Journal of the Korean Data & Information Science Society
2014, **25**(1), 245–254

http://dx.doi.org/10.7465/jkdi.2014.25.1.245
한국데이터정보과학회지

Stochastic simulation of daily precipitation: A copula approach†

Changhui Choi[1] · Bangwon Ko[2]

[1]Korea Insurance Research Institute
[2]Department of Statistics and Actuarial Science, Soongsil University
Received 2 December 2013, revised 3 January 2014, accepted 12 January 2014

Abstract

The traditional methods of simulating daily precipitation have paid little attention to the inherent dependence structure between the total precipitation amount and the precipitation frequency for a fixed period of time. To address this issue, we propose a new simulation algorithm using copula in order to incorporate the dependence into the traditional methods. The algorithm consists of two parts: First, while reflecting the observed dependence, we generate the total precipitation amount (S) and the frequency (N) during the period of interest; then we simulate the daily precipitation whose aggregation matches the pair of (N, S) generated in the first part. Our result shows that the proposed method substantially improves the traditional methods.

Keywords: Copula, daily precipitation simulation, gamma distribution, Markov chain.

1. Introduction

Simulating daily precipitation data is an important task with a wide range of applications, e.g., agriculture (Sharpley and Williamson, 1990a, 1990b), ecology (Kittel *et al.*, 1995), hydrological systems (Pickering *et al.*, 1988) and finance (Leobacher and Ngare, 2011). Traditionally, the simulation has been governed by two independent stochastic models; one for precipitation occurrence and the other for daily precipitation amount. For the former, the Markov chain models have been used extensively and, for the latter, various stochastic models such as exponential, gamma, skewed normal and mixed exponential distributions have been used. The readers are referred to Chin (1977), Stern (1980), Richardson (1981), Stern and Coe (1984), Richardson and Wright (1984), Nicks and Gander (1994), Duan *et al.* (1998), Katz and Parlange (1998), Wilks (1999), Hayhoe (2000) and Wan *et al.* (2005) for some early discussion.

However, a problem may arise under the traditional approach because the two stochastic models are assumed to be independent. Let us illustrate this point with an example using the traditional daily precipitation model for a fixed period of m days, where the number of

† This work was funded by the Korea Meteorological Industry Promotion Agency, Grant number KMIPA 2012-1207.
[1] Research fellow, Korea Insurance Research Institute, Seoul 150-010, Korea.
[2] Corresponding author: Assistant professor, Department of Statistics and Actuarial Science, Soongsil University, Seoul 156-43, Korea. E-mail: bko@ssu.ac.kr

precipitation occurrences (in days) during the period is denoted by N and the amount of daily precipitation in the i-th occurrence is denoted by X_i, where $i = 1, 2, \cdots, N$. The usual assumption is that the random variables X_is are independent and identically distributed (i.i.d.) and are independent of N. Then, for the total precipitation denoted by $S = \sum_{i=1}^{N} X_i$, it holds that

$$Cov(N, S) = Cov(N, E[S|N]) = Cov(N, NE[X_i]) = E[X_i]Var(N). \tag{1.1}$$

For the first identity, see, e.g., Ross (2007). Equation (1.1) implies that the individual moments of X_i and N determine the covariance between N and S regardless of their actual or observed dependent structure.

In Table 1.1, we present Pearson's correlation coefficients between N and S based on the daily observations of the month of July during the recent 30 years (1983∼2012) from 9 major cities (Busan, Daegu, Daejeon, Gangneung, Gwangju, Incheon, Jeju, Seoul, Ulsan) in South Korea. We have used the Korea Meteorological Administration (KMA) data. For the observed monthly pairs $(n_i, s_i), i = 1, 2, ..., 30$ for each city, the correlation coefficients can be obtained in two ways. By definition they can be calculated directly as $\widehat{\rho} = \sum_{i=1}^{30}(s_i - \bar{s})(n_i - \bar{n})/\sqrt{\sum_{i=1}^{30}(s_i - \bar{s})^2}\sqrt{\sum_{i=1}^{30}(n_i - \bar{n})^2}$ and the result is shown in the second column. Or, by using equation (1.1), they can be obtained as $\widehat{\rho}^{(1.1)} = \bar{x}\sqrt{\sum_{i=1}^{30}(n_i - \bar{n})^2}/\sqrt{\sum_{i=1}^{30}(s_i - \bar{s})^2}$ and the result is shown in the third column. The values $\widehat{\rho}^{(1.1)}$ can be considered as the correlation coefficients expected for the simulated data under the traditional approach. The forth column contains the difference between the two. Note that the values in the third column are consistently smaller than those in the second, and the differences are quite large in some cases. This means that the traditional approach tends to underestimate the actual dependence between N and S, and hence the simulation results may deviate from reality.

Table 1.1 Pearson's correlation coefficients

City	$\widehat{\rho}$	$\widehat{\rho}^{(1.1)}$	$\widehat{\rho} - \widehat{\rho}^{(1.1)}$
Busan	0.7601	0.5726	0.2467
Daegu	0.656	0.5554	0.1533
Daejeon	0.5315	0.525	0.0122
Gangneung	0.5324	0.5079	0.046
Gwangju	0.6117	0.5337	0.1275
Incheon	0.7106	0.5149	0.2755
Jeju	0.6831	0.5891	0.1375
Seoul	0.6939	0.518	0.2536
Ulsan	0.7227	0.5633	0.2205
Average	0.6558	0.5422	0.1637

The main contribution of this paper is to present a new copula-based simulation algorithm that takes into account the actual dependence structure within the traditional framework. Our simulation algorithm consists of two parts: First, reflecting the observed dependence, we generate values of N and S for the given period and then generate the amounts of daily precipitation using the simulated values of N and S. In order to inherit the simplicity of the traditional approach, we will assume the widely used probability distributions for N and S, i.e., the first-order Markov chain and the gamma distribution. Once the total precipitation is generated, we will exploit the distributional property of gamma random variables to simulate daily precipitation. We will focus on simulating within a certain period of the year, say one month, to minimize the seasonal effect.

The observed dependence structure between N and S will be modeled by employing a multivariate distribution function called copula. The main advantage of using a copula is due to Sklar's theorem (1959) which states that any multivariate distribution can be separated into its marginal distributions and a copula function. In other words, a copula model can be useful in analyzing correlated (not necessarily identically distributed) random variables, say $X_1, X_2, \cdots, X_n$, whose joint probability distribution is unknown or analytically intractable. The copula models have been popular in the field of finance and insurance (e.g. Cherubini, 2004; Malevergne and Sornette, 2006; Kim and Lee, 2011; Choi *et al.*, 2013), and lately introduced in the fields of meteorology and hydrology (e.g., Favre *et al.*, 2004; Zhang and Singh, 2007; Genest and Favre, 2007). For a general discussion of copula, the readers are referred to Joe (1997) and Nelsen (2006).

The paper is constructed as follows. Section 2 explains some theoretical background, modeling assumptions, and the idea behind the proposed simulation algorithm. Section 3 introduces a performance measure for comparing simulation methods and summarizes the numerical result based on the performance measure. Finally, Section 4 concludes the paper with further comments.

2. Stochastic models, parameter estimation and simulation algorithm

2.1. Precipitation models

Let us assume that the fixed period of interest consists of m days. As mentioned earlier, the precipitation occurrence is modeled by the first-order Markov chain $\{Y_k\}_{k=1}^m$ with two states $\{0, 1\}$. The state-0 on day k, i.e., $Y_k = 0$, means that there is no rain (the amount of precipitation less than 0.1 mm) on day k and state 1 represents the opposite. The transition probabilities are denoted by $p_{qr} = P(Y_k = r|Y_{k-1} = q)$ for $q, r \in \{0, 1\}$ and the transition probability matrix by $M = \begin{bmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{bmatrix}$. Obviously, if N denotes the number of rainy days during the period, it should hold that $N = \sum_{k=1}^m Y_k$.

For a given $N = n > 0$, the amount of precipitation on the i-th rainy day, $X_i, i = 1, 2, \cdots, n$, is modeled by i.i.d. gamma random variables with a shape parameter of α and a scale parameter of β. That is, the common probability density function (p.d.f.) is given by

$$f_{X_i}(x) = \frac{1}{\Gamma(\alpha)\beta^\alpha} x^{\alpha-1} e^{-x/\beta}, x > 0.$$

Then it follows from the additive property of the gamma distribution that $S = \sum_{i=1}^n X_i$ has a gamma distribution with a shape parameter of $n\alpha$ and a scale parameter of β. It is well known that $(X_1, \cdots, X_n)$ given that $S = s$ reduces to a Dirichlet distribution with the p.d.f.

$$f_{X_1,\cdots,X_n|S}(x_1, \cdots, x_n|s) = \frac{\Gamma(n\alpha)}{\{\Gamma(\alpha)\}^n} \frac{(x_1)^{\alpha-1} \cdots (x_n)^{\alpha-1}}{s^{n\alpha-1}},$$
$$x_1 > 0, \cdots, x_n > 0, x_1 + \cdots + x_n = s > 0.$$

Note that the scale parameter β is irrelevant. Moreover, if $W_1, \cdots, W_n$ are i.i.d. gamma random variables with a shape parameter of α and an arbitrary scale parameter, it is also known that $\left(\frac{sW_1}{W_1+\cdots+W_n}, \cdots, \frac{sW_n}{W_1+\cdots+W_n}\right)$ has the same Dirichlet distribution (see Hogg and Craig, 1978). These distributional properties will be used later for generating daily precipitation of $(X_1, \cdots, X_n)$ conditional on total precipitation $S = s$.

Unconditionally, the total precipitation S may not be gamma distributed, but it is assumed to follow the gamma distribution for simulation purposes in the first part of our simulation algorithm. This assumption can be justified through empirical investigation, if necessary. The shape parameter and scale parameter of the distribution of S will be denoted by a and b, respectively.

2.2. Copula models

This subsection briefly reviews the bivariate copulas that will be used later.

A bivariate copula function $C(u, v)$ is a two-dimensional joint distribution function whose univariate marginal distributions are uniform in the interval [0, 1]. For a given copula $C(u, v)$, a joint distribution of N and S can be defined as

$$F_{N,S}(n, s) = C\left(F_N(n), F_S(s)\right), \tag{2.1}$$

where $F_N(n)$ and $F_S(s)$ are the cumulative distribution functions for N and S, respectively. Conversely, there always exists a copula function $C(u, v)$ satisfying (2.1) for a given joint distribution $F_{N,S}(n, s)$. In our case, $C(u, v)$ may not be unique because N is a discrete random variable. See Nelsen (2006) for more detail.

Among the numerous copula models, we tried MATLAB-supplied copulas such as the Clayton copula, the Frank copula, the Gumbel copula, the Gaussian copula and the Student-t copula. The former three copulas belong to the class of Archimedean copula with the following explicit functional forms; the Clayton copula with parameter $\theta > 0$ takes the form

$$C(u, v) = (u^{-\theta} + v^{-\theta} - 1)^{-1/\theta};$$

the Frank copula with parameter $\theta \in (-\infty, \infty) \setminus \{0\}$ is given by

$$C(u, v) = -\frac{1}{\theta}\ln\left(1 + \frac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{e^{-\theta} - 1}\right);$$

and the Gumbel copula with parameter $\theta \geq 1$ is of the form

$$C(u, v) = \exp\left(-\left[(-\ln u)^{\theta} + (-\ln v)^{\theta}\right]^{1/\theta}\right).$$

The Archimedean copulas are very popular in many areas and have been applied for analyzing meteorological data by earlier researchers. See, for instance, Favre *et al*. (2004) and Zhang and Singh (2007). On the other hand, not allowing for compact functional forms, the Gaussian and the Student-t copulas are also extensively used in applied fields. In particular, the Gaussian copula is useful for simulating values with prescribed correlations and the Student-t copula appropriate for modeling tail dependence.

2.3. Estimation procedure

The simulation algorithm in the following subsection will require the estimates of the gamma parameters (α, β) for the daily precipitation amount and the gamma parameters (a, b) for the total precipitation amount and the transition probability matrix M. To estimate the gamma parameters, we have used the maximum likelihood estimation (MLE) method. The transition probabilities have also been obtained by MLE, i.e., by counting the number of rainy days in the given data. So, for each $q, r \in \{0, 1\}$,

$$\widehat{p}_{qr} = \frac{\sum_{h=1}^{T}\sum_{k=2}^{m} I(Y_{k-1}^h = q \text{ and } Y_k^h = r)}{\sum_{h=1}^{T}\sum_{k=2}^{m} I(Y_{k-1}^h = q \text{ and } Y_k^h = q) + \sum_{h=1}^{T}\sum_{k=2}^{m} I(Y_{k-1}^h = q \text{ and } Y_k^h = r)}$$

where Y_k^h takes on 0 if there is no rain on day k in year h and 1 otherwise, and $I(\cdot)$ denotes an indicator function. Here T and m represent the number of years and the number of days within a specific month, respectively. In this paper, $T = 30$ and $m = 30$ or 31 depending on the number of days in the month.

In the first part of the simulation algorithm below, we will generate random samples of N and S by the inverse transformation method. Since S is assumed to follow the gamma distribution with parameters a and b, the values of S can be simulated as $F_S^{-1}(U; \widehat{a}, \widehat{b})$ with a uniformly distributed random variable U on (0, 1) and MLEs $\widehat{a}$ and $\widehat{b}$. However, it is not so easy to find the distribution function of N. Hence we are going to generate a large number, say, 10^6 of Markov chains $\{Y_1, Y_2, ..., Y_m\}$ using the estimated transition probability matrix $\widehat{M}$ and estimate the distribution function of N from the simulated values. Let us denote the estimated empirical distribution function by $\widehat{F}_N(n; \widehat{M})$ based on the 10^6 simulated values. Now the values of N can be simulated as $\widehat{F}_N^{-1}(V; \widehat{M})$ with a uniformly distributed random variable V on (0, 1).

On the other hand, each copula model can be fitted by using the MATLAB function 'copulafit' (see `"http://www.mathworks.co.kr/kr/help/stats/copulafit.html"` for more detail). For an observed data set of $\{(n_1, s_1), (n_2, s_2), \cdots, (n_T, s_T)\}$, the input of the function is the pairs of the empirical distributions given by

$$F_N^e(n) = \frac{1}{T}\sum_{i=1}^{T} I(n_i \le n) \text{ and } F_S^e(s) = \frac{1}{T}\sum_{i=1}^{T} I(s_i \le s).$$

Then the function 'copulafit' returns MLEs for copula parameters. We repeat this procedure for the 5 candidate copulas and determine the best copula in Subsection 3.b.

2.4. Simulation algorithm

Our simulation algorithm consists of two parts. First we generate a random sample of (N, S) and then we generate two sets of daily precipitation amounts and occurrences, $\{X_1, X_2, \cdots, X_N\}$ and $\{Y_1, Y_2, \cdots, Y_m\}$, satisfying $S = \sum_{i=1}^{N} X_i$ and $N = \sum_{k=1}^{m} Y_k$.

Algorithm-A: generating the total precipitation

(A1) Generate a uniform random sample of (U, V) whose joint distribution is the fitted copula.

(A2) Set $N = \widehat{F}_N^{-1}(V; \widehat{M})$ and $S = F_S^{-1}(U; \widehat{a}, \widehat{b})$.

To implement (A1), we have used the MATLAB function 'copularnd'. For the simulated pair of (n, s), the daily precipitations can be obtained as follows.

Algorithm-B: generating the daily precipitation

(B1) Generate a random sample of $\{W_1, W_2, \cdots, W_n\}$ from the gamma distribution with the estimated shape parameter $\widehat{\alpha}$ and arbitrary scale parameter.

(B2) Set $X_i = \frac{sW_i}{\sum_{j=1}^n W_j}$ for each $i = 1, 2, \cdots, n$.

(B3) Repeat generation of a random sample of $\{Y_1, Y_2, \cdots, Y_m\}$ until we get $n = \sum_{k=1}^m Y_k$.

(B4) For each $j = 1, 2, \cdots, m$, if $Y_j = 1$, set $h_j = \sum_{k=1}^j Y_k$ and $D_j = X_{h_j}$. Otherwise, set $D_j = 0$.

The computational burden of (B3) is not too heavy because, if we let $q = P(\sum_{i=1}^m Y_i = n) > 0$, then the probability of not getting a random sample of $\{Y_1, Y_2, \cdots, Y_m\}$ satisfying $n = \sum_{k=1}^m Y_k$ until the r-th trial is $(1-q)^r$ and expected running time of the algorithm is q^{-1}. Even if q is small, the probability will converge to zero exponentially fast, and hence it can be expected that the random sample will be obtained within a reasonable amount of time. Finally, (B4) combines the generated data sets to construct a random sample of the unconditional daily precipitation $\{D_1, D_2, \cdots, D_m\}$.

Figure 2.1 illustrates Algorithm-A for a given data (Busan, July): Figure 2.1-(a) plots the pairs of the given data $\{(n_i, s_i)\}_{i=1}^{30}$; Figure 2.1-(b) the pairs of the empirical distributions $\{(F_N^e(n_i), F_S^e(s_i))\}_{i=1}^{30}$; Figure 2.1-(c) 100 pairs of uniform random numbers from the fitted copula; and Figure 2.1-(d) 100 pairs of (N, S).

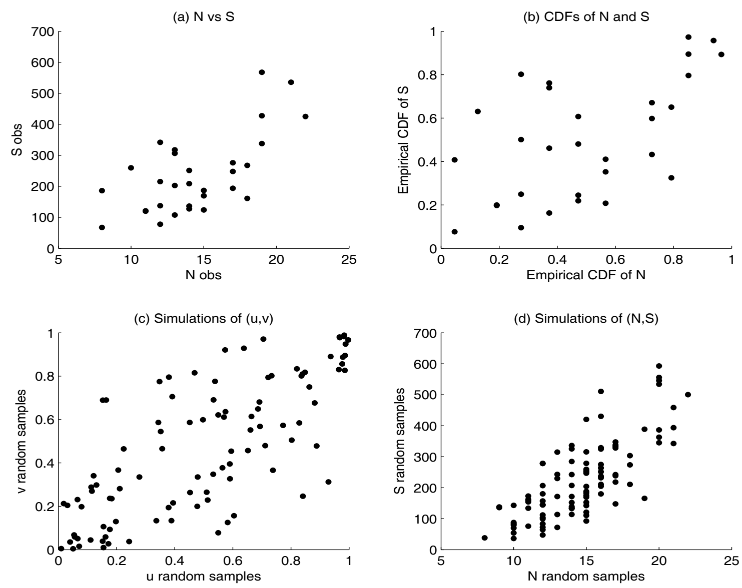

Figure 2.1 100 simulated values of (N, S) using Algorithm-A, Busan, July, 1983∼2012

3. Model comparison

For comparing models in this section, we used the KMA 30-year (1983∼2012) daily precipitation data of 9 cities (Busan, Daegu, Daejeon, Gangneung, Gwangju, Incheon, Jeju, Seoul, Ulsan) in South Korea from the month of July to September. We simulate daily precipitation under each model and estimate from the simulated data set the transition probability matrix, the gamma parameters and the correlation coefficient between N and S. If a simulation method results in the parameter estimates closest to the estimates from the observed data, then it will be considered the best among the copula models considered. The measure of closeness will be introduced in Subsection 3.1, and the result summarized in Subsection 3.2.

3.1. Performance measure

Throughout this section, we will use the symbol Θ to represent which data set is in use for the parameter calculation. If the calculation is based on the observed data, it will be denoted by $\Theta = O$; if on the simulated data under the traditional model, $\Theta = Tr$; and if on the simulated data under a copula model, $\Theta = C1 \sim C5$ ($C1$ (Gaussian), $C2$ (Student-t), $C3$ (Clayton), $C4$ (Frank), $C5$ (Gumbel)). So the notations $\widehat{M}^{\Theta}, (\widehat{\alpha}^{\Theta}, \widehat{\beta}^{\Theta})$ and $\widehat{\rho}^{\Theta}$ represent the parameter estimates and reveal which data set is being used. Our performance measure of a simulation method is defined through the following relative gap.

Definition 3.1 For two matrices A and B of identical dimensions, we define the relative gap with respect to A as $g(A, B) = \|A - B\|_F / \|A\|_F$ where $\| \cdot \|_F$ denotes the Frobenius norm defined by $\|A\|_F = \sqrt{\sum_{i=1}^{m} \sum_{j=1}^{n} A_{ij}^2}$ where $A \in R^{m \times n}$.

The concept of relative gap (or relative error) is widely used for measuring relative distance between two mathematical quantities such as matrices (see "`http://www.netlib.org/lapack/lug/node75.html`" for more detail). Obviously, the relative gap $g(A, B)$ is not symmetric, but it can be considered as an adjusted distance between A and B in terms of the magnitude of A. So, if $g(A, B) < g(A, C)$, we can say that B is closer to A than C is.

Suppose that we have estimated parameters $\left\{\widehat{M}^{\Theta}, (\widehat{\alpha}^{\Theta}, \widehat{\beta}^{\Theta}), \widehat{\rho}^{\Theta}\right\}$. Our performance measure is defined as

$$G(O, \Theta) = g(\widehat{M}^{O}, \widehat{M}^{\Theta}) + g((\widehat{\alpha}^{O}, \widehat{\beta}^{O}), (\widehat{\alpha}^{\Theta}, \widehat{\beta}^{\Theta})) + g(\widehat{\rho}^{O}, \widehat{\rho}^{\Theta}).$$

It simply adds the relative gaps with equal weights being given to the estimates of the Markov chain parameters, the gamma parameters, and the correlation coefficient. The motivation behind the performance measure is that, for an ideal simulation method, we can expect that the value of $G(O, \Theta)$ would be small because the parameter estimates from the simulated values under the ideal method would be close to those from the observed data. Thus a simulation method Θ_1 can be considered better than Θ_2 if $G(O, \Theta_1) < G(O, \Theta_2)$.

3.2. Computational result

In this subsection, we compare our copula models ($\Theta = C1 \sim C5$) with the traditional method ($\Theta = Tr$). In order to do so, we generated 10^5 sets of monthly precipitation (or

$10^5 \times m$ daily precipitation amounts) for each region and each month, obtained the parameter estimates, and calculated the performance measure $G(O, \Theta)$.

Table 3.1 shows which copula produces the smallest $G(O, \Theta)$ among the 5 copula models. The Gumbel copula and the Gaussian copula were selected as the best for 25 out of 27 cases (9 regions, 3 months), and the Student-t copula selected twice. The Clayton copula and the Frank copula did not produce the best result in any of the cases.

Table 3.1 Copula models with the smallest $G(O, \Theta)$

	July	August	September
Busan	Student-t	Gaussian	Gaussian
Daegu	Gumbel	Gumbel	Gumbel
Daejeon	Gaussian	Gumbel	Student-t
Gangneung	Gumbel	Gumbel	Gaussian
Gwangju	Gumbel	Gaussian	Gumbel
Incheon	Gaussian	Gumbel	Gaussian
Jeju	Gumbel	Gaussian	Gaussian
Seoul	Gaussian	Gumbel	Gumbel
Ulsan	Gumbel	Gaussian	Gaussian

Table 3.2 summarizes the performance measures of the traditional method and the best copula models selected in the previous step. The result shows that, on average, our best copula model reduces the performance measure by about 50%, which implies that our method can generate daily precipitation with statistical characteristics closer to the observed data than the traditional method. Observe that there are 6 cases (indicated by the asterisk in Table 3.2) where the traditional method performs better than our copula method. So we investigated these cases more carefully and found that it could happen when the traditional method generates random samples whose correlation coefficient is close to that of the observed data. Therefore we recommend that our method be used if there is a significant discrepancy between the correlation coefficients from the observed data and the traditional method as illustrated in Table 1.1 Although not presented in this paper, the decomposition of the values of $G(O, \Theta)$ shows that, when our copula method is used, the relative gaps of $g(\widehat{M}^O, \widehat{M}^\Theta)$ and $g((\widehat{\alpha}^O, \widehat{\beta}^O), (\widehat{\alpha}^\Theta, \widehat{\beta}^\Theta))$ are slightly larger but the relative gap of $g(\widehat{\rho}^O, \widehat{\rho}^\Theta)$ is significantly smaller than the traditional counterparts.

Table 3.2 The values of $G(O, \Theta)$ under the traditional method and the best copula method

	July		August		September	
	$G(O, Tr)$	$G(O, C)$	$G(O, Tr)$	$G(O, C)$	$G(O, Tr)$	$G(O, C)$
Busan	0.2975	0.1111	0.2875	0.1124	0.1893	0.0908
Daegu	0.2071	0.1326	0.1693	0.0972	0.1962	0.0914
Daejeon	0.0626	0.0893*	0.1600	0.1017	0.2762	0.1172
Gangneung	0.0928	0.0582	0.5149	0.2475	0.0702	0.0712*
Gwangju	0.1800	0.1301	0.2394	0.1433	0.1831	0.0714
Incheon	0.3208	0.0606	0.0696	0.1175*	0.1412	0.1461*
Jeju	0.1978	0.0901	0.0860	0.1413*	0.2514	0.0593
Seoul	0.2984	0.0839	0.2565	0.0540	0.0685	0.1143*
Ulsan	0.2677	0.1453	0.1092	0.0739	0.2595	0.0839
Average	0.2139	0.1001	0.2103	0.1210	0.1817	0.0940

* indicates the cases where the traditional method performs better than our copula method.

In Figure 3.1, we plotted the 100 simulated pairs of (N, S) under the traditional method and our copula method for the month of July in Seoul. As shown in the figure, the traditional method tends to underestimate the dependence between N and S, and hence the simulated pairs are more widely spread out than the observed data. On the other hand, our method could overcome this shortcoming by modeling the observed dependence structure through copula.

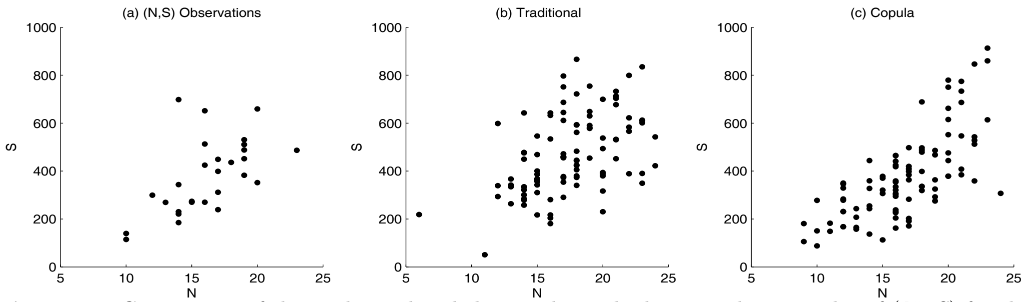

Figure 3.1 Comparison of the traditional and the copula method:100 random samples of (N, S) for the month of July in Seoul

4. Concluding remarks

This paper proposes a new simulation algorithm for generating daily precipitation within a fixed period of time. Unlike the traditional approach where the dependence between the total precipitation amount and the precipitation frequency has not been accounted for, the proposed method incorporates the dependence structure into the traditional methods by using copulas. The result shows that our simulation method can be more effective if there is a noticeable difference between the correlation coefficients of the observed values and those of the traditionally simulated values.

To inherit the simplicity of the traditional method, we have considered the first-order Markov chain and the gamma distributions. It is possible that the versatility of copulas will allow us to accommodate other stochastic models without much difficulty. For instance, the Markov chain can be replaced by the binomial model. If this is the case, then the algorithm for generating the precipitation frequency would be much simpler.

As mentioned earlier, this paper has focused on simulating daily precipitation within a fixed period of time, say, one month in order to minimize the seasonal effect. In particular, we have used the precipitation data for the month of rainy season in the hope that our algorithm could be used to develop and price precipitation insurance. However, we think that it is worthwhile to extend our copula method by reflecting seasonality in the future research.

Finally, we also would like to mention that our copula method for simulating daily precipitation is applicable to other areas such as the collective risk model for casualty and property insurance. The traditional precipitation model has a probabilistic structure similar to the collective risk model where S is the aggregate amount of losses for insurance company, X_i is the i.i.d. individual amount of loss, and N is the number of accidents independent of X_i. If the observed data of S and N reveals a certain dependence structure other than that implied by Equation (1.1), then our simulation method can be adopted for generating the individual amount of loss.

References

Cherubini, U., Luciano, E. and Vecchiato, W. (2004). *Copula methods in finance*, John Wiley & Sons, New Jersey.

Chin, E. H. (1977). Modeling daily precipitation occurrence process with Markov chain. *Water Resources Research*, **13**, 949-956.

Choi. C, Lee, H. and Joo, H. (2013), Analyzing rainfall patterns and pricing rainfall insurance using copula. *Journal of the Korean Data & Information Science Society*, **24**, 1598-9402.

Duan, J., Selker, J. and Gordon, E. G. (1998). Evaluation of probability density function in precipitation models for the Pacific Northwest. *Journal of the American Water Resources Association*, **24**, 617-627.

Favre, A.-C., El Adouni, S., Perreault, L., Thiémonge, N. and Bobée, B. (2004). Multivariate hydrological frequency analysis using copulas. *Water Resources Research*, **40** .

Genest, C. and Favre, A.-C. (2007). Everything you always wanted to know about copula modeling but were afraid to ask. *Journal of Hydrologic Engineering*, **12**, 347-368.

Hayhoe, H. N. (2000). Improvements of stochastic weather data generators for diverse climates. *Climate Research*, **14**, 75-87.

Hogg, R. V. and Craig, A. T. (1978). *Introduction to mathematical statistics*, 4th Edition, Macmillan, New York.

Joe, H. (1997) *Multivariate models and dependence concepts*, Chapman and Hall, New York.

Katz, R. W. and Parlange, M. B. (1998) Over dispersion phenomenon in stochastic modeling of precipitation. *Journal of Climate*, **11**, 591-601.

Kim, E. and Lee, T. (2011). A numerical study on portfolio VaR forecasting based on conditional copula, *Journal of the Korean Data & Information Science Society*, **22**, 1065-1074.

Kittel, T. G. F., Rosenbloom, N. A., Painter, T. H., Schimel, D. S. and VEMAP Modeling Participants. (1995). The VEMAP integrated database for modeling United States ecosystem/vegetation sensitivity to climate change. *Journal of Biogeography*, **22**, 857-862.

Leobacher, G. and Ngare, P. (2011). On modeling and pricing rainfall derivatives with seasonality. *Applied Mathematical Finance*, **18**, 71-91.

Malevergne, Y. and Sornette, D. (2006). *Extreme financial risks: From dependence to risk management*, Springer, New York.

Nelsen, R. B. (2006). *An introduction to copulas*, Springer, New York.

Nicks, A. D. and Gander, G. A. (1994). CLIGEN: A weather generator for climate inputs to water resource and other models. In *Proceedings Fifth International Conference on Computers in Agriculture*, American Society of Agricultural Engineers, Orlando, FL, 903-909.

Pickering, N. B., Stedinger, J. R. and Haith, D. A. (1988). Weather input for nonpoint-source pollution models. *Journal of Irrigation and Drainage Engineering*, **114**, 674-690.

Richardson C. W. (1981). Stochastic simulation of daily precipitation, temperature, and solar radiation. *Water Resources Research*, **17**, 182-190.

Richardson, C. W. and Wright, D. A. (1984). WGEN: A model for generating daily weather variables. Agricultural Research Service, ARS-8, **83**, Washington DC, USA.

Ross, S. M. (2007). *Introduction to probability models*, 9th Edition, Academic Press, New York.

Sharpley, A. N. and Williams, J. R. (1990a). *Erosion/productivity impact calculator 1. Model documentation*, Technical Bulletin 1768, U. S. Department of Agriculture, Washington DC, USA.

Sharpley, A. N. and Williams, J. R. (1990b). *Erosion/productivity impact calculator 2. User manual*, Technical Bulletin 1768, U.S. Department of Agriculture, Washington DC, USA.

Sklar, A. (1959). Fonctions de répartition à n dimensions et leurs marges. *Paris Institute of Statistics*, **8**, 229-231.

Stern, R. D. (1980). The calculation of probability distributions for models of daily precipitation. *Archiv für Meteorologie, Geophysik und Bioklimatologie B*, **28**, 137-137.

Stern, R. D. and Coe, R. (1984). A model fitting analysis of daily rainfall data. *Journals of the Royal Statistical Society A*, **147**, 1-34.

Wan, H., Zhang, X. and Barrow, E. M. (2005). Stochastic modeling of daily precipitation for Canada. *Atmosphere-Ocean*, **43**, 23-32.

Wilks, D. S. (1998). Multisite generalization of a daily stochastic precipitation generation model. *Journal of Hydrology*, **210**, 178-191.

Zhang, L. and Singh, V. P. (2007). Bivariate rainfall frequency distributions using Archimedean copulas. *Journal of Hydrology*, **332**, 93-109.