Comparison of Methods to Estimate Missing Rainfall Data for Short Term Period at UMP Gambang

To cite this article: A A G Nadiatul Adilah and H Hannani 2021 *IOP Conf. Ser.: Earth Environ. Sci.* **682** 012027

Comparison of Methods to Estimate Missing Rainfall Data for Short Term Period at UMP Gambang

A A G Nadiatul Adilah* and H Hannani

Civil Engineering Department, College Engineering, Universiti Malaysia Pahang, 26300 Kuantan, Pahang, Malaysia

*Corresponding Author: nadiatul@ump.edu.my

Abstract. Hydrological analysis is crucial for planning or designing engineering structures and water resources system as stated by World Meteorological Organization (WMO). A long and complete data set of rainfall is required in order to carry out the hydrological study successfully. However, the most common problem that always come up in data collection is the lacking of data due to some occurrence involving the equipment installed that may cause the error to occur. For this study, another problem that has been a concern is the availability of the rainfall data is insufficient as the studied area is newly developed so the limitation of the existing data may affect to the relevance of this study. Therefore, this study is mainly conducted to compare three existing methods from previous scholars which are Arithmetic Mean Method, Normal Ratio Method and Inverse Distance Weighting Method to prove whether these methods are capable to fill the gaps of rainfall data for short term period of study. The existing data collected at the target station are then compared with the three methods whether the results gained are similar or otherwise. Analysis shows that the study for short term period is irrelevant due to the insufficiency of data collection.

1. Introduction

In meteorological and hydrological researches, missing rainfall data has always been one of the most challenging problems which need to be faced by the researchers. The problems of missing rainfall data are due to the wrong technique used when measuring the rainfall, relocation of the rain station and malfunctioned of instrument [1]. The "presence" of gaps in environmental data time series represents a very common, but extremely critical problem, since it can produce biased results and in the worst case, it can even prevent important analyses of the considered variables from being carried out [2]. For performing the effective rainfall analysis, it is essential to estimate the missing value in rainfall series [3]. It is often found that the rainfall stations might recorded short breaks or missing data that believed are due to the instrumental errors or the absence of data collector. The missing data however may affect to the reliability of the rainfall data recorded for the study by future researchers. As a procedure suggested by U.S. Weather Bureau, the missing rainfall data from any particular station may be estimated from the observations of the rainfall from some other stations as close to and as evenly spaced around the station with the missing record as possible. The station which has the missing data is called as target station whereas the nearby gauging stations which data are used to calculate the missing station data are called neighbouring stations. It is observed that there are missing data from time to time throughout the collected years. Therefore, the missing data has to be solved in order to improve the reliability of the recorded rainfall data at UMP Gambang campus to be used for analysis by future researchers. Most studies from previous researchers [2, 4, 5 and 6] have chosen a station of a minimum of 30 years of data as recommended by the World Meteorological Organization (WMO) in

2011 to be the target station. However, for this study, the data available of the target station is only of short-term period of less than five (5) years which may or may not affect the success of the estimation of missing rainfall data. The methods that are selected to retrieve missing rainfall data for this study are Arithmetic Mean Method, Normal Ratio Method and Inverse Distance Weighting Method. These methods are selected since they are usually used by another scholars in the previous studies to estimate missing rainfall data at a particular target station based on the available rainfall values recorded at the neighbouring stations. Moreover, these two methods are found to be very useful in the case where the neighbouring stations are very close and highly correlated with the target station.

1.1 Objectives of Study

i. To estimate the missing rainfall from the data collected at UMP Gambang
ii. To test the reliability of the methods used in filling data gaps for short term period of study

1.2 Scopes of Study

i. This study is focusing on finding the missing daily rainfall from the collected data
ii. The rainfall data collected for this study is from the rain gauge at Kolej Kediaman 2 (KK2), UMP Gambang and nearby hydrological stations from Department of Irrigation and Drainage (DID) Malaysia
iii. The estimated missing rainfall data from nearby hydrological stations of UMP Gambang will be compared with the existing data at UMP Gambang
iv. Several statistical methods will be implemented to retrieve and recover the missing rainfall data

2. Study Area

UMP Gambang has been selected as the study area for this research to be done since there is an existing rain gauge station installed at Kolej Kediaman 2, UMP Gambang. This station then will become the target station that has missing rainfall data before the estimation of rainfall data can be done by selecting the neighbouring stations for temporal pattern analysis. Figure 1 shows the location of rain gauge that has been installed at Kolej Kediaman 2, UMP Gambang.

Figure 1. Rain gauge located at UMP Gambang

Nine nearby hydrological stations with a radius of 50 km to the target station are selected to be the neighbouring stations. This distance is considered the best and most optimal for areas in Peninsular Malaysia [4]. This consideration is also done based on the usual situation where there is lacking number of installed hydrological stations nearby the target station to be analysed for this kind of study. The selection of neighbouring stations is also done by considering the availability of rainfall data of a period 2015 to 2020 that can be used for this study. The list of selected neighbouring stations along

with their distance in km and Euclidean distance to the target station are shown in Table 1. Euclidean distance is the straight-line distance between two points in Euclidean space.

Table 1. The neighbouring stations with their location and distance to the target station

Station's Name	Latitude	Longitude	Euclidean Distance (km)
Kolej Kediaman 2, UMP Gambang	**3.73**	**103.13**	
Kg. Pulau Manis	3.65	103.12	0.078 (8.7)
Paya Besar, Kuantan	3.77	103.28	0.161 (17.7)
Ldg. Kuala Reman	3.90	103.13	0.169 (18.8)
Rancangan Pam Paya Pinang	3.84	103.26	0.173 (19.2)
Kg. Sg. Soi	3.73	103.30	0.175 (19.4)
Ldg. Nada	3.91	103.11	0.179 (19.9)
Sg. Lembing P.C.C.L. Mill	3.92	103.04	0.206 (22.9)
Pejabat JPS Negeri Pahang	3.81	103.33	0.218 (24.3)
Balok, Kuantan	3.94	103.38	0.335 (37.2)

Data that is collected at the target station is of year 2015 to 2020 since the rain gauge at the study area is recently installed for previous studies. For the data of the neighbouring stations, however, due to some circumstances, the available data that can be used are only of year 2015 and 2017. All the rainfall data for the neighbouring stations is collected from Department of Irrigation and Drainage Malaysia.

3. Methods

Three different methods for estimating missing data are used to evaluate the suitability of each method to estimate missing data of short term period. The three methods that related to deterministic methods are listed below.

3.1. Arithmetic Mean Method

If the normal annual rainfalls at surrounding gauges are within 10% of the normal annual precipitation at the stations concerned, then the arithmetic procedure could be adopted to estimate the missing data [7]. This assumes equal weights from all nearby rain gauge stations and uses the arithmetic mean of the precipitation data.

$$p_x = \frac{1}{m}\sum_{i=1}^{m} p_i \qquad (1)$$

3.2. Normal Ratio (NR) Method

This method is used if the normal annual precipitation of any surrounding gauges exceeds 10% of the gauge that is under consideration. This weighs the effect of each surrounding station [8].

$$p_x = \frac{1}{m}\sum_{i=1}^{m} \left(\frac{N_x}{N_i}\right) p_i \qquad (2)$$

3.3. Inverse Distance Weighting (IDW) Method

In this method, the weight for each station is assumed to be inversely proportional to its squared distance of the target station from the neighbouring station with data [9].

$$p_x = \frac{\sum_{i=1}^{m} \frac{1}{d_i^2} p_i}{\sum_{i=1}^{m} \frac{1}{d_i^2}} \qquad (3)$$

p_x = Estimate for the target station (X)
p_i = Rainfall values of rain gauges used for estimation
m = Number of surrounding stations

N_y = Normal annual precipitation of the X station
N_i = Normal annual precipitation of the surrounding stations
d_i = Distance from each location to the point being estimated

Another method that is used to determine the suitability of the three methods to estimate missing data is by comparing the calculated data using these three methods with the actual existing data of the target station.

4. Results and Discussions

The Arithmetic Mean method could not be applied to estimate the missing rainfall at the target station since the average annual rainfalls at their neighbouring stations are not within the 10% range of the normal annual rainfall of the target station. This means that the annual rainfall values can be significantly different among the neighbouring stations that is most probably happened because they are far from one another. Since the available data of the target station exceeded 10% of the normal annual precipitation of its surrounding gauges, NR method is relevant to be used. It is assumed that the weight for each station to be inversely proportional to the squared distance of the target station to allow the IDW method to be used. The results for the estimation of missing rainfall data are presented in scatter graph to clearly show the comparison of the resulted daily rainfall value in mm for every month for both NR and IDW methods. The choose of rainfall data from the neighbouring stations to be used for analysis to estimate missing rainfall is based on the absence of daily rainfall recorded from the collected rainfall data at the target station. Figures 2 - 4 show the missing rainfall values estimated using both NR and IDW methods of January, February and October 2015.

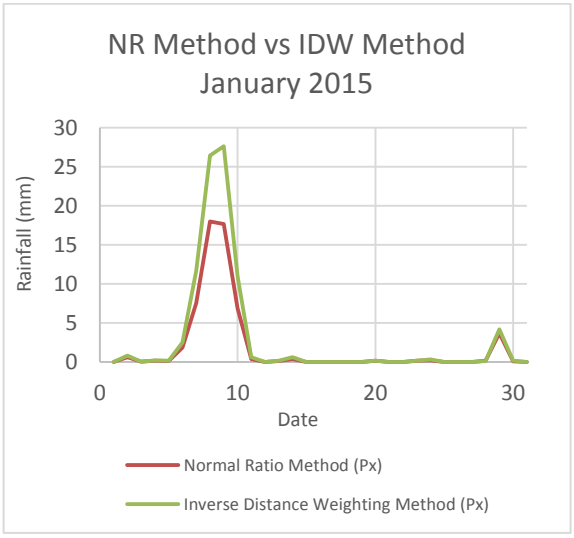

Figure 2. Difference between the estimation of missing rainfall data using NR method and IDW method of January 2015

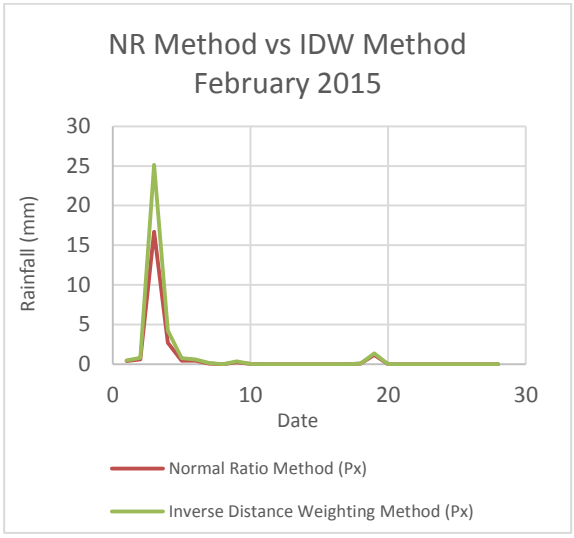

Figure 3. Difference between the estimation of missing rainfall data using NR method and IDW method of February 2015

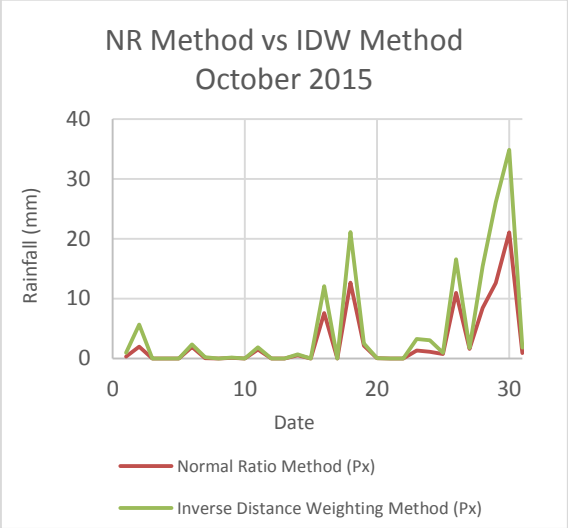

Figure 4. Difference between the estimation of missing rainfall data using NR method and IDW method of October 2015

The outcome from the estimation of missing rainfall value using NR method and IDW method of January, February and October 2015 show significant dissimilarity where every estimation shows huge difference of result for missing rainfall estimation for every analysed month. From Figures 2 - 4, it is observed that IDW method always estimated higher value of rainfall to compare with NR method. Figures 5 - 7 moreover show the next comparison of missing rainfall estimation using NR method and IDW method of January, February and March 2017 respectively.

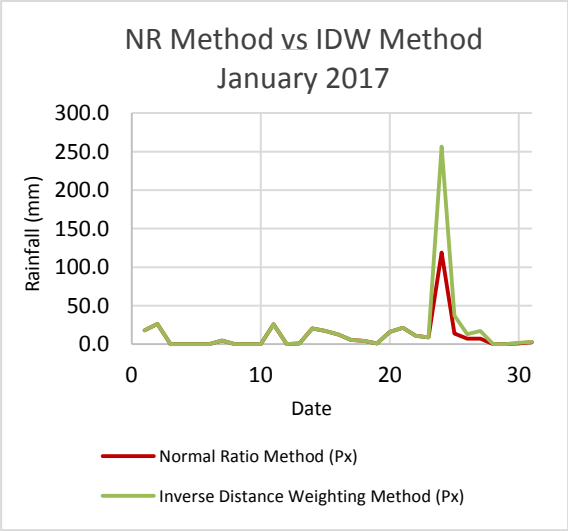

Figure 5. Difference between the estimation of missing rainfall data using NR method and IDW method of January 2017

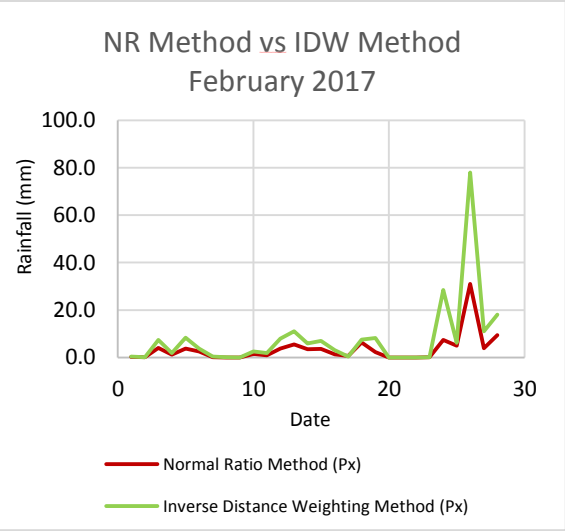

Figure 6. Difference between the estimation of missing rainfall data using NR method and IDW method of February 2017

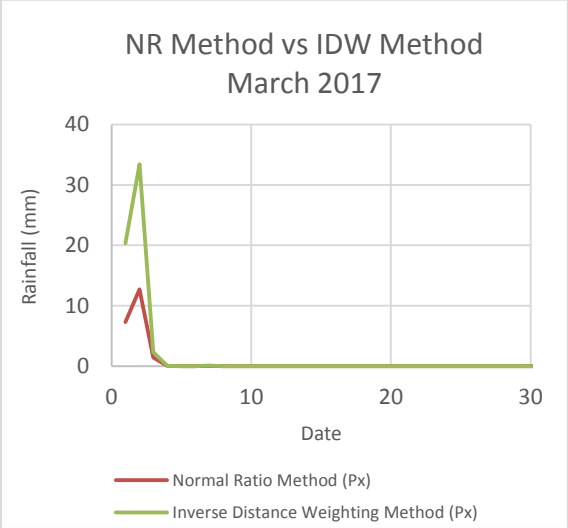

Figure 7. Difference between the estimation of missing rainfall data using NR method and IDW method of March 2017

The missing rainfall data estimation in year 2017 using NR method and IDW method as presented in Figure 3, Figure 4 and Figure 5 show the same pattern with the rainfall data estimated in year 2015 where the two methods estimated different values of rainfall data and not even close to one another. IDW method also estimated higher value compare to NR method. This clearly proves that the NR method and IDW method will not produce the same outcome of the estimation of missing rainfall data despite the change of month where the collected data at the target station contain gaps for each year respectively. The latter presented graph that is shown in Figures 8 - 10 show the comparison of the actual existing collected data from the target station of May, June and July 2015 whereas Figures 11 - 13 show the same parameter of May, June and July 2017.

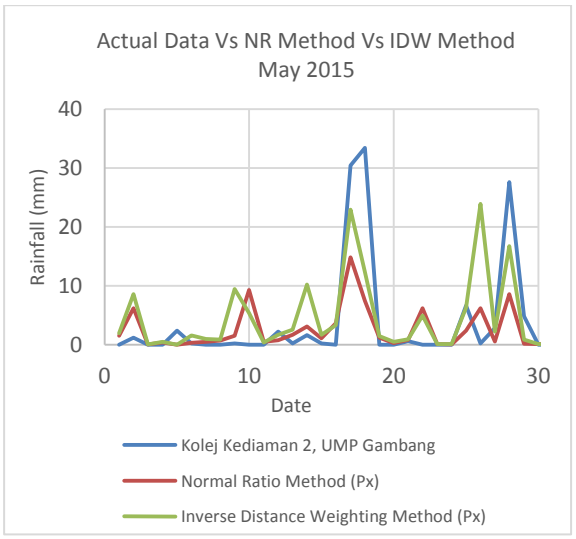

Figure 8. Comparison of the estimation of missing rainfall data using NR method and IDW method with actual data of May 2015

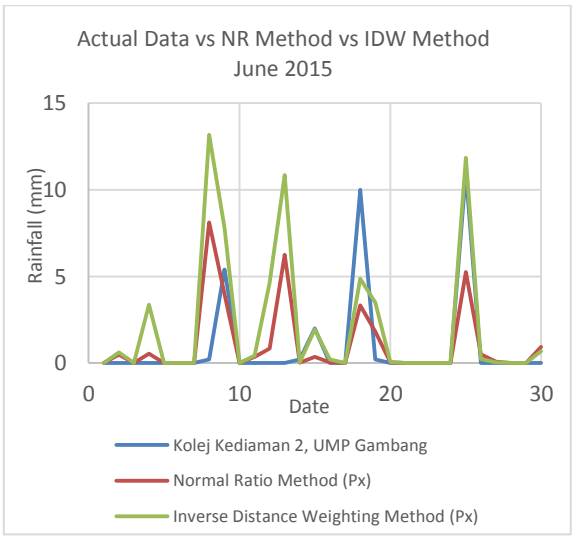

Figure 9. Comparison of the estimation of missing rainfall data using NR method and IDW method with existing data of June 2015

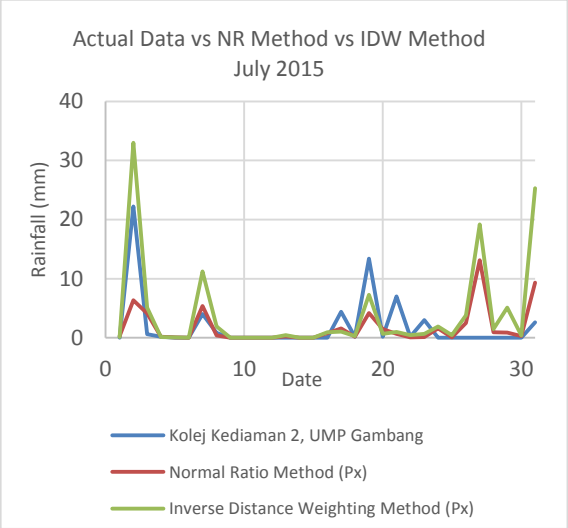

Figure 10. Comparison of the estimation of missing rainfall data using NR method and IDW method with existing data of July 2015

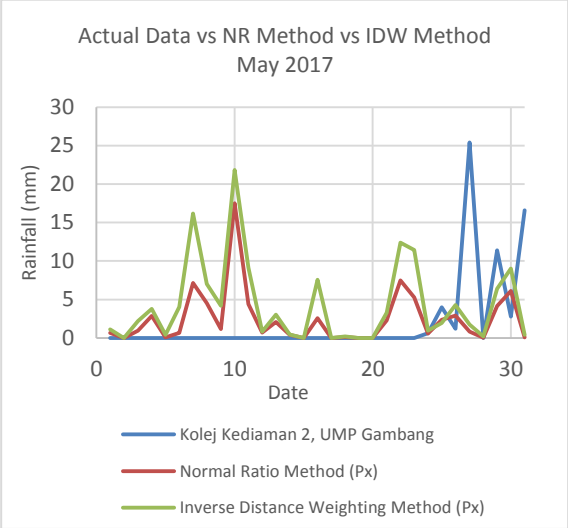

Figure 11. Comparison of the estimation of missing rainfall data using NR method and IDW method with existing data of May 2017

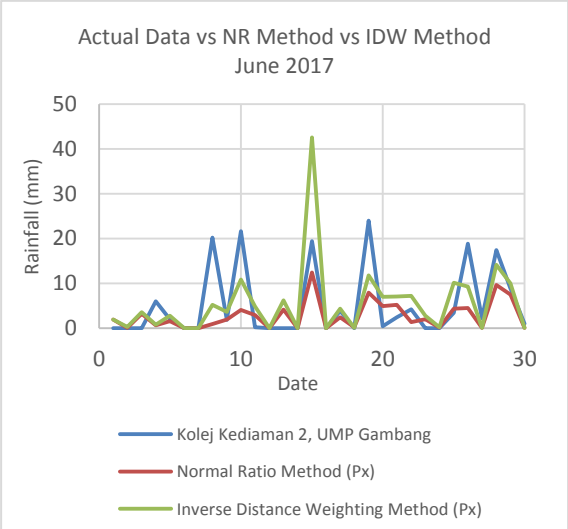

Figure 12. Comparison of the estimation of missing rainfall data using NR method and IDW method with existing data of June 2017

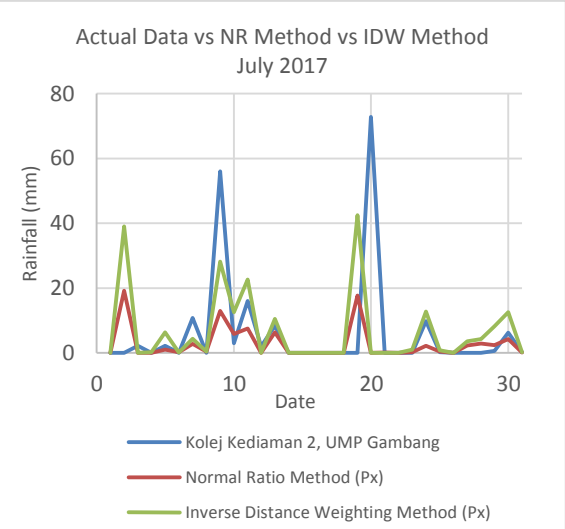

Figure 13. Comparison of the estimation of missing rainfall data using NR Method and IDW Method with existing data of July 2017

By referring to the graphs in Figures 8 - 13 for the comparison of actual existing data at the target station with the estimation of rainfall data using NR method and IDW method of year 2015 and 2017 respectively, it is obvious that both NR and IDW methods estimated different values than the actual rainfall data. The estimated values of missing rainfall are not likely to be precise in comparison with the actual rainfall data [10, 11]. It can be seen that both the methods estimated fluctuating and inconsistent value of rainfall data where sometimes both the methods estimated higher value and vice versa.

5. Conclusion

Studies involving the use of continuous time series data will normally face the problem of missing value. Mostly the existing data series are too short to perform a good and meaningful analyses and often contain a large number of missing values. Normally lack of data and inhomogeneity problem are

due to rainfall station relocation, changes in the environment, instrument malfunctions and network reorganizations. By referring to the analysis done in this study, it can be concluded that it is impossible to decide which of all the methods used in this study which are AM method, NR method and IDW method to be capable to estimate missing rainfall data for short term period of study. This is based on the observation done from the results as they shown that each of the methods did not estimate the missing rainfall value as precise as the actual existing data at the target station. On the other hand, the NR method will always estimate lower value of rainfall data to be compared with the estimation using IDW method.

6. References

[1] Azman M A, Zakaria R and Radi N F A 2015 *AIP Conference Proceeding* **65** p 1643
[2] Presti R L, Barca E and Passarella G 2008 *Environmental Monitoring and Assessment* **160** pp 1– 22
[3] Romman Z A, Al-Bakri J and Kuisi M A 2019 *International Journal of Engineering and Applications* **9** pp 76-80
[4] Suhaila J, Sayang M D and Jemain A A 2008 *Asia-Pacific Journal of Atmospheric Sciences* **44** pp 93-104
[5] Kamaruzaman I F, Zin W Z W and Ariff N M 2017 *Malaysian Journal of Fundamental and Applied Sciences* **13** pp 375–380
[6] Morales J L, Rangel F A H., Segovia-Domínguez I, Morua A R and Hernández J H 2019 *Atmósfera* **32** pp 237–259
[7] Chow V T, Maidment D R and Mays L W 1988 *Applied Hydrology, McGraw-Hill Book Company, Singpore*
[8] Singh V P 1994 *Elementary Hydrology, Prentice Hall of India, New Delhi*
[9] Chen F W and Liu C W 2012 *Paddy Water Environment* **10**
[10] Bennett N D, Newham L T H, Croke B F W and Jakeman A J 2007 *International Congress on Modelling and Simulation (MODSIM 2007)* University of Canterbury, Christchurch, New Zealand pp 2520–2526
[11] Elshorbagy A A, Panu U S and Simonovic S P 2000 *Hydrological Sciences* **45** pp 849–866

Acknowledgement
The authors sincerely acknowledged the Department of Irrigation and Drainage Malaysia (DID), for providing the daily rainfall data that is used in this study. The authors also acknowledged Universiti Malaysia Pahang for the financial support given for this project (RDU180350).