SPATIAL INTERPOLATION OF MONTHLY PRECIPITATION IN SELANGOR, MALAYSIA – COMPARISON AND EVALUATION OF METHODS

1Nurul Nadiah Mohd Firdaus and 2Suhaimi Abdul Talib
1Faculty of Applied Sciences, 2Faculty of Civil Engineering, Universiti Teknologi MARA, 40450, Shah Alam, Malaysia
Correspondence: 1nadi0978@gmail.com

ABSTRACT
There are many spatial interpolations schemes, but none of them can perform best in all cases. Hence, this study aims to find an optimal interpolation scheme for precipitation in Selangor and Langat basin of which are the two major basins in Selangor. In order to obtain spatially distributed precipitation data, 21 measured rain gauges points are interpolated. Five interpolation methods have been tested after exploring data and cross-validation was used as the criterion to evaluate the accuracy of the various methods. The best method was obtained by the kriging method while the inverse distance weighting (IDW) perform worst.

Keywords: Precipitation, Interpolation, Cross-Validation, Kriging, IDW

1. INTRODUCTION
Hydrology and water quality applications in catchment areas no doubt require data on the most important parameter which is the precipitation. Since these data are often collected using the rain gauge, they are then considered as point data. However, the use of single rain gauges as precipitation inputs carries great uncertainties regarding runoff estimation (Faurès et al., 1995 and Chaubey et al., 1999). This presents a great problem for the prediction of discharge, groundwater level and soil moisture, especially if the rain gauge is located outside of the catchment (Schuurmans and Bierkens, 2007). As a result, some applications such as precipitation mapping on erosions (Aronica and Ferro, 1997; Goovaerts, 1999; Hoyos et al., 2005; Nyssen et al., 2005; Angulo-Martinez and Begueria, 2009; Angulo-Martinez et al., 2009) and hydrological modelling (Syed et al., 2003; Kobold and Suselj, 2005; Gabellani et al, 2007; Cole and Moore, 2008; Collischonn et al., 2008; Ruelland et al., 2008; Moulin et al., 2009) require precipitation data that are spatially continuous. The quality of such result is thus determined by the quality of the continuous spatial precipitation (Singh, 1997; Andreassian et al., 2001; Kobold and Suselj, 2005; Leander et al., 2008; Moulin et al., 2009).

Spatial interpolation can be used to estimate precipitation variables at other locations. Although there are several methods to perform this, it can be a challenging task to determine which of the methods produce the results closest to the actual conditions. Each methods advantage and their disadvantages hence depend strongly on the characteristics of the data set used in order to define their suitability. Thus, criteria must be fund to decide whether the method chosen is suited for the point data set. Besides, it is also important to define the aims of the interpolation as different aims can represent different criteria for the evaluation of the interpolation method used.

The most frequently used deterministic methods in spatial interpolation for precipitation are the Thiessen polygon and Inverse Distance Weighting (IDW) while the geostatistical method constitutes a discipline involving mathematics and earth sciences. Dirks et al. (1998) compared Inverse Distance Weighting, the Thiessen polygon and kriging in interpolating precipitation data from a network of thirteen rain gauges on Norfolk Island. They recommended the use of IDW for interpolations for spatially dense networks. Nalder and Wein (1998) used cross validation to evaluate four forms of kriging and three simple alternatives for spatial interpolation of climatic data. They found that IDW had a smaller error of estimates than Ordinary Kriging (ORK) and Universal Kriging (UNK) in interpolating monthly precipitation in the Canadian boreal forest. Buytaert et al. (2006) studied the variability of spatial and
temporal precipitation in the south Ecuadorian Andes using the Thiessen polygon and kriging. Their study suggested that spatial interpolation with kriging gives better a result than Thiessen polygon, and the accuracy of both methods improves when external trends are incorporated. Basistha et al. (2008) analysed the spatial distribution of precipitation in the Indian Himalayas using both deterministic and geostatistical methods. They reported that UNK was the most suitable method, followed by ORK and IDW.

Although comparisons of interpolation methods for precipitation in numerous areas of the world has been studied, (Holdaway, 1996; Dodson and Marks, 1997; Thornton et al., 1997; Couralt and Monestiez, 1999; Xia et al., 1999; Shen et al., 2001; Xia et al., 2001; Jarvis and Stuart, 2001; Hasenauer et al., 2003; Berne et al., (2004); Garen and Marks, 2005; Stahl et al., 2006; Attorre et al., 2007; Hofstra et al., 2008; Mehdi Keblouti et al., 2012; Wang et al., 2014), review of the literature reveals that no interpolation study of climatic variables has been applied to the study area. However, there is a single attempt to introduce a method for estimating mean monthly precipitation in the Langat River Basin, Selangor by analysing its precipitation trend (Palizdan et al., 2013).

In this study five GIS-based spatial interpolation methods were compared to determine their suitability for estimating mean monthly precipitation, from data recorded at nearly 21 rain gauges in Selangor. This study is then constructed in two parts: the first consists of the presentation of the precipitation network in Selangor, the available data and the different methods used to spatially interpolate the precipitation of the area. While the second part presents the methodological of the analysis and the results obtained from the evaluation method. In this part, cross-validation method is used to assess which method gives the best interpolation.

### 2. MATERIALS AND METHODS

#### Study Area

The Selangor and Langat River basin in the Selangor state is used as the case study basin. The total area of the study basin is around 2514.13 sq. km. and the climate is Tropical with hot, dry season and wet monsoon season. The basin is divided into 34 sub catchment areas (20 in Selangor catchment and 14 in Langat catchment), each of which is a source of surface runoff as well as an independent groundwater aquifer. Fig. 1 shows the schematic localities of the basins in the study area.

![Schematic Locality of the Basins in Selangor](image.jpg)
Selangor catchment
Sungai Selangor basin which is located in the state of Selangor is approximately 70 km long and 30 km wide that stretches from Bukit Fraser on the northeast to the Straits of Malacca in the west with a catchment area of 1820 sq. km. Among the main tributaries in the basin area are Sg Batang Kali, Sg Serendah, Sg Kuang, Sg Ranching, Sg Buloh, Sg Kerling and Sg Garing. The flow of the Sungai Selangor in general is in the south west direction before ending into the Straits of Malacca via the town of Kuala Selangor. Alluvial soils accounted 55% of the area which mainly are in the coastal plains and riverine areas, although part of this area especially the coastal peat area have been drained for paddy cultivation while some hilly areas are cultivated for oil palm and rubber plantation (Department of Environment, 1994). Selangor basin experienced high temperature and humidity with a relatively small amount of seasonal variations besides subjected to two monsoonal periods which are the Northeast Monsoon (October till January) and the Southwest Monsoon (May to September). Average precipitation is in the range of 2000 mm to 3500 mm with the largest peak during the Northeast Monsoon.

Langat Catchment
The Langat basin is located at the southern part of Klang Valley which is the most urbanized river basin in Malaysia. It is believed that the Langat basin is currently experiencing “spill over” effects due to excessive development in the Klang Valley. Hydro meteorologically, the Langat basin is affected by two types of monsoon, i.e. the northeast (November – March) and the southwest (May – September) monsoons (Hadi et al. 2012). Average annual precipitation is about 2400mm. The wettest months are April and November with average monthly precipitation exceeding 250mm, while the driest month is June with average monthly precipitation not exceeding 100mm. Topographically, the Langat basin can be divided into three distinct areas with reference to the Langat River: the mountainous area in the upstream, undulating land in the centre and flat flood plain in the downstream. The basin has a rich density of landforms; surface features and land cover (Hadi et al., 2012, Palizdan et al., 2013).

Data Collection
The data used in this study comprise continuous records of mean monthly precipitation for the period (1970 - 2010) in 21 stations scattered throughout the study area (Fig. 2). These data have been originally provided by the Department of Irrigation and Drainage Malaysia for all precipitation gauges in the study area.

Fig. 2. Location of the 21 rain gauges stations used in the study
2. INTERPOLATION METHODS

The interpolation methods used in this study were performed by ESRI ArcGIS® Geostatistical Analyst 10.2. Geostatistical Analyst is an extension to the ArcGIS Desktop that provides a powerful suite of tools for spatial data exploration and surface generation using sophisticated statistical methods. Geostatistical Analyst provides two groups of interpolation techniques: deterministic and geostatistical. All methods rely on the similarity of nearby sample points to create the surface. Deterministic techniques use mathematical functions for interpolation. Geostatistics relies on both statistical and mathematical methods, which can be used to create surfaces and assess the uncertainty of the predictions. This section briefly introduces the different interpolation methods used in this study, detailed descriptions of these methods are reported elsewhere (ESRI, 2001; Li and Heap, 2008; Chang, 2010; Lloyd, 2010).

Deterministic Methods

Deterministic interpolation methods create surfaces from measured points, based on either the extent of similarity like Inverse Distance Weighted or the degree of smoothing such as Radial Basis Functions. Deterministic interpolation methods can be divided into two groups: global and local. Global methods calculate predictions using the entire dataset. Local methods calculate predictions from the measured points within neighbourhoods, which are smaller spatial areas within the larger study area. Geostatistical Analyst provides the Global Polynomial as a global interpolator and the Inverse Distance Weighted, Local Polynomial, and Radial Basis Functions as local interpolators. Deterministic interpolation techniques may be exact or inexact interpolators. Exact interpolators such as Inverse Distance Weighted Interpolation and Radial Basis Functions generate a surface that passes through the control points. In contrast, inexact interpolators such as Global and Local Polynomial predict a value at the point location that differs from its known value.

A-Inverse Distance Weighted (IDW) Interpolation

IDW is the workhorse of spatial interpolation, the method that is most often used by GIS analysts. It employs the Tobler’s First Law of Geography by estimating unknown measurements as weighted averages over the known measurements at nearby points, giving the greatest weight to the nearest points (Longley et al., 2011). The general equation for IDW method is shown in equation (1):

$$Z_0 = \frac{\sum_{i=1}^{n} \frac{Z_i}{d_i^k}}{\sum_{i=1}^{n} \frac{1}{d_i^k}}$$

(1)

Where $Z_0$ is the estimated value at point 0, $Z_i$ is the Z value at known point $i$, $d_i$ is the distance between point $i$ and point 0, $n$ is the number of known points used in estimation, and $k$ is the specified power which controls the degree of local influence (Chang, 2010).

B-Global Polynomial (GP) Interpolation

GP interpolation simply uses multiple regression methods on all of the data. A response or trend surface is fitted to the x-and y-coordinates, which are the covariates. A first-order Global Polynomial (linear) fits a single plane through the data as shown in equation (2):

$$Z (X_i, Y_i) = \beta_0 + \beta_1 X_i + \beta_2 Y_i + \epsilon (X_i, Y_i)$$

(2)

Where $Z (X_i, Y_i)$ is the datum at location $(X_i, Y_i)$, $\beta_i$ are parameters, and $\epsilon (X_i, Y_i)$ is a random error. A second-order Global Polynomial (quadratic) fits a surface with a bend in it, allowing surfaces representing valleys; a third-order Global Polynomial (cubic) allows for two bends; and so forth, up to a 10 are allowed in Geostatistical Analyst (ESRI, 2001).

C-Local Polynomial (LP) Interpolation

As with global polynomials a least square polynomial fit to the data is applied, with options for Order 1, 2 or 3 equations. However, instead of fitting the polynomial to the entire dataset it is
fitted to a local subset defined by a window. The size of this window needs to be large enough for a reasonable number of data points to be included in the process. One further adjustment is made to this procedure — a measure of distance-based weighting is included, so the least squares model is in fact a weighted least squares fit. The weights are computed using a power function of distance as a fraction of the window size. The simplest case is where the moving window is a circle with radius R. If the distance between grid point \((X_i, Y_i)\) and a data point \((x,y)\) within the circle is denoted \(d_i\), then the weight \(w_i\) is given by equation (3) and the least squares procedure then involves minimizing the expression given by equation (4) (De Smith et al., 2011):

\[
w_i = (1 - \frac{d_i}{R})^p
\]

\[
\sum_{i=1}^{n} w_i(f(x_i, y_i) - z_i)^2
\]

Where \(p\) is a user definable power and if \(p = 0\) all the weights are 1.

**Geostatistical Methods**

Geostatistical interpolation methods create surfaces incorporating the statistical properties of the measured data. These techniques produce not only prediction surfaces but also error or uncertainty surfaces, giving the analyst an indication of how good the predictions are. Many methods are associated with geostatistics, but all are in the Kriging family. Kriging assumes that the spatial variation of an attribute is neither totally random (stochastic) nor deterministic. Instead, the spatial variation may consist of three components: a spatially correlated component, variation representation of the regionalized variable in the form of a “drift” or structure that represents a trend; and a random error term. The interpretation of these components has led to development of different Kriging methods for spatial interpolation. In this study, Ordinary and Universal Kriging was used.

**A-Ordinary Kriging**

Assuming the absence of a drift, Ordinary Kriging (OK) focuses on the spatially correlated component and uses the fitted semivariogram, a diagram relating the semivariance to the distance between sample points used in Kriging, directly for interpolation. The estimator of ordinary Kriging is given by equation (6):

\[
Z^*(x_0) = \sum_{i=1}^{n} \lambda_i Z(x_i)
\]

Where \(Z^*(x_0)\) is the estimate value at \(x_0\), \(Z(x_i)\) is the measure value at the \(x_i\) and \(\lambda_i\) is the weight assigned for the residual of \(Z(x_i)\) (Sun et al., 2009).

**B-Universal Kriging**

Universal Kriging (UK) assumes that the spatial variation in \(z\) values has a drift or a trend in addition to the spatial correlation between the sample points. By definition of the drift component, the expected value \(m(x)\) of \(z(x)\) at point \(z\) is given by equation (7) and the estimator of universal Kriging is given by equation (8) (Sun et al., 2009):

\[
E[Z(\chi)] = m(\chi)
\]

\[
Z^*(x_0) = \sum_{i=1}^{n} \lambda_a Z_a
\]
Where \( n \) the number of is available sampling data, \( Z^*(x_0) \) is the estimate value, \( Z_a \) is the measured value at sampling point \( a (a = 1, \ldots, n) \), and \( \lambda_a \) is the weighting coefficient, which is calculated with unbiased and minimum error variance.

**Cross-validation**

Cross-Validation was used to evaluate the performance of each interpolation method. It is one of the most commonly used statistical techniques for comparing interpolation methods. Cross-Validation compares the interpolation methods by repeating the following procedure for each interpolation method to be compared (Chang, 2010): (1) Remove a known point from the data set, (2) Use the remaining points to estimate the value at the point previously removed, and (3) Calculate the predicted error of the estimation by comparing the estimated with the known value. After completing the procedure for each known point, two common diagnostic statistics, Root Mean Square Error (RMSE) and the standardized RMSE, are calculated to assess the accuracy of the interpolation method as shown in equations (9) and (10):

\[
\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (Z_i - Z)^2}
\]  
\[
\text{Standardized RMSE} = \frac{\text{RMSE}}{S}
\]

Where \( Z_i \) and \( Z \) are the measured and the estimated value at the sampling point \( i (i = 1, 2, \ldots, n) \); \( n \) is the number of values used for the estimation; and \( S \) is the standard error. The RMSE statistic is available for all exact local methods, but the Standardized RMSE is only available for Kriging because the variance is required for computation. A better interpolation method should yield a smaller RMSE and a better Kriging method should yield a smaller RMSE and a Standardized RMSE closer to 1 (Chang, 2010).

### 3. RESULTS AND DISCUSSION

Mean Monthly Precipitation for two selected months (Jun and October) as representatives of the two monsoon seasons (South West Monsoon and South East Monsoon) in Selangor was interpolated in turn using five GIS based interpolation techniques (IDW, LP, GP, OK, UK). Fig. 3 shows samples of interpolated surfaces using different methods. RMSE (for the six methods) and Standardized RMSE (for only OK and UK) were then calculated using Cross-Validation as shown in Fig. 4. The minimal RMSE are obtained by OK and UK, which have almost the same RMSE and Standardized RMSE for all months. Thus this method is the optimal method for interpolating mean monthly precipitation in the study area. Similar results were obtained from Nusret and Dug (2012), Ayandale and Odekunle (2009), Basistha et al. (2008), Atkinson (1998) and Tabios and Salas (1985) where their studies found that interpolation using OK and UK were the optimal method in addressing the precipitation of their studied area.
Ordinary Kriging and Universal Kriging are the most optimal methods for interpolating mean monthly precipitation in Selangor. This conclusion is based on available precipitation data recorded 21 rain gauges stations representing two main catchments; Langat and Selangor during the period (1970 - 2010), which were in turn interpolated using five GIS-based interpolation methods. Cross-Validation was used to compare the various interpolation methods. Diagnostic

Statistic indicated that Ordinary and Universal Kriging had the smallest RMSE and thus they are considered the optimal methods for interpolating precipitation in Selangor.

4. ACKNOWLEDGEMENTS
This research was supported by the International Foundation of Sciences (IFS) grant (W/5238-1), the Department of Irrigation and Drainage Malaysia for the precipitation data and Universiti Teknologi MARA Malaysia.

5. REFERENCES


