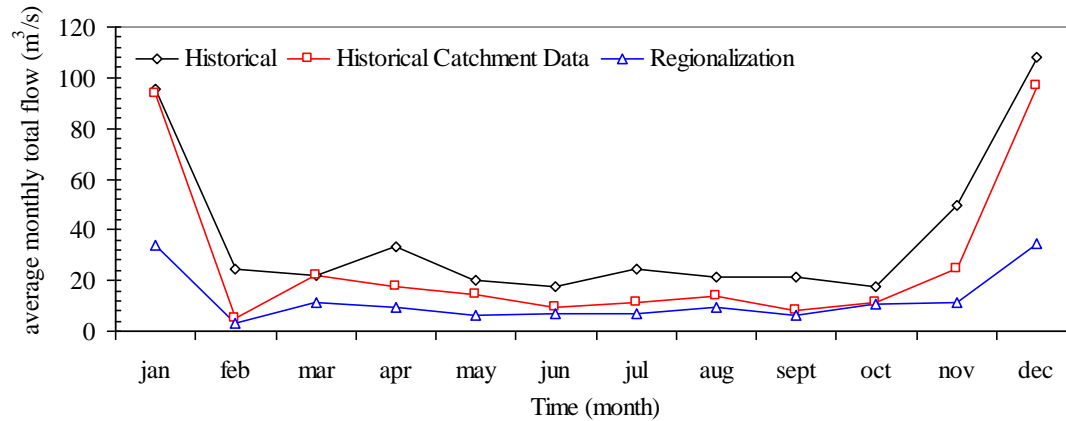
Catchment	IHACRES - Calibration				IHACRES - Validation			
	R <sup>2</sup>	ME	ME/Q <sub>o</sub>	RMSE	R <sup>2</sup>	ME	ME/Q <sub>o</sub>	RMSE
Kesang	0.55	-0.04	-0.02	1.69	0.55	-0.19	-0.02	1.69
Linggi	0.67	-0.29	-0.02	7.39	0.44	1.27	0.13	7.70
Lui	0.53	-0.03	-0.02	0.63	0.53	-0.41	-0.22	0.91
Sayong	0.78	-0.43	-0.03	6.00	0.83	-1.87	-0.12	7.17
Selangor	0.63	-0.27	-0.02	5.46	0.56	0.15	-0.02	6.34
Sembrong	0.76	-0.67	-0.04	6.29	0.57	0.57	-0.16	3.66
Tasoh	0.73	0.03	0.02	1.13	0.70	0.40	0.29	1.26
Tebrau	0.75	-0.48	-0.17	1.93	0.53	-0.09	-0.03	1.86

**Table 3:** Statistical performance of regionalization model during calibration and validation

	Calibration		Validation	
	Historical Catchment Data	Regionalization	Historical Catchment Data	Regionalization
R <sup>2</sup>	0.795	0.565	0.558	0.488
ME	0.395	-0.276	-0.611	-0.483
$\frac{ME}{Q_o}$	0.353	-0.247	-0.415	-0.327
RMSE	1.359	1.978	1.910	2.056

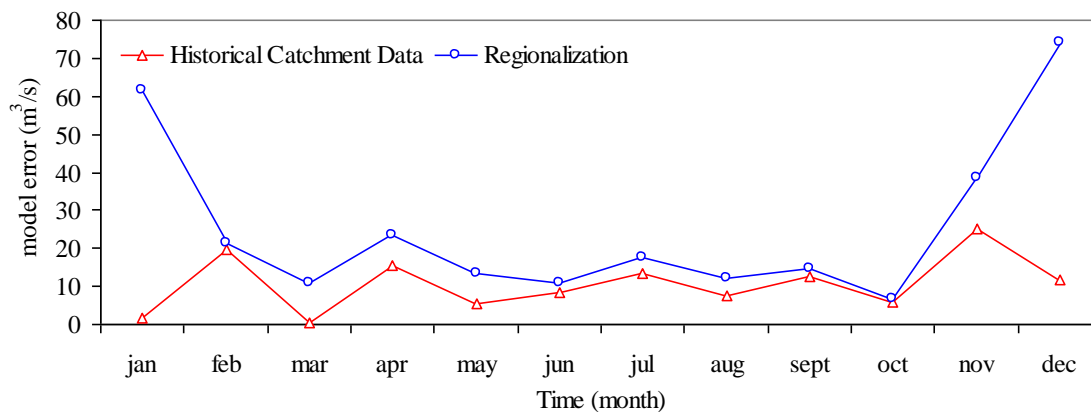
The results of two modeling approaches were also assessed in terms of their ability to simulate monthly flows (Figure 3). Both approaches show similar flow pattern compared to the historical monthly average flow. However, the flow pattern from regionalization approach consistently exhibits more deviation as compared to the flow simulated from model that was directly calibrated and validated from observed flow. Higher average monthly inflow for the months of December to January is due to more rainfall consistently recorded for those months during the period of analysis.





**Figure 3:** Comparison on average monthly total flow simulated by two models at Upper Sg. Layang Catchment

The performance criteria in model selection were also complemented by the assessment of model error from monthly average total flow of seven years historical data. According to graphical comparison of model error in Figure 4, it can be seen that the regionalization model errors are higher for all months compared to error revealed from model that was directly calibrated and validated from historical observed flow.



**Figure 4:** Comparison of errors in two models in simulating streamflow at Upper Sg Layang Catchment

#### 4. Conclusion

IHACRES rainfall-runoff model was used to develop the relationship between model parameters and physical catchment descriptors of eight gauged catchments distributed over the west coast of peninsular Malaysia. The result of the regional model was compared with observed monthly stream flow data of the catchment to assess the ability of regional model.

The obtained results revealed that the regional model was able to replicate the historical monthly average flow. However, the flow pattern from regionalization approach consistently exhibits more deviation, particularly for the months of December and January. It can be concluded that the relationship between the catchment area and hydrologic response characteristics were not fully understood by regional model. The rational performance of regional model was probably due to homogeneity of the gauged catchment in terms of their hydrologic response and meteorological factors. The study emphasizes to consider other dominant catchment factors in order to obtain a better regional model for prediction in ungauged basin.

### **Acknowledgments**

We are grateful to the Ministry of Higher Education (MOHE) Malaysia and Universiti Teknologi Malaysia (UTM) for financial support of this research through the FRGS research project (Vote No. R.J130000.7822.4F683).

### **References**

- Croke, B. F. W., Andrews, F., Jakeman, A. J., Cuddy, S. and Luddy, A. (2005). Redesign of the IHACRES rainfall-runoff model. Engineers Australia. 29th Hydrology and Water Resources Symposium 21-23 February 2005, Canberra
- Evans, J.P., and Jakeman, A.J. (1998). Development of a simple, catchment-scale, rainfall evapotranspiration-runoff model. *Environmental Modelling & Software*. 13, 385–393.
- Jakeman AJ, Hornberger GM (1993) How much complexity is warranted in a rainfall-runoff model?. *Water Resources Research*, vol 29, pp 2637–2649
- Jakeman AJ, Littlewood IG, Whitehead PG (1990) Computation of the instantaneous unit hydrograph and identifiable component flows with application to two small upland catchments. *Journal of Hydrology*, vol 117, pp 275-300
- Kokkonen T, Jakeman AJ, Young PC, Koivusalo HJ (2003) Predicting daily flows in ungauged catchments: model regionalization from catchment descriptors at the Coweeta Hydrologic Laboratory, North Carolina. *Hydrological Processes*, vol 17, pp 2219-2238
- Kovacs A, Honti M, Eder A, Zessner M, Clement A and Blöschl G (2012). Identification of phosphorus emission hotspots in agricultural catchments, *Sci. Total Environ.*, 433: 74–88.

- Littlewood IG, Down K, Parker JR, Post DA (1997) IHACRES – Catchment-scale rainfall-streamflow modelling (PC version) Version 1.0 – April 1997. The Australian National University, Institute of Hydrology and Centre for Ecology and Hydrology
- Parajka J, Viglione A, Rogger M, Salinas JL, Sivapalan M, Bloschl G (2013). Comparative assessment of predictions in ungauged basins – Part 1: Runoff-hydrograph studies. *Hydrol. Earth Syst. Sci.*, 17, 1783–1795
- Schreider, S.Y., Smith, D.I., and Jakeman, A.J. (2002). Climate change impacts on urban flooding. *Climate Change*. vol 47, pp 1-23
- Sefton CEM, Howarth SM (1998) Relationships between dynamic response characteristics and physical descriptors of catchments in England and Wales. *Journal of Hydrology*, vol 211, pp 1-16
- Wang, G.Q, Zhang JY, Jin JL, Liu YL, He RM., Bao ZX, Li Y (2014). Regional calibration of a water balance model for estimating stream flow in ungauged areas of the yellow river basin. *Quaternary International*, 336, 65-72.
- Ye W, Bates BC, Viney NR, Sivapalan M, Jakeman AJ (1997) Performance of conceptual rainfall runoff models in low-yielding ephemeral catchments. *Water Resources Research*, vol 33, pp 153-16