Menemui Matematik (Discovering Mathematics)
Vol. 41, No. 2: 80 - 87 (2019)

Rainfall Modelling Using Bartlett Lewis Rectangular Pulse Models

Nur Allysa Binti Mohd Redzuan Chan[1], Wendy Ling Shinyie[1, a)] and Wah June Leong[1]

[1]*Department of Mathematics, Faculty of Science, Universiti Putra Malaysia,*
43400 UPM Serdang, Selangor, Malaysia.
[a)]Corresponding author: sy_ling@upm.edu.my

ABSTRACT

Stochastic rainfall model can be very useful to solve problem related in hydrology, agricultural science, and engineering. The objective of this study is to study the hourly rainfall in Peninsular Malaysia, to estimate the parameters of the models using the generalized method of moments (GMM), and to identify the performance of Bartlett Lewis rectangular pulse model (BLRPM) for the rainfall data using mean absolute percentage error (MAPE). Hourly rainfall datasets of five raingauge stations which are located at different region of the Peninsular Malaysia (North West, South West, West, Center and East) are used. In this study, two types of Bartlett Lewis rectangular pulse model used are Original Bartlett Lewis rectangular pulse model (OBLRPM) and Modified Bartlett Lewis rectangular pulse model (MBLRPM). The mean hourly rainfall of all the stations has been observed and it is found that Pekan station has the greatest amount of mean hourly rainfall due to the effect of the monsoon winds. Also, Johor Bahru is not very influenced by the monsoon winds because the change of the mean hourly rainfall in Johor Bahru is small. Parameter estimation has been performed for each model by using generalized method of moment (GMM) and Nelder-Mead optimization algorithm. The performance of the models is observed by using the MAPE. Overall, performance in reproducing the observed rainfall data is better by using MBLRPM.

Keywords: Bartlett Lewis rectangular pulse model, generalized method of moments, hourly rainfall

INTRODUCTION

Malaysia is a country of Southeast Asia which located near the equator; therefore, the climate of Malaysia is categorised as tropical rainforest climate. Tropical rainforest climate is typically hot and humid throughout the year. There are only two seasons in tropical rainforest climates, which are wet season and dry season.

The rainfall in Malaysia is caused by the monsoon winds, which namely as Southwest monsoon and Northeast monsoon. According to Malaysia Meteorological Department's classification that the Southwest monsoon occurs between May to September, while Northeast monsoon occurs between November to March. Southwest monsoon came from the desserts of Australia will bring very little rain to the west part of Peninsular Malaysia because the monsoon is generally dry. The monsoon is striking between May and September. For Northeast monsoon, this monsoon came from China and north Pacific will bring heavy rain to the east part of

Peninsular Malaysia while the west part of Peninsular Malaysia experiences sunny and dry weather. This monsoon is striking between November and March.

The application of the Bartlett Lewis rectangular pulse model (BLRPM) for describing rainfall processes has been done for countries with 4 seasons but not much work has been done for countries with 2 seasons. BLRPM has done many modifications, such as Modified Bartlett Lewis rectangular pulse model (MBLRPM), Bartlett Lewis model with 2 cell types (BL2n), Bartlett Lewis rectangular pulse gamma model (BLRPGM), and Bartlett Lewis rectangular with cell depth distribution dependent on duration (BLRD). In this study, Original Bartlett Lewis rectangular pulse model (OBLRPM) and MBLRPM are used.

Stochastic models have been widely used to generate rainfall in hydrology world. Hanaish et al. (2013) stated that stochastic models have the ability to describe the rainfall process well and the other properties of the natural process based on the small number of parameters can be deduced. Currently, the most popular stochastic models is the pulse-based models with representative NSRPM (Rodriguez-Iturbe et al., 1987; Cowperwait & O' Connell, 1997) and BLRPM (Hanaish et al. 2013; Onof & Wheater, 1994; Verhoest et al., 1997).

According to Rodriguez-Iturbe et al. (1987), the OBLRPM with five parameters is unable to reproduce the proportion of dry periods correctly; therefore, they proposed a new version of the model which is MBLRPM with six parameters. Hanaish et al. (2013) performed a study to determine which model between OBLRPM and MBLRPM can accurately describe rainfall processes in Peninsular Malaysia and found that the MBLRPM can accurately describe the rainfall processes in Peninsular Malaysia. They used GMM and Nelder-Mead optimization method for parameter estimation. Smithers et al. (2002) had done a study on South Africa using two version of BLRPM which is MBLRPM and BLRPGM. The researchers used method of moment and least squares objective function for parameter estimation and relative error for comparing the model performance. They found that the performance of the MBLRPM was more sensitive to the moments used to estimate the model parameters but the performance of BLRPGM were better when using short duration rainfall data. The objectives of this study are to study the hourly rainfall in Peninsular Malaysia using OBLRPM and MBLRPM and to assess their performance using mean absolute relative error.

The contents of this paper are structured as follows. Next section describes the material and methods. The results and discussions are explained in the following section. Finally, the conclusions are provided.

MATERIAL AND METHODS

Data Description

In this study, the hourly rainfall dataset was used for a period from 1970 to 2008. The rainfall data were collected from 5 different regions of Peninsular Malaysia. The first rain gauge station is located at Alor Setar is the station on the North West region. The second station is located at Ampang is the station on the West region. The third station is located at Johor Bahru is the station on the South West region. The fourth station is located at Pekan is the station on the East region. The last rain gauge station is located at Chanis is the station on the Center region of Peninsular Malaysia. The exact location is shown in Figure 1.

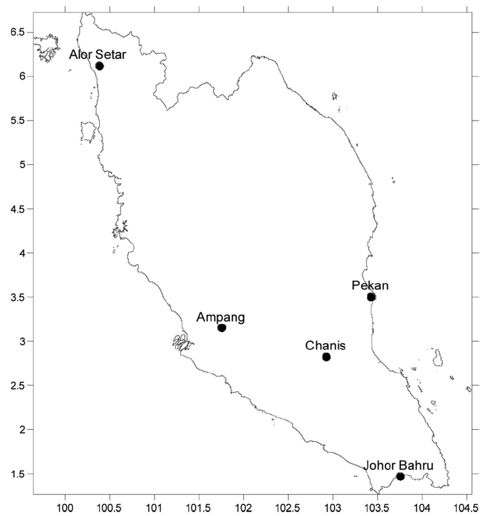

Figure 1: Location of stations

BLRPM Model

According to Kossieris et al. (2018), the measurement with respect to time does not allow fitting the pulse models directly. Therefore, the statistics of the datasets is obtained for each month of each year. The statistics used in this study are mean, variance, lag-1 autocorrelation, and probability of dry. The probability of dry were obtained from the amount of rainfall that less then 1mm.

Two stochastic models that used in this study are OBLRPM and MBLRPM. OBLRPM is a cluster-based rectangular pulses model with five parameters $(\lambda, \beta, \gamma, \eta, \mu_x)$. The algorithm is described as the storms arrivals occur randomly according to a Poisson distribution with parameter λ. Within each storm, there is a cells arrive randomly following a Poisson distribution with rate β. The storm obtains within the rectangular cells that stop after a given time is exponentially distributed with parameter γ. The duration of the rectangular pulse of each cell is exponentially distributed with parameter η. Each cell depth is a random constant exponentially distributed with mean, $1/\mu_x$ and the number of cells per storm follows a geometric distribution with mean $\mu_c = 1 - \kappa/\phi$. The equations used for the modelled statistics to obtain the simulated data are:

$$E\left(Y_i^h\right) = \frac{\lambda}{\eta} h \mu_c \mu_x$$

$$\text{var}\left(Y_i^h\right) = \frac{2\frac{\lambda}{\eta}\mu_c\left(E(X^2)+\frac{\beta\mu_x^2}{\gamma}\right)h}{\eta} + \frac{2\frac{\lambda}{\eta}\mu_c\mu_x^2\beta\eta\left(1-e^{-\gamma h}\right)}{\gamma^2(\gamma^2-\eta^2)} - 2\frac{\lambda}{\eta}\mu_c\left(E(X^2)+\frac{\beta\mu_x^2\gamma}{(\gamma^2-\eta^2)}\right)(1-e^{-\eta h})/\eta^2$$

For a lag $k \geq 1$, the covariance is

$$\text{cov}\left(Y_i^h, Y_{i+k}^h\right) = \frac{\lambda}{\eta}\frac{\mu_c\left(E(X^2)+\beta\gamma\mu_x^2(\gamma^2-\eta^2)^{-1}\right)e^{\eta(k-1)h}\left(1-e^{-nh}\right)^2}{\eta^2} - \frac{\lambda}{\eta}\mu_c\,\mu_x^2\,\beta\eta(1-e^{-\gamma h})^2 e^{\gamma(k-1)h}(\gamma^2-\eta^2)^{-1}/\gamma^2$$

The probability of dry for a period of length h is

$P(h)' = \exp\left(-\lambda(h+\mu_T) + \frac{\lambda G_p^*(0,0)\left(\gamma+\beta e^{-(\beta+\gamma)h}\right)}{\beta+\gamma}\right)$
where $G_p^*(0,0) = \eta^{-1}e^{-k}\int_0^1 t^{\phi-1}(1-t)e^{kt}\,dt$

MBLRPM is a six-parameter model $(\lambda, \mu_x, \alpha, \nu, \kappa, \phi)$. The algorithm is described as the arrival rate of a storm following a Poisson distribution with the parameter λ. Within each storm, there is cells arrive randomly following a Poisson distribution with rate β. The duration of the storm is distributed according to an exponential distribution with parameter γ. The duration of the rectangular pulse of each cell with parameter η is following a gamma distribution with shape parameter α and scale parameter ν. Each cell depth of the rectangular pulse of each cell is exponentially distributed with mean μ_x and the number of cells per storm follows a geometric distribution with mean $\mu_c = 1 + \kappa/\phi$, where κ and ϕ are a dimensionless parameter. The parameters β and γ change randomly when keeping κ and ϕ constant. The equations used for the modelled statistics to obtain the simulated data are:

$E\left(Y_i^h\right) = \frac{\lambda h \nu \mu_x}{\alpha-1}\left(1+\frac{\kappa}{\phi}\right)$
$\mathrm{var}\left(Y_i^h\right) = 2A1\left((\alpha-3)h\nu^{2-\alpha} - \nu^{3-\alpha} + (\nu+h)^{\alpha}\right) - 2A2(\phi(\alpha-3)h\nu^{2-\alpha} - \nu^{3-\alpha} + (\nu+\phi h)^{3-\alpha})$
For a lag $k \geq 1$ the covariance is

$$\mathrm{cov}\left(Y_i^h, Y_{i+k}^h\right) = A1((\nu+(k+1)h)^{3-\alpha} - 2(\nu+kh)^{3-\alpha} + (\nu+(k-1)h)^{3-\alpha}) \\ -A2((\nu+(k-1)\phi h)^{3-\alpha} - 2(\nu+k\phi h)^{3-\alpha} + (\nu+(k-1)\phi h)^{3-\alpha})$$

where
$A1 = \frac{\lambda\mu_c\nu^{\alpha}}{(\alpha-1)(\alpha-2)(\alpha-3)}\left(2\mu_x^2 + \frac{\kappa\phi\mu_x^2}{\phi^2-1}\right)$ and
$A2 = \frac{\lambda\mu_c\kappa\mu_x^2\nu^{\alpha}}{\phi^2(\phi^2-1)(\alpha-1)(\alpha-2)(\alpha-3)}$

The probability of dry for a period of length h is
$P(h)' = \exp\left(-\lambda h - \lambda\mu_T + \lambda G_p^*(0,0)\frac{\phi+\kappa\left(\frac{\nu}{\nu+(\kappa+\phi)h}\right)^{\alpha-1}}{\phi+\kappa}\right)$
where
$\mu_T \approx \frac{\nu}{\phi(\alpha-1)}\left(1+\phi(\kappa+\phi) - \frac{1}{4}\phi(\kappa+\phi)(\kappa+4\phi) + \frac{1}{72}\phi(\kappa+\phi)(4\kappa^2+27\kappa\phi+72\phi^2)\right)$
$\mu_T =$ the expected duration of a single cell storm.

Parameter Estimation

In this section, five parameters need to be estimated for OBLRPM while MBLRPM need to estimate six parameters. The parameter estimation for the OBLRPM and MBLRPM was conducted using "MOMFIT" package, developed in R. In this package, the GMM is used to estimate the parameters. Lu and Qin (2012) stated that the GMM is a fundamental method to choose the best parameter that can minimize an objective function:

$$S(\theta) = \left(T-\tau(\theta)\right)' W\left(T-\tau(\theta)\right)$$

where $\theta = (\theta_1, \theta_2, \ldots)$ is the parameter vector , $T = (T_1, T_2, \ldots)$ is the vector of the observed rainfall summary statistics (mean, variance, autocorrelation, etc), $\tau(\theta) = (\tau_1(\theta), \tau_2(\theta), \ldots)$ be the expected value of T according to the model, θ, and W is the weight matrix. The weight matrix is obtained as where stand as $W_i = 1/\text{var}(T_i)$ where $\text{var}(T_i)$ stand as the ith diagonal elements of the covariance matrix of the summary statistics. The Nelder-Mead optimization algorithm is used to minimize the objective function, $S(\theta)$ which is already included in the "MOMFIT" package.

Smithers et al. (2002) stated that the choose set of moment should have relatively small sampling errors and not be highly mutually correlated. Therefore, as suggested by Rodriguez-Iturbe et al. (1987), the statistic that will be considered are 1-hour mean, 1-hour variance, 6-hour variance, 24-hour variance, 1-hour autocorrelation of lag-1, 24-hour autocorrelation of lag-1, 1-hour probability of wet, and 24-hour probability of wet.

Model Performance

In this section, the performance of OBLRPM and MBLRPM will be compared using MAPE. As suggested by Velghe et al. (1994):

$$\text{MAPE} = \frac{1}{m}\sum_{i=1}^{m}\left|1 - \frac{X_i}{X_{his,i}}\right| \times 100$$

where X_i is the value of the ith estimated statistic, $X_{his,i}$ is the value of the ith observed statistic and m is the number of statistic evaluated. MAPE is the most common method used to measure forecast accuracy. Large MAPE indicates that the errors are much greater then the actual values. Conversely, when the result obtained is a small number, this means the performance of the model is more appropriate.

RESULTS AND DISCUSSIONS

Figure 2 shows the mean hourly rainfall depth for all month in Johor Bahru, Ampang, Alor Setar, Chanis, and Pekan. We can see that all the station has different rainfall pattern because of the influence of Northeast monsoon and Southwest monsoon. Rain gauge station at Pekan has the highest amount of mean hourly rainfall compared to other rain gauge stations because of the location of the station. Pekan which located at the east part of the Peninsular Malaysia which received a lot of rainfall on October to February due to the Northeast monsoon. At the month of May and September, the rainfall received in Ampang, the West region of Peninsular Malaysia is less due to the Southwest monsoon. Johor Bahru which located at the South West region has the least amount of rainfall compared with other rain gauge stations. Johor Bahru is not very influenced by the monsoon because the change of the mean hourly rainfall in every month is too small.

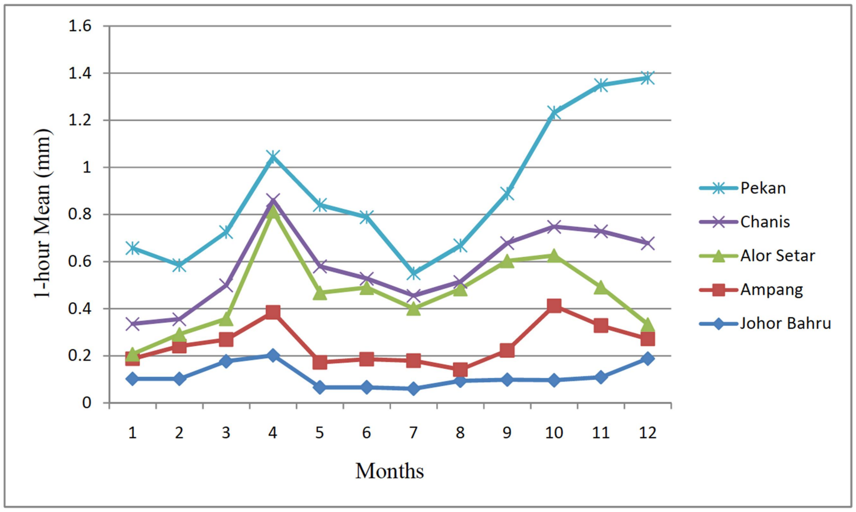

Figure 2: Mean hourly rainfall depth for all months in five raingauge stations

Estimated parameters for OBLRPM and MBLRPM for July and November in 5 stations are shown in Table 1 and 2. Based on result for OBLRPM, the rate of storm arrival for both months are quite stable except for the month of July in Chanis has relatively smaller storm arrival rate. Alor Setar has the highest rate of storm arrival in July while other stations has high storm arrival rate in November. Other than that, the mean cell depth, μ_x at Chanis and Ampang is highest in November while for Johor Bahru, Pekan and Alor Setar, the highest mean cell depth is in July. By comparing the result of MBLRPM, the rate of storm arrival for both months are quite stable except for the month of November in Alor Setar has relatively higher storm arrival rate. Ampang and Pekan have high rate of storm arrival in July while Johor Bahru, Chanis and Alor Setar has high storm arrival rate in November. Nevertheless, the mean cell depth at Alor Setar is the highest in November while Johor Bahru, Chanis, Ampang and Pekan has the highest mean cell depth in July. A certain location with high value of mean cell depth and small value of storm duration experiences a convectional rainfall, which is a heavy rainfall in a short duration. Johor Bahru has the highest value of mean cell depth with a small value of storm duration at the month of July. Alor Setar in November experiences many rainfall because the value of storm duration is the highest.

Table 1: Estimated parameters of OBLRPM in 5 stations in July and November

	July				
	$\lambda\ (h^{-1})$	$\beta\ (h^{-1})$	$\gamma\ (h^{-1})$	$\mu_x (mmh^{-1})$	$\eta\ (h^{-1})$
Johor Bahru	2.00e-03	7.48e+12	7.73e+19	2.32e+16	6.29e+14
Chanis	1.56e-06	2.34e+01	7.54e-05	1.41e-02	3.79e+07
Ampang	5.99e-04	1.66e+09	2.68e+06	4.04e+06	6.22e+07
Alor Setar	4.98e-03	3.56e+17	1.19e+11	5.46e+23	2.28e+28
Pekan	1.41e-04	2.86e+07	3.07e-03	3.33e+04	1.03e+14
	November				
	$\lambda\ (h^{-1})$	$\beta\ (h^{-1})$	$\gamma\ (h^{-1})$	$\mu_x\ (mmh^{-1})$	$\eta\ (h^{-1})$
Johor Bahru	5.00e-03	3.94e+09	1.64e+07	4.24e+07	1.35e+08
Chanis	1.60e-03	1.00e-06	1.07e+11	1.07e+15	3.24e+13
Ampang	1.03e-03	1.06e+05	7.05e+07	2.88e+13	7.62e+11
Alor Setar	2.91e-03	2.18e-06	4.88e+11	1.52e+14	2.32e+12
Pekan	6.80e-04	3.23e+14	1.32e+01	5.91e+01	7.36e+12

Table 2: Estimated parameters of MBLRPM in 5 stations in July and November

	July					
	$\lambda\,(h^{-1})$	κ	ϕ	$\mu_x(mmh^{-1})$	α	ν
Johor Bahru	1.00e-03	6.99e+01	1.28e-06	1.34e+13	5.09e+17	1.97e+01
Chanis	7.70e-05	3.67e+01	6.21e-02	6.90e+09	2.96e+05	5.01e+04
Ampang	5.91e-04	3.76e-03	1.48e-04	3.76e-03	5.34e+17	2.70e-14
Alor Setar	1.57e-03	3.49e-04	4.17e-05	2.42	1.08e+17	1.73e-18
Pekan	1.15e-03	1.00e-06	1.00e-06	2.28e+04	6.96e+11	1.18
	November					
	$\lambda\,(h^{-1})$	κ	ϕ	$\mu_x(mmh^{-1})$	α	ν
Johor Bahru	1.00e-03	4.68e+01	1.20e-02	4.43e+12	2.48e+05	9.21e+08
Chanis	1.57e-03	4.88e+01	2.97e-06	6.12e+09	4.22e+05	7.71e+09
Ampang	7.50e-05	6.58e-05	5.45e-06	2.90e-02	1.16e+10	3.19e-10
Alor Setar	1.66	4.98e+09	2.06e+09	8.61e+07	6.11e+04	3.75e+03
Pekan	6.77e-04	2.50e+01	2.91e-05	1.04e+02	9.78e+06	4.60e-02

By comparing each of the statistics is difficult to identify which model is more suitable to the rainfall pattern in Malaysia. Therefore, Figure 3 shows the overall MAPE for July and November for all the stations based on the OBLRPM and MBLRPM. Only for the month of November in Ampang and July in Pekan show that OBLRPM has better performance while the rest of the raingauge stations indicate that MBLRPM has better performance. Therefore, we can conclude that the overall performance in reproducing the observed rainfall data is better using MBLRPM.

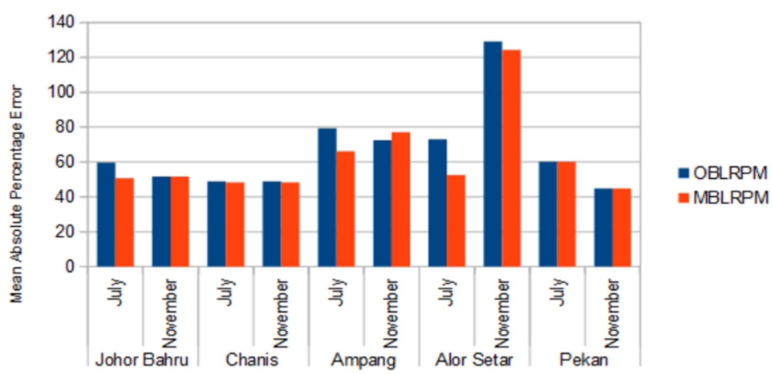

Figure 3: Overall MAPE for July and November for all stations

CONCLUSION

In conclusion, this study is to identify which Bartlett Lewis model that can describe Peninsular Malaysia rainfall accurately. Two Bartlett Lewis models (OBLRPM and MBLRPM) were used to achieve the purpose of this study. To estimate the parameter of the models, GMM was used to estimate parameter and Nelder-Mead optimization algorithm was implemented to minimize the objective function. We have observed the mean hourly rainfall for all the stations and found out the Pekan has the greatest amount of mean hourly rainfall due to the effect of the Northeast monsoon and Southwest monsoon. Also, the change of mean hourly rainfall in Johor Bahru is small; therefore, Johor Bahru is not vey influenced by the monsoon winds. The performance of the models is examined by using the MAPE. The MAPE is smaller for the OBLRPM in Johor Bahru while the other stations have the smallest MAPE for the MBLRPM. In overall, performance in reproducing the observed rainfall data is better by using MBLRPM.

ACKNOWLEDGMENTS

We acknowledge the financial support of the Universiti Putra Malaysia (9587800). We thank the Department of Irrigation and Drainage for providing the hourly rainfall data.

REFERENCES

Cowpertwait, P. S. P., & O'Connell, P. E. (1997). A regionalised Neyman-Scott model of rainfall with convective and stratiform cells. *Hydrology and Earth Sciences,* 1:71-80.

Hanaish, I. S., Ibrahim, K., & Jemain, A. A. (2013). On the potential of Bartlett Lewis rectangular pulse models for simulating rainfall in Peninsular Malaysia. *Hydrological sciences journal*, 58(8), 1690-1703.

Kossieris, P., Makropoulos, C., Onof, C., & Koutsoyiannis, D. (2018). A rainfall disaggregation scheme for sub-hourly time scales: Coupling a Bartlett-Lewis based model with adjusting procedures. *Journal of Hydrology*, 556, 980-992.

Lu, Y., & Qin, X. S. (2012). Comparison of Stochastic Point Process Models of Rainfall in Singapore. In Proceedings of World Academy of Science, Engineering and Technology (No. 68). World Academy of Science, Engineering and Technology.

Onof, C., & Wheater, H. S. (1994). Improved fitting of the Bartlett-Lewis Rectangular Pulse Model for hourly rainfall. *Hydrological sciences journal*, 39(6), 663-680.

Rodriguez□Iturbe, I., De Power, B. F., & Valdes, J. B. (1987). Rectangular pulses point process models for rainfall: analysis of empirical data. *Journal of Geophysical Research: Atmospheres*, 92(D8), 9645-9656.

Smithers, J. C., Pegram, G. G. S., & Schulze, R. E. (2002). Design rainfall estimation in South Africa using Bartlett–Lewis rectangular pulse rainfall models. *Journal of Hydrology*, 258(1-4), 83-99.

Velghe, T., Troch, P. A., De Troch, F. P. & Van de Velde, J. (1994) Evaluation of cluster-based rectangular pulses point process models for rainfall. *Water Resources Research*, 30(10), 2847-2857.

Verhoest, N., Troch, P. A., & De Troch, F. P. (1997). On the applicability of Bartlett–Lewis rectangular pulses models in the modeling of design storms at a point. *Journal of hydrology*, 202(1-4), 108-120.