EVALUATION OF THE QUANTILES OF THE 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 (Katsura, Kyoto 615-8530, Japan) E-mail:mondonedo@flood.dpri.kyoto-u.ac.jp

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

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

The serial application of a stochastic rainfall time series model and a rainfall-runoff model is an effective option in the event when historical discharge records are inadequate for determining quantile flood flows crucial in flood control applications. In this study, the Neyman-Scott clustered Poisson rectangular pulse rainfall model was applied to generate synthetic hourly and daily rainfall time series. This model was evaluated for quantiles for the different areas of Kamishiiba (Kyushu), Naha (Okinawa), and Sapporo (Hokkaido) in Japan. Sets of parameters were determined based on combinations of moments of the historical records. The quantiles of each synthetic time series were compared to the historical counterparts, yielding mixed results. In general, the model proved effective in matching the historical quantile except in periods in which a combination of rainfall sources such as convective, typhoon and/or frontal rainfall was dominant. An ideal combination of historical moments for generating model parameters for the objective was determined for the mentioned study areas, although it is recommended to employ a similar methodical test of moment combinations for regions elsewhere.

Key Words: *stochastic rainfall modeling, Neyman-Scott rainfall model, design flood, quantile rainfall*

1. INTRODUCTION

Flood control decision variables connected to design and operation depend on an appropriate availability of discharge records. Ideally, such decision variables should be determined based on a quantile of flood discharge. Unfortunately, discharge records may be short (less than 30 years) and/or nonstationary due to land use change and/or river intrusive construction, restricting effective extreme value analysis. Generating this data, via input of synthetic rainfall time series into a rainfall-runoff model provides a valuable option in this case. As such, stochastic techniques based on the point process[1, 2] approach can be employed to generate synthetic rainfall. However, to be applicable in hydraulic design applications, the quantiles of the historical rainfall record should be well represented in the synthetic time series as well. This study emphasizes on the quality of the synthetic quantiles generated by one stochastic rainfall model of interest.

The Neyman–Scott clustered Poisson rectangular pulse rainfall model applied in this study follows the basic formulation of Rodriguez-Iturbe, *et al.*[3], NSM here for brevity. Based on a clustered point process[1, 2] approach, this model is theoretically capable of generating time series that are consistent at several levels of aggregation. The NSM was widely studied to model the moment characteristics of historical rainfall records. The matching of the synthetic extreme values to the historical ones however, has been examined less. Unlike previous studies therefore, the emphasis here is to investigate the ability of the NSM to preserve the historical quantile rainfall depths of 1-hour and 24-hour duration, which may be basic information necessary for risk analysis in flood control applications. In so doing, longer NSM rainfall records, although synthetic, can be available as reliable bases of quantile flood events.

The historical data used in this study were obtained from three locations in Japan to incorporate the effect on the rainfall by local conditions such as susceptibility to fronts and/or typhoons. Sixteen yearly records (1988 to 2003) of hourly rainfall were collected from Kamishiiba (Kyushu Island), while 27 years (1976-2002) each of hourly rainfall were collected from Naha

(Okinawa) and Sapporo (Hokkaido) as the main historical records of this study.

Section 2 of this paper presents a review of the NSM in its governing equations and parameter estimation requirements. The details of the model are limited here, although past studies (Rodriguez-Iturbe, *et al.*[3]), Burlando and Rosso[4]), Cowpertwait[5])) present a more thorough discussion. The applications of the model to the above-mentioned study areas appear in section 3, followed by a discussion in section 4. Section 5 presents the conclusions of this study.

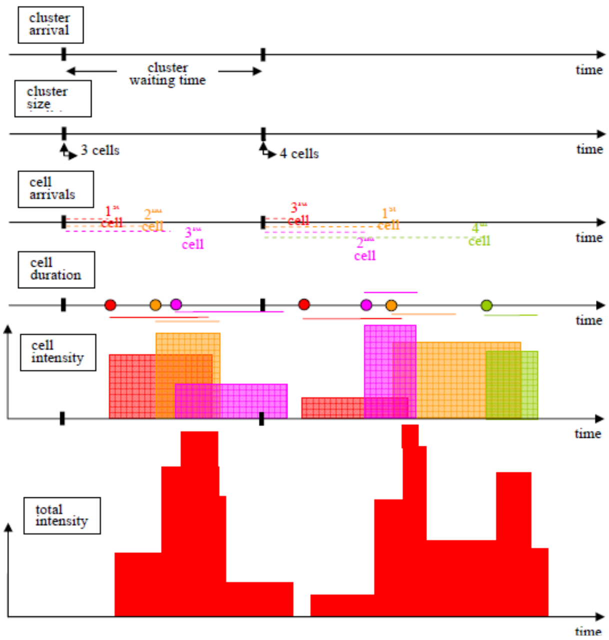

Fig. 1. Schematic drawing of the Neyman-Scott Rainfall Model.

2. NEYMAN-SCOTT RAINFALL MODEL

(1) Governing equations

This version of the NSM consists of five probability distributions, based on Rodriguez-Iturbe's original formulation[3]) (see Fig. 1). Clusters of rain cells are linked integrally to a storm origin with mean occurrence rate λ, regarded as a Poisson process. Each cluster can have a random number of rain cells described by a geometric distribution (with all clusters containing at least one cell) with mean cell number μ_c. Relative to the cluster origin, the random arrival of each cell is based on an exponential distribution with mean displacement 1/β. Each cell has a corresponding independent identically distributed (iid) random intensity and duration, also based on the exponential distribution with mean intensity μ_x and mean cell life span 1/η. The total rainfall intensity is then the superposition of the effects of these random cell intensities.

Inherent in the model is the assumption of stationarity in the mean and variance. In applying the model therefore, it would be beneficial to have as long a historical rainfall record as possible (about 20 years if possible). The historical rainfall record can be expressed in terms of the model parameters as shown in Eqs. (1) – (3) (Rodriguez-Iturbe, *et al.*[3])).

$$E\langle Y_i^{(h)}\rangle = \lambda\mu_c h/(\eta\mu_x) \quad (1)$$

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

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

The previous equations were derived from the method of moments[3]) in which:

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

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

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

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

where:

i = time interval counter

h = integer specifying time step interval of data (<u>1</u> for 1 hour, <u>24</u> for 1 day, etc.)

Y_i^h = rainfall depth in the i-th time of interval h

$E\langle Y_i^{(h)}\rangle$ = mean rainfall depth record at h-hours

$\mathrm{var}\langle Y_i^{(h)}\rangle$ = variance of rainfall record at h-hours

$\mathrm{cov}(Y_i^h, Y_{i+k}^h)$ = covariance of rainfall record at h-hours at lag k

(2) Parameter Estimation

Parameter estimation following the method of moments approach was adopted here although maximum likelihood estimators may be used as well[4]). Nontrivial combinations of Eqs. (1) – (3) are required to form the objective function for the estimation of parameters λ, η, μ_c, μ_x, and β in NSM. Normally, the system formed by these combinations is solved for the required parameters by minimization. At least five equations are therefore required in one combination. For the estimation, the objective function was adopted from previous studies[6] to ensure that large numerical values do not dominate the fitting procedure. This objective function appears here as Eq. (4).

$$F = \sum_{j=1}^{M} \left(\frac{f_j(Y_i)}{W_j} - 1 \right)^2 \quad (4)$$

where:

$f_j(Y_i)$ = j-th NS moment equation of rainfall time series Y_i (from Eqs. (1) – (3)).

W_j = historical moment value from rainfall record.

M = number of equations to be adopted in the estimation.

Combinations of Eqs. (1) – (3) are required in (4), depending on the target use of the final NSM. Previous combinations include those employed by Rodriguez-Iturbe *et al.*[3], Burlando and Rosso[3], Cowpertwait, *et al.*[6], Calenda and Napolitano[7], and Favre *et al.*[8], to cite a few. For instance, the determination of the five parameters can be based on the basic combination Test Set I (see Table 1).

The frequency distributions of the hourly and daily maximum in one month per record of the

Table 1. Proposed test sets of historical moment combinations for NSM parameter estimation problem.

Test Set		1	6	12	24	48
I	Mean	○				
	Var	○			○	
	Cov1*	○			○	
II	Mean	○				
	Var	○	○		○	
	Cov1	○	○		○	
III	Mean	○				
	Var	○		○	○	
	Cov1	○		○	○	
IV	Mean	○				
	Var	○			○	○
	Cov1	○			○	○
V	Mean	○				
	Var	○	○	○	○	
	Cov1	○	○	○	○	
VI	Mean	○				
	Var	○		○	○	○
	Cov1	○		○	○	○

*Covariance at lag 1

Table 2. Range of NSM parameters used for optimization.

Parameter	Min	Max
λ (1/h)	0.001	0.050
β (1/h)	0.01	0.50
μ_x (mm/h)	0.30	15.0
μ_c	2.00	100.00
η(1/h)	0.10	5.0

Table 3. Sample NSM parameters for July.

NSM Param.	Kamishiiba	Naha	Sapporo
λ (1/h)	0.00636	0.00185	0.00828
β (1/h)	0.07107	0.01000	0.21714
μ_x (mm/h)	4.49481	10.51604	2.99253
μ_c	44.33524	13.93186	10.46855
η (1/h)	2.17691	1.34515	2.65969

generated synthetic data should be sufficiently close to those of the historical maxima in this study (see Section 4). Since Eqs. (1), (2), and (3) does not directly employ information from the maxima in the estimation of the parameters, it would be appropriate to assume that Test Set I may not cover this requirement for all possible regional conditions of the selected study areas. To cover such possible dependencies of the target maxima on the short-term and long-term moments of the historical data, five more combination test sets were adopted, as shown in Table 1.

The minimization techniques adopted here employ the constraints shown in Table 2. Calenda and Napolitano[7] adopted these ranges in their study on unbiased parameter estimates for the NSM. The application of these ranges in this study include the use of the Nelder-Mead simplex[9] and the Levenberg-Marquardt[9] minimization techniques in tandem for initial and refined estimation, respectively.

A sample of the estimated July parameters Test Set I for the three study areas appear in Table 3. In general, the characteristics of the study areas are manifest through this table of sample parameters (see (1)). For instance, the higher λ and the moderate β in the Kamishiiba region indicates more frequent clusters with moderate gaps between storm birth and rainfall initiation, respectively. When a cluster does arrive in this mountainous region, it is on the average densely packed (high μ_c), with moderate intensity (μ_x) and short duration (η), possibly characterizing the orographic effect. In contrast, clusters are less dense in Naha (moderate μ_c), where the orographic effect is less prevalent. Although clusters are moderately frequent, rain cells occur long after the storm origin arrival, with higher intensity and longer duration, possibly reflecting the relatively high vapor storage capacity in this

Table 4. Sample moments generated from July NSM.

	Kamishiiba		Naha		Sapporo	
	His.	Syn.	His.	Syn.	His.	Syn.
HM	0.583	0.578	0.202	0.186	0.097	0.097
HSD	2.531	2.487	1.746	1.737	0.698	0.681
HC1	0.672	0.672	0.489	0.500	0.517	0.511
DM	13.982	13.870	4.846	4.471	2.339	2.328
DSD	37.600	36.762	15.928	16.031	7.116	6.856
DC1	0.348	0.352	0.376	0.369	0.096	0.093

HM = Hourly Mean, HSD, Hourly Stand. Dev., HC1 = Hourly Covariance at lag 1
DM = Daily Mean, DSD, Daily Stand. Dev., DC1 = Daily Covariance at lag 1

tropical maritime region. Sapporo historically experiences less rainfall, with relatively low rain cell count, low intensity and short duration. Also, despite the moderate storm frequency, the gap from origin to rain initiation is quite small.

Table 4 shows sample moments that were generated for Kamishiiba from the NSM model using TS I parameters for July (Table 3). In general, the moments of the historical data were well-maintained in the generated synthetic data.

3. MODEL APPLICATION

It was necessary to divide the historical records on a monthly basis from June to October, the accepted “rainy season” of Japan. This was done to affect a stationarity of historical data required in the NSM[3]. Parameter sets were then estimated for each combination for each monthly historical record of each study area (see Section 4).

The uniform deviate generator used here was developed from the Park-Miller “minimal standard” generator based on the simple multiplicative congruential algorithm[9]. The inverse CDF method[10] (ICDFM) was used for generating continuous and discrete random variables (through a look-up table implementation of the ICDFM for the latter).

One hundred month-hours of rainfall were then generated for each month (i.e.: 72,000 hours of synthetic records for June, 74,400 for July, etc.). Each synthetic hourly time series was then aggregated for every 24 hours to generate its corresponding synthetic daily time series.

The maxima of each month was searched from each synthetic record and compared to its historical counterpart. A Kolmogorov-Smirnov (KS) test was conducted to check for significance of match. The KS probability P_{KS} of 95% was designated to be the ideal result at which the synthetic maxima and historical maxima were most likely from the same population. The governing equations of the KS test applied here appear in Eqs. (5) – (8) (see Press, *et al.*[9]).

$$Q_{KS}(\gamma) = 2\sum_{j=1}^{\infty}\left(4j^2\gamma^2 - 1\right)e^{-2j^2\gamma^2} \tag{5}$$

$$Q_{KS}(0) = 1\ ;\ Q_{KS}(\infty) = 0 \tag{6}$$

$$P_{KS} = Q_{KS}([\sqrt{N_e} + 0.155 + 0.24/\sqrt{N_e}]D) \tag{7}$$

$$N_e = \frac{N_h N_s}{N_h + N_s} \tag{8}$$

where:
P_{KS} = Kolmogorov-Smirnov probability
D = maximum deviation between cumulative frequency distributions of historical and synthetic rainfall maxima.
N_h = length of historical record.
N_s = length of synthetic record.

This test was performed for each synthetic time series and its corresponding historical counterpart.

4. APPLICATION RESULTS

The three regions in which the NSM was applied have varying regional rainfall characteristics. Isolated from the effects of most typhoons, the Sapporo area of Hokkaido in Northern Japan has a smaller average rainfall during the rainy season (with a 2-14 mm/day average rainfall during the rainy months of June to October). In contrast, the southern tropical maritime region of Naha in Okinawa is more frequented by typhoons (with a 5-30 mm/day average rainfall during the rainy months). Finally, located closer to Okinawa, the mountainous Kamishiiba region is the area that receives the most rainfall among the three study areas (with a 4-30 mm/day average rainfall during the rainy months). Based on section 2 (2), these conditions directly affected the simulation results.

The mean and variances of the synthetic data match those of the historical records, based on Student’s t-tests for significantly different means[9] and F-tests for significantly different variances[9]. The results of these tests were omitted here as this was the expected result in the conception of the NSM. The previously described differing regional conditions appear to have no effect on the proximity of the synthetic moments to the historical moments. This indicates in general that the clustered rectangular rainfall pulses used in NSM can account for the varying rainfall properties inherent in the study areas.

However, not all the maxima of the synthetic

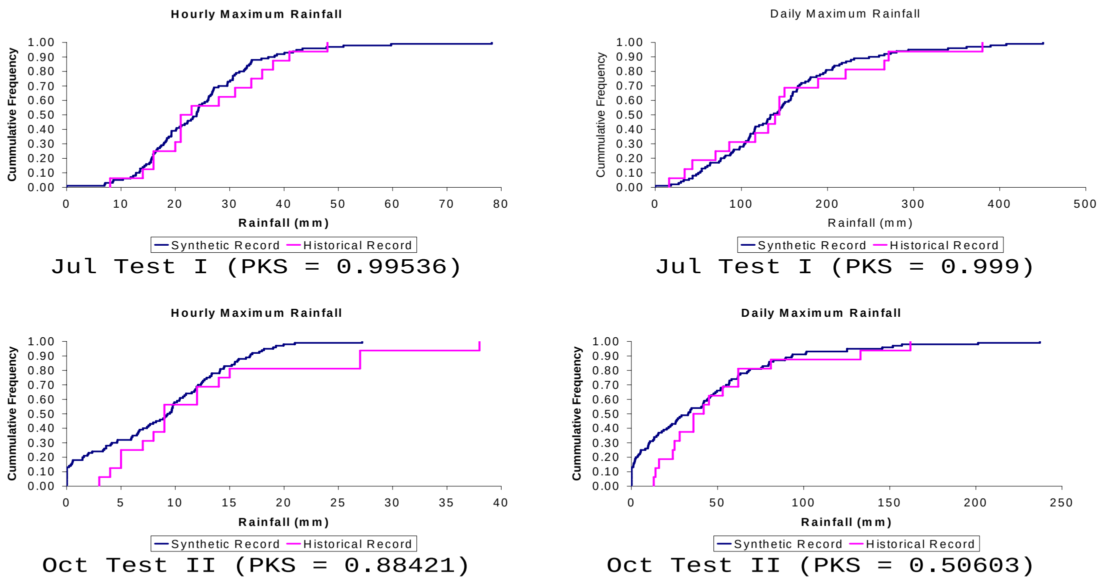

Fig. 2. Sample KS-Test results from NSM (taken from Kamishiiba application).

records appear to match the maxima of the historical records. Figure 2 shows a sample of Kamishiiba cumulative distributions of historical and synthetic monthly maximum rainfall intensity in which 16 historical monthly maximum were compared to 100 synthetic monthly maximum for the hourly and daily records. KS test results gathered were classified into 3 types. Type I results reflect a 95% KS probability for the hourly and daily maxima, indicating significant similarity between historical and synthetic records. Type II results reflect KS probabilities below 95% for both the hourly and daily maxima. Type III results reflect at least a 95% KS probability for either the hourly or daily synthetic maxima.

Table 5 summarizes the results of KS tests for hourly and daily data at the three study regions. Although the majority of the KS results were Type I, most Type II and III results were concentrated in the August-September period of Kamishiiba, and during June and October in Naha. However, during these periods, the arrival and duration of rain cells of a cluster of the present formulation may not be adequate. In particular, during August and September in Kamishiiba, it is possible for the region to experience the usual share of convective rainfall as well as the irregular passing of a typhoon. Under such conditions, the rain cell composition of the historical records is a mix from the two rainfall sources. This affected the maxima of the synthetic record although the synthetic moments were still closely matched to those of the historical records.

Type II and III results were likewise concentrated on the months of June and October in Naha. A similar pattern in which the rainfall seems to be mixed in sources as typhoon and convective rainfall appears to occur in October. However, although typhoons can occur in the rainy month of June over Naha, it is possible that this region also experiences a frontal rainfall source.

Incorporating covariances of higher order may rectify the limitations due to the mix of sources, although an adequate representation of the mixed temporal structure appears to be a more thorough approach. In his formulation, Cowpertwait[5] proposed a cluster composition consisting of so-called “light cells” of long expected duration and “heavy cells” of short duration. This formulation appears to be more appropriate for the mixed-source conditions.

Test IV and VI could not be applied successfully to estimate the NSM parameters from the Sapporo historical data. It appears that the rainfall is poorly correlated at the 48 hour period. Thus, no synthetic records were generated for these combinations. The June synthetic time series of this region were also ignored since the rainfall during this period appeared to be very little and poorly correlated as well.

Although the proposed methodology led to synthetic time series that closely matched the moments of the historical records, a few combinations appeared to yield synthetic maxima that were closely matched to the historical maxima consistently. However, these ideal combinations did not appear to be consistent for all the regions involved. The reason for this appears to be the varied correlation structure of each regional historical record.

Within a region however, certain combinations seemed to model the maxima better than other combinations. It is thus possible that a dominant correlation structure of the historical record exists

Table 5. Kolmogorov-Smirnoff Tests for synthetic and historical maxima.

KAMISHIIBA	JUNE		JULY		AUGUST		SEPTEMBER		OCTOBER	
TEST	HOURLY	DAILY	HOURLY	DAILY	HOURLY	DAILY	HOURLY	DAILY	HOURLY	DAILY
I	0.998	0.998	0.995	0.999	0.624	0.864	0.999	0.669	0.997	0.982
II	0.994	0.996	0.999	0.999	0.981	0.945	0.884	0.990	0.884	0.506
III	0.884	0.996	0.999	0.997	0.449	0.903	0.994	0.550	0.995	0.999
IV	0.449	0.991	0.999	0.864	0.793	0.934	0.976	0.911	0.726	0.998
V	0.994	0.998	0.998	0.988	0.740	0.972	0.972	0.874	0.999	0.999
VI	0.968	0.999	0.999	0.994	0.683	0.754	0.976	0.286	0.999	0.948
NAHA	JUNE		JULY		AUGUST		SEPTEMBER		OCTOBER	
TEST	HOURLY	DAILY	HOURLY	DAILY	HOURLY	DAILY	HOURLY	DAILY	HOURLY	DAILY
I	0.844	0.126	0.743	0.601	0.999	0.976	0.999	0.999	0.536	0.965
II	0.993	0.241	0.957	0.980	0.999	0.999	0.999	0.928	0.630	0.993
III	0.949	0.481	0.994	0.979	0.999	0.998	0.999	0.972	0.354	0.966
IV	0.982	0.196	0.957	0.999	0.991	0.856	0.996	0.959	0.920	0.579
V	0.936	0.697	0.992	0.816	0.999	0.994	0.999	0.959	0.927	0.723
VI	0.982	0.077	0.999	0.996	0.772	0.798	0.995	0.971	0.920	0.848
SAPPORO	JUNE		JULY		AUGUST		SEPTEMBER		OCTOBER	
TEST	HOURLY	DAILY	HOURLY	DAILY	HOURLY	DAILY	HOURLY	DAILY	HOURLY	DAILY
I	0.997	0.805	0.995	0.966	0.998	0.953	0.993	0.909	0.603	1.000
II	0.891	0.971	0.998	0.999	0.998	0.406	0.964	0.957	0.999	0.999
III	0.947	0.998	0.973	0.999	0.999	0.953	0.999	0.971	0.999	0.947
IV	-not applied-		-not applied-		-not applied-		-not applied-		-not applied-	
V	0.999	0.944	0.999	0.999	0.999	0.544	0.999	0.997	0.999	0.999
VI	-not applied-		-not applied-		-not applied-		-not applied-		-not applied-	

for all the months within a region. The ideal combinations appear to be Tests I, II, and V for Kamishiiba, Tests II and III for Naha, and Test III for Sapporo. However, the ideal combination of historical moments to parameterize a NSM that sufficiently and consistently model historical rainfall maxima is not straightforward. The systematic check of combinations conducted here is recommended in future applications.

5. CONCLUSION

The Neyman-Scott Poisson rectangular pulse rainfall model was applied to generate synthetic rainfall time series for the regions of Kamishiiba, Naha, and Sapporo in Japan. A majority of the resulting synthetic records appear to match the historical moments and maxima. However, the maxima in several months in which the rainfall was due to a mix of sources such as convective rainfall with typhoons and/or fronts were inadequately matched. A more detailed NSM formulation, incorporating two rain cell types with short and long expected durations, is expected to rectify the limitations of the current study.

The concept of an ideal test set for estimating NSM parameters that lead to well-matching synthetic maxima seems ill-advised as of this writing. Although all combinations appear to yield synthetic data with moments closely matching the historical moments, the same is not necessarily true for the case of the maxima. Since a historical data set will have a correlation structure that is more dominant in several aggregation levels, it is also possible to have more than one ideal combination. A systematic check of possible combinations, followed by a KS-test for each synthetic and historical maximum set is thus recommended for regions elsewhere.

ACKNOWLEDGMENTS

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

REFERENCES

1) Cox, D. and Isham, V.: Point processes, Chapman and Hall, London, 1980.
2) Cox, D. and Isham, V.: Stochastic spatial temporal models for rain. Stochastic methods in hydrology: rain, landforms and floods (Barndorff-Nielsen, O. *et al*, ed), World Scientific, 1998.
3) Rodriguez-Iturbe, I. *et al*.: Some models for rainfall based on stochastic point processes, *Proc. R. Soc. London A*, Vol. 410, pp. 269-288, 1987.
4) Burlando, P. and Rosso, R.: Stochastic models of temporal rainfall: Reproducibility, Estimation and Prediction of Extreme Events, *Stochastic hydrology and its use in water resources systems simulation and optimization* (Marco, J. *et al*. ed.), NATO ASI Series, Kluwer Academic Pub., 1993.
5) Cowpertwait, P.: A generalized point process for rainfall, *Proc. R. Soc. London A.*, Vol. 447, pp. 23-27, 1994.
6) Cowpertwait, P. *et al*.: Neyman-Scott modeling of rainfall time series: 1. Fitting procedures for hourly and daily data, 2. Regionalization and disaggregation procedures, *J. Hydrol.* Vol. 175 (1-4), pp. 17-46, pp. 47-65, 1996.
7) Calenda, G. and Napolitano, F. Parameter estimation of Neyman-Scott processes for temporal point rainfall simulation, *J. Hydrol.* Vol. 225, pp. 45-66, 1999.
8) Favre, A. C., *et al.* Unbiased Parameter Estimation of the Neyman-Scott Model for rainfall simulation with related confidence interval, *J. Hydrol.*, Vol. 286, pp. 168-178, 2004.
9) Press, W. H. *et al*.: Numerical Recipes in Fortran, 2nd Ed., Cambridge University Press, London, 1992.
10) Gentle, J.: Random number generation and Monte Carlo methods, Springer-Verlag, New York, 1998.

(Received September 30, 2006)