

Suroso, 2018_ Multisite daily precipitation simulation in

by Suroso Suroso

Submission date: 28-Mar-2023 09:47PM (UTC+0700)

Submission ID: 2049069977

File name: Suroso,_2018__Multisite_daily_precipitation_simulation_in.pdf (823.1K)

Word count: 3493

Character count: 19300

Multisite daily precipitation simulation in Singapore

Suroso Suroso^{1,*} and András Bárdossy²

¹Department of Civil Engineering, Jenderal Soedirman University, Purwokerto, Indonesia

²Institute for Modelling Hydraulic and Environmental Systems, University of Stuttgart, Germany

Abstract. Stochastic precipitation simulation is of great importance for the design and operation of water infrastructures. The objective of this research is to develop a stochastic simulation method for daily precipitation. Daily precipitation generation needs special treatment because of many zero values appearing due to dry days. It is implemented for 26 rain gauge stations located in Singapore. This research follows three steps. First, a hidden autoregressive (AR) model is fitted to time series data at each gauging station using a power transformation. Zero precipitation amounts are treated as censored values of the power-transformed Gaussian process. The hidden AR has four parameters: mean, autocorrelation, power transformation, and variance of error. Second, a conditional multivariate Gaussian distribution is fitted to residuals of the AR models and used to fill in censored values corresponding to errors of the AR at dry events. Third, stochastic simulations from the created spatial-temporal model are carried out. Single and multi-site statistical characteristics such as empirical distribution function, cross-correlation coefficient and entropy are used for evaluation of the model. The results of this research show that the developed model produces synthetic precipitation amounts having statistical characteristics very similar to the observed ones.

1 Introduction

Stochastic precipitation simulation is of great importance for the design and operation of water infrastructure projects because precipitation is one of the key inputs of the hydrologic systems analysis which is required in every step of projects such as planning, design, operation, and monitoring. The basic idea of a precipitation generator is to reproduce long synthetic precipitation data series which preserve some important statistical characteristics of observed precipitation data not only at single station such as mean value, variances, median, minimum and maximum values and some important quantiles of rainfall amounts, but also at multiple locations such as interdependence between two stations at different locations, namely Pearson and Spearman rank correlations of precipitation values. Recently, the entropy-based measure is also used to evaluate the performance of precipitation simulations [2,6]. It provides a measure of dispersion, uncertainty, disorder and diversification of precipitation [5].

* Corresponding author: surososipil@gmail.com

Daily precipitation model development is more demanding and needs special treatment because daily precipitation has unique characteristics including zero-inflated data due to dry days. Unlike annual and monthly precipitation amounts which can be modelled with simple autoregressive moving average (ARMA) processes, daily precipitation amounts cannot be directly modelled by the ARMA model [2,3]. The main difficulty with modelling daily precipitation amounts arises from the intermittent property of precipitation values both in space and time. Dry days with high probability due to zero-inflated data can be modelled by a discrete distribution and the rainfall amounts on rainy days at a selected location can be described by a continuous distribution [1], but the continuous distribution is usually skewed.

Two additional contrasting properties that make developing daily precipitation model more challenging are the long memory visible in the autocorrelation function and the sudden changes in the series [3]. Long memory refers to a not negligible dependence between distant observations in a time series [7]. Precipitation intensities itself follow exponential decay which indicates that precipitation series can be change rapidly.

The purpose of this study is to develop daily precipitation models at multiple locations incorporating both temporal and spatial dependencies. Zero-inflated data for dry days are treated as latent variables and estimated using the MCMC approach. The models are implemented in Singapore. The paper is divided into five sections. After the introduction, the spatial and temporal model developments are described in section 2. Section 3 presents the data and study location. Section 4 explains the simulation steps. The empirical findings of the investigation in Singapore are discussed in section 5. In the end, a summary of the results is given in section 6.

2 Model developments

2.1 Temporal model

An autoregressive model (AR) with lag-1 is used to model mixed daily precipitation occurrence and amount simultaneously. The general equations for this model are defined as follows.

$$Y_t = \mu + r \cdot Y_{t-1} + \varepsilon_t \quad (1)$$

Where Y_t is daily precipitation at time t , Y_{t-1} is daily precipitation at time $t-1$, μ is the mean parameter, r is autocorrelation, and ε_t is the residual error of AR (1) process at time t . The residual error ε_t is independent, identic, and normally distributed (white noise process). By taking ε_t and putting on the left side and also putting Y_t on the right side, the equation (1) can be written in a different formula.

$$\varepsilon_t = Y_t - (\mu + r \cdot Y_{t-1}) \quad (2)$$

Since ε_t follows the normal distribution, the probability density function (PDF) of ε_t can be written as a simple normal distribution function.

$$f(\varepsilon_t) = \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right) \exp \left[-\frac{1}{2\sigma^2} (\varepsilon_t - \text{mean})^2 \right] \quad (3)$$

Where σ^2 is the variance of the residual error ε_t , ε_t is the random variable of residual error which is equal to $Y_t - (\mu + r \cdot Y_{t-1})$ and the mean is zero. By substitution ε_t with $Y_t - (\mu + r \cdot Y_{t-1})$, equation (3) can be written in a different formula.

$$f(Y_t) = \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right) \exp \left[-\frac{1}{2\sigma^2} (Y_t - (\mu + r \cdot Y_{t-1}) - 0)^2 \right] \quad (4)$$

Since the daily precipitation Y_t follows the linear regression model AR (1) process, Y_t also follows the Gaussian process due to the assumption that its errors are white noise processes. In fact, the daily precipitation values present a skewed distribution characterized by zero-inflated data. To model the zeroes in the AR (1) process, a latent variable approach is employed and zeroes are considered as censored values of realizations in the negative part of the distribution whereas non-zeroes precipitation values are considered the positive realizations of a random variable. For the true daily precipitation ($Y_t > 0$), a likelihood function of density function derived from equation (4) can be expressed as follows.

$$L(\mu, r, \sigma^2) = \prod_{t=2}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[-\frac{1}{2\sigma^2} (Y_t - (\mu + r \cdot Y_{t-1}))^2 \right] \quad (5)$$

Since negative values of Y_t are considered as censored values, the likelihood function is modified. The likelihood function is used not only for parameter estimation purpose but also for filling in censored values.

$$L(\mu, r, \sigma^2, Y_{k1}, Y_{k2}, \dots, Y_{kj}) = \prod_{t=2}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[-\frac{1}{2\sigma^2} (Y_t - (\mu + r \cdot Y_{t-1}))^2 \right] \quad (6)$$

Y_k corresponding to variables for which is just known $Y_t \leq 0$. These correspond to no rain observation.

2.2 Power transformed AR (1) model

Daily precipitation amounts follow the skewed process instead of the normal distribution [1]. The skewed precipitation amounts are transformed to normality by using a beta β power transformation. This beta β power transformation is applied only to positive rainfall ($Z_t > 0$) values. In the following, rainfall values are represented by Z_t , whereas Y_t represents the underlying transformed into normalised AR. By considering the beta β power transformation, the transformed rainfall amount values are plugged into the AR (1) model.

$$Y_t = \begin{cases} Z_t^{\frac{1}{\beta}}, & \text{for } Z_t > 0 \\ Y_t, & \text{for } Z_t = 0 \end{cases} \quad (7)$$

The probability density function of each residual error ε_t as a function of daily precipitation Z_t is changed into equation 8 and 9 due to the beta β power transformation.

$$f(Y_t) = \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right) \exp \left[-\frac{1}{2\sigma^2} \left(Z_t^{\frac{1}{\beta}} - (\mu + r \cdot Y_{t-1}) \right)^2 \right] \left(\frac{Z_t^{\frac{1}{\beta}-1}}{\beta} \right), \text{ for } Z_t > 0 \quad (8)$$

$$f(Y_t) = \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right) \exp \left[-\frac{1}{2\sigma^2} (Y_t - (\mu + r \cdot Y_{t-1}))^2 \right], \text{ for } Z_t = 0 \quad (9)$$

The likelihood function of equation 8 and 9 can be expressed as shown in equation 10. The likelihood function is used for estimating parameters on the basis of rainfall values $Z_t > 0$ and latent variables Y_t corresponding to $Z_t = 0$.

$$L(\mu, r, \beta, \sigma^2, Y_{k1}, \dots, Y_{kj}) = \prod_{t: Z_t > 0}^{N_{pos}} \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[-\frac{1}{2\sigma^2} \left(Z_t \frac{1}{\beta} - \left(\mu + r \cdot Y_{t-1} \frac{1}{\beta} \right) \right)^2 \right] \left(\frac{Z_t \frac{1}{\beta}}{\beta} \right)^* \prod_{t: Z_t < 0}^{N_{neg}} \frac{1}{\sqrt{2\pi\sigma^2}} \exp \left[-\frac{1}{2\sigma^2} (Y_t - (\mu + r \cdot Y_{t-1}))^2 \right] \quad (10)$$

Where $N_{neg} = \{k_1, \dots, k_j\}$ corresponds to the times for which no rain was observed ($Z_{k1} = \dots = Z_{kj} = 0$). This likelihood function is used both for the purpose of parameter estimation and also for filling in censored daily precipitation by including as arguments of the likelihood function of latent variables Y_{k1}, \dots, Y_{kj} corresponding to dry days, $Z_{k1} = Z_{kj} = 0$.

2.3 Spatial gaussian model

Residual errors obtained from times series model AR (1) are modelled spatially by spatial Gaussian distribution. A vector-valued random variable $X = [X_1, X_2, \dots, X_d]$ is said to have a multivariate normal (or Gaussian) distribution with mean vector μ and covariance matrix Σ denoted by $X = [X_1, X_2, \dots, X_d] \sim N(\mu, \Sigma)$, if its PDF is given by this following formula.

$$p(x; \mu, \Sigma) = \left(\frac{1}{2\pi} \right)^{\frac{d}{2}} \left[\frac{1}{\det(\Sigma)} \right]^{\frac{1}{2}} \exp \left[-\frac{1}{2} (x - \mu)^T \Sigma^{-1} (x - \mu) \right] \quad (11)$$

Where Σ is a $d \times d$ symmetric, positive definite of a covariance matrix, μ is expected value, d is dimensional multivariate Gaussian model, T is matrix transpose operator, and d is the multivariate dimensional. In order to estimate parameters μ and Σ , the likelihood function is used. Given an independent and identically-distributed (iid) sample $X = \{X_1, X_2, \dots, X_d\}$ of random vectors, the likelihood of the sample, assuming data is normally distributed, is given by the equation below.

$$L(\mu, \Sigma) = f(X_1, \dots, X_d; \mu, \Sigma) = \prod_{j=1}^n \left\{ \frac{1}{(2\pi)^{\frac{d}{2}} |\Sigma|^{\frac{1}{2}}} \exp \left[-\frac{(X_j - \mu)^T \Sigma^{-1} (X_j - \mu)}{2} \right] \right\} \quad (12)$$

Where n is the length of series data and Σ^{-1} is the inverse of the covariance matrix, also called a concentration matrix or precision. Zero mean vectors and covariance matrix Σ are used to model the residuals of the fitted individual AR (1) models using multivariate Gaussian distribution. These residual errors consist of censored errors since some of them come from the censored daily precipitation that is non-positive daily precipitation. In order to fill in the censored residual errors, conditional normal distribution is adopted.

$$\begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \sim N \left(\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix} \right) \quad (13)$$

Where X_1 is the censored residual errors, X_2 is the true residual errors, μ_1 is the mean value of the censored residual errors, μ_2 is mean value of true residual errors, Σ_{11} is variance of the censored residual errors, Σ_{12} is covariance matrix between the censored residual errors,

and Σ_{21} is the covariance matrix between the true residual errors and the censored residual errors, and Σ_{22} the is variance of the true residual errors.

3 Data and study location

The models are implemented in Singapore using daily precipitation amounts which are simultaneously measured from 126 stations during the period of 1980-2010 (31 years). The stations cover a wide range of distances from 1.9 to 38.7 km as presented in Fig. 1.

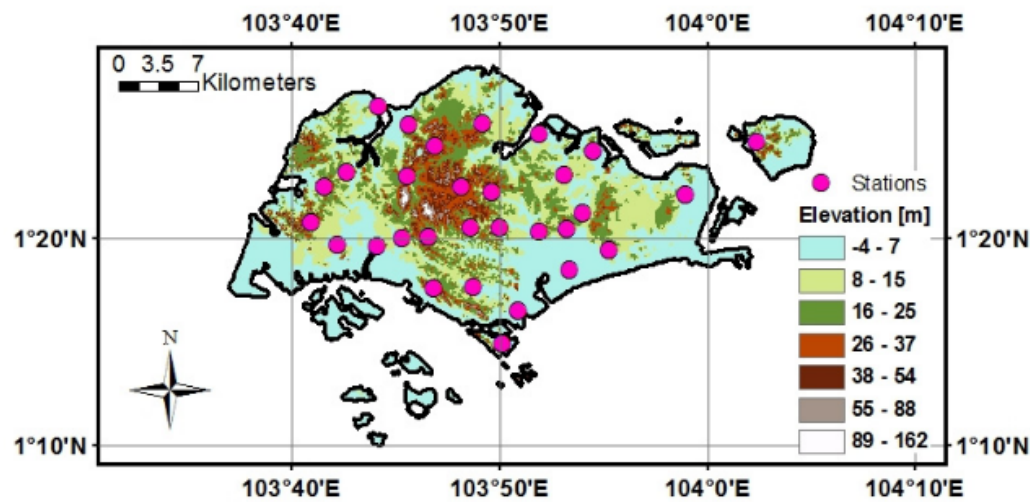


Fig. 1. Study location of Singapore.

4 Simulation steps

4.1 Temporal model

The time series model is developed separately for each station. To capture seasonal behaviour, the temporal models are constructed for every month separately. The detailed steps for developing the temporal models are illustrated as follows.

1. Observed daily precipitation variables Z_t are transformed using beta power transformation into normal variates Y_t , as given in equation 7.
2. Transformed daily precipitation values are formulated into AR (1) underlying Gaussian process as in equation 1.
3. Power transformed normal density functions are formulated as equations 8 ($Z_t > 0$) and 9 ($Z_t = 0$).
4. Likelihood functions of AR (1) are formulated as equation 10.
5. The parameters of AR (1) are estimated and the censored values are filled in by MCMC using metropolis Hastings algorithm.

4.2 Spatial model

Residual errors resulted from the marginal time series AR (1) models are modelled spatially using a multivariate Gaussian distribution with a mean vector of zeroes and covariance matrix Σ . The residual errors include censored data coming from censored data in time

series AR (1). Hence, conditional multivariate normal distribution is applied to impute censored residual errors based on the true residual errors. To estimate covariance matrix Σ as well as to fill in censored residual errors, the MCMC approach is implemented. The detailed steps for the fitting spatial model are illustrated as follows.

1. For each station, observed daily precipitation Z_t data are transformed into normal variates Y_t using estimated beta power transformation (β) as equation 7.
2. Simulated N number synthetic daily precipitation data using fitted temporal model AR (1) as equation 1. Variable N corresponds to the length of observed daily precipitation minus one data because it is taken as initial value Y_{t-1} in the first simulation ($Y_{t=0}$).
3. If daily precipitation given at time t is observed positive ($Z_t > 0$), calculate the true residual errors $\hat{\epsilon}$ using subtracting transformed observed daily data (step-1) with simulated synthetic data (step-2).
4. If daily precipitation given time is observed non-positive ($Z_t \leq 0$), initiate the censored errors $\tilde{\epsilon}$.
5. Estimate parameters of multivariate normal distribution (the covariance matrix Σ and correlation matrix Γ) and impute the censored errors using MCMC approach with metropolis Hastings algorithm.

4.3 Model simulation

Based on the temporal and spatial models constructed, synthetic daily precipitation data can be generated. The detailed steps for developing simulation models are described below.

1. Set up a number of values to simulate as N . For comparison purposes, this number should be the same as the observed values.
2. The observed daily precipitation at time $t=1$ ($Z_{t=1}$) is transformed into normal variate Y using estimated beta power transformation as equation 7. This value Y is used as starting value for simulation ($Y_{t=0}$).
3. Generate residual errors based on fitted multivariate Gaussian distribution.
4. Simulate synthetic daily precipitation data using fitted time series model AR (1) at each station.
5. If simulated daily precipitation at time t from step-4 is positive ($Y_t > 0$), transform this value Y_t to the power fitted beta (β).
6. If simulated daily precipitation at time t from step-4 is non-positive ($Y_t \leq 0$), set up simulated daily precipitation to zero.
7. Set up this simulated data Y_t as previous value Y_{t-1} for the next step simulation.
8. Repeat steps 3-7 until N number of simulations.

5 Results and discussion

To evaluate whether or not the model performs well, some measures such as empirical cumulative distribution function (ECDF), cross-correlation coefficient (CCC) and entropy-based measure [2, 4] are taken. A good model will reproduce synthetic daily precipitation series which have similar statistical key parameters characteristics to observed data.

5.1 Single site evaluation

Mean daily precipitation is a basic measure to be used for measuring information of ¹³ precipitation generation model between simulated and observed daily precipitation. The mean is a measure of central tendency of data distribution. The developed model can

synthetically reproduce daily precipitation very well in terms of mean value for all stations. Like the mean value, the standard deviation is also used to measure whether or not simulated daily precipitation resembles observed daily precipitation. The standard deviation is defined as the positive square root of the variance. The variance of a given set of data gives a measure of the variability of the data with respect to the mean. The model performs very well in that there are no significant differences between simulated and observed daily precipitations.

Cumulative relative frequencies both observed and simulated daily precipitation for all rain gauge stations are shown in Fig. 2. Overall, simulated ones can satisfactorily follow the behaviour of observed ones. Using the Kolmogorov Smirnov test with a significant level 95%, there is no significant difference between the two different datasets. But interestingly, for extreme values, some simulated cumulative relative frequencies tend to be located slightly above the observed cumulative relative frequencies, indicating an underestimation of the probability of very high rainfall values. It can occur due to the strange precipitation events that are influenced by many factors, for example, climate change.

5.2 Cross-correlation coefficient (CCC)-based evaluation

Spatial correlation exhibited by observed daily precipitation should be mimicked by the good multi-site daily precipitation generation model. Multi-site simulated daily precipitation should reveal similar two-dimensional cross-correlations of a pair of daily precipitation series data to the real data. The correlations measure the strength of association between two continuous variables.

In this study, three different types of CCCs, namely Pearson, Spearman, and Kendall cross-correlation, are applied to evaluate the multi-site precipitation generation model. The Pearson correlation is also called the linear correlation coefficient because it measures the linear association between two variables. The Pearson correlation is not resistant to outliers because it is computed using non-resistant measures, namely means and standard deviations.

In contrast, the Spearman correlation is a rank correlation. It depends only on the ranks of the data and not the values themselves. Thus, the Spearman correlation is more resistant to outliers than Pearson correlation. Fig. 3 reveals that there is only a slight difference for all CCCs between simulated and observed daily precipitation. It is found that simulated daily precipitation exhibits lower cross-correlations compared to cross-correlations observed in data.

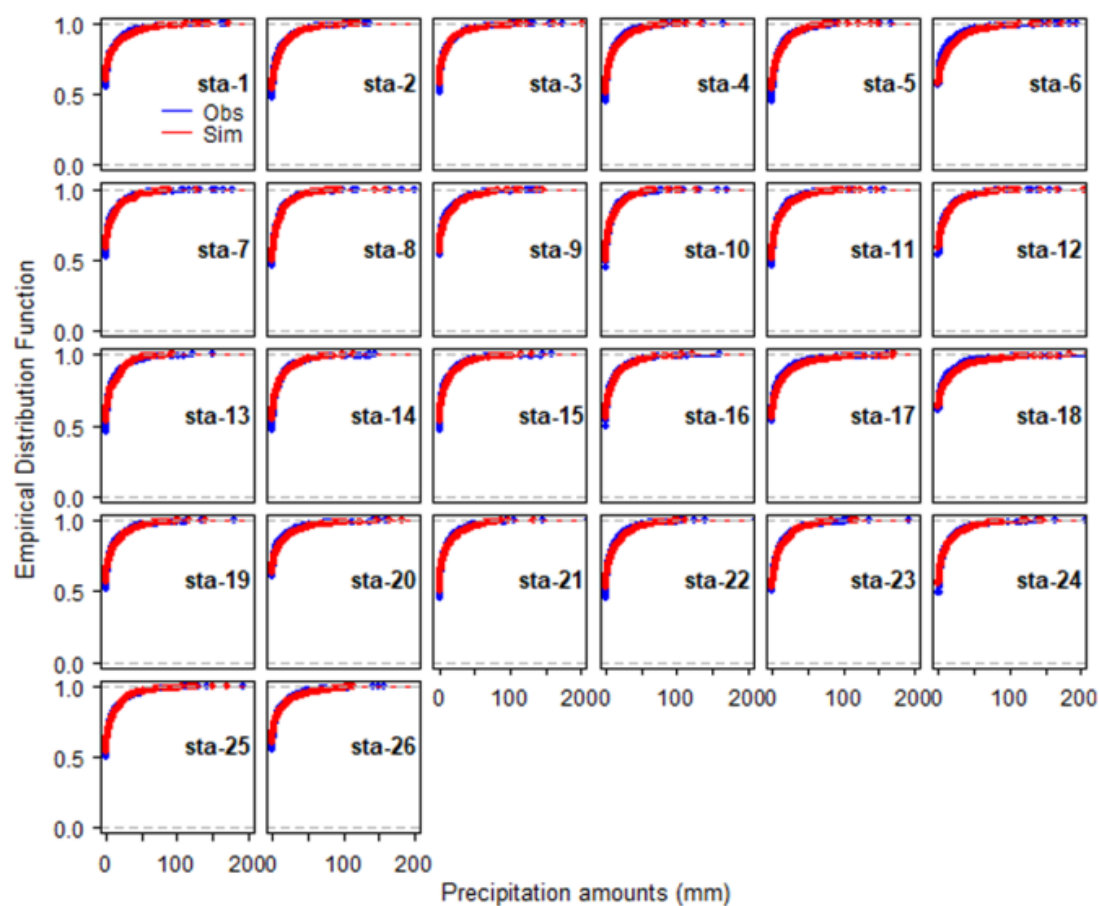


Fig. 2. The empirical cumulative distribution function for both observed and simulated precipitation amounts.

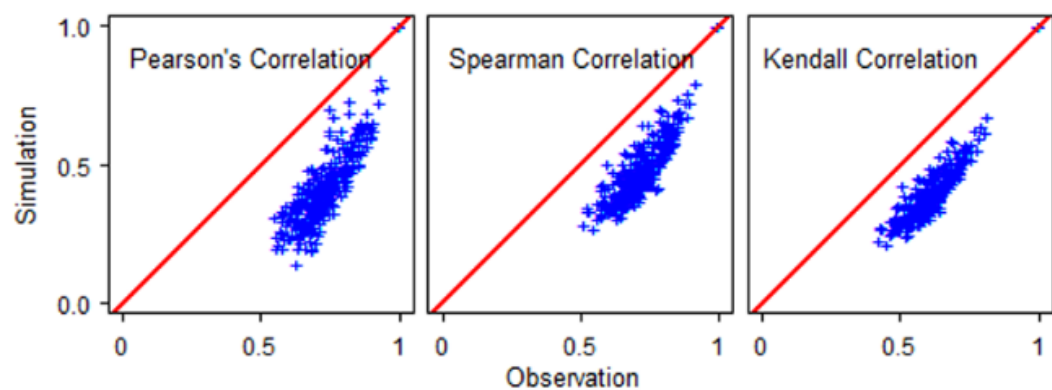


Fig. 3. Cross-correlation coefficient between two stations at different locations for Pearson, Spearman, and Kendall).

5.3 Entropy-based measure

Traditional measures such as mean, variance, cross-correlations, and other conventional criterions are not enough to evaluate multi-site daily precipitation model. The entropy method to measure uncertainty information especially used to validate daily precipitation model independently is the breakthrough approach in any hydrologic application. Therefore, the entropy-based measure is also used for both observed and simulated daily

precipitation amounts for some important quantiles (0.5, 0.6, 0.75, 0.8, 0.9, and 0.95). A number of triangles created each month are combinations of all stations taken from three stations at a time. All simulated daily precipitation for all thresholds systematically gives a slightly bigger entropy value than observed data. This fact indicates that simulated daily precipitation yield greater uncertainty or lower association than the observed one. This issue has become a hot topic in hydrological application. This finding also confirms what Bardossy [2, 4] discovered about the entropy approach used for validation of daily precipitation generation models. As already noted, the entropy of simulated daily precipitation is consistently higher than the entropy of observed daily precipitation.

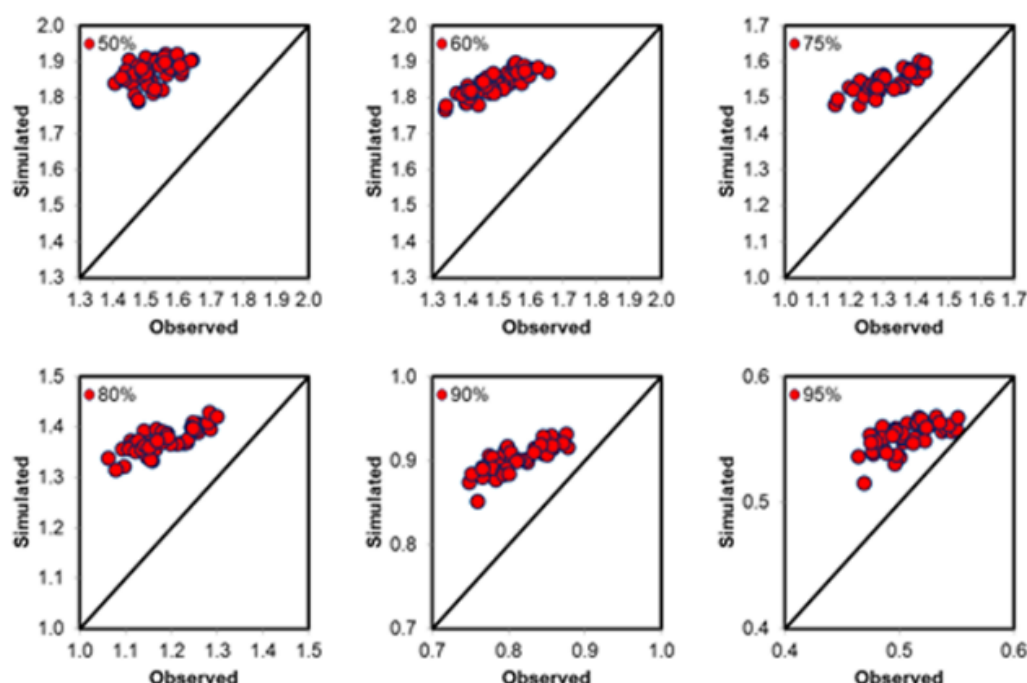


Fig. 4. Comparison of observed and simulated entropy for all triples combinations.

6 Conclusions

The results of the research show that the model reproduces synthetic daily precipitation having very similar statistical characteristics to the observed daily precipitation in terms of mean, variance, and cumulative frequency relative. There is only slight a difference value for all CCCs between simulated and observed daily precipitation. It is found that simulated daily precipitation exhibits lower CCCs compared to observed data. All simulated daily precipitation amounts for all thresholds give larger entropy than observed data. This fact indicates that simulated daily precipitation gives greater uncertainty or lower association than the observed one.

References

1. A. Bardossy, E.J. Plate. *Space-time model for daily rainfall using atmospheric circulation patterns*. Water Resources Research, **28** (5), pp. 1247–1259, (1992)
2. A. Bárdossy, GGS. Pegram. *Copula-based multisite model for daily precipitation simulation*. Hydrology and Earth System Sciences, **13** (12), p. 2299, (2009)

3. A. Bárdossy. *Generating precipitation time series using simulated annealing*. Water Resources Research, **34** (7), pp. 1737–1744, (1998)
4. A. Bárdossy, Pegram, Geoffrey. *Multiscale spatial recorrelation of RCM precipitation to produce unbiased climate change scenarios over large areas and small*. Water Resources Research, **48** (9), (2012)
5. T. Kawachi, T. Maruyama, V.P. Singh. *Rainfall entropy for delineation of water resources zones in Japan*. Journal of Hydrology, **246** (1-4), pp. 36–44, (2001)
6. F. Serinaldi, A. Bárdossy, CG. Kilsby. *Upper tail dependence in rainfall extremes. Would we know it if we saw it?*, Stochastic environmental research and risk assessment, **29** (4), pp. 1211–1233, (2015).
7. W. Wang, P Van Gelder, JK. Vrijling, Chen, X. *Detecting long-memory. Monte Carlo simulations and application to daily streamflow processes*. Hydrology and Earth System Sciences Discussions, **11** (2), pp. 851–862, (2007)

Suroso, 2018_ Multisite daily precipitation simulation in

ORIGINALITY REPORT

7%

SIMILARITY INDEX

1%

INTERNET SOURCES

6%

PUBLICATIONS

0%

STUDENT PAPERS

PRIMARY SOURCES

1	Suroso Suroso, András Bárdossy. "Investigation of asymmetric spatial dependence of precipitation using empirical bivariate copulas", Journal of Hydrology, 2018 Publication	2%
2	Deshpande A.V., Patil S.N.. "Assessment of groundwater quality by using statistical analysis from kopargaon taluka, Ahmednagar, India", International Journal of Advanced Geosciences, 2016 Publication	1%
3	Stuart Gatehouse, William Noble. "The Speech, Spatial and Qualities of Hearing Scale (SSQ)", International Journal of Audiology, 2009 Publication	1%
4	Jiří Stehlík, András Bárdossy. "Multivariate stochastic downscaling model for generating daily precipitation series based on atmospheric circulation", Journal of Hydrology, 2002 Publication	1%
5	W. Wang, P. H. A. J. M. van Gelder, J. K. Vrijling, X. Chen. "Detecting long-memory: Monte Carlo simulations and application to daily streamflow processes", Copernicus GmbH, 2006 Publication	<1%
6	www.iws.uni-stuttgart.de Internet Source	<1%

7	Ashok K. Mishra, Mehmet Özger, Vijay P. Singh. "An entropy-based investigation into the variability of precipitation", Journal of Hydrology, 2009 Publication	<1 %
8	András Bárdossy. "Generating precipitation time series using simulated annealing", Water Resources Research, 1998 Publication	<1 %
9	Engineering Risk in Natural Resources Management, 1994. Publication	<1 %
10	Hou-Liang Lu, Fang-Fang Li, Tongliang Gong, Yanhong Gao, Jianfeng Li, Jun Qiu. "Reasons behind seasonal and monthly precipitation variability in the Qinghai-Tibet Plateau and its surrounding areas during 1979–2017", Journal of Hydrology, 2023 Publication	<1 %
11	Hristos Tyralis, Georgia Papacharalampous, Apostolos Burnetas, Andreas Langousis. "Hydrological post-processing using stacked generalization of quantile regression algorithms: Large-scale application over CONUS", Journal of Hydrology, 2019 Publication	<1 %
12	Zengchao Hao, Vijay P. Singh. "Review of dependence modeling in hydrology and water resources", Progress in Physical Geography, 2016 Publication	<1 %
13	Andras Bardossy, Erich J. Plate. "Space-time model for daily rainfall using atmospheric circulation patterns", Water Resources Research, 1992 Publication	<1 %

14

Langousis, Andreas, and Vassilios Kaleris. "Statistical framework to simulate daily rainfall series conditional on upper-air predictor variables", Water Resources Research, 2014.

Publication

<1 %

15

Qing Zhang, Bhawan Singh, Sébastien Gagnon, Jean Rousselle, Noël Evora, Steven Weyman. "The Application of WGEN to Simulate Daily Climatic Data for Several Canadian Stations", Canadian Water Resources Journal, 2004

Publication

<1 %

Exclude quotes On

Exclude bibliography On

Exclude matches Off