Some models for rainfall focused on the inner correlation structure

Sangdan Kim*

Abstract

In this study, new stochastic point rainfall models which can consider the correlation structure between rainfall intensity and duration are developed. In order to consider the negative and positive correlation simultaneously, the Gumbels type-II bivariate distribution is applied, and for the cluster structure of rainfall events, the Neyman-Scott cluster point process is selected. In the theoretical point of view, it is shown that the models considering the dependent structure between rainfall intensity and duration have slightly heavier tail autocorrelation functions than the corresponding independent models. Results from generating long time rainfall events show that the dependent models better reproduce historical rainfall time series than the corresponding independent models in the sense of autocorrelation structures, zero rainfall probabilities and extreme rainfall events

Key words: Rainfall model, Intensity, Duration, Correlation

1. Introduction

A stochastic rainfall model has a considerable portion in any stochastic hydrologic analysis. Cluster point processes have been used to model hourly rainfall events. However, in order to easily derive the model covariance structure, many models have assumed the independent structure between rainfall intensity and duration except a few models although indeed the independence is totally unrealistic and gives rise to critical model limitations. Even though above models can consider the dependence, they have some important problems, and they only applied the Poisson process to rainfall occurrence without considering any clustering feature of rainfall. In particular, Singh and Singh (1991), Bacchi et al. (1994), and Kurtothe et al. (1997) applied the Gumbel's type-I bivariate exponential distribution (Gumbel, 1960). Therefore, they just considered the negative correlation between rainfall intensity and duration by reason that this bivariate distribution always has the negative correlation between the variables. In the cases of Cordova and Rodriguez-Iturbe (1985), and Goel et al. (2000), the Downton's bivariate exponential distribution (Downton, 1970) was applied in order to consider the positive correlation between rainfall intensity and duration. However, Cordova and Rodriguez-Iturbe (1985) just applied it to compute the rainfall amount (not intensity) with the intention of computing surface runoff. Although Goel et al. (2000) insisted that their model could consider the negative correlation, the Downton's distribution can consider only the positive correlation between the variables (Downton, 1970; Nagao and Kadoya, 1971; Cordova and Rodriguez-Iturbe, 1985). In this study we intend to develop stochastic point rainfall models which can explain the correlation between rainfall intensity and duration and the clustering feature of rainfall occurrence.

* Research Fellow, Kyonggi Research Institute, E-mail: skim@kri.re.kr

2. Poisson Rectangular Pulse Model

The model representing rainfall intensity $X(t)$ is based upon the following assumptions. There are storms that occur in the Poisson process with parameter λ; each event is a rectangular pulse of random intensity i_r and duration t_r. It is assumed that the event characteristics are independent of the times of occurrence and that they are identically distributed and mutually independent random variables. In the traditional case, i_r and t_r for each event are also independent. However, in this study it is assumed that in each event, i_r and t_r are considered correlated random variables. In order to consider such correlations, we consider i_r and t_r to be represented by the bivariate exponential distribution of the type II proposed by Gumbel (1960).

$$f_{I_rT_r}(i_r, t_r) = \mu e^{-\mu i_r}\eta e^{-\eta t_r}[1 + r(2e^{-\mu i_r} - 1)(2e^{-\eta t_r} - 1)] \quad (1)$$

where μ, η, and r are parameters which should be estimated later. The marginal distributions of i^r and t_r are given by the following.

$$f_{I_r}(i_r) = \mu e^{-\mu i_r}, \qquad \mu > 0 \quad (2)$$

$$f_{T_r}(t_r) = \eta e^{-\eta t_r}, \qquad \eta > 0 \quad (3)$$

Hence, the second-order properties of the aggregated model process Y_T, the cumulative rainfall amount over the time interval T are the following.

$$E[Y_T] = \frac{\lambda(4+r)T}{4\mu\eta} \quad (4)$$

$$Var[Y_T] = \frac{\lambda}{8\mu^2\eta^3}\begin{pmatrix} -21r + 32 + 18r\eta T + 32\eta T - 3re^{-2\eta T} \\ + 32e^{-\eta T} + 24re^{-\eta T} \end{pmatrix} \quad (5)$$

$$Cov[Y_T, Y_{T+1}] = \frac{\lambda}{16\mu^2\eta^3}\begin{pmatrix} 21r + 32 - 3re^{-4\eta T} + 32e^{-2\eta T} - 48re^{-\eta T} \\ + 30re^{-2\eta T} - 64e^{-\eta T} \end{pmatrix} \quad (6)$$

The detailed derivation without considering the dependent structure between i_r and t_r is given by Rodriguez-Iturbe et al. (1984). The results of equations (4)~(6) can be derived by the same manner with Rodriguez-Iturbe et al. (1984) except for considering the dependent structure. If r would be zero, these formulas perfectly agree with the corresponding equations of the traditional Poisson rectangular pulses model without considering the dependent structure between rainfall intensity and duration.

3. Neyman-Scott Rectangular Pulse Model

The Poisson process with parameter λ governs storm arrivals. The storm is conceptualized as many rain cells according to a time-placement probability density function, $f(t)$.

$$f(t) = \beta e^{-\beta t} \qquad \beta > 0 \quad (7)$$

where β is a parameter which also should be estimated later. The number of cells per a storm is a random variable ν, which is independent of the storm properties, and is assumed to be governed by the Poisson distribution. These cells are distributed by the Neyman-Scott process. A detailed discussion on this process is given by Kavvas and Delleur (1981), and Waymire and Gupta (1981). Each cell is the rectangular pulse of random intensity i_r and duration t_r as the previous case. That is, we consider i_r and t_r to be represented by

$$f_{I_r T_r}(i_r, t_r) = \mu e^{-\mu i_r} \eta e^{-\eta t_r}[1 + r(2e^{-\mu i_r} - 1)(2e^{-\eta t_r} - 1)] \tag{8}$$

Hence, the second-order properties of the aggregated model process Y_T are the following.

$$E[Y_T] = \frac{\lambda E[\nu](4+r)T}{4\mu\eta} \tag{9}$$

$$\begin{aligned} Var[Y_T] &= \frac{\lambda E[\nu]}{8\mu^2\eta^3}\begin{pmatrix} -21r + 32 + 18r\eta T + 32\eta T - 3re^{-2\eta T} \\ + 32e^{-\eta T} + 24re^{-\eta T} \end{pmatrix} \\ &+ \frac{\lambda E^2[\nu]}{4\beta\mu^2(\eta^2 - \beta^2)(4\eta^2 - \beta^2)}\begin{pmatrix} -r^2\eta^2 + r^2\eta^2 - 8r\eta^2 \\ + 8r\eta^2 - 16\eta^2 + 16\eta^2 \\ + 4\beta^2 - 4\beta^3 T + r^2\eta^2 e^{-} \\ + 8r\eta^2 e^{-} + 16\eta^2 e^{-} - 4\beta^2 e^{-} \end{pmatrix} \end{aligned} \tag{10}$$

$$\begin{aligned} Cov[Y_T, Y_{T+1}] &= \frac{\lambda E[\nu]}{16\mu^2\eta^3}\begin{pmatrix} 21r + 32 - 3re^{-4\eta T} + 32e^{-2\eta T} - 48re^{-\eta T} \\ + 30re^{-2\eta T} - 64e^{-\eta T} \end{pmatrix} \\ &+ \frac{\lambda E^2[\nu](1 - e^{-\beta T})}{4\beta\mu^2(\eta^2 - \beta^2)(4\eta^2 - \beta^2)}\begin{pmatrix} r^2\eta^2 + 8r\eta^2 + 16\eta^2 - 4\beta^2 \\ - r^2\eta^2 e^{-} - 8r\eta^2 e^{-} \\ - 16\eta^2 e^{-} + 4\beta^2 e^{-} \end{pmatrix} \end{aligned} \tag{11}$$

The detailed derivation without considering the dependent structure between i_r and t_r is given by Rodriguez-Iturbe (1986). The results of equations (9)~(11) can be derived by the same manner with Rodriguez-Iturbe (1986) except for considering the dependent structure. As one should expect, if r would be zero, these formulas completely agree with the corresponding equations of the traditional Neyman-Scott rectangular pulses model without considering the dependent structure between rainfall intensity and duration.

4. Parameter Estimation

The above models have four or six parameters: λ, μ, η, r, β, and $E[\nu]$, which should be estimated using the method of moments. Various combinations of first- and second-order statistics from historical rainfall data can be equated to their theoretical expressions, resulting in a set of four or six highly nonlinear equations with four or six unknowns. A minimum least square technique has been employed to obtain estimates of the model parameters. Let $F(X)$ be the set of nonlinear equations in parameter X that must satisfy the observation vector θ:

$$F(\hat{X}) - \theta = 0 \tag{12}$$

where $F(\hat{X})$ is the best estimate of θ. The elements in θ have different order of magnitudes and hence their sum of the squares tends to be biased toward higher values. To circumvent such problem, every element of $F(X)$ is normalized by the corresponding element of θ. Now, the solution of equation (12) may be derived through a nonlinear minimization:

$$\min X\{(\frac{f_1(X)}{\theta_1}-1)^2+(\frac{f_2(X)}{\theta_2}-1)^2+...+(\frac{f_i(X)}{\theta_i}-1)^2+...\} \quad (13)$$

The equation (13) is applied to the traditional Poisson rectangular pulse model (PRPM), the dependent Poisson rectangular pulse model (DPPRPM), the tradition Neyman-Scott rectangular pulse model (NSRPM), and the dependent Neyman-Scott rectangular pulse model (DPNSRPM). The results of the parameter estimation are given in Table 1. In order to satisfy the temporal homogeneity with regard to the rainfall characteristics, we used only July (wet season) data for Seoul.

5. Model Performances

Using estimated parameters, about 1,000 days rainfall depths are generated for each model. Fig. 1 shows the probability of zero depth from original data and generated data. As can be seen in Fig. 1, the dependent models have more similar dry-wet time structure of rainfall than the corresponding independent models. This feature is thought to be important in the hydrologic applications where the soil moisture variation and evaporation of a basin are of interest. The further study about the soil moisture variation using the dependent and independent models is recommendable and being explored by the first author. The extreme value analysis for various durations using the above generations is represented in Fig. 2. As can be seen in Fig. 2, the results of the extreme value analysis for various durations show that the dependent models have more suitable structures than those of the corresponding independent models.

6. Conclusions

In this study we constructed four stochastic point rainfall models: 1) with or without considering the correlation between rainfall intensity and duration and 2) with or without considering the cluster structure of rainfall. Especially in the case of considering the correlation between rainfall intensity and duration, such models were developed to be able to simulate both the positive or negative correlation, which depends on historical rainfall events. The parameters of each model can be estimated by using the nonlinear optimization technique combined with the method of moments in the sense of comparing the statistics from historical rainfall events with the theoretical model statistics. As a result from generating long time rainfall events with the estimated model parameters, the dependent-clustering model shows the best performances in the sense of reproducing the statistics of historical rainfall time series. Then it comes the independent-clustering model, the dependent-nonclustering model, and the independent-nonclustering model in that order. Especially, it is noticeable that the problems with the overestimation of zero rainfall probabilities and the underestimation of extreme rainfall events can be improved by considering the correlation structure between rainfall intensity and duration. In addition, if such inner correlation structures of rainfall can be coupled with modeling the soil moisture behavior, another significant results will be expected.

References

1. Bacchi, B., Becciu, G. and Kottegoda, N.T. (1994). "Bivariate exponential model applied to

intensities and durations of extreme rainfall." J. Hydrol., Vol. 155, pp. 225-236.
2. Cordova, J.R. and I. Rodriguez-Iturbe, I. (1985). "On the probabilistic structure of storm surface runoff." Water Resour. Res., Vol. 21, pp. 755-763.
3. Downton, F. (1970). "Bivariate exponential distributions in reliability theory." J. Royal Stat. Soc., Ser. B, Vol. 32, pp. 408-417.
4. Goel, N.K., Kurothe, R.S., Mathur, B.S. and Vogel, R.M. (2000). "A derived flood frequency distribution for correlated rainfall intensity and duration." J. Hydrol., Vol. 228, pp. 56-67.
5. Gumbel, E.J. (1960). "Bivariate exponential distributions." J. Am, Stat. Assoc., 55, pp. 698-707.
6. Kavvas, M.L., and J.W. Delleur (1981). "A stochastic cluster model of daily rainfall sequences." Water Resour. Res., Vol. 17, pp. 1151-1160.
7. Kurothe, R.S., Goel, N.K. and Mathur,B.S. (1997). "Derived flood frequency distribution for correlated rainfall intensity and duration." Water Resour. Res., Vol. 33, pp. 2103-2107.
8. Rodriguez-Iturbe, I. (1986). "Scale of fluctuation of rainfall models." Water Resour. Res., Vol. 22, 15s-37s.
9. Rodriguez-Iturbe, I., Gupta, V.K. and Waymire, E. (1984). "Scale considerations in the modeling of temporal rainfall." Water Resour. Res., Vol. 20, pp. 1611-1619.
10. Nagao, M., and Kadoya, M. (1971). "Two-variate exponential distribution and its numerical table of engineering applications." Bull. Disaster Prev. Res. Inst., Kyoto Univ., Vol. 20, pp. 183-215.
11. Singh, K. and Singh, V.P. (1991). "Derivation of bivariate probability density functions with exponential marginal." Stoch. Hydrol. Hydraul., Vol. 5, pp. 55-68.
12. Waymire, E. and Gupta, V.K. (1981). "The mathematical structure of rainfall representations, 3, Some applications of the point process theory to rainfall process." Water Resour. Res., Vol. 17, pp. 1287-1294.

Table 1. The results of the parameters estimation

Station	Parameter	PRPM	DPPRPM	NSRPM	DPNSRPM
Seoul	λ, 1/hour	0.051836	0.035667	0.006686	0.0070001
	μ, 1/hour	0.11303	0.093143	0.13149	0.10426
	η, 1/hour	0.92059	0.58952	1.3427	0.83888
	$E[\nu]$			13.1552	8.1427
	β			0.086004	0.064729
	r		−0.93218		−0.94242

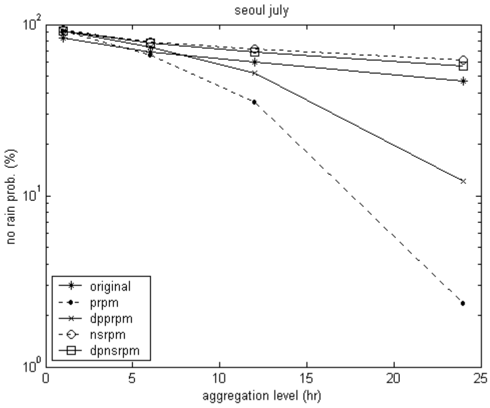

Fig. 1. The probability of zero depth

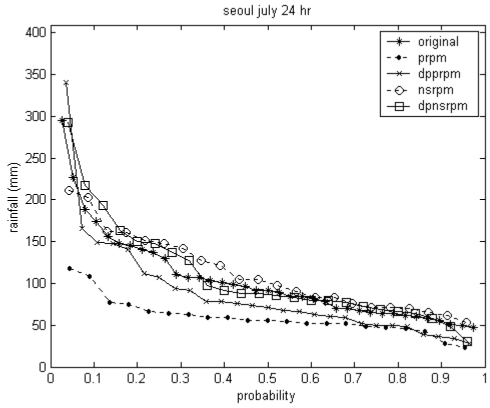

Fig. 2. The extreme value analysis