京都大学防災研究所年報　第 50 号 B　平成 19 年 4 月
Annuals of Disas. Prev. Res. Inst., Kyoto Univ., No. 50 B, 2007

NSM Modifications for Improved Generation of Rainfall Maxima

Carlo Arturo MONDONEDO*, Yasuto TACHIKAWA* and Kaoru TAKARA

*Graduate School of Urban Environmental Engineering, Kyoto University

Synopsis

A previous application of the Neyman-Scott Rectangular Clustered Poisson Rainfall Model, NSM here for brevity, focused on the generation of synthetic rainfall records that yielded block maxima (monthly hourly and daily maximum rainfall values) that matched the historical counterpart. Among the findings, it was determined that historical moments used to identify the NSM model parameters such as the hourly mean, hourly variance, hourly covariance at lag 1, 24-hourly variance and 24-hourly covariance at lag-1 were sufficient for modeling the Kamishiiba Region in Kyushu. To affect parsimony in the synthetic rainfall generation in this study, the historical data from the said region was pooled into the frontal rainfall season (June to July) and typhoon season (September to October). Results were further improved when the search information included the hourly third central moment of the NSM. As of this writing, a limited improvement was observed in the Peaks Over Threshold Maxima as well.

Keywords: stochastic hydrology, rainfall model, design flood, time series generation, simulation

1. Introduction

Synthetic rainfall records generated via the Neyman-Scott Rectangular Clustered Poisson Rainfall Model, NSM here for brevity, must have block maxima (hourly maxima and 24-hourly maxima for instance) that match the historical counterparts to be sufficient. In a previous study (Mondonedo *et al*, 2007), it was shown that a proper combination of moments and covariances from the historical records is critical in meeting this objective. From the same study, it was found that (among other combinations) the hourly mean, hourly variance, hourly covariance at lag 1, 24-hourly variance and 24-hourly covariance at lag 1 were sufficient for modeling the monthly rainfall block maxima of hourly and 24-hourly duration in the Kamishiiba Region in Kyushu. Based on this region, this study focuses on the improvement of the parameter search and the synthetic rainfall generation and on the extension of maxima modeling to the Peaks Over Threshold (POT) definition as well.

2. NSM Equations

The schematic of the NSM appears in Figure 1. This model consists of essentially five probability distributions. A storm is characterized as a cluster of cells that follows a Poisson process with mean occurrence rate λ. Each storm consists of a random number of cells with all storms containing at least one cell. Candidate distributions for this purpose are the Poisson and geometric distributions. In this study, the geometric distribution was used extensively with mean μ_c. The arrival of each cell is an exponential variate with mean β. The duration and intensity of each cell are assumed independent identically distributed (iid) based on the exponential distribution with means η and $1/\mu_x$, respectively. The total rainfall intensity is then the superposition of the effects of these random cell intensities. However, this study also involved the case when the intensity of each cell is rendered as a gamma variable with shape parameter α and scale parameter θ.

From the method of moments, the NSM parameters can be linked to actual record data moments. Such expressions include the following (Rodriguez-Iturbe, 1987):

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

$$\mathrm{var}\left(Y_i^{(h)}\right) = \frac{2\lambda \mu_c E\langle X^2 \rangle (\mu_c - 1)\left[\beta^3 A_1(h) - \eta^3 B_1(h)\right]}{\beta \eta^3 \left(\beta^2 - \eta^2\right)} + \frac{4\lambda \mu_c {\mu_x}^2 A_1(h)}{\eta^3}$$

$$\mathrm{cov}\left(Y_i^{(h)}, Y_{i+k}^{(h)}\right) = \frac{2\lambda \mu_c E\langle X^2 \rangle (\mu_c - 1)\left[\beta^3 A_2(h,k) - \eta^3 B_2(h,k)\right]}{\beta \eta^3 \left(\beta^2 - \eta^2\right)} + \frac{4\lambda \mu_c {\mu_x}^2 A_2(h,k)}{\eta^3}$$

in which:

$$A_1(h) = \eta h - 1 + e^{-\eta h}$$

$$B_1(h) = \beta h - 1 + e^{-\beta h}$$

$$A_2(h,k) = 0.5\left(1 - e^{-\eta h}\right)^2 e^{-\eta h(k-1)}$$

$$B_2(h,k) = 0.5\left(1 - e^{-\beta h}\right)^2 e^{-\beta h(k-1)}$$

where:

i = time interval counter

h = integer specifying time step interval of data (1-hour, 24-hour, etc.)

Y_i^h = rainfall depth in the ith time interval of length h

$E\left(Y_i^h\right)$ = mean rainfall depth

$\mathrm{var}\left(Y_i^h\right)$ = variance of rainfall record

$\mathrm{cov}\left(Y_i^h, Y_{i+k}^h\right)$ = covariance or rainfall record at lag k.

The term $E\langle X^2 \rangle$ refers to the expectation of the square of the cell intensity random variable which can be based on either an exponential or a gamma distribution (see Section 3 for details).

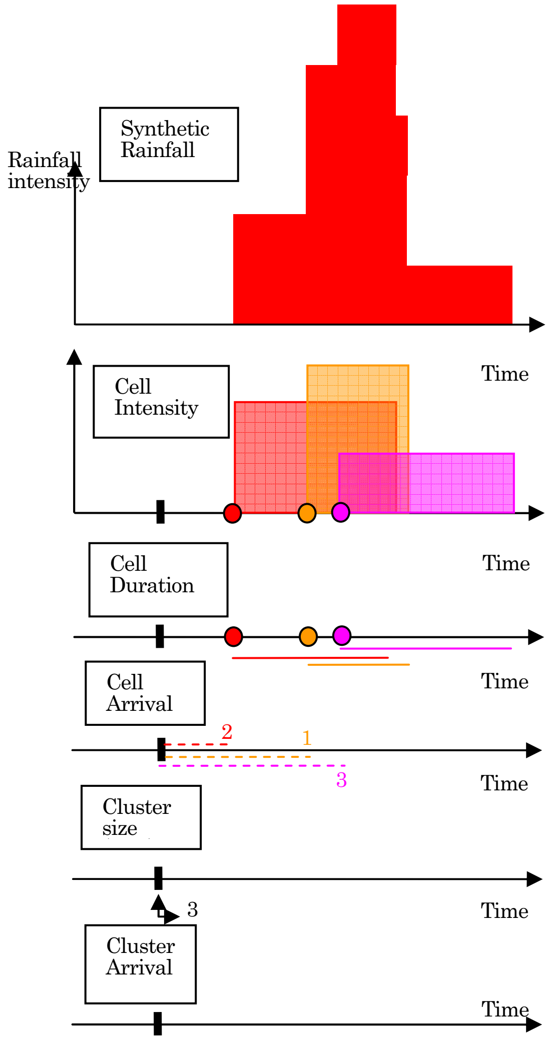

Fig. 1 Schematic diagram of the Neyman Scott Clustered Poisson Rectangular Rainfall Model (NSM).

As of this writing, the formulation of the third central moment included a cell intensity based solely on the gamma distribution. A formulation using the original exponential distribution and other distributions for cell intensity will appear in a future edition of this study. The third central moment of rainfall intensity appears here as (Cowpertwait, 1998):

$$E\left\langle \left(Y_i^h - \overline{Y_i^h}\right)^3 \right\rangle = 6\lambda \mu_c \left(\frac{\theta^3 \Gamma(3+\alpha)}{\Gamma(\alpha)}\right) * \left(\eta h - 2 + \eta h e^{-\eta h} + 2e^{-\eta h}\right) / \eta^4$$
$$+ 6\lambda \alpha \theta \left(\alpha(1+\alpha)\theta^2\right) \mu_c (\mu_c - 1) f(\eta, \beta, h) / \left[2\eta^4 \beta \left(\beta^2 - \eta^2\right)^2\right]$$
$$+ 6\lambda (\alpha\theta)^3 \mu_c (\mu_c - 1)^2 g(\eta, \beta, h) / \left[2\eta^4 \beta \left(\eta^2 - \beta^2\right)(\eta - \beta)(2\beta + \eta)(2\eta + \beta)\right]$$

$$
\begin{aligned}
f(\eta,\beta,h) &= -2\eta^3\beta^2 e^{-\eta h} - 2\eta^3\beta^2 e^{-\beta h} \\
&+ \eta^2\beta^3 e^{-2\eta h} + 2\eta^4\beta e^{-\eta h} + 2\eta^3\beta^2 e^{-(\eta+\beta)h} \\
&- 2\eta^4\beta e^{-(\eta+\beta)h} - 8\eta^3\beta^3 h + 11\eta^2\beta^3 \\
&- 2\eta^4\beta + 2\eta^3\beta^2 + 4\eta\beta^5 h + 4\beta\eta^5 h \\
&- 7\beta^5 - 4\eta^5 + 8\beta^5 e^{-\eta h} - \beta^5 e^{-2\eta h} \\
&- 2h\eta^3\beta^3 e^{-\eta h} - 12\eta^2\beta^3 e^{-\eta h} + 2h\eta\beta^5 e^{-\eta h} \\
&+ 4\eta^5 e^{-\beta h}
\end{aligned}
$$

$$
\begin{aligned}
g(\eta,\beta,h) &= 12\eta^5\beta e^{-\beta h} + 9\eta^4\beta^2 \\
&+ 12\eta\beta^5 e^{-\eta h} + 9\eta^2\beta^4 + 12\eta^3\beta^3 e^{-(\eta+\beta)h} \\
&- \eta^2\beta^4 e^{-2\eta h} - 12\eta^3\beta^3 e^{-\beta h} - 9\eta^5\beta \\
&- 9\beta^5\eta - 3\eta\beta^5 e^{-2\eta h} - \eta^4\beta^2 e^{-2\beta h} \\
&- 12\eta^3\beta^3 e^{-\eta h} + 6\beta^2\eta^5 h - 10\beta^4\eta^3 h \\
&+ 6\beta^5\eta^2 h - 10\beta^3\eta^4 h + 4\beta^6\eta h \\
&- 8\beta^2\eta^4 e^{-\beta h} + 4h\eta^6\beta + 12\eta^3\beta^3 \\
&- 8\eta^2\beta^4 e^{-\eta h} - 6\eta^6 - 6\beta^6 - 2\eta^6 e^{-2\beta h} \\
&- 2\beta^6 e^{-2\eta h} + 8\beta^6 e^{-\eta h} + 8\eta^6 e^{-\beta h} \\
&- 3\beta\eta^5 e^{-2\beta h}
\end{aligned}
$$

where:

$E\left\langle \left(Y_i^h - \overline{Y_i^h}\right)^3 \right\rangle$ = third central moment of rainfall depth.

$\overline{Y_i^h}$ = mean value of rainfall at aggregation level h.

3. NSM Parameter Search

The objective function used in the estimation follows the form:

$$
\mathrm{F} = \sum_{j=1}^{5}\left(\frac{f_{\mathrm{j}}(Y_i)}{W_j} - 1\right)^2
$$

where:

$f_{\mathrm{j}}(Y_i)$ = jth NSM moment or equation covariance equation of rainfall depth Y_i.

W_j = actual moment value from rainfall record.

This choice of the objective function was made to ensure that large numerical values do not dominate the fitting procedure (Favre *et al*., 2004).

Table 1. NSM Parameter search limits.

NSM Parameter	Upper limit	Lower limit
λ, 1/h	0.001	0.05
β, 1/h	0.05	0.99
η, 1/η	0.5	60
μ_x, mm/h	0.3	15
μ_c	1	50
α	0.01	20
θ, mm/h	N/A	N/A

Two key parameter searches based on a basic reconfiguration of the model were used in this study. The first search, or S1, did not involve the third central moment with rain cell intensity X rendered as an exponential variate, requiring parameters λ, β, μ_c, μ_x, and η. In this case, the term $E\langle X^2\rangle$ in the variance and covariance equations corresponded to $2\mu_x^2$. This search required the following historical information: hourly mean rainfall, hourly rainfall variance, hourly rainfall covariance at lag 1, 24-hourly rainfall variance, and 24-hourly rainfall covariance at lag 1.

The second search, or S2, involved the third central moment with rain cell intensity X rendered as a gamma distribution variable, requiring parameters λ, β, μ_c, α, θ, and η. However, since $\mu_x = \alpha\theta$, the parameter θ was rendered in terms of the historical hourly rainfall mean, leaving only 5 parameters in the search. In addition, the term $E\langle X^2\rangle$ in the variance and covariance equations corresponded to $\alpha(1+\alpha)\theta^2$. This search required the following historical information: hourly rainfall variance, hourly correlation at lag 1, hourly 3rd central moment, 24-hourly variance and 24-hourly correlation at lag 1. NSM correlation equations are the ratio of the covariance equation and the variance equation at appropriate duration h and lag k.

A 15 year record of hourly rainfall from 1988 to 2002 from a rain gauge based in Kamishiiba served as the historical data in the study. Rainfall from months June to July was pooled to form a 15-year long record of the frontal season rainfall. The same was done with the rainfall from September to October, covering the typhoon season. It was assumed that stationarity was still observed within the pooled data. Each block of the 15 year historical record then contained 61 days worth of hourly rainfall. Four searches where then conducted in total: S1 and S2 for June-July and S1 and S2 for September-October. For brevity, the former pair was called SF1 and SF2 while the latter pair was called ST1 and ST2.

Table 1 show the limits used in the constrained search. A simplex algorithm was used for an initial estimate and the Levenberg-Marquardt search was used for refinement (Press, 1986). Table 2 shows the results of the estimation obtained from all searches.

Table 2. Parameters determined from NSM searches.

NSM Parameter	SF1	SF2	ST1	ST2
λ, 1/h	0.0106	0.0106	0.0038	0.0040
β, 1/h	0.0946	0.0946	0.0556	0.0543
η, 1/η	2.179	2.179	1.037	0.994
μ_x, mm/h	3.784		4.112	
μ_c	36.214	40.774	20.124	50.000
α		0.794		0.220
θ, mm/h		4.217		6.965

4. Synthetic Rainfall Generation

The synthetic record generated from each search consisted of 100 realizations of 61-day rainfall. A summary of the moments and correlations involved in the searches and the historical counterparts appear in Table 3.

Table 3. Historical and Synthetic Rainfall Moments

Kamishiiba rainfall	Jun-Jul	SF1	SF2	Sep-Oct	ST1	ST2
Mean (1H), mm	0.665	0.643	0.647	0.305	0.306	0.298
Variance (1H), mm^2	6.384	6.280	6.383	3.042	3.113	3.120
Lag-1 Co-variance (1H), mm^2	4.315	4.266	4.346	2.165	2.233	2.237
Lag-1 Cor-relation (1H)	0.676	0.679	0.681	0.712	0.717	0.717
3rd Central Moment (1H), mm^3	6.614	N/A	6.711	10.653	N/A	10.604
Variance (24H), mm^2	1288.50	1249.05	1314.54	606.99	643.84	636.03
Lag-1 Co-variance (24H), mm^2	347.46	349.98	354.48	228.65	249.25	246.51
Lag-1 Cor-relation (24H)	0.270	0.280	0.270	0.377	0.387	0.388

5. Preliminary Analysis

Rainfall maxima obtained from synthetic and historical rainfall records were checked here following two tests. The standard Kolmogorov-Smirnov or KS Test in which the unbinned distributions of the two data sets are compared for maximum deviation served as the main test at this phase of the study. This test was applicable to the study data in that for each record, each maximum can be assigned to a single independent variable (Press, 1986). Based on the maximum deviation of historical and synthetic cumulative distributions, the KS probability approaches unity when the null hypothesis that two sets did not derive from the same population can be rejected (i.e.: statistical difference is improbable). An alternative test based on the fit of the same maxima to extreme value function distributions (Gumbel, Pareto, Pearson, etc.) is underway as of this writing.

Both block maxima and POT maxima were checked. The historical block maxima for a period were used as a basis for determining the POT maxima. For instance, Table 4 shows the historical block maxima for the 15 year Kamishiiba record. The threshold used to determine the POT maxima were the 8th to the 11th largest values from this record.

Table 4. Block Maxima from Kamishiiba rainfall records.

Kamishiba Maximum Rainfall (1988-2002) in mm				
	Jun-Jul		Sep-Oct	
Rank	Hourly	24-Hourly	Hourly	24-Hourly
1	15	100	9	21
2	20	101	12	28
3	21	120	13	32
4	22	127	15	55
5	23	128	18	55
6	28	131	19	62
7	30	144	22	62
8	31	145	24	81
9	31	150	27	89
10	34	189	30	162
11	36	221	31	163
12	38	225	34	182
13	41	266	36	212
14	48	271	38	232
15	55	380	39	318

Table 5 shows the KS Test results for the 4 searches conducted in the study for the block maxima and POT maxima. It appears that both S1 and S2 can adequately model the block maxima for the June-July period. S2 appears to have a slight advantage over S1 for the September-October period in which skewness is more prominent.

S1 and S2 appear to perform consistently for the frontal rainfall season. This suggests that the basic structure of storm and rain cells can be further improved for the typhoon season. A possible correction here is to modify the NSM storm into a collection of cells that suits the typhoon season better than the current configuration (Cowpertwait, 1996).

S2 can be modified into a similar search S2` that includes additional historical information such as lag 2 correlations or the third central moment of daily rainfall depth for the typhoon season. However, such an approach may strain the objective functions such that S2` may perform worse than the original S2.

While S2 suggests an improved search for NSM parameters, it is not clear how S2 affects the synthetic rainfall generation. Parameters determined from S2 may have improved from 1) the choice of distribution for cell intensity, 2) the inclusion of the third moment of rainfall depth, or 3) both.

Table 5. Summary of KS Test Results using generated synthetic and historical Kamishiiba rainfall

Search	SF1	SF2	ST1	ST2
Block Maxima				
1H	0.992	1.000	1.000	0.994
24H	1.000	1.000	0.918	0.966
Hourly POT Maxima				
8/15	0.957	0.997	0.992	0.464
9/15	0.957	0.997	0.992	0.991
10/15	0.778	0.944	0.997	0.863
11/15	0.987	0.986	0.985	0.776
24-Hourly POT Maxima				
8/15	0.999	0.975	0.961	0.963
9/15	0.933	1.000	0.620	0.699
10/15	0.336	0.984	0.975	1.000
11/15	1.000	0.965	0.726	0.998

As expected from previous NSM performance studies, moments, covariances, and block maxima of the historical records were closely matched for all searches used here. From this information and KS Test results, it appears that pooling together the frontal rainfall record of June and July and the typhoon rainfall record of September and October is justified.

In addition, the POT maxima do not appear to be well-matched. Neither S1 nor S2 yielded parameters that led to synthetic rainfall that closely match the POT maxima of the historical record consistently. However, this limitation may be rectified by including more moments or correlations in further investigations.

In a later edition of this study, the historical and synthetic maxima will be checked through the fit of an extreme value distribution such as the Gumbel distribution. Confirmation by this test is based on the proximity of the fitted Gumbel parameters with SLSE

Table 6. Fitted Gumbel parameters from Block Maxima

Gumbel Distribution: $P[X<x]=\mathrm{Exp}\{\mathrm{Exp}[-(x-\rho)/\varepsilon]\}$						
Search	Jun-Jul	SF1	SF2	Sep-Oct	ST1	ST2
ρ						
1H	0.488	0.547	0.558	0.392	0.316	0.35
24H	0.332	0.384	0.336	-0.172	0.047	-0.052
ε						
1H	0.329	0.228	0.273	0.448	0.415	0.356
24H	0.356	0.351	0.384	0.716	0.496	0.465

(Takara *et al*, 1994) taken from historical and synthetic data. A preliminary summary of this data appears in Table 6.

6. Further Studies

Modeling typhoon rainfall based on the current NSM configurations that require searches S1 and S2 was limited. Approaches to remedy this limitation will be conducted.

The KS Tests of the block maxima indicate that the inclusion of the third central moment is beneficial for the objectives of the study. Parameter searches where rain cell intensity is exponentially distributed, gamma distributed, or by some other distribution with and without the third central moment of rain fall depth will be assessed.

Neither S1 nor S2 searches convincingly yield POT maxima that match the historical POT maxima. A solution to this, whether in the form of a new NSM equation or the search for ideal combination in the parameter search will be investigated.

Additional block maxima from the synthetic data will also be checked relative to the historical counterparts (6-hour maxima, 12-hour maxima, etc.).

A secondary confirmation, by way of extreme value distribution parameters, appears to be a necessity to reinforce findings based on the KS tests.

A practical application of this synthetic rainfall data in rainfall-runoff modeling in the Kamishiiba area will be conducted to pool synthetic stream flow peak discharges that correspond to each storm.

References

Cowpertwait, P. S. P., O'Connell, P. E. , Metcalfe, A. V. , and Mawdsley, J. A., 1996. Stochastic point process modeling of rainfall. I. Single-site fitting and validation. Journal of Hydrology. 175(1-4): 47-65.

Cowpertwait, 1998, A Poisson-cluster model of rainfall: high-order moments and extreme values. J. R. Soc. Lond. A 454, pp.885-898.

Favre, A. C., Musy, A., Morgenthaler, S., 2004, Unbiased parameter estimation of the Neyman-Scott Model for rainfall simulation with related confidence interval, Journal of Hydrology Vol. 286, pp. 168-178.

Mondonedo, C. 2007, Evaluation of the quantiles of the Neyman-Scott rainfall model, Annual Journal of Hydraulic Engineering, JSCE, vol. 51, pp. 79-84.

Press, W., Teukolsky, S., Vertterling, W., Flannery, B. 1986, Numerical Recipes in FORTRAN, Cambridge University Press, London.

Rodriguez-Iturbe, I., Cox, D., Isham, V. 1987, Some models for rainfall based on stochastic point processes, Proc. R. Soc. London A 410, pp. 269-288.

Takara, K., Stedinger, J., 1994, Recent Japanese contributions to frequency analysis and quantile lower bound estimators, Stochastic and Statistical Methods in Hydrology and Environmental Engineering, Vol 1, Kluwer Academic Publishers, The Netherlands, pp. 217-234.

極値水文量の再現性向上のためのNSM（ノイマン・スコットモデル）の改良

Carlo Arturo MONDONEDO*・立川康人*・宝　馨

＊京都大学大学院工学研究科都市環境工学専攻

要　旨

ノイマン・スコットモデルによって得られる降雨時系列から得られる月最大時間降水量および月最大日降水量が，観測降水時系列から得られるそれらの値とよく対応することを，これまでの研究により筆者らは確認した。また，ノイマン・スコットモデルのパラメータ決定において，時間降水量の平均値，分散，時間差1時間の自己相関係数，24時間雨量の分散と時間差24時間の自己相関係数を用いることで，九州上椎葉地点の降水量時系列を適切にモデル化できることを確認した。本研究では，さらに再現性の向上を図るために，降雨成因の異なる梅雨期（6月7月）と台風期（9月10月）とに分けて降水時系列データを発生させた。モデルパラメータ推定に時間降水量の３次モーメントを導入することによって，極値水文量の適合性が向上することを確認した。また，POT最大値についても，その再現性が向上することを確認した。

キーワード：確率水文学，降雨モデル，計画洪水，時系列発生，シミュレーション