

UC Irvine

UC Irvine Previously Published Works

Title

Conditional simulation of remotely sensed rainfall data using a non-Gaussian v-transformed copula

Permalink

<https://escholarship.org/uc/item/2bj4d4zv>

Journal

Advances in Water Resources, 33(6)

ISSN

03091708

Authors

AghaKouchak, Amir
BÃairdossy, AndrÃais
Habib, Emad

Publication Date

2010-06-01

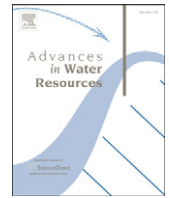
DOI

10.1016/j.advwatres.2010.02.010

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, availalbe at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed



Conditional simulation of remotely sensed rainfall data using a non-Gaussian v-transformed copula

Amir AghaKouchak^{a,*}, András Bárdossy^b, Emad Habib^a

^a Department of Civil Engineering, University of Louisiana at Lafayette, PO Box 42291, Lafayette, Louisiana, 70504, USA

^b Institute of Hydraulic Engineering, University of Stuttgart, Pfaffenwaldring 61, D-70569, Stuttgart, Germany

ARTICLE INFO

Article history:

Received 4 August 2009

Received in revised form 21 February 2010

Accepted 22 February 2010

Available online 1 April 2010

Keywords:

Copula

V-transformed copula

Multivariate simulation

Remotely sensed rainfall data

Ensemble generation

ABSTRACT

Quantification of rainfall and its spatial and temporal variability is extremely important for reliable hydrological and meteorological modeling. While rain gauge measurements do not provide reasonable areal representation of rainfall, remotely sensed precipitation estimates offer much higher spatial resolution. However, uncertainties associated with remotely sensed rainfall estimates are not well quantified. This issue is important considering the fact that uncertainties in input rainfall are the main sources of error in hydrologic processes. Using an ensemble of rainfall estimates that resembles multiple realizations of possible true rainfall, one can assess uncertainties associated with remotely sensed rainfall data. In this paper, ensembles are generated by imposing rainfall error fields over remotely sensed rainfall estimates. A non-Gaussian copula-based model is introduced for simulation of rainfall error fields. The v-transformed copula is employed to describe the dependence structure of rainfall error estimates without the influence of the marginal distribution. Simulations using this model can be performed unconditionally or conditioned on ground reference measurements such that rain gauge data are honored at their locations. The presented model is implemented for simulation of rainfall ensembles across the Little Washita watershed, Oklahoma. The results indicate that the model generates rainfall fields with similar spatio-temporal characteristics and stochastic properties to those of observed rainfall data.

© 2010 Elsevier Ltd. All rights reserved.

1. Introduction

Reliable direct measurements of ground surface rainfall can be obtained only at very limited spatial scales using rain gauge stations. However, rainfall is known to be highly variable in space and time [54,10,13]. In recent years, satellites, weather radar systems and remote sensing techniques have provided rainfall information at higher temporal and spatial resolutions than was previously possible from rain gauge measurements. Application of remotely sensed rainfall data in hydrologic and meteorological predictions has increased significantly in the past few decades. Despite extensive research, however, the uncertainties associated with remotely sensed rainfall data have not yet been well quantified [27,34,16,26,33,5,11]. Characterization and quantification of such uncertainties are extremely important as it is believed that spatial and temporal variability in input rainfall is one of the main sources of error in rainfall-runoff processes and hydrologic predictions [43,53,3,23].

Estimates of spatio-temporal uncertainties in remotely sensed rainfall data can be obtained by simulation of a reasonable rainfall

ensemble which consists of a large number of rainfall realizations, each of which represents an equiprobable rainfall event that can occur. Unlike geostatistical interpolation techniques [31,32], stochastic simulation models preserve the typical variability and fluctuation patterns of rainfall [15]. The generated rainfall ensembles can be used as input to hydrological and meteorological models to assess model prediction uncertainties. A great deal of effort has been made to develop stochastic models for simulation of multivariate rainfall fields (e.g., gauge, radar and satellite). In a number of studies, rainfall fields are directly simulated using different multivariate stochastic techniques (e.g., [40,58,28,37,45,9,6]). Ciach et al. [8] developed an operational approach based on empirical investigations of joint samples of radar and ground surface data whereby the radar rainfall uncertainties consist of a systematic distortion function and a stochastic component. The radar error components are then estimated using non-parametric estimation schemes. Based on the work presented by [8,56] developed a radar rainfall field generator and a model to produce probability of exceedance maps of radar rainfall conditioned on given radar rainfall estimates. Ensemble of rainfall fields can be obtained by imposing simulated random error fields over original rainfall estimates. [22] proposed that radar estimates can be perturbed with stochastic fields in order to obtain an ensemble of radar estimates. In a recent work, [2] presented a random error model in which radar fields are perturbed using two error components: a proportion error field to account for

* Corresponding author. Present address: Department of Civil & Environmental Engineering, University of California Irvine, Irvine, CA, USA.

E-mail address: amir.a@uci.edu (A. AghaKouchak).

errors that are proportional to the magnitude of rain rate and a purely random error field for the sum of errors from different sources.

A reasonable and robust model for rainfall simulation is expected to produce rainfall fields with similar spatial and temporal dependencies to those of the observed. It is well known that hydrological and meteorological data are dependent in both space and time. Traditional measures of dependence (e.g., Pearson linear correlation coefficient) involve only the second moment statistics and thus can capture limited aspect of the dependence among random variables. Copulas, on the other hand, can be employed to describe the dependencies among n random variables on an n dimensional unit cube (uniform). Description of the spatial dependence structure independent of the marginal distribution is one of the most attractive features of copulas. The application of copulas in simulation of multivariate data and nonlinear dependence structures has become popular in hydrological and meteorological studies. In recent years, a number of copulas-based models have been introduced for multivariate frequency analysis, risk assessment, geostatistical interpolation and multivariate extreme value analyses (see [14,21,42,4,49,20,60,46,1]). Using bivariate mixed distributions [48] and a copula-based Markov approach, [47] and [57] introduced a model for radar rainfall estimation uncertainties. The most recent applications of copulas in hydrology and meteorology are limited to multivariate Gaussian or bivariate non-Gaussian copulas. The well-known normal distribution can be extended to the multivariate normal as described in [36]; however, the Gaussian copula cannot capture the upper tail dependence that may exist among hydrometeorological data [41]. On the other hand, some, but not all, non-Gaussian copulas, such as the commonly used Archimedean copulas are restricted for large dimensions. In this paper, a multivariate non-Gaussian copula-based model is introduced for simulation of remotely sensed rainfall data. The v -transformed copula is used to describe the dependence structure of the rainfall data. Having described the dependencies using copulas, the empirical distribution function of observed rainfall error is numerically approximated and applied on the simulated error fields so that simulated realizations are similar to those of the observed in terms of the distribution function. The ensemble-based approach, presented here, can be applied for the simulation of radar estimates, satellite images and also multi-site rain gauge data. In an example, the model is implemented for the simulation of 4 km gridded Stage IV radar data, also known as Next Generation Weather Radar (NEXRAD) multi-sensor precipitation estimates (Hereinafter, radar estimates). The NEXRAD system is installed and maintained by the National Weather Service (NWS). It consists of a network of Weather Surveillance Radars – 1988 Doppler (WSR-88D) that covers the entire United States [59,12]. Using the presented model, ensembles of radar estimates are simulated with conditions based on available observations. The generated rainfall ensembles can then be used as input to hydrological and meteorological models to assess model prediction uncertainties. Subsequent runs of a hydrological or meteorological model using simulated realizations of radar fields would then allow an assessment of uncertainty propagation due to the precipitation input. In the following sections, after a brief review on the study area and data resources, the mathematical formulation of the model is explained in detail. The model is then implemented over a relatively large watershed in Oklahoma, USA.

2. Data and study area

The Little Washita watershed located in the southwestern part of the state of Oklahoma, USA was selected for this study. The climate of the region is classified as moist sub humid and the average annual rainfall of the watershed is approximately 760 mm, while the mean annual temperature has been determined to be 16°C [18]. The area of the watershed is approximately 611km² and the surface water drains into the Little Washita River which is a tributary of the Washita River. Stage IV radar estimates used in this study are provided by NCAR/EOL under sponsorship of the National Science Foundation. The data is available

with a temporal resolution of 1 hour and a spatial resolution of 4 × 4 km. The closest NEXRAD radar station is the Oklahoma City station located approximately 70 km from the center of the watershed.

Hydrological and meteorological variables of the watershed have been measured since the 1930s. The Micronet network, operated by the Agricultural Research Service (ARS), is located within and around the Little Washita boundaries. The Micronet network includes 42 rain gauge stations, which are almost uniformly distributed throughout the watershed. Fig. 1 shows the location of the watershed and the position of the rain gauges and radar pixels. In addition to the Micronet network, there are three stations of the Oklahoma Mesonet located within the study area that are not included in the analysis due to the fact that they are used in the radar precipitation estimation processing. The Micronet network is equipped with tipping-bucket gauges that record rainfall data with a temporal resolution of 5 min and an accuracy of 0.254 mm [59]. The 5 min accumulations are aggregated to hourly intervals in order to synchronize the rain gauge measurements with radar estimates. The reference surface rainfall data are then used to obtain estimates of the radar rainfall error across the study area. The differences between the reference surface rainfall data and the radar estimates are considered and termed as observed error. It is noted that the observed error is estimated after removing the overall bias of the rainfall estimates with respect to the rain gauge measurements as described in [8].

3. V-transformed copula

The v -transformed copula is obtained through a non-monotonic transformation of the multivariate Gaussian copula [4]. Eq. (1) describes the multivariate Gaussian copula with correlation matrix ρ as its only parameter [36]:

$$C_{\rho}^n(u_1, \dots, u_n) = \Phi_{\rho}^n(\Phi^{-1}(u_1), \dots, \Phi^{-1}(u_n)) \tag{1}$$

where: Φ_{ρ}^n = multivariate standard Gaussian distribution function.

The variable N is defined as an n -dimensional normal random variable with a mean of zero, unit standard deviation and correlation matrix ρ . The v -transformed copula can be derived with the following transformation [4]:

$$X_i = \begin{cases} k(N_i - m) & \text{if } N_i \geq m \\ m - N_i & \text{if } N_i < m \end{cases} \tag{2}$$

where: k and m = copula parameters

The univariate marginal distribution of X_i can be expressed as [4]:

$$F(x) = \Phi\left(\frac{x}{k} + m\right) - \Phi(-x + m) \tag{3}$$

with the density function being [4]:

$$f(x) = \frac{1}{k} \varphi\left(\frac{x}{k} + m\right) + \varphi(-x + m) \tag{4}$$

where Φ is the standard normal distribution and the term φ denotes the probability density function of the normal distribution:

$$\varphi = \frac{1}{\sqrt{2\pi\sigma^2}} \text{EXP}\left(-\frac{(x-\mu)^2}{2\sigma^2}\right). \tag{5}$$

There are many families of copulas developed for different practical contexts. Each family of copulas has a number of parameters to describe the dependencies. The main difference associated with different copulas is in the detail of the dependence they represent. For instance, various copula families may differ in the part of their distributions (upper tail/

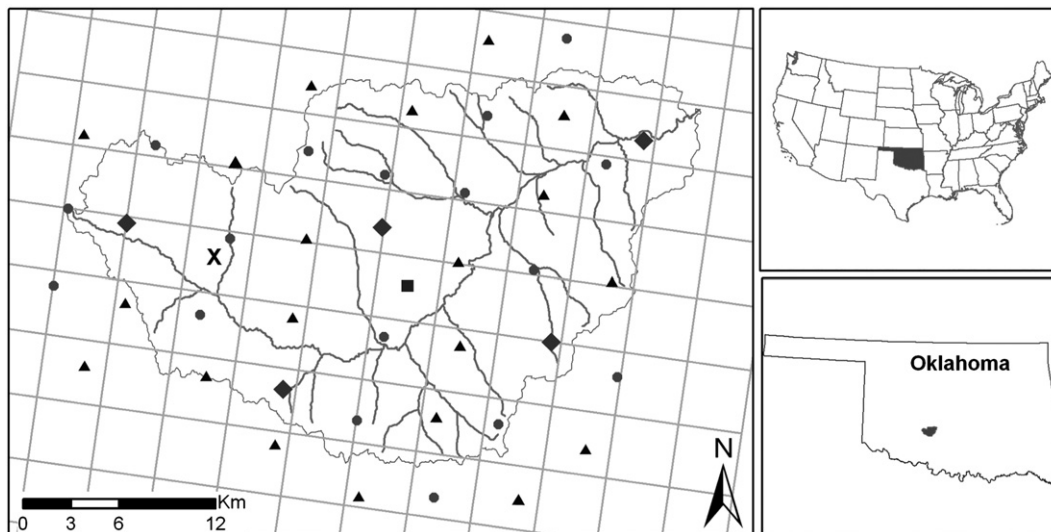


Fig. 1. Little Washita research watershed, Oklahoma, USA.

lower tail) where the association is strongest/weakest. Tail dependence describes the significance of the dependence in lower left quantile or upper right quantile of a multivariate distribution function. The upper tail expresses the probability occurrence of positive large values (extremes) at multiple locations jointly. The tail behavior depends solely on the type of copula and not on the choice of marginal distribution. Thus, in copula-based simulation, the type of copula strongly affects the tail dependence of simulated realizations. While the Gaussian copula does not have upper tail nor lower tail dependence, a number of non-Gaussian copulas have lower tail dependency, upper tail dependency, or even both [35]. Thus a non-Gaussian copula is used in this paper so that the model can capture the tail dependence more accurately. For additional information regarding different copula families, the reader is referred to [36] and [29].

4. Model description

As mentioned before, one way to account for uncertainties associated with rainfall estimates is to generate an ensemble of rainfall realizations that is a representation of possible variabilities in precipitation data. Multiple realizations of rainfall fields can be obtained by imposing random error fields over the original rainfall estimates. In this study, the following formulation is used to generate an ensemble of rainfall estimates:

$$R_s = R_i + R_i \times \varepsilon \quad (6)$$

where :

R_s	simulated rainfall (rain rate)
R_i	rain rate
ε	rainfall error

The term, $R_i \times \varepsilon$, represents the error component which is known to be proportional to the magnitude of rain rate [2,24,8]. Intuitively, one may expect that small rain rates are not subject to very large random errors. Similar behavior has been observed for measurement error of rating curves (e.g., [38]). Previous studies confirm that when error is proportional to the magnitude of the indicator variable (here, rainfall), the associated uncertainties can be modeled using a multiplicative error term (see [44] and [38]). A similar approach is adopted here by using a multiplicative component ($R_i \times \varepsilon$) to describe the uncertainties of remotely sensed precipitation estimates. This approach guarantees

that simulated values to be proportional to the magnitude of rain rate. That is, one can avoid unrealistically large errors where the rain rate is insignificant.

It should be noted that copula parameters are to be estimated based on the available observations which are typically obtained from pair of radar estimates and rain gauge measurements. Having estimated the parameters, the v -transformed copula is used for simulation of multivariate error fields with a similar dependence structure to that of the observed data. Using the method introduced here, radar error fields can also be simulated unconditionally (without conditioning on observations in the simulation process). Nonetheless, in either case, the simulated error fields are transformed by applying the empirical cumulative distribution function (CDF) of the observed rainfall error to the simulated error fields. This can be achieved by approximating the CDF of the observed error with a stairs function such that the step heights are simply the data that fall within the step. It is worth remembering that copulas are invariant to monotonic transformations and hence the simulated error fields will have the same spatial dependence structure after the transformation. This issue is illustrated in Fig. 2 using an example with synthetic data. Fig. 2(a) shows the correlation matrix (20×20) of 20 randomly simulated variables ($\xi \equiv (\xi_1, \dots, \xi_{20})$). Fig. 2(b) displays the empirical rank correlation [36] of the same variables (notice that Fig. 2(a) and (b) are not identical). Fig. 2(c) and (d) present the Pearson correlation matrix and empirical copula of the transformed variables ($\ln(\xi)$), respectively. As shown, by a logarithmic transformation of the variables the Pearson correlation matrix changes significantly (compare Fig. 2(a) and (c)), whereas the empirical copula remains unchanged (compare Fig. 2(b) and (d)).

Simulation of multivariate random fields can be performed based on any given copula. In order to obtain a simulated field of $x(x_1, \dots, x_n)$ with marginals of F_1, \dots, F_n , the following three steps are required:

1. Estimation of the copula (C^n) parameters.
2. Simulation of uniform random variables $u(u_1, \dots, u_n)$ using the copula C^n .
3. Transformation of the univariate marginals to F_1, \dots, F_n using Sklar's theorem ([50], see Appendix A; Eq. (10)): $x_i = F_i^{-1}(u_i)$.

As mentioned above, due to invariance of copulas to monotonic transformations, applying the marginals (Step 3) does not affect the underlying dependence structure. Furthermore, F_1, \dots, F_n do not necessarily need to have the same distribution family. This is a great advantage in simulation, as the variables may belong to different probability distribution families or may have significantly different

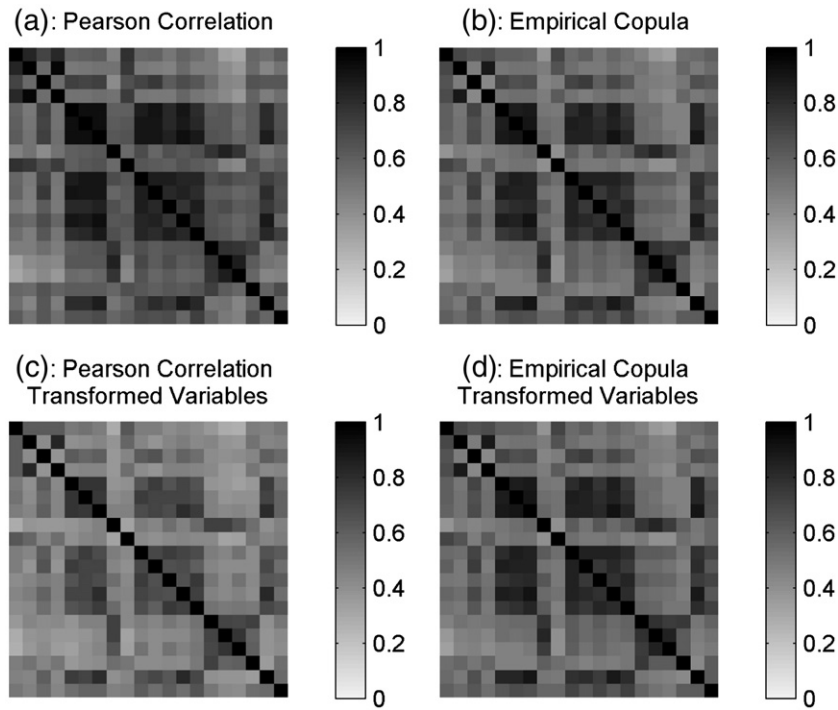


Fig. 2. Invariance of copulas to monotonic transformations: (a) Pearson correlation matrix of the data; (b) empirical copula of the data; (c) Pearson correlation matrix of the transformed data; and (d) empirical copula of the transformed data.

empirical distribution functions. In other words, using copulas and the Sklar theorem (see Appendix A), one can simulate random variables with the same probability distribution as that of the input data while preserving the dependence structure of the variables. This is one of the main attractive features of copulas in simulation of spatially dependent random fields. The interested reader is referred to [41] for a detailed overview of the step-by-step conditional simulation using copulas.

5. Case study

In the following, an example implementation of the model for the simulation of two-dimensional radar rainfall fields is presented. In the example, unconditional and conditional simulations of rainfall fields are demonstrated for three rainfall events occurring in August 2003, September 2005 and September 2006. During the first event, 42 rain gauges were available while during the second and third events, 22 and 20 rain gauges were in operation and available for analysis. Table 1 displays summary statistics including the mean, standard deviation, and 10 and 90% quantiles of the lumped rainfall accumulation for each storm.

Using the available rain gauge measurements and radar estimates, the parameters of the v-copula model are to be estimated first. The v-transformed copula is parameterized with the correlation matrix and two other parameters, *k* and *m*, that are used to transform the multivariate normal copula. In this study the parameter estimation method is based on [4] whereby the correlation matrix is described as a

parameter dependent on different distance vectors of *d*. Based on this assumption, the correlation matrix can be obtained using the spherical function. Here, the spherical function is defined with two sills (*s*₁ and *s*₂), two ranges (*r*₁ and *r*₂) and a nugget parameter (*s*₀) as shown in Eq. (7):

$$\begin{aligned} \gamma(x) &= \gamma(x_1) + \gamma(x_2) \\ &= s_0 + s_1 \left(\left(\frac{3x_1}{2r_1} - \frac{x_1^3}{r_1^3} \right) I_{(0,r_1)}(x_1) + I_{[r_1,\infty)}(x_1) \right) \\ &\quad + s_2 \left(\left(\frac{3x_2}{2r_2} - \frac{x_2^3}{r_2^3} \right) I_{(0,r_2)}(x_2) + I_{[r_2,\infty)}(x_2) \right) \end{aligned} \tag{7}$$

where the indicator function *I_D*(*x*) is 1 if *x* ∈ *D* and 0 otherwise. A likelihood approach is used where the likelihood function itself, consists of the product of multiple likelihood functions obtained from disjoint subsets of data (see [4] for details). Table 2 lists the parameters (*s*₀, *s*₁, *s*₂, *r*₁, *r*₂, *k* and *m*) of the v-transformed copula used for simulations. Notice that for each event, the parameters are estimated based on the seasonal observations (e.g., summer 2003, summer 2005) in which the event occurs. Using the estimated parameters, multiple rainfall fields are simulated for the selected storms. Fig. 3(a) presents a radar rainfall image occurring during Event 1 and Fig. 3(b) shows the corresponding rain gauge measurements. Fig. 3(c) and (d) display two realizations of conditionally simulated radar fields using the v-transformed copula. A visual comparison confirms that the model reproduces the rain gauge values at their locations.

Table 1
Summary statistics of rainfall accumulations for the selected storms.

Event	Date Time	Duration [h]	Mean [mm]	Standard Deviation	Q10 [mm]	Q90 [mm]
Event 1	8/30/03 07:00	12:00	4.3	5.5	0.8	8.8
Event 2	9/15/05 06:00	14:00	3.8	4.3	1.0	10.7
Event 3	9/17/06 15:00	10:00	10.1	15.5	1.0	22.7

Table 2
The parameters of the spherical function.

Event	<i>s</i> ₀	<i>s</i> ₁	<i>s</i> ₂	<i>r</i> ₁ [m]	<i>r</i> ₂ [m]	<i>k</i>	<i>m</i>
Event 1	0.01	0.14	0.43	3400	14,000	1.2	2.0
Event 2	0.03	0.12	0.41	2800	14,500	0.9	3.1
Event 3	0.01	0.11	0.53	3800	14,000	1.0	1.5

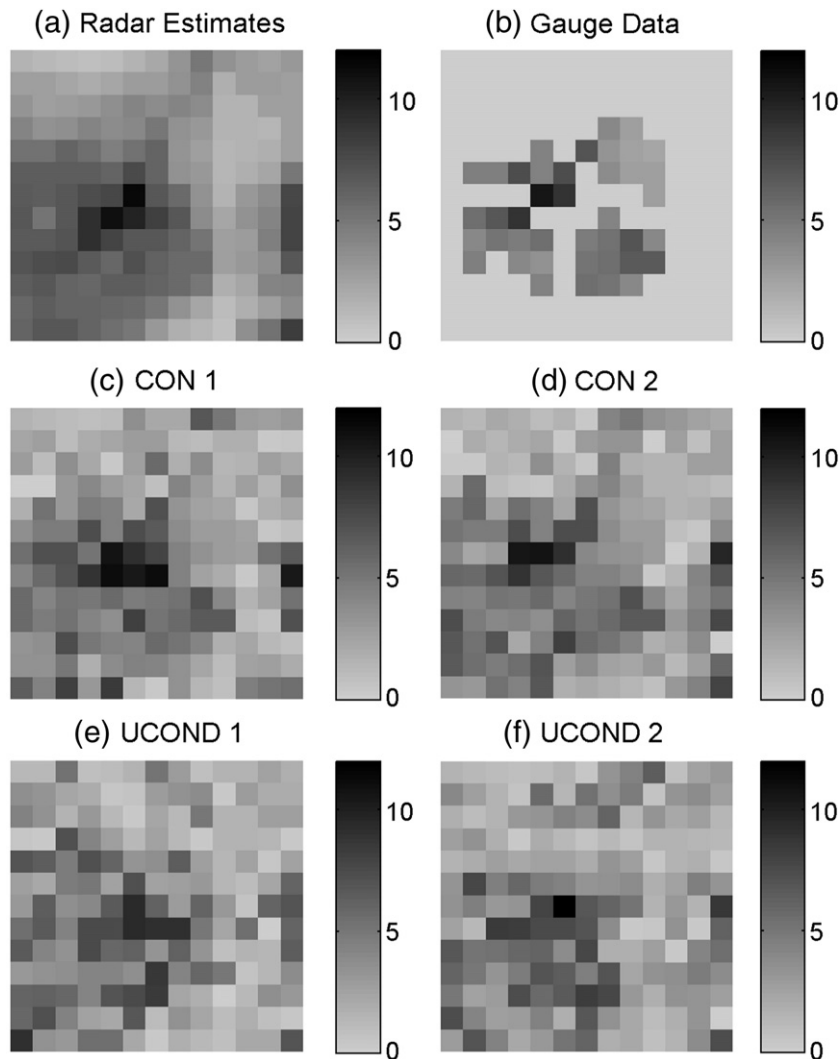


Fig. 3. Event 1: (a) one radar image; (b) corresponding gauge measurements; (c) and (d) two realizations of conditionally (CON) simulated rainfall fields; (e) and (f) two realizations of unconditionally (UNCOND) simulated rainfall fields.

The presented model can also be used without conditioning on observed errors estimated from rain gauge data and radar estimates. For illustration, two radar fields are simulated unconditionally using the aforementioned parameters and shown in Fig. 3(e) and (f). One can see that in Fig. 3(c) and (d), simulated fields at gauge locations show similar values to those of the rain gauge measurements, while in Fig. 3(e) and (f) ground reference measurements are not reproduced at their locations. Similarly, Figs. 4 and 5 present the results of conditional and unconditional simulations for Events 2 and 3, respectively. In general, conditional simulation is more desirable as it reflects reference measurements.

For Events 1 to 3, rainfall ensembles are obtained by imposing 500 conditionally simulated error fields over radar estimates. Fig. 6(a) to (c) plots radar estimates and simulated rainfall ensembles (500 realizations) over one radar pixel (the square-marked pixel shown in Fig. 1). The solid black line represents radar estimates and the gray lines are simulated rainfall realizations over the length of the storm, or in other words, estimated uncertainty associated with rainfall data. The estimated ensemble (gray area) describes the uncertainty associated with rainfall estimates. A statistical ensemble of rainfall includes a significantly large number of equiprobable rainfall realizations, where each realization could be a possible true observation. The figure shows that using the multiplicative error term helps to account for the proportionality of error to the magnitude of the rain rate. As mentioned previously, an attractive feature of this model is that the empirical

marginal distribution of the observed rainfall error is applied to simulated error fields so that simulated values will have the same distribution function as that of observed. This can be achieved when copulas are used for describing the dependencies as they are invariant to monotonic transformations (see Eq. (11)).

In order to validate the model, rainfall ensembles were generated with different numbers of gauges to evaluate whether or not the estimated uncertainty encloses rain gauge measurements if less numbers of gauges were available. For Event 1, Fig. 7(a) to (c) demonstrates estimated rainfall uncertainty over the x-marked pixel which contained one of the removed rain gauges. In Fig. 7(a) all available gauges were included, whereas in Fig. 7(b) and (c), 20 gauges (marked with triangle and diamond in Fig. 1) and 5 gauges (marked with diamond in Fig. 1) were used for simulations. The solid and dashed lines represent radar estimates and rain gauge measurements, respectively. The gray lines are simulated radar realizations over the entire storm. The results confirm that simulated realizations (estimated uncertainty) reasonably encompass ground reference measurements. It is noted that as the number of gauges decreases, the sample size of the observed error decreases. This may, but not necessarily, result in a shorter range of error (the difference between minimum and maximum error). One can intuitively conclude that any change in the error range, and thus the empirical distribution of error, will affect the estimated uncertainty. One can argue that if the empirical distribution functions of observed error do not change drastically as the gauges are reduced, the

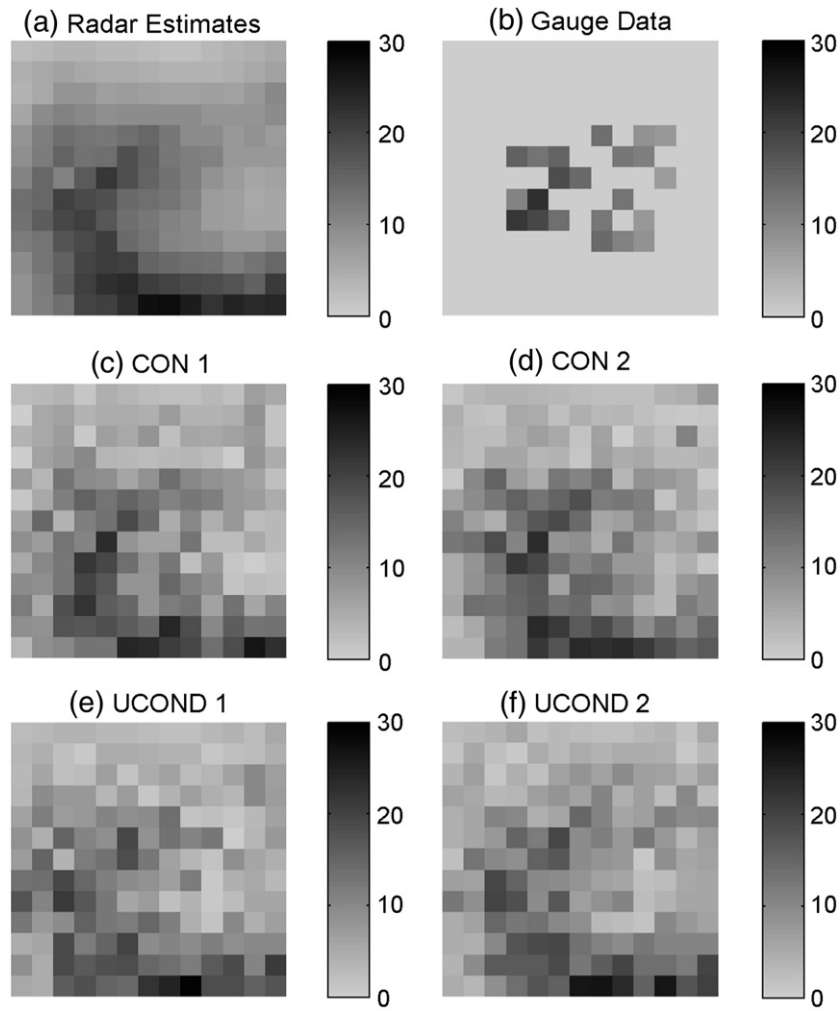


Fig. 4. Event 2: (a) one radar image; (b) corresponding gauge measurements; (c) and (d) two realizations of conditionally (CON) simulated rainfall fields; (e) and (f) two realizations of unconditionally (UNCOND) simulated rainfall fields.

distribution functions of simulated error fields remain similar, although their spatial dependence structure may be significantly different from each other. The generated ensembles (rainfall uncertainty) using different numbers of gauges are evaluated by counting the number of time steps (n_{out}) where the ground reference measurements do not fall inside the 95% confidence bounds of the estimated uncertainty. Table 3 lists the number of time steps (n_{out}) in percentage. As shown, reducing the available gauges to half (≈ 1 gauge in 31 km^2) results in an increase in the n_{out} of approximately 3% to 5%. Further reduction of gauges to 12% (≈ 1 gauge in 130 km^2), results in an increase of n_{out} to a maximum of 6.9%. Overall the table indicates that even with few rain gauges across the watershed, the estimated uncertainty obtained from the v-copula model reasonably encloses the ground reference measurements.

In most previous works, temporal and spatial dependencies of rainfall error fields were neglected, mainly due to unavailability of true representation of rainfall in space and time. [51] modeled rainfall random error assuming no spatial or temporal correlation beyond the size of one radar pixel ($4 \text{ km} \times 4 \text{ km}$). However, [30] argued that significant dependencies may exist in rainfall error fields both in space and time. [7] recognized this issue and indicated the need to account for error dependencies. [30] presented a random cascade model for simulation of error fields taking into account error correlations. The results confirmed that the dependencies in the error fields were non-negligible, particularly in space. By analyzing radar rainfall fields, [8] showed that the rainfall random error component was correlated in space and time and the estimated correlations were higher in the cold

season. Additionally, [24] discussed that the spatial dependence of radar error is not negligible and can have a significant effect on streamflow hydrologic predictions. [56] described the spatio-temporal dependencies of rainfall random error using the modified exponential function. They reported that temporal and spatial dependencies were captured reasonably well using the modified exponential function.

As mentioned before, in the present study, radar rainfall uncertainty is assumed to be correlated in space and the v-transformed copula is used for describing the dependencies. In order to evaluate the presented model with respect to spatial dependencies, the Spearman's ρ_s rank correlation matrix is used to assess the dependence structure of the simulated radar rainfall fields after imposing the error fields. The Rank correlation matrix, unlike Pearson product-moment correlation, evaluates the degree of association in terms of ranks, which is known as concordance. The Spearman rank correlation is a nonparametric method of describing the dependencies in terms of ranks and is independent of the marginal distributions [52,25]:

$$\rho_s = \frac{n \sum \chi_i \xi_i - \sum \chi_i \sum \xi_i}{\sqrt{n \sum \chi_i^2 - (\sum \chi_i)^2} \sqrt{n \sum \xi_i^2 - (\sum \xi_i)^2}} \quad (8)$$

where:

χ_i rank of x_i in X
 ξ_i rank of y_i in Y

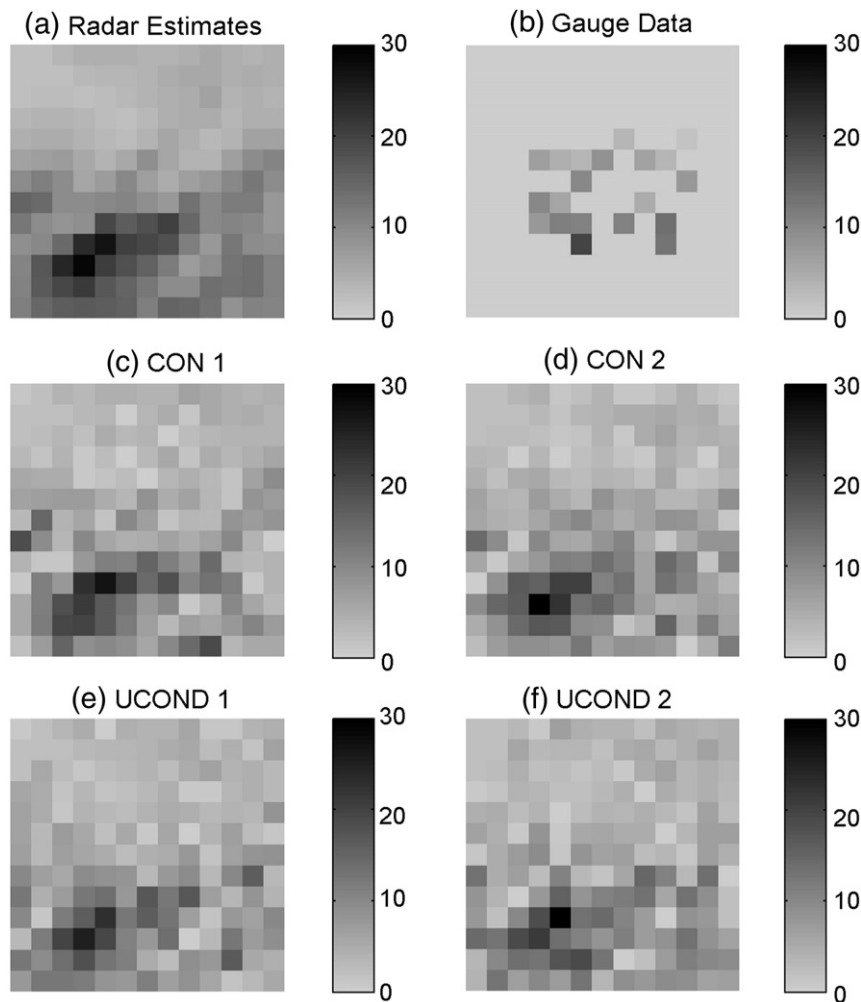


Fig. 5. Event 3: (a) one radar image; (b) corresponding gauge measurements; (c) and (d) two realizations of conditionally (CON) simulated rainfall fields; (e) and (f) two realizations of unconditionally (UNCOND) simulated rainfall fields.

Fig. 8(a) to (c) shows the Spearman correlation matrices of the rainfall events selected for this study. The figures show symmetrical matrices that describe the rank correlation between pairs of radar rainfall pixels. Fig. 8(d) to (f) demonstrates the correlation matrices of one set of simulated rainfall data after imposing the rainfall uncertainty on radar rainfall estimates. One can see that the rank correlation matrices of simulated rainfall fields are not considerably different than those of the radar rainfall estimates and the overall dependence structure is preserved.

In addition to spatial dependence structure, temporal dependencies of simulated rainfall fields are also tested. In a recent work, [24] reported that the temporal autocorrelations of the rainfall error fields were rather low at the first time lag and close to zero for larger time lags. For this reason, ε was assumed to be temporally uncorrelated in the formulation of the model (Eq. (6)). However, after imposing the uncertainty term over rainfall estimates, the autocorrelations of the rainfall field will be carried forward and simulated fields will have quite similar temporal autocorrelations. Fig. 9(a) to (c) compares the temporal autocorrelation of the radar estimates and the 500 simulated rainfall fields for the square-marked pixel in Fig. 1. One can see that simulated fields have slightly weaker temporal autocorrelations due to perturbation. However, the overall trend of the temporal autocorrelations in all cases are similar to those of the original rainfall estimates, indicating that the random error term added to the radar rainfall estimates did not destroy the underlying temporal autocorrelations. The figures confirm that even without explicitly accounting for temporal dependencies in the model employed here, the

simulated rainfall data have realistic temporal self-correlation characteristics.

A major difference associated with different copula families is the inherent tail dependence. A particular type of copula may provide strong upper tail dependence, whereas another copula may represent strong lower tail dependence. In the following, the v -copula model is tested for probability occurrence of extreme values. In this study, the 90th percentile of the radar error estimates is assumed as the extreme value threshold. In order to investigate the v -copula model with respect to extremes, occurrences of rainfall above the thresholds in the simulated fields are compared with those of the observations. Fig. 10(a) to (c) displays the sum of the number of occurrences above the 90th percentile threshold (solid black lines), the mean of the number of occurrences above the 90th percentile in 500 realizations simulated using the v -copula (dashed lines) and the number of occurrences above the threshold in each simulated rainfall error realization (gray lines). In the figures, the x-axis shows the number of simulated realizations (here 500) and the y-axis represents the number of occurrences above the threshold (90th percentile) of the observations. One can see that for Events 1 and 2 the v -copula model slightly over estimates the number of joint occurrences, which indicates that the v -copula shows a positive tail dependence.

6. Summary and concluding remarks

It is well known that remotely sensed rainfall data are subject to different types of errors that affect the quality of rainfall estimates. We

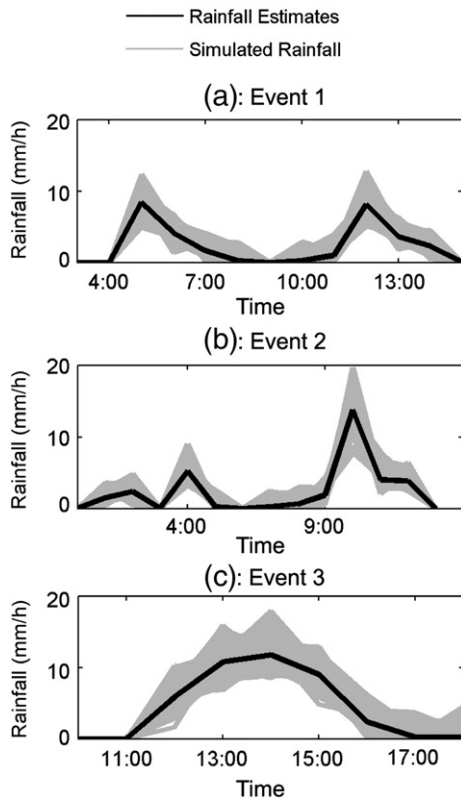


Fig. 6. 500 realizations of conditionally simulated rainfall data over the square-marked pixel in Fig. 1.

presented a non-Gaussian copula-based model for simulation of remotely sensed rainfall data through imposing simulated error fields over remotely sensed rainfall estimates. In this paper, Stage IV radar data are used as the input of the model to generate an ensemble of rainfall estimates. Similarly, satellite data and reflectivity fields can also be used as the input. In this model, the v -transformed copula was employed to describe the dependence structure of the rainfall data without the influence of the marginal distribution. Spatial dependencies of the simulated radar rainfall fields were then investigated by calculating their rank correlation matrix. The results revealed that using this copula-based model, one can preserve the spatial dependencies of simulated rainfall fields. It should be noted that no explicit accounting for error temporal autocorrelation was included in the model. However, the results showed that the trend of temporal autocorrelations in simulated fields will be similar to the temporal autocorrelation of the rainfall estimates, mainly due to the fact that the temporal autocorrelations of the rainfall estimates were dominant.

The issue of accounting for spatial dependencies using copulas requires further investigations, as the choice of copula itself plays an important role [35,55,19]. Using cross validation and by taking the mean absolute error (MAE) as the estimator, the fitted copula was tested based on the available observations. Cross-validation is a well known approach that can be used to evaluate fitted models to observations and also to compare performances of different predictive modeling procedures Picard and Cook [39]; Efron and Tibshirani [17]. Table 4 summarizes the results of the repeated random sub-sampling cross validation Picard and Cook [39] for the fitted v -copula to the observations. For the case where only 5 reference gauges were used in simulations, the leave-one-out cross-validation (LOOCV) was employed due to the limited observations (see column 4 in Table 4). In this method, a single observation from the original sample is put aside for validation, and the remaining observations are used as the

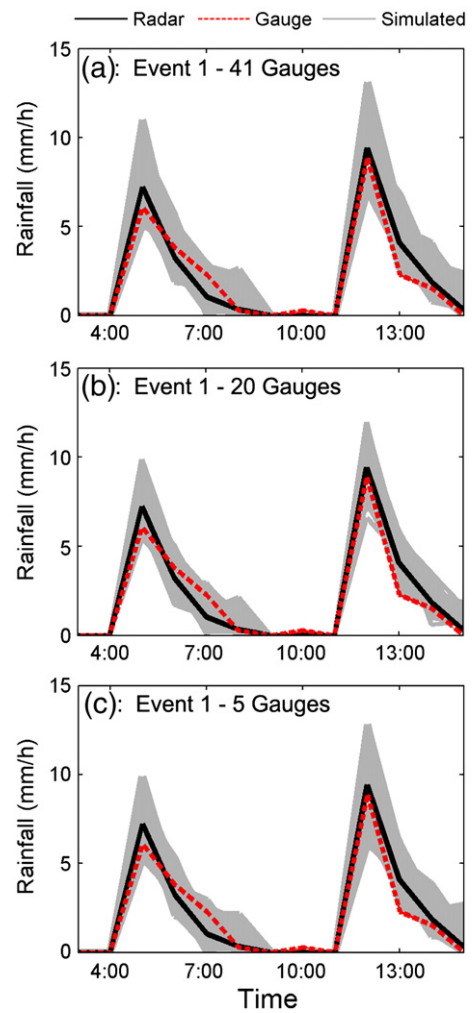


Fig. 7. Estimated rainfall uncertainty over the x -marked pixel when: (a) all available gauges; (b) 20 gauges (marked with triangle and diamond in Fig. 1); and (c) 5 gauges (marked with diamond in Fig. 1) were used for simulations.

training data. This is repeated for the number of observations such that each observation is used as the validation data once. It is noted that the mean absolute error does not change considerably as the number of gauges are reduced. Overall, it seemed that the v -transformed copula was an appropriate choice for modeling spatial dependencies. However, the parameter estimation and simulation process is computationally demanding. Other copula families should also be investigated to evaluate their reliability and robustness.

In most rainfall simulation models, a standard distribution function is fit to the data and used for simulation in order to simplify the process. Often, the normal distribution is used for its simplicity in terms of parameter estimation and simulation of multivariate fields [2]. In this work, however, the empirical marginal distribution of the observed rainfall error is applied so that simulated

Table 3

The number of time steps, n_{out} (%), that the 95% uncertainty bound did not enclose the rain gauge measurements.

	All Gauges	50% of Gauges	12% of Gauges
Event 1	1.7	4.7	6.4
Event 2	1.9	3.6	6.9
Event 3	1.5	3.0	5.7

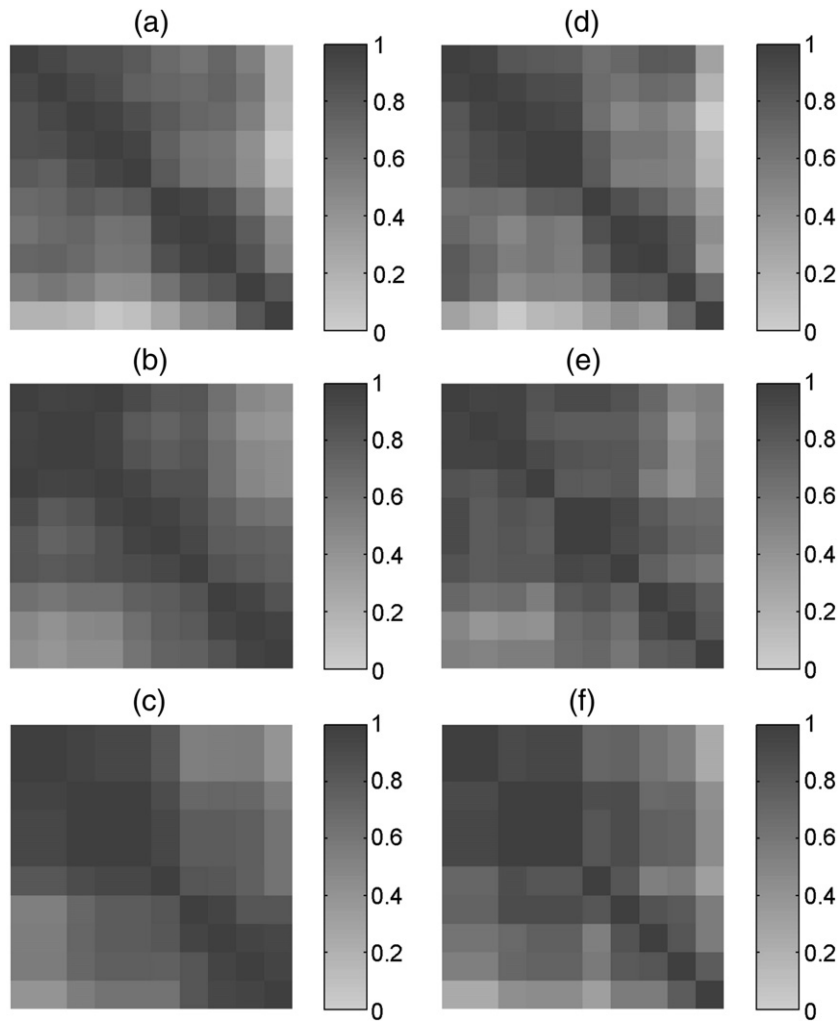


Fig. 8. Spearman rank correlation matrices of (a) to (c): rainfall Events (1) to (3); (d) to (f): one set of simulated data for Events (1) to (3).

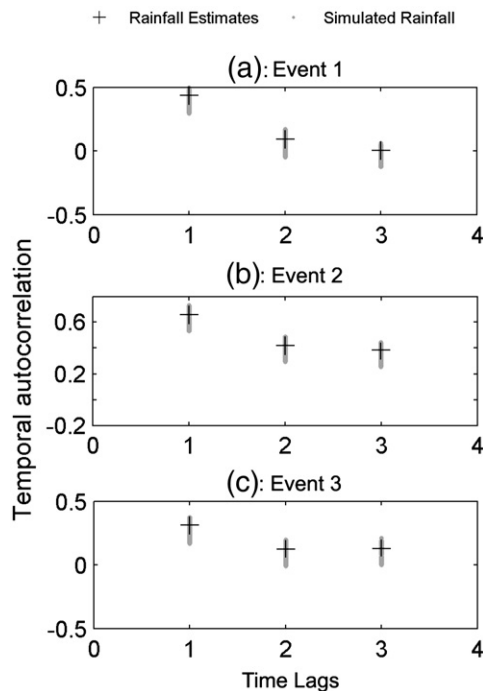


Fig. 9. Temporal autocorrelation of rainfall estimates and 500 simulated rainfall realizations for the square-marked pixel.

error values will have the same distribution function as that of the observed. This can be achieved using copulas, as they are invariant to monotonic transformations (see Eq. (11)). The presented model was tested through investigation of the marginal distribution functions of observed and simulated rainfall error using the Two-sample Kolmogorov-Smirnov test. The results confirmed that the empirical distribution function of the observed error was reasonably reproduced (figures not included). The model was validated by generating rainfall ensembles with different numbers of gauges to investigate the estimated uncertainty when fewer gauges were available. The number of gauges used for simulations were reduced to 20 and 5 gauges (approximately 1/2 and 1/8 of the total number of measurement stations) and rainfall uncertainty was estimated using the selected gauges. Verification of 95% confidence bounds of the estimated uncertainty showed that even with few gauges the estimated uncertainty reasonably encompassed the ground reference observations.

It is noted that in this work, we have used one individual gauge in a pixel to represent the true area average rainfall over a pixel size of (4 km × 4 km), which may not be very accurate. However, based on the available rain gauge stations, this was the best possible approximation of the true area average rainfall values. Unreliable rain gauge measurements may result in inaccurate true area average rainfall representation and thus errors in parameter estimation. Consequently, this will affect the simulated rainfall fields, especially

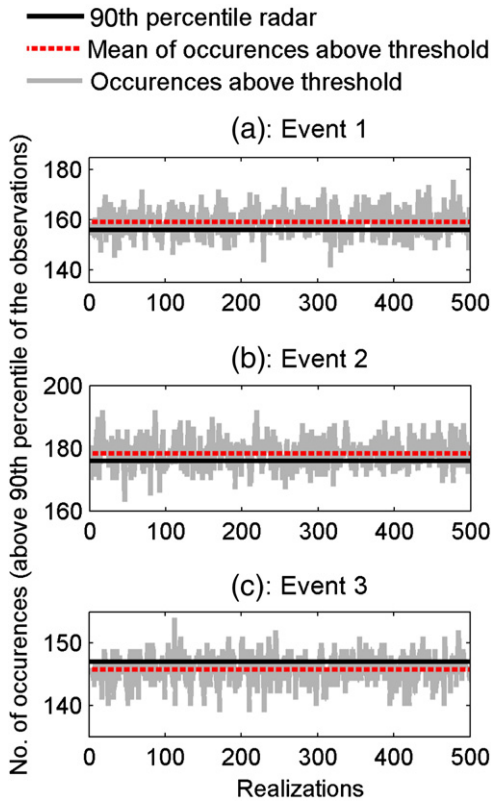


Fig. 10. Total number of occurrences above the 90th percentile of radar estimates (solid black lines), mean number of occurrences above the same threshold in 500 realizations (dashed lines) and the number of occurrences above the 90th percentile of radar estimates in each simulated realizations.

Table 4
The results of cross validation for the v-copula model (MAE estimator).

Event ID	All Gauges	50% of Gauges	12% of Gauges
Event 1	0.90	0.91	2.40
Event 2	0.72	1.14	2.44
Event 3	1.06	1.56	2.66

in the case of conditional simulation. Overall, the results presented here indicated that one can generate ensembles of rainfall fields through perturbation of remotely sensed rainfall data with simulated random error fields. Subsequent simulations of a hydrologic model using a large number of generated radar rainfall fields can be used for the purpose of investigating error propagation in modeling hydrologic processes and also for uncertainty analysis of the model with respect to the input rainfall. In order to verify the robustness of the model, further research including simulations over different spatio-temporal scales and different gauge network setups are required. Additionally, in order to evaluate the effects of sampling error on simulated rainfall fields, the sensitivity of the presented model to the number of gauges available needs to be investigated. Various issues regarding this model are still being investigated by the authors to ensure the validity of the model's performance and its statistical robustness.

Acknowledgments

This research was funded by a grant provided to the third author from the Research Competitiveness Subprogram of the Louisiana Board of Regents Support Fund.

Appendix A. Theoretical Background

Copulas are joint cumulative distribution functions that describe dependencies among variables independent of their marginals ([29] and [36]):

$$C^n(u_1, \dots, u_n) = Pr(U_1 \leq u_1, \dots, U_n \leq u_n) \tag{9}$$

where C is an n -dimensional joint cumulative distribution function of a multivariate random vector $U \equiv (U_1, \dots, U_n)$ whose marginals are $u [0,1]$. Note that throughout this paper, a common statistical convention is used in which uppercase characters denote random variables and lowercase characters are their specified variables. [50] showed that each continuous multivariate distribution $F(F_1, \dots, F_n)$ can be represented with a unique copula C :

$$F(x_1, \dots, x_n) = C^n(F_1(x_1), \dots, F_n(x_n)). \tag{10}$$

The copula can capture the dependence structure of multivariate distributions. Having described the dependencies using a copula, a transformation function can be applied to each variable in order to transform the marginal distribution into the desired marginals [36]:

$$C^n(u_1, \dots, u_n) = F(F_1^{-1}(u_1), \dots, F_n^{-1}(u_n)) \tag{11}$$

where $F(F_1, \dots, F_n)$ is the multivariate cumulative distribution function (CDF) with marginals F_1, \dots, F_n belonging to different distribution families. The n -th partial derivative of an n -dimensional copula C^n is termed as copula density (for derivations, the reader is pointed to [36] and [19]):

$$c^n(u_1, \dots, u_n) = \frac{\partial^n}{\partial u_1 \cdot \partial u_2 \dots \partial u_n} C^n(u_1, \dots, u_n) \tag{12}$$

References

- [1] AghaKouchak A, Bárdossy A, Habib H. Copula-based uncertainty modeling: application to multi-sensor precipitation estimates. *Hydrol Process* 2010;24, doi: 10.1002/hyp. 7632.
- [2] AghaKouchak A, Habib E, Bárdossy A. Modeling radar rainfall estimation uncertainties: a random error model. *J Hydrol Eng* 2009, doi:10.1061/(ASCE)HE.1943-5584.0000185.
- [3] Arnaud P, Bouvier C, Cisner L, Dominguez R. Influence of rainfall spatial variability on flood prediction. *J Hydrol* 2002;260:216–30.
- [4] Bárdossy A, Li J. Geostatistical interpolation using copulas. *Water Resour Res* 2008;44(7):W07412.1–W07412.15, doi:10.1029/2007WR006115.
- [5] Bastin L, Fisher P, Wood J. Visualizing uncertainty in multi-spectral remotely sensed imagery. *Comput Geosci* 2002;28(3):337–50.
- [6] Bras R, Rodriguez-Iturbe I. Rainfall generation: a nonstationary time varying multidimensional model. *Water Resour Res* 1976;12:450–6.
- [7] Carpenter T, Georgakakos K. Impacts of parametric and radar rainfall uncertainty on the ensemble streamflow simulation. *J Hydrol* 2004;298:242–66.
- [8] Ciach G, Krajewski W, Villarini. Product-error-driven uncertainty model for probabilistic quantitative precipitation estimation with nexrad data. *J Hydrometeorol* 2007;8:1325–47.
- [9] Clark M, Gangopadhyay S, Brandon D, Werner K, Hay L, Rajagopalan B, et al. A resampling procedure for generating conditioned daily weather sequences. *Water Resour Res* 2004;40(4):W04304.1–W04304.15, doi:10.1029/2003WR002747.
- [10] Corradini C, Singh V. Effect of spatial variability of effective rainfall on direct runoff by geomorphologic approach. *J Hydrol* 1985;81:27–43.
- [11] Crosetto M, Ruiz J, Crippa B. Uncertainty propagation in models driven by remotely sensed data. *Remote Sens Environ* 2001;76(3):373–85.
- [12] Crum T, Albany R. The wsr-88d and the wsr-88d operational support facility. *Bull Am Meteorol Soc* 1993;74:1669–87.
- [13] Dawdy D, Bergman J. Effect of rainfall variability on streamflow simulation. *Water Resour Res* 1969;5:958–66.
- [14] De Michele C, Salvadori J. A generalized pareto intensity-duration model of storm rainfall exploiting 2-copulas. *J Geophys Res Atmos* 2003;108(D2):1–11, doi: 10.1029/2002JD002534.
- [15] Deutsch V, Journel A. GSLIB – Geostatistical Software Library and User's Guide. Oxford University Press; 1998.
- [16] Dong J, Walker J, Houser P. Factors affecting remotely sensed snow water equivalent uncertainty. *Remote Sens Environ* 2005;97(1):68–82.
- [17] Efron B, Tibshirani R. Improvements on cross-validation: The. 632 + bootstrap method. *J Am Stat Assoc* 1990;92(438):548–60.

- [18] Elliott R, Schiebe F, Crawford K, Peter K, Puckett W. A unique data capability for natural resources studies. Paper No. 932529, International Winter Meeting, Chicago, IL: American Society of Agricultural Engineers; 1993. Dec. 14–17.
- [19] Embrechts P, Lindskog F, McNeil A. Modelling Dependence with Copulas and Applications to Risk Management. Zurich, Switzerland: Department of Mathematics, ETHZ; 2001.
- [20] Gebremichael M, Krajewski W. Application of copulas to modeling temporal sampling errors in satellite-derived rainfall estimates. *J Hydrol Eng* 2007;12(4):404–8.
- [21] Genest C, Favre A, Béliveau J, Jacques C. Metaelliptical copulas and their use in frequency analysis of multivariate hydrological data. *Water Resour Res* 2007;43(9):W09401.1–W09401.12, doi:10.1029/2006WR005275.
- [22] Germann U, Berenguer M, Sempere-Torres D, Salvade G. Ensemble radar precipitation – a new topic on the radar horizon. Proceeding of the 4th European Conference on Radar in Meteorology and Hydrology ERAD, Barcelona, Spain, 18–22 September 2006; 2006.
- [23] Goodrich D, Faures J, Woolhiser D, Lane L, Sorooshian S. Measurement and analysis of small-scale convective storm rainfall variability. *J Hydrol* 1995;173:283–308.
- [24] Habib E, Aduvala A, Meselhe E. Analysis of radar rainfall error characteristics and implications for streamflow simulations uncertainty. *J Hydrol Sci* 2008;53(3):568–87.
- [25] Hollander M, Wolfe D. Nonparametric Statistical Methods. New York: Wiley; 1973.
- [26] Hossain F, Anagnostou E. Numerical investigation of the impact of uncertainties in satellite rainfall estimation and land surface model parameters on simulation of soil moisture. *Adv Water Resour* 2005;28(12):1336–50.
- [27] Hossain F, Huffman G. Investigating error metrics for satellite rainfall data at hydrologically relevant scales. *J Hydrometeorol* 2008;9(3):563–75.
- [28] Hutchinson M. Stochastic space time weather models from ground based data. *Agric For Meteorol* 1995;73:237–64.
- [29] Joe H. Multivariate Models and Dependence Concepts. London: Chapman Hall; 1997.
- [30] Jordan P, Seed A, Weinmann P. A stochastic model of radar measurement errors in rainfall accumulations at catchment scale. *J Hydrometeorol* 2003;4(5):841–55.
- [31] Journel A, Huijbregts C. Mining geostatistics. New York: Academic Press; 1978.
- [32] Krajewski W, Smith J. Sampling properties of parameter estimation for a storm field rainfall model. *Water Resour Res* 1989;25(9):2067–75.
- [33] Lucieer A, Kraak M. Interactive and visual fuzzy classification of remotely sensed imagery for exploration of uncertainty. *Int J Geogr Inf Sci* 2004;18(5):491–512.
- [34] Margulis S, Entekhabi D, McLaughlin D. Spatiotemporal disaggregation of remotely sensed precipitation for ensemble hydrologic modeling and data assimilation. *J Hydrometeorol* 2006;7(3):511–33.
- [35] Melchiori M. Which Archimedean Copula is the right one? Third Version, YeldCurve e-Journal, September 2003; 2003.
- [36] Nelsen R. An Introduction to Copulas (Springer Series in Statistics). New York: Springer Verlag; 2006.
- [37] Pegram G, Clothier A. Downscaling rainfields in space and time using the string of beads model in time series mode. *Hydrol Earth Syst Sci* 2001;5(2):175–86.
- [38] Petersen-Øverleir A. Accounting for heteroscedasticity in rating curve estimates. *J Hydrol* 2004;292:173–81.
- [39] Picard R, Cook D. Cross-validation of regression models. *J Am Stat Assoc* 1997;79(387):575–83.
- [40] Rajagopalan B, Lall U. A k-nearest neighbour simulator for daily precipitation and other weather variables. *Water Resour Res* 1999;35:3089–101.
- [41] Salvadori G, DeMichele C, Kottegoda N, Rosso R. Extremes in Nature: an approach using copulas. Berlin: Springer; 2007.
- [42] Schölzel C, Friederichs P. Multivariate non-normally distributed random variables in climate research – introduction to the copula approach. *Nonlinear Process Geophys* 2008;15(5):761–72.
- [43] Schuurmans J, Bierkens M. Effect of spatial distribution of daily rainfall on interior catchment response of a distributed hydrological model. *Hydrol Earth Syst Sci* 2007;11:677–93.
- [44] Seber G, Wild C. Nonlinear Regression. Chichester: Wiley; 1998.
- [45] Seed A, Srikanthan R. A space and time model for design storm rainfall. *J Geophys Res* 1999;104(D24):31623–30.
- [46] Serinaldi F. Analysis of inter-gauge dependence by Kendall's τ , upper tail dependence coefficient, and 2-copulas with application to rainfall fields. *Stoch Env Res Risk A* 2008;22:671–88.
- [47] Serinaldi F. Copula-based mixed models for bivariate rainfall data: an empirical study in regression perspective. *Stoch Env Res Risk A* 2009;23(5):677–93, doi:10.1007/s00477-008-0249-z.
- [48] Shimizu K. A bivariate mixed lognormal distribution with an analysis of rainfall data. *J Appl Meteorol* 1993;32:161–71.
- [49] Singh V, Zhang L. Idf curves using the frank archimedean copula. *J Hydrol Eng* 2007;12:651–62.
- [50] Sklar A. Fonctions de répartition à n dimensions et leurs marges. *Publications de l'Institut de Statistique de L'Université de Paris* 8; 1959. p. 229–31.
- [51] Smith J, Shed R, Walton M. Parameter estimation for the nexrad hydrology sequence. Preprints, 24nd Conference on Radar Meteorology. Tallahassee, FL: American Meteorological Society; 1989. p. 259–63.
- [52] Spearman C. The proof and measurement of association between two things. *Am J Psychol* 1904;15:72–101.
- [53] Syed K, Goodrich D, Myers D, Sorooshian S. Spatial characteristics of thunderstorm rainfall fields and their relation to runoff. *J Hydrol* 2003;271:1–21.
- [54] Troutman BM. Runoff prediction errors and bias in parameter estimation induced by spatial variability of precipitation. *Water Resour Res* 1983;19(3):791–810.
- [55] Venter G. Tails of copulas. Proceedings of the Casualty Actuarial Society, LXXXIX 2002; 2002.
- [56] Villarini G, Krajewski W, Ciach G, Zimmerman D. Product-error-driven generator of probable rainfall conditioned on wsr-88d precipitation estimates. *Water Resour Res* 2009;45(1):W01404.1–11, doi:10.1029/2008WR006946.
- [57] Villarini G, Serinaldi F, Krajewski W. Modeling radar-rainfall estimation uncertainties using parametric and non-parametric approaches. *Adv Water Resour* 2008;31(12):1674–86, doi:10.1016/j.advwatres.2008.08.002.
- [58] Wilks D. Multisite generalization of a daily stochastic precipitation generation model. *J Hydrol* 1998;210:178–91.
- [59] Young C, Bradley A, Krajewski W. Evaluating nexrad multisensor precipitation estimates for operational hydrologic forecasting. *J Hydrometeorol* 2000;1(3):241–54.
- [60] Zhang L, S.R., Singh V. Bivariate rainfall frequency distributions using archimedean copulas. *J Hydrol* 2008;32(1–2):93–109.