QUANTILE STREAMFLOW ESTIMATES BASED ON THE NEYMAN-SCOTT RAINFALL MODEL

Carlo Mondonedo¹, Yasuto Tachikawa², and Kaoru Takara³

¹ Graduate Student, Department of Urban and Environmental Engineering, Kyoto University
Nishikyo-ku, Kyoto, 615-8540, Japan, e-mail: mondonedo@flood.dpri.kyoto-u.ac.jp
² Associate Professor, Department of Urban and Environmental Engineering, Kyoto University
Nishikyo-ku, Kyoto, 615-8540, Japan, e-mail: tachikawa@mbox.kudpc.kyoto-u.ac.jp
³ Professor, Disaster Prevention Research Institute, Kyoto University
Uji-shi, Kyoto, 611-0011, Japan, e-mail: takara@flood.dpri.kyoto-u.ac.jp

ABSTRACT

Flood control decisions in areas that are poorly gauged can be based on synthetic data if and only if such data can be made consistent to the historical counterpart. However, established methods that do not involve synthetic data are preferred for this problem such as the procedure adopted by the Japan Ministry of Land, Infrastructure, Transport, and Tourism (MLIT). Here, we develop an alternative methodology based on the Neyman-Scott clustered Poisson rectangular pulse rainfall model (NSM) suited for rainfall extreme value consistency. It is shown that simulated streamflow based on the NSM rainfall is a competent estimate to extreme flood magnitudes. We observe that the latter method is more rational due to its careful emphasis on extreme value rainfall.

Keywords: Synthetic Rainfall, Synthetic Streamflow, Point Processes, Neyman-Scott model

1. INTRODUCTION

Historical streamflow records are the prime basis of effective flood control decisions such that a safe coexistence can be established between river basin and inhabitants. However, the lack of recorded data in some river basins complicates the determination of design flooding events. Fortunately, methodologies are available to systematically solve this problem. The approach we adopt in this study incorporates techniques that simulate the transformation of rainfall into streamflow in the form of distributed hydrological modeling (Beven, 2002). More importantly, we couple this tool with an enhanced model for generating synthetic rainfall from scarce historical rainfall records based on the Neyman-Scott clustered Poisson rectangular pulse rainfall model, or NSM (Rodriguez-Iturbe et al., 1987). Consistency of NSM synthetic extreme rainfall to those of historical counterpart has been explored in the past (Cowpertwait, 1998). We have also contributed to this problem by developing the so-called NSM Fano factor exponent, or FFE (see Sec 2). Emphasis on consistency in rainfall extremes leads to a sufficient synthetic basis that is sound, reliable and economical. Only then can one base design flood evaluation on a synthetic technique such that the lack of historical data is no longer a hindering factor in the decision-making process.

Our interest here is in showing the advantage of synthetic rainfall generation in evaluating design streamflow by comparing two associated methodologies. One method excluding synthetic rainfall generation is based on the Japan Ministry of Land, Infrastructure, Transport, and Tourism (MLIT). Another method is developed by the authors following the NSM framework. Both methods involve distributed hydrological modeling based on Kyoto University’s Object Oriented Hydrologic Modeling System (OHyMOS).
2. METHODOLOGY

2.1 The NSM and its governing equations

The NSM is a clustered Poisson point process (Sefozo, 1990) in which: a) storms arrive following a temporal Poisson process (mean recurrence rate $\lambda$), b) storms consist of a geometric random number of rain cells, no storm containing zero cells (mean cell number $\mu_c$), c) each cell arrival relative to the storm arrival follows an exponential distribution (mean lag time $1/\beta$), d) each cell duration follows an exponential duration (mean duration $1/\eta$), and e) each cell intensity follows a two-parameter gamma distribution (shape parameter $\alpha$ and scale parameter $\theta$). The superposition of these rain cell pulses in the rainfall intensity-time plane results in the target synthetic rainfall. Six parameters are therefore required to tune the NSM for a particular application. The following T-duration aggregated moments were derived for parameter estimation in the NSM (Rodriguez-Iturbe et al., 1987 and Cowpertwait, 1998).

\[ E[Y] = \frac{\lambda \mu_c \alpha \theta}{\eta} T \]

\[ \text{Var}[Y] = \frac{4 \lambda \mu_c (\alpha \theta)^2 A_1(T) + 2 \lambda \mu_c \alpha \theta^2 (1 + \alpha)(\mu_c - 1) \left[\beta^2 A_1(T) - \eta^2 B_1(T)\right]}{\frac{\eta^3}{\beta^3 (\beta^2 - \eta^2)} + \frac{2 \lambda \mu_c \alpha \theta^2}{\beta^3 (\beta^2 - \eta^2)}} \]

\[ A_1(T) = \eta T - 1 + \exp(-\eta T); \quad B_1(T) = \beta T - 1 + \exp(-\beta T) \]

\[ \text{Cov}[Y_T, Y_{T+k}] = \frac{4 \lambda \mu_c (\alpha \theta)^2 A_2(T, k) + 2 \lambda \mu_c \alpha \theta^2 (1 + \alpha)(\mu_c - 1) \left[\beta^2 A_2(T, k) - \eta^2 B_2(T, k)\right]}{\frac{\eta^3}{\beta^3 (\beta^2 - \eta^2)}} \]

\[ A_2(T, k, \eta) = \frac{1}{2} (1 - \eta T)^2 \exp\left(\eta T - k - 1\right); \quad B_2(T, k, \beta) = \frac{1}{2} (1 - \beta T)^2 \exp\left(-\beta T - k - 1\right) \]

\[ \text{Cor}[Y_T, Y_{T+k}] = \frac{\text{Cov}[Y_T, Y_{T+k}]}{\text{Var}[Y]} \]

\[ \text{TCM}[Y] = \frac{6 \lambda \mu_c E[I^2] (\eta T - 2 + \eta T e^{-\eta T} + 2 e^{-\eta T})}{\eta^4} + \frac{3 \lambda \mu_c E[I] E[C(C-1)] f(\eta, \beta, T)}{2 \eta^4 \beta^2 (\beta^2 - \eta^2)} \]

\[ + \frac{\lambda (E[I])^2 E[C(C-1)(C-2)] g(\eta, \beta, T)}{2 \eta^4 \beta (\beta^2 - \eta^2) (2 + \eta) (2 + \beta)} \]

\[ E[I] = \alpha \theta; \quad E[I^2] = (1 + \alpha) \theta^2; \quad E[I^3] = \alpha^3 \frac{\Gamma(3 + \alpha)}{\Gamma(\alpha)} \]

\[ E[C(C-1)] = 2 \mu_c (\mu_c - 1); \quad E[C(C-1)(C-2)] = 6 \mu_c (\mu_c - 1)^2 \]

\[ f(\eta, \beta, T) = -2 \eta^2 \beta^2 e^{-\eta T} - 2 \eta^2 \beta^2 e^{-\beta T} + 2 \eta^2 \beta^3 e^{-2\eta T} + 2 \eta^4 \beta e^{-\eta T} + 2 \eta^3 \beta^2 e^{-(\eta + \beta) T} - 2 \eta^4 \beta e^{-(\eta + \beta) T} - 2 \eta^3 \beta^2 e^{-2\eta T} - 2 \eta^4 \beta e^{-2\eta T} - 8 \eta^3 \beta^3 T + 11 \eta^2 \beta^3 - 2 \eta^4 \beta T + 4 \eta^3 \beta^2 T + 4 \eta^2 \beta^3 T - 7 \beta^4 - 4 \eta^5 + 8 \eta^5 e^{-\eta T} - \beta^5 e^{-2\eta T} - 2 T \eta^3 \beta^3 e^{-\eta T} - 12 \eta^2 \beta^3 e^{-\eta T} + 2 T \eta \beta^5 e^{-\eta T} + 4 \eta^5 e^{-\eta T} \]
\[ g(\eta, \beta, T) = 12\eta^3\beta T e^{-\eta T} + 9\eta^4\beta^2 + 12\eta^5\beta e^{-\eta T} + 9\eta^6\beta^4 + 12\eta^7\beta^3 e^{-(\eta+\beta)T} - \eta^3\beta^4 e^{-2\eta T} \\
- 9\eta^4\beta - 12\eta^5\beta^3 e^{-\eta T} - 9\beta^5 \eta - 3\eta^5\beta e^{-2\eta T} - \eta^4\beta^2 e^{-2\eta T} - 12\eta^3\beta^3 e^{-\eta T} - 3\beta^5 \eta e^{-2\eta T} \\
- 10\beta^4 \eta^3 + 6\beta^5 \eta^2 T - 10\beta^5 \eta T - 8\beta^5 \eta^4 e^{-\eta T} + 4T\eta^5\beta + 12\eta^3\beta^3 - 8\beta^5 \eta^4 e^{-\eta T} \\
- 6\eta^6 - 6\beta^6 - 2\eta^6 e^{-2\eta T} - 2\beta^6 e^{-2\eta T} + 8\beta^6 e^{-\eta T} + 8\eta^6 e^{-\beta T} + 6\beta^2 \eta^5 T \]

where \( E[Y] \) is mean rainfall depth, \( \text{Var}[Y] \) is the rainfall depth variance, \( \text{Cov}[Y_T, Y_{T+k}] \) is the rainfall depth autocovariance at lag-\( k \), \( \text{Cov}[Y_T, Y_{T+k}] \) is the rainfall depth autocorrelation at lag-\( k \), \( \text{TCM}[Y] \) = third central moment (TCM) of rainfall depth.

### 2.2 The Fano factor exponent of the NSM

We also include the expression for the Fano factor exponent (FFE) of the NSM (Mondonedo et al., 2008) in the parameter search. The derivation of the Fano Factor exponent expression starts with the Peaks Over Threshold (POT) rainfall point process, shown in Figure 1. In this case, historical hourly data pooled monthly \( U_i(T) \) is determined using the threshold \( z_i \), taken as the smallest monthly maximum rainfall (Figure 1a). The POT rainfall \( Q_i(T) \), the corresponding hourly historical rainfall greater than or equal to this threshold (Figure 1b), is given a unit count \( B_i(T) \) (Figure 1c). The point process \( N_i(T^*) \) (Figure 1d) is the sum of the unit counts within adjacent non-overlapping windows of size \( T^* \). \( N_i(T^*) \) is our basis for the FFE.

![Figure 1. Determining the Peaks Over Threshold (POT) rainfall point process and counting process \( N_i(T^*) \).](image)

(a) Sample \( U_i(T) \)

(b) Sample \( Q_i(T) \)

(c) Sample \( B_i(T) \)

(d) Sample \( N_i(T^* = 20 \text{ hours}) \)

The Fano factor (Fano, 1947) is defined as the ratio of the variance of \( N_i(T^*) \) and its mean, or:
\[
\text{FF}_i(T^*) = \frac{E\left[N_i(T^*)^2\right] - E\left[N_i(T^*)\right]^2}{E\left[N_i(T^*)\right]}
\]

where:

\[E\left\{\right\} \text{ is operation to obtain expected value.}\]

When the Fano factor is evaluated for the set of windows \( W = \{2, 10, 20, \ldots 100 \text{ hours}\} \) we can observe the typical plot shown in Figure 2. A power law relationship was proposed for this scaling behavior by Lowen and Teich (1995, 2005) as well as Telesca et al. (2007), shown here as Eq. 6.

\[
\text{FF}(T^*) = 1 + \left(\frac{T^*}{T_0}\right)^\xi
\]

Figure 2 Scaling in the counting process \( N_i(T^*) \) obtained from Kamishiiba POT series \( Q_i(t) \) in June.

where \( \xi \) is the Fano factor exponent, or the FFE. Alternatively, we assume that the average of the point process \( N_i(T^*) \) has as a power law relationship with respect to \( T^* \), while the average can be taken as a linear one, then the following approximation for the overall FFE scaling behavior holds:

\[
\text{Var}\left\{N_i(T^*)\right\} = A_i T^{b_i}
\]

(7)

\[
E\left\{N_i(T^*)\right\} = C_i T^*
\]

(8)

\[
\text{FF}_i(T^*) \approx \frac{A_i}{C_i} T^{b_i - 1}
\]

(9)

After curve fitting operations for Eqs. 7 and 8 (i.e.: linear regression of historical \( Q_i(T) \) for determining \( A_i, B_i, \) and \( C_i \)), we may use Eq. 9 to write the approximate Fano Factor of historical \( Q_i(T) \), \( \text{FF}_{Hi}(T_{Mi}) \), at window \( T^* = \text{mean storm duration} \ T_{Mi} \) as:

\[
\text{FF}_{Hi}(T_{Mi}) \approx \frac{A_i}{C_i} T^{b_i - 1}
\]

(10)

Similarly, the same Fano factor at mean storm duration \( \text{FF}_{Si}(T_{Mi}) \) can be written explicitly using Eq. 6 such that:
For our purposes, an ideal simulation should yield synthetic rainfall with $Q_i(T)$ such that historical and synthetic Fano factors at mean storm duration are equal, or based on Eqs. 10 and 11:

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

After some algebra, we may isolate the synthetic Fano factor exponent $\xi_{Si}$ shown here as Eq. 13:

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

The expression for $T_{M_i}$ in terms of the NSM parameters is based on Cowpertwait’s (1991):

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

where $\gamma$ is the Euler constant (0.577...). Substituting this approximation for $T_{M_i}$ in Eq. 13, we have:

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

Normally, plots similar to Figure 2 taken from the historical data have shown that the value of $T_0$ in Eq. 15 can be taken as 1 hour. In this case, our NSM FFE reduces to:

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

Moreover, using Eq. 6 to estimate the historical counterpart $\xi_{Hi}$, it is now possible to incorporate the FFE in the NSM parameter estimation (shown in the next subsection).
2.3 Parameter estimation of the NSM

The parameter estimation of the NSM is based on the solution of the following objective function:

\[ O(\Omega) = \left( 1 - \frac{\text{NSM}_i}{\text{HIS}_i} \right)^2 \]  

(17)

where \(\text{NSM}_i\) pertains to the component expression listed in Eqs. 1 – 5, and 16 while \(\text{HIS}_i\) pertains to the historical counterpart (see Appendix for historical counterpart of eq. (6)). Two combinations of components are used for parameter estimation for synthetic rainfall generation from the NSM. Scheme A is a configuration based on the TCM in which eq. (17) adopts the hourly mean, hourly variance, hourly autocorrelation at lag-1, hourly third central moment, 12-hourly autocorrelation at lag-1, 24-hourly variance, and 24-hourly autocorrelation at lag-1. Scheme B is essentially Scheme A, only that eq. (17) adopts the hourly FFE instead of the hourly TCM.

2.4 Description of adopted streamflow modeling method

Two distributed hydrologic models (DHM) were developed from OHyMOS (Ichikawa, 2000). One model was developed for the Kamo river basin located in Kyoto Prefecture while another one was developed for Kamishiiba river basin in Miyazaki Prefectures in Japan. Both study areas were less than 300 Km² in size for which the use of a single rain gauge is justified. Rainfall-runoff conversion is implemented following a kinematic wave model based on a function similar to a discharge-stage relationship (Tachikawa et al., 2007). Parameters for these conversion equations are assumed constant for all the calculation elements in the river basin although internal systems are available for changing these parameters per sub-basin. These parameters were estimated in separate studies (Tachikawa et al., 2007 for Kamo and Lee et al., 2007 for Kamishiiba). The simulations conducted with these tuned river basin models were set for monthly runs at 5 minute computation time with streamflow results displayed per hour.

3. EXPERIMENTAL SETUP

Methods for calculating extreme floods are based on two procedures. One procedure is based on the Japan Ministry of Land, Infrastructure, Transport and Tourism (MLIT) while another is based on the NSM synthetic rainfall. For stationarity considerations, both methods are applied for historical rainfall pooled monthly during June, assumed to be representative of the rainy season in all study regions.

3.1 MLIT method

A procedure for estimating q-return period design floods from historical rainfall and streamflow simulation based on the Japan Ministry of Land, Infrastructure, Transport, and Tourism (MLIT, 2008) is designated here as M-I. This procedure starts with assigning a design duration from which the quantile rainfall for all simulations is based. For river basin areas such as Kamo and Kamishiiba, each less than 300 Km² in size, this duration is specified as 24 hours. The corresponding q-return period quantile rainfall depth is then calculated
based on 24-hourly aggregated historical rainfall record. A 24-hourly distribution to this total magnitude is assigned by searching through the historical records for the corresponding 24-hourly maximum rainfall per month. These maximum storms are then proportionately modified such that the total rainfall within the 24-hour period is the basis quantile rainfall.

The MLIT then runs each 24-hourly storm in an appropriate model to simulate streamflow. There are thus as many resulting hydrographs from these simulations as there are historical rainfall records. Each hydrograph is then searched for its maximum value, yielding the estimate for the design flood, say $S_{mq}$. Consequently, M-I gives several q-return period design flood estimates based on the $S_{mq}$ of each simulated hydrograph.

3.2 Method based on NSM rainfall

A second streamflow generating procedure is designated here as M-II. We generate 100 synthetic records of the target month for our applications here where return periods are within the 100 year value. Each monthly record is then run through the appropriate river basin model for streamflow simulation.

The resulting hydrographs are searched for maximum streamflow, resulting in 100 values for flood frequency analysis. The required quantile event, say $Q_{mq}$, is then estimated from the cumulative distribution that best fits these empirical maxima (a log-normal distribution fitted by least squares is adopted although other distributions may be used). Two variants of this procedure correspond to the two NSM parameter estimation schemes of Sec. 2.3. M-IIA adopts NSM O($\Omega$) Scheme A while M-II adopts that of Scheme B for parameter generation.

4. RESULTING STREAMFLOW ESTIMATES

Historical streamflow data was limited in this study. Each historical June rainfall record is also run through the appropriate DHM and is considered a suitable substitute for historical streamflow. Resulting quantile estimates from the synthetic streamflow and pseudo-historical counterparts shown here are limited to the hourly duration. Quantile streamflow are estimated for the 10-, 20-, 30-, 50-, and 100-year return periods.

4.1 Kamo river basin streamflow

The quantile-quantile (q-q) plot of Kamo pseudo-historical hourly streamflow maxima ($Dis$) appears in Figures 3a-b. Each maximum streamflow value is given a plotting position $pp$ proportionate to its rank in the overall record ($pp=1/(i+1)$ in which $i$ is rank). The log-normal distribution quantiles are used in this figure such that the independent variable (related to the event return period) corresponds to the inverse of the standard normal distribution (Gaussian distribution with zero mean and variance of unity) $\Phi^{-1}$ of $pp$ while the dependent variable corresponds to the logarithm of $Dis$. Linear regression gives us the parameters for the fit of this log-normal model that leads to the 95% confidence bands.

We may then project the best fit line for the pseudo-historical streamflow (Figure 3a) to extrapolate the trend at return periods 10-, 20-, 30-, 50-, and 100-years. There is a tendency of M-I to give a wide range of estimates for these target quantile floods in Figure 3a. In fact, the variation of M-I estimates becomes wider along with increasing return period. There is thus a pronounced ambiguity in the quantile estimates of M-I, making it disadvantageous despite its simple approach of using historical data alone. In other words, it would be difficult to depend on M-I to quantify the quantile events given that we cannot justify which among the multiple estimates is the most likely value.
This ambiguity does not appear in M-II, shown in Figure 3b. Estimates appear to be quite consistent to the projections of pseudo-historical data. Scheme B results appear to be the more rational estimate since this scheme includes POT rainfall maxima information in the FFE. The advantage of using M-II, which involves synthetic rainfall generation, is therefore its clearer and unambiguous estimates of the quantile events.

4.2 Kamishiiba river basin streamflow

Similar q-q plots based on Kamishiiba results appear in Figures 4a-b. Not all quantile estimates generated from M-I are within reasonable proximity to what can be drawn from the pseudo-historical counterpart, as shown in the lower return periods (10-year and 20-year estimates), indicating poor performance. In fact, quantiles should be evaluated at higher return periods (i.e.: higher than 100 years) before the M-I method yields estimates with high variation that lie along the pseudo-historical 95% region. This however is not the ideal application since at times, one needs an estimate of lower return periods (i.e.: urban conditions/low priority flood protection works). The M-I scheme therefore generates poor estimates of the required quantiles in Kamishiiba River Basin.

Figure 4b shows the same estimates generated from M-IIA and M-IIB. Most quantile estimates are within the historical 95% region of historical quantiles, indicating better performance over the M-I estimates (of Figure 4a). In fact, both schemes perform appreciably well given that both yield almost the same low return period estimates and gradually diverge at higher return periods within the 95% historical region. Therefore, either adopting M-IIA or M-IIB yields reasonable quantile estimates.
5. CONCLUSION

A comparison between several methods was conducted that led to the advantage of using synthetic rainfall in the estimation of critical streamflow in river basins with limited historical rainfall and/or streamflow data. Synthetic rainfall was based on the Neyman-Scott clustered Poisson rectangular pulse rainfall model (NSM). The streamflow generated from modeling historical rainfall through a distributed hydrological modeling (DHM) was assumed as an equivalent to historical streamflow (referred to as pseudo-historical streamflow).

Results indicate that an established method from the Japan Ministry of Land, Infrastructure, Transport, and Tourism (MLIT) for estimating design floods from historical rainfall have several limitations. Estimates from this method (for the Kamo River Basin in Kyoto) vary widely for any return period due the use of multiple design rainfall that are each plausible occurrences of the quantile event. In one application (Kamiishiha River Basin in Miyazaki), results were in gross error for low return periods. Though this method involved only historical rainfall, estimates were found to be generally unreliable.

Another method involving NSM synthetic rainfall generation appears to be more rational in both form and delivered results. There were no ambiguous or erroneous estimates from the results of this method. Estimates based on this method are therefore more reliable and recommendable than those of the former type.

REFERENCES

