Annual Journal of Hydraulic Engineering, JSCE, Vol.52, 2008, February

POT NORMALIZED VARIANCE PARAMETER SEARCH OF THE TEMPORAL NEYMAN-SCOTT RAINFALL MODEL

Carlo Arturo MONDONEDO [1], Yasuto TACHIKAWA[2], and Kaoru TAKARA[3]

[1]Student Member, Graduate student, Dept. of Urban and Environmental Eng., Kyoto University
(Kyoto 615-8540, Japan) E-mail:mondonedo@flood.dpri.kyoto-u.ac.jp

[2]Member of JSCE, Dr. Eng., Associate Professor, Dept. of Urban and Environmental Eng., Kyoto University
(Kyoto 615-8540, Japan)

[3]Fellow of JSCE, Dr. Eng., Professor, DPRI, Kyoto University (Gokasho, Uji 611-0011, Japan)

In previous investigations on the Neyman-Scott clustered Poisson rectangular pulse rainfall model (NSM), model parameters were often estimated using rainfall moments, covariances, correlations, and dry moments while model performance was often assessed through the accuracy of extreme values. No former contribution included extreme value information in improving the parameter search. This was the motivation of the authors in finding a link between the scaling apparent in the historical rainfall maxima and the NSM parameters. The normalized variance or Fano factor of the Peaks Over Threshold (POT) rainfall process was adopted to define this link. In general, the results obtained from this approach showed acceptable performance with minor limitations.

Key Words: point processes, Neyman-Scott rainfall model, Fano factor, fractals

1. INTRODUCTION

Stochastic rainfall modeling under point processes[1)] is an essential tool used by hydrologists for numerous applications. A thorough review of this subject was presented by Waymire and Gupta[2, 3, 4)]. A variant of this approach, the Neyman-Scott clustered Poisson point process, NSM here for brevity, was further developed by Rodriguez-Iturbe *et al.* [5)]

For flood control applications, a synthetic rainfall record generated from a point process such as the NSM should replicate the historical extreme values. It would thus be advantageous to consider the rainfall extreme values in estimating the parameters of the NSM. However, previous investigations did not employ this information in the parameter search, adopting mainly a method of moments approach which incorporated sample moments and correlations instead[5)]. A means to include the maximal information in the method of moments parameter search is proposed here.

Often, the scaling of statistics in point processes leads mathematically to power-law dependencies. This tendency for scaling has been a key focus in the investigations of fractal stochastic point process[6, 7)]. In fact, studies such as Telesca *et al.*[8)] demonstrated scaling in rainfall occurrence vs. counting time in historical data.

This study focuses on linking the scaling of historical extremes to the NSM parameters. Historical rainfall from the Automated Meteorological Data Acquisition System (AMeDAS) station at Kamishiiba in the southwestern island of Kyushu, Japan, was gathered to include mechanisms such as fronts, typhoons, etc. A brief description of the NSM appears in Section 2 followed by a discussion on the Peaks Over Threshold (POT) rainfall Fano factor and its eventual inclusion in the NSM in Section 3. Model application and verification appear in Section 4 followed by concluding remarks in Section 5.

2. MODEL DESCRIPTION

In the NSM, rainfall is formulated as a temporal cluster of rain cells which occur based on a Poisson process with occurrence rate λ. Cluster size is based on a geometric distribution with mean μ_c. Rain cells lag cluster occurrences randomly following an exponential distribution with mean $1/\beta$. Rain cell durations follow an exponential distribution as well with mean $1/\eta$. Rain cell intensities are modeled by a gamma

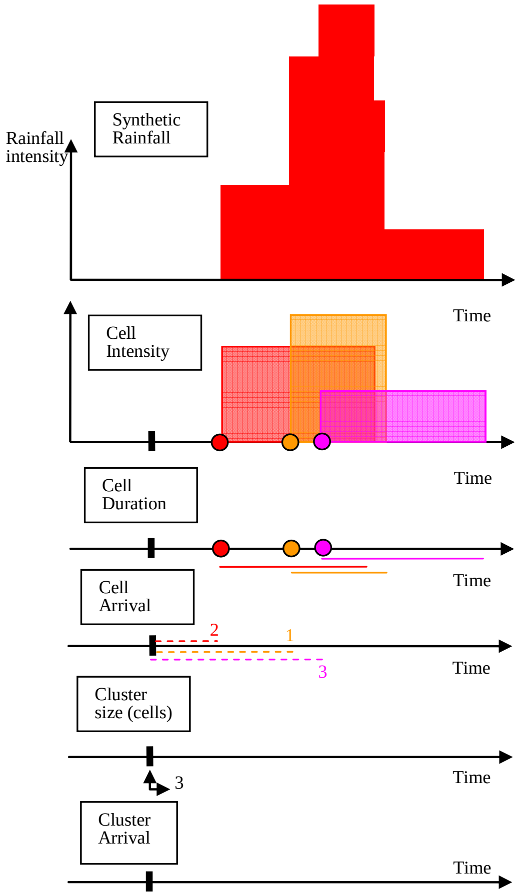


Fig. 1. Schematic drawing of the Neyman-Scott Rainfall Model.

distribution with scale parameter θ and shape parameter α. The resulting rainfall is the superposition of these processes as shown in **Fig. 1**.

The moment and correlation expressions for the aggregated rainfall point process in terms of the NSM parameters were derived by Rodriguez-Iturbe *et al.*[9] The succeeding Equations (1) – (3) list these expressions in the aggregated form, specific to the distribution parameters mentioned: β, λ, μ_c, η, α, and θ.

$$E\left\langle Y_n^{(h)}\right\rangle = \lambda\mu_c(\alpha\theta)h/\eta \tag{1}$$

$$\operatorname{var}\left\langle Y_n^{(h)}\right\rangle = \frac{2\lambda\mu_c\alpha\theta^2(1+\alpha)(\mu_c-1)}{\beta\eta^3(\beta^2-\eta^2)}\times \left[\beta^3 A_1(h)-\eta^3 B_1(h)\right]+\frac{4\lambda\mu_c(\alpha\theta)^2 A_1(h)}{\eta^3} \tag{2}$$

$$\operatorname{cor}\left\langle Y_n^{(h)},Y_{n+k}^{(h)}\right\rangle = \frac{2\lambda\mu_c\alpha\theta^2(1+\alpha)(\mu_c-1)}{\operatorname{var}\left\langle Y_i^{(h)}\right\rangle\beta\eta^3(\beta^2-\eta^2)}\times \left[\beta^3 A_2(h,k)-\eta^3 B_2(h,k)\right] + \frac{4\lambda\mu_c(\alpha\theta)^2 A_2(h,k)}{\eta^3} \tag{3}$$

$$A_1(h) = \eta h - 1 + e^{-\eta h}$$
$$B_1(h) = \beta h - 1 + e^{-\beta h}$$
$$A_2(h,k) = 0.5\left(1-e^{-\eta h}\right)^2 e^{-\eta h(k-1)}$$
$$B_2(h,k) = 0.5\left(1-e^{-\beta h}\right)^2 e^{-\beta h(k-1)}$$

where:

n = time interval counter
h = integer specifying time step interval of data (1 for 1 hour, 24 for 1 day, etc.)
Y_n^h = rainfall depth in the n-th time of interval h
$E\left\langle Y_n^{(h)}\right\rangle$ = mean rainfall depth record at h-hours
$\operatorname{var}\left\langle Y_n^{(h)}\right\rangle$ = variance of rainfall record at h-hours
$\operatorname{cor}\left\langle Y_n^{(h)},Y_{n+k}^{(h)}\right\rangle$ = autocorrelation of rainfall record at h-hours at lag k

3. THE POT RAINFALL AND NSM FANO FACTOR EXPONENT

(1) POT Rainfall count process $N_i(t)$

Figure 2 shows the construction of the special count process $N_i(t)$ to be used in determining the Fano factor $FF(T)$ at an arbitrary window T. With rainfall pooled for each month i (i=1: January, i=2: February, etc.), we define the process $U_i(t)$ (**Fig. 2a**), as the rainfall magnitude per fixed duration t (i.e.: t=1 hour in this study). For the M months in each $U_i(t)$, we determine the M monthly or block maxima $Z_i(t)$ and threshold value z_i, the minimum of $Z_i(t)$. We define the POT rainfall point process $Q_i(t)$ based on Eq. (4) (**Fig. 2b**).

$$Q_i(t) = \begin{cases} U_i(t); & U_i(t) \ge z_i \\ 0\ ; & U_i(t) < z_i \end{cases} \tag{4}$$

For each rainfall occurrence in $Q_i(t)$, we define the binary process $B_i(t)$ based on Eq.(5) (**Fig. 2c**). This process counts all durations of rainfall greater than or equal to the threshold z_i and serves as the basis for defining the Fano factor used here:

$$B_i(t) = \begin{cases} 1; & Q_i(t) > 0 \\ 0; & Q_i(t) = 0 \end{cases} \tag{5}$$

Using adjacent windows of arbitrary value T, we define the count process $N_i(T)$, the sum of $B_i(t)$ within each segment T (**Fig. 2d**).

(2) Fano factor and governing relationships

The Fano factor $FF(T)$ is defined as the ratio of variance of count $N_i(T)$ and mean of $N_i(T)$, or:

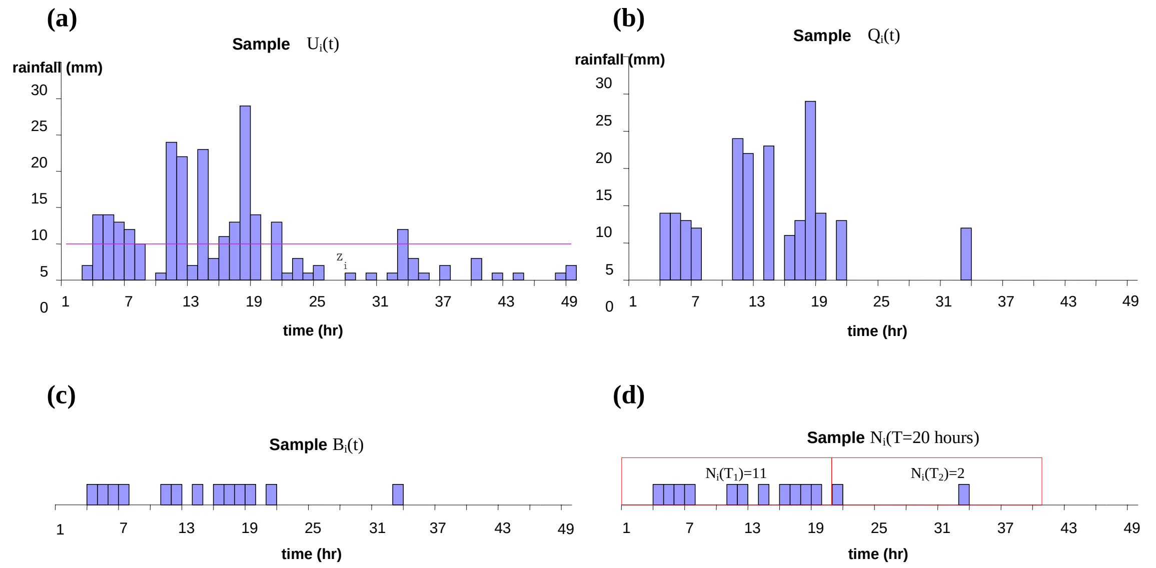


Fig. 2. Determining the Peaks Over Threshold (POT) rainfall point process and count process $N_i(t)$. (a) sample rainfall $U_i(t)$ with previously determined threshold z_i, (b) POT rainfall $U_i(t)$ values greater than or equal to z_i (c) unit counts $B_i(t)$ assigned for each rainfall occurrence, (d) count process $N_i(t)$.

$$FF_i(T) = \frac{E\langle [N_i(T)]^2 \rangle - E\langle N_i(T) \rangle^2}{E\langle N_i(T) \rangle} \quad (6)$$

where:

$E\langle \ \rangle$ = operation to obtain expected value.

Based on data from independent studies, Lowen and Teich[6,7] and Telesca *et al.*[8] proposed the power law relationship of Eq. (7) to describe the scaling that occurs over several decadal values of T (**Fig. 3**):

$$FF(T) = 1 + \left(\frac{T}{T_0}\right)^{\xi} \quad (7)$$

in which T_0 is the basic data duration (i.e.: $T_0 = 1$ hour here). Strictly speaking, the rainfall data used by Telesca *et al.*[8] were pooled yearly ($z_i = 0$) instead of $Q_i(t)$. It was assumed here that Eq. (7) is valid for $Q_i(t)$ throughout the set of windows $T \in \Omega = \{2, 10, 20, \ldots, 100 \text{ hours}\}$, a conservative range for possible mean storm durations (**Fig. 3b**).

Alternatively, within Ω, the approximations for variance and mean of $N_i(T)$ are proposed here as Eqs. (8) and (9) such that the Fano factor can be independently estimated as Eq. (10).

$$Var\langle N_i(T) \rangle = A_i T^{B_i} \quad (8)$$

$$E\langle N_i(T) \rangle = C_i T \quad (9)$$

$$FF_i(T) \approx \frac{A_i}{C_i} T^{B_i - 1} \quad (10)$$

(3) NSM Fano factor exponent

Curve fitting operations for Eqs. (8), (9), and Eq. (10) (i.e.: least squares fit for A_i, B_i, and C_i) are used to explicitly determine $FF_{Hi}(T_{Mi})$, the Fano factor of the historical $Q_i(t)$ at window T = mean storm duration T_{Mi} such that:

$$FF_{Hi}(T_{Mi}) \approx \frac{A_i}{C_i} T_{Mi}^{B_i - 1} \quad (11)$$

The relationship for the synthetic equivalent $FF_{Si}(T_{Mi})$ can be written explicitly using Eq. (7) such that:

$$FF_{Si}(T_{Mi}) = 1 + \left(\frac{T_{Mi}}{T_0}\right)^{\xi_{Si}} \quad (12)$$

For our purposes, an ideal simulation should yield synthetic rainfall with $Q_i(t)$ such that historical Fano factor $FF_{Hi}(T_{Mi})$ and synthetic Fano factor $FF_{Si}(T_{Mi})$ are equal, or based on Eqs. (10) and (12):

$$\frac{A_i}{C_i} T_{Mi}^{B_i - 1} = 1 + \left(\frac{T_{Mi}}{T_0}\right)^{\xi_{Si}} \quad (13)$$

Isolating the synthetic Fano factor exponent ξ_{Si} leads to an expression relating properties of the historical $Q_i(t)$ to the unknown mean duration T_{Mi}.

$$\xi_{Si} = \frac{\ln\left[\left|\frac{A_i}{C_i} T_{Mi}^{(B_i - 1)} - 1\right|\right]}{\ln[T_{Mi}] - \ln[T_0]} \quad (14)$$

Actual mean storm duration T_{Mi} was estimated here based on Cowpertwait's expression[10] derived from the NSM parameters shown here as Eq. (15).

$$T_{Mi} = \frac{1}{\eta}\left(\gamma + \ln\left[(\mu_c - 1)\frac{\eta}{\eta - \beta}\right]\right) \quad (15)$$

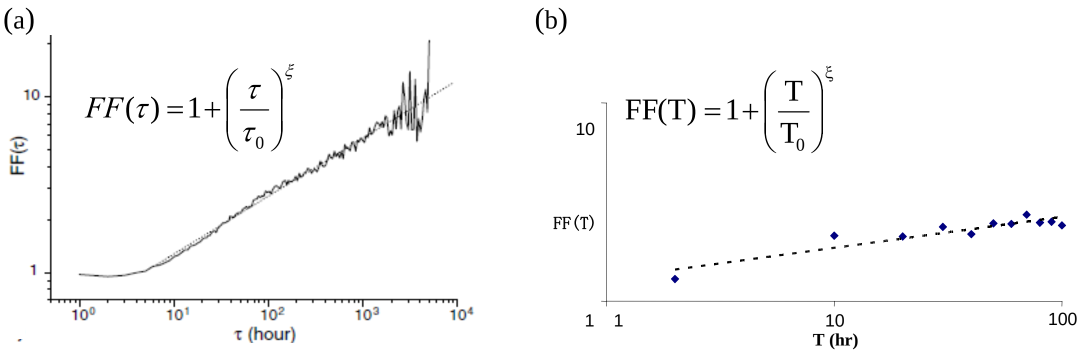


Fig. 3. (a) Scaling observed between count of spikes from a nerve fiber FF(τ) vs. counting time τ (reproduced from Lowen and Teich[6]). (b) Scaling in the count process $N_i(t)$ obtained from Kamishiiba POT series $Q_i(t)$ in June.

Table 1. Estimated NSM Parameters for Kamishiiba Region.

	α	β (1/hr)	η (1/hr)	λ (1/hr)	μ_c	θ (mm/hr)	T_{Mi} (hr)
Jan	0.0295	0.2096	0.5000	0.0155	50.0000	2.2066	23.9182
Feb	20.0000	0.2741	60.0000	0.0116	50.0000	0.7379	16.3187
Mar	20.0000	0.3141	60.0000	0.0205	21.6258	1.7363	11.4900
Apr	20.0000	0.3548	60.0000	0.0208	28.1992	1.3535	10.9542
May	20.0000	0.2302	24.5687	0.0113	50.0000	0.6774	19.4561
Jun	20.0000	0.1925	60.0000	0.0202	37.3840	2.9474	21.6899
Jul	20.0000	0.0771	60.0000	0.0063	44.6919	6.3261	56.4790
Aug	20.0000	0.0687	26.6501	0.0052	50.0000	2.4956	65.0530
Sep	20.0000	0.0762	3.4522	0.0051	33.2218	0.4549	53.4124
Oct	20.0000	0.1163	58.2798	0.0041	38.1742	3.1718	36.0710
Nov	20.0000	0.1432	9.9924	0.0053	24.2467	0.3746	26.0940
Dec	20.0000	0.1800	60.0000	0.0049	31.0726	1.4719	22.1366

By substituting Eq. (15) for T_{Mi} in Eq. (14), a direct link between historical $Q_i(t)$ and the NSM parameters is established as Eq. (16).

$$\xi_{Si} = \frac{\ln\left[\left|\frac{A_i}{C_i}\left\{\frac{1}{\eta}\left(\gamma + \ln\left[(\mu_c - 1)\frac{\eta}{\eta-\beta}\right]\right)\right\}^{(B_i-1)} - 1\right|\right]}{\ln\left[\frac{1}{\eta}\left(\gamma + \ln\left[(\mu_c - 1)\frac{\eta}{\eta-\beta}\right]\right)\right] - \ln[T_0]} \qquad (16)$$

Moreover, with the historical exponent ξ_{Hi} estimable through curve fitting Eq. (7) to the historical $Q_i(t)$, it is now possible to include ξ_{Si}, the NSM Fano factor exponent, in the NSM parameter estimation. Thus, in a limited manner the proposed parameter search contains information about historical extreme values that should lead to improved synthetic maxima.

4. MODEL APPLICATION

(1) Parameter Estimation

A basic scheme consisting of the existing NSM moments and correlation equations and the previously derived NSM Fano factor exponent ξ_{Si} is proposed here. The objective function in the parameter search is given by Eq. (17).

$$O = \sum_{m=1}^{6}\left(1 - \frac{NSM_m}{HIS_m}\right)^2 \qquad (17)$$

$$\theta = \frac{E\langle Y_n^{(1)}\rangle \eta}{\lambda \mu_c \alpha} \qquad (18)$$

where:
NSM_m = any of equations (1), (2), (3), and (16)
HIS_m = equivalent historical value of moment.

To determine β, λ, μ_c, η, α, and θ, six expressions were required in the search: (1) hourly variance, (2) hourly correlation at lag 1, (3) 12-hourly correlation at lag 1, (4) 24-hourly variance, (5) 24-hourly correlation at lag 1, and (6) hourly Fano exponent ξ_{Si}. Eq. (18), the scale parameter θ in terms of the mean hourly rainfall, was used to limit the computational difficulty in numerically solving Eq. (17). Historical hourly rainfall data was obtained from the AMeDAS station in Kamishiiba, Kyushu, Japan from 1988 to 2002. Estimated parameters appear in **Table 1**, indicating T_{Mi} values within the range of Ω.

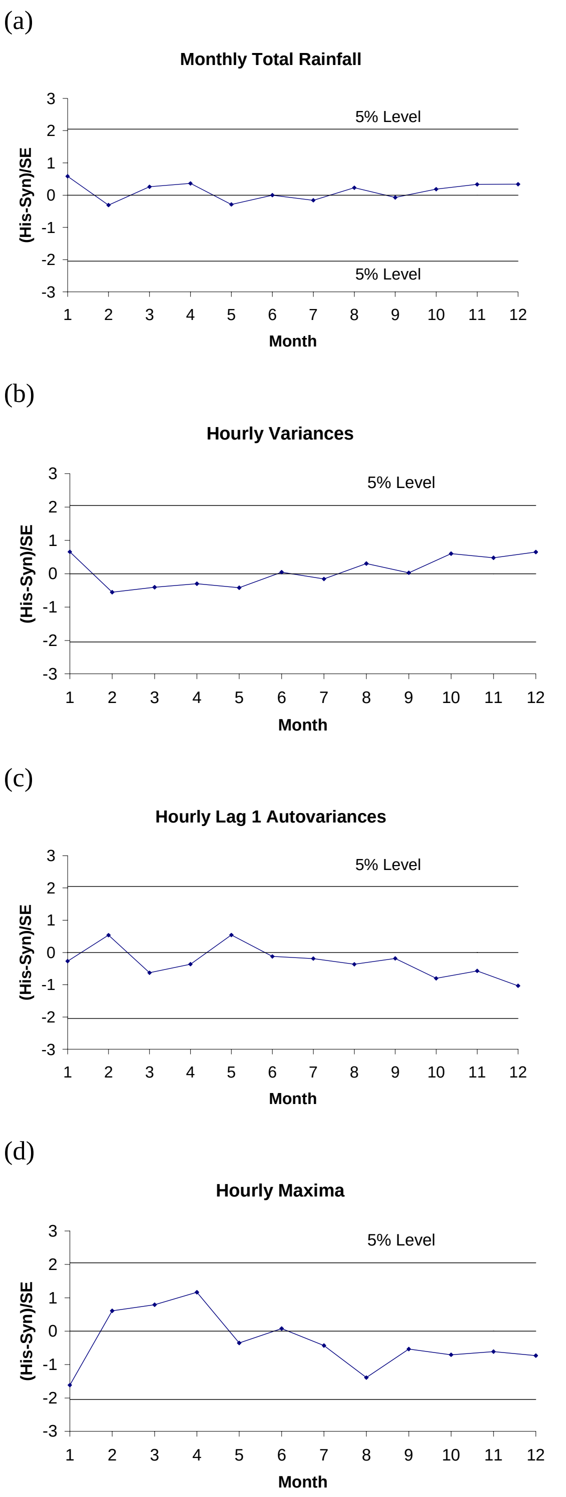


Fig. 4. Student's T tests for the hourly rainfall point process for (a) monthly totals, (b) hourly variances, (c) lag 1 autocovariances, and (d) hourly maxima.

(2) Model Performance

Fifteen synthetic records were generated per month corresponding to the 15 monthly historical records. Each student's *t* ordinate t_S in **Fig. 4a** correspond to the ratio of the difference between average historical monthly total rainfall and its synthetic counterpart and the estimate of the standard error. The values were observed to lie within the 5% significance value of 2.048 (for the sample size $N_{His} + N_{Syn} - 2 = 28$), which indicates acceptable model performance. The same is true for other statistics such as those shown in **Figs. 4b-c** and **5**. Monthly maxima of lower rank were also within the 5% significance levels and were omitted here.

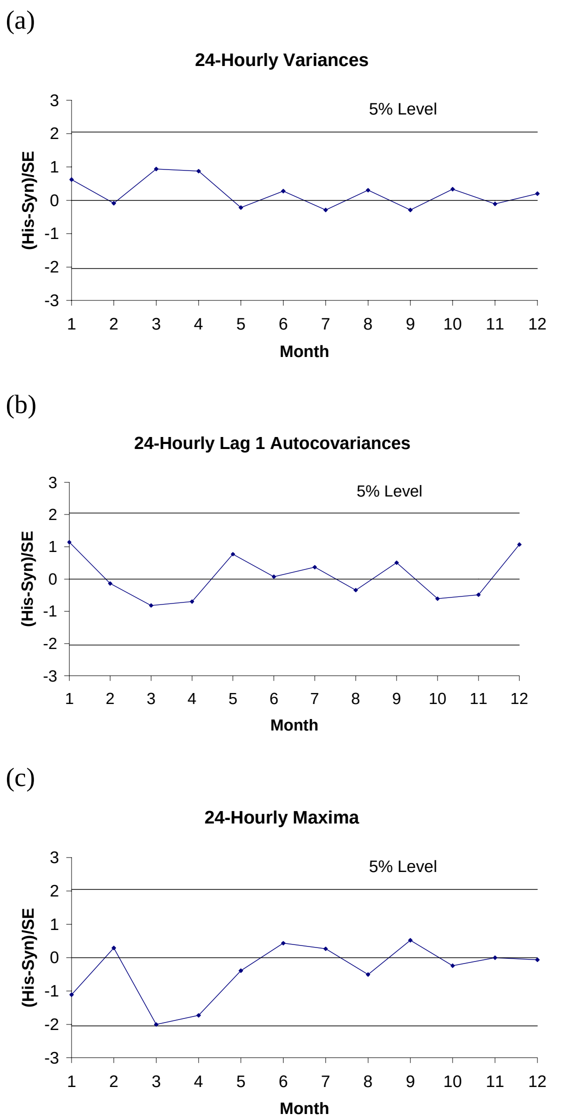
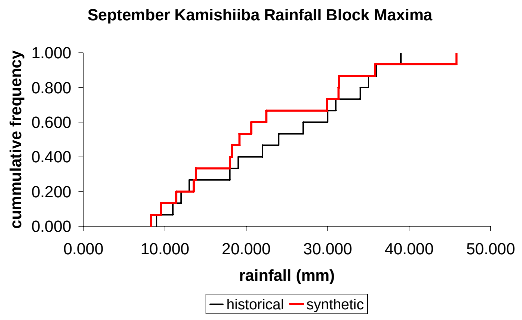


Fig. 5. Student's T tests for the hourly rainfall point process for (a) 24 hourly variances, (b) lag 1 autocovariances, and (c) 24-hourly maxima.

(3) Model Limitation

Figure 6 shows a sample Kolmogorov -Smirnov test (KS test) applied to the block maxima of September. The test was formulated such that 2 data sets that are more consistent in probabilistic distribution receive a KS probability approaching unity[11]. Based on **Table 2**, the synthetic block maxima were significantly different from the historical counterparts for several months (i.e.: values below 0.95). During these months, rain cell information may by more relevant than storm cluster information. Therefore, we propose replacing the 12-hourly lag 1 correlation with the 6-hourly lag 1 correlation to improve the results for these months.



KS significance prob. = 0.99

Fig. 6. Historical vs. synthetic rainfall Kolmogorov-Smirnov test using 15 monthly maximum values per set.

5. CONCLUSION

A new expression referred to here as the Fano factor exponent was determined for the Neyman-Scott Clustered Poisson Rectangular Rainfall Model, or NSM. Unlike previous parameter search methods for the NSM, this version linked the model parameters to the Fano factor of a portion of the historical data based on the POT rainfall series. Including this Fano factor exponent expression in the search estimated NSM parameters that were satisfactory for generating synthetic rainfall with significant mean total rainfall, variances, autocovariances and monthly maxima.

However, in several cases the same parameters generated synthetic block maxima that were inconsistent with the historical block maxima in empirical distribution. To rectify this limitation during these months, rain cell based information should be prioritized in the parameter search. For this purpose, the 6-hourly lag 1 autocorrelation should be used instead of the 12-hourly lag 1 autocorrelation.

ACKNOWLEDGMENTS

This study was supported by Grant-in-Aid for Scientific Reserch (C) 18560497 (PI: Assoc. Prof. Y. Tachikawa, Kyoto Univ.).

Table 2. KS Test results for Kamishiiba rainfall.

	Block Maxima KS Probability	
Month	1-Hour	24-Hour
1 (Jan)	0.83	0.83
2	0.98	0.98
3	0.83	0.55
4	0.98	0.28
5	0.99	0.99
6	0.98	0.99
7	0.98	0.99
8	0.55	0.98
9	0.99	0.83
10	0.99	0.99
11	0.98	0.99
12 (Dec)	0.98	0.99

REFERENCES

1) Cox, D. and Isham, V.: *Point processes,* Chapman and Hall, London, 1980.
2) Waymire, E. and Gupta, V.: The mathematical structure of rainfall representations 1: A review of stochastic rainfall models, *Water Resources Research*, Vol. 17, No. 5., pp. 1261-1272, 1981a.
3) Waymire, E. and Gupta, V.: The mathematical structure of rainfall representations 2: A review of the theory of point processes, *Water Resources Research*, Vol. 17, No. 5., pp. 1273-1285, 1981b.
4) Waymire, E. and Gupta, V.: The mathematical structure of rainfall representations 3: Some applications of the point process theory to rainfall processes, *Water Resources Research*, Vol. 17, No. 5., pp. 1287-1294, 1981c.
5) Rodriguez-Iturbe, I., Gupta, V. and Waymire, E.: Scale considerations in the modeling of temporal rainfall, *Water Resources Research*, Vol. 20, No. 11., pp. 1611-1619, 1984.
6) Lowen, S. and Teich, M.: Estimation and simulation of fractal stochastic processes, *Fractals*, Vol. 3., No. 1, pp. 183-210, 1995.
7) Lowen, S. and Teich, M.: *Fractal-Based Point Processes*, Wiley Series in Probability and Statistics, New Jersey, 2005.
8) Telesca, L., Lapenna, V., Scalcione, E. and Summa, D.: Searching for time-scaling features in rainfall sequences, *Chaos, Solitons, and Fractals*, Vol. 32, pp. 35-41, 2007.
9) Rodriguez-Iturbe, I., Cox, D. and Isham, V. : Some models for rainfall based on stochastic point processes, *Proc. R. Soc. London A*, Vol. 410, pp. 269-288, 1987.
10) Cowpertwait, P.: Further developments of the Neyman-Scott clustered Poisson process for modeling rainfall, *Water Resources Research*, Vol. 27, No. 7, pp. 1431-1438, 1991.
11) Press, W., Teukolsky, S., Vetterling, W. and Flannery, B.: *Numerical Recipes in FORTRAN*, 2nd ed. Cambridge University Press, London, 1992.

(Received September 30, 2007)