京都大学防災研究所年報　第 49 号 B　平成 18 年 4 月
Annuals of Disas. Prev. Res. Inst., Kyoto Univ., No. 49 B, 2006

Quantiles of the Neyman-Scott Rectangular Pulse Rainfall Model for Hydrologic Design

Carlo Arturo MONDONEDO*, Yasuto TACHIKAWA, and Kaoru TAKARA

***Graduate School of Urban and Environmental Engineering, Kyoto University**

Synopsis

Following the clustered Poisson point process approach, we applied a Neyman-Scott rainfall Model (NSM) to generate synthetic extreme value rainfall depths. The main motivation for this study is on the acceptability of synthetic quantile rainfall depths as a possible decision-making aid in hydrologic design in lieu of historical records of appropriate length. Unlike previous studies, the emphasis here is to access the ability of the NSM to preserve the historical quantile rainfall depths of 1-hour and 24-hour duration such that longer NSM rainfall records, although synthetic, can be used as reliable bases of quantile events in hydraulic structure design (i.e.: impounding structures, sewer systems, etc). The historical data used in this study were obtained from 16 yearly records (1988 to 2003) of hourly and daily rainfall taken from the Kamishiiba Observatory in Japan. Following some stationarity assumptions inherent to NSM, it is necessary to obtain parameter sets for each month. Specifically, on a monthly basis, five NSM parameters were obtained by optimizing (by the Levenberg-Marquardt method) an objective function based on historical moments (namely mean, variance, covariance lag 1, etc.) and equivalent NSM moments, which were expressions in terms of the target parameters. By limiting historical information to five moments, the parameter sets obtained for several months (as of this writing) yielded adequate hourly quantiles and poor daily quantiles. Based on the framework of the NSM, it may be necessary to supply more historical moments (hence more NSM equations in the objective function) so that the five NSM parameters can capture more of the properties of the historical data. The determination of this ideal set of historical moments is the current pursuit of the authors and will appear in sequels of this paper.

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

1. Introduction

In synthetic rainfall generation, several stochastic methods are often used with Monte Carlo simulation in the event when recorded data is insufficient for effective analysis. Such synthetic data can be used in design storm evaluation for small impoundment structures and sewer systems. Detailed quantile design storm events of short and long duration can also be used in the preliminary evaluation of design conditions for large flood control structures such as dams and levees. The generation of such synthetic rainfall records is thus an effective decision-making aid to water resources engineers.

For such purposes, several families of stochastic models are available such as the Poisson Marks Model (PMM) (Rodriguez-Iturbe *et al.*

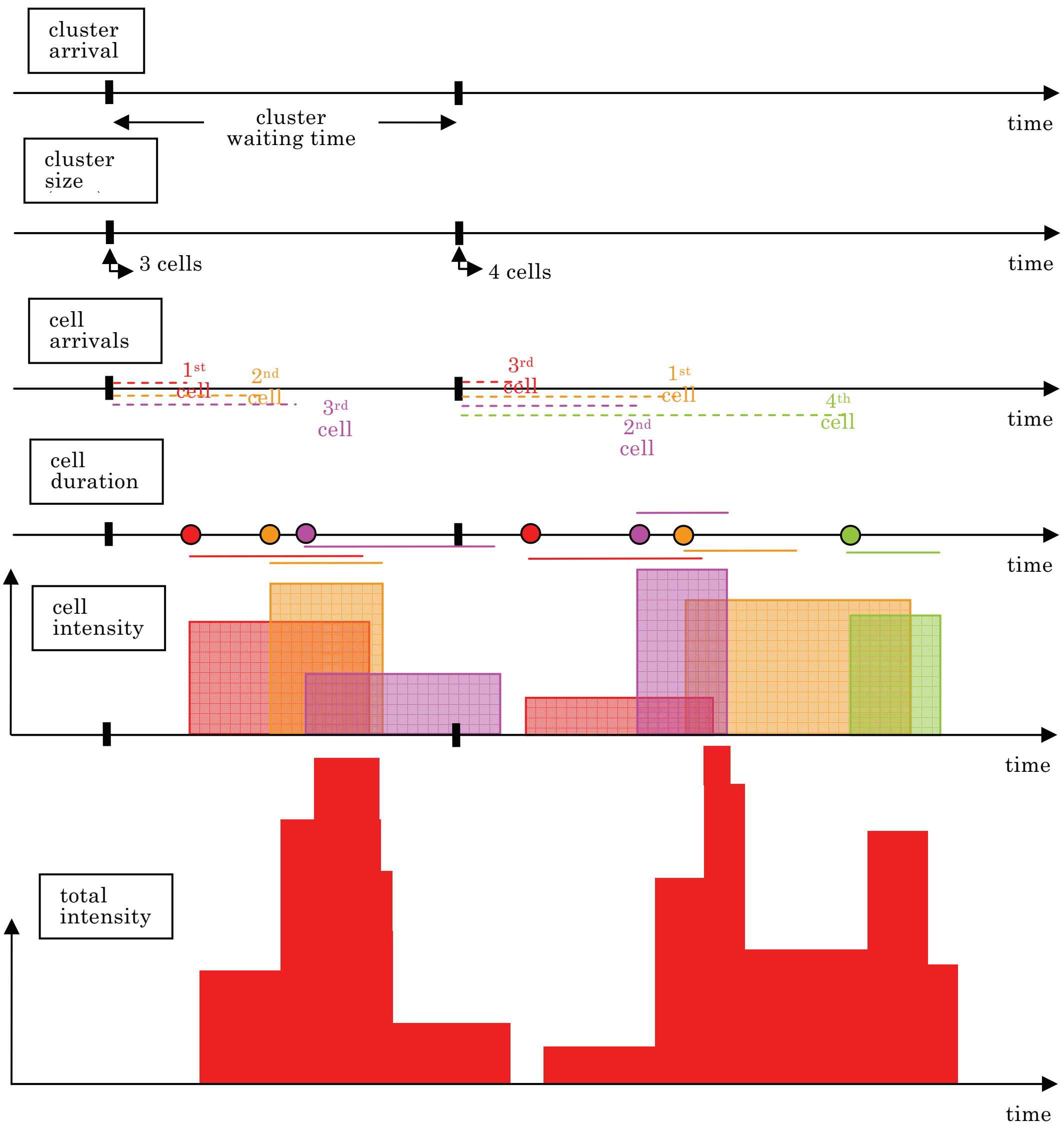

Fig 1. Schematic diagram of the Neyman-Scott Model.

(1987)), Poisson Rectangular Pulse Models (PRPM) (Rodriguez-Iturbe *et al.* (1987)), and Clustered Poisson Rectangular Pulse Models (CPRPM) (Rodriguez-Iturbe *et al.* (1987) and Burlando and Rosso (1993)). A model from the latter type was the key technique used in this study.

All previously mentioned models were based on the theory of point processes (Cox and Isham, 1980). Under this theory, a set of probabilities can be mapped to random occurrences of point events, such as rainfall. Both the PMM and PRPM follow Poisson occurrences. Under each model, rainfall arrival, intensity, and duration are random variables. However, the PMM models rainfall with no possibility of overlap while the PRPM permits overlap of occurrences. Applications of the models have appeared in the literature (see Burlando and Rosso (1993) for an example) with the shortcoming that the models developed could not be consistent with more than one aggregation time period.

A Clustered Poisson Rectangular Pulse Rainfall Model (CPRPRM) is a special point process in which rainfall is generated from a system of cell clusters of random size, and arrival time. Upon the arrival of each cluster, rain cells of random birth, intensity, and duration are generated to produce the total effective rainfall intensity by superposition. The Neyman-Scott model (NSM) is a CPRPRM in which rain cells arrive subsequent to the arrival of a cell cluster's origin. It was used in this study due to its consistency with more than one aggregation level (Rodriguez-Iturbe *et al.*, 1987).

Based on the method of moments, the NSM parameters were estimated by optimization. The NSM equations were then supplied with random numbers (uniform deviates) in a Monte-Carlo fashion to generate synthetic rainfall data yielding first order and second order moments (mean, standard deviation, autocorrelation coefficient order 1, etc.) similar to those of the original (historical) data. These NSM synthetic records, in the form of rainfall depth values were then analyzed for quantile-based events. The main motivation for this study was on the acceptability of synthetic quantile rainfall depths as a possible decision-making aid in hydrologic design in lieu of historical records of appropriate length.

Unlike previous studies, the emphasis here was the ability of the NSM to preserve the historical quantile rainfall depths of 1-hour and 24-hour duration such that longer NSM rainfall records, although synthetic, can be used as reliable bases of quantile events in hydraulic structure design (i.e.: impounding structures, sewer systems, etc). The historical data used in this study were obtained from 16 yearly records (1988 to 2003) of hourly and daily rainfall taken from the Kamishiiba Observatory in Japan.

2. NSM Rainfall Time Series Generation Method

Fig. 1 show the random processes involved in the concept of the Neyman-Scott model. This model consists of essentially five probability distributions. In this NSM, clusters of cells are linked integrally to a storm origin with mean occurrence rate λ, regarded as a Poisson process, where waiting times are exponential in λ. The arrivals of these clusters are shown in the first time line of Fig. 1. Each storm can have a random number of cells described by a geometric distribution, as shown in the second time line. The arrival of each cell is based on an exponential distribution, as shown in the third time line. Each cell has a corresponding independent identically distributed (iid) random intensity and duration, characterized by the exponential distribution, shown in the fourth and fifth time line, respectively. The total rainfall intensity is then the superposition of the effects of these random cell intensities, as shown in the sixth time line.

Table 1. Parameters of Neyman-Scott Model

SYMBOL	NAME	DISTRIBUTION
λ	mean arrival rate of a storm	Poisson
μ_c	mean number of cells in a	Geometric
μ_x	mean intensity of a cell	Exponential
$1/\beta$	mean displacement	Exponential
$1/\delta$	mean cell life span	Exponential

A succinct representation of the previously mentioned distributions can be written as:

$$p(N = n) = \frac{v^n e^{-v}}{n!} \tag{1}$$

$$f(t_s) = 1/\lambda \exp(-\lambda t_s) \tag{2}$$

$$p[C = c] = \frac{(1 - 1/\mu_c)^{c-1}}{\mu_c} \tag{3}$$

$$f(t_d) = \beta \exp(-\beta t_d) \tag{4}$$

$$f(i_c) = 1/\mu_x \exp(-1/\mu_x i_c) \tag{5}$$

$$f(t_c) = \delta \exp(-\delta t_c) \tag{6}$$

where:

v = mean number of occurrences
= λT; T is the time period in consideration

$p[N=n]$ = probability that the number of clusters N is equal to n

t_s = storm arrival time

λ = mean arrival rate of a storm

$f(t_s)$ = probability that the arrival of a storm origin is t_s

$p[C=c]$ = probability that the number of cells of a storm C is equal to c

μ_c = mean number of cells in a storm

$f(t_d)$ = probability that the arrival of a cell from the storm origin is t_d

$1/\beta$ = mean displacement of a cell from the storm origin

$f(i_c)$ = probability that the intensity of a cell is equal to i_c

μ_x = mean intensity of a cell

$f(t_c)$ = probability that the duration of a cell's life is equal to t_c

$1/\delta$ = mean cell life span

It is from the method of moments that the NS parameters can be linked to actual record data moments. Such expressions include the following (Rodriguez-Iturbe, 1987):

$$E\left\langle Y_i^{(h)}\right\rangle = \lambda(\mu_c)h/(\delta\mu_x) \quad (7)$$

$$\mathrm{var}\left\langle Y_i^{(h)}\right\rangle = \frac{\lambda(\mu_c{}^2-1)[\beta^3 A_1(h)-\delta^3 B_1(h)]}{\beta\mu_x{}^2\delta^3(\beta^2-\delta^2)} + \frac{4\lambda\mu_c A_1(h)}{\mu_x{}^2\delta^3} \quad (8)$$

$$\mathrm{cov}\left(Y_i^{(h)}, Y_{i+k}^{(h)}\right) = \frac{4\lambda\mu_c A_2(h,k)}{\mu_x{}^2\delta^3} + \frac{\lambda(\mu_c{}^2-1)[\beta^3 A_2(h,k)-\delta^3 B_2(h,k)]}{\beta\mu_x{}^2\delta^3(\beta^2-\delta^2)} \quad (9)$$

in which:

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

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

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

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

where:

i = time interval counter

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

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

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

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

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

3. Parameter Estimation

Five parameters are required in the NSM: λ, δ, μ_c, μ_x, and β. As shown previously, these parameters are directly connected to sample moments of the rainfall records in the form of equations (7)-(9). Several nontrivial combinations of these equations are available in the literature of the NSM.

Normally, these systems are solved for the required parameters by unconstrained minimization of an objective function. Such combinations include those used in the studies of Rodriguez-Iturbe *et al.* (1987), Burlando and Rosso (1993), Cowpertwait *et al.* (1996), Calenda and Napolitano (1999), and Favre *et al.* (2002), to cite a few. These combinations range from the most basic (5 equations in the objective function to solve for the 5 parameters are used) to the more thorough (more than 5 equations in the objective function to solve for the 5 parameters are used). Only the former combinations were considered for the preliminary output of this study. Thus, as of this writing, the determination the five parameters of the NS model included the following equations:

1. hourly mean of rainfall depth ((7) cast in h = 1 hr)
2. variance of hourly rainfall depth ((8) cast in h = 1 hr)
3. variance of daily rainfall depth ((8) cast in h = 24 hrs)
4. lag-1 covariance of hourly rainfall depth ((9) cast in h = 1 hr, and lag k = 1)
5. lag-1 covariance of daily rainfall depth ((9) cast in h = 24 hr, and lag k = 1)

In succeeding phases of this study, it may be necessary to include more moments of the records such as variances at different levels of aggregation, lag-2 covariances, the probability of dry time intervals, and transition probabilities (Cowpertwait ,1996). The expressions of these moments are deferred to future editions of this report for brevity.

The objective function used in the estimation follows the form:

$$\mathrm{F} = \sum_{j=1}^{5} \left(\frac{f_{\mathrm{j}}(Y_i)}{W_j} - 1 \right)^2 \tag{10}$$

where:

$f_{\mathrm{j}}(Y_i)$ = jth NS moment equation (items 1-5 previously mentioned) 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)). Weight factors can also be included for each term to integrate precedence of the modeled moment. For instance, Cowpertwait (1996) used a weight factor of 100 for the mean rainfall depth term of F (in (7)).

Table 2. NSM Parameters for Kamishiiba rainfall.

Month	λ	μ_c	μ_x	η	β
	1/hr		mm/hr	1/hr	1/hr
1	0.0112	5.20	1.208	0.708	0.168
2	Not evaluated (as of this preliminary stage)				
3					
4					
5					
6					
7					
8	0.00426	16.54	0.511	6.306	0.0658
9	0.00399	50.30	3.539	1.771	0.0723
10	0.00366	8.46	3.802	0.705	0.0436
11	0.00504	19.77	1.904	1.863	0.126
12	0.00466	2.58	2.055	0.427	0.0728

Following some stationarity assumptions inherent to NSM, it was necessary to obtain specific parameter sets of the NSM for each month. Such practice is now considered standard as homogeneity and stationarity assumptions seem to hold only in periods within seasons or months rather than years (Cox and Isham, 1998). The objective function was then minimized for each month of the 16-year long rainfall record of the Kamishiiba Observatory in Japan. The search algorithm used was the Levenberg-Marquardt method, a gradient based method following the framework of Gauss-Newton method with a Jacobian estimate for the Hessian Matrix. The results of the estimation are shown in Table 2.

4. NSM Model Implementation

To use the NSM parameters, a uniform deviate generator (random numbers within (0,1)) was developed from the method described in Press *et al.* (1987) which made use of the Park-Miller "Minimal Standard" generator based on the simple multiplicative congruential algorithm:

$$I_{j+1} = \xi I_j \left(\mathrm{mod}\ \ m\right) \tag{11}$$

where:

ξ = multiplier = 7^5 = 16,807

m = modulus = $2^{31} - 1$ = 2,147,483,647

mod = modulus operator

I_j = previous random integer between 0 and m-1

I_{j+1} = succeeding random integer between 0 and m-1.

To generate the required random variables, a uniform deviate was used in the inverse of a required distribution's CDF. In the case of the exponential distribution, a continuous distribution for the rainfall intensity, position, and cell life, the solution was algebraic. On the other hand, the discrete Poisson and geometric distributions were handled by deriving an empirical CDF. The uniform deviates were then applied in a look-up table fashion. Facilitating the task of generating uniform deviates and random variables, tallying rainfall intensity time series and bookkeeping synthetic data moments was a simple FORTRAN program dubbed *"Rainmaker.exe"*.

Using this program, 16 synthetic rainfall record depths were generated for each target

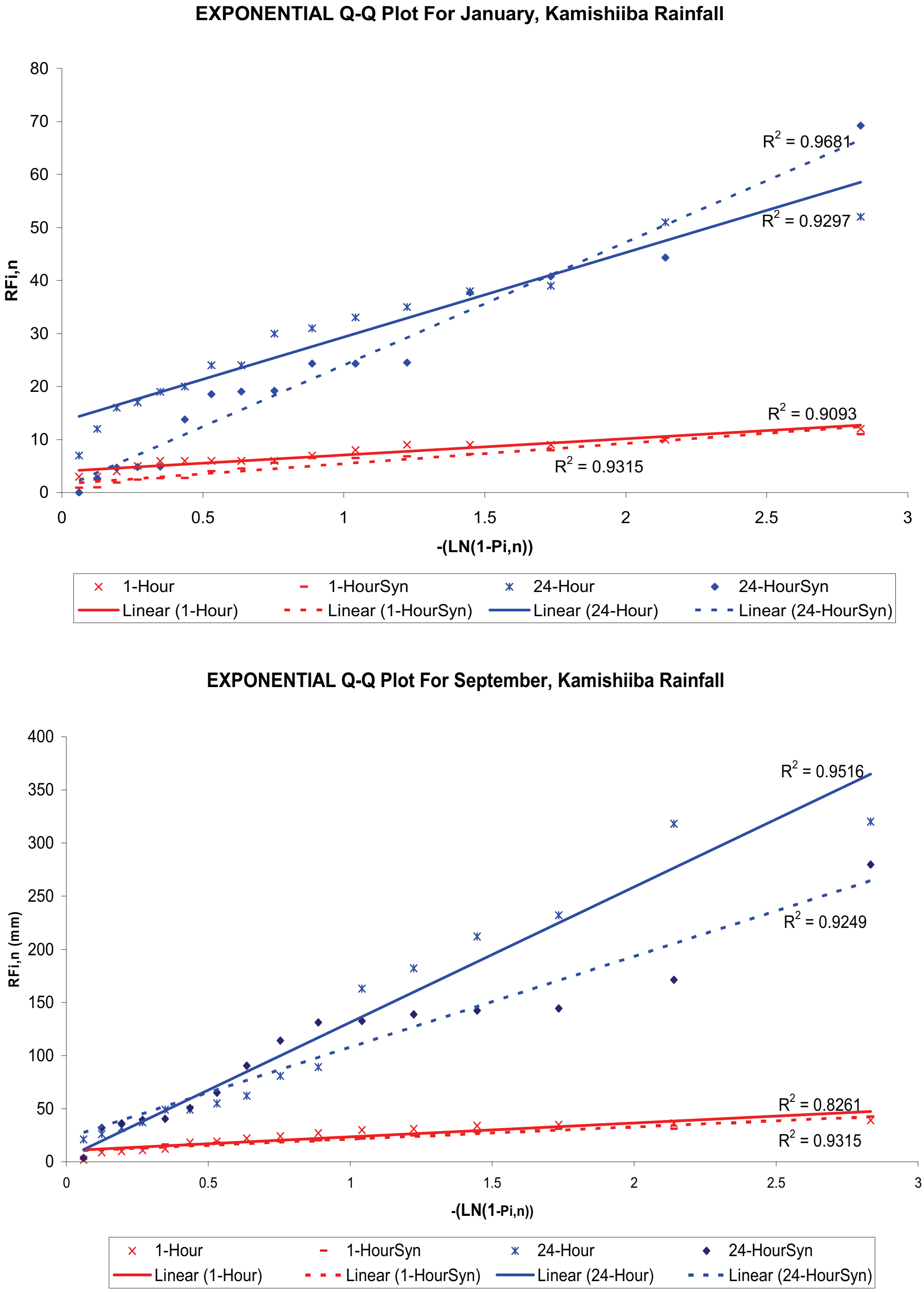

Fig. 2. Deviations in historical and synthetic rainfall quantiles from Kamishiiba Obsevatory data.

duration (1- and 24-hour). Each record was then searched to obtain a sample of synthetic maxima. Only several sets of parameters were available as of this writing. A brief analysis of these maxima appears in the succeeding section.

5. Preliminary Analysis

5.1. Neyman-Scott Parameters

Following the previous work of Calenda and Napolitano (1999), the Neyman-Scott Parameters obtained by the Levenberg-Marquardt method. Based on the mentioned study's framework, applications using 1-hour and 1-day target durations may be regarded as applications with separations large enough such that any minimization algorithm (if it converges) would yield an estimate irrelevant of starting values (no bias). To test this, for the month of January, a relatively coarse grid of the five parameters of NSM was constructed from a range of allowable values. The centroids of these 5x5 "volumes" were regarded as the starting point of a search. In this study, the range used was (see Napolitano and Calenda, (1999) is shown in Table 3.

Dividing each range by two, a set of 32 trial centroids was obtained. Not all these initial values yielded answers, indicating the complexity inherent to changing the historical components of the objective function (the W_j terms in equation 10). In cases when a solution was obtained, the solutions were practically indistinguishable from those of other centroids, indicating weak or no bias, as cited previously (Napolitano and Calenda, 1999).

Table 3. Range of NSM parameters for optimization.

Parameter	Min	Max
λ (1/h)	0.001	0.050
μ_c	2.0	100.0
b (1/h)	0.01	0.50
μ_c (mm/h)	0.30	15.0
η (1/h)	0.10	5.0

5.2. QQ Plot Representation

For a preliminary test of synthetic data, the quantile-quantile plot (QQ plot) of the historical and synthetic rainfall for January and September are shown in Fig. 2. The vertical coordinates of each point is based on the maximum of each monthly record while the horizontal coordinates are the corresponding quantile of each maximum. The calculation of quantiles is based upon a ranking of each monthly maximum (from smallest to largest), whereby a probability of exceedance was assigned via the equation:

$$EP_i = \frac{i}{N_{tot} + 1} \tag{12}$$

where:

i = ith largest maxima of monthly record

EP_i = exceedance probability of ith largest maximum

N_{tot} = total number of maximums in the monthly record

Under this *Weibull plotting position*, a larger maximum would receive a lower probability of exceedance and vice-versa. Other plotting position equations are possible and may appear in the sequel of this report (such as Cunane's plotting position).

Upon obtaining the exceedance probability of each maximum, the quantile is calculated as the inverse of the survival function. In this case, the exponential distribution's CDF yields:

$$Q_i = -Ln(1 - EP_i) \tag{13}$$

where:

Ln(a) = natural logarithm of a

Q_i = exponential quantile value

For the quantile-analyzing purpose of this study, the advantage of using a QQ plot in general stems from its inherent *linearity*. In the case when an extreme value larger than the largest maximum of the sample is in question, this linearity gives a means to quantify its quantile, and hence its probability of exceedance.

Note that this QQ plot yields a linear least-squares fitting problem. This indicates that the *goodness of fit* of a presumed model (exponential, Pareto, Weibull, etc.), can be easily checked visually and quantified by means of the linear correlation coefficient. In sequels of this study, a means to include the confidence interval of a presumed model will be included in these QQ plots (possibly by block bootstrap methods).

The hourly quantiles are quite close in distribution, evidenced by the apparent overlaying of the fitted straight line equations of the historical and synthetic data. This may indicate that more standard tests such as the Komolgorov-Smirnoff Test will show that the two samples are essentially drawn from the same population. Proper plots of

Table 4. Proposed test sets for NSM parameter estimation problem.

Test Set I	Hours of aggregation to be used				
	1	**6**	**12**	**24**	**48**
Mean	○				
Variance	○			○	○
Covariance	○			○	○
Test Set II	**Hours of aggregation to be used**				
	1	**6**	**12**	**24**	**48**
Mean	○				
Variance	○		○	○	
Covariance	○		○	○	
Dry Interval Probability				○	
Test Set III	**Hours of aggregation to be used**				
	1	**6**	**12**	**24**	**48**
Mean	○				
Variance	○	○	○	○	○
Covariance	○			○	
Dry Interval Probability				○	
Test Set IV	**Hours of aggregation to be used**				
	1	**6**	**12**	**24**	**48**
Mean	○				
Variance	○			○	
Covariance	○	○	○	○	○
Dry Interval Probability				○	

KS tests are deferred to sequels of this study.

For preliminary purposes though, the high correlation coefficients of the lines of best fit for the synthetic and historical data suggest that the exponential CDF is a good candidate distribution for the hourly rainfall maxima. Essentially, the parameters determined can yield synthetic data with maximum rainfall depths at 1 hour similar to the historical data. It would be sensible to check larger records of synthetic data for acceptability as well (see Section 6).

However, the lines of best fit of historical daily and synthetic daily maxima shows a drastic deviation. It may be possible that the NSM parameters estimated from the previously cited set of equations (see Section 3) may not be appropriate enough for modeling the daily maxima. In this case, it may be necessary to include intermediate moments (6-hour variance, 12-hour variance, 6-hour covariance lag 1, 12-hour covariance lag 1, etc.) as well as special daily dry probabilities in the search for the NSM parameters.

6. Further Considerations

Current results indicate that the selected set of historical moments is inadequate to model daily maxima. It is therefore a primary concern to find this ideal set of historical moments. However, the current equations used were based primarily on rainfall occurrences. The dry probability of the NSM will also be used in the sequel (Cowpertwait, 1996).

$$PD(h) = \exp\left\{-\lambda h - \lambda \int_0^{\infty} [1 - p_t(h)]dt\right\} \tag{14}$$

$$p_t(h) = \left(1 - e^{-\beta t} + e^{-\beta(t+h)}\right)\left(1 - \beta \frac{e^{-\beta t} - e^{-\eta t}}{\eta - \beta}\right) \exp\left\{(\mu_c - 1)\left(e^{-\beta(t+h)} - e^{-\beta t} - \beta \frac{e^{-\beta t} - e^{-\eta t}}{\eta - \beta}\right)\right\} \tag{15}$$

where:

PD(h) = probability that the duration h within a series is dry.

Table 4 presents several tests to be used for further studies. Test Set I examines the case of using the original test set as well as the 48-hour moments in modeling of the daily maxima. Only this test does not make use of the dry interval probability given by eqn. (14) and eqn. (15). Test Set II modifies the original set by including the 12-hour data and the daily dry interval probability, or DIP (total number of days dry over total number of days). Test Set III is a variance test examining the possibility that due emphasis to variance, along with the original historical moments will yield better parameter estimates. Test IV is a similarly themed test set focused on covariances.

References

Burlando, P., Rosso, R. (1993): *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.

Calenda, G. and Napolitano, F. (1999): Parameter estimation of Neyman-Scott processes for temporal point rainfall simulation. *Journal of Hydrology* Vol. 225 pp 45-66.

Cowpertwait, P. *et al* (1996): Neyman-Scott modeling of rainfall time series: 1. Fitting procedures for hourly and daily data. 2. Regionalization and disaggregation procedures. *Journal of Hydrology* Vol. 175 (1-4), 17-46, 47-65.

Cox, D. and Isham, V. (1980): Point processes. Chapman and Hall, London.

Cox, D. and Isham, V. (1998): *Stochastic spatial temporal models for rain*. Stochastic methods in hydrology: rain, landforms and floods. (Barndorff-Nielsen, O. *et al*, ed). World Scientific.

Favre, A. C., *et al* (2004): Unbiased Parameter Estimation of the Neyman-Scott Model for rainfall simulation with related confidence interval. *Journal of Hydrology* Vol. 286, pp. 168-178.

Press, W. H. *et al*. (1986): Numerical Recipes in Fortran. Cambridge University Press, London.

Rodriguez-Iturbe, I. *et al*. (1987): Some models for rainfall based on stochastic point processes. *Proc. R. Soc. London* A 410, 269-288.

水工計画のためのノイマン－スコット型矩形パルス降雨発生モデルによる確率水文量の発生

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

Clustered Poisson Point Process 理論を用いたNeyman-Scott Rainfall生成モデル(NSM)を用いて降雨時系列を確率的に発生させる。本研究の動機は，実際のデータから得られる確率水文量を反映させることができる降雨発生モデルを構築することにある。ここでは水工構造物の設計規模を決める基礎となる年最大時間降水量と年最大24時間降水量に着目し，NSMがから得られる確率水文量と実際のデータから得られる確率水文量との違いを分析する。九州の上椎葉アメダス観測所の16年分(1988から2003年)の降雨データ用いて，NSMの5つのモデルパラメータを月毎に決定した。そのNSMを用いて時間降水量データを発生させ時間降水量と日降水量の確率水文量を算定し，観測時系列データから得られる確率水文量と比較したところ，時間降水量に関しては確率水文量の発生特性を満たすが，日降水量に関しては十分でないことがわかった。これにより，NSMによる降雨発生を決定付けるパラメータ決定において，さらに異なるモーメント特性を導入してモデルパラメータを決定する必要があることがわかった。

キーワード：確率水文学，降雨モデル，基本高水，時系列発