

Copula-based downscaling of spatial rainfall: a proof of concept

M. J. van den Berg¹, S. Vandenberghe¹, B. De Baets², and N. E. C. Verhoest¹

¹Laboratory of Hydrology and Water Management, Ghent University, Coupure links 653, 9000 Ghent, Belgium

²Department of Applied Mathematics, Biometrics and Process Control, Ghent University, Coupure links 653, 9000 Ghent, Belgium

Received: 23 December 2010 – Published in Hydrol. Earth Syst. Sci. Discuss.: 13 January 2011

Revised: 18 April 2011 – Accepted: 2 May 2011 – Published: 11 May 2011

Abstract. Fine-scale rainfall data is important for many hydrological applications. However, often the only data available is at a coarse scale. To bridge this gap in resolution, stochastic disaggregation methods can be used. Such methods generally assume that the distribution of the field is stationary, i.e. the distribution for the entire (fine-scale) field is the same as the distribution of a smaller region within the field. This assumption is generally incorrect and we provide a proof of concept of a method to estimate the distribution of a smaller region. In this method, a copula is used to construct a bivariate distribution describing the relation between the scales. This distribution is then used to estimate the distribution of the fine-scale rainfall within a single coarse-scale pixel, by conditioning on the coarse-scale rainfall depth.

1 Introduction

Hydrological models are used in a variety of small-scale applications, including flood forecasting (Winchell et al., 1998), water management (Varis et al., 2004), and land slide risk assessment (Collison et al., 2000). These models often simulate hydrological processes and water fluxes at a small scale ($<100 \text{ km}^2$), requiring rainfall data with a high spatio-temporal resolution. However, such data is not available for many parts of the world as rain-gage networks are lacking in coverage and General Circulation Models (GCMs) and remote sensing cannot provide data at the required resolution (Onof et al., 1998; Deidda, 2000). This lack in resolution results in a scale gap where observations are only available at a coarser scale, but not at the required finer scale.

The scale gap between the observed data and the hydrological model can be mitigated using stochastic methods. This is done by investigating the coarse-scale rainfall from which parameters for a suitable synthetic rainfall model are derived. A first approach consists of representing rainfall as a stochastic process of cluster-based rectangular pulses, and found its origin in the mathematical description of rainfall, starting with a seminal paper by Le Cam (1961). This has led to the development of rainfall models based on Neyman-Scott and Bartlett-Lewis point processes (Onof et al., 2000). These models, which recently have been expanded into the spatial domain (e.g. Cowpertwait et al., 2002; Burton et al., 2008), use parameters that have a physical interpretation, such as the mean intensity and the mean duration of a raincell (Onof et al., 2000).

A second approach to bridge the scale gap between observed rainfall and the hydrological model applies cascade models. In contrast to the Neyman-Scott and Bartlett-Lewis models, cascade models do not use physically-based variables, but rather a physically relevant model structure: the (multiplicative) cascade. These models are derived from statistical fractals (Lovejoy and Mandelbrot, 1985) and constitute the first spatial synthetic rainfall models (Schertzer and Lovejoy, 1987). Further model development (Deidda et al., 1999) resulted in spatio-temporal synthetic rainfall models. One advantage of these models over the Neyman-Scott and Bartlett-Lewis models is that the resulting cascade-based synthetic data better represent the spatial structure of rainfall, which may prove important in small-scale applications (Willems, 2001).

A third approach consists of purely stochastic methods. These models are not based on physically-based concepts or constraints, but rather attempt to model the rainfall process by stochastic methods. Spatial variants of these methods



Correspondence to: M. J. van den Berg
(martinus.vandenberg@ugent.be)

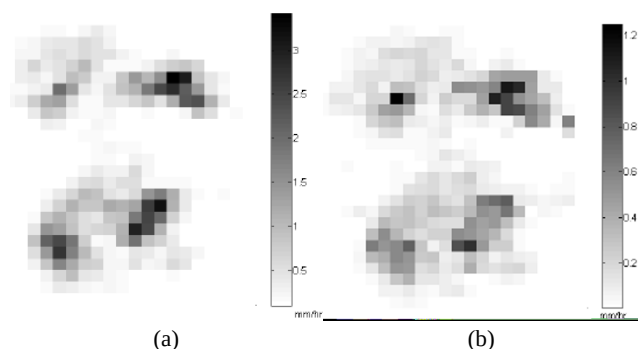


Fig. 1. A coarse-scale image of the rainfall depth (mm) (a), and its corresponding sub-pixel standard deviation (b).

include Markov Random Fields (Mackay et al., 2001; Allcroft and Glasbey, 2003), filtered autoregressive processes, and the superimposing of random fields (Ferraris et al., 2003a). Examples of the application of these methods include Rebora et al. (2006) and Pegram and Clothier (2001) where an autoregressive model is used to generate a random field which is passed through a non-linear filter to obtain a rainfall field that has similar spatial patterns as the large-scale rainfall field. For a more elaborate introduction to the various methods of simulating rainfall, we refer to Onof et al. (2000), and Ferraris et al. (2003a).

Many different downscaling methods exist, however, the above three categories give a brief summary of the most common methods. All these methods can be used to downscale a rainfall field, however, forcing the model to some observed coarse-scale data has proven to be difficult. Generally, only the last group of methods offers a way to take into account other assumptions and forcing data (Mackay et al., 2001; Rebora et al., 2006). Here, the specific point of interest is the use of small-scale sub-pixel distributions to force the fine-scale field to follow the coarse-scale field. For example, Mackay et al. (2001) use a Gamma distribution for subgrid variability, and Rebora et al. (2006) use a single log-normal distribution across the entire field. However, a cursory glance at a rainfall field shows that it is likely that the local variability is not stationary throughout the field, and that it varies with the depth of the coarse-scale field (see Fig. 1). In this paper, we present a method to estimate the subgrid variability distribution needed in the downscaling of coarse-scale rainfall. Therefore, a novel method is explored that describes the dependence between the coarse- and the fine-scale rainfall depth using a copula. Furthermore, it allows for modeling the sub-pixel probability distribution within one coarse-scale pixel.

This paper is structured as follows. The data used for empirical testing is introduced first. Section 3 presents the framework of the copula-based methodology, and Sect. 4 introduces some basics on copulas. Subsequently, the frame-

Table 1. The seven events for which data was available, together with the number of rainy hours on that day, the average depth (mm) for all active pixels. The ratio expresses the proportion of dry to wet pixels within the radar image.

Date (dd-mm-yyyy)	Rainy Hours	Average Depth	Ratio
29-08-2003	14	0.713	0.765
18-11-2004	6	0.498	1.07
29-06-2005	14	1.23	0.742
29-07-2005	7	1.11	0.738
10-09-2005	6	0.914	0.951
03-08-2006	8	0.867	0.558
09-08-2007	19	0.567	1.07

work is tested and validated in Sect. 5. Finally, the results are discussed and conclusions are drawn.

2 Data

The data for this study were acquired by a weather radar near Wideumont, Belgium, operated by the Belgian Royal Meteorological Institute (RMI). The installation covers a circular area with a radius of 240 km, producing a scan every 5 min. The region covered includes coastal landscapes to the west, and a low mountain range, the Ardennes, to the east with land cover mostly composed of forests, urbanization and agriculture. The entire region has a temperate climate and receives about 800 mm of rain annually, almost uniformly distributed throughout the year (De Jongh et al., 2006) and a mean monthly temperature which varies between 18 °C in June and 3 °C in January.

The dates, at which radar imagery is available, are selected based on rain-gage readings from a network across Belgium, where the rainfall was required to have a minimum Peak-over-Threshold return period of 10 yr to be included in the dataset. This resulted in a dataset consisting of the days listed in Table 1. For these days, the number of rainy hours has been listed, along with the average intensity for each rainstorm, and the ratio of dry pixels to wet pixels. Images that did not display rainfall were removed from the dataset as they hold no information.

The raw radar data are stored as digital numbers, representing reflectivities ranging from −31.5 dB to 95.5 dB in steps of 0.5 dB. Because of the 0.5 dB step, conversion of these reflectivities into rainfall rates results in discrete rainfall intensity values. Such data is likely to result in repeated values (often referred to as ties) which can lead to problems later on in the analysis. Similar to Vandenberghe et al. (2010b), uniform random noise ranging from −0.25 dB to 0.25 dB is introduced to perturb the data and to overcome the problem of ties. This perturbed data is then converted

into rainfall intensities using the Marshall-Palmer relationship (Marshall and Palmer, 1948)

$$R = \sqrt[b]{\frac{10^{0.1 \cdot Z_{\text{dB}}}}{a}}, \quad (1)$$

where Z_{dB} is the reflectivity [dB] and a and b are dimensionless parameters, respectively equal to 200 and 1.6 as suggested by Heylen and Maenhout (1994). This results in a set of spatially distributed rainfall intensities, each representing a surface over which the estimate is valid. However, the resolution of the image is not constant but changes with distance from the radar station (Heylen and Maenhout, 1994). As this is likely to influence the analysis, the data has been re-sampled to a grid of square pixels of $600 \text{ m} \times 600 \text{ m}$ and the 12 images within one hour are summed to form the total depth for a single hour.

The processed data was artificially downgraded to obtain the coarse-scale images with a pixel size of $19.2 \text{ km} \times 19.2 \text{ km}$. Each coarse-scale pixel is obtained by spatially averaging over the rainfall depths of the 1024 fine-scale pixels (hereafter referred to as sub-pixels) covered by a single coarse-scale pixel (Ferraris et al., 2003a; Deidda, 2000). Thus, the relation between the coarse scale field D_λ and the fine scale field $D'_{\lambda'}$, can be expressed as

$$D_\lambda(i, j) = \frac{1}{n^2} \sum_{k=(i-1)n+1}^{in} \sum_{l=(j-1)n+1}^{jn} D'_{\lambda'}(k, l), \quad (2)$$

where i and j are location indices at the coarse scale, k and l denote the location at the finer scale and n is the number of rows (and columns) of sub-pixels within one coarse-scale pixel.

3 Framework

Downscaling of rainfall fields at the image scale, as it is generally done in literature, does not specify a local cumulative distribution function (CDF) for the sub-pixels belonging to a coarse-scale pixel. It does not model the local fluctuations in the fine-scale distribution of the entire field. However, one can expect that the distribution of the local fine-scale rainfall, e.g. within a single coarse-scale pixel, will vary across the entire rainfall field. These local distributions may deviate (strongly) from the total field fine-scale distribution, and it can be expected that the shape of these distributions will depend on the corresponding coarse-scale observation, among other factors, as can be concluded from Fig. 1. When the coarse-scale rainfall depth is shown together with the corresponding sub-pixel standard deviation field, the relation between the distribution and the coarse-scale depth becomes evident. Therefore, to model these varying sub-pixel probability density functions, one should account for the observed

rainfall depth in the coarse-scale pixel by considering the dependence between the observed coarse-scale depth and the corresponding fine-scale sub-pixels.

In this paper, a framework is introduced which allows for the derivation of the modeled cumulative distribution function of the sub-pixel rainfall depth at a small scale, given a coarse-scale rainfall depth; this framework is depicted in Fig. 2. Paramount in this framework is the use of a copula to model the dependence between the coarse- and fine-scale rainfall depths given their marginal distribution functions.

Within the framework, the coarse-scale distribution function is first fitted to the entire rainfall image (Fig. 2e). The field-wide fine-scale cumulative distribution function (Fig. 2c) is then derived from the coarse-scale cumulative distribution function through scaling laws. Further, the dependence between the coarse- and fine-scale variables can then be described by means of a copula $\mathbf{C}$ (Fig. 2d):

$$F_{\lambda'\lambda}(d_{\lambda'}, d_\lambda) = \mathbf{C}(F_{\lambda'}(d_{\lambda'}), F_\lambda(d_\lambda)) = \mathbf{C}(u, v), \quad (3)$$

which is obtained through fitting to a set of corresponding coarse- and fine-scale wet pixels. Note that the upper case D has been replaced by the lower case d to denote the use of a single pixel; also, in this context the spatial organization of the field is disregarded and thus they are referred to as variables. $\mathbf{C}$ is a copula, a bivariate distribution function on the unit square with uniform marginals (see Sect. 4). A copula, together with the coarse- and fine-scale marginal distribution functions and their inverses, can describe the joint probability distribution. The validity of this approach is assured by Sklar's theorem (see Nelsen, 2006), which states that for every bivariate distribution with continuous marginal distributions, there exists a unique copula.

The modeled sub-pixel distribution of a wet coarse-scale pixel with observed depth d_m can be calculated from the joint probability distribution. To do this, the copula is conditioned to the value $v_m = F_\lambda(d_m)$, leading to the conditional cumulative distribution function (Fig. 2a), calculated as:

$$F_{U|V}(u|v_m) = \mathbb{P}(U \leq u | V = v_m) = \left. \frac{\partial \mathbf{C}(u, v)}{\partial v} \right|_{v=v_m}. \quad (4)$$

In this equation U and V correspond to the cumulative probabilities of $D_{\lambda'}$ and D_λ , i.e. the depth at the coarse and fine scale. Then using the fine-scale marginal distribution, U can be transformed into its fine-scale rainfall depth $D_{\lambda'}$ (Fig. 2c) (e.g., $x' = F_{\lambda'}^{-1}(u')$). This is possible by virtue of Eq. (3) which shows that corresponding values of U and $D_{\lambda'}$ have the same probability of occurrence, i.e.

$$F_{U|V}(u|v_m) = F_{\lambda'|\lambda}(d_{\lambda'}|d_m). \quad (5)$$

Hence, through a transformation using the cumulative distribution function (CDF) at the fine scale, the modeled fine-scale sub-pixel distribution, as given in Fig. 2b, is obtained. Moreover, the resulting distribution function is specific to the coarse-scale value d_m .

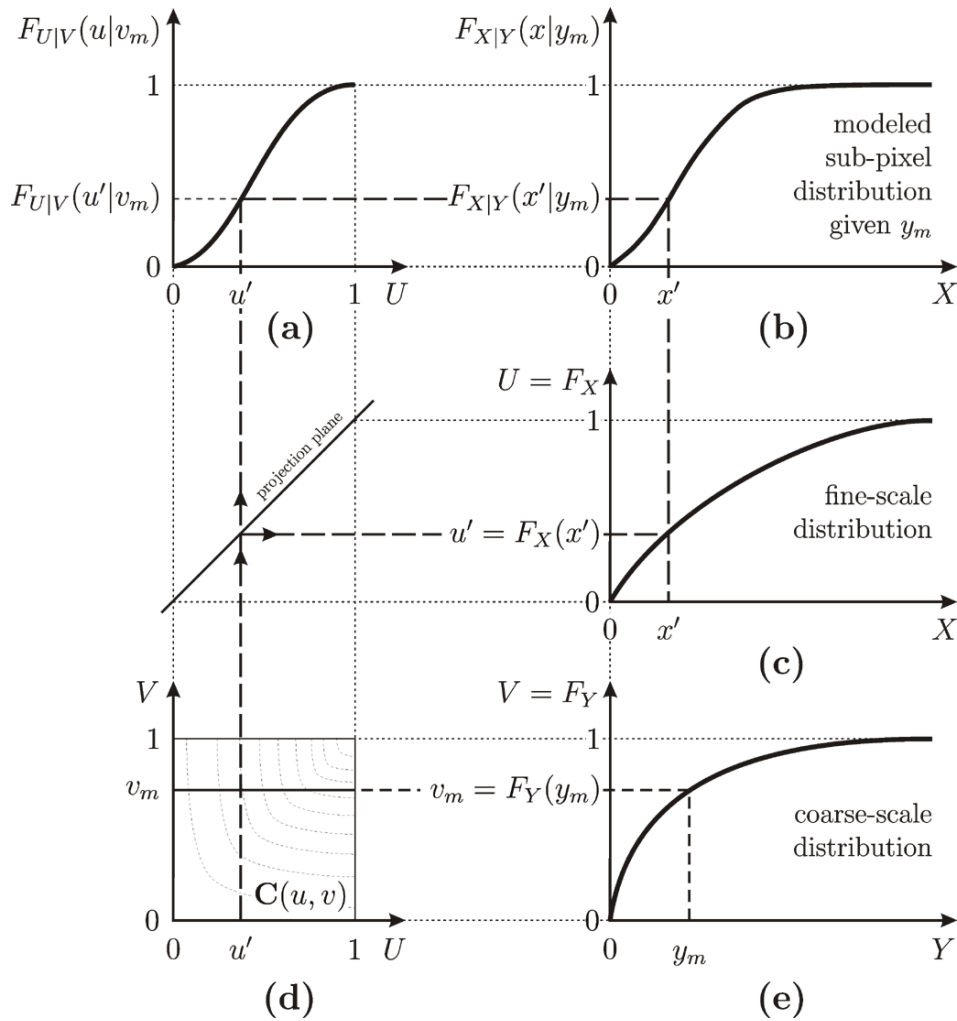


Fig. 2. A graphical representation of the framework used to retrieve the sub-pixel distribution. In the lower left, the copula is depicted, and in the top left, a slice of the copula representing $P(U \leq u, V \leq v)$. Starting at the coarse-scale distribution, one obtains the probability (rank) of the value used for conditioning. This value is used in the copula to find the conditional probability of the fine-scale ranks, which are then 'projected' using the fine-scale distribution to find the conditional probability of the sub-pixel rainfall depths.

The cumulative distribution function $\mathbb{P}(U \leq u | V = v_m)$ only represents the wet regions of the rainfall field. To extend this to the entire rainfall field within a coarse-scale pixel, including the intermittency, the fraction of dry fine-scale pixels should be accounted for. This fraction is modeled using an exponential function

$$f(v) = e^{av}, \quad (6)$$

which was fitted to several coarse-scale pixels. In Fig. 4 the relation between the fraction of dry pixels and the coarse-scale rainfall depth is depicted, as well as the fit of Eq. (6). Note that this function is not probabilistic in nature, but describes the relationship between the fraction of dry pixels $P(X = 0)$ and the coarse-scale value. Finally, the sub-pixel

distribution, including zero-valued pixels, is found as

$$F_{U|V}^+(u|v_m) = f(v_m)H(u) + (1 - f(v_m))F_{U|V}(u|v_m), \quad (7)$$

where $H(u)$ is the Heaviside function which is equal to one for all $u \geq 0$ and zero elsewhere.

4 Copulas

Copulas have gained increased attention in hydrological studies, including investigations into rainfall (Serinaldi, 2009; De Michele and Salvadori, 2003; Vandenberghe et al., 2010b; Grimaldi and Serinaldi, 2006), storm generators (Kao and Govindaraju, 2008; Salvadori and De Michele, 2006; Vandenberghe et al., 2010a) and data assimilation (Gao et al.,

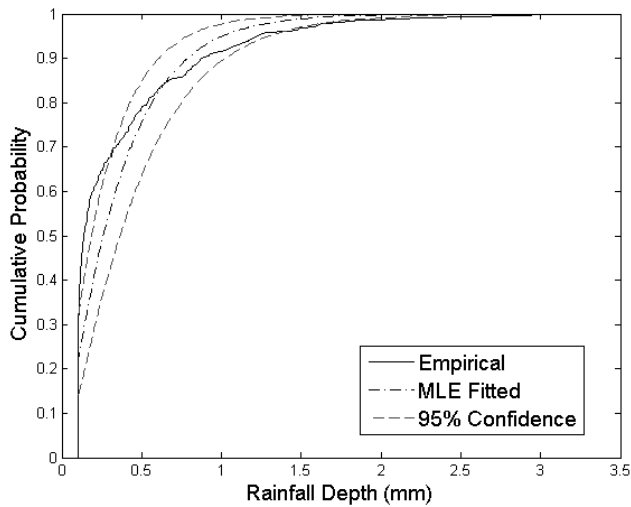


Fig. 3. The coarse-scale empirical marginal distribution and the fitted coarse-scale marginal distribution.

2007). This increased popularity can be attributed to the flexibility copulas offer for describing multivariate dependence (Genest and Favre, 2007).

The increased tractability offered by the use of copulas is due to the fact that the fitting of a bivariate distribution function is split into two more manageable problems, the marginal distribution functions on the one hand, and the dependence between them on the other hand. To illustrate this, a simple example in the context of this paper is presented. For more mathematical details, the reader is referred to Nelsen (2006), Genest and Favre (2007) and Salvadori and De Michele (2007).

Consider two scale levels, a fine-scale level and a coarse-scale level, respectively with resolutions λ' and λ , both of the same rainfall field. These fields are represented as couples $(d_\lambda, d_{\lambda'})$, where a single coarse-scale pixel d_λ can have several corresponding fine-scale pixels $d_{\lambda'}$, i.e. each coarse scale pixel leads to N couples where N is the number of fine-scale pixels within the coarse-scale pixel. Assume that both scale levels are observable and that their rainfall depths are described by cumulative distribution functions F_λ and $F_{\lambda'}$. These marginal distribution functions map the random variables to the unit interval $\mathbb{I}$, according to:

$$\begin{cases} u = F_{\lambda'}(d_{\lambda'}) \\ v = F_\lambda(d_\lambda) \end{cases} \iff \begin{cases} d_\lambda = F_\lambda^{[-1]}(u) \\ d_{\lambda'} = F_{\lambda'}^{[-1]}(v) \end{cases}, \quad (8)$$

where d denotes a single pixel from the field D , and u and v are single pixels from the probability-transformed fields U and V . $F_\lambda^{[-1]}$ denotes the pseudo-inverse, which corresponds to the inverse if it is defined everywhere on $\mathbb{R}$. Since the transformed fields U and V are both uniformly distributed,

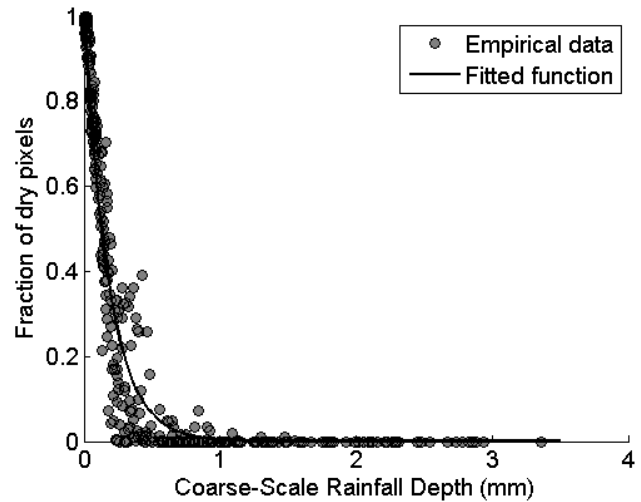


Fig. 4. The fraction of zero rainfall cells at fine scale compared to the coarse-scale value.

their pixels u and v can be approximated according to:

$$u = \frac{\sum_{j=1}^n \mathbf{I}(d_{\lambda'}^j < d_{\lambda'})}{n}, \quad (9)$$

$$v = \frac{\sum_{j=1}^n \mathbf{I}(d_\lambda^j < d_\lambda)}{n},$$

for all pixels d in D , where $\mathbf{I}()$ is the indicator function (Genest and Favre, 2007). The superscript j is an index number identifying the pixels within a field.

The bivariate copula of the transformed variables u, v is a function $\mathbf{C} : \mathbb{I} \times \mathbb{I} \rightarrow \mathbb{I}$, i.e. it maps values from the unit square $\mathbb{I} \times \mathbb{I}$ to the unit interval $\mathbb{I}$. Moreover, it satisfies the following conditions:

– for all $u, v \in \mathbb{I}$

$$\mathbf{C}(u, 0) = 0, \mathbf{C}(0, v) = 0, \mathbf{C}(u, 1) = u, \mathbf{C}(1, v) = v. \quad (10)$$

– for all $u_1, u_2, v_1, v_2 \in \mathbb{I}$ for which $u_1 \leq u_2$ and $v_1 \leq v_2$

$$\mathbf{C}(u_2, v_2) - \mathbf{C}(u_2, v_1) - \mathbf{C}(u_1, v_2) + \mathbf{C}(u_1, v_1) \geq 0. \quad (11)$$

These conditions imply that $\mathbf{C}$ is an increasing function. The link between bivariate copulas and bivariate distribution functions is expressed by the theorem of Sklar (see Nelsen, 2006) as given in Eq. (3). This theorem states that a bivariate copula results in the same probability as the multivariate distribution function, however it is calculated from the uniform marginal distribution functions of the variables U and V .

The significance of Sklar's theorem is that the dependence between D_λ and $D_{\lambda'}$ can be described independently from their marginal distribution functions; this simplifies the calculations as the new marginal distributions are uniform and

parameter free. Furthermore, the copula is invariant under strictly increasing transformations of the marginals and the bivariate distribution function is no longer dependent on the marginals allowing any (continuous) distribution functions to be joined with the copula. In this paper, the coarse- and fine-scale marginals are coupled to form a joint distribution function using the dependence model provided by a copula. A large number of different copulas exist, covering a wide range of different behaviors. For example, the tail behavior is difficult to capture using data and often described by means of a parametrical copula. Also, more common copulas exist such as the Gaussian copula which, when combined with Gaussian marginals, describes the bivariate Gaussian joint probability distribution function.

However, fitting a parametrical copula has proven to be non-trivial, and research is still ongoing (Schölzel and Friderichs, 2008). Because of this, an empirical, non-parametrical copula provides an appealing parsimony, and retains all information found in the data as well. Therefore, we opted to work with the empirical copula, although future research will investigate the use of parametrical copulas in the proposed framework. The function $\mathbf{C}$ can be empirically represented by the joint rank of the couples (u_i, v_i) , given by

$$\mathbf{C}_n(u, v) = \frac{1}{n} \sum_{i=1}^n \mathbf{I}(u_i \leq u, v_i \leq v). \quad (12)$$

This equation calculates the cumulative probability of each point in the dataset, representing the theoretical copula $\mathbf{C}$ as a set of empirical points $\mathbf{C}_n(u, v)$.

5 Results

In this section, the application of the framework is discussed and validated. First, the scaling laws and the empirical copula will be shown to be effective for modeling the relations observed in the data. Then, several practical issues that were observed will be explained.

5.1 Assessment of the scaling laws

The distribution of spatial rainfall fields is often represented by means of a Gamma distribution (Wilks, 1999; Vrac and Naveau, 2007; Mackay et al., 2001). This distribution is also used in this study and exhibits a good fit (see Fig. 3). Moreover, if the rainfall depth follows the Gamma distribution at coarse scale λ and shows simple scaling, or fractal behavior (Gupta and Waymire, 1990), then the distribution at the fine scale λ' is known as well. This relation will be used to downscale the field-wide distribution in this study and is defined as

$$t \cdot \Gamma(k, \theta_\lambda) = \Gamma(k, t \cdot \theta_\lambda). \quad (13)$$

$\Gamma(k, \theta_\lambda)$ denotes the Gamma distribution with shape parameter k and scale parameter θ_λ . t is a variable factor that describes the relation between scale levels λ and λ' . Furthermore, while k is constant over all scales, the scale parameter changes with each scale such that $\theta_{\lambda'} = t \cdot \theta_\lambda$. As mentioned before, the scaling of the Gamma distribution is tied to fractal behavior (Gupta and Waymire, 1990). Therefore, in order to find t , the fractal scaling laws need to be examined (see Veneziano et al., 2006; Gupta and Waymire, 1990; Menabde et al., 1999).

Consider the coarse-scale rainfall field D_λ whose scaling behavior is fractal (Veneziano et al., 2006; Ferraris et al., 2003b). Then, the scaling behavior of this field can be described as (Veneziano et al., 2006)

$$D_\lambda \stackrel{d}{=} r^{-H} D_{r\lambda}. \quad (14)$$

H is a constant over all scales, the scaling factor $r = \lambda'/\lambda$ and $\stackrel{d}{=}$ denotes equality in distribution. Since $D_\lambda \stackrel{d}{=} \Gamma(k, \theta_\lambda)$, Eq. (14) becomes

$$r^H \Gamma(k, \theta_\lambda) \stackrel{d}{=} \Gamma(k, \theta_{\lambda'}), \quad (15)$$

and thus $t = r^H$. The value of H can be found by computing the expected value on both sides of Eq. (15):

$$r^H E[\Gamma(k, \theta_\lambda)] = E[\Gamma(k, \theta_{\lambda'})], \quad (16)$$

where $E[\cdot]$ denotes the expectation of the distribution. By log transforming Eq. (16), one obtains:

$$H \cdot \log_{10}(r) = \log_{10} \left(\frac{E[\Gamma(k, \theta_{\lambda'})]}{E[\Gamma(k, \theta_\lambda)]} \right). \quad (17)$$

Hence, H can be found as the slope of the linear fit to this relation. This relation has been fitted to a series of ten randomly selected images, and the results are shown in Fig. 5. The fitted line shows a good consistency with the results, and the resulting $H \approx -0.10$ has been used throughout this study; this value is consistent with results obtained by Gupta and Waymire (1990). Summarizing, the marginal distribution of the coarse-scale image is determined by fitting a Gamma distribution to the empirical CDF. From this coarse-scale image, the fine-scale CDF is determined by applying Eq. (15) with $H = -0.10$.

5.2 Construction of the empirical copula

Once the marginal distributions of the rain fields on both scales are obtained, the dependence structure between the coarse-scale pixels and their corresponding sub-pixels needs to be determined. Since this paper is considered a proof of concept, an empirical copula will be used to describe the dependence, as a parametrical copula may introduce errors due to an improper representation of the actual dependence. However, the empirical copula, if it is to represent the dependence properly, should be constructed from carefully chosen

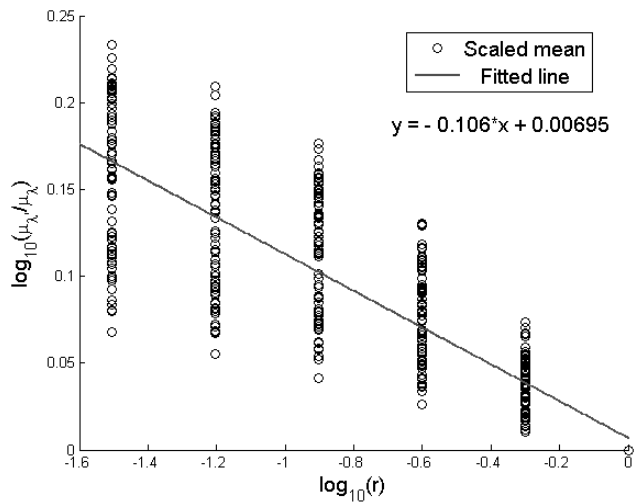


Fig. 5. Regression analysis between $\log_{10}(r)$ and $\log_{10}(\frac{\mu_{\lambda'}}{\mu_{\lambda}})$, where $\mu_{\lambda} = E[\Gamma(k, \theta_{\lambda})]$.

data. Moreover, issues such as not representing every possible point in the domain and noise need to be taken into account. In the following paragraphs, the practical application and implementation of the empirical copula in our framework will be outlined.

The marginal distributions, as well as the copula itself, need to be continuous. However, rainfall cannot be fully described by a single, continuous distribution because of the intermittency (Onof et al., 1998), which causes a singularity at zero rainfall in the probability distribution function. Therefore, rainfall is often described as a discrete (binomial) distribution that determines whether or not it is raining in a pixel, and a distribution to determine the rainfall depth of the wet pixels. Hence, a copula cannot be fitted directly to a rainfall field if it contains pixels without rainfall, i.e. dry pixels. To cope with this, all dry pixels were removed from the data used, making the copula only applicable to rainfall in a wet pixel both on the coarse and the fine scale.

In Sect. 3, a framework for deriving the sub-pixel distribution was described which requires that the copula is continuous. However, this condition is not necessarily fulfilled when using an empirical copula. For example, the derivative in Eq. (4) cannot be directly assessed from the empirical copula because it consists of points instead of a continuous surface. This issue is solved by linearly interpolating the available points to estimate the cumulative probability of a point wherever no empirical value is available. However, as in any empirical approach, if insufficient data are available, noisy and inaccurate results are obtained. This is especially an issue near the upper edge of the copula domain.

For each coarse-scale pixel the cumulative probability $v_m = \Gamma_{\lambda}(d_m)$ is estimated from a fitted Gamma distribution. This value is then used to derive the conditional probability

distribution function using Eq. (7). However, as an empirical copula is used, the derivative in this equation is approximated as:

$$F_{U|V}(u, v_m) \approx \frac{C(u, v_m + \delta) - C(u, v_m)}{\delta}, \quad (18)$$

requiring that two slices need to be extracted from the copula, i.e. $C(u, v_m + \delta)$ and $C(u, v_m)$, for which, when needed, linear interpolation on the empirical copula is performed. Finally, the obtained CDF is rescaled to the fine-scale domain through the application of the inverse Gamma distribution of the fine-scale rainfall field resulting in the modeled fine scale distribution of the non-zero rainfall sub-pixels.

In order to test the framework, an empirical copula needs to be used which preserves the scale dependencies valid for the storm at hand. However, constructing these copulas from an analysis of the coarse- and fine-scale pixels within the image for which the sub-pixel distribution functions are to be derived may introduce false optimism. We therefore derived an empirical copula for the rainfall field of the same storm as observed 5 h prior to the image to be downscaled; this approach is followed unless mentioned otherwise. For this time frame, it is assumed that the scaling dependencies remain similar, but that the actual rainfall field has already evolved sufficiently to prevent over-fitting on a specific image. The validity of this approach will be further investigated in Sect. 5.4.4.

The copulas obtained from various images show a consistent behavior in time and between different scales, see e.g. Fig. 6, and some common patterns are observed. These patterns are easily explained as an effect of the changing scale difference and the numerical properties of scaling, although they are not the result of the data treatment itself.

5.3 Construction of the intermittence model

As the copula only describes the distribution of the wet sub-pixels, another model is needed to describe the fraction of dry sub-pixels. As described in Sect. 3, a simple exponential model was used for this purpose. Fitting this function to a data set, which was obtained by randomly drawing coarse-scale pixels from the complete data set, the value $a = -5.25$ was found (see Fig. 4). This model generally performs well, although large errors, up to 50 %, occurred. This can be attributed to the fact that the intermittency function actually varies between storms. Given the fact that modeling this variability is not trivial (Onof et al., 1998) and that the application in this paper supports a proof of concept, this additional modeling exercise was not performed.

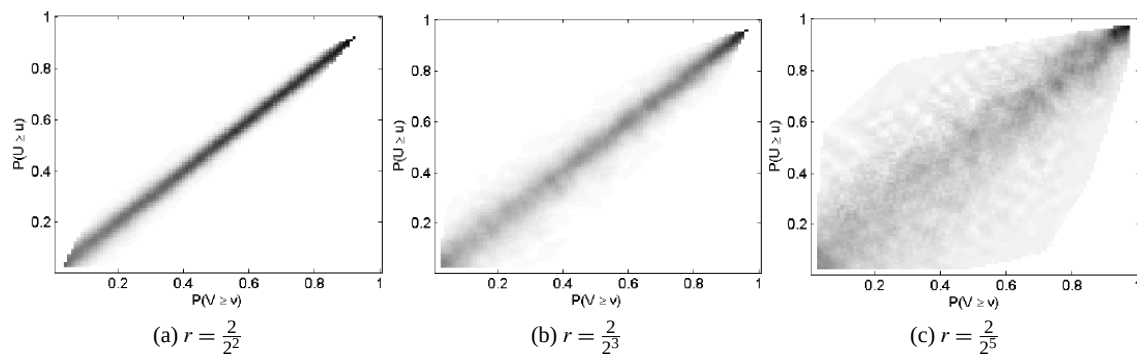


Fig. 6. Three copula densities, constructed from the same image, but using different scale steps r .

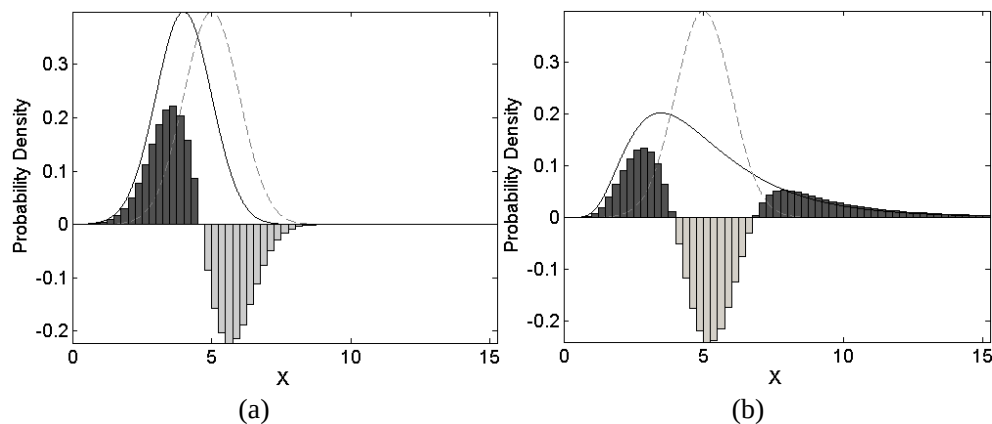


Fig. 7. An illustration of the EMD. The difference between the functions is shown as mass bars. These bars, or parts of them, need to be moved around in such a way that the dark gray bars (surplus) fill all the light gray bars (deficit). The distance they need to be moved is taken into account when calculating the EMD.

5.4 Application of the framework

5.4.1 Assessment of the resulting probability distribution functions

Generally, the sub-pixel cumulative distribution obtained through the copula framework resembles the empirical distribution. However, when statistical tests are applied, it is found that both distributions are not the same, nor does the scaled marginal CDF represent the empirical distribution. Yet, to verify which of the assumed CDFs generally better compares to the observed one, a distance function, called the Earth Movers Distance (EMD) or Wasserstein metric (Rubner et al., 2002; Gibbs and Su, 2002), is applied.

The EMD is a distance function which is a natural choice for comparing probability density functions (Rubner et al., 2002). It differs from classical measures such as the absolute error or the root mean squared error, which only take into account the vertical distance between both distributions. In

contrast, the EMD also includes the horizontal difference, i.e. the location of the deviations between both distributions.

The EMD is a variation on the transportation problem (Cha and Srihari, 2002; Rubner et al., 2002) and as such requires that the integral of the functions to be compared is equal. Therefore, we start by transforming the modeled CDFs to their PDFs, using a small amount of smoothing to ensure that noise does not lead to spurious results. Then, through subtracting the modeled PDF (either the modeled sub-pixel distribution or the field-wide marginal distribution) from the empirical PDF, a difference function showing the positive and negative differences for the considered variable is obtained (see Fig. 7 for an example). If two PDFs resemble each other, the difference function will display small values, with negative and positive values occurring close to each other, i.e. for sub-pixel fields which do not differ much in density distribution (see Fig. 7a). When the PDFs are very different, less overlap between both functions occurs and the resulting difference function will display positive and negative values for very distinct rainfall values (see Fig. 7b).

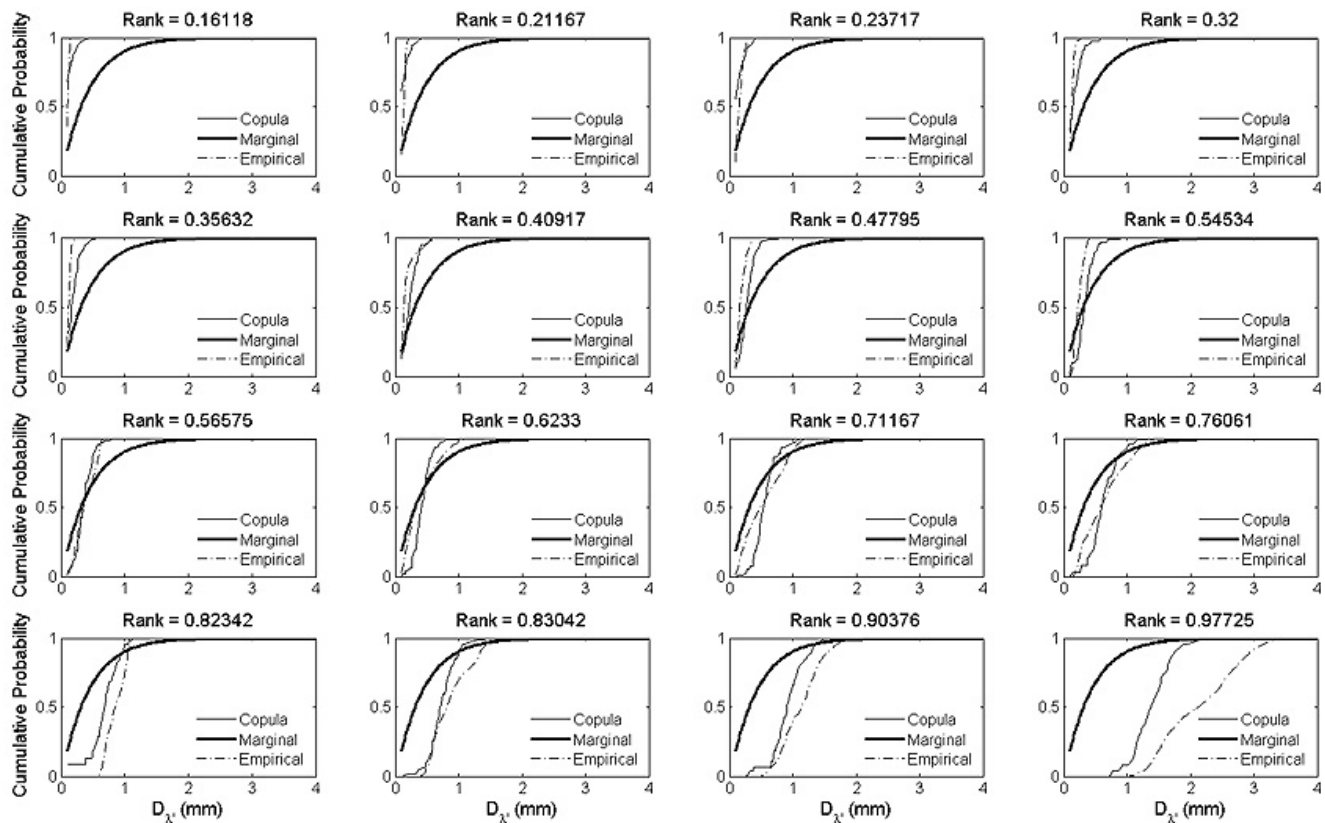


Fig. 8. Sixteen plots of actual coarse-scale pixels with their marginal (field-wide) distribution (thick, black), the empirical distribution of their sub-pixels (thin, dashed) and the modeled distribution of their sub-pixels (thin, solid). The rank of the coarse-scale observation used for conditioning is displayed above each plot.

The EMD is a measure for the minimal cost needed to move all surplus mass to deficit areas, transforming one function into the other. This cost not only accounts for the total mass to be transported (i.e. the vertical difference) but also for the distance between the surplus and the deficit area (i.e. the horizontal difference). Thus, the cost equals the vertical distance (the mass) multiplied by the distance it needs to be transported. For the distance measure, weighing functions can be used, however, for this study, all data were binned and the distance was calculated as the absolute difference between the two bin centers. This cost function needs to be minimized in order to find the EMD. This problem is generally known as the transportation problem (Cha and Srihari, 2002; Rubner et al., 2002), and many different specific solutions exist such as the Hungarian method (Kuhn, 1955) or the simplex method. These methods find the minimal distance over which the mass needs to be transported, such that the EMD can be found by multiplying the mass transported by the distance it needs to be moved.

5.4.2 Comparison to the field-wide sub-pixel CDF

In order to assess whether the copula-based sub-pixel distribution better represents the actual distribution compared to what would be obtained with classical scaling, the resulting PDFs are visually inspected before summarizing the results using the EMD. Figure 8 plots the cumulative sub-pixel distribution for 16 different coarse-scale pixels. These plots, which are representative for most of the pixels analyzed, show that the cumulative distributions obtained with the proposed methodology generally better represent the observed empirical distributions, whereas the sub-pixel distribution obtained through the scaled Gamma distribution does not allow for discriminating between the different coarse-scale cells (as this distribution represents the distribution of the complete rainfall field). The fit of Γ_{λ} deviates in several ways from the empirical marginal distribution. First, the empirical distribution evidently changes and differs from the field wide marginal. Second, small-sample distributions associated with low rainfall values tend to be underestimated (i.e. the distribution predicts more values lower than the empirical distribution), whereas those with high coarse-scale

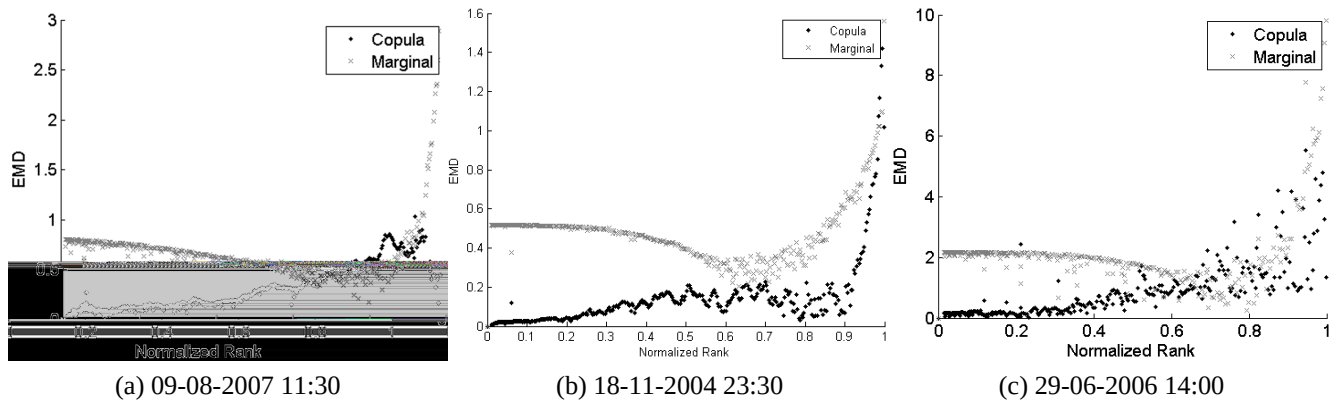


Fig. 9. The EMD between the marginal distribution and the empirical sub-pixel distribution (black crosses) compared to the EMD of the copula-based, modeled distribution (gray dots).

values are overestimated (i.e. more high-valued pixels are predicted than should be) by the field-wide distribution. As can be seen from Fig. 8, the copula-based methodology is able to better estimate the varying sub-pixel distributions, providing a closer match to the empirical distribution.

To quantify the performance, the EMD has been computed for each coarse-scale pixel for all images analyzed. These values are plotted against the normalized ranks, resulting in Fig. 9a, displaying the performance of both the marginal distribution, and that of the copula-based distribution. Compared to the marginal distribution, the copula-based distribution has a much more consistent performance over the entire domain and for almost all cases the copula-based method outperforms the fractal-based scaling of the field-wide marginal distribution. Despite this, the copula-based performance decreases towards the upper end of the coarse-scale CDF. This trend is likely due to the imperfect fit of the marginal distributions and the empirical nature of the framework (see Sect. 5.2).

The performance has been observed to vary between different rainstorms and even within storms. As an example, consider Fig. 9a and 9b where two EMD plots are shown for different storms. Figure 9b shows a fairly typical performance, doing better than the marginal distribution and showing a fairly typical pattern. However, Fig. 9c shows a different picture, with a less stable performance. This has been observed in several cases and seems to be partially due to instability in the copula as a result of the data. Moreover, the behavior of the performance observed over the domain changes between images and storms. Nevertheless, the copula-based approach generally outperforms the classical scaling, or has an equal performance. Also, the current dataset is not amenable to an analysis of the changing patterns prohibiting any further investigation.

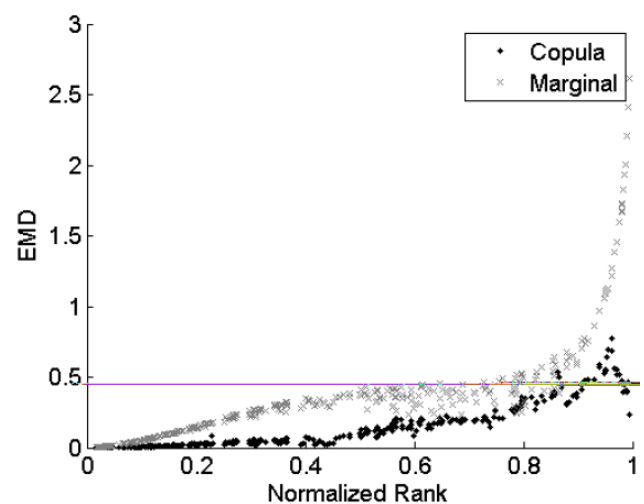


Fig. 10. The EMD after correction for the fraction of dry pixels.

5.4.3 Effects of including intermittency

The above section investigated the performance of the copula-based method without including the intermittence model. In this section, this model is included in the results and the EMD computed in a similar fashion to the previous experiment. The results of this experiment are displayed in Fig. 10, for the same storm as used in Fig. 8 (results are similar for all storms). Note that the EMD for the two approaches is very close for the first part of the curve, clearly due to the compression of the non-zero part of the distribution. The patterns in the remainder of the domain remain similar, although a smoothing effect is observed possibly as a result of the rescaling of the non-zero parts of the CDF. Thus, as would be expected, the inclusion of intermittency does not change the performance patterns significantly.

5.4.4 Temporal robustness of copula

The copula is likely to be specific to a certain storm or image and the performance might be limited by the time lag between the image on which the copula is fitted and the image to be downscaled. To test this, the copula was fitted to the first image of a storm and was used to downscale all sequential radar images within that same storm. For each of these images the EMD is calculated on a per pixel basis and all values are then averaged. In Fig. 12, these values are displayed for all storms in our dataset, ordered according to their time-lag. Note that at zero time-lag, i.e. the image from which the copula is constructed is the same as the image to be downscaled, the error is not zero. This is partially due to the empirical nature of the framework, and possibly due to the generalizations inherent to fitting a copula. Despite this, the curves do not appear to have an upward trend, suggesting that an increasing time-lag does not have a negative effect on the performance.

5.4.5 Storm dependence of the copula

Until now, the copula which was used was obtained from the same rain event. To investigate whether the storm used to construct the empirical copula is of importance, a copula was built from 10 different images picked at random from the entire dataset. This ensures that the dependence structure generated by different storm types is mixed in order to provide a more general copula. This copula is then applied to several storms and the EMD is calculated for each coarse-scale pixel. The results are displayed in Fig. 11. As can be seen, using the general copula only slightly decreases the overall performance, but still outperforms the sub-pixel distribution based on scaled marginal distributions. Although more research in this respect is needed, this simple example demonstrates the potential of this scaling technique for downscaling rainfall images in an operational framework, as it may only require to once construct a copula based on a variety of storms.

6 Conclusions

In this paper, a novel technique is introduced which allows for downscaling coarse-scale rainfall images. This method is based on the simultaneous use of fractal-based scaling of the marginal probability distribution functions and a copula which describes the dependence between both scales. The introduction of the dependence allows for a better estimation of the actual shape of sub-pixel probability functions compared to the scaled marginal distribution function.

The proposed method can strengthen current downscaling methods which assume unrealistic sub-pixel distributions. This will require at least a moderate degree of temporal stability of the copula, which has been tentatively shown in this study. Future research will focus on this temporal stability and it will be investigated whether different copulas are

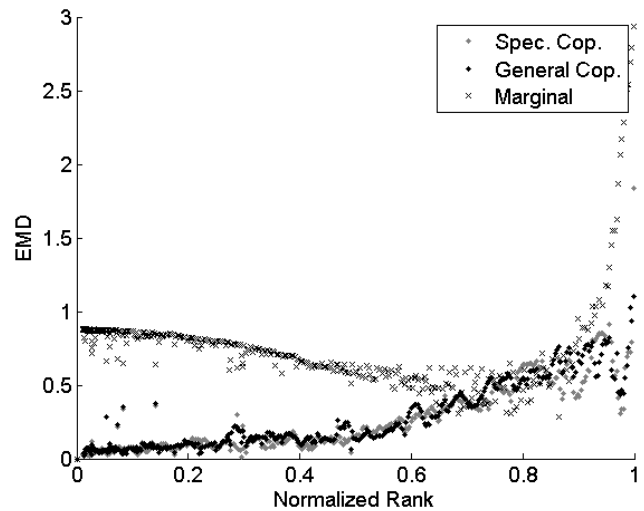


Fig. 11. The EMD between the empirical sub-pixel distribution and the specifically fitted copula (gray dots), the more general copula (black dots), or the field-wide marginal distribution (black crosses).

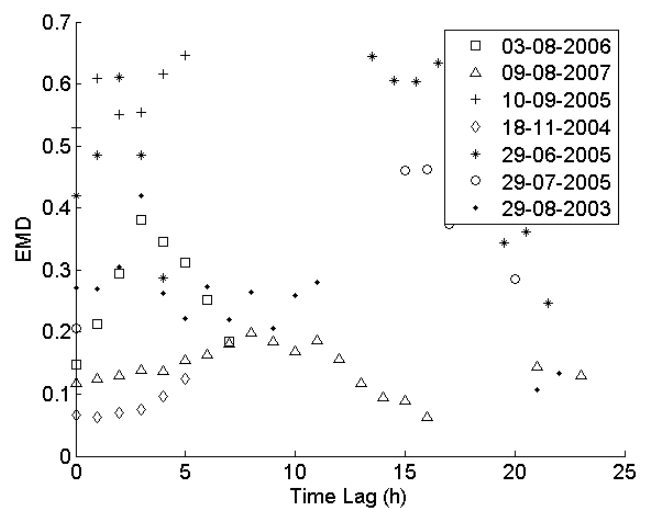


Fig. 12. The EMD determined for every image of each storm (dates listed in the legend) with the copula used for all images in a single storm based on the first image of that storm.

needed for different storm types and whether the same scaling can be applied for different storm types. As such, future work should not only include a larger dataset, but also involve storm classification to investigate differences in dependence structure between storms.

The current method is, as mentioned, still a proof of concept. However, to obtain a workable downscaling method, a parametrical copula should be identified that best describes the dependence between two scales. Furthermore, the way the copula changes as a result of a change in scales needs to

be known. Therefore, a function is needed that relates the parameters of the parametrical copulas to the scales between which these are valid. Future research should also focus on whether a parametrical copula can be fitted such that problems at the extremes of the copula, where little data is available, are solved.

It was shown that the copula-based framework is sufficiently robust to be applied to different storms and time steps, demonstrating its large potential for statistical downscaling and hydrological modeling. Yet, more research is needed to further elaborate the proposed methodology.

Acknowledgements. The first author was supported by the Special Research Fund of Ghent University, and the second author was funded by Research Foundation Flanders (FWO). The authors kindly thank the Royal Meteorological Institute of Belgium for providing the data.

Edited by: F. Pappenberger

References

- Allcroft, D. and Glasbey, C.: A latent Gaussian Markov random-field model for spatiotemporal rainfall disaggregation, *J. R. Stat. Soc. C-App.*, 52, 487–498, doi:10.1111/1467-9876.00419, 2003.
- Burton, A., Kilsby, C. G., Fowler, H. J., Cowpertwait, P. S. P., and Connell, P. E. O.: RainSim : A spatial temporal stochastic rainfall modelling system, *Environ. Modell. Softw.*, 23, 1356–1369, doi:10.1016/j.envsoft.2008.04.003, 2008.
- Cha, S. H. and Srihari, S. N.: On measuring the distance between histograms, *Pattern Recogn.*, 35, 1355–1370, doi:10.1016/S0031-3203(01)00118-2, 2002.
- Collison, A., Steven, W., Griffiths, J., and Dehn, M.: Modelling the impact of predicted climate change on landslide frequency and magnitude in SE England, *Eng. Geol.*, 55, 205–218, doi:10.1016/S0013-7952(99)00121-0, 2000.
- Cowpertwait, P. S. P., Kilsby, C. G., and O'Connell, P. E.: A space-time Neyman-Scott model of rainfall: Empirical analysis of extremes, *Water Resour. Res.*, 38, 1131, doi:10.1029/2001WR000709, 2002.
- De Jongh, I. L. M., Verhoest, N. E. C., and De Troch, F. P.: Analysis of a 105-year time series of precepitation observed at Uccle, Belgium, *Int. J. Climatol.*, 26, 2023–2039, doi:10.1002/joc.1352, 2006.
- De Michele, C. and Salvadori, G.: A generalized pareto intensity-duration model of storm rainfall exploiting 2-copulas, *J. Geophys. Res.*, 108, 4067, doi:10.1029/2002JD002534, 2003.
- Deidda, R.: Rainfall downscaling in a space-time multifractal framework, *Water Resour. Res.*, 36, 1779–1794, doi:10.1029/2000WR900038, 2000.
- Deidda, R., Benzi, R., and Siccardi, F.: Multifractal modeling of anomalous scaling laws in rainfall, *Water Resour. Res.*, 35, 1853–1867, doi:10.1029/1999WR900036, 1999.
- Ferraris, L., Gabellani, S., Rebora, N., and Provenzale, A.: A comparison of stochastic models for spatial rainfall downscaling, *Water Resour. Res.*, 39(12), 1368, doi:10.1029/2003WR002504, 2003a.
- Ferraris, L., Gabellani, S., Parodi, U., Rebora, N., von Hardenberg, J., and Provenzale, A.: Revisiting multifractality in rainfall fields, *J. Hydrometeorol.*, 4, 544–551, doi:10.1175/1525-7541(2003)004<0544:RMIRF>2.0.CO;2, 2003b.
- Gao, H., Wood, E. F., Drusch, M., and McCabe, M. F.: Copula-derived observation operators for assimilating TMI and AMSR-E retrieved soil moisture into land surface models, *J. Hydrometeorol.*, 8, 413–429, doi:10.1175/JHM570.1, 2007.
- Genest, C. and Favre, A.: Everything you always wanted to know about copula modeling but were afraid to ask, *J. Hydrol. Eng.*, 12, 347–368, doi:10.1061/(ASCE)1084-0699(2007)12:4(347), 2007.
- Gibbs, A. L. and Su, F. E.: On Choosing and Bounding Probability Metrics, *International Statistical Review*, 70, 419–435, doi:10.1111/j.1751-5823.2002.tb00178.x, 2002.
- Grimaldi, S. and Serinaldi, F.: Asymmetric copula in multivariate flood frequency analysis, *Adv. Water Resour.*, 29, 1155–1167, doi:10.1016/j.advwatres.2005.09.005, 2006.
- Gupta, V. K. and Waymire, E.: Multiscaling properties of spatial rainfall and river flow distribution, *J. Geophys. Res.*, 95, 1999–2009, doi:10.1029/JD095iD03p01999, 1990.
- Heylen, R. and Maenhout, A.: Inleiding tot de radarmeteorologie, Koninklijk Meteorologisch instituut van België, Brussels, 69 pp., 1994 (in Dutch).
- Kao, S. and Govindaraju, R. S.: Trivariate statistical analysis of extreme rainfall events via the Plackett family of copulas, *Water Resour. Res.*, 44, W02415, doi:10.1029/2007WR006261, 2008.
- Kuhn, H. W.: The Hungarian method for the assignment problem, *Nav. Res. Logist. Q.*, 2, 83–97, doi:10.1002/nav.3800020109, 1955.
- Le Cam, L.: A stochastic description of precipitation, in: *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, vol. 3, 165–186, 1961.
- Lovejoy, S. and Mandelbrot, B.: Fractal properties of rain and a fractal model, *Tellus*, 37, 209–232, doi:10.1111/j.1600-0870.1985.tb00423.x, 1985.
- Mackay, N. G., Chandler, R. E., Onof, C., and Wheeler, H. S.: Disaggregation of spatial rainfall fields for hydrological modelling, *Hydrol. Earth Syst. Sci.*, 5, 165–173, doi:10.5194/hess-5-165-2001, 2001.
- Marshall, J. and Palmer, W.: The distribution of raindrops with size, *J. Atmos. Sci.*, 5, 165–166, doi:10.1175/1520-0469(1948)005<0165:TDORWS>2.0.CO;2, 1948.
- Menabde, M., Seed, A., and Pegram, G.: A simple scaling model for extreme rainfall, *Water Resour. Res.*, 35, 335–339, doi:10.1029/1998WR900012, 1999.
- Nelsen, R. B.: *An introduction to copulas*, Springer Verlag, 2006.
- Onof, C., Mackay, N. G., Oh, L., and Wheeler, H. S.: An improved rainfall disaggregation technique for GCMs, *J. Geophys. Res.*, 103, 19577–19586, doi:10.1029/98JD01147, 1998.
- Onof, C., Chandler, R. E., Kakou, A., Northrop, P., Wheeler, H. S., and Isham, V.: Rainfall modelling using Poisson-cluster processes: a review of developments, *Stoch. Env. Res. Risk A.*, 14, 384–411, doi:10.1007/s004770000043, 2000.
- Pegram, G. G. S. and Clothier, A. N.: Downscaling rainfields in space and time, using the String of Beads model in time series mode, *Hydrol. Earth Syst. Sci.*, 5, 175–186, doi:10.5194/hess-5-175-2001, 2001.
- Rebora, N., Ferraris, L., von Hardenberg, J., and Provenzale, A.:

- RainFARM: rainfall downscaling by a filtered autoregressive model, *J. Hydrometeorol.*, 7, 724–738, doi:10.1175/JHM517.1, 2006.
- Rubner, Y., Tomasi, C., and Guibas, L.: A metric for distributions with applications to image databases, in: *Computer Vision, Sixth International Conference on*, 59–66, IEEE, 2002.
- Salvadori, G. and De Michele, C.: Statistical characterization of temporal structure of storms, *Adv. Water Resour.*, 29, 827–842, doi:10.1016/j.advwatres.2005.07.013, 2006.
- Salvadori, G. and De Michele, C.: On the use of copulas in hydrology: theory and practice, *J. Hydrol. Eng.*, 12, 369–380, doi:10.1061/(ASCE)1084-0699(2007)12:4(369), 2007.
- Schertzer, D. and Lovejoy, S.: Physical modeling and analysis of rain and clouds by anisotropic scaling multiplicative processes, *J. Geophys. Res.*, 92, 9693–9714, doi:10.1029/JD092iD08p09693, 1987.
- Schölzel, C. and Friederichs, P.: Multivariate non-normally distributed random variables in climate research – introduction to the copula approach, *Nonlin. Processes Geophys.*, 15, 761–772, doi:10.5194/npg-15-761-2008, 2008.
- Serinaldi, F.: Copula-based mixed models for bivariate rainfall data: an empirical study in regression perspective, *Stoch. Env. Res. Risk A.*, 23, 677–693, doi:10.1007/s00477-008-0249-z, 2009.
- Vandenberghe, S., Verhoest, N. E. C., Buyse, E., and De Baets, B.: A stochastic design rainfall generator based on copulas and mass curves, *Hydrol. Earth Syst. Sci.*, 14, 2429–2442, doi:10.5194/hess-14-2429-2010, 2010a.
- Vandenberghe, S., Verhoest, N. E. C., and De Baets, B.: Fitting bivariate copulas to the dependence structure between storm characteristics: A detailed analysis based on 105 year 10 min rainfall, *Water Resour. Res.*, 46, W01512, doi:10.1029/2009WR007857, 2010b.
- Varis, O., Kajander, T., and Lemmelä, R.: Climate and water: from climate models to water resources management and vice versa, *Climatic Change*, 66, 321–344, doi:10.1023/B:CLIM.0000044622.42657.d4, 2004.
- Veneziano, D., Langousis, A., and Furcolo, P.: Multifractality and rainfall extremes: A review, *Water Resour. Res.*, 42, W06D15, doi:10.1029/2005WR004716, 2006.
- Vrac, M. and Naveau, P.: Stochastic downscaling of precipitation: From dry events to heavy rainfalls, *Water Resour. Res.*, 43, W07402, doi:10.1029/2006WR005308, 2007.
- Wilks, D. S.: Multisite downscaling of daily precipitation with a stochastic weather generator, *Clim. Res.*, 11, 125–136, doi:10.3354/cr011125, 1999.
- Willems, P.: A spatial rainfall generator for small spatial scales, *J. Hydrol.*, 252, 126–144, doi:10.1016/S0022-1694(01)00446-2, 2001.
- Winchell, M., Gupta, H., and Sorooshian, S.: On the simulation of infiltration-and saturation-excess runoff using radar-based rainfall estimates: effects of algorithm uncertainty and pixel aggregation, *Water Resour. Res.*, 34, 2655–2670, doi:10.1029/98WR02009, 1998.