Integration of statistical and spatial methods for distributing precipitation in tropical areas

Mandana Abedini, Md Azlin Md Said and Fauziah Ahmad

ABSTRACT

The high spatial resolution of precipitation distribution is a major concern for experts in environmental research and planning. This paper establishes a combination of multivariate regression algorithm and spatial analysis to predict distribution of precipitation, considering the four topographical factors of altitude, slope, aspect and location. Annual average and seasonal rainfall data were collected in nine rain gauges in Ulu Kinta Catchment in East Malaysia from 1974 to 2010. To examine records and fill gaps from long-term rain gauges, homogeneity analysis was performed using the double-mass curve method. Estimated missing rainfall data were also tested using index gauges from network rainfall stations. Multivariate regression analysis was conducted to propose an empirical equation for the study area. Topographical factors were considered from a 90 m resolution digital elevation model. The multivariate regression model was found to clarify 74% of spatial variability of precipitation on annual average and 78% during wet season. However, the correlation coefficient for the dry season decreased sharply to 63%. By using the kriging interpolation method, the estimated annual average improved to 78.4%; the average improved to 65.2 and 80.3% in the dry and wet seasons, respectively. This confirms the efficiency and significance of the model and its potential for use in other tropical catchments.

Key words | double-mass curve, geostatistical analysis, kriging, multivariate regression, precipitation, spatial distribution

Mandana Abedini
Md Azlin Md Said (corresponding author)
Fauziah Ahmad
School of Civil Engineering,
Universiti Sains Malaysia,
Seri Ampangan,
Malaysia
E-mail: *azlin@eng.usm.my*

INTRODUCTION

Determining the spatial distribution of rainfall is a prerequisite in scientific disciplines such as environmental research planning, risk and disaster analysis, water resources management and agricultural modelling. Appropriate estimation and prediction of rainfall distribution requires a dense network of instruments and large installations at large operational cost (Goovaerts 2000). Accordingly, optimising estimation and prediction of precipitation distribution is an important concern.

Many studies have been conducted on spatial distribution analysis using different interpolation methods. These include traditional approaches such as Thiessen polygon, inverse distance weighting (IDW), statistical analysis and isohyets to estimate point rainfall or station average (Shepard 1968; Handbook 1996; Arnold *et al.* 1998; Mair & Fares 2011). Geostatistical approaches, on the other hand, are based on the theory of regionalised variables and provide statistical correlation on data observation and data processing (Goovaerts 2000). These approaches represent spatial correlation using synchronous observations from surrounding stations to predict value at the unsampled location. Teegavarapu & Chandramouli (2005) introduced several conceptual improvements to IDW. They proposed six different spatial interpolation techniques that improve estimation of prediction. Tabios III & Salas (1985) and Phillips *et al.* (1992) showed that the geostatistical prediction technique, like kriging group analysis, provides better estimation of rainfall than conventional methods. Moreover, Goovaerts (2000) attempted to predict rainfall spatial distribution using two different approaches. The first involved the univariate interpolation models (Thiessen polygon, inverse

doi: 10.2166/nh.2012.159

square distance and ordinary kriging or OK) without considering elevation; the second was the multivariate geostatistical algorithm incorporating the digital elevation model (DEM). The multivariate geostatistical algorithms yielded a lower error using these methods.

Preliminary work that incorporated elevation to precipitation was undertaken by Spreen (1947). He identified elevation as a major contributing factor for predicting rainfall. This theory was analysed in detail by Smith (1979). The relation between precipitation and elevation was further investigated by different researchers (Phillips *et al.* 1992; Daly *et al.* 1994; Goovaerts 2000). All designed a model that had a small estimated predictive error. In addition, some discussed the relation between precipitation and other topographic variables such as altitude, latitude and slope by using regression (Basist *et al.* 1994; Daly *et al.* 1994; MarquInez *et al.* 2003; Guan *et al.* 2005; Ninyerola *et al.* 2007; Ranhao *et al.* 2008). Local climate conditions are observed to be strongly influenced by topography. However, the regression equations for tropical and extra tropical regions indicated distinctive slope coefficient in different regions (MarquInez *et al.* 2003). Many published studies discussed distribution patterns or filled in gaps in Malaysia (Desa & Niemczynowicz 1996; Tang *et al.* 1996; Wong *et al.* 2009; Suhaila *et al.* 2010; Varikoden *et al.* 2011). However, no research has been conducted into a comprehensive method incorporating the effects of topographic variables to spatial prediction of rainfall.

Consequently, the present study sought to develop a multivariate regression model in Ulu Kinta Catchment, a tropical area. The model was used to estimate long average rainfall (LAR) in monthly and seasonal (dry season and wet season) periods. The model combined a high-resolution DEM and statistical analysis to predict precipitation, considering the elevation, slope, aspect and location of rain gauges relative to precipitation amount at each reference point.

DATA COLLECTION

Study area

The Ulu Kinta Catchment is located in Perak, Peninsular Malaysia between latitudes 4°29′ and 4°31′ N and longitudes 101°03′ and 101°12′ E (Figure 1). This catchment is a narrow river valley trending northwest and has two right angle bends. The northwest flow path of the river appears offset by a rectangular meander with an area of approximately 30,752 ha. The study area varies from 38 m in the west to 2,160 m in the north and east parts of the hilly area. Figure 2 depicts the profile of the (west–east) elevation trend in the study area.

Dataset collection and pre-processing

Precipitation data used in this study were provided by the Department of Irrigation and Drainage (DID) of the Water Resource Management and Hydrology Division, Malaysia. The initial sample consisted of daily precipitation data collected from January 1974 to December 2010 from 9 operating rain gauges (Table 1). Table 2 summarises the statistical properties of each station based on data series. Almost 70% of the average rainfall occurred from March to May and from September to December. These periods were the wet season; the other months comprised the dry season. Insufficient rainfall data measurement in five gauges was initially caused by failure in the station or by missing recorded data.

Given that precipitation is strongly affected by the quantity measured at different periods, the construction of the models considered time series and seasonal variability (Figure 3). Generally, two dry seasons are in between two wet seasons. Three separate detailed survey processes were therefore conducted for monthly average, dry season and wet season.

Datasets filling-in gaps

Rainfall data were aggregated monthly for homogeneity and to adjust inconsistency and fill in gaps. Double-mass curve analysis was conducted (Larson & Peck 1974; Tang *et al.* 1996; Mair & Fares 2010) during accumulation from 1974 to 1994. Plotting the accumulation of two quantities of precipitation results in a straight line; this represents the constant of proportionality between the two qualities.

R5 was chosen as an index gauge because of extensive history, relative location and homogenous recorded data. Consistency analysis suggested that some gauges enhanced

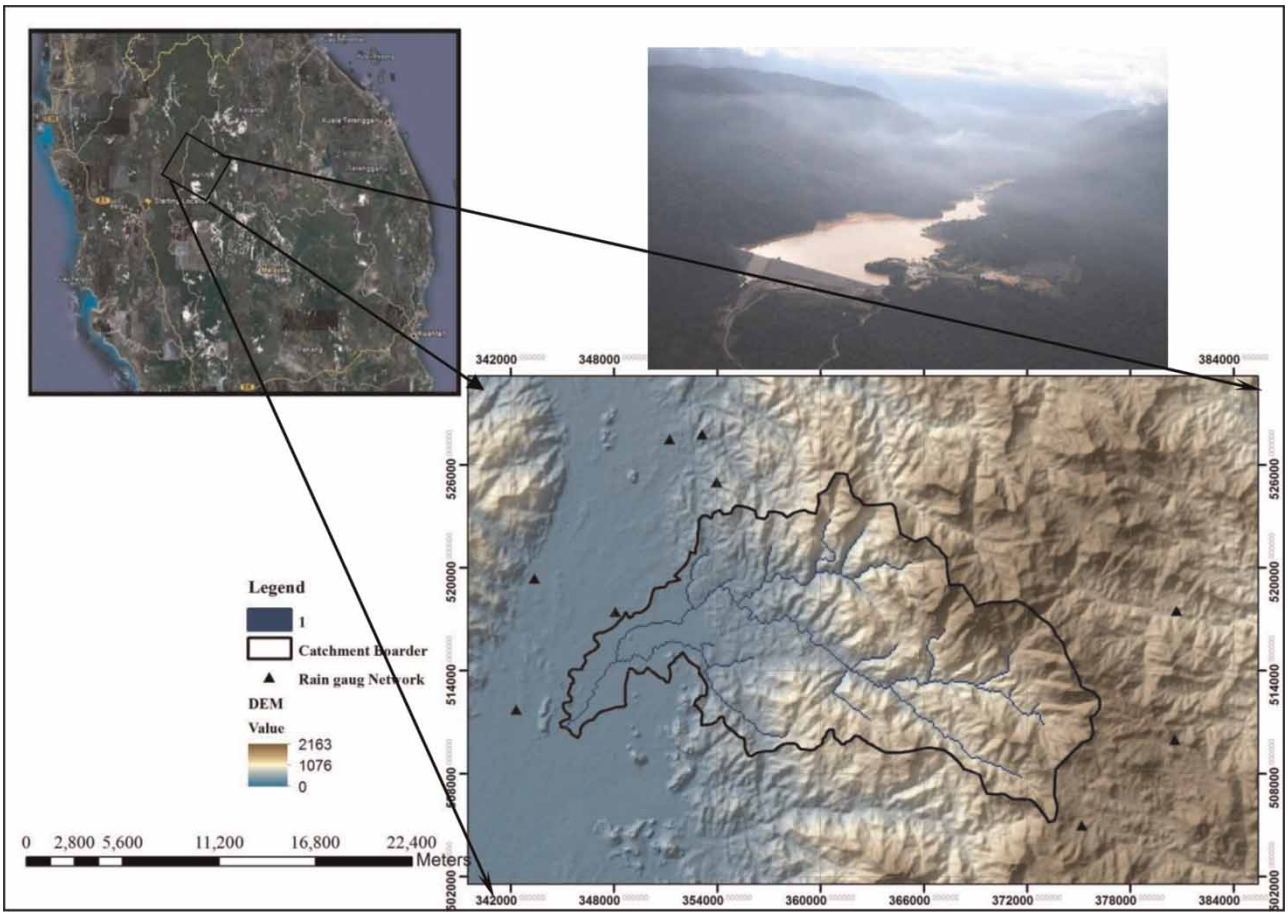

Figure 1 | Location of the Ulu Kinta catchment (Ipoh, Malaysia) and rain gauges network.

regularity in operation time. For instance, R1, R2 and R9 were corrected by one correction factor every month; others such as R3, R6, R7 and R8 represented fluctuation and two different correction factors. Figure 4 illustrates an example of double-mass curve analysis. Table 3 shows the modification factor in the precipitation gauges, so data were estimated for the missing period or extrapolated beyond the existing length of record (Searcy & Hardison 1960; Albert 2004).

To identify homogeneity of data prediction, a standard deviation and variance of the variable in LAR group were compared by examining a scatter plot of predicted versus observed data (Figure 5). Consequently, the LAR data did not indicate differences before and after filling-in and imputing data.

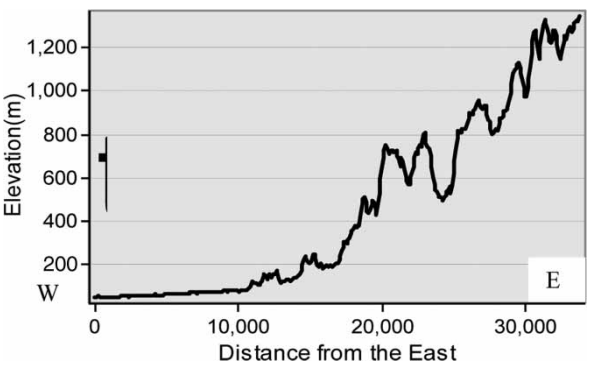

Figure 2 | Profile of elevation trend (west–east) in the study.

METHODS AND TECHNIQUE

Multivariate regression was used to obtain the distribution of precipitation at an ungauged location X and to determine the best model. The efficiency and capability of the model were analysed via spatial analysis.

Table 1 | Rain gauges in the study area

Station ref.	Station no.	Station name	Location Latitude	Longitude	Elevation (m)
R1	4413034	STN KAJICUACA	04°28′25	101°22′40	1,474
R2	4510117	LDG, PINJI	04°31′50	101°04′20	31
R3	4511111	POLITECHNIK UNGKU OMAR, IPOH	04°35′20	101°07′30	44
R4	4514031	LDG, THE BLUE VALLEY	04°35′10	101°25′10	1,410
R5	4514032	LDG, THE SG.PALAS CAM, HIGHLANDS	04°31′00	101°25′00	1,492
R6	4610116	HOSPITAL, IPOH	04°36′10	101°05′20	40
R7	4611001	LDG, KUDA KEB, ULU KINTA	04°40′50	101°10′10	93
R8	4611114	KERJA AIR KINTA, TG, RAMBUTAN	04°39′15	101°10′30	98
R9	4611115	TAMAN BAHAGIA TG, RAMBUTAN	04°40′30	101°09′15	76

Table 2 | Long-term statistics result for the monthly and annual rainfall data (mm) for 1974–2010

Period	Max	Min	Mean	Std dev.	Rainfall distribution
January	467.50	0.00	92.62	22.96	5.71
February	492.00	0.00	105.50	25.14	5.28
March	609.00	0.00	165.90	30.05	7.8
April	614.50	0.00	189.65	26.35	9.9
May	554.70	0.00	194.50	25.92	9.58
June	374.50	0.00	137.50	85.88	6.1
July	383.00	0.00	14.00	63.85	6.4
August	478.90	0.00	138.00	24.39	7.02
September	528.00	0.00	200.00	24.65	9.22
October	766.50	0.00	287.30	39.26	11.58
November	636.50	0.00	249.70	32.10	11.66
December	546.00	0.00	153.15	27.03	9.25

Statistical analysis method

A statistical method was considered to determine how precipitation was affected by topographical factors. To determine topographical variables, ERDAS 9.1 software was used and variables were extracted from the Shuttle Radar Topography Mission (SRTM) map. The DEM had 90 m resolution with a project coordinate system of RSO-Malaysia. For data analysis, IBM SPSS Statistics 19, using a polynomial function, was used.

The best model is the function of correlation matrix for the variables to explain precipitation pattern with a multivariate non-linear regression (Daly *et al.* 1994; Ranhao *et al.* 2008):

$$P = b_0 + b_1X + b_2X^2 + b_3X^3 + b_4Y + b_5Y^2 + b_6Y^3 + b_7\mathrm{DEM} + b_8\mathrm{DEM}^2 + b_9\mathrm{DEM}^3 + b_{10}\mathrm{slope} + b_{11}\mathrm{slope}^2 + b_{12}\mathrm{slope}^3 + b_{13}\mathrm{aspect} + b_{14}\mathrm{aspect}^2 + b_{15}\mathrm{aspect}^3 \quad (1)$$

where P represent precipitation (mm); b_0 is a constant; b_1, …., b_{15} represent the coefficient obtained for each independent variable; X, Y, DEM, slope and aspect represent the variables of location (X and Y), altitude (m), slope (°) and aspect (°), respectively. The general types of models to be employed were determined from the idea that a time series is regarded as being made up of two types of variation, namely (Gemmer *et al.* 2010): (1) a trend describing the long-term average of series; and (2) a periodic or seasonal component describing, for example, the variations within a cycle of duration 1 year.

Spatial analysis method

Surface interpolation and spatial analysis of precipitation and topographical factors were conducted to verify the efficiency and capability of the statistical analysis. The stochastic interpolation method of kriging, a standard approach for surface interpolation (Teegavarapu & Chandramouli 2005), was also adopted in the present study.

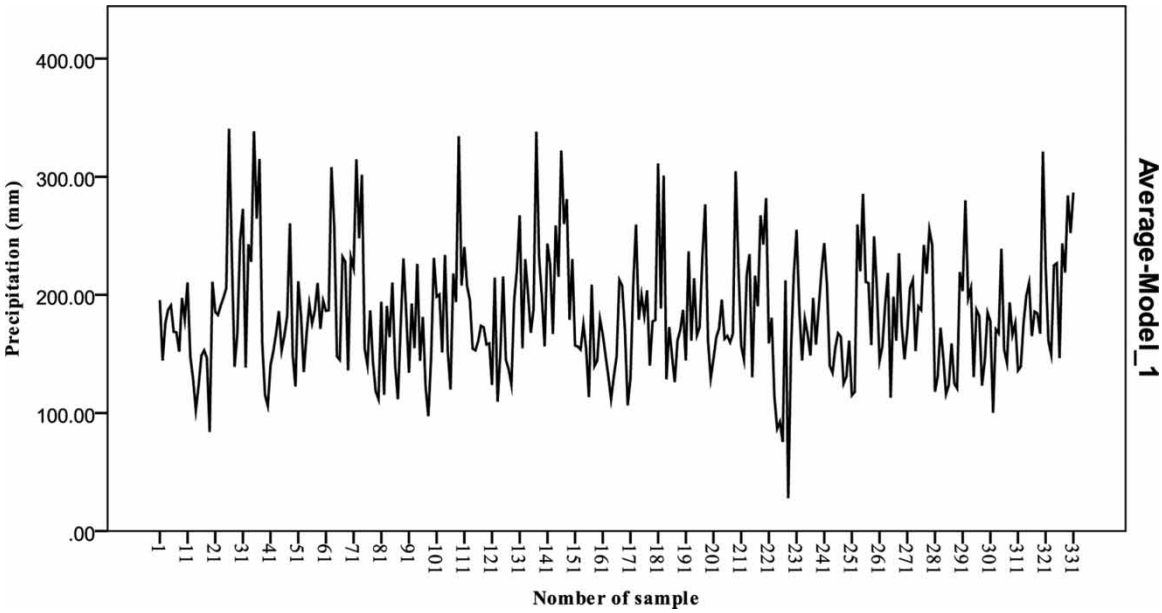

Figure 3 | Long-term average rainfall between 1974 and 2010.

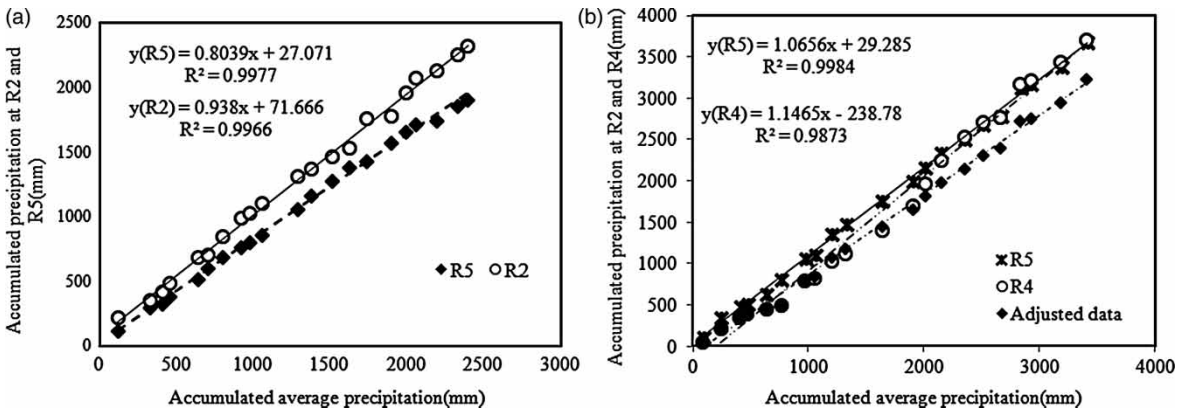

Figure 4 | Double-mass curve analysis: accumulated precipitation at (a) R5 and R2 versus LAR at January and (b) R4 and R2 versus LAR at March.

Kriging is a group of statistical techniques which estimate interactive surface from a set of points with z-values. Spatial variation of a continuous climate variable is considered too irregular (Vicente-Serrano *et al.* 2003); kriging can determine the best linear unbiased spatial behaviour of the phenomenon represented by the z-value (Zhang & Srinivasan 2010). Moreover, previous research on the performance of different interpolation methods has indicated that kriging provided improved validation results (Goovaerts 2000; Teegavarapu & Chandramouli 2005; Webster & Oliver 2007). This model is similar to IDW; it weighs the surrounding measured values to derive a prediction for an unmeasured location. The general formula for both interpolators is formed as a weighted sum of the data:

$$\hat{Z}(s_0) = \sum_{i=1}^{N} \lambda_i z(s_i) \tag{2}$$

where $z(s_i)$ is the measured value at the ith location; λ_i is an unknown weight for the measured value at its location; s_0 is

Table 3 | Modification factor in the precipitation gauges based on the double-mass curve

		Correction factor							
Month	**Index gauge**	**R1**	**R2**	**R3**	**R4**	**R6**	**R7**	**R8**	**R9**
January	R5	1.23	1.16	0.79	1.9	1.23	0.68	0.9	0.97
February	R5	1.05	1.07	1.114	1.33	1.18	1.69	0.88	0.92
March	R5	1.0	0.82	1.10	[h]1.01/0.93	0.94	1.47	1.04	1.20
April	R5	0.84	1.045	094	0.98	0.86	1.06	0.92	1.02
May	R5	0.91	1.37	1.27	1.16	1.26	1.0	1.15	0.98
June	R5	0.84	1.08	1.5	0.95	0.98	0.93	0.86	1.03
July	R5	1.05	1.07	1.103	0.93	[b]0.74/1.09	0.86	0.89	0.8
August	R5	0.85	1.27	1.21	1.12	1.18	1.03	0.98	1.12
September	R5	0.78	0.91	[a]1.05/0.96	0.89	1.02	[d]0.76/0.88	[f]0.71/0.89	1.3
October	R5	0.83	0.96	1.13	0.96	[c]1.08/06	1.25	1.02	0.95
November	R5	0.84	1.02	1	0.89	0.88	[e]1.05/1.09	[g]0.74/0.89	1.67
December	R5	1.03	1.03	1.04	0.64	0.94	1.22	1.14	1.17

[a]Before 1982/after 1983.
[b]Before 1983/after 1984.
[c]Before 1988/after 1989.
[d]Before 1979/after 1980.
[e]Before 1988/after 1989.
[f]Before 1980/after 1981.
[g]Before 1982/after 1983.
[h]Before 1985/after 1988.

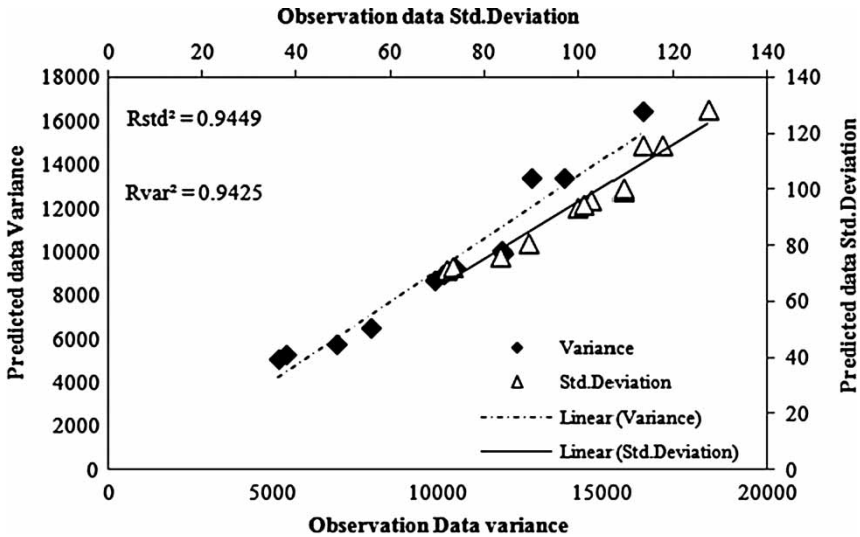

Figure 5 | Homogeneity of variance and standard deviation of predicted data versus observation data.

the prediction location; and N is the number of measured values.

The results from interpolation were evaluated by using statistical analysis to indicate the degree of concordance between model and reality (ESRI 2008; Hillier 2011).

Evaluation of different algorithms

The performance of the statistical and spatial analysis model was evaluated and compared using different indices to rescale the empirical equation to higher R^2 values.

Furthermore, predicted data based on spatial analysis and that based on the rainfall empirical equation were compared using indices. These measurements provide initial ideas on the model's validity, and can measure the efficiency of the estimation method (MarquInez *et al.* 2003; Di Piazza *et al.* 2011). For this study, the following indices were used.

- Root mean squared error (RMSE) was used to compare and validate models. It indicates how closely the model predicts the measured values. A smaller error indicates a better result (Goovaerts 2000; Mair & Fares 2011):

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n}(P_i - O_i)^2}{n}} \tag{3}$$

- The mean absolute error (MAE) indicates the average absolute difference between real rainfall and its estimation in the validation points:

$$\mathrm{MAE} = N^{-1}\sum_{i=1}^{N}|P_i - O_i| \tag{4}$$

Furthermore, to describe how predicted data are spread out from the mean value of observed precipitation, mean absolute deviation (MAD) was estimated. Accordingly, smaller values result in data that are more consistent:

$$\mathrm{MAD} = \sum_{i=1}^{N}\frac{|P - \bar{O}|}{n} \tag{5}$$

where N is the number of observations; O is the observation value; P is the predicted value; and $\bar{O}$ is the mean of the observation value. A conceptual schematic of the implementation of the method for predicting distribution of precipitation and designing process is provided by Figure 6.

RESULTS AND DISCUSSION

Statistical result analysis

To determine the relation between variables, ERDAS 9.1 software was used and topographical variables were extracted from SRTM data. The DEM was of 90 m resolution, with RSO-Malaysia project coordinate system.

Topographical variables and their Pearson's relation with precipitation are listed in Table 4 and the model parameters are provided in Table 5. Two topographic variables, location and DEM, are significant at a level of

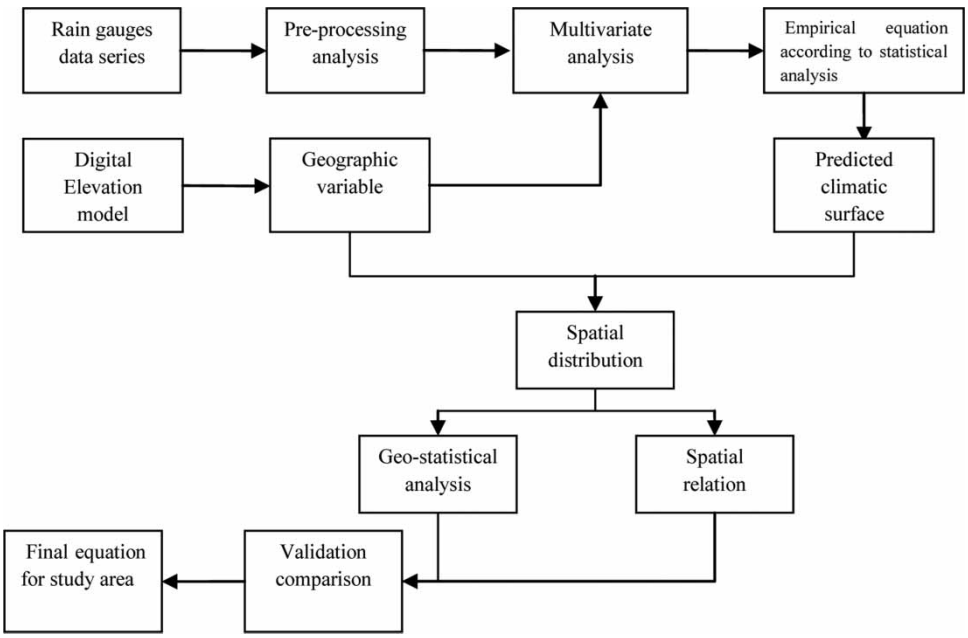

Figure 6 | Schematic of the methodological components.

Table 4 | Pearson correlation coefficient matrix for independent variables and precipitation

	X	Y	Elevation	Slope	Aspect	Moving average	Moving DRYseas	Moving WETseas
X	1	–0.473	0.969**	0.080	–0.394	–0.427	–0.287	–0.296
Y	–0.473	1	–0.658	–0.605	0.649	0.065	–0.127	0.062
Elevation	0.969**	–0.658	1	0.208	–0.473	–0.360	–0.163	–0.259
Slope	0.080	–0.605	0.208	1	–0.425	0.186	0.121	0.202
Aspect	–0.394	0.649	–0.473	–0.425	1	0.398	0.434	0.361
Moving average	–0.427	0.065	–0.360	0.186	0.398	1	0.793*	0.976**
Moving DRYseas	–0.287	–0.127	–0.163	0.121	0.434	0.793*	1	0.688*
Moving WETseas	–0.296	0.062	–0.259	0.202	0.361	0.976**	0.688*	1

Correlation is significant at the **0.05 and *0.1 level.

Table 5 | Summary of model characteristics

		Average				Dry season				Wet season			
Model	Topographical variable	ANOVA	R^2	Adj R^2	SE	ANOVA	R^2	Adj R^2	SE	ANOVA	R^2	Adj R^2	SE
1	Rain, DEM, slope	0.051	0.201	–0.07	9.59	0.08	0.052	0.264	5.39	0.06	0.136	0.15	18.16
2	Rain, DEM, slope, aspect	0.05	0.352	–0.31	9.47	0.058	0.305	0.112	5.06	0.059	0.296	0.13	17.96
3	Rain, DEM, slope, aspect, location	0.08	0.420	–0.55	11.57	0.051	0.328	0.031	4.72	0.09	0.313	0.83	22.90
4	Equation (1) variables	0.03	0.740	0.721	21.79	0.073	0.63	0.51	3.86	0.041	0.781	0.702	16.68

$p < 0.1$ and above. The respond rate for elevation and monthly average during the wet and the dry season is between 12 and 43% (Table 4). Preliminary analysis of different combinations of variables did not indicate a significant relation to spatial variability.

Due to poor relations between precipitation and independent factors, comparative analyses was conducted in a more 'natural' setting with description of other variables (Table 5). To illustrate this, four different scenarios from associated variables were developed (Table 5).

For models 1–4, the mean annual variability of precipitation was 20–74%; the spatial variability of mean precipitation during the wet season was within the range 13–78%. During the dry season, it dropped from 5 to 63%.

In model 4, the best response rate for precipitation prediction of four scenarios was 74, 78 and 63% for mean annual, dry season and wet season, respectively. Table 6 provides an evaluation of model 4, a multivariate regression model of the rain gauges, to determine the coefficient of the adjusted model. The best is the equation generated for both mean annual and wet season, which is significant at the $p <= 0.05$ level. Furthermore, the high correlation coefficient R^2 and adjusted R^2 indicate a good description of Equation (1).

The presence of autocorrelation in residuals (Durbin–Watson) test was indicated by 3.51, 3.34 and 3.76% for monthly average, wet season and dry season, respectively. The model can therefore predict spatial variability of precipitation in the Ulu Kinta Catchment. To improve the results of the statistical model, a spatial analysis was conducted.

SPATIAL DISTRIBUTION OF PRECIPITATION TREND

Trend surface and 3D plot analysis were useful supplements and provided pre-predictors of precipitation distribution patterns in three periods (Figure 7). The trend analysis showed different patterns for the periods of mean precipitation from 1974 to 2010. The annual average estimate illustrated a steady increase from west to northeast. During the dry season, precipitation was mostly distributed in the west, representing low values in highlands. However, during the

Table 6 | Multivariate regression models for rain gauges in the study area during 1975–2010

	Monthly annual average		Dry season		Wet season	
Variable of the model	**Unstandardised coefficient**	**Standardised coefficient**	**Unstandardised coefficient**	**Standardised coefficient**	**Unstandardised coefficient**	**Standardised coefficient**
b_0	–2,275.056		–453.848		–4,696.221	
b_1X	–0.003	–5.32	–0.001	–3.164	–0.005	–4.968
b_2X^2	0.0	0.0	0.0	0.0	0.0	0.0
b_3X^3	0.0	0.0	0.0	0.0	0.0	0.0
b_4Y	0.007	5.986	0.002	3.047	0.013	6.292
b_5Y^2	0.0	0.0	0.0	0.0	0.0	0.0
b_6Y^3	0.0	0.0	0.0	0.0	0.0	0.0
b_7DEM	–0.525	–39.404	–0.152	–22.057	–1.038	–42.815
b_8DEM^2	0.0	0.0	0.0	0.0	0.0	0.0
b_9DEM^3	2.84×10^{-7}	48.172	8.691×10^{-8}	28.237	5.56×10^{-7}	51.270
b_{10}slope	–2.299	–1.935	–2.268	–3.697	–3.043	–1.408
b_{11}slope^2	0.0	0.0	0.0	0.0	0.0	0.0
b_{12}slope^3	0.005	2.699	0.004	4.130	0.007	2.235
b_{13}aspect	–0.245	–2.794	–0.045	–1.004	–0.503	–3.153
b_{14}aspect^2	0.0	0.0	0.0	0.0	0.0	0.0
b_{15}aspect^3	7.185×10^{-7}	0.962	1.822×10^{-7}	0.473	1.593×10^{-6}	1.173
R^2	0.74	0.056	0.63		0.781	
Adj R^2	0.721		0.51		0.702	
SE	21.79		3.86		16.68	
DW	3.51		3.34		3.76	

wet season, precipitation is mostly distributed with high values in the middle part of the study area. The amount of rainfall decreased in the east and west. During this season, the north becomes the wettest area.

To understand the capability of model 4, a spatial distribution of precipitation pattern was considered for the annual average, dry season and wet season. This interpolation is a combination of a multi-regression and geospatial interpolation. An exponential model was performed. A lag size after trial procedure was fixed at 12 km, totalling 12 lags. Figure 8 depicts an example of an advanced surface modelling method. A digital elevation map (Figure 8(a)) presents the elevation pattern and its relation to precipitation. In trend analysis, a similar behaviour with clearer distribution is evident in three scenarios. The spatial distribution of annual average precipitation was related to elevation, slope and aspect. A west–northeast gradual increase in the amount of rainfall was also represented. Three different zones can therefore be extracted from the distribution map. The west zone recorded low precipitation, which gradually increased in areas of higher altitude situated in the northeast. Moreover, the high land in the eastern part showed a constant decrease in precipitation, whereas an increase was noted in the northeast.

During the dry season, the negative trends of precipitation were observed with elevation, slope and aspect (compare Figure 8(a) and 8(c)). During the wet season, precipitation gradual increased from south to north and from west to east. Meanwhile, the wettest area is in the north and in a small south-eastern portion of the Ulu Kinta Catchment.

Validation of spatial analysis

To determine the accuracy of the estimated statistical and spatial model, RMSE, MAE and MAD were considered. Results of the analysis are presented in Table 7. The most

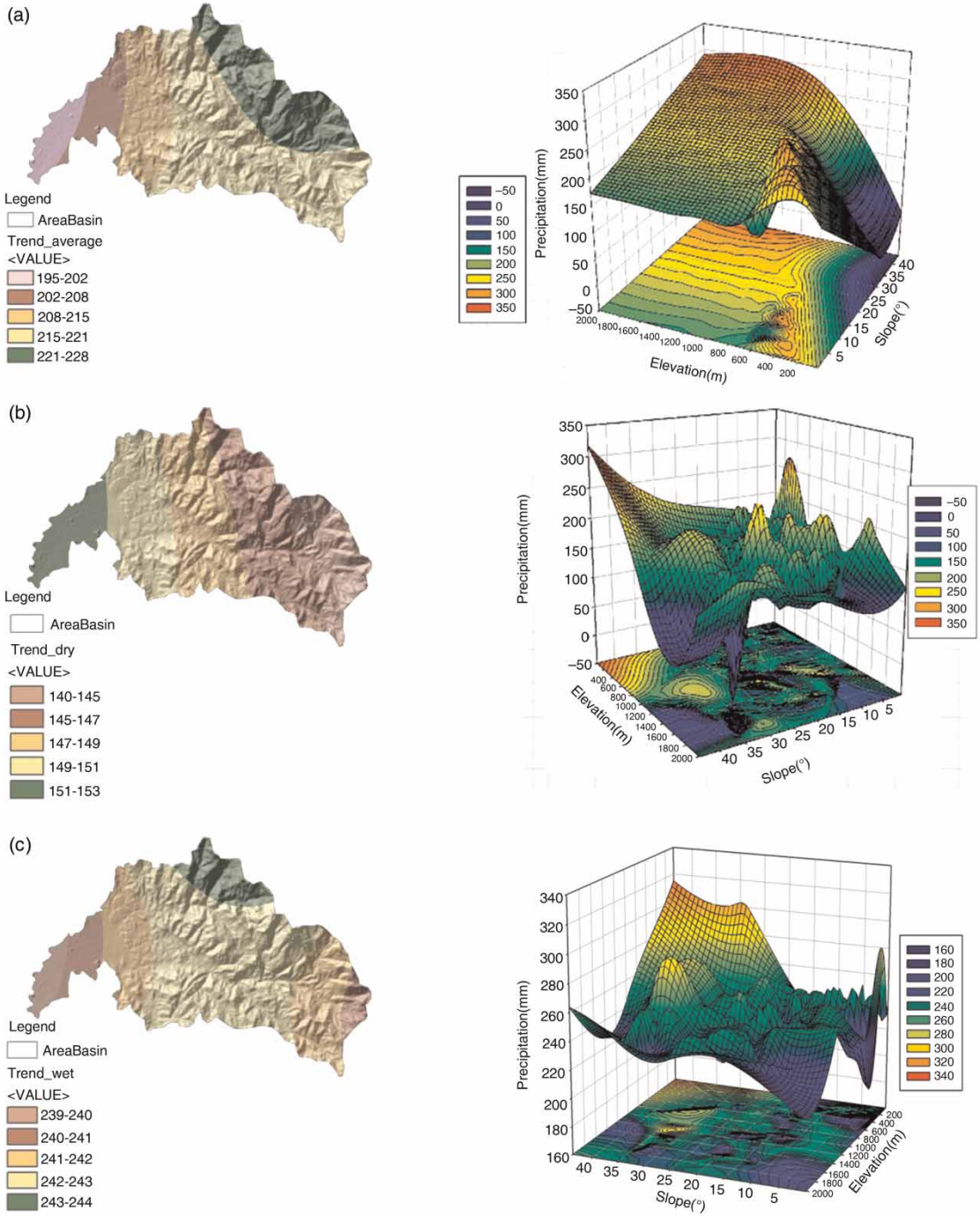

Figure 7 | Precipitation trend in the study area by surface map and 3D plot: (a) annual avergae rainfall trend; (b) dry season trend; and (c) wet season trend.

accurate result is that for the wet season. Based on statistical analysis, the response rate for the annual average and dry season is 74 and 63%, respectively. Furthermore, the coefficient of determination between two models (statistical and spatial model) improved in all cases. The regression model has a lower RMSE, MAE and MAD value in spatial analysis than in statistical analysis. Based on the kriging method, the predicted precipitation value indicated a slight increase by

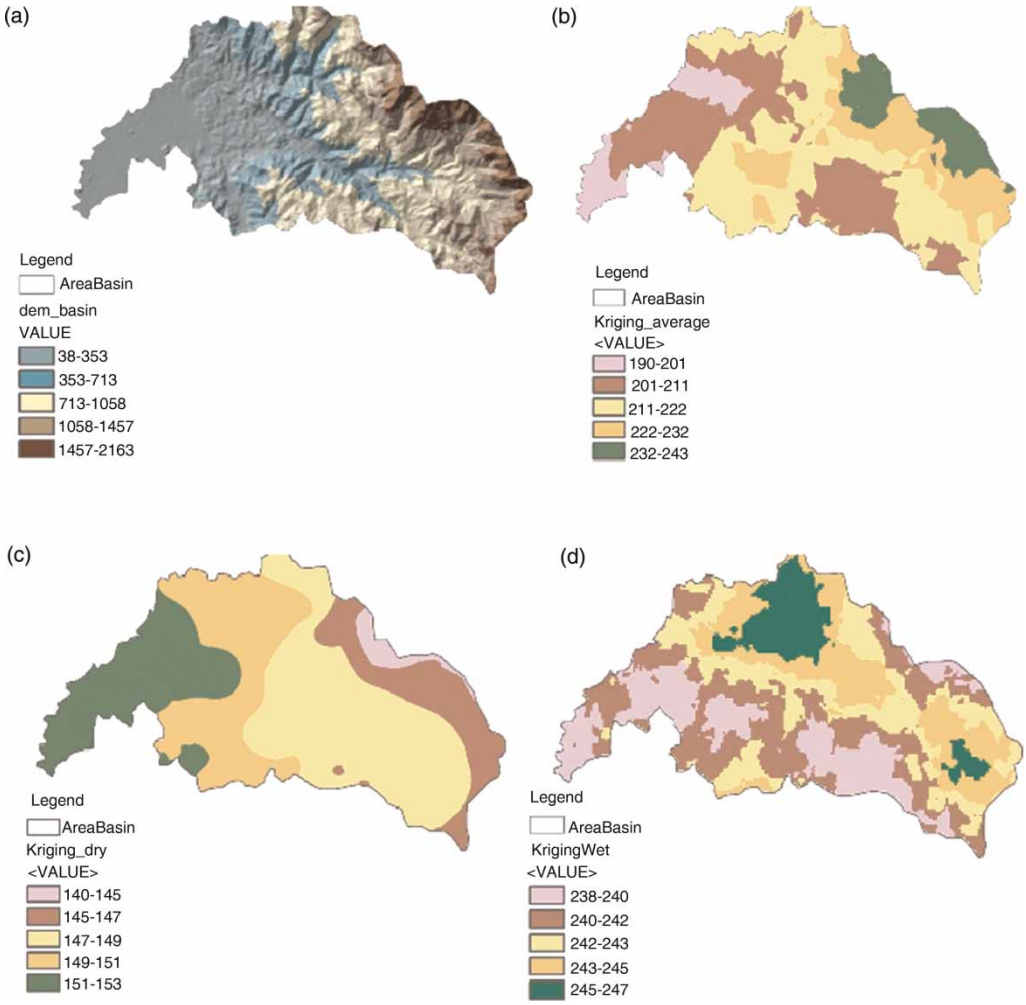

Figure 8 | Result of kriging interpolation methods: (a) DEM of the study area; (b) annual average precipitation distribution; (c) precipitation distribution in the dry season; and (d) precipitation distribution in the wet season.

Table 7 | Accuracy measurement for precipitation models

	According to regression result				After geostatistical analysis (kriging)			
Model	R^2	**RMSE (mm)**	**MAE (mm)**	**MAD (mm)**	R^2	**RMSE (mm)**	**MAE (mm)**	**MAD (mm)**
Annual average	0.86	21.88	24.32	11.23	0.88	19.38	17.79	10.97
Dry season	0.79	20.58	24.85	15.87	0.80	23.77	14.97	9.19
Wet season	0.88	26.38	21.58	13.46	0.89	19.52	15.81	13.28

3–7 mm for RMSE, MAE and MAD. This suggests that the residuals are random errors, and that the regression model is effective in interpreting the spatial variability of precipitation.

CONCLUSION

Topographical variables such as elevation, slope, aspect and location were used to estimate precipitation distribution

using synchronous observation from surrounding stations to predict values at unsampled locations. A combination of statistical and spatial analyses was used to estimate rainfall through a catchment. Findings suggest that a multivariate regression method that uses limited rain gauge data in precipitation distribution should be used.

Results from multivariate regression indicate the capability of the model to distribute rainfall through a catchment. The accuracy indices showed 63–78% of the spatial variability of annual average precipitation in the study region. Furthermore, the model is more efficient during the wet season than the dry season.

The most direct approach comprises the use of multivariate regression analysis for precipitation value and the application of geostatistical regression model. These provided better results for interpreting the spatial variability of precipitation for the study area (i.e. in wet season up to 35%).

However, these findings are limited by the use of topographical factors and precipitation data from an insufficient number of stations. Nevertheless, the findings can be applied to other regions to estimate rainfall distribution. Further investigation on precipitation, especially concerning satellite imagery, is strongly recommended.

REFERENCES

Albert, J. M. 2004 Hydraulic analysis and double mass curves of the Middle Rio Grande from Cochiti to San Marcial, New Mexico. Master of Science, Colorado State University, Fort Collins, Colorado, USA.

ASCE 1996 *Hydrology Handbook*. American Society of Civil Engineers, Reston, VA.

Arnold, J. G., Srinivasan, R., Muttiah, R. S. & Williams, J. 1998 Large area hydrologic modeling and assessment Part I: model development. *JAWRA Journal of the American Water Resources Association* **34**, 73–89.

Basist, A., Bell, G. D. & Meentemeyer, V. 1994 Statistical relationships between topography and precipitation patterns. *Journal of Climate* **7**, 1305–1315.

Daly, C., Neilson, R. P. & Phillips, D. L. 1994 A statistical-topographic model for mapping climatological precipitation over mountainous terrain. *Journal of Applied Meteorology* **33**, 140–158.

Desa, M. N. & Niemczynowicz, J. 1996 Temporal and spatial characteristics of rainfall in Kuala Lumpur, Malaysia. *Atmospheric Research* **42**, 263–277.

Di Piazza, A., Conti, F. L., Noto, L. V., Viola, F. & La Loggia, G. 2011 Comparative analysis of different techniques for spatial interpolation of rainfall data to create a serially complete monthly time series of precipitation for Sicily, Italy. *International Journal of Applied Earth Observation and Geoinformation* **13**, 396–408.

ESRI 2008 ArcEditor 9.3.1 (Computer software) (Online). Redlands, CA. Available from: http://resources.arcgis.com/content/web/about.

Gemmer, M., Fischer, T., Jiang, T., Su, B. & Liu, L. L. 2010 Trends in precipitation extremes in the Zhujiang River Basin, South China. *Journal of Climate* **24**, 750–761.

Goovaerts, P. 2000 Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall. *Journal of Hydrology* **228**, 113–129.

Guan, H., Wilson, J. L. & Makhnin, O. 2005 Geostatistical mapping of mountain precipitation incorporating autosearched effects of terrain and climatic characteristics. *Journal of Hydrometeorology* **6**, 1018–1031.

Hillier, A. 2011 *Manual for Working with ArcGIS 10, The Selected Works of Amy Hillier*. University of Pennsylvania, Pennsylvania.

Larson, L. W. & Peck, E. L. 1974 Accuracy of precipitation measurements for hydrologic modeling. *Water Resources Research* **10**, 857–863.

Mair, A. & Fares, A. 2010 Assessing rainfall data homogeneity and estimating missing records in Makaha valley, O'ahu, Hawai'i. *Journal of Hydrologic Engineering* **15**, 61–66.

Mair, A. & Fares, A. 2011 Comparison of rainfall interpolation methods in a mountainous region of a tropical island. *Journal of Hydrologic Engineering* **16**, 371–383.

Marquinez, J., Lastra, J. & Garcia, P. 2003 Estimation models for precipitation in mountainous regions: the use of GIS and multivariate analysis. *Journal of Hydrology* **270**, 1–11.

Ninyerola, M., Pons, X. & Roure, J. M. 2007 Monthly precipitation mapping of the Iberian Peninsula using spatial interpolation tools implemented in a Geographic Information System. *Theoretical and Applied Climatology* **89**, 195–209.

Phillips, D. L., Dolph, J. & Marks, D. 1992 A comparison of geostatistical procedures for spatial analysis of precipitation in mountainous terrain. *Agricultural and Forest Meteorology* **58**, 119–141.

Ranhao, S., Baiping, Z. & Tan, J. 2008 A multivariate regression model for predicting precipitation in the Daqing Mountains. *Mountain Research and Development* **28**, 318–325.

Searcy, J. K. & Hardison, C. H. 1960 *Double-Mass Curves, Manual of Hydrology. Part 1*. US Geological Survey Water-Supply, Washington.

Shepard, D. 1968 A two-dimensional interpolation function for irregularly-spaced data. In: *Proceedings of 23rd Association for Computing Machinery National Conference*, ACM, New York, pp. 517–524.

Smith, R. B. 1979 The influence of mountains on the atmosphere. *Advances in Geophysics* **21**, 87–230.

Spreen, W. C. 1947 A determination of the effect of topography upon precipitation. *Transactions of American Geophysics Union* **28**, 285–290.

Suhaila, J., Deni, S. M., Zin, W. Z. W. & Jemain, A. A. 2010 Spatial patterns and trends of daily rainfall regime in Peninsular Malaysia during the southwest and northeast monsoons: 1975–2004. *Meteorology and Atmospheric Physics* **110**, 1–18.

Tabios III, G. Q. & Salas, J. D. 1985 A comparative analysis of techniques for spatial interpolation of precipitation. *Journal of the American Water Resources Association (JAWRA)* **21**, 365–380.

Tang, W. Y., Kassim, A. H. M. & Abubakar, S. H. 1996 Comparative studies of various missing data treatment methods – Malaysian experience. *Atmospheric Research* **42**, 247–262.

Teegavarapu, R. S. V. & Chandramouli, V. 2005 Improved weighting methods, deterministic and stochastic data-driven models for estimation of missing precipitation records. *Journal of Hydrology* **312**, 191–206.

Varikoden, H., Preethi, B., Samah, A. A. & Babu, C. A. 2011 Seasonal variation of rainfall characteristics in different intensity classes over Peninsular Malaysia. *Journal of Hydrology* **404**, 99–108.

Vicente-Serrano, S. M., Saz-Sánchez, M. A. & Cuadrat, J. M. 2003 Comparative analysis of interpolation methods in the middle Ebro Valley (Spain): application to annual precipitation and temperature. *Climate Research* **24**, 161–180.

Webster, R. & Oliver, M. A. 2007 *Geostatistics for Environmental Scientists*. John Wiley & Sons, Chichester.

Wong, C. L., Venneker, R., Uhlenbrook, S., M. Jamil, A. B. & Zhou, Y. 2009 Variability of rainfall in Peninsular Malaysia. *Hydrology and Earth System Sciences Discussions* **6**, 5471–5503.

Zhang, X. & Srinivasan, R. 2010 GIS-based spatial precipitation estimation using next generation radar and raingauge data. *Environmental Modelling & Software* **25**, 1781–1788.

First received 1 November 2011; accepted in revised form 10 July 2012. Available online 5 December 2012