

# Comparison of Stochastic Point Process Models of Rainfall in Singapore

Y. Lu, X. S. Qin

**Abstract**—Extensive rainfall disaggregation approaches have been developed and applied in climate change impact studies such as flood risk assessment and urban storm water management. In this study, five rainfall models that were capable of disaggregating daily rainfall data into hourly one were investigated for the rainfall record in the Changi Airport, Singapore. The objectives of this study were (i) to study the temporal characteristics of hourly rainfall in Singapore, and (ii) to evaluate the performance of various disaggregation models. The used models included: (i) Rectangular pulse Poisson model (RPPM), (ii) Bartlett-Lewis Rectangular pulse model (BLRPM), (iii) Bartlett-Lewis model with 2 cell types (BL2C), (iv) Bartlett-Lewis Rectangular with cell depth distribution dependent on duration (BLRD), and (v) Neyman-Scott Rectangular pulse model (NSRPM). All of these models were fitted using hourly rainfall data ranging from 1980 to 2005 (which was obtained from Changi meteorological station). The study results indicated that the weight scheme of inversely proportional variance could deliver more accurate outputs for fitting rainfall patterns in tropical areas, and BLRPM performed relatively better than other disaggregation models.

**Keywords**—rainfall disaggregation, statistical properties, Poisson process, Bartlett-Lewis model, Neyman-Scott model

## I. INTRODUCTION

RAINFALL is one of the most important inputs for hydrological modeling, and indicators for climate change impact studies. The rainfall data at finer timescales is generally more useful for hydrological process research, especially in the field of extreme rainfall-event evaluation and flood risk management [1]. In the world wide, rainfall data are usually available at timescales of daily or monthly. The short interval records are limited due to various reasons such as costly data procurement and complex geographical conditions. Over the past decades, a large variety of disaggregation methods were proposed and used to provide possible realization of hourly data which were aggregated up to the given daily data. Rodriguez-Iturbe et al. [2] first introduced the temporal rainfall models based on the clustered Poisson process. Based on such a theory, rainfall event was divided into a cluster of rain cells, where the temporal location of the cells relative to the event origin was specified by either Bartlett-Lewis or Neyman-Scott clustering mechanism [3]. The cluster models were suited in summarizing rainfall statistical characteristics by a group of parameters and regenerating the hierarchical rainfall structures conveniently [3].

Y. Lu is with the School of Civil & Environmental Engineering, Nanyang Technological University, 50 Nanyang Ave., 679798 Singapore (e-mail: ylu11@e.ntu.edu.sg).

X. S. Qin is with the School of Civil & Environmental Engineering and Earth Observatory of Singapore (EOS), Nanyang Technological University, Singapore (\*Correspondence author e-mail: xsqin@ntu.edu.sg).

Rodriguez-Iturbe et al. [1], [4], [5] described three models, including rectangular pulse Poisson model (RPPM), Bartlett-Lewis rectangular pulse model (BLRPM) and Neyman-Scott rectangular pulse model (NSRPM), to represent rainfall intensity where the clustering effect of rainfall process was accounted for. BLRPM has many modifications, such as the modified Bartlett-Lewis model (MBL) [6], Bartlett-Lewis model with 2 cell types (BL2C), and Bartlett-Lewis Rectangular with cell depth distribution dependent on duration (BLRD) [7]. Extensions to NSRPM include specialization of rain cell [8] and a spatial-temporal NSRPM [9] etc. NSRPM was also used in RainSim software [10]. However, there were limited studies for comparing multiple versions of BLRPM and NSRPM.

Many of the above-mentioned models have been successfully applied in US, Canada and UK, at different temporal scales and ranges; whereas, very limited studies were reported for Southeast Asia. One attempt was made by Hanaish et al. [11], where three versions of Bartlett-Lewis rectangular pulse models were applied in Malaysia. Singapore is a typical region with tropical climate. Characterized by heavy rainfall with short durations, its overall climatic features is important to the global climate change studies. Moreover, for evaluating the suitability of current storm water management systems under changing climatic conditions, the rainfall patterns in finer time resolutions for future conditions will need to be investigated. In this study, five models that could disaggregate rainfall from daily to hourly time series were evaluated. The rainfall dataset of Changi meteorological station, which is located in the eastern part of Singapore, was chosen for demonstration. We also compared the fitted precision using different weight schemes in the process of parameter estimations.

## II. MODEL DESCRIPTION

(1) *Rectangular pulse Poisson model (RPPM)*. The RPPM was developed with the following assumption. The occurrence of rainfall events follows the Poisson process as follows [4].

$$P_i = \frac{\lambda^i e^{-\lambda}}{i!}, i = 0, 1, 2, \dots \quad (1)$$

where  $e$  is the base of the natural logarithm,  $i$  is the number of occurrences of an event,  $\lambda$  is a rate parameter (a positive real number) which is equal to the expected number of occurrences during the given time interval. The mean and standard deviation of the single pulse intensity are denoted by parameters  $\mu_x$  and  $\sigma_x$ , respectively. The corresponding quantity for duration  $L$  is denoted as  $\mu_L$ . The RPPM can be characterized by a set of parameters:  $(\lambda, \mu_x, \sigma_x, \mu_L)$ . Readers are referred to Rodriguez-Iturbe et al. [4] for details.

(2) *Bartlett-Lewis rectangular pulse model (BLRPM)*. This cluster-based model was proposed by Rodriguez-Iturbe et al. [2]. For BLRPM, the intervals between successive cells are independent and identically distributed [4]. The rainfall arrives in a Poisson process at a rate of  $\lambda$ . Each rain cell also follows a Poisson distribution at a rate of  $\beta$ . In BLRPM, the rainfall and cell durations are by default taken as having exponential distributions, and denoted by parameters  $\gamma$  and  $\eta$ , respectively [5]. The mean cell intensity distribution is characterized by  $\mu_x$ . The BLRPM can therefore be described by a set of parameters  $(\lambda, \beta, \mu_x, \gamma, \eta)$ . The related theory was fully described by Rodriguez-Iturbe et al. [4]. The explanatory sketches of BLRPM were shown in Figure 1(a).

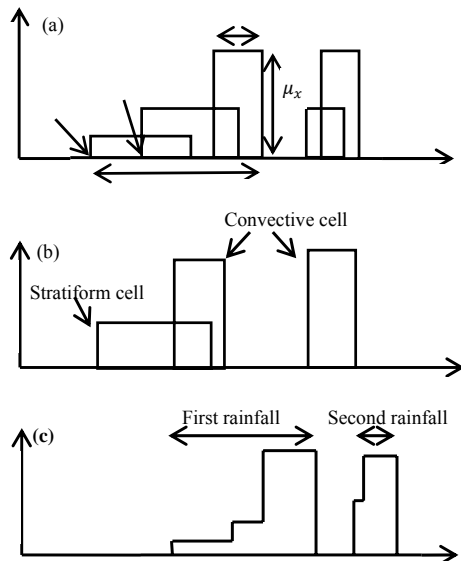


Fig. 1 Explanatory sketch of (a) BLRPM, (b) BL2C and (c) NSRPM [4] [5] [11]

(3) *Bartlett-Lewis model with 2 cell types (BL2C)*. The original Bartlett-Lewis model only shows one type of rain cell within the same rainfall event. However, observational studies on precipitation field have shown that multiple types of cells may exist in one storm, such as heavy convective rain cell and light stratiform rain cell. They are featured by: (i) two random variables for the intensity distribution  $\{x_i, i = 1, 2\}$ , with means  $\{\mu_{x_i}, i = 1, 2\}$  and mean square intensities  $\{\mu_{x_i^2}, i = 1, 2\}$ ; (ii) two duration distributions with exponential parameters  $\{\eta_i, i = 1, 2\}$ ; (iii) two probabilities  $\{\Psi_i, i = 1, 2\}$  of occurrences for cell type,  $\sum_{i=1}^2 \Psi_i = 1$ . This model is defined by 9 parameters, i.e.  $(\lambda, \mu_{x_1}, \mu_{x_2}, \eta_1, \eta_2, \Psi_1, \Psi_2, \beta, \gamma)$  [11]. The explanatory sketch of cell types are shown in Figure 1(b).

(4) *Bartlett-Lewis Rectangular with cell depth distribution dependent on duration (BLRD)*. The dependent depth-duration model can introduce dependence between cell intensity and cell duration distribution to improve the wet-dry properties [12]. The parameter set of  $(\lambda, \beta, \gamma, \eta)$  of BLRPM are also available for this model, representing the rainfall arrival, cell arrival, cell duration and cell duration, respectively. But the

cell intensities  $X$  are now specified through the distribution of  $X$  conditional upon the cell duration  $L$ . The conditional mean and standard deviation of cell intensity limit for 0 cell duration are denoted by  $\mu_{x|0}$  and  $\sigma_{x|0}$ . The mean durations of cell and rainfall event are distributed with parameters  $\delta_c$  and  $\delta_s$ , where  $\delta_c = 1/\eta$  and  $\delta_s = 1/\gamma$ . The mean number of cells in each rainfall event is denoted by  $\mu_c$ , and  $\mu_c = 1 + \beta/\gamma$ . Therefore, the mechanistic parameters for BLRD are  $(\lambda, \mu_x, \mu_{x|0}, \mu_c, \sigma_x, \sigma_{x|0}, \delta_c, \delta_s)$  [12].

(5) *Neyman-Scott rectangular pulse model (NSRPM)*. In this model, the position of cells is determined by a set of independent and identically distributed random variables, where the location of this distribution is given by the rainfall origin [5]. The parameters of  $\lambda, \beta, \gamma$ , and  $\eta$  also represent the rainfall arrival, cell arrival, rainfall duration and cell duration, respectively. The number of cells per rainfall event is an independent random variable, where the corresponding mean is  $\mu_c$ . The explanatory of NSRPM is depicted in Figure 1(c).

Among the above-mentioned models, RPPM is the fundamental form for all other versions that are characterized as rectangular pulses models. BL2C and BLRD are modified versions based on BLRPM. Both BLRPM-type models and NSRPM are empirical and it is normally difficult to simply choose which one is superior for specific applications [4].

### III. STUDY AREA AND DATA INPUT

In this study, the hourly rainfall dataset based on ground observation from Changi meteorological station was used. The station is located in the eastern part of Singapore and the period of the dataset is from 1980 to 2005. Singapore's climate is classified as tropical rainforest climate (*Köppen climate classification Af*), without obvious seasonal variations. The annual rainfall normally exceeds 2300mm [13].

There are two monsoon seasons for Singapore. The Northeast Monsoon occurs from December to early March, and Southwest Monsoon season occurs from June to September [14]. Based on rainfall levels, the wet season is from November to January and the rest months are relatively dry. The December and February are the wettest and driest month per year, respectively. Therefore, these two months will be chosen as typical months to fit different weight schemes when estimating parameters. Since both months are in the Northeast Monsoon, we also take June into consideration for comparison, as it falls in the Southwest monsoon which has relatively constant rainfall patterns.

### IV. PARAMETER ESTIMATION AND MODEL FITTING

The parameters are estimated based on the method of moments. Specifically, let  $T = (T_1, T_2, \dots)$  be the vector of the observed rainfall summary statistics, and let  $\tau(\theta) = [\tau_1(\theta), \tau_2(\theta), \dots]$  be the expected value of  $T$  according to the model,  $\theta$  is a parameter vector. The fundamental method of moments is to choose the best  $\theta$  that can minimize a quadratic function [11]:

$$S(\theta) = [T - \tau(\theta)]' W [T - \tau(\theta)] \quad (2)$$

Where  $W$  is the weight matrix.

TABLE I

THE OBSERVED AND FITTED RAINFALL STATISTICAL PROPERTIES FOR FEBRUARY UNDER TWO WEIGHT SCHEMES

Properties	Timescales	Observed	RPPM		BLRPM		BL2C		BLRD		NSRPM	
			W <sub>1</sub>	W <sub>2</sub>	W <sub>1</sub>	W <sub>2</sub>	W <sub>1</sub>	W <sub>2</sub>	W <sub>1</sub>	W <sub>2</sub>	W <sub>1</sub>	W <sub>2</sub>
Mean	1 hour	0.163	0.163	0.163	0.163	0.187	0.162	0.163	0.162	0.163	0.163	0.151
	1 hour	2.638	1.899	1.937	2.449	2.168	2.429	2.144	2.422	2.145	2.448	2.173
	6 hour	29.872	36.707	32.105	33.236	31.484	32.844	32.154	32.611	32.200	33.273	30.612
Variance	24 hour	143.783	197.603	158.716	151.208	144.124	169.685	153.233	172.408	152.950	152.225	158.168
	1 hour	0.268	0.292	0.332	0.269	0.273	0.266	0.253	0.272	0.274	0.271	0.278
Lag-1 autocorrelation	24 hour	0.080	0.009	0.011	0.079	0.079	0.070	0.076	0.084	0.077	0.076	0.083
	1 hour	0.046	0.043	0.040	0.046	0.047	0.046	0.046	0.046	0.046	0.046	0.048
Probability of wet	24 hour	0.306	0.314	0.345	0.306	0.263	0.306	0.306	0.307	0.307	0.307	0.328

TABLE II

THE OBSERVED AND FITTED RAINFALL STATISTICAL PROPERTIES FOR DECEMBER UNDER TWO WEIGHT SCHEMES

Properties	Timescales	Observed	RPPM		BLRPM		BL2C		BLRD		NSRPM	
			W <sub>1</sub>	W <sub>2</sub>	W <sub>1</sub>	W <sub>2</sub>	W <sub>1</sub>	W <sub>2</sub>	W <sub>1</sub>	W <sub>2</sub>	W <sub>1</sub>	W <sub>2</sub>
Mean	1 hour	0.418	0.418	0.518	0.417	0.414	0.418	0.423	0.467	0.494	0.418	0.517
	1 hour	6.185	3.650	2.127	5.214	4.631	5.330	4.588	7.019	6.951	5.220	2.127
	6 hour	67.084	78.127	55.207	75.414	75.709	75.901	78.296	61.087	52.195	77.298	55.207
Variance	24 hour	390.982	454.415	470.450	403.761	408.910	377.871	387.395	343.920	359.788	365.269	470.450
	1 hour	0.280	0.328	0.331	0.283	0.286	0.285	0.283	0.285	0.282	0.278	0.286
Lag-1 autocorrelation	24 hour	0.182	0.011	0.011	0.180	0.181	0.183	0.181	0.188	0.178	0.177	0.182
	1 hour	0.131	0.127	0.145	0.131	0.130	0.131	0.131	0.137	0.132	0.131	0.145
Probability of wet	24 hour	0.624	0.637	0.472	0.624	0.638	0.624	0.622	0.535	0.501	0.624	0.472

TABLE III

THE OBSERVED AND FITTED RAINFALL STATISTICAL PROPERTIES FOR JUNE UNDER TWO WEIGHT SCHEMES

Properties	Timescales	Observed	RPPM		BLRPM		BL2C		BLRD		NSRPM	
			W <sub>1</sub>	W <sub>2</sub>	W <sub>1</sub>	W <sub>2</sub>	W <sub>1</sub>	W <sub>2</sub>	W <sub>1</sub>	W <sub>2</sub>	W <sub>1</sub>	W <sub>2</sub>
Mean	1 hour	0.174	0.175	0.175	0.174	0.202	0.174	0.175	0.175	0.188	0.174	0.197
	1 hour	2.818	2.873	2.858	2.814	1.805	2.401	1.600	2.296	1.946	2.058	1.771
	6 hour	25.472	24.337	25.977	25.587	22.660	28.005	21.443	28.193	23.711	28.492	22.223
Variance	24 hour	114.008	101.939	109.901	123.358	136.471	123.430	147.371	126.001	136.599	131.830	139.496
	1 hour	0.228	0.235	0.284	0.228	0.228	0.226	0.225	0.228	0.229	0.228	0.229
Lag-1 autocorrelation	24 hour	0.018	0.007	0.009	0.018	0.018	0.013	0.017	0.022	0.018	0.018	0.018
	1 hour	0.047	0.045	0.039	0.047	0.050	0.047	0.047	0.046	0.052	0.045	0.050
Probability of wet	24 hour	0.423	0.434	0.494	0.423	0.315	0.423	0.422	0.426	0.313	0.434	0.322

The Nelder-Mead optimization algorithm can be used to minimize the objective function [11]. The observed rainfall summary statistics applied in the study are: 1-hour mean (Mean1), 1-hour variance (Var1), 6-hours variance (Var6), 24-hours variance (Var24), 1-hour autocorrelation of lag 1 (Ac1), 24-hours autocorrelation of lag 1 (Ac24), 1-hour probability of wet (Pwet1), and 24-hours probability of wet (Pwet24).

These indicators were fully discussed by Chandler et al. [15], [16] and Hanaish et al. [11]

V. RESULTS AND DISCUSSIONS

Tables I to III provided the disaggregation analysis of February, December, and June; they are representative for the dry, wet, and medium seasons, respectively. To compare the fitness of models, we employed two weight schemes to

obtain different rainfall statistical indicators. The first weight scheme is based on the diagonal element, expressed as  $w_i = 1/var(T_i)$ , where  $var(T_i)$  is the variance of the related monthly statistics across all years. The second weight scheme is given by  $w_i = [1/\bar{T}_i]^2$ , where  $\bar{T}_i$  is the mean of the related monthly statistics across all years [11]. In the tables,  $w_1$  represents the weight from the inversely proportional variance and  $w_2$  represents the weight from inversely proportional mean squared.

From Table 1, the value of variance in timescales of 6-hour and 24-hour are both over-estimations for February. BL2C and BLRD have a higher fitting result in estimating variances of 6-hour and 24-hour under  $w_1$  weight scheme compared with those under  $w_2$ . All models perform well in fitting 1-hour mean. From Table 2, all models show a lower accuracy under  $w_2$  scheme for December.

And, generally, the fitted values of 24-hour variance are much less accurate than those of other time resolutions. This may be due to the fact that a larger value of time interval leads to smaller weight in the object function. All versions of Bartlett-Lewis model have better performance than Neyman-Scott model. This could be the reason that Bartlett-Lewis models assume independent and identically distributed intervals between consecutive rain cells, which may be more suitable in describing local rainfall patterns. Due to the Southwest Monsoon influence, the rainfall variation is smoother from June to September than those in other months. From Table 3, it appears that the performance of fitness is better than other two months in terms of Var6 and Var24; this probably because the rainfall distribution is more homogeneous in June.

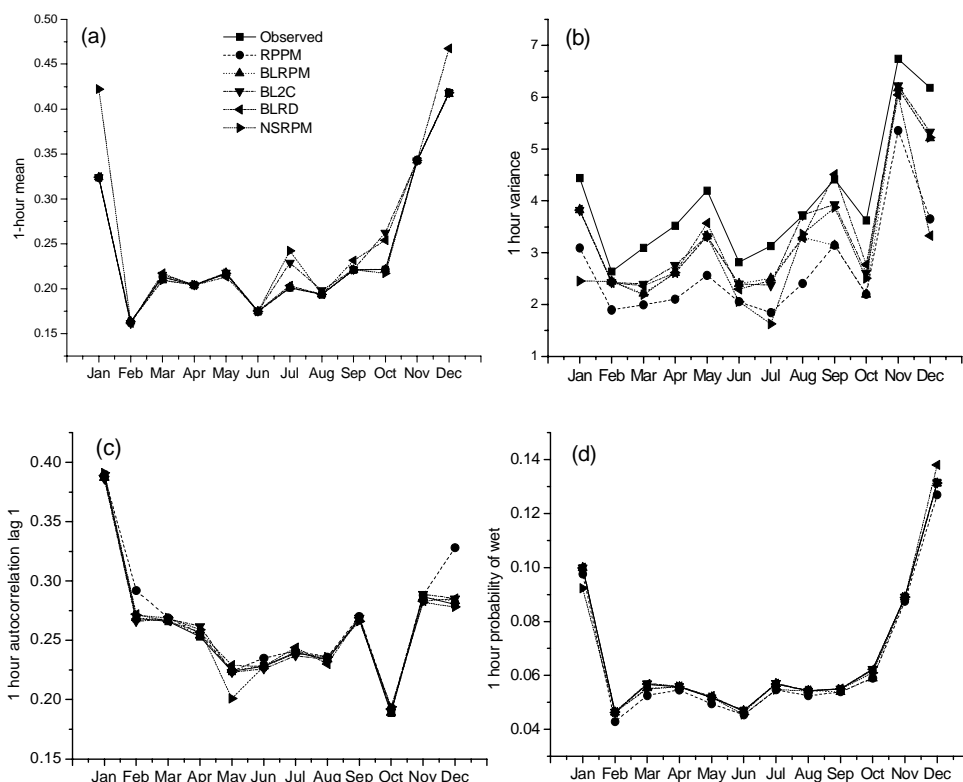


Fig. 2 Comparison between observed and fitted rainfall statistical properties under  $w_1$  weight scheme. (a) Mean1, (b) Var1, (c) Ac1, (d) Pwet1.

From Tables 1 to 3,  $w_1$  weight scheme is considered relatively superior than  $w_2$  one for most models in terms of 1-hr statistics. Figure 2 shows the performance of five models under  $w_1$  weight scheme for 1-hr statistics over all months. The values of Mean, Ac1 and Pwet1 all fit very well (mean relative error < 5%). The 1-hr variance has more significant variations, as its value is generally bigger than other indicators and a lower weight would be assigned to the objective function. Rainfall in Singapore is characterized by convective type cell, which demonstrating heavy rainfalls in short durations. Therefore, BLRPM which only allow for one type of rectangular pulse is more suitable for application in Singapore than BL2C, which represents 2 types of cell.

Another finding is that RPPM is relatively more accurate than others in estimating the 1-hour mean values, especially under  $w_1$  weight scheme.

VI. CONCLUSION

The temporal characteristics of hourly rainfall in Changi meteorological station of Singapore were investigated. The study results indicated that the characteristics of rainfall in Singapore were featured by large amounts in short duration, which belongs to convective rain. The performances of different disaggregation models were also evaluated. In terms of the short timescales properties, all Bartlett-Lewis models have good performances.

Generally, comparing the results from different models under two weight schemes, BLRPM has a higher accuracy than other models under both weight schemes; the scheme with weight of inversely proportional variance performs better than that from inversely proportional mean squared. In further studies, the description of extremes values from different models needs to be improved.

#### ACKNOWLEDGMENT

This research was supported by Earth Observatory of Singapore Project (M4080707.B50), and Singapore's Ministry of Education (MOM) AcRF Tier 1 Project (M4010973.030).

#### REFERENCES

- [1] C. G. Kilsby, A. Butron, S. J. Birkinshaw, A. M. Hashemi, P. E. O'Connell, "Extreme rainfall and flood frequency distribution modeling for present and future climates," Proceedings of the British Hydrological Society Seventh National Hydrology Symposium, 2000, pp. 3.51-3.56.
- [2] I. Rodriguez-Iturbe, V. K. Gupta, E. Waymire, "Scale consideration in the modeling of temporal rainfall," Water Resources Research 20 (11), 1984, 1611-1619.
- [3] C. Onof, R. E. Chandler, A. Kakou, P. Northrop, H. S. Wheater, V. Isham, "Rainfall modeling using Poisson-Cluster processes: a review of developments," Stochastic Environmental Research and Risk Assessment 14(6), 2000, 384-411.
- [4] I. Rodriguez-Iturbe, D. R. Cox, V. Isham, "Some models for rainfall based on stochastic point processes," Proceedings of the Royal Society of London, Series A 410, 1987a, 269-288.
- [5] I. Rodriguez-Iturbe, B. Febres de Power, J. B. Valdes, "Rectangular pulses point process models for rainfall: analysis of empirical data," Journal of Geophysical Research 92, 1987b, 9645-9656.
- [6] A. N. Engida, M. Esteves, "Characterization and disaggregation of daily rainfall in the Upper Blue Nile Basin in Ethiopia," Journal of Hydrology 399, 2011, 226-234.
- [7] C. Onof, "DEFRA Project: Improved methods for national spatial-temporal rainfall and evaporation modeling for BSM," 2003, unpublished.
- [8] P. S. P. Cowpertwait, "A generalized point process model for rainfall," Proceedings of Royal Society of London 447, 1994, 23-27.
- [9] P. S. P. Cowpertwait, "A generalized spatial-temporal model of rainfall based on a clustered point process," Proceedings of Royal Society of London 450, 1995, 163-175.
- [10] A. Burton, C. G. Kilsby, H. J. Fowler, P. S. P. Cowpertwait, P. E. O'Connell, "RainSim: A spatial-temporal stochastic rainfall modeling system," Environmental Modelling & Software 23, 2008, 1356-1369.
- [11] I. S. Hanaish, K. Ibrahim, A. A. Jemain, "Stochastic modeling of rainfall in Peninsular Malaysia using Bartlett-Lewis rectangular pulse models," Modelling and Simulation in Engineering, 2011.
- [12] A. Kakou, "Point process based models for rainfall," Ph.D. thesis, University of London, 1997, unpublished.
- [13] Meteorological Services Division, National Environmental Agency, <http://www.nea.gov.sg>
- [14] Geography of Singapore, WIKIPEDIA, [http://en.wikipedia.org/wiki/Geography\\_of\\_Singapore](http://en.wikipedia.org/wiki/Geography_of_Singapore)
- [15] R. Chandler, G. Lourmas, J. Jesus, "Software for moment-based on fitting of single-site stochastic rainfall models," 2010, unpublished.
- [16] J. Jesus and R. Chandler, "Estimating functions and the generalized method of moments," Interface Focus, 1(6), 2011, 871-885.