Proceeding of 2nd International Science Postgraduate Conference 2014 (ISPC2014)

GENERALIZED LINEAR MODELS (GLMs) APPROACH IN MODELLING RAINFALL DATA OVER JOHOR AND KELANTAN AREA

[1]NOR HANISAH SUHAIMI AND [2]SHARIFFAH SUHAILA SYED JAMALUDIN

[1,2]Department of Mathematical Sciences, Faculty of Science
Universiti Teknologi Malaysia,
81310 UTM Johor Bahru, Johor, Malaysia

[1]nhanisah87@gmail.com, [2*]suhailasj@utm.my

*Corresponding author

Abstract. Observations of rainfall data are always changing over time. With the concern over climate change, this study is done to demonstrate how Generalized Linear Models (GLMs) could be utilized to model daily rainfall amount over Johor and Kelantan areas. Hence, in modeling rainfall amount, Fourier series are used as the smoothing technique. This research concentrated on the daily rainfall series with the duration period of 1985 to 2011 from three rainfall stations in Johor and another three in Kelantan area. The results indicated that the rainfall stations demonstrate different behaviours of rainfall patterns. One harmonic is sufficient to model the mean rainfall per rainy day at the stations that are located at the Johor area while four harmonics are best described the rainfall pattern at Kelantan area. Based on the resulting curves with fitted smoothing parameters, a good summary of statistics of the six stations were obtained. The results from the model will then be used to compare the rainfall patterns among the stations.

Keywords daily rainfall series; smoothing technique; Generalized Linear Model; Fourier series

1.0 INTRODUCTION

Peninsular Malaysia experiences rainfall that varies seasonally. These seasonal variations in rainfall have lead to a situation in which the parameters of rainfall occurrence and rainfall amount keep changing throughout the year. Both rainfall occurrence and rainfall amount models respectively are the two types of the stochastic rainfall models. Rainfall occurrence is a model that generates the sequence of wet and dry days, while rainfall amount is a model that simulates the rainfall amount on wet days.

This variation is normally handled by generating separate parameters for each month of the year [1]. However, many parameters need to be estimated from the models. So, a more efficient and sophisticated approach is proposed by using the method of Fourier series in smoothing the parameters of the model. Fourier series is convenient for the seasonally fluctuating values of parameters in rainfall models [2]. They applied Fourier series to smooth the model parameters for the stations that are located in continental United States.

The seasonal variation that occurs in Malaysia is influenced by four main seasons that is due to the uniform periodic changes in the wind flow patterns. The four main seasons are the southwest monsoon, northeast monsoon and two shorter periods of inter-monsoon seasons. The Southwest monsoon season is usually commences between May and August while the Northeast monsoon usually occurs between November and February.

As been mentioned before, this seasonal variation is also influenced by two shorter periods of inter-monsoon which is the transition period between the monsoon that occur during March to April and September to October. Northeasterly winds bring heavy rainfall to the east coast area. As the distance between the areas and the east coast increases, the areas would be less affected with its influences. Besides, the Titiwangsa Range and other mountain ranges might block the northeasterly winds from bringing the heavy rainfall to those areas.

This study will discuss only on modeling the rainfall amount on wet days. Generalized Linear Model will be implemented to model the rainfall distributions. The daily amount of rainfall will be analyzed and Fourier series will be fitted to the mean rainfall of the gamma distributions. The results from the model obtained will be utilized in comparing the rainfall patterns among the stations selected, specifically Johor and Kelantan area. Particularly, the comparison will be analysed based on the number of harmonics that best described the rainfall patterns of the rainfall stations and the differences in term of the seasonal rainfall peaks between the stations.

Figure 1 Physical Map and Selected Rainfall Stations

2.0 DATA

Based on the completeness of the data, six rain gauge stations that are located in both Kelantan and Johor area were selected for this study. Rainfall data were obtained from the Malaysian Meteorological Department. In this study, daily rainfall series from the period of 1985 to 2011 are analysed. In this study, a wet day is defined as a day with rainfall of at least 1 mm ($R \geq 1$ mm). To overcome the situation when there is no rain in certain day, the daily rainfall data were combined for every five days. The values of mean obtained are the mean of the rainfall amount per five days for 32 years. Therefore, the number of days, T = 73 days. The locations of those stations are as shown in Figure 1 and Table 1 displays the descriptive statistics for each rain gauge station, along with its latitude and longitude.

Table 1 Summary Statistics of Annual Rainfall for Studied Stations during Year 1985 to 2011

Stations	Latitude	Longitude	Amount (mm)	Intensity (mm day^{-1})	CV (%)
Johor					
Hospital Pontian	01° 29' N	103° 23' E	2370.98	16.46	12.0
Kluang	02° 01' N	103° 19' E	2192.67	15.12	14.6
Senai	01° 38' N	103° 40' E	2512.67	15.42	13.1
Kelantan					
Kota Bharu	06° 10'N	102° 17'E	2588.05	19.17	18.3
Pusat Pertanian Pasir Mas	06° 02' N	102° 07' E	2680.38	20.35	19.8
Kuala Krai	05° 32' N	102° 12' E	2525.76	16.70	14.4

Table 1 shows that Kelantan areas received high average of rainfall amount and intensity compared to the stations that are located in Johor region. Coefficient of Variation (CV) of the annual rainfall intensity reflects the ratio between the standard deviation and the average of the rainfall intensity annually. The stations at Kelantan area show the largest variability of rainfall amount which is between 14% to 20%. This indicates that the variation of the rainfall amount at Kelantan is quite different every year.

3.0 PROBLEM STATEMENT

This section is divided into two sub-sections. The two sub-sections will discuss about Fourier fitting and the methods in evaluating the deviances.

3.1 Fourier Fitting as the Smoothing Function

The model for rainfall amount only describes the distribution of rainfall on wet days. Several distributions have been used by other researchers in modeling rainfall amount. The distributions are gamma distribution [3], exponential distribution [4], log normal distribution [5] and others.

Considering $X(t)$ as the amount of rain on day t with a condition that day t is wet, gamma distributions have been choose for modeling the rainfall amount on the wet days. Between the gamma and other models, gamma model is slightly better, in term of its efficacy [6]. Besides, gamma distributions have been identified to fit well with the distribution of $X(t)$ which is highly skewed. The density function of the gamma distribution is as follows:

$$f(x) = \left(\frac{k}{\mu(t)}\right)^k x_i(t)^{k-1}\exp\left(-\frac{kx_i(t)}{\mu(t)}\right)/\Gamma(k) \quad (1)$$

$E\big(X(t)\big) = \mu(t)$, is the mean rainfall on day t where $t = t_1, t_2, .., t_T$ and T= 73. $x_i(t), i = 1,2, .., n(t)$, where $n(t)$ is the number of years in which day t had rained, is the amount of rain on day t at year i. $1/\sqrt{k}$ is a constant coefficient of variation for the distribution.

The response variable of a generalized linear model may come from the exponential family [7]. Since gamma distribution is also included in the exponential family, then the generalized linear model can be utilized to fit the distribution. A log link is taken to the $\mu(t)$ because the mean rainfall must be positive. Then, the function can be written as $\ln\big(\mu(t)\big) = g(t)$. If $g(t)$ is linear when the parameters are unknown, then once again this model is a generalized linear model. Fourier series is used as the smoothing function. The Fourier series is as follows:

$$g(t) = A_0 + \sum_{j=1}^{m}(A_j \sin(jt') + B_j \cos(jt')) \quad (2)$$

where A_j and B_j are the parameter coefficients, j is the number of harmonics, m is the maximum harmonic required for the series and $t' = \pi(t - 183)/183$.

3.2 Evaluating the Deviances

In order to evaluate the adequacy of the generalized linear models, the deviances are calculated. There are various ways that could be done in measuring the discrepancy or goodness of fit [7]. Deviance is one type of measuring from the logarithm of a ratio of likelihoods. Recently, it is classified into two components which is 'between-day deviance' and 'within-day deviance'. For 'between-day deviance', the equation is as follow:

$$D_B = 2\sum_t n(t)[\ln \hat{\mu}(t) - \ln \mu(t)] \tag{3}$$

$\hat{\mu}(t)$ is the fitted value of $\mu(t)$. This 'between-day deviance' will contribute to the result of number of harmonics and the value of residual. If the deviance has a distribution that is approximately a multiple of a χ^2 distribution, then the model is correct. The mean deviance for each harmonic is calculated by dividing the value of deviance with degree of freedom. The ratio of mean deviance will have an approximate F distribution [8]. By taking residual of between-day deviance/degree of freedom for the residual term, as denominator of the ratio, F-distributions could be approximated. When there are no further harmonics that reduces the deviance significantly, the maximum number of harmonics that best described the model could be determined. Below is the formula for the 'within-day deviance':

$$D_W = 2\sum_t n(t)[\ln \mu(t) - \overline{\ln x}\,(t)] \tag{4}$$

For the above formula, $\overline{\ln x}\,(t) = \sum_{i=1}^{n(t)} \ln x_i(t)/n(t)$.

4.0 SIMULATIONS AND RESULTS

4.1 The Number of Harmonics

Table 2 shows the results of the analysis of deviance for Kluang station in the Johor area. In the record period, the total numbers of rainy days were 1644 days. Based on the table, Fourier series with one harmonic was found to be reasonable. Since the deviance was not reduced significantly when the second harmonics were applied, it indicates that one harmonic is sufficient to model the mean rainfall per rainy day at this station.

Table 2 Analysis of Deviance for Modeling Mean Rainfall per Rainy Day at Kluang produced by fitting the Fourier series

Source	Degrees of Freedom	Deviance	Mean Deviance	F	P-value
Between Day	72	165.81			
One Harmonic	2	64.05	32.02	22.14	0.0000
Two Harmonics	2	5.81	2.90	2.01	0.1481
Three Harmonics	2	3.37	1.68	1.16	0.3255
Four Harmonics	2	2.04	1.02	0.70	0.5044
Five Harmonics	2	0.87	0.43	0.30	0.7462
Residual	62	89.69	1.45		
Within Days	1582	1690.27	1.07		
Total	1644	1856.09			

The probability value (P-value) indicates the number of harmonics required to model the mean rainfall per five rainy days for the station. When the P-values are less than 0.05 (significance level), then that would be the indicator on the maximum number of harmonics that best fit the model. Based on the value in Table 3, four harmonics are sufficient for Kota Bharu station. No further harmonic is required in the model since they do not reduce the deviance significantly. The station has recorded 1532 rainy days from the period of 1985 to 2011.

Table 3 Analysis of Deviance for Modeling Mean Rainfall per Rainy Day at Kota Bharu produced by fitting the Fourier series

Source	Degrees of Freedom	Deviance	Mean Deviance	F	P-value
Between Day	72	697.53			
One Harmonic	2	312.22	156.11	72.06	0.0000
Two Harmonics	2	69.76	34.88	16.10	0.0000
Three Harmonics	2	138.23	69.11	31.90	0.0000
Four Harmonics	2	42.04	21.02	9.70	0.0002
Five Harmonics	2	0.97	0.49	0.22	0.8019
Residual	62	134.32	2.17		
Within Days	1460	1948.52	1.33		
Total	1532	2646.05			

The observed and fitted values of the mean rainfall per rainy day for all studied stations have been plotted in Figure 3. Table 4 shows the number of harmonics required for the model at each station and also the coefficient of the Fourier series for all stations. If one harmonic is fitted, then three parameters are estimated which include the constant value, a sine coefficient and also a cosine coefficient. So, if four harmonics is sufficient, then there would be nine parameters estimated in the model.

Table 4. Number of Harmonics and Coefficient of the Fourier series for all Stations

Station	Har-mon ics	Coefficient of the Fourier series								
		A_0	A_1	B_1	A_2	B_2	A_3	B_3	A_4	B_4
Hospital Pontian	1	3.632	0.109	-0.053						
Kluang	1	3.556	-0.048	-0.268						
Senai	1	3.661	0.002	-0.131						
Kota Bharu	4	3.586	0.271	-0.359	-0.317	-0.039	0.396	0.189	-0.234	-0.045
Pusat Pertani-an Pasir Mas	4	3.744	0.358	-0.221	-0.255	0.14	0.243	0.013	-0.207	-0.040
Kuala Krai	4	3.611	0.304	-0.280	-0.205	0.122	0.154	-0.009	-0.228	-0.032

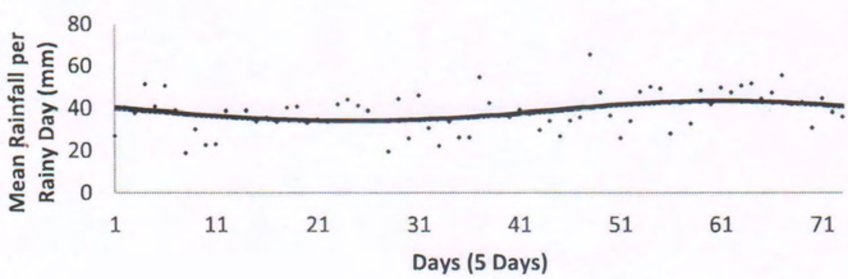
Based on the results, stations that are located at Kelantan area are best described with four harmonics, while one harmonic is sufficient to model the mean rainfall per rainy day for the stations at Johor area.

4.2 The Seasonal Rainfall Peaks

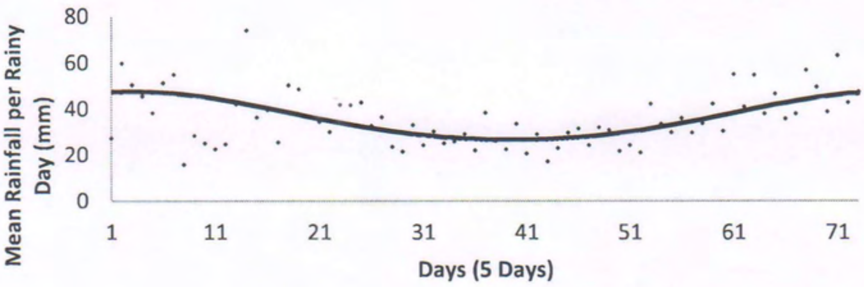
Figures 2(a), 2(b) and 2(c) describe the fitted curves for stations that are located in Johor while Figures 2(d), 2(e) and 2(f) describe stations in Kelantan. From the curves, the minimum amount of rain recorded in the Johor area for every five days is 26.6 mm whereas its maximum value is 47.5 mm. For stations that are located in the Johor area which is Hospital Pontian, Senai and Kluang, the highest peaks are recorded during the months of December to January. The figure depicts a bimodal pattern of rainfall for the stations in Johor and a unimodal pattern for the stations in Kelantan.

On the other hand, the fitted values for stations in Kelantan range approximately from 17 to 139 mm per day. The extreme value of the mean rainfall is recorded during the months of December. This extreme value is mainly influenced by Northeast monsoon that occurs between November and February. Thus, it can be said that the rainfall pattern for all stations are strongly affected by the northeast monsoon. Generally, northeast monsoon brings heavy rainfall to all three stations in Kelantan. Since there is no mountain ranges located around the stations that could block the northeasterly winds, then the winds easily bring the heavy rainfall to those areas.

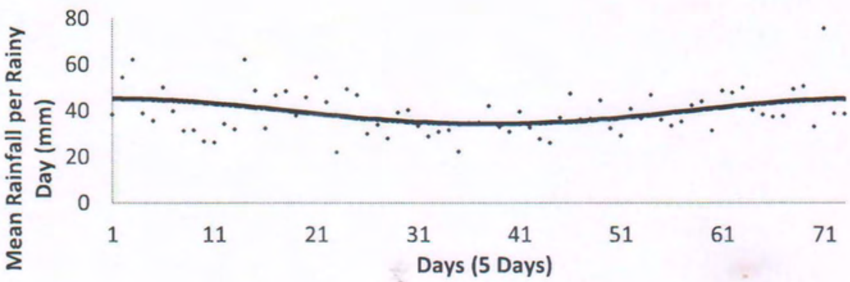
(a)



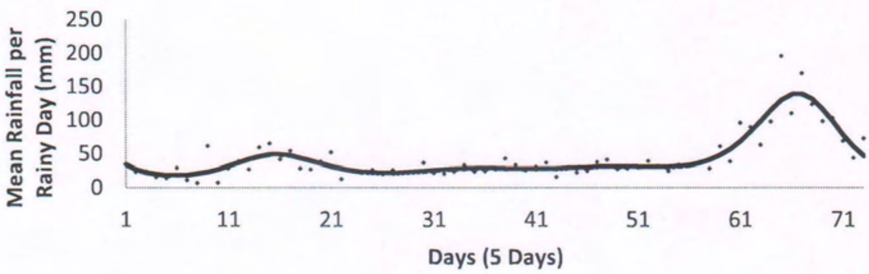
(b)


Mean Rainfall per Rainy Day (mm)
0
20
40
60
80
1
11
21
31
41
51
61
71
Days (5 Days)

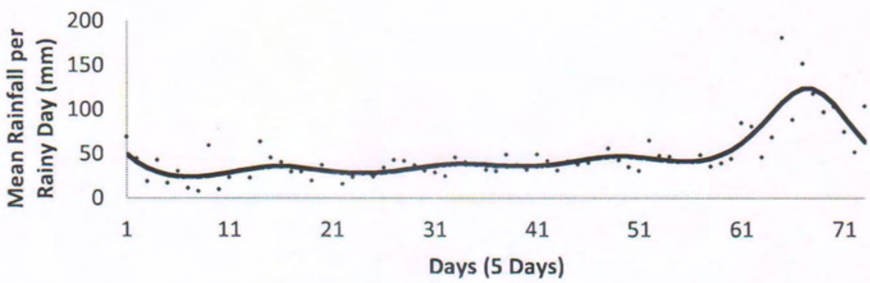
(c)


Mean Rainfall per Rainy Day (mm)
0
20
40
60
80
1
11
21
31
41
51
61
71
Days (5 Days)

(d)


Mean Rainfall per Rainy Day (mm)
0
50
100
150
200
250
1
11
21
31
41
51
61
71
Days (5 Days)

(e)


Mean Rainfall per Rainy Day (mm)
0
50
100
150
200
1
11
21
31
41
51
61
71
Days (5 Days)

(f)

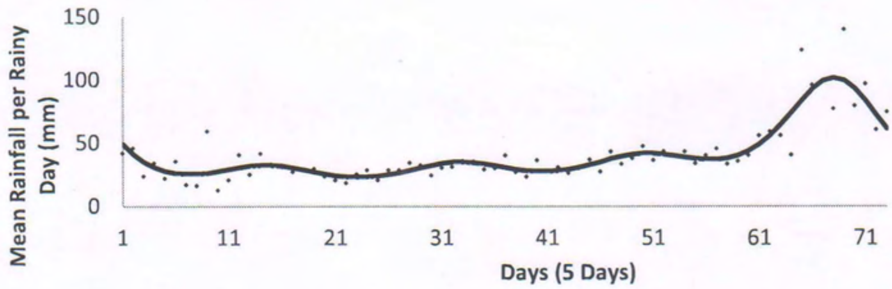


Figure 2 Observed and Fitted Mean rainfall per Rainy Day for each Station (a) Hospital Pontian (b) Kluang (c) Senai (d) Kota Bharu (e) Pertanian Pasir Mas (f) Kuala Krai

5.0 SUMMARY AND CONCLUSION

The stations that are located in the Kelantan area are best described with four harmonics while all the stations in Johor required one harmonic to model the mean rainfall per five rainy days. All the stations in Kelantan have unimodal rainfall patterns. On the other hand, bimodal patterns are best described with one harmonic for the stations in Johor. The wettest month for the stations in Kelantan is in December. However, for Johor area, the highest peak was recorded in December-January.

There are several limitations to this study. Initially, this study did not consider the events of the previous days. The results might be affected by the events of the previous days which could either be dry or wet events. Secondly, deep observations and analysis regarding to the seasonal rainfall peaks have not been emphasized in this study. Comparison of the rainfall patterns between the regions could be seen clearly if the period of wet and dry days with the dates, and also the maximum and minimum rainfall values also been analysed in this study. For upcoming studies, these issues that have been highlighted will be included into the future analysis.

ACKNOWLEDGMENTS

The authors are indebted to the staff of the Malaysian Meteorological Department for providing the daily rainfall data used in this study. The comments of an anonymous referee are also acknowledged. The authors would like to extend their sincere gratitude to the Ministry of Higher Education Malaysia (MOHE) for the financial supports received for this work under (UTM—FRGS 4F173). We are also grateful to the Universiti Teknologi Malaysia for supporting the project.

REFERENCES

[1] J. Suhaila and A. A. Jemain, A comparison of the rainfall patterns between stations on the East and the West coasts of Peninsular Malaysia using the smoothing model of rainfall amounts. *Meteorol. Appl.* 16: 391-401, 2009.

[2] D. A. Woolhiser and G.G.S. Pegram, Maximum Likehood Estimation of Fourier Coeeficients to Describe Seasonal Variations of Parameters in Stochastic Daily Precipitation Models. *J. Appl. Meteor.* 18: 34-42, 1978.

[3] Z. Hussain, Z. Mahmood and Y. Hayat, Modelling the Rainfall Amounts of North-West Pakistan for Agricultural Planning *Sarhad J. Agric.* 27(2): 313-321, 2011.

[4] P. Todorovic and D.A. Woolhiser, A Stochastic Model of n-Day Precipitation. *J. Appl. Meteor.* 14(1): 17-24, 1975.

[5] H.K. Cho, K. P. Bowman and G. R. North, A Comparison of Gamma and Lognormal Distributions for Characterizing Satellite Rain Rates from the Tropical Rainfall Measuring Mission. *J. Appl. Meteor.*43: 1586-1597, 2004.

[6] D. Firth, Multiplicative errors: lognormal or gamma?. *Journal of the Royal Statistical Society Series B*, 50: 266-268, 1988.

[7] P. McCullagh and J. A. Nelder, *Generalized Linear Models*, London: Chapman and Hall, 1989, pp. 1-35.

[8] R. J. Baker and J. A. Nelder, *The GLIM system release 3*, Oxford: Num. Algo. Group. Coe, 1978.