

Comparison of Rainfall Interpolation Methods in Langat River Basin

M Hassim^{*1}, A Yuzir¹, M N Razali², F C Ros¹, M F Chow³ and F Othman⁴

¹ Malaysia-Japan International Institute of Technology (MJIT), Universiti Teknologi Malaysia, Jalan Sultan Yahya Petra, 54100 Kuala Lumpur, Malaysia.

² Universiti Teknologi Malaysia, 81310 Johor Bahru, Johor, Malaysia.

³ Universiti Tenaga Nasional, Jalan Ikram-Uniten, 43000 Kajang, Selangor, Malaysia

⁴ University of Malaya, Jalan Universiti, 50603 Kuala Lumpur, Malaysia

*Corresponding author e-mail: muhammadhassim37@gmail.com

Abstract. Rainfall is an element of climate that can be measured by a rain gauge. The rain gauge was set up for every station predefined by the Department of Irrigation and Drainage (DID) Malaysia. One millimeter (mm) of rainfall means that within a square meter of a flat surface, water can be as high as one mm. In the hydrology model, the rainfall data is very important in order to predict the flood or assist in the disaster mitigation plan. In this case, the availability of complete rainfall data in a region is essential. By performing spatial interpolation, rainfall data can predict values from the empty data at each point. In this study, Inverse Distance Weighting (IDW), Ordinary Kriging (OK), Simple Kriging (SK) and Kernel Smoothing (KS) method were considered in the rainfall interpolation for this area. Rainfall data at 20 points in Langat River Basin that obtained from DID Ampang for the period 2008–2017 were used as reference data. This study aimed to compare IDW, Kriging and Spline methods to obtain better interpolation methods. The interpolation is done by running a cross-validation using a geostatistical wizard in ArcGIS. The method effectiveness was evaluated by the calculation of mean error (ME), Root Mean Square Error (RMSE), Root mean Square Standardize Error (RMSSE) and Average Standard Error (ASE). For IDW method, only the ME and RMSE results are available. From the result, it can be seen that SK method outperforms the IDW, OK and KS method for these rainfall interpolations in Langat River Basin by showing better statistical evaluation.

1. Introduction

Rain is a liquid form of precipitation that falls to the ground as a result of water vapor condensation in the atmosphere. Rain is an important source of natural water. Although rain provides many benefits, excess rainfall is a factor that contributes to many natural disasters such as flash floods and landslides. In addition, rainfall deficiency can result in dry or drought situations that can affect human daily life, both animals and plants. Therefore, an index for measuring the excessive or shortage of rainfall is important to provide information on the nature of rainfall [1].

The amount of rainfall that falls to the ground can be measured by a rain gauge. More than 60 mm of rainfall in two (2) to four (4) hours typically can cause flash floods. However, monsoon rains for long periods of time with heavy intervals of high intensity within 24 hours can cause a monsoon flood [2]. The amount of rainfall is recorded in inches or millimeters (1 inch = 25.4 mm), where 1 mm of rainfall indicates the height of rainwater covering the surface of 1 mm if the water does not absorb into the soil or evaporate into the atmosphere [3].



In order to determine the beginning of the rainy season and the beginning of the dry season, it is necessary to have complete and accurate rainfall data. Limitations in data collection have made it almost impossible to directly measure continuous spatial data. To determine a new value or point on a surface from a given value or point, interpolation process is required, where this technique is also used to predict unobservable variables in a variety of locations. In the interpolation method, the probabilistic function is applied to generate the value of the variable to be predicted [4].

2. Methodology

2.1. Study area

The study uses rainfall data in 20 monitoring stations in Langat River Basin from DID Malaysia. The area with an overall length of 200 km and 20-52 meters wide (average). The overall catchment area is 2,423 sq. km (Figure 1). The Langat River basin, which runs from Mount Nuang, Titiwangsa Range is located in the southern part of the State of Selangor and is the second largest basin in the State of Selangor. It is located south of Kuala Lumpur City and flows through three states namely Selangor, Federal Territory of Putrajaya and Negeri Sembilan before reaching its destination or its peak in the Straits of Malacca near Bandar Banting, Kuala Langat. The Langat River Basin is one of the fastest growing economic and physical development areas [5]. As such, the river basin saw ongoing development processes, especially mega developments such as KLIA, Putrajaya, Cyberjaya, F1 Circuit and the opening of new areas to meet the housing, business and industrial needs.

The limitations of data in analyzing are often created some issues. The data obtained is sometimes not complete as we require some interpolation method to estimate the data to be as good as desired. Interpolation is a mathematical method or function that assumes values in locations where data is not available [6]. This interpolation assumes that the data attributes are continuous in the area and they are also spatially dependent [7]. These assumptions indicate that the estimation and assumption of data attribute can be made based on the surrounding locations and that the values at adjacent points will be more similar than the values at the more distant points.

3. Rainfall Interpolation

Interpolation is a method or analysis that estimates the values in locations with no observed data [6] [8]. In the context of mapping, interpolation is a process of estimating values in regions that are not being sampled or measured for the purpose of mapping or distributing values across mapped regions [9]. The choice of spatial interpolation technique is very crucial in scientific research because there is no spatial interpolation method that always performs well everywhere or anytime.

In ArcGIS, this interpolation process can be done using either a 3D Analyst extension or a Spatial Analyst. The interpolation methods provided are various, including Spline, IDW and Kriging. Each of these methods uses different approaches and methodologies for assuming the value of data in a given location, which means that using the same data sample input, we will get different raster output when using Spline, IDW, Kriging and others.

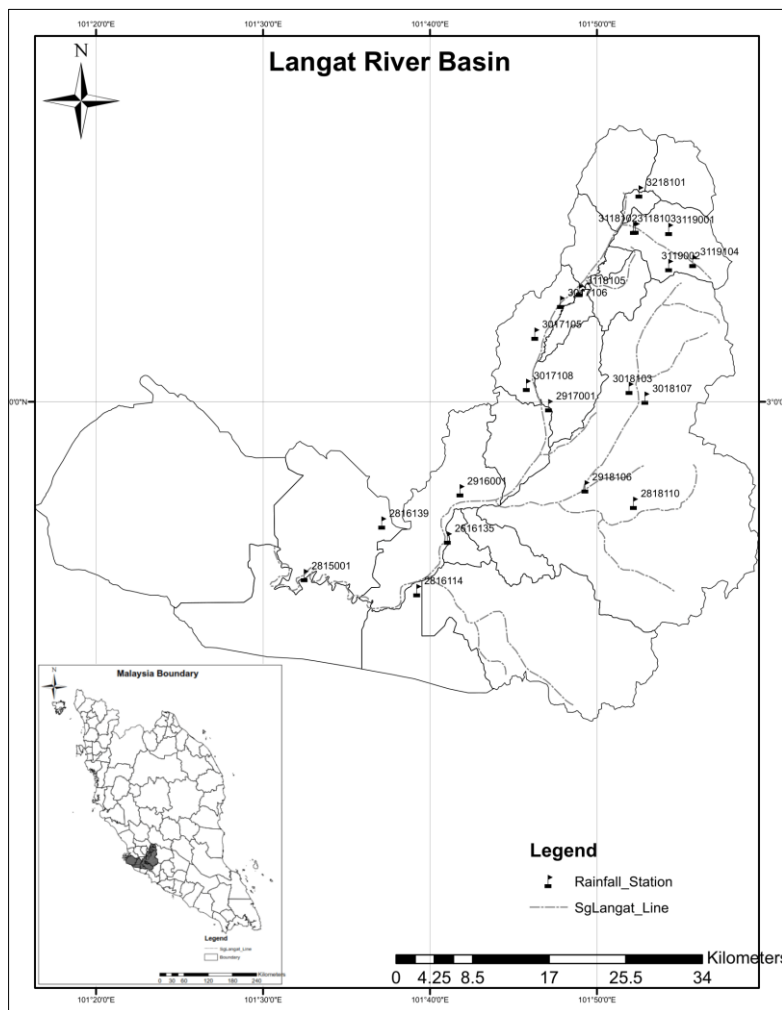


Figure 1. The location of Langat River Basin

The assumptions or predictions used are based on the principle of Spatial Autocorrelation in which the value of a data in a location is determined based on the relationship and distance between the sample point to the location. The output to be generated is in GRID raster data. There are many factors that will influence the selection of the appropriate interpolation method. No ultimate method that is fitted for all the problems. It all depends on the variable features and time of the variable is represented. All the methods have their own specialty and also good at certain analysis. Each method works differently for example, in a study by [4], and [10] show that IDW got the smallest value of RMSE and [11] stated that IDW gives the best result for precipitation, and maximum temperature was better handled by Spline. A study by [12] stated that Kriging got the highest value of Mean Absolute Deviation (MAD) and also [13] show that OK produce the lowest error and more accurate prediction. Due to the discrepancy in the results, further research is needed to obtain better interpolation methods.

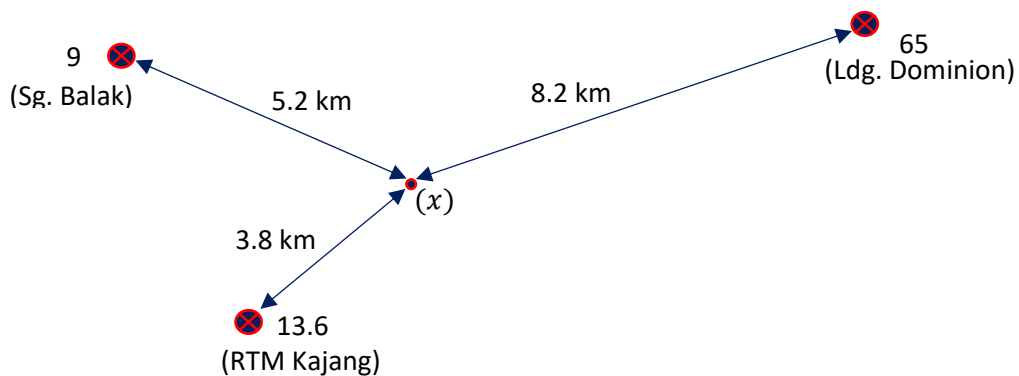
Interpolation is also a method of obtaining data based on some known data. Within the scope of mapping, interpolation is the process of estimating values in unstructured or measured regions, thus creating maps or distributing values across regions. In the interpolation method, there are definitely bias and error. Errors that are generated before interpolation can be due to errors in the method of data sampling, measurement error and analysis error in the field [14]. There are several interpolation methods commonly used in the data process, which each of the methods has its own characteristics and pro and cons. Interpolation methods include Natural Neighbour, Kriging, Spline and Inverse Distance Weighted (IDW).

3.1 Inverse Distance Weighted (IDW)

The Inverse Distance Weighted (IDW) method assumes that each input point has a relation with the distance [15]. The IDW interpolation method is generally influenced by the inverse distance obtained from the mathematical equations. The effect will be greater with the input point closer to the point resulting in a more detailed surface. But as the spacing increases, the detail decreases and it becomes smaller.

$$Z(x) = \frac{\sum w_i z_i}{\sum w_i} \quad (1)$$

example:



$$Z(x) = \frac{9/5.2^2 + 13.6/3.8^2 + 65/8.2^2}{1/5.2^2 + 1/3.8^2 + 1/8.2^2} = 18.5$$

with

W_i = Inverse square of distance ($1/\text{distance}^2$)

Z_i = Weight of a point differs

3.2 Kriging

The Kriging method is a stochastic estimation which quite similar to Inverse Distance Weighted (IDW) which uses linear combinations of weights to estimate values between sample data [16]. The assumption of this method is that the distances and orientations between the sample data show a significant spatial correlation in the interpolation results [17]. Kriging methods use a lot of computer systems in calculations. Calculation speed depends on the amount of sample data used and the coverage of the area under consideration.

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

with

$Z(s_i)$ = Value at i th location

λ_i = An unknown weight for the measured value at the i th location

s_0 = Predicted location

N = Number of measured value

4. Cross Validation

This study then performs cross-validation (CV) technique to compare which method is the better method in rainfall interpolation. The concept of CV is it removes one data location and predicts the associated data using the data at the rest of the locations. This tool will compare the observed value and predicted value to obtain the information in the model parameters [18]. The information that will obtain from the analysis is ME, RMSE, RMSSE and ASE.

4.1 Mean Error (ME)

Mean Error is the averaged difference between the measured and the predicted values.

$$\frac{\sum_{i=1}^n (\hat{Z}(s_i) - z(s_i))}{n} \quad (3)$$

4.2 Root Mean Square Error (RMSE)

Root Mean Square Error is to indicates how closely your model predicts the measured values. The smaller this error, the better.

$$\sqrt{\frac{\sum_{i=1}^n (\hat{Z}(s_i) - z(s_i))^2}{n}} \quad (4)$$

4.3 Root Mean Square Standardize Error (RMSSE)

Root Mean Square Standardized Error should be close to 1 if the prediction standard errors are valid. If the root-mean-squared standardized error is greater than 1, you are underestimating the variability in your predictions. If the root-mean-square-standardized error is less than 1, you are overestimating the variability in your predictions

$$\sqrt{\frac{\sum_{i=1}^n [(\hat{Z}(s_i) - z(s_i)) / \hat{\sigma}(s_i)]^2}{n}} \quad (5)$$

4.4 Average Standard Error (ASE)

and Average Standard Error is the average of the prediction standard errors.

$$\sqrt{\frac{\sum_{i=1}^n \sigma^2(s_i)}{n}} \quad (6)$$

5. Precipitation Data

Data from the monitoring station were obtained from the Department of Irrigation and Drainage Malaysia (DID). The data provide the coordinate for each station and also the type of the station. Data measurement was made along the stream in various locations. This data consists information of rainfall, water level, and also streamflow.

There are various methods of precipitation used in meteorological sciences. In Malaysia, two methods are used for the rainfall observation which is using conventional tools or automatic equipment. Conventional rain observation uses a tool known as rain gauges. Conventional rain gauges are 5 inches in diameter and 18 inches high [19]. Rain collection and readings are done daily.

There are several types of tools that are used for automatic observation such as weighing precipitating as well as tipping bucket. For rain detection, advance equipment that uses laser technology was used. Due to its economically viable prices, consistency of the data measured and simple calibration, tipping bucket are widely being used in Malaysia. The working principle of the tipping bucket includes a funnel receiver that will accumulate the rainwater in two containers in which the containers are positioned in a see saw position. After the collected amount of the rainwater reaches 0.2 mm, the mass of the water causes the container to tumble down and free up space. An electrical signal was sent every time the container was tilted down and this allows the amount of rainfall to be recorded over time. The maximum rainfall detectable per hour is 200mm for this very method [19].

6. Results and Discussion

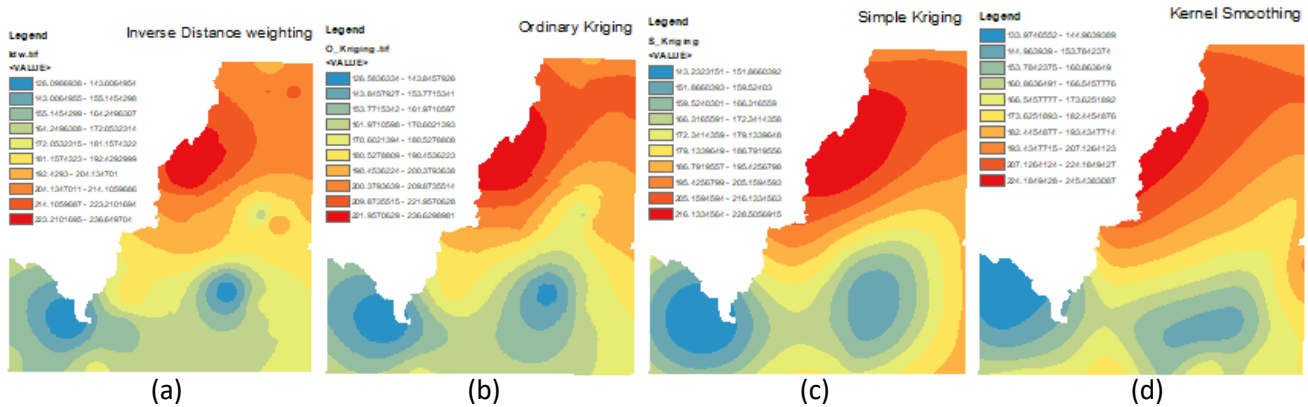


Figure 2. Rainfall maps obtained by the interpolation of 20 observations station (a) Inverse Distance Weighting, (b) Ordinary Kriging, (c) Simple Kriging, (d) Kernel Smoothing

The above result shows the precipitation map for each method used (Figure 2). From the map, we can see that the IDW method shows the smallest distribution area for the rainfall. The visual difference between the interpolation methods is presented in the maps of interpolated rainfall. IDW and Kriging method produced a map that appears quite similar. This because of the concept of Kriging and IDW that based on the surrounding measured values and assumes distance as a reflection of spatial correlation.

Table 1. Monthly error analysis of rainfall precipitation for each method

Method	Error	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Average	Total
IDW	Mean Error	-0.002	2.166	3.756	2.706	6.6	2.447	1.112	6.645	3.125	-1.121	3.301	-3.886	2.237417	26.849
	Root Mean Square Error	39.57	23.1	34.74	37	28.09	27.33	25.2	21.94	29.99	28.75	36.05	50.1	31.82167	381.86
	Root Mean Square Standardize Error														
	Average Standard error														
Ordinary Kriging	Mean Error	1.232	1.178	0.232	-1.501	-1.402	2.015	0.507	5.556	1.88	-1.815	3.613	-1.16	0.86125	10.335
	Root Mean Square Error	39.71	23.98	41.18	34.66	30.41	27.27	27.42	24.25	30.11	27.34	35.61	46.67	32.38417	388.61
	Root Mean Square Standardize Error	1.151	1.225	1.828	1.158	1.182	1.202	1.167	1.545	1.306	1.006	1.122	1.106	1.249833	14.998
	Average Standard error	33.408	18.98	27.48	31.03	25.77	23.76	21.94	18.51	22.39	27.13	30.3	41.87	26.88067	322.568
Simple Kriging	Mean Error	0.356	1.82	2.655	2.425	1.666	2.455	2.318	3.928	0.334	-0.589	2.692	0.054	1.676167	20.114
	Root Mean Square Error	34.72	21.33	35.16	37.39	36.12	27.19	26.26	21.96	27.82	26.42	34.37	41.07	30.8175	369.81
	Root Mean Square Standardize Error	0.994	1.015	0.957	0.867	2.08	0.922	0.994	0.834	1.038	0.964	0.945	0.99	1.05	12.6
	Average Standard error	34.96	20.41	36.14	40.42	20.43	29.2	25.41	24.75	26.81	27.58	35.81	41.49	30.28417	363.41
Kernel Smoothing	Mean Error	-5.2	-0.767	-8.44	-3.34	-1.354	0.646	2.354	-1.069	-0.545	-2.889	-2.034	-1.683	-2.02675	-24.321
	Root Mean Square Error	37.46	24.19	37.5	32.26	30.42	26.47	29.77	19.43	32.41	30.3	38.06	45.66	31.99417	383.93
	Root Mean Square Standardize Error	1.073	1.072	1.01	0.972	1.018	1.093	1.252	0.966	1.25	1.082	1.045	0.959	1.066	12.792
	Average Standard error	34.93	22.06	37.69	34.26	29.16	23.95	23.28	20.83	25.73	26.56	35.94	48.68	30.25583	363.07

By comparing the results of the ME, RMSE, RMSSE and ASE calculations of the four methods used, it can be seen that the SK method produces smaller error values than the other methods. The table above is the result of the error calculations on a monthly basis, that is, from January to December, indicating that the SK method is better than the OK, IDW and KS methods. This is because the values of ME, RMSE, RMSSE and ASE in the Simple Kriging method are always lower than the values in the OK,

IDW and KS methods. For IDW method, only the Mean Error and Root Mean Square Error results are available.

Even when testing the method, not all the data shows that the SK method is better than the others. It is because there are still some IDW, OK and KS error values smaller than SK as shown in Table 1. This can occur in months with relatively high rainfall intensity.

Table 2. Overall error analysis of rainfall precipitation for each method

	ME	RMSE	RMSSE	ASE
IDW	2.237417	31.82167		
OK	0.86125	32.38417	1.249833	26.88067
SK	1.676167	30.8175	1.05	30.28417
KS	-2.02675	31.99417	1.066	30.25583

From the result also we can see that SK method have the smallest RMSE value with 30.82 (Table 2). This can be reinforced by the RMSE SK boxplot results which are always shorter than the RMSE of IDW, OK and KS which means that in the SK method, the diversity or dissemination of observational data is smaller.

Table 3. RMSE quartile of rainfall precipitation for each method

	IDW	OK	SK	KS
Minimum	21.94	23.98	21.33	19.43
Q1	26.7975	27.3225	26.38	28.945
Median	29.37	30.26	31.095	31.34
Q3	36.2875	36.635	35.4	37.47
Maximum	50.1	46.67	41.07	45.66
Mean	31.82167	32.38417	30.8175	31.99417
Range	28.16	22.69	19.74	26.23

Boxplot is one of the ways in descriptive statistics to graphically draw from numerical data. Boxplot can also convey variation and placement or location information to predefined data, primarily to detect and visualize variations in location between different data sets. In Figure 3 below, the boxplot results show that the SK method is better than the other methods that can be seen from the box length in its vertical direction, which is the length of the SK box is not longer than IDW, OK and KS. From the result also we can see that SK method has the smallest mean value which means that the data dissemination is smaller than the other methods.

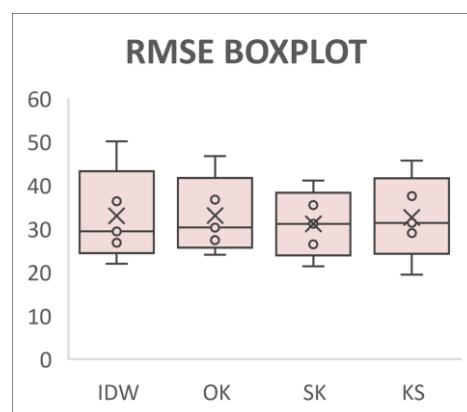


Figure 3. RMSE boxplot of rainfall precipitation for each method

7. Conclusion

It is almost impossible to directly measure continuous spatial data due to limitations in data collection. Therefore, the interpolation process is required in order to determine a new value or point on a surface from a given value or point. This method also used to predict unobservable variables in a variety of locations. The value of the variable was to be predicted using the probabilistic function that applied in the interpolation method.

From the research done through ME, RMSE, RMSSE and ASE monthly analysis, RMSE boxplot and Quartile table, it can be seen that the error value for the Kriging method is never greater than the IDW and KS method. This indicates that the error rate in the Kriging method is smaller than the IDW and KS method so that the Kriging method has a better accuracy rate especially the SK method. After various tests, it is known that the SK method is better than the IDW, OK and KS methods. Although there are certain months that the IDW method is better than the Kriging method, this is extremely rare.

However, these results are limited for rainfall interpolation in Langat River Basin area only. The other method also has its own specialty in the interpolation process not only for the rainfall but for other cases like water quality, river discharge, fuel consumption and so on.

Reference

- [1] Soliman M M 2010 *Engineering hydrology of arid and semi-arid regions*: CRC Press)
- [2] F. Ahmad, Tomoki Ushiyama and Sayama T 2017 Determination Of Z-R Relationship And Inundation Analysis For Kuantan River Basin *Malaysian Meteorological Department (MMD), Ministry of Science, Technology and Innovation (MOSTI). Volume 2*
- [3] Ibrahim A, Basarudin H, Ramli A, Yaakop M, Sulaiman M and Abu M A 2019 Development of rain gauge using weight measurement for microwave network. In: *AIP Conference Proceedings*: AIP Publishing) p 020126
- [4] Kurniadi H, Aprilia E, Utomo J B, Kurniawan A and Safril A 2018 Perbandingan Metode IDW dan Spline dalam Interpolasi Data Curah Hujan (Studi Kasus Curah Hujan Bulanan di Jawa Timur Periode 2012-2016). Seminar Nasional GEOTIK 2018)
- [5] Fung K F, Huang Y F, Koo C H and Mirzaei M 2019 Improved SVR machine learning models for agricultural drought prediction at downstream of Langat River Basin, Malaysia *Journal of Water and Climate Change*
- [6] Hu Q, Li Z, Wang L, Huang Y, Wang Y and Li L 2019 Rainfall Spatial Estimations: a review from spatial interpolation to multi-source data merging *Water* **11** 579
- [7] Li J and Heap A D 2008 A review of spatial interpolation methods for environmental scientists
- [8] Kamińska A and Grzywna A 2014 Comparison of deterministic interpolation methods for the estimation of groundwater level *Journal of Ecological Engineering* **15** 55--60
- [9] Lambie J M, Dahl J, Kennel J, Tonkin M, Karanovic M and Huber W A 2019 Simultaneous multi-event universal kriging methods for spatio-temporal data analysis and mapping. Google Patents)
- [10] Walter M 2001 MAPPING PRECIPITATION: ANALYSIS OF GIS INTERPOLATION TECHNIQUES. ed T U Department of Civil and Environmental Engineering
- [11] Hartkamp A D, De Beurs K, Stein A and White J W 1999 Interpolation techniques for climate variables. Cimmyt)
- [12] Fridayani N M S, KENCANA I P E N and SUKARSA K G 2012 Perbandingan Interpolasi Spasial Dengan Metode Ordinary dan Robust Kriging pada Data Spasial Berpencilan (Studi Kasus: Curah Hujan di Kabupaten Karangasem) *E-Jurnal Matematika*
- [13] Mair A and Fares A 2010 Comparison of rainfall interpolation methods in a mountainous region of a tropical island *Journal of Hydrologic Engineering* **16** 371-83
- [14] Chen T, Ren L, Yuan F, Yang X, Jiang S, Tang T, Liu Y, Zhao C and Zhang L 2017 Comparison of spatial interpolation schemes for rainfall data and application in hydrological modeling *Water* **9** 342
- [15] Chen F-W and Liu C-W 2012 Estimation of the spatial rainfall distribution using inverse distance weighting (IDW) in the middle of Taiwan *Paddy and Water Environment* **10** 209-22

- [16] S.M. Gandhi and B.C. Sarkar 2016 Chapter 12 - Geostatistical Resource/Reserve Estimation *Essentials of Mineral Exploration and Evaluation*,
- [17] Gumiaux C, Gapais D and Brun J 2003 Geostatistics applied to best-fit interpolation of orientation data *Tectonophysics* **376** 241-59
- [18] Blanchet J, Paquet E, Vaittinada Ayar P and Penot D 2019 Mapping rainfall hazard based on rain gauge data: an objective cross-validation framework for model selection *Hydrology and Earth System Sciences* **23** 829-49
- [19] DID D o I a D M 2018 *HYDROLOGICAL STANDARD FOR RAINFALL STATION INSTRUMENTATION* (National Library Malaysia: National Library Malaysia)