MATEMATIKA, 2008, Volume 24, Number 2, 243–257

Fourier Series in a Neyman Scott Rectangular Pulse Model

[1]**Fadhilah Yusof**, [2]**Norzaida Abas** & [3]**Zalina Mohd. Daud**
[1]Department of Mathematics, Faculty of Science, Universiti Teknologi Malaysia,
81310 UTM Skudai, Johor, Malaysia
[2,3]Universiti Teknologi Malaysia, Jalan Semarak, 54100 Kuala Lumpur
e-mail: fadhilah@mel.fs.utm.my

Abstract The ability of Fourier Series to exhibit seasonal fluctuation of rainfall process is presented. The Neyman Scott Rectangular Pulse Model with mixed exponential distribution for cell intensity is selected to describe the rainfall process. The model's parameters were estimated by employing the Shuffle Complex Evolution (SCE-UA) method. Seasonal variation is dealt with by fitting Fourier Series to the parameters. Significant harmonics for each parameter is determined using the cumulative fraction of total variance explained by significant harmonic. Results indicate seasonal fluctuations of parameters were sufficiently represented by the Fourier Series. Comparison between Fourier Series estimations and observed data of 10 years demonstrated the ability of Fourier Series in capturing the statistical characteristics of rainfall process.

Keywords Fourier series; Neyman Scott rectangular pulse model; harmonic.

1 Introduction

Stochastic rainfall models based on point process theory have been used for some time to analyse and predict the occurrences of rainfall in space and time. A detailed description of point process theory is given by Cox and Isham [4]. In hydrological studies and water resources planning, effective modeling of rainfall normally requires a long series of rainfall data. Stochastic models which produce synthetic rainfall data help in the development of these studies by generating data with statistical traits such as mean, variance and skewness resembling those of the observed data.

In this study a clustered point process model namely the Neyman Scott Rectangular Pulse model with mixed exponential distribution is adopted to model the rainfall process of the Klang Valley. A Fourier Series is then fitted to the parameters in an effort to reduce the number of parameters estimated to account for seasonal fluctuation.

2 Methodology

2.1 Neyman Scott Rectangular Pulse

Storm origins are assumed occur according to a Poisson process of rate λ. Each storm origin creates a random number of rain cells C ($C = 1, 2, ...$) and the distribution of C is either Poisson or Geometric with mean $E(C) = \mu_c$. The waiting time of each rain cell after the occurrence of a storm origin is determined by the exponential distribution with parameter β. None of the cell origin is located at the storm origin. Every rain cell is represented by a rectangular pulse for the purpose of parsimonious parameterization and mathematical tractability. The pulse is made up of a random duration L and a random intensity$X$, where the intensity is assumed constant throughout the cell duration. The parameters for duration and intensity are η and ξ respectively. According to Rodriguez-Iturbe et al. [8], the distribution for the rain cell intensity is arbitrary. The total rainfall intensity at any point in time is defined as the total amount of the intensities of all active cells at that time.

The total intensity $Y(t)$ is the sum of contributions from all active cells at time t.

$$Y(t) = \int_{u=0}^{\infty} X_{t-u}(u)dN(t-u)$$

where

- $X_{t-u}(u)$ is the random intensity of a pulse originating at time u measured a time $(t-u)$ later and $X_{t-u}(u) = \begin{cases} X & \text{with probability } P(x) \\ 0 & \text{with probability } 1 - P(x) \end{cases}$
- $N(t\}$ represents the counting stochastic process of the arrivals of the individual rain cells.
- $dN(t-u) = \begin{cases} 1 & \text{if there is a cell origin at } t-u \\ 0 & \text{orthewise.} \end{cases}$

The intensity of $N(t)$ is the product of the rate of storm origin and the mean number of cells per storm $\lambda\mu_c$. The first and second order properties of $Y(t)$ are as follow,

$$E[Y(t)] = \frac{\lambda\mu_c\mu_x}{\eta}$$

$$Var[Y(t)] = \frac{\lambda\mu_c E[X]}{\eta} + \frac{\lambda\beta\mu_x^2 E[C^2 - C]}{2\eta(\beta+\eta)} \quad - 2\lambda(\beta h - 1 + e^{-\beta h})\frac{\mu_c\mu_x^2(\mu_c - 1)}{\beta(\beta^2 - \eta^2)})$$

Auto covariance at lag τ,

$$\begin{aligned} c_Y(\tau) &= Cov[Y(t), Y(t+\tau)] \\ &= \frac{\lambda e^{-\pi\tau}}{\eta}\left(\mu_c E[X^2]\frac{\beta^2\mu_x^2 E[C^2 - C]}{2(\beta^2 - \eta^2)}\right) - \frac{\lambda\beta\mu_x^2 e^{-\beta\tau} E[C^2 - C]}{2(\beta^2 - \eta^2)} \end{aligned}$$

Rainfall data are normally presented in the form of rainfall intensities in discrete time intervals such as hourly or daily totals. Therefore the aggregated or cumulative properties are used in the process of parameter estimations of the model. The aggregated rainfall totals in disjoint time interval i of some fixed length h is as follows

$$Y_i^{(h)} = \int_{(i-1)h}^{ih} Y(t)dt$$

where h refers to the h-hour of aggregation.

The second-order properties of the aggregated process for $Y_i^{(h)}$ as derived by Rodriguez-Iturbe et al. [9] are as follow,

Mean, $\mu(h)$**:**

$$E[Y_i^{(h)}] = \frac{h\lambda E[C]E[X]}{\eta} \tag{1}$$

Variance, $\gamma(h)$**:**

$$\begin{aligned} Var[Y_i^{(h)}] = \lambda\eta^{-3}(\eta h - 1 + e^{-\eta h})[2\mu_c E[X^2] + E[C^2 - C]\mu_x^2\beta^2/(\beta^2 - \eta^2)] \\ - \lambda(\beta h - 1 + e^{-\beta h})E[C^2 - C]\mu_x^2\beta^{-1}/(\beta^2 - \eta^2) \end{aligned} \tag{2}$$

Covariance, $\gamma(h,k)$:

$$\begin{aligned} Cov\{Y_i^{(h)}, Y_{i+k}^{(h)}\} = \lambda\eta^{-3}(1 - e^{-\eta h})^2 e^{-\eta(k-1)h} \times [\mu_c E[X^2] + \tfrac{1}{2}E[C^2 - C]\mu_x^2\beta^2/(\beta^2 - \eta^2)] \\ - \lambda(1 - e^{-\beta h})^2 e^{-\beta(k-1)h} E[C^2 - C]\mu_x^2/[2\beta(\beta^2 - \eta^2)] \end{aligned} \tag{3}$$

The lag k autocorrelation function, $\rho(h,k)$:

$$\rho(h,k) = \frac{\gamma(h,k)}{\gamma(h)} \tag{4}$$

Work by Fadilah [6] indicated that mixed exponential distribution for cell intensity is better at describing the distribution of hourly rainfall amount compared to exponential, Gamma or Weibull distribution. The mixed-exponential distribution is a weighted average of two one-parameter exponential distributions. This distribution has three parameters namely α representing the mixing probability, ξ and θ representing the scale parameters.

Given that x is the hourly rainfall amounts per hour, the probability distribution function for the mixed exponential distribution is,

$$f(x) = \frac{\alpha}{\xi}e^{(-x/\xi)} + \frac{(1-\alpha)}{\theta}e^{(-x/\theta)}$$

$$x > 0; \quad 0 \le \alpha \le 1; \quad 0 < \xi < \theta$$

and the distribution function is,

$$F(x) = \alpha e^{-x/\xi} + (1-\alpha)e^{-x/\theta}$$

In this study the number of rain cells C follow the Poisson distribution whilst rain cell intensities X follow the mixed exponential distribution. To ensure that each storm generate at least one rain cell, $(C-1)$ is assumed to follow the Poisson distribution with mean $\nu - 1$.

Thus adjustments to equations (1) - (3) are made by replacing the following expressions,

$$\begin{aligned}
&\mu_c = E(C) = \nu \\
&E(C^2 - C) = \mu_c^2 - 1 = \nu^2 - 1 \\
&\mu_x = E(X) = \alpha(\xi) + (1-\alpha)(\theta) \\
&E(X^2) = 2\alpha(\xi^2) + 2(1-\alpha)(\theta^2)
\end{aligned}$$

Studies by Cowpertwait et al. [2] and Fadhilah [6] show that using transition probabilities in the estimation process tend to produce better estimations of statistical characteristics of rainfall.

Probability of dry interval of length h, $\phi(h)$ as derived by Cowpertwait [1] is

$$P\left(Y_i^{(h)} = 0\right) = \exp\left[\begin{array}{l} -\lambda h + \lambda\beta^{-1}(\nu-1)^{-1}\{1 - \exp[1 - \nu + (\nu-1)e^{-\beta h}]\} \\ -\lambda \int\limits_0^\infty [1 - p_h(t)]dt) \end{array}\right] \tag{5}$$

where

$$\begin{array}{l} p_h(t) = \left\{e^{-\beta(t+h)} + 1 - (\eta e^{-\beta t} - \beta e^{-\eta t}/(\eta - \beta)\right\} \times \\ \exp\left\{-(\nu-1)\beta(e^{-\beta t} - e^{-\eta t})/(\eta-\beta) - (\nu-1)e^{-\beta t} + (\nu-1)e^{-\beta(t+h)}\right\} \end{array}$$

and

$$\int\limits_0^\infty [1 - p_h(t)]dt = \frac{1}{\beta}\left\{\gamma + \ln[(\tfrac{\eta}{\eta-\beta} - e^{-\beta h}) \cdot (\nu - 1)]\right\}$$

where γ is a Euler's constant $= 0.5772$.

The transition probabilities are as follow,

$$\text{probability of wet-wet hour :}\phi_{WW}(h) = P(Y_{i+1}^{(h)} > 0 \left| Y_i^{(h)} > 0\right.) \tag{6}$$

$$\text{probability of dry-dry hour :}\phi_{DD}(h) = P(Y_{i+1}^{(h)} = 0 \left| Y_i^{(h)} = 0\right.) \tag{7}$$

Equations (6) and (7) can be expressed in terms of $\phi(h)$= $P(Y_i^{(h)} = 0)$ ([2]):

$$\phi_{DD}(h) = \phi(2h)/\phi(h)$$

$$\phi(h) = \phi_{DD}(h)\phi(h) + \{1 - \phi_{WW}(h)\}\{1 - \phi(h)\}$$

$$\phi_{WW}(h) = \{1 - 2\phi(h) + \phi(2h)\}/\{1 - \phi(h)\}$$

The NSRP model with mixed exponential distribution has seven parameters, namely

λ, ν , β, η , α, ξ, and θ that characterize respectively the origin of storm, the number of cells, the positions of cells relative to the storm origin, the duration of rain cells, the mixing probability and the intensity of the rain cells that is described by the last two parameters. For brevity, the NSRP model with mixed exponential distribution representing rain cell will be known as NSRP (Mexp).

2.2 Parameter Estimation

Statistics of selected observed hourly series and transition probabilities aggregated at three different temporal scales are,

$$\mu(1), \gamma(1), \gamma(6), \gamma(24), \quad \phi(24), \phi_{DD}(1), \quad \phi_{DD}(24)\phi_{WW}(1) \text{ and } \phi_{DD}(24) \tag{8}$$

The parameters of the NSRP(Mexp) model are estimated by minimizing the sum of squares (SS) of selected observed statistics as given in (10) and their equivalent theoretical expressions (equations (1)-(7)).

Let $K_i \equiv K_i\left(\lambda, \beta, \eta, \nu, \xi, \alpha, \theta\right)$ be a function of NSRP(Mexp) and $\hat{K}_i$ is the observed statistics, and the sum of squares is given by

$$SS = \sum_{i=1}^{k} w_i \left[1 - \frac{K_i}{\hat{K}_i}\right]^2 \quad k \geq 7, \quad \hat{K}_i > 0, \quad \lambda, \beta, \eta, \xi \geq 0 \text{ and } \nu \geq 1 \tag{9}$$

with w_i as a weight which gives greater weight to the fitting of some sample moments relative to others. The minimization of equation (9) was done using Shuffled Complex Evolution-University of Arizona (SCE-UA) method by Duan et al. [5].

2.3 Harmonic Analysis of the Fourier Series

The NSRP model is stationary whereas the rainfall data are not. To account for the seasonal pattern of rainfall, the usual practice is to estimate each parameter according to months hence permitting the parameters to vary from one month to another. For a period of one year, this would require a total of 12 estimates per parameter. The NSRP(Mexp) model with 7 parameters would then require a total of 84 estimates for one year duration. However to be parsimonious with respect to the number of estimates, it is assumed that the 12 estimates of each parameter vary periodically during the year. Harmonic analysis of the Fourier Series provides an alternative technique to present the seasonal pattern of the parameters with smaller number of coefficients to be estimated.

For the NSRP(Mexp) model , the parameter set for each month m is

$$\Phi(m) = \{\lambda(m), v(m), \beta(m), \eta(m), \alpha(m), \xi(m), \theta(m)\}$$

The periodic component Ω_m representing $\Phi(m)$ could be expressed in the Fourier form as follows:

$$\Omega_m = \bar{U}_m + \sum_{q=1}^{Q} \left[A_q \cos\left(\frac{2\pi\ qm}{w}\right) + B_q \sin\left(\frac{2\pi\ qm}{w}\right)\right]; \ m = 1, ..., w \tag{10}$$

where $\bar{U}_m$ represents the periodic mean of each parameter, A_q and B_q are the coefficients of the Fourier Series, m represents to a particular month for example m=1 is refers to January , m=2 refers to Febuary etc with w=12. q is the number of harmonic with the maximum number of harmonic Q yet to be determined.

$\bar{U}_m$ and the coefficients A_q and $B_q$are determined as follow:

$$\bar{U}_m = \frac{1}{w} \sum_{m=1}^{w} U_m$$

$$A_q = \frac{2}{w}\sum_{m=1}^{w} U_m \cos\left(\frac{2\pi_q m}{w}\right)$$

$$B_q = \frac{2}{w}\sum_{m=1}^{w} U_m \sin\left(\frac{2\pi_q m}{w}\right)$$

Equation (10) can also be presented in the polar form

$$\Omega_m = \bar{U}_m + \sum_{q=1}^{Q} C_q \cos\left(\frac{2\pi\ qm}{w} + \theta_q\right) \quad ; \quad m = 1, ..., w \tag{11}$$

with the amplitude $C_q = \sqrt{A_q^2 + B_q^2}$ and phase angle $\theta_q = \tan^{-1}\left[-\frac{B_q}{A_q}\right]$

2.4 Determination of Significant Harmonics

There are a few ways to determine the number of significant harmonics such as the use of cumulative periodogram as suggested by Salas et al. [11] and the use of cumulative fraction of the total variance explained by significant harmonic.

The cumulative periodogram P_Q is defined as the explained variance yield by the first Q harmonics,

$$P_Q = \frac{\sum\limits_{q=1}^{Q} V_q}{MS(U)}$$

where $MS(U) = \frac{1}{w}\sum\limits_{m=1}^{w}(U_m - \bar{U}_m)^2$ is the total mean square deviation of U_m about $\bar{U}_m$.

The graph of P_Q vs q is called cumulative periodogram plot. Variance V_q for each harmonic is given by $V_q = \frac{1}{2}C_q^2$ and the cumulative sum of the variances up to and including the harmonic Q is $VH_Q = \sum\limits_{q=1}^{Q} V_q$ where $Q = 1, ..., Q_{\max}$.

The percentage of the total variance is $\dfrac{VH_Q}{VH_{Q_{\max}}}$.

This study will adopt the second method in which the percentage contribution of the total variance explained by a particular harmonic can be determined by forming a ratio of the sequence of the amplitude of that harmonic to the sum of the squares of the amplitudes of all the harmonics, the sum being the total variance.

3 Results and Discussions

Data taken at station Km 16 Gombak, Selangor (Station 321700) which consist of rainfall series of 10 year period (1981-1990) were used this study. Selected statistics as in (8) were derived from the data series. Parameters of the NSRP(Mexp) were then estimated following the procedure described in Section 2.2. The data for selected statistics and estimated parameters are given in Fadhilah [6].

Different number of harmonics were fitted to these parameter estimates in order to determine the number of significant harmonic which best describe the seasonal variation of the parameters. The determination of significant harmonics for each parameter was performed according to the process described in Section 2.4. The findings are presented in Table 1.

Table 1: The Number of Significant Harmonics for Each Parameter of the NSRP(Mexp) Model

Parameters Parameters	λ (Lamda)	η (Eta)	ν (Nu)	ξ (Xi)	β (Beta)	α (Alpha)	θ (Teta)
Significant Harmonics	2	5	4	5	4	5	5

Coefficients of the Fourier series (A_q and B_q) of each parameter based on the individual number of significant harmonics were evaluated using equation (10) and the result are given in Table 2. The equations for each parameter could then be easily formed with these values of A_q and B_q.

For example the Fourier Series equation for the parameter λ is

$$\begin{aligned}\lambda_m = 0.0218 &- 0.0009\cos\left(\frac{2\pi\ m}{12}\right) - 0.0035\sin\left(\frac{2\pi\ m}{12}\right)\\ &- 0.0048\cos\left(\frac{4\pi\ m}{12}\right) - 0.0086\sin\left(\frac{4\pi\ m}{12}\right)\end{aligned}$$

Graphic representation of Fourier Series estimation for selected parameters are shown in Figure 1: (a),(b) and (c). On the whole the Fourier Series managed to capture the seasonal pattern of rainfall activity in the Klang Valley. The graph of λ (Figure 1a) shows that the arrival of storms are the highest during the months of April and October. This finding agrees with the rainfall pattern of the Klang Valey [13]. The climate of Kuala Lumpur and its neighboring area is characterized by the northeast and southwest monsoons which occur in the months of December-March and June-September respectively, and two relatively short inter-monsoon seasons which occur during the months of April-May and October-November.

Generally, convective rains which are caused by short duration storms produce high amount of rainfall whilst stratiform rain which are brought by the northeast and southwest monsoon tend to have longer durations. Heavy thunderstorms are common during the inter-monsoon periods which explains why λ is relatively high in April and October . The parameters ξ and θ which represent the light rain cells and heavy rain cells respectively, also tend to be the highest for the month of October.

Graphical comparison between the moments of the observed data and the Fourier Series estimations is done by means of box plots. Box plots for each sample moments were constructed using 10 years of data. Using the Fourier Series coefficients of Table 2, the equivalent moments were derived and plotted. These are shown in Figure 2: (a) and (b), Figure 3: (c) and (d), Figure 4: (e) and (f) and Figure 5: (g) and (h). It is very obvious that the moments of the Fourier Series are captured within the range of the box plots. For each statistic the Fourier series estimations was practically within the desired limits. Although the autocorrelations at one, six and twenty four hours are not used in the estimation process, the Fourier Series coefficients were able to produce estimates reasonably well.

Table 2: Fourier Coefficients for Each Parameter of the NSRP(Mexp) Model

Parameters	Significant Harmonics	U_m	A_q	B_q
λ (Lamda)	1	0.0218	−0.0009	−0.0035
	2		−0.0048	−0.0086
η	1	3.0074	0.6893	−0.0854
	2		−0.7873	0.4432
	3		−0.4361	0.3790
	4		−0.4510	−0.1964
	5		−0.4338	0.9127
ν	1	7.5066	1.7428	−0.4031
	2		−0.3624	0.4526
	3		−0.1791	1.0150
	4		−1.0386	−1.4959
ξ	1	1.1716	−0.3361	−0.6400
	2		0.6775	−0.5402
	3		−0.3462	−0.2793
	4		−0.1742	0.3824
	5		0.4986	0.3614
β	1	0.3260	−0.0588	0.0184
	2		−0.0865	−0.0464
	3		0.0131	0.0113
	4		0.0356	0.0364
α	1	0.7214	−0.0179	−0.0746
	2		0.1262	−0.0568
	3		−0.0057	0.0116
	4		−0.0295	0.0085
	5		0.0853	−0.0175
θ	1	15.5630	0.6886	−2.0947
	2		−1.4675	0.7052
	3		−2.7306	−0.0400
	4		−1.8811	1.1123
	5		1.2411	4.1166

As such we can conclude that Fourier Series is able to adequately capture the statistical characteristics of the rainfall process.

4 Conclusion

This study demonstrated the ability of Fourier Series to describe the periodic seasonal fluctuation of parameters of NSRP(Mexp) model. Fourier Series does not only offer the convenience to express the seasonally fluctuating values of parameter but provide an economical alternative in terms of the reduced number of parameters needed to describe rainfall. Simulation of rainfall data with the reduction in parameters would be less cumbersome and could be done in a shorter time frame. It is anticipated that the proposed method be extended for multisite estimation.

Acknowledgement

We thank the Department of Irrigation and Drainage Malaysia for providing the rainfall data used in this study.

References

[1] P.S.P. Cowpertwait, *Further Developments of the Neyman-Scott Clustered Point Process for Modeling Rainfall,* Water Resources Research, 27(1991), 1431-1438.

[2] P.S.P. Cowpertwait, P.E. O'Connel, A.V. Metcalfe & J.A. Mawdsley, *Stochastic Point Process Modeling of Rainfall: (I) Single-site Fitting and Validation (II) Regionalization and Disaggregation,* Journal of Hydrology, 175(1996), 17-65

[3] P.S.P. Cowpertwait & C.G. Kilsby, *A Space-time Neyman-Scott Model of Rainfall: Empirical Analysis of extremes,* Water Resources Research, 38(8)(2002).

[4] D.R. Cox & V. Isham, *Point Processes,* Chapman and Hall, 1980.

[5] Q. Duan, S. Sorooshian & V. Gupta, *Effective and Efficient Global Optimization for Conceptual Rainfall-runoff Models,* Water Resources Research, 28(4)(1992), 1015-1031.

[6] Fadhilah Yusof, *Stochastic Modeling of Hourly Rainfall Processes,* Ph.D. Thesis, Universiti Teknologi Malaysia, 2008.

[7] C. Onof & H.S. Wheater, *Modeling of British Rainfall Using a Random Parameter Bartlett-Lewis Rectangular Pulse,* Journal of Hydrology, 149(1993), 67-95

[8] I. Rodriguez-Iturbe, *Scale of Fluctuation of Rainfall Models,* Water Resources Research, 22(1986), 15S–37S.

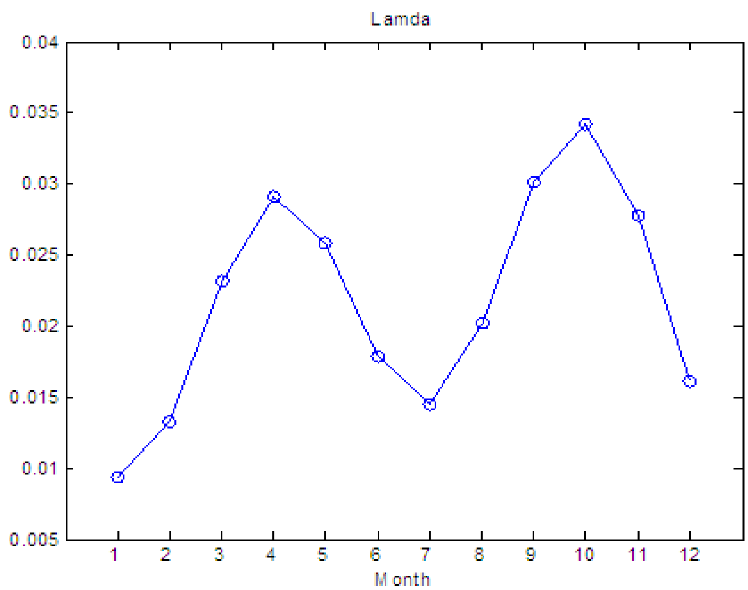
[9] I. Rodriguez-Iturbe, D.R. Cox & V. Isham, *Some Models for Rainfall Based on Stochastic Point Processes,* Proc. Roy. Soc. London Ser. A, 410(1839)(1987a), 269-288.

[10] I. Rodriguez-Iturbe, B. Febres De Power & J. Valdes, *Rectangular Pulses Point Process Models for Rainfall: Analysis of Empirical Data,* Journal of Geophysical Resources, 92(D8)(1987b), 9645-9656.

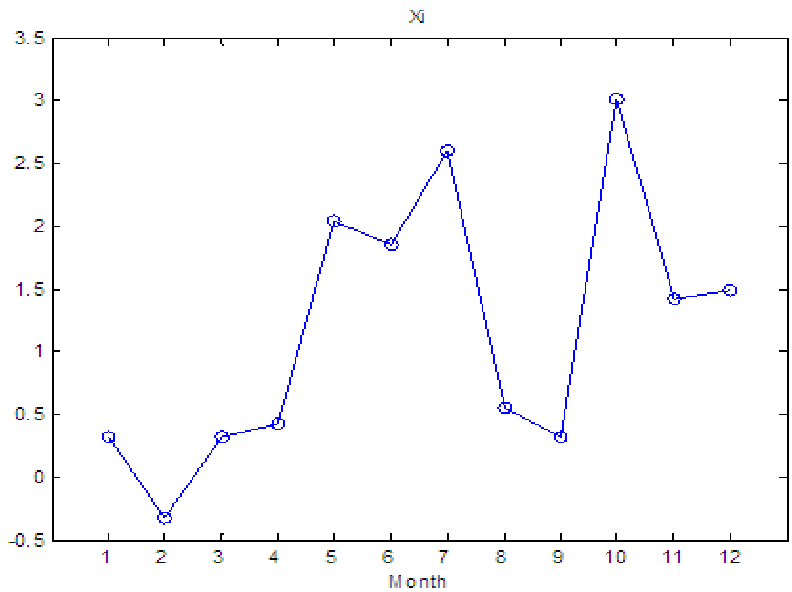
[11] J.D. Salas, J.W. Delleur, V. Yevjevich & W.L. Lane,*Applied Modeling of Hydrologic Series,* Water Resources Publications, 1980.

[12] J.A. Smith & A.F. Karr, *Parameter Estimation for a Model of Space-time Rainfall,* Water Resources Research 21(1985), 1251–1257.

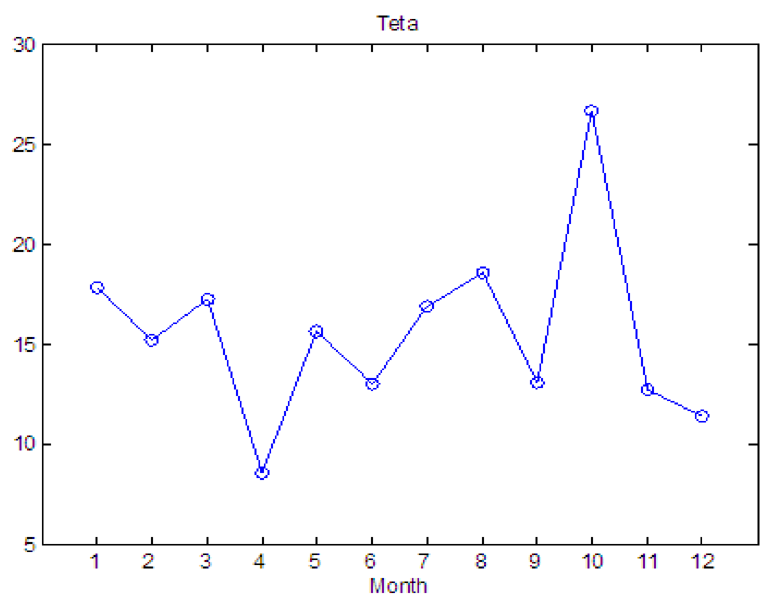
[13] M.D. Zalina, M.A. Nurul Huda, Y. Fadhilah, A. Maizah Hura, A. Robiah & W. Azli, *Exploring the Anomalies of Rainfall Events in the Klang Valley,* National Conference World Water Day 2007, Terengganu, Malaysia 15-16 April 2007.



(a) Lamda

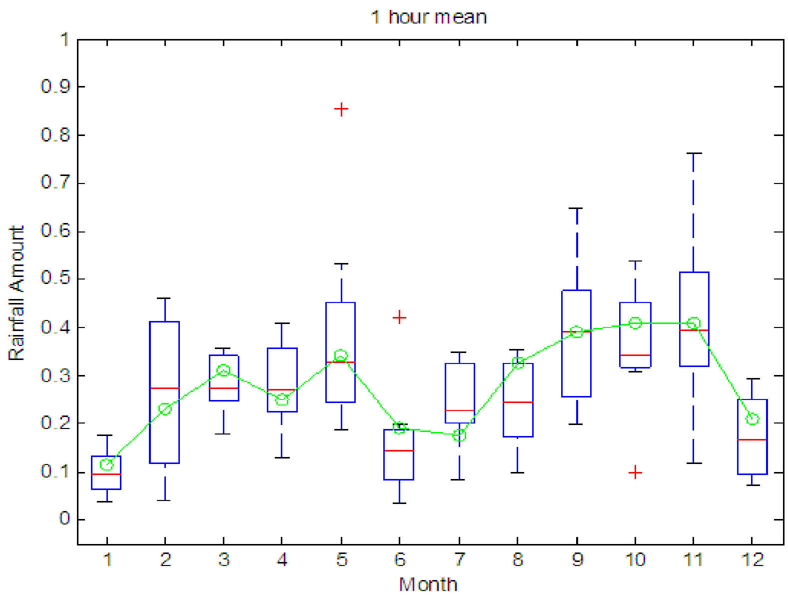


(b) Xi

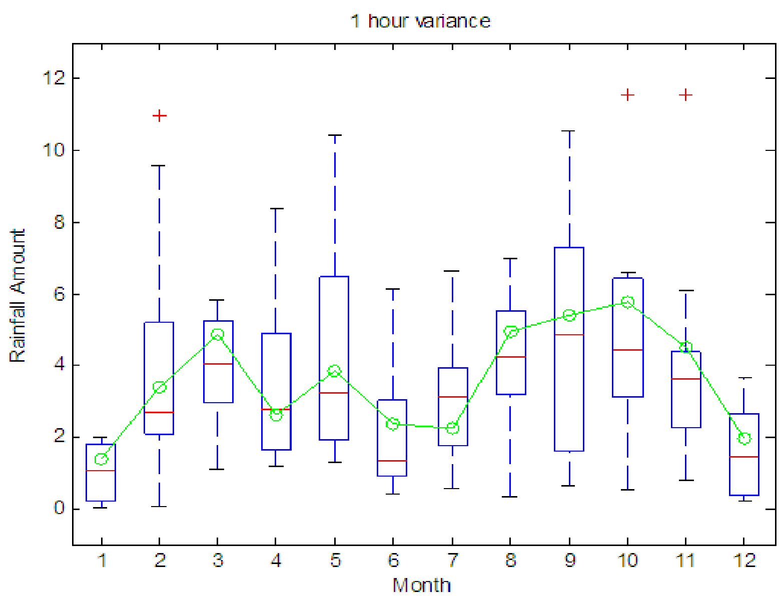


(c) Teta

Figure 1: (a)-(c) Shows Graphs of Fourier Series Estimation for the Parameters λ, ξ and θ

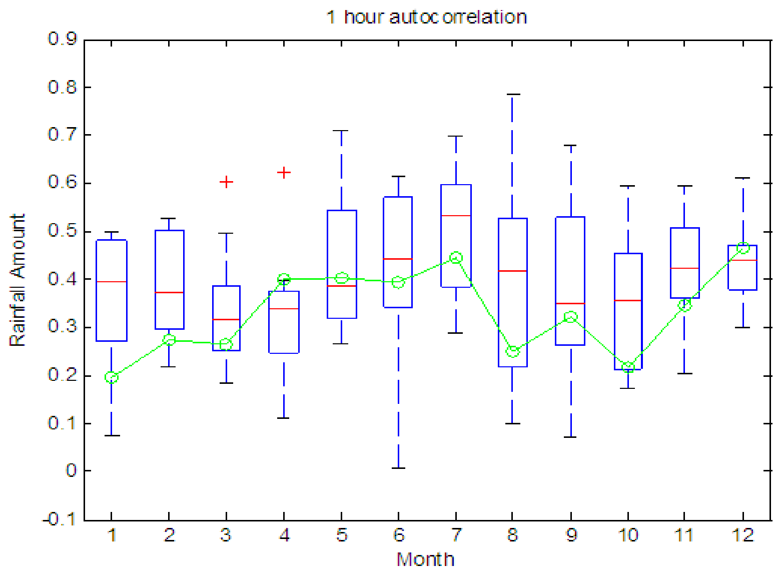


(a)

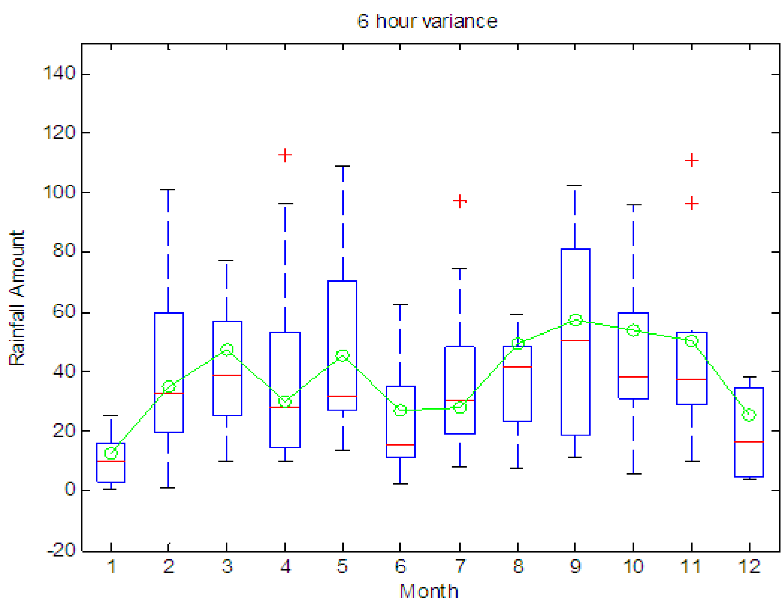


(b)

Figure 2: Comparison of Sample Moments of the Observed Properties (Box Plots) and Fourier

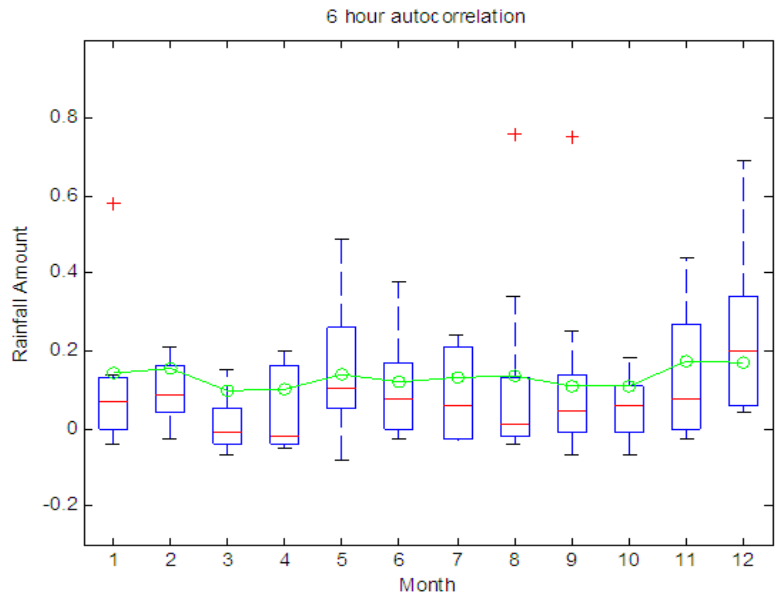


(c)

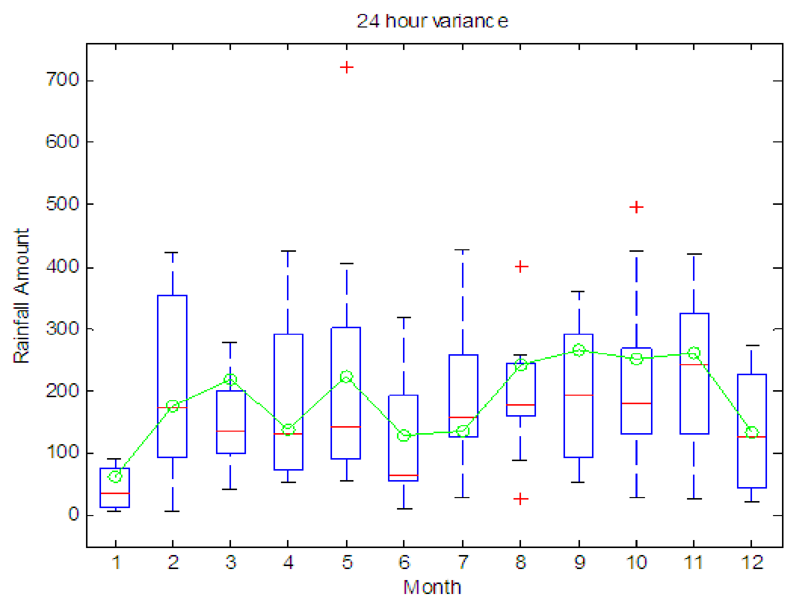


(d)

Figure 3: Comparison of Sample Moments of the Observed Properties (Box Plots) and Fourier(continued)

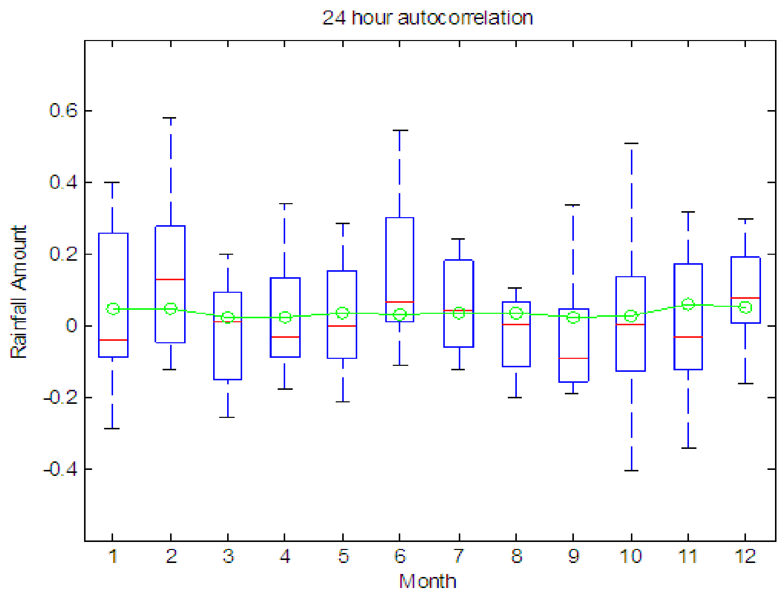


(e)

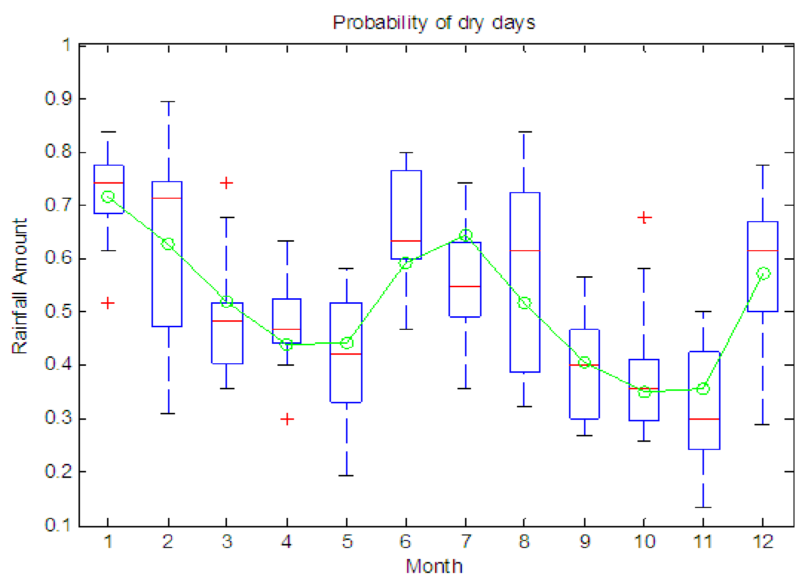


(f)

Figure 4: Comparison of Sample Moments of the Observed Properties (Box Plots) and Fourier(continued)



(g)



(h)

Figure 5: Comparison of Sample Moments of the Observed Properties (Box Plots) and Fourier(continued)