MATEMATIKA, 2013, Volume 29, Number 1c, 233-240

Functional Data Analysis Technique on Daily Rainfall Data: A Case Study at North Region of Peninsular Malaysia.

[1]Muhammad Fauzee Hamdan, [2]Jamaludin Suhaila and [3]Abdul Aziz Jemain
[1,2]Department of Mathematical Sciences, Faculty of Science, Universiti Teknologi Malaysia,
81310 UTM, Johor Bahru, Malaysia.
[3]School of Mathematical Sciences, Faculty of Science & Technology, Universiti Kebangsaan Malaysia,
43600 Bangi, Selangor, Malaysia.
email: [1]mfauzee@utm.my, [2]suhailasj@utm.my, [3]azizj@ukm.my

Abstract: The study of rainfall features and patterns are very useful for water management systems, water resources engineering and also in agricultural planning. It can be beneficial in order to reduce the risks and losses. Functional data analysis technique is one of the method can be used to explore and display the pattern and variation of the rainfall data. This technique displays the pattern in the form of curves. The first and second derivatives of the curves represent the rate of change and the acceleration of the curves. The objective of the study is to model two rainfall features; rainfall amount and rainfall occurrence by using functional data analysis technique at eight rainfall stations from the north part of Peninsular Malaysia. Markov chain model has been used to model the rainfall occurrence and Fourier basis to smoothing the data. The results show that both of the rainfall features have similar bimodal pattern. Although the mean curves are slightly similar, the first peak of variance curve for rainfall occurrence is higher than the second peak which is difference with variance curve for rainfall amount. The relationship between rainfall amount and rainfall occurrence for both observed and estimated curve is also discussed.

Keywords Functional Data Analysis; Markov Chain; Rainfall Amount; Probability of Rainfall Occurrence.

2010 Mathematics Subject Classification 62P12

1 Introduction

In agricultural planning, water resources management played a very importance role in order to maximize the production through effectiveness uses of water resource facilities and to minimize losses due to unfavorable weather conditions. However the water resource activities always affected by climate change elements especially the changes of rainfall pattern. It is beneficial to take into consideration of the rainfall distribution into water resource programs. An efficient utilization of rainfall during the rainy season could result in saving irrigation water for use during the dry season provided water saved could be stored for use later [1]. The efficient utilize of the rainfall is only possible if we can capture the features of rainfall pattern. Two main features of rainfall are the amount of the rainfall and the occurrence of the rainfall.

Most of the traditional approaches used a single value of mean rainfall on monthly or annually to indicate the rainfall pattern throughout the year on a certain area. The uncertainties in climate change can led to that value are no longer inadequate especially in agricultural planning. Fortunately, the development in computer technology make the analyzed possible to been carried throughout the year. One of the methods that can capture the interesting pattern of the data is functional data analysis. Instead of analyze the discrete data; the functional data analysis method can change the discrete data into a curve or function. An explanation in [2] gives a very good guidance on several functional methods such as principle component analysis, linear model and canonical correlation and discriminant analysis. It has been used in so many applications such as in biology [3], environmental problem [4] and economy [5].

This approach has been widely explored and used in other statistical branches such as nonparametric statistics [6], functional analysis of variance [7] and functional clustering technique [8]. In this paper, we only focus on the exploring of the rainfall pattern using the classical summary statistics (mean and variance) and the derivative of the curve obtain from the smoothing technique.

2 Materials and Methods

2.1 Data

Northern region of peninsular Malaysia consists the states of Perlis, Kedah, Perak and Penang. Kedah and Perlis are also known as Rice Bowl of Malaysia because of its massive rice production. Since the irrigation system can be controlled, understanding the rainfall pattern could be beneficial in cropping planning. The daily rainfall data were obtained from the Department of Irrigation and Drainage Malaysia with a different period of records. The periods are varying from 33 years to 54 years. Table 1 shows the location and the length of the period for each station.

Table 1 The list of eight rainfall stations with their geographical coordinates and length of the period

Station	Longitude	Latitude	Period
Alor Setar	100.40	6.20	1975 – 2008
Kangar	100.19	6.45	1975 – 2008
Kodiang	100.3	6.37	1954 – 2008
Arau	100.27	6.43	1956 – 2008
Pendang	100.48	5.99	1959 – 2008
Guar Nangka	100.28	6.48	1960 – 2008
Sik	100.73	5.81	1954-2008
Kaki Bukit	100.21	6.64	1954 – 2008

First of all, we set the threshold to differentiate between the rainy day and the dry day. In this study we set the daily rainfall amount of at least 1mm as a rainy day and less from 1mm as a dry day.

2.2 Functional Data Analysis

The first stage of FDA is to represent the discrete observe proportion (y_i, t_i) into a functional form, $x(t_i)$ by using a suitable basis function. We can express this in notation as

$$y_i = x(t_i) + \varepsilon_i \tag{1}$$

where ε_i is an error term which is assumed to be independently distributed with mean zero and constant variance. The functional form $x(t_i)$ is approximated by a finite linear combination of K basis function, $\phi_k(t)$

$$x(t) = \sum_{k=1}^{K} c_k \phi_k(t) \tag{2}$$

By using Fourier basis, the functional form, $x(t)$ can we stated as follow

$$x(t) = c_0 + c_1 \sin \omega t + c_2 \cos \omega t + c_3 \sin 2\omega t + c_4 \cos 2\omega t + \ldots \quad (3)$$

In selecting the value of *K*, too many basis functions will over fit the data while too few basis functions fail to capture the interesting pattern of curves. Too many basis functions also mean small bias but large sampling variance. On the other hand, if too few basis functions are used, it will result in small sampling variance but large bias. One reasonable way for choosing the number of basis is to add basis functions until the deviance, s^2 fails to decrease substantially [2].

$$s^2 = \frac{1}{n-K} \sum_{i=1}^{n} \left(y_i - \hat{y}_i(t) \right)^2 \quad (4)$$

After several calculations, we found that eleven basis is the most suitable number of basis function for all stations.

The derivatives of order *M* are evaluated as

$$D^M \hat{x}(t) = \sum_{k=1}^{K} c_k D^M \phi_k(t) \quad (5)$$

We only focus on first derivative as a rate of change of the rainfall amount and probability of rain and the second derivative as acceleration which indicate how fast the changes in rainfall amount curve and probability of rain curve.

2.3 Markov Chain Model

The implementation of Markov chain approach to model the occurrence of daily rainfall has been extensively used over decades start with [9] and been used in [10] and [11] in order to found the optimum order of Markov chain. In this study, a wet day is defined as a day with a rainfall amount of at least 1mm. A day with the rainfall amount less than 1mm is considered as a dry day. Markov chain model of zero order were used to model the rainfall occurrence for each station. A sequence of consecutive wet or dry day is obtained from the daily record as $S=\{Y_1, Y_2, Y_3, \ldots, Y_{j-1}, Y_j\}$ where Y_1 , Y_2 , Y_3 , or , Y_j is either 0 or 1 and the suffixes denotes the Julian day. The similar steps were done in finding the probability of rainfall occurrence. The overall probability of rain on any given date is estimated by the proportion of rainy days on that day (Markov chain zero order).

3.0 Result and Discussion

A plot of smoothed curves of mean rainfall amount and probability of rainfall occurrence for each station is given in Figure 1 and Figure 2 respectively. Figure 3 shows the curve of mean rainfall amount and probability of occurrence for north region of peninsular Malaysia.

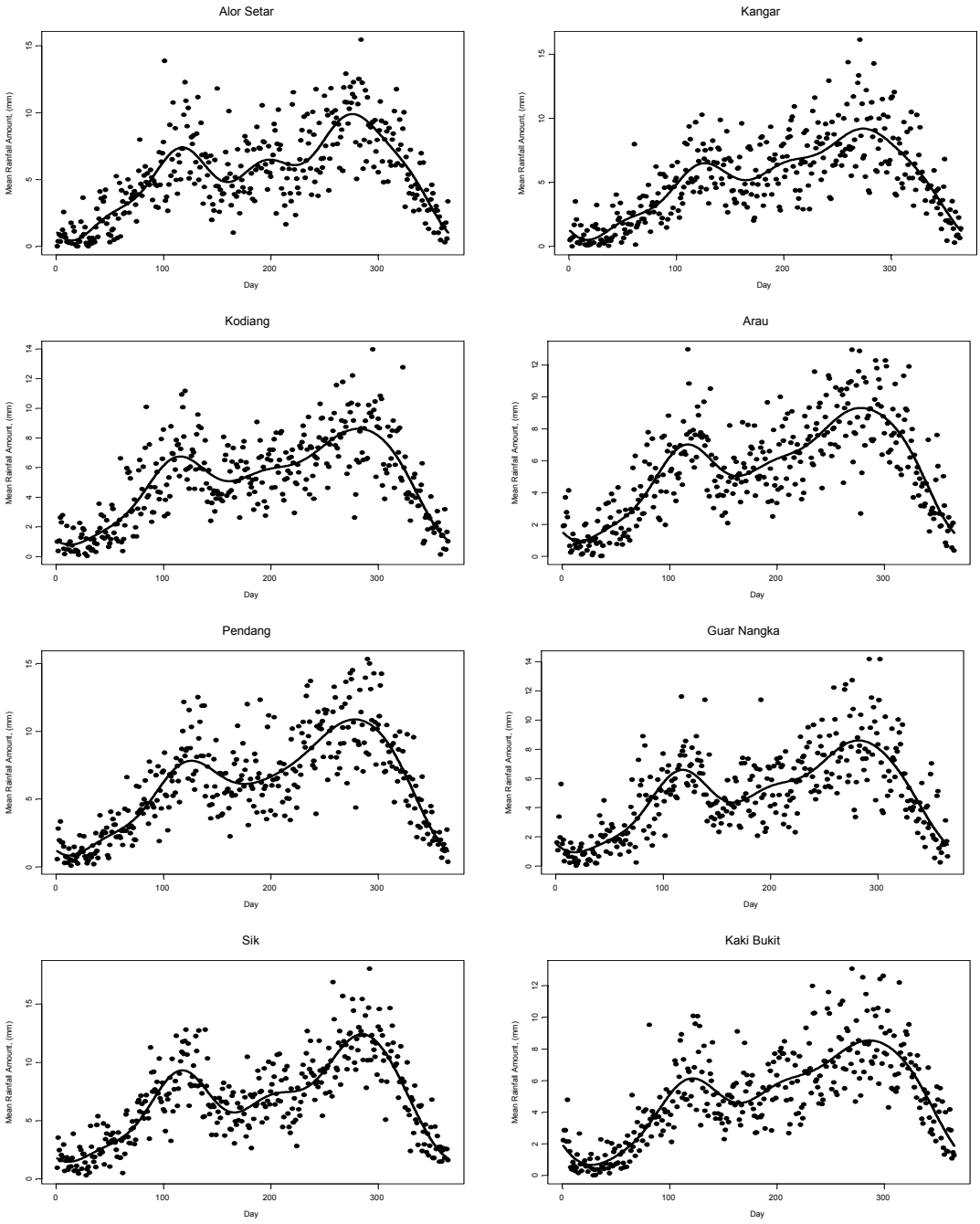


Figure 1 Mean rainfall amount for each station.

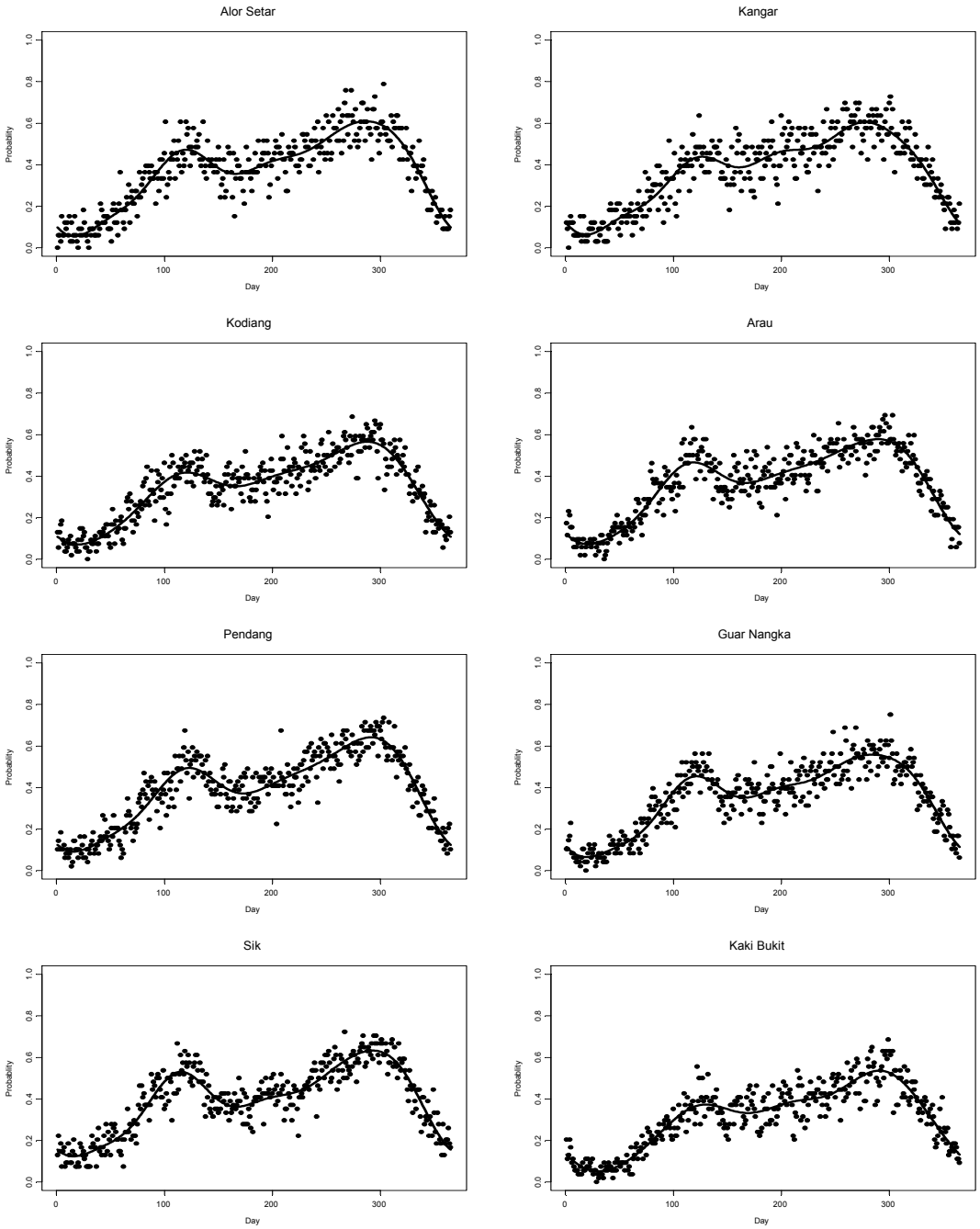
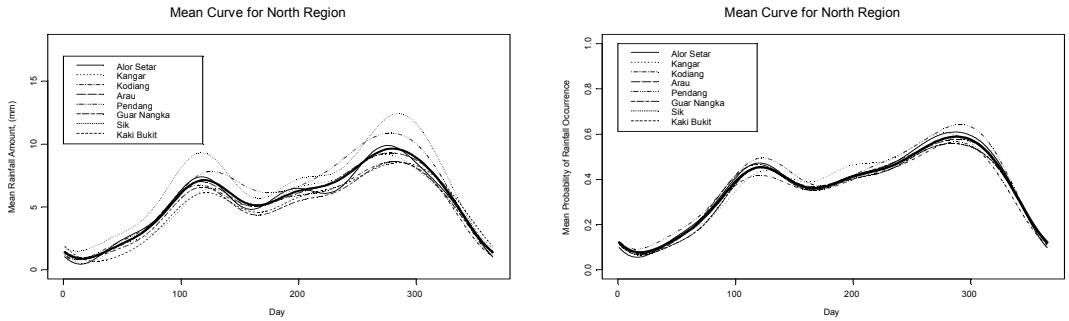


Figure 2 Probability of rainfall occurrence for each station.

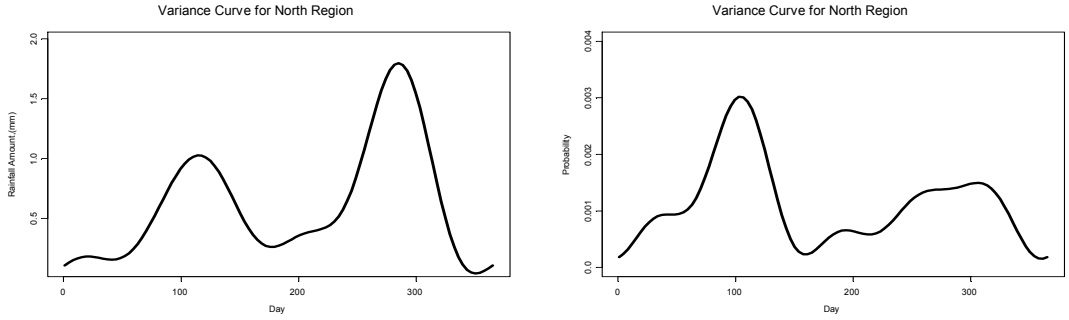


(a) (b)

Figure 3: (a) Mean curve for rainfall amount, (b) Mean curve for probability of rainfall occurrence

The curve of rainfall amount gently increase start from the start of the year and achieved the first peak in early April. The same pattern discovered in the probability of rainfall occurrence curve. The mean curve of rainfall amount showed that the rainfall amount changes in range 5mm to 10mm per day from early April to October. Meanwhile, for probability of rainfall occurrence, the changes are in range 40% to 60% per day. Both of the rainfall features achieved the second peak of the curve at the end of October. After that the curves start decrease until the end of the year.

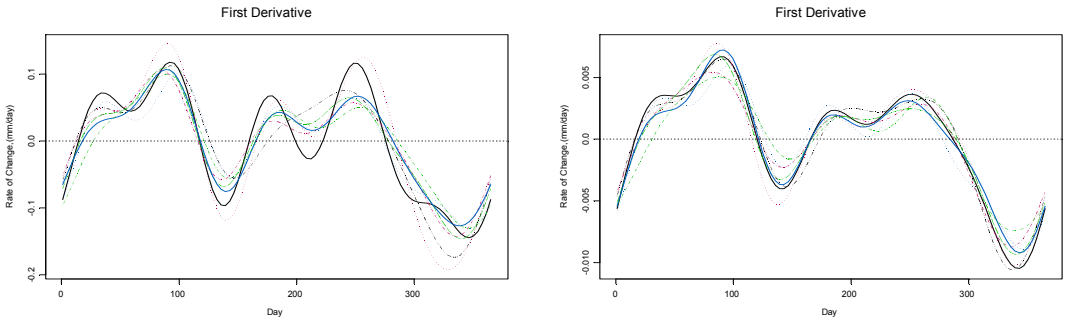
The variance curve for rainfall amount and probability of rain is given in Figure 4. For the variance rainfall amount curve, the difference on the second peak is higher than the difference on the first peak among the stations. Nevertheless, the variance curve for probability of rain for the first peak is higher than the second peak. This show that the curves of probability of rain among station are more vary on the first peak while for the rainfall amount is on the second peak.



(a) (b)

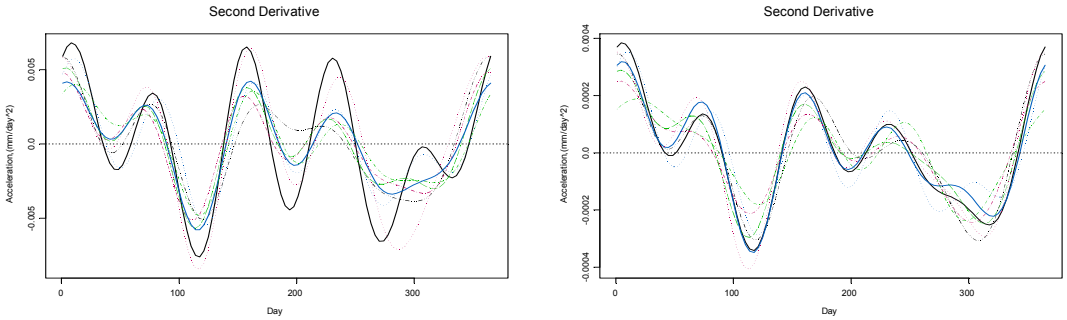
Figure 4 (a) Variance curve for rainfall amount. (b) Variance curve for probability of rainfall occurrence.

The first derivative of the curve could tell us how the rainfall amount curve and probability of rain change in the matter of time. Figure 5 shows the fluctuated of the rainfall amount and probability of rainfall occurrence curves over the year. It shows that the change can decrease to 2mm per day on December and also can increase up to 1mm per day on March and September for rainfall amount. For the probability of rain, the percentage to have rain can change up to 0.5% on February and decrease to 1% on December. In addition, Figure 6 shows the second derivative for both rainfall features. The graph indicated the acceleration on how fast the rainfall amount curve can be change.



(a) (b)

Figure 5 First derivative of (a) Rainfall amount (b) Probability of rainfall occurrence.



(a) (b)

Figure 6 Second derivative of (a) Rainfall amount (b) Probability of rainfall occurrence.

Finally we plotted the value of rainfall amount fitted curve versus the value of probability of rain fitted curve as shown in Figure 7. The graph shows a strong positive linear relationship between both rainfall features. The relation can be used to find the rainfall amount at certain value of the probability to have rain.

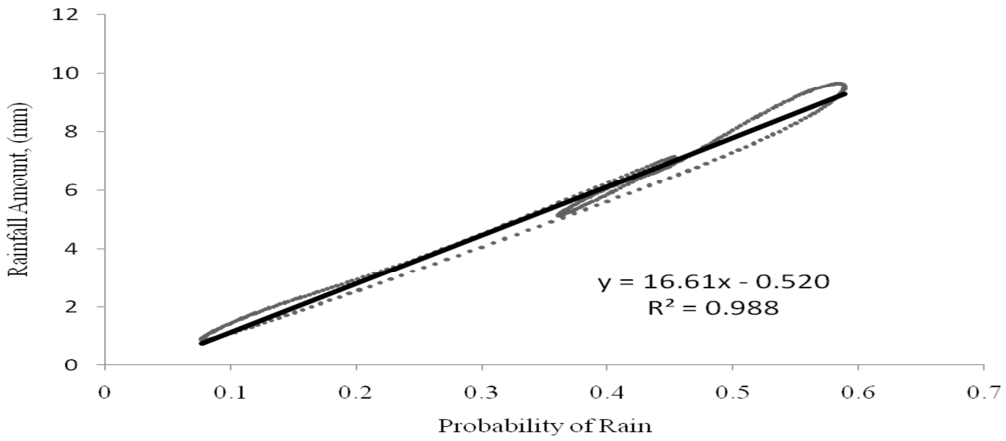


Figure 7 Relationship between rainfall amount and probability of rainfall occurrence.

4 Conclusion

In this paper, we have analyzed a daily rainfall data from eight different rainfall stations at north region of peninsular Malaysia by using functional data analysis technique. With this method, we have represented the discrete rainfall data into smoothed curves in order to investigate the rainfall pattern. We used Fourier basis to smoothing the data since rainfall always found to be seasonal. We found that all eight stations give a bimodal pattern for rainfall amount and probability of rainfall occurrence. Moreover, rainfall amount have strong positive linear relationship with probability of rainfall occurrence. The result found in this study can provided very informative information for crop scheduling (seedling and harvesting), irrigation system (time to released or block the water) and other agricultural planning.

Acknowledgement

The authors would like to thanks the Department of Irrigation and Drainage Malaysia for supplying the data for this study.

References

[1] Lee, T. S., Haque, M.A., Najim, M.M.M. 2005. Scheduling the cropping calendar in wet-seeded rice schemes in Malaysia. Agricultural Water Management. 71:71-84.
[2] Ramsay, J.O. & Silverman, B.W. 2005. *Functional Data Analysis*, second ed. Springer, New York.
[3] Newell, J., McMillan, K., Grant, S. & McCabe, G. 2006. Using functional data analysis to summarise and interpret lactate curves. Computers in Biology and Medicine. 36:262-275.
[4] Gao, H.O. & Niemeier, D.A. 2008. Using functional data analysis of diurnal ozone and NOx cycles to inform transportation emissions control. Transportation Research Part D. 13:221-238.
[5] Laukaitis, A. & Rackauskas, A. 2005. Functional data analysis for clients segmentation tasks. European Journal of Operational Research. 163:210-216.
[6] Ferraty, F. & Vieu, P. 2006. *Nonparametric functional data analysis*: *theory and practice*. New York: Springer.
[7] Cuevas, A., Febrero, M. & Fraiman, R. 2004. An anova test for functional data. *Computational Statistics and Data Analysis* 47: 111-122.
[8] Mizuta, M. 2004. Clustering method for functional data. Proceedings in Computational Statistics, Physica-Verlag, A Springer Company, 1503-1510.
[9] Gabriel, K.R., Neumann, J., 1962. A Markov Chain model for daily rainfall occurrence at Tel Aviv. Quart. J**.** R. Met. Soc**.,** London, 88, 90–95.
[10] Jimoh, O.D., Webster, P., 1999. Stochastic modeling of daily rainfall in Nigeria: intra-annual variation of model parameters. J. Hydrol. 222, 1–17.
[11] Deni, S.M., Jemain, A.A., Ibrahim, K., 2009. Fitting optimum order of Markov chain models for daily rainfall occurrences in Peninsular Malaysia. Theor. Appl. Climatol. 97,109-121.