Research Article

Stochastic Modeling of Rainfall in Peninsular Malaysia Using Bartlett Lewis Rectangular Pulses Models

Ibrahim Suliman Hanaish,^{1,2} Kamarulzaman Ibrahim,² and Abdul Aziz Jemain²

¹ Department of Statistics, Faculty of Science, Misurata University, 4th Rounded Street, Misurata 051, Libya

² School of Mathematical Sciences, Faculty of Science and Technology, Universiti Kebangsaan Malaysia (UKM),
Selangor, 43600 Bangi, Malaysia

Correspondence should be addressed to Ibrahim Suliman Hanaish, henaish@yahoo.com
and Kamarulzaman Ibrahim, kamarulz@ukm.my

Received 5 July 2011; Revised 16 September 2011; Accepted 17 September 2011

Academic Editor: A. Mohamed

Copyright © 2011 Ibrahim Suliman Hanaish et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Three versions of Bartlett Lewis rectangular pulse rainfall models, namely, the Original Bartlett Lewis (**OBL**), Modified Bartlett Lewis (**MBL**), and 2N-cell-type Bartlett Lewis model (**BL2n**), are considered. These models are fitted to the hourly rainfall data from 1970 to 2008 obtained from Petaling Jaya rain gauge station, located in Peninsular Malaysia. The generalized method of moments is used to estimate the model parameters. Under this method, minimization of two different objective functions which involve different weight functions, one weight is inversely proportional to the variance and another one is inversely proportional to the mean squared, is carried out using Nelder-Mead optimization technique. For the purpose of comparison of the performance of the three different models, the results found for the months of July and November are used for illustration. This performance is assessed based on the goodness of fit of the models. In addition, the sensitivity of the parameter estimates to the choice of the objective function is also investigated. It is found that **BL2n** slightly outperforms **OBL**. However, the best model is the Modified Bartlett Lewis **MBL**, particularly when the objective function considered involves weight which is inversely proportional to the variance.

1. Introduction

Stochastic models can be used to generate rainfall across a range of different timescales for the purpose of reservoir design, flood studies, and design of sewerage systems. Recent developments in the area of hydrological modeling emphasize on characterizing rainfall processes at a single site or an areal average, as described in several works by Rodriguez-Iturbe et al. [1, 2], Cowpertwait [3, 4], Cowpertwait et al. [5], Onof et al. [6, 7], and Agizew and Michel [8]. These authors found that the rectangular pulse Poisson process models, particularly, the Bartlett-Lewis models, are successful in describing the rainfall processes for a wide range of temporal scale for countries such as UK, America, and Australia. The nature of rainfall in Peninsular Malaysia which consists of heavy rainfall in a short duration and light rainfall in a long duration can be well explained by some stochastic models. However, not much works have been done on the application

of Bartlett Lewis models for rainfall occurrence in Peninsular Malaysia. In this study, three different forms of Bartlett Lewis rectangular pulse rainfall models are fitted to the data which are available at Petaling Jaya rainfall gauge station in Peninsular Malaysia for the period from 1971 to 2008 using the generalized method of moments. In addition, a comparison of the performance of the models is made with the allowance for two objective functions considered under the generalized method of moments approach. For illustrative purposes, the analysis for July and November is provided since these two months are typical for the two different monsoons periods in Peninsular Malaysia.

2. Rainfall Models Considered

2.1. The Original BL (OBL) Model. The original BL model for point rainfall had been developed by Rodriguez-Iturbe et al. [1]. In this model, storms arrive randomly following

FIGURE 1: (a) Explanatory sketch for the structures of rainfall storms. (b) Explanatory sketch for the parameters of Bartlett-Lewis rectangular pulses model.

a Poisson distribution with parameter λ . Within each storm, cells arrive following a Poisson distribution with rate β . The generation of rectangular cells within a storm (cluster) stops after a given time. This storm duration is exponentially distributed with parameter γ . Each cell arrival is associated with a rectangular pulse which has an exponentially distributed duration with parameter η and an exponentially distributed depth of mean μ_x . The number of cells per storm follows a geometric distribution with mean $\mu_c = 1 + \beta/\gamma$. Thus, the **OBL** is a model which can be characterized by the set of parameters $(\lambda, \mu_x, \eta, \beta, \gamma)$, as given by Rodriguez-Iturbe et al. [1].

2.2. The Modified BL (MBL) Model. The Modified Bartlett-Lewis rectangular pulses model (**MBL**) was considered in many works due to its wide applicability for describing various different climates. The diagrammatic explanation of the Modified Bartlett-Lewis rectangular pulses model (**MBL**) is depicted in Figures 1(a) and 1(b), and the assumptions for the model are as follows: the storm origins are assumed to follow a Poisson process with rate λ and the cell origins follow a Poisson process with rate β . Cell arrivals terminate after a particular time, and this length of period is exponentially distributed with parameter γ . Each cell has a duration which is exponentially distributed with parameter η . The distribution of the uniform intensity is typically assumed exponential with a parameter μ_x . For each storm, the parameter η is randomly varied from storm to storm with a gamma distribution with shape parameter α and scale parameter ν ,

FIGURE 2: Two types of cells.

such that $E(\eta) = \alpha/\nu$ and $\text{var}(\eta) = \alpha/\nu^2$. Subsequently, parameters β and γ also vary in a manner that the ratios $\kappa = \beta/\eta$ and $\phi = \gamma/\eta$ be constant. Therefore, a 6 parameter **MBL** is described by the set of parameters $(\lambda, \mu_x, \alpha, \nu, \kappa, \phi)$. The equations of the BL model, in its original or the modified (random parameter) configuration, may be found in the appropriate references, such as Rodriguez-Iturbe et al. [1].

2.3. 2N-Cell-Type Bartlett Lewis Model (BL2n). The original and modified Bartlett Lewis Models only allow for the existence of one type of rectangular pulse, that is, one type of cell. Therefore, the parameter estimates for the intensity and duration of the cells in Bartlett Lewis Models are likely to be average values over the various types of precipitation that can occur in the same precipitation field. So, another possible modification to the original Bartlett Lewis Model (**OBL**) involves classification of more than one type of cells. The simplest possible model uses only one cell type but the literature suggests that the precipitation field can be broadly classed into two types: convective and stratiform precipitations. Many climates tend to experience both types precipitations, as shown in Figure 2. **BL2n** can be characterized by two random variables which represent the intensity of two types of cell, denoted as X_i , $i = 1, 2$. The random variables X_i 's follow an exponential with means μ_{x_i} , $i = 1, 2$. The duration of i th cell is exponentially distributed with parameters η_i , $i = 1, 2$. The probabilities of occurrence for the cell type is denoted by ϕ_i , $i = 1, 2$ such that $\sum_{i=1}^2 \phi_i = 1$. Thus, the 2-cell-type Bartlett Lewis Model (**BL2n**) is described by the set of parameters $(\lambda, \mu_{x_1}, \mu_{x_2}, \eta_1, \eta_2, \phi_1, \phi_2, \beta, \gamma)$.

3. Study Area and Input Data

In this study, hourly rainfall data were obtained from Petaling Jaya rain gauge station which is located in the midlands of Peninsular Malaysia. The hourly data ranging from the period of 1970 to 2008 were collected from the database of Malaysian Meteorological Service (MMS). This city experiences an equatorial climate and is very much influenced by the monsoons. Rainfall at Petaling Jaya is characterized by stratiform rainfall which occurs from December to February during the northeast monsoon and from May to August during the southwest monsoon, while the convective rainfall occurs during the months of March to April and September to November, which are the two intermonsoon periods. In particular, the intermonsoon periods are distinguishable by the higher mean hourly rainfall, as shown in Figure 3. Moreover, we can also see that the mean hourly rainfall for July is the lowest when compared to the other months of the year while for the period of August to December an increasing trend can be observed. July is the driest month of

FIGURE 3: Observed and fitted properties for the original and modified models.

TABLE 1: Fitting properties and weights for the two objective functions for July and November.

Fitting properties	Time scales	Weight			
		July		November	
		OF1	OF2	OF1	OF2
Mean	1 hour	81.14521	1.026457313	20.76888	0.1567646
	24 hour	0.00092619	0.00005140	0.00057534	0.00001170
Variance	6 hour	0.00000969	0.00000060	0.00000395	0.00000009
	24 hour	0.00000046	0.00000002	0.00000024	0.000000004
Lag-1 autocorrelation	1 hour	625.5488	10.64053	1070.488	8.021788
	24 hour	1125.354	337.9444	1063.028	708.9913
Probability of wet	1 hour	16496.9	481.3714	4358.29	83.92131
	24 hour	1520.041	11.02459	841.5213	3.078955

TABLE 2: Optimal values based on the two objective functions for each model fitted to July and November data.

Model	No. of parameters	July		November	
		OF1	OF2	OF1	OF2
1	5	3.62524015	0.08939851	4.30727375	0.06258780
2	6	3.17993491	0.08559614	4.23920085	0.06369667
3	9	3.60225088	0.00245042	0.39245869	0.02786144

TABLE 3: Observed rainfall statistics for July with corresponding properties for the fitted models.

Fitting properties	Time scales	Observed	Fitting using model					
			Model 1		Model 2		Model 3	
			OF1	OF2	OF1	OF2	OF1	OF2
Mean	1 hour	0.987	1.11	1.176	1.075	1.147	1.129	0.986
	1 hour	139.47	110.59	121.42	112.36	121.26	120.23	140.22
Variance	6 hour	1285.54	1137.15	1240.99	1188.2	1255.12	1122.33	1324.92
	24 hour	6055.21	5749.17	6240.82	6083.84	6298.91	4912.63	5823.81
Lag-1 autocorrelation	1 hour	0.306	0.306	0.301	0.305	0.300	0.299	0.307
	24 hour	0.054	0.057	0.054	0.064	0.054	0.051	0.054
Probability of wet	1 hour	0.045	0.036	0.037	0.035	0.037	0.047	0.045
	24 hour	0.301	0.301	0.311	0.306	0.320	0.279	0.301

the year for the Peninsular. The autocorrelation coefficient of lag 1 (ACF(1)) for the hourly data is relatively greater than 0.35; that is, $ACF(1) > 0.35$ for the months of November to February as compared to the other periods of the year. This is not surprising since many locations in the country, particularly, northeast areas, experience heavy rainfall over a long duration during this period of time.

4. Model Fitting

There is some difficulty in applying the standard techniques of the Bartlett Lewis model such as maximum likelihood method for estimating the parameters of model since the likelihood is not available in a closed form. Estimation of parameters in these models is usually carried out using a generalized method of moments (GMM). Specifically, let $\theta = (\theta_1, \theta_2, \dots, \theta_p)$ be the parameter vector for the model given the observed rainfall data y , let $\mathbf{T}(y) = (T_1(y), T_2(y), \dots, T_k(y))$ be a vector of summary statistics computed from the data, and let $\tau(\theta) = (\tau_1(\theta), \tau_2(\theta), \dots, \tau_k(\theta))$ denote a vector of the fitted value of $\mathbf{T}$ under the model. The idea behind the method of moments is to choose θ to minimize the objective function given by

$$S(\theta) = [\mathbf{T} - \tau(\theta)]' \mathbf{W} [\mathbf{T} - \tau(\theta)], \quad (1)$$

where $\mathbf{w}$ is a $k \times k$ positive definite weighting matrix of “weights” which is determined based on historical data. A special case of (1) can be given by

$$S(\theta) = \sum_{i=1}^k w_i [T_i(y) - \tau_i(\theta)]^2, \quad (2)$$

where $(w_i; i = 1, 2, \dots, k)$ is a collection of positive weights. Therefore, this objective function is minimized in order to

reduce the error between the observed statistics (T_i) and the fitted statistics (τ_i). Minimization of the objective function $S(\theta)$ is done using the Nelder-Mead optimization algorithm. Accordingly, the statistics considered in this work are 1-hour mean (Mean1), 1-hour-variance (Var1), 6-hour-variance (Var6), 24-hour-variance (Var24), 1-hour-autocorrelation of lag-1 (ACF1(1)), 24-hour-autocorrelation of lag-1 (ACF1(24)), 1-hour probability of wet (Pwet1), and 24-hour-probability of wet (Pwet24) as suggested by Rodriguez-Iturbe et al. [1]. The parameter estimates are estimated using the software by Chandler and Lourmas [9].

5. Model Performance

The three Bartlett Lewis models considered require the assumption that the historical time series is stationary, as suggested by Cox and Isham [10]. Therefore, based on Rodriguez-Iturbe et al. [11], we calibrate the models for each month separately so that monthly rainfall forms a stationary time series. In addition to comparison of the model fits, we have also explored the effect of using different objective functions of the form (2). The two objective functions considered have different weights. The first objective function involves the weight which is based on the i th diagonal elements of the matrix $\mathbf{W}$, given by $w_i = 1/\text{Var}(T_i(y))$, where $\text{Var}(T_i(y))$ represents the variance of the i th statistics across all years. The second objective function involves the weight which is $w_i = [1/\overline{T_i(y)}]^2$, where $\overline{T_i(y)}$ represents the mean of the i th statistics across all years. These two objective functions are denoted by OF1 and OF2. In assessing the different models and objective functions, the criteria that are considered include the reproduction of fitting properties and sensitivity of the parameter estimates to the choice of objective functions.

TABLE 4: Observed rainfall statistics for November with corresponding properties for the fitted models.

Fitting properties	Time scales	Observed	Fitting using model					
			Model 1		Model 2		Model 3	
			OF1	OF2	OF1	OF2	OF1	OF2
Mean	1 hour	2.525	2.757	2.907	2.731	2.883	2.529	2.764
Variance	1 hour	292.340	267.44	274.67	266.84	275.27	293.04	306.103
	6 hour	3323.101	2956.60	3051.27	2975.96	3063.33	3229.11	3120.63
	24 hour	14358.21	14767.31	14977.62	14958.63	14992.73	14636.98	13696.22
Lag-1 autocorrelation	1 hour	0.353	0.353	0.353	0.353	0.353	0.354	0.361
	24 hour	0.037	0.045	0.037	0.048	0.037	0.019	0.037
Probability of wet	1 hour	0.109	0.086	0.092	0.086	0.091	0.109	0.110
	24 hour	0.569	0.574	0.603	0.577	0.612	0.569	0.512

TABLE 5: A comparison of results found using OF1 and OF2 on the parameter estimates and 95% confidence intervals for data of July based on model 2.

Parameter	Estimate	OF1		Estimate	OF2	
		Lower	Upper		Lower	Upper
λ	0.01373	0.01274	0.01508	0.01446	0.00009	234.10951
μ_x	111.47220	94.31399	807	108.16010	0.00009	171.89142
α	5.89948	2	20	8.75090	3.16774	20
α/ν	1.24472	0.52487	10	2.25727	0.3	10
κ	0.12259	0.00999	0.23827	0.11244	0.010	10
ϕ	0.06943	0.02999	0.15593	0.0740	0.3	1

6. Results and Discussion

Three models have been fitted for the data of each month of the year across all the years using the two objective functions considered. For illustration of the results, we provide the analysis for the months of July and November, as representing results for the dry and wet periods, respectively. The values of weights for the two objective functions with respect to the different properties and time scales that are considered for July and November are given in Table 1.

6.1. Reproduction of Fitting Properties. As a check on the minimization, it is worth comparing the values of objective functions for each nested models. Both models 2 and 3 are extensions of model 1. Based on each objective function, the optimal values, that is, the minimum values attained for each model, for the months of July and November are reported in Table 2. For the month of July, the optimal values are found smallest for model 3 when OF2 is considered; however, the optimal values are found smallest for model 2 when OF1 is considered. On the hand, for the month of November, we found that the optimal values are smallest for model 3 for both objective functions. Therefore, it is not very clear for the choice of the best model based on the optimal values that are found. In addition, although we observed a lower optimal value for model 3 for the month of November, it is not always guaranteed that it would be the case since model 3 is not an extension of model 2.

To compare the models based on the influence of objective function, as suggested by Wheater et al. [12], it is useful

to calculate the objective function thresholds. Unfortunately, due to numerical instabilities, it has not been possible to calculate the thresholds for model 3, and this indicates that model 3 may be overparameterised. The 95% thresholds found for model 2 based on July data are **23.2092** and **1.65164** for OF1 and OF2, respectively. These thresholds are similar to those obtained from model 1, and it therefore seems that there is little justification for using models with more than 6 parameters, which is model 2. The thresholds for model 2: 95% thresholds are **25.05906** and **2.730417**, respectively for November. The objective function values give an overall measure of model fit, but do not indicate how well the models reproduce the particular properties. Tables 3 and 4 show all the observed fitting properties for July and November, respectively, for all the models using FO1 and FO2.

When the fitted means are compared to the observed means for all the months, it is found that the fitted statistics are generally larger. This could possibly be due to many extreme values in the data set due to the large readings of rainfall amount. There is a tendency for the models to underestimate the 6-hour variance and overestimate the 24-hour variance in both months. FO2 is found to be more accurate than FO1 in estimating the variance since the estimated statistics are closer to the observed values under FO2, particularly for model 2. It is also clear from Tables 3 and 4 that, however, there are no remarkable results in the fitted autocorrelations and probability of wet when the different time scales are considered. Overall, model 2 is considered as the best model and it will be referred for further discussion.

TABLE 6: A comparison of results found using OF1 and OF2 on the parameter estimates and 95% confidence intervals for data of November based on model 2.

Parameter	Estimate	OF1		Estimate	OF2	
		Lower	Upper		Lower	Upper
λ	0.03233	0.03188	0.03255	0.03622	0.00078	0.76313
μ_x	84.9667	82.9582	87.17459	82.0044	0.00009	708
α	16.0523	11.1561	20	20	2	20
α/ν	5.73276	2.55788	8.04747	7.1379	0.3	10
κ	0.16848	0.14244	0.19581	0.2035	0.01	10
ϕ	0.10462	0.09577	0.11424	0.12849	0.03	1

The plots of the fitted properties for the original and modified models against the observed properties are shown in Figure 3.

6.2. Effect of Objective Function on Parameter Estimates. The sensitivity of the estimated parameters to the choice of objective functions is presented using the estimates for model 2. The model fitting in this stage is mainly to ensure the stability in the numerical optimization of the objective function. The parameter estimates under each of the objective functions are given in Tables 5 and 6 for July and November, respectively. The effect of reweighting the fitting properties can be seen in moving from OF1 to OF2. Overall, the parameter estimates of both July and November are more accurate under OF1 than OF2 because the lengths of the intervals are found shorter.

6.3. Conclusion. Short duration heavy rainfall and long duration light rainfall often occur in Peninsular Malaysia, and it seems that a good fitting to standard statistics is unlikely to be reproduced adequately using the original form of Bartlett Lewis model (**OBL**). A new modified Bartlett Lewis stochastic rainfall model known as 2n-cell-type Bartlett Lewis model (**BL2n**) is considered. The **BL2n** model structure is an extension to the Original Bartlett-Lewis model, which allows for more than one type of cells. The hourly rain gauge data obtained from Petaling Jaya was applied in model fitting for all months. On the basis of comparison for dry and wet periods, July and November were considered. Two different sets of weights, based on the reciprocal of the i th diagonal elements w_i of the matrix **W** in (2), that is, $1/\text{Var}(T_i(y))$, and based on the inverse of the squared of mean values, two objective functions were considered for the model fitting. The **BL2n** model outperforms the **OBL** possibly due to the nature of rainfall in Malaysia which consists of two types, that is, convective and stratiform. However, the best model is the Modified Bartlett Lewis **MBL** since it is able to reproduce the fitting properties better than other models. Overall, the parameter estimates of both July and November under objective function 1 are more accurate than objective function 2. There is also a tendency for parameters to be less well identified in July than in November.

References

- [1] I. Rodriguez-Iturbe, D. R. Cox, and V. Isham, "Some models for rainfall based on stochastic point processes," *Proceedings of The Royal Society of London, Series A*, vol. 410, no. 1839, pp. 269–288, 1987.
- [2] I. Rodriguez-Iturbe, D. Febres, and J. B. Valdes, "Rectangular pulses point process models for rainfall: analysis of empirical data," *Journal of Geophysical Research*, vol. 92, no. 8, pp. 9645–9656, 1987.
- [3] P. S. P. Cowpertwait, "A generalized point process model for rainfall," *Proceedings of The Royal Society of London, Series A*, vol. 447, no. 1929, pp. 23–37, 1994.
- [4] P. S. P. Cowpertwait, "A spatial-temporal point process model of rainfall for the Thames catchment, UK," *Journal of Hydrology*, vol. 330, no. 3–4, pp. 586–595, 2006.
- [5] P. S. P. Cowpertwait, P. E. O'Connell, A. V. Metcalfe, and J. A. Mawdsley, "Stochastic point process modelling of rainfall," *Journal of Hydrology*, vol. 175, pp. 17–65, 1996.
- [6] C. Onof, R. E. Chandler, A. Kakou, P. Northrop, H. S. Wheeler, and V. Isham, "Rainfall modelling using poisson-cluster processes: a review of developments," *Stochastic Environmental Research and Risk Assessment*, vol. 14, no. 6, pp. 384–411, 2000.
- [7] C. Onof, T. John, and K. Richard, "Comparison of two hourly to 5-min rainfall disaggregators," *Atmospheric Research*, vol. 77, no. 1–4, pp. 176–187, 2005.
- [8] N. E. Agizew and E. Michel, "Characterization and disaggregation of daily rainfall in the Upper Blue Nile Basin in Ethiopia," *Journal of Hydrology*, vol. 399, no. 3–4, pp. 226–234, 2011.
- [9] R. E. Chandler and Lourmas, "Software for moment-based fitting of single-site stochastic rainfall model fitting, user guide," in *DEFRA Project FD2105*, Department of Statistical Science, University College London, London, UK, 2010.
- [10] D. R. Cox and V. Isham, *Point Processes*, Chapman & Hall, London, UK, 1980.
- [11] I. Rodriguez-Iturbe, B. Febres de Power, and J. B. Valdes, "Rectangular pulses point process models for rainfall: analysis of empirical data," *Journal of Geophysical Research*, vol. 92, no. 8, pp. 9645–9656, 1987.
- [12] H. S. Wheeler et al., "Improved methods for national spatial-temporal rainfall and evaporation modeling for BSM," in *Joint Defra/ EA Flood and Coastal Erosion Risk Management R&D Programme F2105/TR*, Department of Civil and Environmental Engineering, Imperial College London, Department of Statistical Science, University College London, Centre for Ecology and Hydrology, Wallingford, UK, 2006.

- [1] I. Rodriguez-Iturbe, D. R. Cox, and V. Isham, "Some models for rainfall based on stochastic point processes," *Proceedings of*