

Analysis of rainfall distribution in Kelantan river basin, Malaysia

Faizah Che Ros^{1*} and Hiroyuki Tosaka²

¹Department of Environmental Engineering & Green Technology, Malaysia-Japan International Institute of Technology, Universiti Teknologi Malaysia Kuala Lumpur, 54100 Kuala Lumpur, Malaysia

²Geosphere Environmental Technology Ltd. (GET) NCO Kanda-Awajicho 3rd Floor, 101-0063 2-1 Awajicho, Chiyoda-ku, Tokyo, Japan

Abstract. Using rainfall gauge on its own as input carries great uncertainties regarding runoff estimation, especially when the area is large and the rainfall is measured and recorded at irregular spaced gauging stations. Hence spatial interpolation is the key to obtain continuous and orderly rainfall distribution at unknown points to be the input to the rainfall runoff processes for distributed and semi-distributed numerical modelling. It is crucial to study and predict the behaviour of rainfall and river runoff to reduce flood damages of the affected area along the Kelantan river. Thus, a good knowledge on rainfall distribution is essential in early flood prediction studies. Forty six rainfall stations and their daily time-series were used to interpolate gridded rainfall surfaces using inverse-distance weighting (IDW), inverse-distance and elevation weighting (IDEW) methods and average rainfall distribution. Sensitivity analysis for distance and elevation parameters were conducted to see the variation produced. The accuracy of these interpolated datasets was examined using cross-validation assessment.

1 Introduction

Assessing rainfall variability is the key element to develop any conceptual and predictive models in many fields such as water resource management and flood disaster prediction [1]. Using rainfall gauge on its own as input carries great uncertainties regarding runoff estimation, especially when the watershed area is large and the rainfall is measured and recorded at irregular spaced gauging stations [2-3]. Therefore, spatial interpolation is the key to obtain continuous and orderly rainfall distribution for the input of distributed and semi-distributed numerical modelling. Few studies [4-6] indicated that using mean areal rainfall instead of spatially distributed rainfall inclines to undervalue the volumes and peak runoffs. However, the impact of distributed rainfall on runoff estimation is complex as it can be dependent on the rainfall's nature, catchment characteristics and the spatial scale used [7-8].

* Corresponding author: crfaizah@utm.my

The estimation of rainfall distribution can be grouped into two; indirect and direct method. Indirect rainfall estimation via satellite products, especially from radar remote sensing is increasingly used as covariates since they offer spatially detailed data information [3, 9]. The direct ground-based method has a wide range of interpolation techniques ranging from simple to more complex calculations, which can be divided into two main groups; deterministic and geostatistical. The mostly known deterministic approaches are Thiessen polygon and inverse distance weighting (IDW) whereas the geostatistical methods constitutes a discipline of mathematics and earth sciences, known as kriging method.

This study aims to investigate the IDW method in estimating rainfall distribution of Kelantan watershed by considering the spatial effect (large vs small scale) and by integrating elevation data into IDW, via inverse distance elevation weightage (IDEW) method. Sensitivity analysis for distance and elevation parameters were conducted to see the variation produced. The accuracy of these interpolated datasets will be examined by cross-validation assessment.

2 Materials and methods

2.1 Study area

In this study, the large Kelantan watershed was chosen as the main research area. Che Ros et al. [10] has conducted a homogeneity study on fifty rainfall stations. Out of fifty gauges, forty six were found homogenous. Thus, a total of forty six rainfall stations and their daily time-series were used to interpolate gridded rainfall surfaces using IDW and IDEW. The rainfall stations are managed by Department of Irrigation and Drainage Malaysia (DID). The IDW and IDEW were tried twice each by: a) applying on whole watershed and b) sub-watershed scale. So, a total of 4 rainfalls distribution gridded surface was produced in one space-time rainfall series for year 2007.

2.2 Methods

In this study, the Kelantan watershed was delineated into eight sub-watersheds as shown in Fig. 1 to develop the sub-areal based of IDW and IDEW, namely, Areal IDW (AIDW) and Areal IDEW (AIDEW). The delineation was based on the hydrological characteristics such as stream network and waterlines, and the manner of rainfall based on reliable long-term records.

2.2.1 Inverse Distance Weighting Method (IDW)

Inverse distance weighting method (IDW) is a famous conventional deterministic approach for spatial interpolation based on similarity or smoothness within a research area other than Thiessen method. The assigned values to unknown points are calculated as a weighted average of the values available at the known points [11]. Weight increases as the distance decreases from the known points to unknown points [2, 12]. The mathematical form of the IDW is given below:

$$p_e = \sum_{i=0}^N \lambda_i p_i \tag{1}$$

where p_e is the interpolated rainfall for a grid cell; N is the number of gauges that are used in the interpolation for the current grid cell; λ_i is the corresponding weight of each known points where it reduces as the distance between known points and predicted points increases; p_i is the rainfall value in mm/day of the i th gauge station. The weighting factors is called Shepard method [13] and determined by:

$$\lambda_i = \frac{d_{i0}^{-\alpha}}{\sum_{i=1}^N d_{i0}^{-\alpha}} \tag{2}$$

$$\sum_{i=1}^N \lambda_i = 1 \tag{3}$$

where the power of exponent α controls the influence of the distance among the sample points or gauge stations on the interpolations results. A low α leads to a greater weight towards a grid point value of rainfall from remote rain gauges. As the α decreases toward zero, the method approximates the Thiessen method [14-17]. Dirks et al. [15] had concluded, to minimize the interpolation errors, α of 2 is used for daily and monthly steps, 3 for hourly and 1 for yearly. However, most study [2, 12] set α to 2 where inverse square distances were used in their estimation. During the process of prediction, the weights of gauge station are proportional and the sum of these weights is equal to 1.

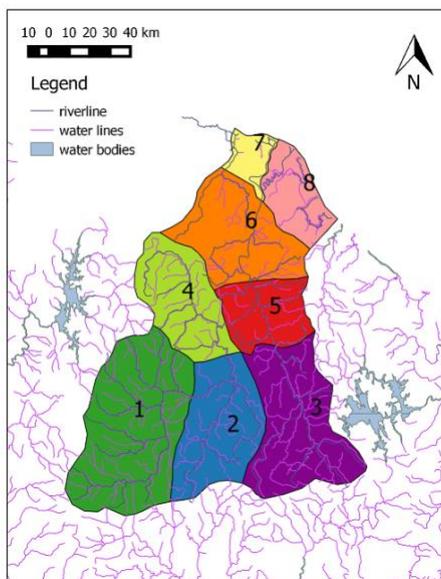


Fig. 1. Kelantan watershed divided into 8 sub-watersheds.

2.2.2 Inverse Distance and Elevation Weighting Method (IDEW)

The IDW interpolation only incorporates a single influence factor, which is horizontal distance. Therefore, it is not suitable to be applied in an area with abrupt changes in elevation, which could create uncertainty in estimating unknown information [27]. Inverse distance and elevation weighting (IDEW) technique provides more suitable results for mountainous regions where topographic impacts on rainfall are important [26,29]. IDEW algorithm is part of hydrological data processing software called HyKit developed by UNESCO-IHE [18]. However, in this study, C programming was used to do the interpolation by adopting the equations as shown below:

$$P_s = W_D \sum_{i=1}^N \frac{1}{D} w(d)_i P_i + W_Z \sum_{i=1}^N \frac{1}{Z} w(z)_i P_i \quad (4)$$

where W_D and W_Z are the importance factors for distances and elevations. Typical values for the importance factor are 0.8 for distance and 0.2 for elevation, respectively [19]. Similarly, $w(d)_i$ and $w(z)_i$ are the individual gauge weighting factors for distance and elevation respectively. Whereas D and Z are the normalization quantities given by the sum of individual weighting factors $w(d)$ and $w(z)$, respectively for all the gauges used in the grid interpolation. The weighting factors of $w(d)_i$ and $w(z)_i$ can be determined by:

$$w(d) = 1/d^\alpha \text{ for } d > 0 \quad (5)$$

$$w(z) = \begin{cases} 1/z^\beta & \text{for } z \leq z_{\min} \\ 1/z^\beta & \text{for } z_{\min} < z < z_{\max} \\ 0 & \text{for } z \geq z_{\max} \end{cases} \quad (6)$$

where d is the distance in kilometer between current grid and the gauge station used for interpolation, z is the absolute elevation difference in meter between the current grid cell and the gauge station used for interpolation, α and β are exponent factors for distance and elevation weightings where the distance weighting exponent is 2 and 1, respectively [19]. z_{\min} and z_{\max} are the minimum and maximum limiting values of elevation differences for computing elevation weightings [18, 19]. Limitations on z_{\min} and z_{\max} help to avoid dominance of the stations having very small elevation difference and enables data point inclusion to be restricted to a local elevation range [18]. Typical values of z_{\min} are 100 to 300 meters while z_{\max} from 500 to 2,500 m depending on the study area [18].

2.2.3 Cross-Validation & Performance Assessment

The performance of the interpolation methods was assessed and compared by cross-validation procedure. The procedure consists of temporarily discarding one observation value at a time from the data sets and re-estimating the discarded value from the remaining sampled point using each interpolation method. Nash-Sutcliffe efficiency (NSE) [20] and root mean square error (RMSE) [21] were adopted to evaluate the performance of interpolation methods. NSE criterion is a form of normalized least squares where perfect agreement between the observed and estimated values yields an efficiency of 1, whereas a

negative efficiency represents lack of agreement where the observed mean is a better predictor than the estimation value. RMSE indicates the deviation between estimated values and observed one hence, smaller RMSE contributes to better estimation capability.

3 Results and discussions

3.1 Cross-validation and performance assessment

The results of four interpolation methods were validated and ranked via RMSE and NSE as shown in Table 1. The RMSE validation indicates that each interpolation method has relatively similar deviations between estimated and observed values with a mean residual of 10.175 ± 0.233 . It is also observed that the range of deviation values is significantly large for all interpolation methods, ranging from 0.05 to 78.5. This shows that despite similar deviation, the confidence level between estimation and observation value is not as high as the external factors need to be considered, such as the quality of the rainfall data. Hence, RMSE is not able to determine the best interpolation method in this case as illustrated in Fig. 2.

Table 1. Cross-validation assessment performance and ranking of different interpolation methods based on root mean square error (RMSE) and Nash-Sutcliffe efficiency (NSE).

Interpolation Methods	Range RMSE	Mean RMSE	Rank of Mean RMSE	Range NSE	Mean NSE	Rank of Mean NSE
IDW	0.05-75.42	9.84	2	-0.87-0.76	0.09	4
IDEW	0.07-75.08	9.82	1	-0.28-0.99	0.93	2
AIDW	0.06-78.46	10.60	4	0.6-0.99	0.98	1
AIDEW	0.07-75.37	10.44	3	-0.97-0.99	0.92	3

On the other hand, the NSE validation shows that the IDW interpolation has the lowest mean value of NSE close to zero, while the rest of interpolation methods demonstrate a high mean value of NSE close to 1. Fig. 3 illustrates the NSE cloud distributions for all interpolation methods and it can be clearly seen that the NSE value for IDW is much scattered within a wide range, both positive and negative efficiency, as compared with others. This shows that the quality of rainfall estimation in Kelantan can be improved by incorporating elevation and smaller sub-watershed in the interpolation. This is obvious since the geophysical and topography of Kelantan is mixed with mountainous terrain and valley. It is also worth to note that integrating elevation in a smaller sub-watershed, i.e. AIDEW, did not significantly change the NSE value of AIDW. This could be due to the smaller elevation change that exists within the sub-watershed; hence elevation gives little impact in estimating the rainfall in smaller watershed. Overall, AIDW shows the best interpolation method to estimate the rainfall in entire Kelantan watershed according to NSE.

3.2 Sensitivity analysis of rainfall distribution parameter

Sensitivity of rainfall distribution parameters was investigated to see variations in the pattern and amount of rainfalls produced by changing the α and z_{min} and z_{max} . Sensitivity analysis of α parameter as shown in Fig. 4 shows that the greater of α will result in greater influence on area or grid that close to the interpolated points, whereas smaller α will result

in interpolated values will be dominated by other points or rainfall gauges that are far away. This can be seen in day 341, year 2007 where when α equals to 1.5, weights of rainfall gauge that has high rainfall records in the east part was lesser compared with α equals 2 and 3. Same goes to day 346, year 2007 where when α equals to 3, both high rainfall and zero rainfall area gave greater influence to surrounding nearby grid compared with 2 and 1.5. Greater α value also results in greater amount of rainfall distributed over Kelantan watershed even though the number of rainfall gauges that have high rainfall value is small. For IDEW elevation parameter, only z_{min} was adjusted since z_{max} for Kelantan watershed is known (around 1800m above sea level). Masih et al [29] suggested the range of z_{min} is around 100 to 300 m. Fig. 6 illustrates the difference by applying z_{min} of 100 and 200 m. This shows that larger range of Δz leads to higher amount of rainfall across the watershed.

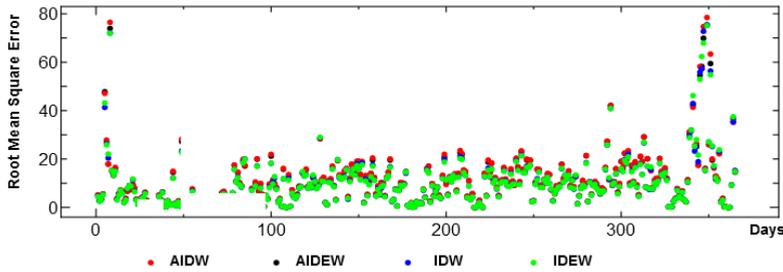


Fig. 2. RMSE clouds from 1st January to 31st December 2007.

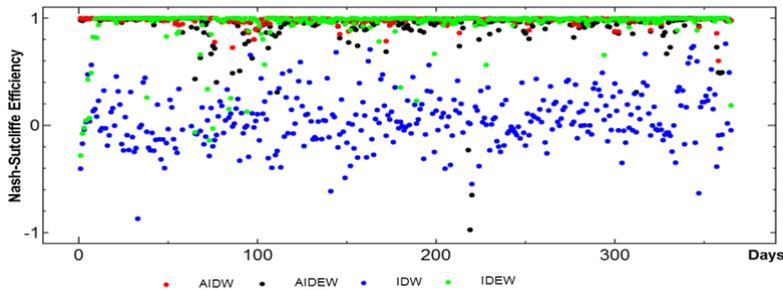


Fig. 3. NSE clouds from 1st January to 31st December 2007.

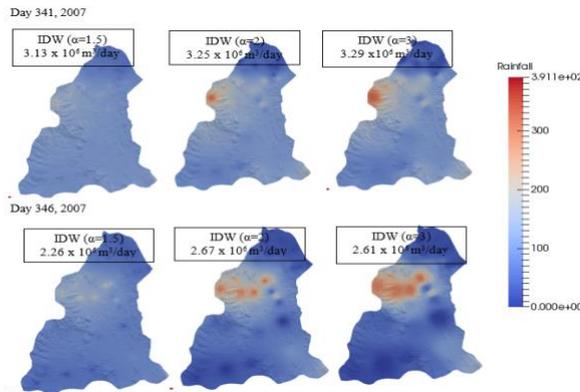


Fig. 4. Sensitivity Analysis of IDW parameter α .

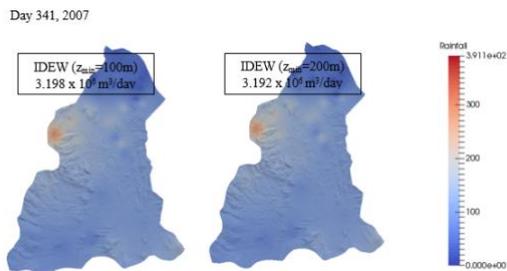


Fig. 5. Sensitivity Analysis of IDEW parameter z_{min} ($\alpha = 2$ and $z_{max} = 1800m$).

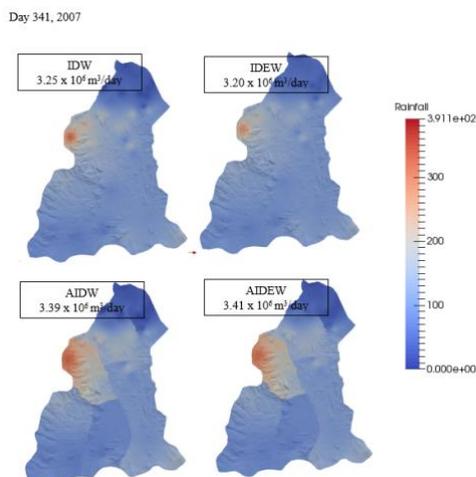


Fig. 6. Sensitivity analysis of rainfall distribution types. ($\alpha = 2$, $z_{min} = 200$ and $z_{max} = 1800m$).

4 Conclusions

Rainfall distributions over the large Kelantan watershed were developed on daily basis. Up to forty six rainfall stations and their daily time-series were used to interpolate gridded rainfall surfaces using inverse distance weighting (IDW), inverse distance and elevation weighting (IDEW). The IDW and IDEW were tried twice each by: a) applying on whole watershed and b) sub-watershed scale. So, a total of five rainfall distribution gridded surface was produced in one space-time rainfall series together with average rainfall distribution. The sensitivity analysis of rainfall distribution parameters was done to see variations in rainfall pattern and amount of rainfall produced by changing α and z_{min} . From the analysis, a low α leads to a greater weight towards a grid point value of rainfall from remote rain gauges. As α decreases toward zero, the method approximates the Thiessen method, whereas larger range of elevation leads to higher amount of rainfall produced across the watershed. For elevation parameter, larger range of Δz leads to higher amount of rainfall across the watershed.

The authors acknowledge the cooperation of the Department of Drainage and Irrigation Malaysia (DID) for providing the data required for this study.

References

1. D. R Dawdy, J. M. Bergman, *Water Resour. Res.* **5**, 958-966 (1969)
2. S. Ly, C. Charles, A. Degre, *Hydrol. Earth Syst. Sci.* **15**, 2259-2274 (2011)
3. P. D. Wagner, P. Fiener, F. Wilken, S. Kumar, K. Schneider, *J. Hydrol.* **464**, 338-400 (2012)
4. P. Arnaud, C. Bouvier, L. Cisner, R. Dominguez, *J. Hydrol.* **260**, 216-230 (2002)
5. J. M. Schuurmans, M. F. P. Bierkens, *Hydrol. Earth Syst. Sci.* **11**, 677-693 (2007)
6. Y. Trambly, C. Bouvier, P.-A., Ayrat, A. Marchandise, *Nat. Hazards Earth Syst. Sci.* **11**, 157-170 (2011).
7. M. L. Segond, H. S. Wheeler, C. Onof, *J. Hydrol.* **347**, 116-131 (2007)
8. T. Tao, B. Chocat, S. Liu, K. Xin, *J. Env. Protection* **1**, 50-58 (2009)
9. A. Verworn, U. Haberlandt, *Hydrol. Earth Syst. Sci.* **15**, 569-584 (2011)
10. F. Che Ros, H. Tosaka, L. M. Sidek, H. Basri, *Int. J. River Basin* **14**, 151-163 (2016)
11. R. Franke, *Math Computations*, **33**, 181-200 (1982).
12. W. Xu. Y. Zou, G. Zhang, M. Linderman, *Int. J. Climatol.* **35**, 2898-2907 (2014)
13. D. Shepard, *ACM Annual Conference/Annual Meeting*, 517-524 (1968)
14. A. Mair, A. Fares, *J. Hydrol. Eng.* **16**, 371-383 (2011)
15. K. N. Dirks, J. E. Hay, C. D. Stow, D. Harris, *J. Hydrol.* **208**, 187-193, (1998)
16. M. Ashraf, L. C. Loftis, K. G. Hubbard, *Agric. For Meteorol.* **84**, 255-271 (1997)
17. S. Ly, C. Charles, A. Degre, *Biotechnol. Agron. Soc. Environ* **17**, 392-406 (2013)
18. I. Masih, S. Maskey, S. Uhlenbrook, V. Smakhtin, *J. Am. Water Resour. Assoc.* **47**, 179-195 (2011)
19. C. Daly, W. P. Gibson, G. H. Taylor, L. G. Johnson, P. Pasteris, *Climate Research* **22**, 99-113 (2002)
20. J. E Nash, J. V. Sutcliffe, *J. Hydrol* **10**, 282-290 (1970)
21. S. H. Hwang, D. H. Ham, J. H. Kim, *Hydrol. Sc. J.* **57**, 1257-1274 (2012)