

Copula-based assimilation of radar and gauge information to derive bias-corrected precipitation fields

S. Vogl, Patrick Laux, W. Qiu, G. Mao, Harald Kunstmann

Angaben zur Veröffentlichung / Publication details:

Vogl, S., Patrick Laux, W. Qiu, G. Mao, and Harald Kunstmann. 2012. "Copula-based assimilation of radar and gauge information to derive bias-corrected precipitation fields." *Hydrology and Earth System Sciences* 16 (7): 2311–28.
<https://doi.org/10.5194/hess-16-2311-2012>.

Nutzungsbedingungen / Terms of use:

CC BY 4.0





Copula-based assimilation of radar and gauge information to derive bias-corrected precipitation fields

S. Vogl^{1,2}, P. Laux², W. Qiu², G. Mao², and H. Kunstmann^{1,2}

¹University of Augsburg, Institute for Geography, Regional Climate and Hydrology, 86135 Augsburg, Germany

²Karlsruhe Institute of Technology (KIT), Institute for Meteorology and Climate Research (IMK-IFU),
Kreuzeckbahnstrasse 19, 82467 Garmisch-Partenkirchen, Germany

Correspondence to: S. Vogl (stefanie.vogl@geo.uni-augsburg.de)

Received: 8 December 2011 – Published in Hydrol. Earth Syst. Sci. Discuss.: 17 January 2012

Revised: 11 April 2012 – Accepted: 27 June 2012 – Published: 25 July 2012

Abstract. This study addresses the problem of combining radar information and gauge measurements. Gauge measurements are the best available source of absolute rainfall intensity albeit their spatial availability is limited. Precipitation information obtained by radar mimics well the spatial patterns but is biased for their absolute values.

In this study copula models are used to describe the dependence structure between gauge observations and rainfall derived from radar reflectivity at the corresponding grid cells. After appropriate time series transformation to generate “iid” variates, only the positive pairs (radar > 0, gauge > 0) of the residuals are considered. As not each grid cell can be assigned to one gauge, the integration of point information, i.e. gauge rainfall intensities, is achieved by considering the structure and the strength of dependence between the radar pixels and all the gauges within the radar image. Two different approaches, namely *Maximum Theta* and *Multiple Theta*, are presented. They finally allow for generating precipitation fields that mimic the spatial patterns of the radar fields and correct them for biases in their absolute rainfall intensities. The performance of the approach, which can be seen as a bias-correction for radar fields, is demonstrated for the Bavarian Alps. The bias-corrected rainfall fields are compared to a field of interpolated gauge values (ordinary kriging) and are validated with available gauge measurements. The simulated precipitation fields are compared to an operationally corrected radar precipitation field (RADOLAN). The copula-based approach performs similarly well as indicated by different validation measures and successfully corrects for errors in the radar precipitation.

1 Introduction

For many hydrological analyses, spatially distributed precipitation information is indispensable.

Whenever gauge data alone is the basis to derive a precipitation field, a wide range of interpolation methods is in use. Some examples include nearest neighbour (e.g. Isaaks and Srivastava, 1989), inverse distance weighting, regression models (e.g. Bourrough and McDonell, 1998), trend surface analysis (e.g. Collins and Bolstadt, 1996), splines (e.g. Hutchinson, 1998a,b; Bourrough and McDonell, 1998) and kriging, for which a large set of sub-methods has been developed (e.g. Isaaks and Srivastava, 1989; Bollerslev, 1986; Goovaerts, 2000; Haberlandt, 2007). There has been also use made of mixed methods such as regression combined with kriging (e.g. Erxleben et al., 2002) and others. Some methods attempt to assimilate additional information such as elevation of the terrain or additional measurements such as radar or from remote sensing (e.g. Haberlandt, 2007).

Even though the methods are different in nature, they have one thing in common: the performance is highly dependent on the density of the observation network and on the complexity of the underlying terrain. This is problematic insofar as especially in regions with large height gradients usually the number of available meteorological stations, providing reliable rainfall measurements, is limited. Radar precipitation fields are supposed to be a good supplement as they are also covering areas with complex terrain in a high spatio-temporal resolution and the patterns of rainfall are assumed to be realistically reproduced.

There are multiple sources of errors when precipitation is derived from radar reflectivities such as the empirical (Z/R) reflectivity-rainfall relationship and errors induced by the radar measurement itself such as backscatter or shadowing effects (e.g. Joss and Lee, 1995). Whenever radar fields are used to drive meteorological or hydrological models, these errors have to be taken into account (e.g. Cole and Moore, 2008; Singh, 1997) as they will directly propagate into the predicted variables. The uncertainty propagation in hydrological equations that directly propagate the probability density functions of uncertain model input parameters into the corresponding probability density functions of model output is studied e.g. by Kunstmann and Karstens (2006). There have been many attempts to quantify the uncertainty of the radar measurement (e.g. Mandapaka et al., 2009; AghaKouchak et al., 2010a). To reduce the uncertainties in the radar-derived precipitation fields, many different approaches exist. It is very common to assimilate gauge information to the radar field (e.g. Brandes, 1975; Krajewski, 1987; Mazzetti and Todini, 2004; Ehret, 2003) to correct for errors in the absolute values. Usually, the existing approaches follow two steps: first the available gauge data are used to generate an interpolated precipitation field, and in the second step radar and interpolated field are combined.

The underlying assumption of Gaussian behaviour is restricting the performance of standard approaches, as many studies showed that the interdependence in hydrological or meteorological datasets is usually more complex (e.g. Gomez-Hernandez and Wen, 1998; Bárdossy, 2006; Bárdossy and Li, 2008).

Alternatively, a copula approach can be used to describe the complex spatio-temporal dependence structure and assess for non-linear behaviour (e.g. Genest and Favre, 2007; Dupuis, 2007). The copula method is advantageous in several respects and has e.g. been used successfully in risk assessment (e.g. Embrechts et al., 2001; Frees and Valdez, 1998). Over the past years, there has been a remarkable increase in applications of copulas in hydrometeorology. Copula-based models have been introduced for bi- and multivariate frequency analysis, geostatistical interpolation and extreme value analyses (e.g. De Michele and Salvadori, 2003; Dupuis, 2007; Bárdossy, 2006; Genest and Favre, 2007; Renard and Lang, 2007; Schölzel and Friederichs, 2008; Bárdossy and Li, 2008; Zhang and Singh, 2008; Laux et al., 2009). For rainfall modelling, De Michele and Salvadori (2003) used copulas to model intensity duration of rainfall events. Favre et al. (2004) utilized copulas for multivariate hydrological frequency analysis. Zhang and Singh (2008) carried out a bivariate rainfall frequency analysis using Archimedean copulas. Renard and Lang (2007) investigated the usefulness of the Gaussian copula in extreme value analysis. Kuhn et al. (2007) employed copulas to describe spatial and temporal dependence of weekly precipitation extremes. Serinaldi (2008) studied the dependence of rain gauge data using the non-parametric Kendall's rank cor-

relation and the upper tail dependence coefficient (TDC). Based on the properties of the Kendall correlation and TDC, a copula-based mixed model for modelling the dependence structure and marginals is suggested. Recently, van den Berg et al. (2011) developed a copula-based approach for statistical downscaling of precipitation fields obtained from radar observations. Copula-based models for estimating error fields of radar information are described e.g. by Villarini et al. (2008); AghaKouchak et al. (2010b,c).

Most of these studies are carried out in the bivariate framework describing dependency between two variates, but there are also few examples of multivariate applications. Bárdossy (2006); Bárdossy and Li (2008); Bárdossy and Pegram (2009) suggested a new method for geostatistical interpolation based on copulas. They used multivariate copulas to describe the spatial variability of groundwater quality parameters and developed a methodology to spatially interpolate these quantities. More information about applications of copulas in hydrometeorology can be found at the website of the International Commission on Statistical Hydrology (ICSH) at www.stahy.org.

In this study, two new copula-based approaches for correcting biases in radar precipitation fields are presented. Information from rain gauges is assimilated by investigating the dependence structure between rain gauges and radar fields.

The article is structured as follows: in Sect. 2 the study area and the database are introduced. Section 3 reviews briefly the basic theory of copulas and the procedure of simulating data using conditional cumulative distribution functions (CDFs). The new copula-based bias-correction approaches are introduced. Results of the application of the methodology in the Alpine space are presented in Sect. 4, followed by the discussion in Sect. 5 and the conclusions in Sect. 6.

2 Study area and data

2.1 Domain

In regions where precipitation reveals high spatio-temporal variability, such as alpine or prealpine terrain, it is a specific challenge to estimate realistic rainfall fields. Therefore, the study area is chosen to be the southern Bavarian Alps and alpine forelands in Germany. There are large gradients in elevation across the domain with lowest values in the north. Within the flat area of the domain, Munich lies at an altitude of 519 m a.s.l. The highest location is the peak Zugspitze at 2962 m a.s.l. Due to the complex orography and heterogeneity in topography, the domain is characterized by strong north southerly differentiations in soils, land-use, and climate. The mean annual temperature is around 7–8 °C in the alpine forelands and 4–5 °C in the southern part of the domain. The mean annual precipitation ranges from about 1100 mm in the

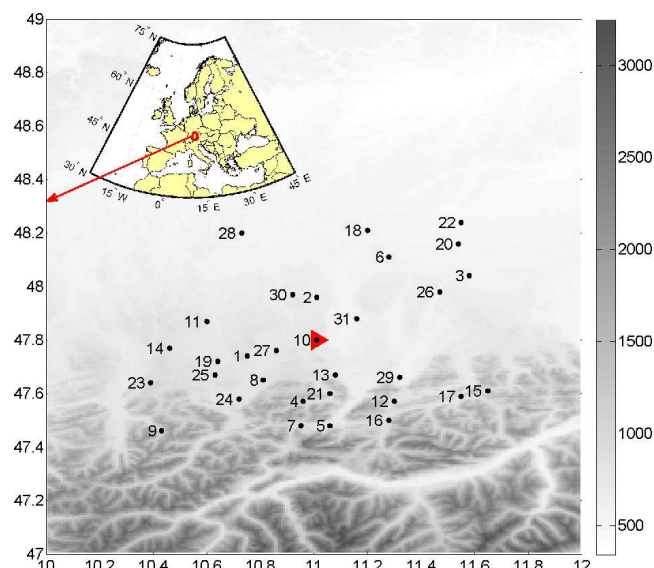


Fig. 1. Research area showing the position of the gauges and the weather radar on Mount Hohenpeissenberg (red triangle). The names of the gauge stations (black dots) can be found in Table 1.

northern part to more than 2000 mm in the south. Figure 1 shows an overview over the study area, highlighting the large north-southerly gradient in altitude.

In this study, radar data from the Meteorological Observatory Hohenpeissenberg (MOHP) are used. The radar is a C-Band research weather radar operated by the German Weather Service (DWD). The observatory is located about 80 km southwest of Munich on Mount Hohenpeissenberg at an altitude of 1000 m a.s.l. The radar installation covers a circular area with a radius of 256 km, producing a scan every 5 min. For the study area, a square region of 100×100 grid cells with $1 \text{ km} \times 1 \text{ km}$ resolution, centred at the radar station at mount Hohenpeissenberg, is selected. Preprocessing steps include a clutter correction of the radar reflectivities, which has been done by DWD. After that, hourly rainfall amounts are derived by using the DWD standard Z (reflectivity in dbz)/ R (rainfall in mm h^{-1}) relationship:

$$Z = 256 * R^{1.42}. \quad (1)$$

Details about the DWD standard Z/R relationship are given e.g. by Riedl (1986) and Seltmann (1997). The data cover the summer months (June, July, and August) of the years 2005–2009. Precipitation data of 31 gauges within the chosen domain are retrieved from DWD, which cover the same period as the radar data. Figure 2 shows a snapshot of a radar-based precipitation field measured at mount Hohenpeissenberg on 14 July 2008 (13:00 CET).

In addition to the radar precipitation derived by a standard Z/R relationship, operationally corrected radar-based hourly rainfall time series are used. DWD developed a routine method for the online adjustment of radar precipitation

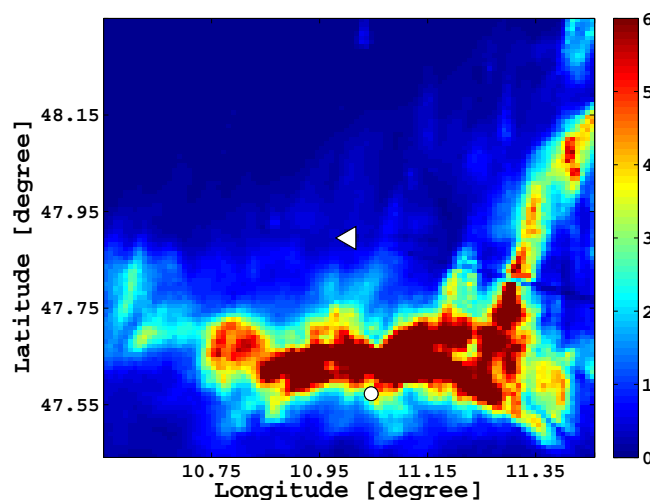


Fig. 2. Precipitation field [mm h^{-1}] derived from radar measurements (standard Z/R relationship) at Mount Hohenpeissenberg on 14 July 2008 (13:00). The position of Hohenpeissenberg (Garmisch-Partenkirchen) is indicated by a white triangle (circle).

by means of automatic surface precipitation stations (ombrometer) of DWD. The correction of DWD (hereinafter referred to as RADOLAN) includes a refined Z/R -relationship, which contains orographic shading correction, statistical reduction of clutter, gradient smoothing and further preprocessing steps. The precipitation amounts observed at surrounding ombrometer stations (more than 1000 for Germany) are interpolated and assimilated resulting in a bias-corrected radar precipitation field (Bartels et al., 2004).

Each of the 31 gauge stations used in this study is assigned to its corresponding grid cell in the radar domain. As the observation network is very sparse, there is no grid cell assigned to more than one gauge station. Figure 1 shows the positions of the available gauge stations and the position of the weather radar at Mount Hohenpeissenberg.

2.2 Data preprocessing and availability

The data pairs are checked for plausibility, and erroneous or significantly anomalous values are removed. This procedure consists of the following steps:

1. High gauge values are checked for plausibility by comparing with nearby gauges.
2. Radar rainfall values smaller than 0.1 mm h^{-1} are set to zero as these measurements are considered to be erroneous.
3. The differences between neighbouring radar grid cells are calculated. Single values with absolute differences exceeding a threshold of 25 mm h^{-1} , revealing unrealistically large gradients in the radar field, are removed (Marx, 2007).

Table 1. Geographical location of gauges and the weather radar as well as statistical measures for gauge & radar (positive pairs only) for period June, July, and August of 2006 and 2007.

ID	Station Name	Altitude [m]	Lat [°]	Long [°]	Gauge _{mean} [mm h ⁻¹]	Gauge _{std} [mm h ⁻¹]	Radar _{mean} [mm h ⁻¹]	Radar _{std} [mm h ⁻¹]
1	Bernbeuren-Prachtsried	936	47.74	10.75	1.76	2.27	1.33	2.23
2	Diessen	658	47.96	11.01	1.51	1.96	1.15	1.74
3	Deisenhofen	585	48.04	11.58	1.62	2.12	1.25	1.60
4	Ettal	940	47.57	10.96	1.59	1.85	1.36	1.88
5	Garmisch-Partenkirchen	719	47.48	11.06	1.61	2.02	1.48	2.13
6	Gilching	550	48.11	11.28	1.56	2.25	1.25	1.75
7	Griesen	801	47.48	10.95	1.48	1.75	1.49	2.13
8	Halblech	780	47.65	10.81	1.68	2.27	1.15	1.42
9	Hindelang	1015	47.46	10.43	1.75	2.28	1.59	2.67
10	Hohenpeissenberg	977	47.80	11.01	5.03	4.10	0.16	0.01
11	Kaufbeuren	716	47.87	10.60	1.66	2.16	1.26	1.93
12	Kochel	805	47.57	11.30	1.51	1.87	1.22	1.60
13	Kohlgrub, Bad	740	47.67	11.08	1.82	2.27	1.25	1.91
14	Kraftisried	831	47.77	10.46	1.70	2.25	1.30	2.12
15	Kreuth	895	47.61	11.65	1.72	2.13	1.43	2.11
16	Krün	873	47.50	11.28	1.71	2.18	1.46	1.97
17	Lenggries	737	47.59	11.55	1.73	2.20	1.55	2.40
18	Maisach	530	48.21	11.20	1.55	2.04	1.19	1.69
19	Marktoberdorf	790	47.72	10.64	1.77	2.30	1.32	2.08
20	Munich	515	48.16	11.54	1.53	2.21	1.43	1.86
21	Oberammergau	835	47.60	11.06	1.58	2.11	1.10	1.88
22	Oberschleissheim	484	48.24	11.55	1.55	2.13	1.24	1.81
23	Oy	885	47.64	10.39	1.84	2.14	1.52	2.09
24	Schwangau	796	47.58	10.72	1.63	2.13	1.47	2.04
25	Seeg	802	47.67	10.63	1.67	2.19	1.36	1.95
26	Schäftlarn	557	47.98	11.47	1.58	2.01	1.24	1.59
27	Steingaden	761	47.76	10.86	1.72	2.18	1.16	1.83
28	Schwaben	538	48.20	10.73	1.54	1.91	1.18	1.67
29	Schlehdorf	609	47.66	11.32	1.76	2.12	1.41	1.82
30	Vilgertshofen	685	47.97	10.92	1.66	2.16	1.21	1.71
31	Wielenbach	550	47.88	11.16	1.45	1.89	1.10	1.44

- Only the remaining positive pairs (radar and gauge) of rainfall intensities are considered for further calculations in this study.

The remaining positive data pairs are divided into two subsets. The first set containing data of June, July, and August of 2006 and 2007 serves as calibration period. The second dataset contains data from the same months of 2008. Mean and standard deviation of gauge and radar (positive pairs) for the calibration period are listed in Table 1. Especially for station Hohenpeissenberg, the number of positive pairs is limited. As the gauge station is located very close to the radar observatory, the radar beam cannot capture the area above this gauge. These are the reasons why mean and standard deviations of the positive pairs differ significantly from those of the other stations.

3 Methodology

In the following section, the theoretical background of copula theory is briefly sketched including information about marginal distributions, a description of the new bias correction algorithms and validation measures.

3.1 Marginal distributions

Modelling the joint dependence structure with copulas requires fitting marginal distributions to the data. The structure of the simulation technique allows to select a suitable distribution for each radar pixel and every gauge. In this study, four different distribution functions are tested to identify the best fit (see Table 2).

To decide which univariate distribution is the best suitable for both radar and gauge data, the parameters are estimated using a standard maximum likelihood approach and the goodness-of-fit is checked with standard goodness-of-fit

Table 2. Marginal distributions.

distribution	formula	parameters
Normal distribution	$f(x) := \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$	mean μ , standard deviation σ ; $\mu, \sigma \in \mathbb{R}$
Exponential distribution	$f_\lambda(x) := \begin{cases} \lambda e^{-\lambda x} & \text{if } x \geq 0, \\ 0 & \text{else} \end{cases}$	$\lambda \in \mathbb{R}_{>0}$
Weibull distribution	$f(x) := \alpha \beta x^{\beta-1} e^{-\alpha x^\beta}$	$\alpha, \beta \in \mathbb{R}_{>0}$
Gamma distribution	$f_\lambda(x) := \begin{cases} \frac{b^p}{\Gamma(p)} x^{p-1} e^{-bx} & \text{if } x \geq 0, \\ 0 & \text{else} \end{cases}$	$b, p \in \mathbb{R}_{>0}$ $\Gamma(p)$: gamma function at p

tests, such as the Kolmogorov-Smirnov test or the Chi Squared test. Additionally, the Akaike and the Bayesian information criteria are used:

$$\text{AIC} = 2k - 2\ln(L) \quad (2)$$

and

$$\text{BIC} = k \ln(n) - 2\ln(L), \quad (3)$$

where k denotes the number of the free parameters of the model, n is the sample size and L is the maximized value of the likelihood function of the estimated model. The smallest value of AIC or BIC respectively suggests the best fitting model/distribution.

For the application of a copula model, it is an indispensable prerequisite that the marginals are “iid” (independent and identically distributed). If this is not the case, an appropriate transformation has to be applied to the data to generate “iid” variates (Laux et al., 2011).

3.2 General introduction to copula theory

In the following, the basic definitions concerning copulas are briefly reviewed.

Let (X_1, \dots, X_n) denote a n -tuple of random variables and (x_1, \dots, x_n) a realization of it. Then, copulas are functions that link the multivariate distribution $F(x_1, \dots, x_n)$ to its univariate marginals $F_{X_i}(x_i)$. Thus, they are often also called dependence functions. Sklar (1959) proved that every multivariate distribution $F(x_1, \dots, x_n)$ can be expressed in terms of a copula C and its marginals $F_{X_i}(x_i)$:

$$F(x_1, \dots, x_n) = C(F_{X_1}(x_1), \dots, F_{X_n}(x_n)) \quad (4)$$

$$C : [0, 1]^n \rightarrow [0, 1]. \quad (5)$$

In turn of linking multivariate distributions to their marginals, copulas allow to merge the dependence structure from the marginal distributions to form their joint multivariate distribution. The copula function is unique when the marginals are steady functions. As the copula is only a reflection of the dependence structure itself, its construction is reduced to

the study of the relationship between the correlated variables, giving freedom for the choice of the univariate marginal distributions. Further information about copulas can be found e.g. in Joe (1997); Frees and Valdez (1998); Nelsen (1999); Salvadori et al. (2007).

The copula approach allows to account for the fact that the dependence structure between two variates (X, Y) is more complex than it can be modelled by the multivariate normal distribution or ordinary dependence measures such as e.g. the Pearson correlation coefficient. Another important property of copula functions is the fact that they are invariant under increasing monotonic transformations. In practice, this means that data may be transformed (e.g. by taking the logarithm or detrending) without changing their copula. More details including the definition of the empirical copula and the theoretical copula functions used in this study can be found in the Appendix A.

3.2.1 Copula parameter estimation

For two variates (X, Y) and a given family of copulas C_Θ , the set of copula parameters Θ can be estimated in different ways, depending on whether $\Theta = \theta \in \mathbb{R}$ or $\Theta = (\theta_1, \dots, \theta_n) \in \mathbb{R}^n$. For one-parametric copulas, the functional relationship between the classical dependence parameters such as Kendall's τ and Spearman's ρ and copula functions, namely

$$\rho = 12 \iint_{[0,1]^2} uv \, dC_\theta(u, v) - 3 \quad (6)$$

and

$$\tau = 4 \iint_{[0,1]^2} C_\theta(u, v) \, dC_\theta(u, v) - 1, \quad (7)$$

can be used to estimate the copula parameter. This is especially useful for Archimedean copulas as for a copula with generator φ that holds

$$\tau = 1 + 4 \int_{[0,1]} \frac{\varphi(t)}{\varphi'(t)} dt. \quad (8)$$

Thus, for example for the Gumbel-Hougaard copula with its generator $\varphi(t) = (-\ln(t))^\theta$, it is found that $\theta = \frac{1}{1-\tau}$. As θ is an increasing function of τ , this empirical link also enables the interpretation of the copula parameter as a measure for the strength of dependence: higher copula parameters reveal a stronger dependence.

For a two dimensional copula parameter $\Theta = (\theta_1, \theta_2) \in \mathbb{R}^2$ as e.g. in the case of the Student-T copula, the method of maximum pseudo-likelihood as described by Genest and Favre (e.g. 2007) can be used as an alternative. In this case, the parameter is obtained by maximizing the rank-based log-likelihood:

$$l(\Theta) = \sum_{i=1}^n \log \left[c_\Theta \left(\frac{r_i}{n+1}, \frac{s_i}{n+1} \right) \right]. \quad (9)$$

3.2.2 Copula goodness-of-fit tests

Goodness-of-fit tests for copulas are applied comparing the empirical copula $C_n(u, v)$ of a n -tuple of random variables (Eq.A1) with the parametric estimate of a theoretical copula model C_θ derived under the null hypothesis. There are different goodness-of-fit tests as e.g. reviewed by Genest and Rémillard (2008); Genest et al. (2009). One of the tests used in this study is based on the Cramér-von Mises statistic (Genest and Favre, 2007):

$$S_n = n \sum_{i=1}^n \{C_\theta(u_i, v_i) - C_n(u_i, v_i)\}^2. \quad (10)$$

As the definition of S_n involves the theoretical copula function, the distribution of the test statistic depends on the unknown value of θ under the null hypothesis that C is from the class C_θ (Grégoire et al., 2008).

Additional goodness-of-fit tests are based on Kendall's transform using K-functions (e.g. Barbe et al., 1996; Genest et al., 2009), and for visual inspection K-plots (e.g. Genest and Favre, 2007) are consulted.

3.2.3 Simulating from copula distributions

In practice, modelling random samples of realisations from the dependence structure of a bivariate dataset consists of several steps. First, the best copula model has to be determined according to the following algorithm:

1. Check the data whether the "iid" condition is valid or not, and in case of violation, apply e.g. an ARMA-GARCH transformation;
2. Transform the positive pairs of the data (or of the ARMA-GARCH residuals) (x_i, y_i) to rank space (r_i, s_i) with $i = 1, \dots, n$ denoting the length of the dataset;
3. Calculate the empirical copula $C_n(u, v)$ based on the ranks (r_i, s_i) ;

4. Estimate the copula parameters for different types of theoretical (Archimedean) copula functions using maximum pseudolikelihood or Kendall's τ ;
5. Apply different selection techniques (e.g. AIC) and suitable goodness-of-fit tests to choose the appropriate copula family.

With the calculated copula dependence function and the marginal distributions – i.e. the copula model consisting in $F_X(x)$, $F_Y(y)$ and $c_\theta(u, v)$ – conditional random samples, called pseudo-observations in the sequel, can be generated using the conditional probability density functions.

The steps for simulating pseudo-observation from observed data are detailed as follows:

1. Compute $u = F_X(x)$, where x is one realisation of the variate X and $F_X(x)$ denotes its marginal distribution;
2. Generate random samples for the variate v^* from the conditional PDF $c_{Y|U}(v|u) = c_u(v)$ and calculate $v = c_u^{-1}(v^*)$, where c_u^{-1} denotes the generalized inverse of c_u (Nelsen, 1999);
3. Calculate the corresponding y -values using the probability integral transformation $F_Y^{-1}(v) = y$, with $F_Y(y)$ being the marginal distribution of the variate Y .

3.3 Copula-based bias correction

Basis for the hereinafter presented copula-based bias correction methods is the estimated copula models derived for all positive radar/gauge pairs in the domain of interest. The copula models are characterized by their copula parameters, which can be visualized as copula maps. For copulas with only one parameter, the copula maps show the magnitude of dependence between gauge stations and corresponding/surrounding radar grids for each station in the domain. Thus, the map is a reflection of the spatial dependence structure between radar and gauge data across the radar domain. To assimilate the statistical characteristics of all gauge stations at the same time, the copula models have to be combined in an appropriate way. Figure 3 schematically illustrates the different methods for one single time step. Note that the *Maximum Theta* approach is only applicable in cases where one single and one-parametric copula model is suitable for all radar/gauge pairs. The *Multiple Theta* approach is flexible in terms of the number of theoretical copula functions and the dimension of their respective parameter space.

In any case, the starting point is the ensemble of θ maps and of estimated marginal distributions. In the sequel, the two different approaches are described in more detail.

3.3.1 Multiple Theta approach

The *Multiple Theta* approach can be applied for a theoretical copula with a parameter set of $\Theta \in \mathbb{R}^n$. The algorithm consists of the following steps:

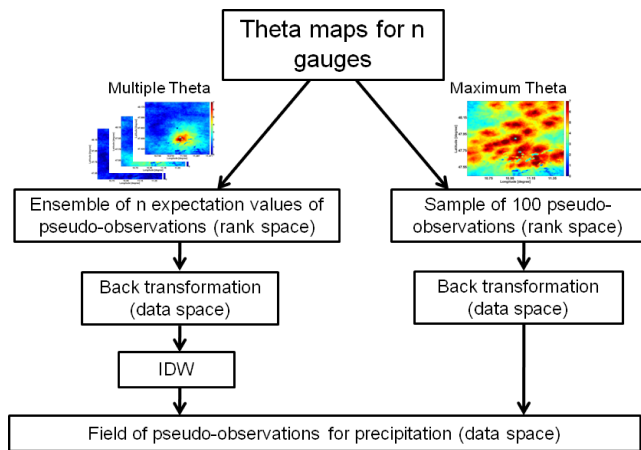


Fig. 3. Overview over the copula-based interpolation methods.

1. Based on the available gauge stations, 31 copula maps are derived, resulting in 31 sets of copula parameters $\Theta \in \mathbb{R}^n$ for one specific radar grid cell.
2. 31 copula parameter sets, 31 marginal distributions of the respective gauge stations and one single marginal distribution of the specific radar grid cell are assigned to a specific radar grid cell.
3. For each set (marginal distribution gauge, marginal distribution radar and copula parameters), a sample of 100 members is simulated in the rank space.
4. The expectation values are calculated from the random samples.
5. The integral transformation is applied to the calculated expectation values to transform back to data space.
6. Inverse distance weighting (IDW) is applied to generate one single value for each radar grid cell.
7. These steps are repeated for all radar grid cells.
8. A field containing the IDW single value for each radar grid cell is obtained.

3.3.2 Maximum Theta approach

The *Maximum Theta* can be applied only in the case of theoretical copulas with one parameter $\theta \in \mathbb{R}$. The approach comprises the following steps:

1. Based on the available gauge stations, 31 copula maps are derived, resulting in a set of 31 copula parameters for a specific radar grid cell.
2. The set showing the maximum copula parameter is assigned to a specific grid cell, retaining the information of the copula parameter and the corresponding marginals (gauge and radar).

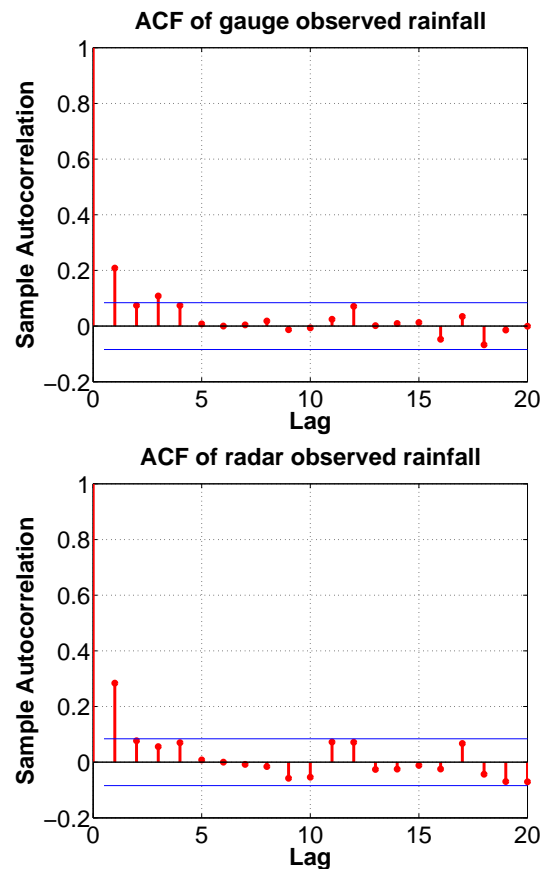


Fig. 4. Autocorrelation function for gauge (top) and corresponding radar grid (bottom) at the station Garmisch-Partenkirchen.

3. One sample of 100 members is simulated in the rank space for this set.
4. The expectation value is calculated from the random sample.
5. The integral transformation is applied to the expectation value to transform back to data space.
6. These steps are repeated for all radar grid cells.
7. A field containing the expectation value for each radar grid cell is obtained.

This approach is restricted to one-parametric copula families as the maximum, required in step 2 of the proposed algorithm, cannot be interpreted as a measure for the strength of the dependence between the marginals for higher dimensional parameter spaces.

3.4 Validation measures

Once fields of pseudo-observations are generated, the efficiency of the copula-based approach has to be tested. The

Table 3. Validation measures used in this study.

Validation measure	Abbreviation	Formula	Range	Perfect Fit
Kendall rank correlation	τ	$\tau = \frac{n_c - n_d}{2^{-1}n(n-1)}$	$[-1, 1]$	$ \tau = 1$
Pearson correlation	r	$r = \frac{\sum_{i=1}^n (o_i - \bar{o}_i)(m_i - \bar{m}_i)}{\sqrt{\sum_{i=1}^n (o_i - \bar{o}_i)^2} \sqrt{\sum_{i=1}^n (m_i - \bar{m}_i)^2}}$	$[-1, 1]$	$ r = 1$
Root-mean-square error	RMSE	$\sqrt{\frac{1}{n} \sum_{i=1}^n (o_i - m_i)^2}$	$[0, \infty[$	RMSE = 0
Nash-Sutcliffe efficiency	NSE	$1 - \frac{\sum_{i=1}^n (o_i - m_i)^2}{\sum_{i=1}^n (o_i - \bar{o}_i)^2}$	$] - \infty, 1]$	NSE = 1

sources that have to be compared are the original radar measurements, the information from the different gauges in the domain and the simulated fields. In order to test the performance of the simulated field, ordinary kriging is applied to the gauge information exclusively to derive an interpolated field. Kriging uses the variogram of the regionalized variable (here precipitation), i.e. the variance between pairs of points that lie different distances apart. The best estimate of the values (BLUE – Best Linear Unbiased Estimator) is calculated considering the layout of the observation network relative to the interpolation grid. The major assumption of ordinary kriging is that the expected value of the regionalized variable is constant across the interpolated precipitation field. This is not the case for precipitation in the Alpine region. Nevertheless, it is often used in a pragmatic way to obtain a first guess on the spatial distribution of rainfall.

Besides a purely visual inspection of the resulting fields, a quantitative validation is done point-wise, using a cross-validation approach. Table 3 shows the different efficiency criteria used in this study, with o_i denoting the value of the observations, m_i the value of the model at time step $i = 1, \dots, n$ and n_c, n_d the number of the concordant/disconcordant values.

4 Results

To apply the *Multiple Theta* and *Maximum Theta* algorithms described above, “iid” data are required. However, as can be seen from Fig. 4, both radar and gauge data show autocorrelation.

Therefore, ARMA-GARCH composite models are fitted to the radar and gauge time series to generate “iid” residuals. In this two-step transformation (Laux et al., 2011), the ARMA model compensates for autocorrelation, while the GARCH (P, Q) model reduces the heteroskedasticity within the respective time series. Following their method, the order of the AR, MA, P and Q components is varied from 0 to 3 and the autocorrelation function and the Ljung-Box Q-test (Box et al., 1994) are applied to test the residuals of the hourly time series for remaining serial dependence and heteroskedasticity. It can be seen from Table 4 that both radar

and gauge data are serially correlated at least up to lag 20. Heteroskedasticity is mainly detected for radar data. Both effects could be removed successfully using an ARMA (1,1)-GARCH (1,1) time series model.

4.1 Marginal distributions

As explained in Sect. 3.2.3 for the probability integral transformation, the single marginal distributions are needed. The appropriate distribution function is estimated for all radar grids and all gauge stations separately. In general, the algorithm presented in Sect. 3.3 allows to fit different theoretical marginal distribution functions. Four univariate distribution functions are considered: normal, exponential, gamma and Weibull. After estimation of the parameters by a maximum likelihood approach, the Kolmogorov-Smirnov and the Chi-Squared goodness-of-fit tests are used. These tests only allow for rejecting the normal distribution at significance levels 0.05 and 0.01. To come to a final decision about the distribution, the AIC/BIC values are calculated for each radar grid and all gauge stations. Table 5 lists exemplarily the results for the stations Garmisch-Partenkirchen, Oberammergau, and Wielenbach, indicating that the Weibull distribution provides the best fit, followed by the gamma and the exponential distribution. It is obvious that neither radar nor gauge time series are normally distributed.

The results shown in Table 5 are representative for the 31 gauge stations where the AIC/BIC criterion confirmed the Weibull distribution to provide the best fit in all cases. This is consistent with the findings for the 10 000 radar grids. The AIC (BIC) criterion allowed to decide for the Weibull distribution for 99.1 % (98.8 %) of the cases, while only 0.9 % (0.8 %) could be assigned to the gamma distribution. Following the results of the goodness-of-fit tests, the Weibull distribution can be considered for all time series without loss of generality.

4.2 Fitting theoretical copula functions

Once the univariate marginal distributions are fitted, the dependence structure between the time series has to be investigated. The first step is to calculate the empirical copulas

Table 4. Ljung-Box Q-test results for three observation stations and their corresponding radar grid cells. The test indicates whether or not the time series exhibit autocorrelation for a given number of lags [hours]. One tests the null hypothesis that a series exhibits no autocorrelation for a fixed number of lags against the alternative hypothesis that the autocorrelation is nonzero. “1” indicates that the null hypothesis is rejected (i.e. autocorrelation (original) and heteroskedasticity (squared)), and “0” indicates no autocorrelation/heteroskedasticity for any given time lag and level of significance (significant at $\alpha = 0.05$ (bold), or $\alpha = 0.01$ level of significance (normal)).

Station	lag [h]	without ARMA-GARCH					with ARMA-GARCH				
		1	5	10	15	20	1	5	10	15	20
Garmisch-Partenkirchen	Gauge (original)	1	1	1	1	1	1	0	0	0	0
	Gauge (squared)	0	0	0	0	0	0	0	0	0	0
	Radar (original)	1	1	1	1	1	0	0	0	0	0
	Radar (squared)	1	1	1	1	1	0	0	0	0	1
Oberammergau	Gauge (original)	1	1	1	1	1	0	0	0	0	0
	Gauge (squared)	0	0	0	1	0	0	0	0	0	0
	Radar (original)	1	1	1	1	1	0	0	0	0	0
	Radar (squared)	1	1	1	1	1	0	0	0	0	0
Wielenbach	Gauge (original)	1	1	1	1	1	0	0	0	0	0
	Gauge (squared)	1	1	0	0	0	0	0	0	0	0
	Radar (original)	1	1	1	1	0	0	0	0	0	0
	Radar (squared)	0	0	0	0	0	0	0	0	0	0

Table 5. Akaike and Bayesian information criteria (AIC and BIC) of radar and gauge (positive pairs only) for selected rain gauge stations and different univariate distribution functions (June, July, and August of 2006 and 2007). Smallest values of AIC/BIC indicate the best fit (bold).

Station		Normal	Exponential	Gamma	Weibull
Garmisch-Partenkirchen	AIC (Radar)	3137	1809	1783	1740
	BIC (Radar)	3149	1815	1794	1751
	AIC (Gauge)	2215	1294	1274	1251
	BIC (Gauge)	3139	1895	1871	1842
Oberammergau	AIC (Radar)	3153	1962	1918	1880
	BIC (Radar)	3164	1968	1929	1891
	AIC (Gauge)	2958	1835	1811	1787
	BIC (Gauge)	3013	1861	1835	1807
Wielenbach	AIC (Radar)	2222	1420	1417	1405
	BIC (Radar)	2233	1425	1427	1416
	AIC (Gauge)	2060	1171	1138	1110
	BIC (Gauge)	2590	1497	1467	1434

for the radar/gauge pairs and then fit a theoretical bivariate copula function with its parameters estimated by either using Kendall's τ or a maximum-pseudolikelihood approach. In this study, five different theoretical copula functions are tested: Gaussian, Student-T, Frank, Clayton and Gumbel-Hougaard. As the Student-T copula has two parameters, the maximum-pseudolikelihood approach is preferable and therefore it is used for all theoretical copulas to ensure consistency. Only the positive pairs of radar and gauge are considered with 0.1 mm being the threshold for a rainy day. Figure 5 shows the probability density functions of the empirical and the estimated theoretical Gumbel-Hougaard copula for radar-gauge pairs exemplarily for the station Garmisch-

Partenkirchen. The empirical copula density is asymmetric with respect to the minor diagonal of the unit square. The density is highest for the upper right corner and shows a second maximum in the lower left corner, indicating a strong upper and lower tail dependence.

For the five different theoretical copula functions, the goodness-of-fit test, which is based on the the Cramér-von Mises statistic, is applied (Genest and Rémillard, 2008). 1000 values of the test statistic are sampled, and the proportion of values larger than S_n is estimated by calculating the corresponding p-values. The results based on S_n only allow to exclude the Gaussian and the Student-T copula, while the three Archimedean copulas, namely Frank, Clayton and

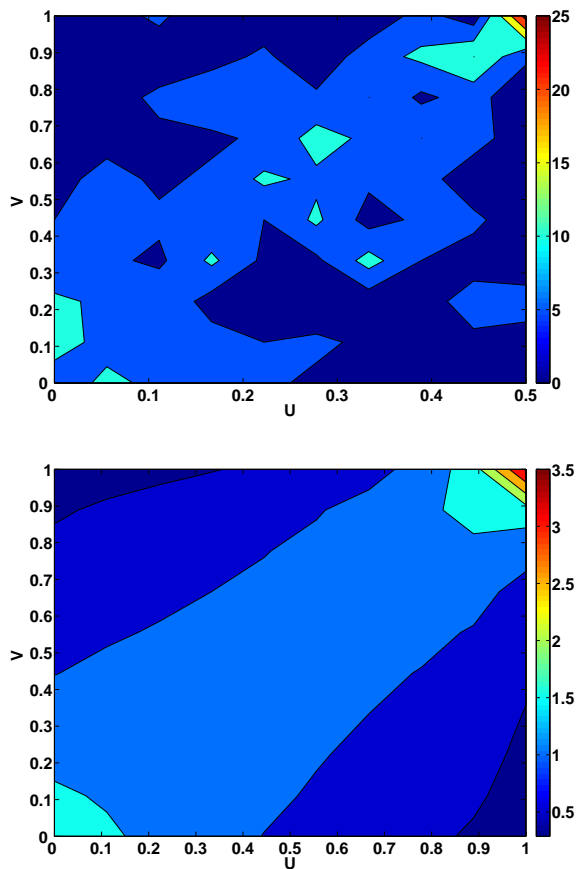


Fig. 5. Probability density function of the empirical copula (top) and theoretical Gumbel-Hougaard copula (bottom) for the station Garmisch-Partenkirchen.

Gumbel-Hougaard, perform similarly well. To decide for one of them, K-functions are used.

The results of the tests are shown for Garmisch-Partenkirchen, Oberammergau and Wielenbach (Table 6), indicating that the Gumbel-Hougaard copula provides the best fit. $\sim 95\%$ out of all radar/gauge pairs (31 stations times 10 000 radar grids) have been identified as the best fit. The remainder of 5 % is assigned to the Frank copula. Following the results of the goodness-of-fit analysis, it can be assumed that the accuracy of the proposed algorithm is not reduced significantly by constraining the model to one specific theoretical copula function. Please note that, for the *Maximum Theta* approach, one single one-parametric copula model must be used to ensure the comparability of the copula maps. Nevertheless, the corresponding copula parameter has to be estimated individually.

4.3 Copula-based simulation of precipitation fields

In this section, the results for the copula-based simulation of precipitation fields are shown, starting with the estimation of the copula maps, followed by the description of

the two different simulation algorithms (*Maximum Theta* and *Multiple Theta*). Bias-corrected precipitation fields are shown for one arbitrarily chosen time step, and the obtained results are validated by visual inspection and different performance measures.

4.3.1 Copula maps

As described in Sect. 3.3, the starting point of the copula-based precipitation simulation is a precipitation field derived from radar reflectivities. The radar field, derived by the standard Z/R relationship, is usually disturbed by e.g. effects of backscatter or spokes caused by obstacles shading the radar beam (see e.g. Fig. 2). Nevertheless, the measured radar field gives a realistic impression of the spatial distribution of precipitation at that certain time step. Even if the precipitation patterns are realistically displayed, the radar does not accurately reproduce the absolute precipitation amounts on the ground. In order to correct the radar field in terms of the absolute precipitation amounts, ground-based measurements are assimilated. The parameter of the Gumbel-Hougaard copula is estimated between the ARMA-GARCH residuals of one selected gauge and all grid cells of the radar field. All copula parameters θ are visualized together building the so-called copula map corresponding to the respective station. These maps show the strength of the dependence between radar and gauges. Figure 6 shows the copula maps for station Garmisch-Partenkirchen (top) and Wielenbach (bottom). In general, the correlation between radar and gauge time series decreases with increasing distance. For station Wielenbach e.g., the copula map is almost radially symmetric. However, considering all available gauge stations in the domain, there are also cases where the dependence structure is highly anisotropic such as for Garmisch-Partenkirchen (see Fig. 6).

For the *Maximum Theta* approach, all 31 available copula maps are combined in the first step. To assign the maximum value out of the set of all available copula parameters to each grid cell is only possible if the goodness-of-fit tests allow to identify one single and one-parametric theoretical copula model for all radar/gauge pairs in the domain. This is the case for the data considered in this study, and Fig. 7 shows the *Maximum Theta* map (estimated parameter of the Gumbel-Hougaard copula) for the whole domain. The asymmetries from the single station maps are reflected by the field, considering all variabilities in the dependence structures.

The anisotropic nature of the dependence structure has to be considered when copula-based precipitation fields are modelled. The results of the two different approaches, namely *Multiple Theta* and *Maximum Theta*, are presented in the sequel.

Table 6. Value of the K-function times 10^4 for radar/gauge pairs at the stations Garmisch-Partenkirchen, Oberammergau and Wielenbach for three Archimedean copulas (Gumbel-Hougaard, Frank, Clayton). Smallest values indicate the best fit (**bold**).

Station	Gumbel-Hougaard	Frank	Clayton
Garmisch-Partenkirchen	1.9	3.8	6.5
Oberammergau	3.5	7.6	12.6
Wielenbach	1.2	7.4	21.3

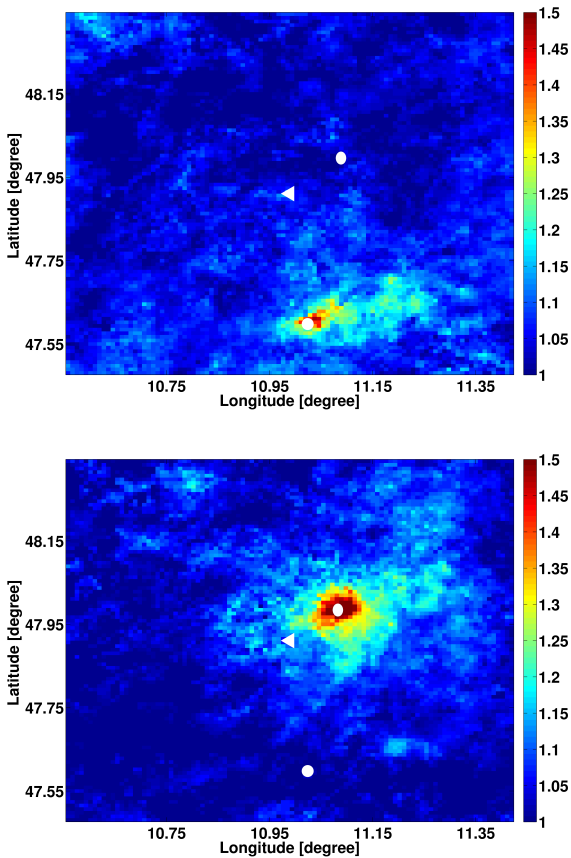


Fig. 6. Copula maps for stations Garmisch-Partenkirchen (top) and Wielenbach (bottom) showing the parameter θ of the Gumbel-Hougaard copula. The position of Hohenpeissenberg is indicated by a white triangle, Garmisch-Partenkirchen (in the south) and Wielenbach (in the north) by a white circle.

4.3.2 Multiple Theta

As described in Sect. 3.3, the single copula maps based on the radar-gauge pairs are combined to simulate a bias-corrected precipitation field.

A precipitation field generated using the *Multiple Theta* approach is shown in Fig. 8. This field is based on 31×100 realisations for each grid cell. The pattern information from the radar precipitation field is preserved, and the absolute values are corrected towards the ground measurements. For the chosen time step, the rainfall intensities and variances are

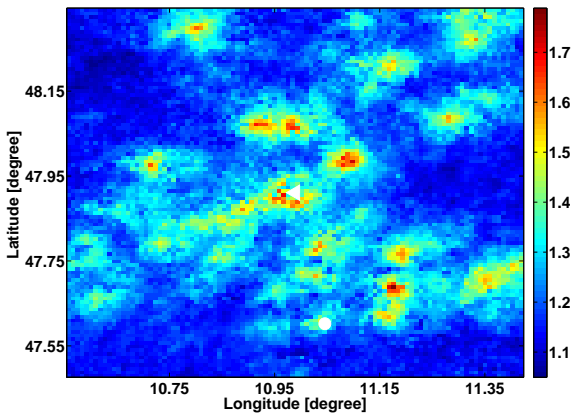


Fig. 7. *Maximum Theta* map derived from all gauge stations in the domain. The position of Hohenpeissenberg (Garmisch-Partenkirchen) is indicated by a white triangle (circle).

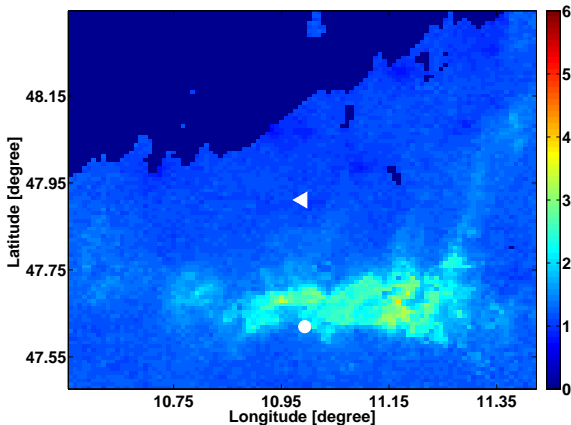


Fig. 8. Precipitation field [mm h^{-1}] in the data space derived with the *Multiple Theta* method for 14 July 2008 (13:00). The position of Hohenpeissenberg (Garmisch-Partenkirchen) is indicated by a white triangle (circle).

reduced. Spokes and backscattering effects could be diminished compared to the original radar image (see Fig. 2).

4.3.3 Maximum Theta

Based on the *Maximum Theta* map, 100 random realisations (rank space) are generated for each grid cell conditioned on the respective radar value. The resulting field is shown in

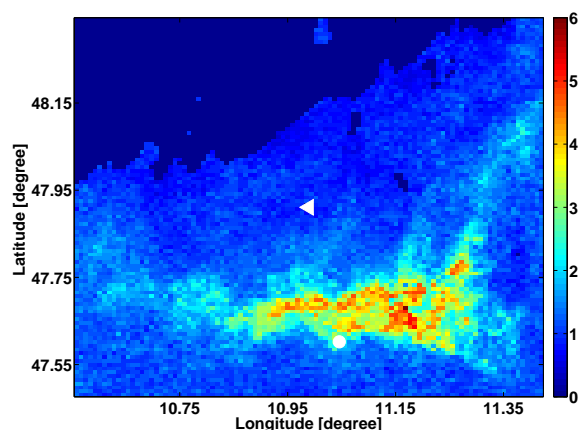


Fig. 9. Precipitation field [mm h^{-1}] in the data space derived with the *Maximum Theta* method for 14 July 2008 (13:00). The position of Hohenpeissenberg (Garmisch-Partenkirchen) is indicated by a white triangle (circle).

Fig. 9. Compared to the results shown for the *Multiple Theta* approach, the patterns from the radar are reproduced similarly well. More small-scale variability is resolved compared to the radar field. Compared to the *Multiple Theta* method, more details are retained and the absolute values are slightly higher.

4.4 Validation of the simulated precipitation fields

4.4.1 Visual inspection

Figure 10 shows an interpolated rainfall field derived from the observed precipitation values on 14 July 2008 (13:00). The observations from the 31 gauge stations in the radar domain are interpolated by application of an ordinary kriging approach (e.g. de Marsily, 1986; Kitanidis, 1997).

Albeit the major assumption of ordinary kriging is not valid throughout the whole research domain, i.e. the expectation value is not constant over space, the interpolated precipitation field roughly reproduces the rainfall patterns revealed by the radar measurement (compare to Fig. 2). The fine structures of the rainband are remarkably smoothed, and the rainfall is concentrated in three local maxima. As gauge stations are only available for the center of the domain (see Fig. 1), there is no information included for the outer regions. This is the reason why the interpolated field cannot reproduce the variability of precipitation in that area. In contrast to the kriging field, the copula-based precipitation fields (see Figs. 8 and 9) not only include information from the rain gauges but also incorporate information from the radar measurement. Consequently, they resolve better the fine structures in the overall pattern of precipitation.

Figure 11 shows the accumulated summer precipitation obtained from radar measurements at Hohenpeissenberg derived by a standard *Z/R*-relationship, the RADOLAN cor-

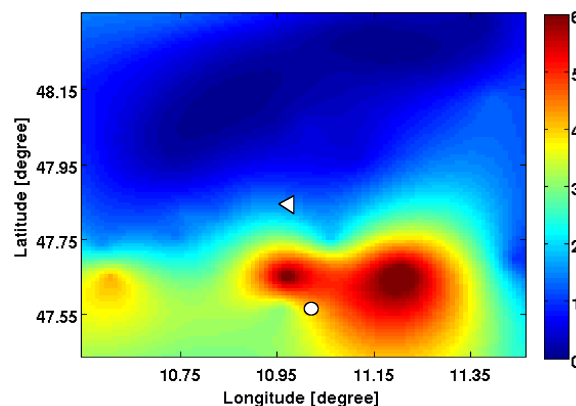


Fig. 10. Interpolated rainfall field [mm h^{-1}] derived from the observed precipitation values on 14 July 2008 (13:00). The observations from the 31 gauge stations in the radar domain are interpolated by application of an ordinary kriging approach. The position of Hohenpeissenberg (Garmisch-Partenkirchen) is indicated by a white triangle (circle).

rection of DWD, and simulated by the *Maximum Theta* and *Multiple Theta* approach for the years 2005–2008. Compared to the rain gauges, the radar observes the liquid and solid water in a certain volume of the atmosphere, which only partly reaches the ground as precipitation. Therefore, the total amount of precipitation measured by the gauges is significantly different from that seen by the radar. In general, the cumulated amount of precipitation obtained by radar is overestimated. This effect is supposed to be stronger in wet summers such as 2006 and 2007. For these seasons, the copula-based methodologies *Maximum Theta* and *Multiple Theta* reduce significantly the total amount of rainfall, while for dry summers the spatial representation of annual amounts is similar for all approaches. While errors due to shading effects are accumulated over time in the radar (standard *Z/R* relationship) and the RADOLAN-corrected fields, these artifacts are reduced by the copula-based approaches.

4.4.2 Quantitative validation

The results of the *Maximum Theta* and *Multiple Theta* approaches are checked by cross-validation for all 31 stations, using the time series of gauge and pseudo-observations in the respective grid cell. Different validation measures are used to obtain a quantitative appraisal for the performance of the simulation algorithms (see Table A). The mean correlation between the simulated pseudo-observations (the respective gauge station is not included in the analysis) and the observed precipitation is 0.6 for both proposed methods indicating equally reasonable results. However, the results for RMSE, MAE and MSE show that the *Multiple Theta* approach is slightly superior. This finding is also supported by the inspection of the respective mean NSE values, being 0.2 for the *Maximum Theta* and 0.1 for the *Multiple Theta*

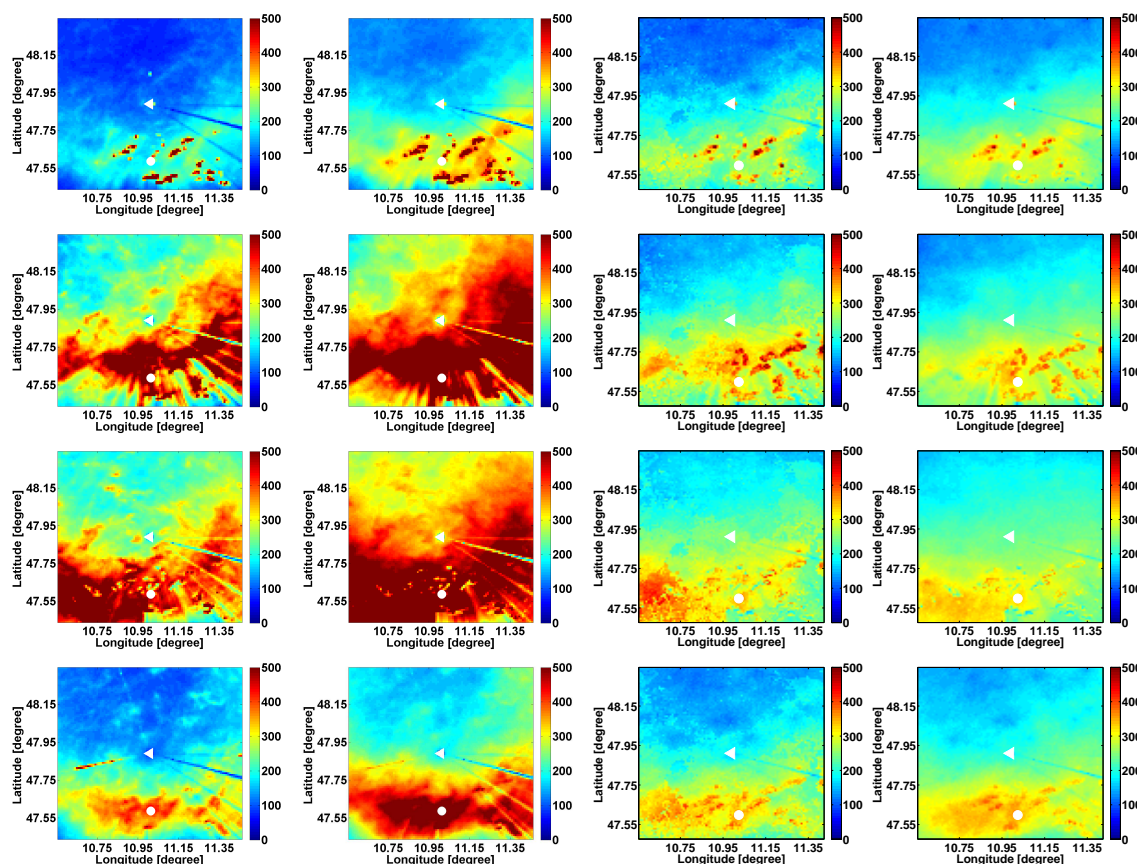


Fig. 11. Accumulated rainfall for summer [mm] obtained by radar (standard Z/R relationship), RADOLAN-correction method, *Maximum Theta*, and *Multiple Theta* approach (from left to right) for the years 2005, 2006, 2007, and 2008 (from top to bottom).

method. Note that the validation results for the station Hohenpeissenberg are not included in the mean values as this station is located close to the radar observatory. For this grid cell, the radar precipitation is significantly biased, resulting in a very low correlation, negative NSE and the highest values for RMSE, MAE and MSE among all gauge stations. It is also important to appreciate how qualified the copula-based bias-correction is compared to operational correction methods. Therefore, precipitation time series corrected with RADOLAN are used for comparison.

Table 7 shows the results of the comparison of radar precipitation derived with the standard Z/R -relationship and the DWD RADOLAN-correction, pseudo-observations simulated with the *Maximum Theta* approach and gauge observations for the three arbitrarily chosen stations Garmisch-Partenkirchen, Oberammergau and Wielenbach (positive pairs, summer 2005–2008). The correlations do not differ significantly between the three methods. However, the RMSE and the NSE indicate that the copula-based approach can improve the results of the standard Z/R relationship similarly well as the RADOLAN correction. Visual inspection of the respective time series shows that high precipitation values adhere to the highest biases. These are most effectively

corrected by the copula approach, which is also confirmed by the calculated NSE.

Overall, the *Maximum Theta* method is found to be an expedient alternative to the RADOLAN correction with equally good performance.

5 Discussion

Radar precipitation fields are a good supplement to gauge measurements as they provide a reasonable representation of spatial rainfall distribution. Nevertheless, precipitation fields from radar reflectivities incorporate significant errors due to the measurement itself (backscatter, shading etc.), the applied Z/R -relationship or the fact that they do not provide values at the ground level. There are sophisticated operational bias-correction algorithms for radar products such as the RADOLAN, which contains orographic shading correction, statistical reduction of clutter, gradient smoothing and further preprocessing steps and assimilates data from ombrometer measurements. Here a purely statistical copula-based approach to assimilate gauge information for bias-correction of radar precipitation fields is proposed, using

Table 7. Comparison of radar precipitation derived with the standard Z/R-relationship, the DWD RADOLAN-correction, pseudo-observations simulated with the *Maximum Theta* approach (all 31 stations included) to gauge observations for Garmisch-Partenkirchen, Oberammergau and Wielenbach (positive pairs, summer 2005–2008).

Methods	Garmisch-P.				Oberammergau				Wielenbach			
	τ [–]	r [–]	RMSE [mm h ^{–1}]	NSE [–]	τ [–]	r [–]	RMSE [mm h ^{–1}]	NSE [–]	τ [–]	r [–]	RMSE [mm h ^{–1}]	NSE [–]
Standard Z/R	0.51	0.64	1.74	0.21	0.49	0.62	1.77	0.21	0.48	0.53	1.72	0.20
RADOLAN	0.51	0.66	1.66	0.28	0.49	0.64	1.66	0.30	0.48	0.58	1.64	0.27
Copula-based	0.47	0.61	1.65	0.29	0.46	0.62	1.67	0.30	0.43	0.55	1.69	0.23

positive radar/gauge pairs only. In general, there exist four cases, namely (0,0), (0,1), (1,0) and (1,1), with 0 denoting dry and 1 wet conditions. It is known that a high number of low precipitation values is measured by the radar that cannot be reaffirmed by ground-based observations. These presumably erroneous datasets are removed by the application of a threshold value, being 0.01 mm for a rain event in this study. Note that the case where the radar does not detect precipitation correctly, i.e. (0,1), cannot be corrected by the proposed algorithms and has to be examined separately. However, a thorough investigation of the radar/gauge data shows that these cases only contribute negligibly as the radar measurement in general tends to overestimate precipitation. This is due to the fact that both liquid and solid water in a certain volume of the atmosphere is observed by the radar, which only partly reaches the ground as precipitation. Therefore, using only positive pairs to derive the dependence structures does not restrict the applicability of the algorithms substantially.

Recently Laux et al. (2011) showed that daily precipitation data incorporate autocorrelation and heteroskedasticity and therefore are not “iid”. They showed that an ARMA-GARCH algorithm can be used to transform the time series so that copula-based methods can be applied.

A thorough investigation of the hourly radar and gauge precipitation used in this study reveals that these hourly data also incorporate autocorrelation and heteroskedasticity to some degree. It is shown by the results of an Ljung-Box Q-test that an ARMA (1,1)-GARCH (1,1) time series model is sufficient to successfully generate “iid” residuals, which are the prerequisite for all copula-based applications.

The copula-based bias correction methods presented in this study tend to apply a pairwise copula analysis to the positive pairs of the data and incorporate anisotropies in the dependence structure by combination of the derived copula maps. While the *Maximum Theta* approach is limited to applications where only one single, one-parametric theoretical copula function can be fitted to the data, the *Multiple Theta* approach is fully flexible with respect to the number of theoretical copula functions and the dimension of their respective parameter space. For the data that were used in this study, one single copula, namely the Gumbel-Hougaard copula, can be

identified by goodness-of-fit tests. As this copula function is one-parametric, the copula maps can be nicely visualized.

In general, the copula parameter (i.e. correlation) for a radar/gauge pair decreases with increasing distance between the respective gauge station and radar grid cell. Besides, the copula maps show more details of the dependence structure. For certain gauge stations (see Fig. 6), the θ -field is strongly asymmetric, revealing the influence of orography on the dependence structure. This asymmetry is disregarded by standard interpolation methods such as inverse distance weighting, although it is getting more important in complex terrain.

Asymmetries in the dependence structure can also be an indicator of air flow directions, dominant for a certain location. This theory could be supported by investigation of different time scales: for small time scales (< 1 h), localized rainfall events are not resolved by the statistical analysis. In that case, the region of strongly correlated grid cells, visualized through high copula parameters in the copula map, is expected to be reduced compared to larger time scales.

For different seasons, different types of rainfall regimes are predominant. Therefore, the investigation of e.g. summer and winter season separately is expected to reveal preferential precipitation types, differentiating large-scale winter precipitation from convective summer events.

As a result from the goodness-of-fit tests for marginal distributions and theoretical copula functions in this study, one single type of marginal distribution (Weibull) and one single theoretical copula function (Gumbel-Hougaard copula) are used for all stations and all radar grid cells.

Both the *Maximum Theta* and the *Multiple Theta* method allow for estimating different marginal distributions for each time series, while only for the *Multiple Theta* approach is it possible to consider more than one theoretical copula function and multidimensional copula parameter spaces. As the Gumbel-Hougaard copula, identified by the goodness-of-fit tests, is one-parametric, both the *Maximum Theta* and the *Multiple Theta* approach can be applied in this study to generate bias-corrected precipitation fields.

Before the integral transformation is applied to the simulated fields, the marginals do not contribute to the simulation results. Therefore, the performance of the copula model itself can be assessed in the rank space. It is found that the

correlation between the pseudo-observations and gauges is significantly higher in that case. Hence, the choice of the marginals is of crucial importance for the overall model results.

Finally, it is shown that the proposed bias-correction methods are able to combine the advantages of the two data sources: the rainfall patterns observed by the radar measurement are retained in the simulated field, while the absolute values are successfully corrected towards the gauge observations.

The results of a cross-validation show that the *Maximum Theta* approach performs slightly better than the *Multiple Theta* method. This is due to the fact that, for each radar grid cell, the simulation is based on the highest θ and the respective marginal distribution, while for the *Multiple Theta* case stations with low correlation they also slightly contribute to the simulated results. However, both methods are potentially useful as not for every application it may be possible to use one single copula function.

As all ingredients for the proposed algorithms, namely the marginal distributions and the theoretical copula functions, they only have to be estimated once the proposed bias-correction methods are computationally not very demanding, which facilitates operational application of the proposed methods in quasi real-time. The effectiveness of the *Maximum Theta* approach is also emphasized by the results of the comparison with the RADOLAN correction. The correlations between corrected rainfall and gauge measurements are nearly the same for the two approaches, and RMSE and NSE indicate that the performance of the copula-based approach is equally good.

6 Conclusions

Two new copula-based methods are proposed to bias-correct radar precipitation fields by assimilation of gauge information. As copula-based applications require “iid” variables, an ARMA-GARCH model is used to generate “iid” residuals. After this transformation for each gauge station and all radar grid cells, a copula map is derived revealing anisotropies in the dependence structure. Asymmetries in the copula maps reflect the complexity of the terrain as well as predominant flow directions and rainfall types.

The goodness-of-fit tests suggest that it is possible to constrain the copula model to one single, one-parametric theoretical copula function, which is a requirement for the *Maximum Theta* approach. The *Multiple Theta* approach is not restricted and can be applied also for more than one theoretical copula family and copulas with multidimensional parameter spaces. The copula-based methods are suitable to merge the advantages of the different data sources: the spatial distribution of the radar rainfall field is preserved, while absolute values are corrected towards gauge observations. As a consequence, they can be used as a computationally

not very demanding alternative method to bias-correct radar precipitation fields. Considering the performance measures applied in this study, the *Maximum Theta* approach models the gauge values with the same performance as the bias-correction method RADOLAN, used as standard procedure by DWD.

Appendix A

Empirical and theoretical copulas

The empirical copula $C_n(u, v)$, which is defined on the ranks of the data, is an estimator for the unknown theoretical copula distribution $C_\theta(u, v)$ associated with the pair (X, Y) , having a set of parameters θ :

$$C_n(u, v) = 1/n \sum_{i=1}^n 1 \left(\frac{r_i}{n+1} \leq u, \frac{s_i}{n+1} \leq v \right) \quad (\text{A1})$$

where $(r_1, s_1), \dots, (r_n, s_n)$ denote the pairs of ranks of the data $(x_1, y_1), \dots, (x_n, y_n)$, and $1(\dots)$ is the indicator function.

The density of a theoretical copula $C(u, v)$ is calculated as

$$c_\theta(u, v) = \frac{\partial^2 C(u, v)}{\partial u \partial v}. \quad (\text{A2})$$

A special family of copula functions is the so-called Archimedean copulas. Some of the copula functions discussed in this study are members of this family.

Let $\varphi: [0, 1] \rightarrow [0, \infty]$ a steady, strict monotonic function with $\varphi(1) = 0$ and let $\varphi^{[-1]}: [0, \infty] \rightarrow [0, 1]$ the pseudo-inverse of φ (Nelsen, 1999):

$$\varphi^{[-1]} := \begin{cases} \varphi^{-1}(t) & \text{if } 0 \leq t \leq \varphi(0), \\ 0 & \text{else} \end{cases} \quad (\text{A3})$$

then the function

$$C: [0, 1]^2 \rightarrow [0, 1] \\ (u, v) \mapsto \varphi^{[-1]}(\varphi(u) + \varphi(v)) \quad (\text{A4})$$

defines a copula only if φ is convex and φ is called the generator of the Archimedean copula C .

In this paper the following Archimedean copulas are used:

the Frank copula with $\varphi(t) = -\ln \left(\frac{e^{-\theta t} - 1}{e^{-\theta} - 1} \right)$ and $\theta > 0$:

$$C_\theta(u, v) = -\frac{1}{\theta} \ln \left(1 + \frac{(e^{-\theta u} - 1)(e^{-\theta v} - 1)}{e^{-\theta} - 1} \right) \quad (\text{A5})$$

the Clayton copula with $\varphi(t) = \frac{1}{\theta} (t^{-\theta} - 1)$ and $\theta > 0$

$$C_\theta(u, v) = (u^{-\theta} + v^{-\theta} - 1)^{-\frac{1}{\theta}} \quad (\text{A6})$$

Validation

Table B1. Validation of the *Multiple Theta* and *Maximum Theta* approach for the 31 stations (summer 2005–2008, wet days only).

ID	Station Name	Multiple Theta				Maximum Theta			
		τ [–]	r [–]	RMSE [mm h ^{–1}]	NSE [–]	τ [–]	r [–]	RMSE [mm h ^{–1}]	NSE [–]
1	Bernbeuren-Prachtsried	0.47	0.58	1.90	0.09	0.44	0.57	1.80	0.17
2	Diessen	0.45	0.61	1.75	0.11	0.46	0.62	1.65	0.21
3	Deisenhofen	0.36	0.47	1.72	0.07	0.34	0.45	1.70	0.13
4	Ettal	0.31	0.52	1.74	0.06	0.36	0.53	1.70	0.10
5	Garmisch-Partenkirchen	0.46	0.59	1.84	0.11	0.46	0.60	1.73	0.21
6	Gilching	0.37	0.63	1.80	0.10	0.33	0.60	1.71	0.19
7	Griesen	0.41	0.57	1.56	0.11	0.36	0.54	1.50	0.18
8	Halblech	0.48	0.61	2.06	0.10	0.45	0.60	1.93	0.21
9	Hindelang	0.34	0.46	2.41	0.00	0.35	0.48	2.27	0.11
10	Hohenpeissenberg	–0.21	–0.25	3.36	–0.11	–0.20	–0.26	4.70	–1.20
11	Kaufbeuren	0.41	0.63	1.83	0.10	0.37	0.61	1.72	0.20
12	Kochel	0.49	0.63	1.86	0.11	0.44	0.60	1.76	0.20
13	Kohlgrub, Bad	0.50	0.68	1.98	0.10	0.46	0.67	1.83	0.23
14	Kraftsried	0.28	0.50	2.07	0.04	0.30	0.48	1.97	0.13
15	Kreuth	0.43	0.65	2.14	0.07	0.42	0.65	1.99	0.20
16	Krün	0.49	0.66	1.84	0.13	0.48	0.64	1.68	0.27
17	Lenggries	0.43	0.63	1.98	0.11	0.43	0.63	1.75	0.31
18	Maisach	0.30	0.56	1.80	0.07	0.37	0.58	1.70	0.18
19	Marktobersdorf	0.45	0.58	1.86	0.11	0.44	0.57	1.78	0.19
20	Munich	0.40	0.56	1.87	0.10	0.40	0.55	1.77	0.19
21	Oberammergau	0.51	0.65	1.81	0.15	0.45	0.66	1.74	0.23
22	Oberschleissheim	0.27	0.51	2.05	0.06	0.35	0.55	1.93	0.17
23	Oy	0.32	0.50	1.98	0.03	0.34	0.52	1.90	0.14
24	Schwangau	0.37	0.49	1.92	0.08	0.34	0.49	1.84	0.15
25	Seeg	0.48	0.53	2.16	0.10	0.45	0.54	2.03	0.20
26	Schäftlarn	0.48	0.60	1.87	0.09	0.43	0.60	1.75	0.21
27	Steingaden	0.54	0.68	2.06	0.14	0.49	0.68	1.83	0.32
28	Schwaben	0.36	0.53	2.00	0.06	0.36	0.53	1.94	0.11
29	Schlehdorf	0.41	0.56	1.91	0.07	0.44	0.57	1.77	0.20
30	Vilgertshofen	0.45	0.59	1.80	0.13	0.45	0.58	1.67	0.24
31	Wielenbach	0.42	0.53	1.69	0.11	0.36	0.52	1.60	0.19

and the Gumbel-Hougaard copula with generator $\varphi(t) = (-\ln(t))^\theta$ and $\theta > 1$

$$C_\theta(u, v) = e^{-((-\ln(u))^\theta + (-\ln(v))^\theta)^{\frac{1}{\theta}}}. \quad (\text{A7})$$

In addition to the Archimedean copulas, the bivariate Gaussian copula

$$C_\theta(u, v) = \int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} \frac{1}{2\pi\sqrt{1-\theta^2}} \exp\left(-\frac{s^2 - 2\theta st + t^2}{2(1-\theta^2)}\right) ds dt. \quad (\text{A8})$$

where $\theta \in [-1, 1]$, and Φ denotes the inverse of the univariate Gaussian distribution, and the Student-T copula

$$C_{\theta, \nu}(u, v) = \int_{-\infty}^{t_v^{-1}(u)} \int_{-\infty}^{t_v^{-1}(v)} \frac{1}{2\pi(1-\theta^2)^{\frac{1}{2}}} \left(1 + \frac{s^2 - 2\theta st + t^2}{\nu(1-\theta^2)}\right)^{-\frac{\nu+2}{2}} ds dt, \quad (\text{A9})$$

where t_v^{-1} is the inverse of the univariate Student-T distribution with ν degrees of freedom considered.

Acknowledgements. This research was funded by the Bavarian State Ministry of the Environment and Public Health (reference number VH-ID:32722/TUF01UF-32722), by the Federal Ministry of Education and Research as part of the research project LUCCi (reference number 01LL0908C), HGF (research project: Terrestrial

Environmental Observatories TERENO – <http://tereno.net>) and by the HGF Impuls und Vernetzungsfond (research project: Regional Precipitation Observation by Cellular Network Microwave Attenuation and Application to Water Resources Management, PROCEMA, Virtual Institute, reference number VH-VI-314). The help of the German Weather Service (DWD), which provided the rain gauge data and the radar data, is highly appreciated. The authors also thank Jörg Seltmann from DWD, Meteorological Observatory Hohenpeissenberg (MOHP), for technical support concerning the radar data.

The service charges for this open access publication have been covered by a Research Centre of the Helmholtz Association.

Edited by: C. de Michele

References

- AghaKouchak, A., Habib, E., and Bárdossy, A.: Modeling Radar Rainfall Estimation Uncertainties: Random Error Model, *J. Hydrol. Eng.*, 15, 265–275, 2010a.
- AghaKouchak, A., Bárdossy, A., and Habib, H.: Copula-based uncertainty modeling: Application to multi-sensor precipitation estimates, *Hydrol. Process.*, 24, 2111–2124, doi:10.1002/hyp.7632, 2010b.
- AghaKouchak, A., Bárdossy, A., and Habib, H.: Conditional simulation of remotely sensed rainfall data using a non-gaussian v-transformed copula, *Adv. Water Resour.*, 33, 624–634, doi:10.1016/j.advwatres.2010.02.010, 2010c.
- Barbe, P., Genest, C., Ghoudi, K., and Rémillard, B.: On Kendall's Process, *J. Multivariate Anal.*, 58, 197–229, 1996.
- Bárdossy, A.: Copula-based geostatistical models for ground-water quality parameters, *Water Resour. Res.*, 42, W11416, doi:10.1029/2005WR004754, 2006.
- Bárdossy, A. and Li, J.: Geostatistical interpolation using copulas, *Water Resour. Res.*, 44, W07412, doi:10.1029/2007WR006115, 2008.
- Bárdossy, A. and Pegram, G. G. S.: Copula based multisite model for daily precipitation simulation, *Hydrol. Earth Syst. Sci.*, 13, 2299–2314, doi:10.5194/hess-13-2299-2009, 2009.
- Bartels, H., Weigl, E., Reich, T., Lang, P., Wagner, A., Kohler, O., and Gerlach, N.: Projekt RADOLAN, Routineverfahren zur Online-Aneicherung der Radarniederschlagsdaten mit Hilfe von automatischen Bodenniederschlagsstationen (Ombrometer), Abschlussbericht, Deutscher Wetterdienst, 2004 (in German).
- Bollerslev, T.: Generalized autoregressive conditional heteroscedasticity, *J. Econometrics*, 31, 307–327, doi:10.1016/0304-4076(86)90063-1, 1986.
- Box, G. E., Jenkins, G. M., and Reinsel, G. C.: Time series analysis: forecasting and control, 3rd Edn., Prentice Hall Englewood Cliffs, New Jersey, 1994.
- Burrough, P. A. and McDonnell, R. A.: Principles of Geographical Information Systems, Oxford University Press, Oxford, 1998.
- Brandes, E. A.: Optimizing Rainfall Estimates with the Aid of Radar, *J. Appl. Meteor.*, 14, 1339–1345, 1975.
- Cole, S. J. and Moore, J.: Hydrological modelling using raingauge and radar based estimators of areal rainfall, *J. Hydrol.*, 358, 159–181, 2008.
- Collins, F. B. and Bolstad, P. V.: A comparison of spatial interpolation techniques in temperature estimation, Proceedings, Third International Conference/Workshop on Integrating GIS and Environmental Modeling, Santa Fe, NM, Santa Barbara, CA: National Center for Geographic Information and Analysis, Santa Barbara, 1996.
- De Michele, C. and Salvadori, J.: A generalized pareto intensity-duration model of storm rainfall exploiting 2-copulas, *J. Geophys. Res.-Atmos.*, 108, 111, doi:10.1029/2002JD002534, 2003.
- Dupuis, D. J.: Using copulas in hydrology: Benefits, cautions, and issues, *J. Hydrol. Eng.*, 12, 381–393, 2007.
- Ehret, U.: Rainfall and Flood Nowcasting in Small Catchments using Weather Radar, PhD dissertation, Stuttgart University, ISBN 3-933761-24-7, 2003.
- Embrechts, P., McNeil, A. J., and Straumann, D.: Correlation and Dependency in Risk Management: Properties and Pitfalls, Cambridge University Press, New York, 2001.
- Erxleben, J., Elder, K., and Davis, R.: Comparison of spatial interpolation methods for estimating snow distribution in the Colorado Rocky Mountains, *Hydrol. Process.*, 16, 3627–3649, 2002.
- Favre, A.-C., El Adlouni, S., Perreault, L., Thiérmonge, N., and Bobee, B.: Multivariate hydrological frequency analysis using copulas, *Water Resour. Res.*, 40, 112, doi:10.1029/2003WR002456, 2004.
- Frees, E. W. and Valdez, E. A.: Understanding relationships using copulas, *North American Actuarial Journal*, 2, 1–25, 1998.
- Genest, C. and Favre, A.-C.: Everything you always wanted to know about copula modeling but were afraid to ask, *J. Hydrol. Eng.*, 12, 347–368, 2007.
- Genest, C. and Rémillard, B.: Validity of the parametric bootstrap for goodness-of-fit testing in semiparametric models, *Annales de l'Institut Henri Poincaré: Probabilités et Statistiques*, 44, 1096–1127, 2008.
- Genest, C., Rémillard, B., and Beaudoin, D.: Goodness-of-fit tests for copulas: A review and a power study, *Insur. Math. Econ.*, 44, 199–214, 2009.
- Gomez-Hernandez, J. and Wen, X.: To be or not to be multi-Gaussian? A reflection on stochastic hydrogeology, *Adv. Water Resour.*, 21, 47–61, 1998.
- Goovaerts, P.: Geostatistical approaches for incorporating elevation into the spatial interpolation of rainfall, *J. Hydrol.*, 228, 113–129, 2000.
- Grégoire, V., Genest, C., and Gendron, M.: Using copulas to model price dependence in energy markets, *Energy Risk*, 5, 58–64, 2008.
- Haberlandt, U.: Geostatistical interpolation of hourly precipitation from rain gauges and radar for a large-scale extreme rainfall event, *J. Hydrol.*, 332, 144–157, 2007.
- Hutchinson, M. F.: Interpolation of rainfall data with thin plate smoothing splines: I Two dimensional smoothing of data with short range correlation, *Journal of Geographic Information and Decision Analysis*, 2, 139–151, 1998a.
- Hutchinson, M. F.: Interpolation of rainfall data with thin plate smoothing splines: II Analysis of topographic dependence, *Journal of Geographic Information and Decision Analysis*, 2, 152–167, 1998b.
- Isaaks, E. H. and Srivastava, R. M.: Applied Geostatistics, Oxford University Press, New York, 1989.

- Joe, H.: *Multivariate Models and Dependence Concepts*, Chapman and Hall, New York, 1997.
- Joss, J. and Lee, R.: The application of radar-gauge comparisons to operational precipitation prole corrections, *J. Appl. Meteorol.*, 34, 2612–2630, 1995.
- Kitanidis, P. K.: *Introduction to Geostatistics*, Cambridge University Press, Cambridge, 1997.
- Krause, P., Boyle, D. P., and Bäse, F.: Comparison of different efficiency criteria for hydrological model assessment, *Adv. Geosci.*, 5, 89–97, 2005, <http://www.adv-geosci.net/5/89/2005/>.
- Krajewski, W. F.: Cokriging radar-rainfall and rain-gauge data, *J. Geophys. Res.-Atmos.*, 92, 9571–9580, 1987.
- Krajewski, W. F. and Smith, J.: Radar hydrology: Rainfall estimation, *J. Hydrol.*, 25, 1387–1394, 2002.
- Kuhn, G., Khan, S., and Ganguly, A.: Geospatial-temporal dependence among weekly precipitation extremes with applications to observations and climate model simulations in south america, *J. Adv. Water Resour.*, 30, 2401–2423, 2007.
- Kunstmann, H. and Karstens, M.: Direct propagation of probability density functions in hydrological equations, *J. Hydrol.*, 325, 82–95, 2006.
- Laux, P., Wagner, S., Wagner, A., Bárdossy, A., Jacobeit, J., and Kunstmann, H.: Modelling Daily Precipitation Features in the Volta Basin of West Africa, *Int. J. Climatol.*, 29, 937–954, doi:10.1002/joc.1852, 2009.
- Laux, P., Vogl, S., Qiu, W., Knoche, H. R., and Kunstmann, H.: Copula-based statistical refinement of precipitation in RCM simulations over complex terrain, *Hydrol. Earth Syst. Sci.*, 15, 2401–2419, doi:10.5194/hess-15-2401-2011, 2011.
- Mandapaka, P. V., Krajewski, W. J., Ciach, G. J., Villarini, G., and Smith, J. A.: Estimation of radar-rainfall error spatial correlation, *Adv. Water Resour.*, 32, 1020–1030, 2009.
- de Marsily, G.: *Quantitative Hydrogeology: Groundwater Hydrology for Engineers*, Academic Press, Inc., New York, 1986.
- Marx, A.: *Einsatz gekoppelter Modelle und Wetterradar zur Abschätzung von Niederschlagsintensitäten und zur Abflussvorhersage*, PhD dissertation, Stuttgart University, ISBN 3-933761-64-6, 2007.
- Mazzetti, C. and Todini, E.: Combining Rain-gauges and Radar Precipitation Measurements using a Bayesian Approach, *Quant. Geo. G.*, 13, 401–412, doi:10.1007/1-4020-2115-1_34, 2004.
- Messer, H., Zinevich, A., and Alpert, P.: Environmental monitoring by wireless communication networks, *Science*, 312, 713, doi:10.1126/science.1120034, 2006.
- Nelsen, R. B.: *An Introduction to Copulas*. Springer-Verlag, New York, 1999.
- Renard, B. and Lang, M.: Use of a gaussian copula for multivariate extreme value analysis: Some case studies in hydrology, *Adv. Water Resour.*, 30, 897–912, 2007.
- Riedl, J.: Radar-Flächenniederschlagsmessung, *Promet*, 2/3, 20–23, 1986.
- Salvadori, G., De Michele, C., Kottegoda, N. T., and Rosso, R.: *Extremes in Nature: an approach using copulas*, Springer-Verlag, New York, 2007.
- Schölzel, C. and Friederichs, P.: Multivariate non-normally distributed random variables in climate research introduction to the copula approach, *Nonlinear Proc. Geoph.*, 15, 761–772, 2008.
- Seltnmann, J.: *Radarforschung im DWD: Vom Scan zum Produkt*, *Promet*, 26, 32–42, 1997.
- 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.*, 22, 671–688, 2008.
- Singh, V. P.: Effects of spatial and temporal variability in rainfall and watershed characteristics on stream flow hydrograph, *Hydrol. Process.*, 11, 1649–1669, 1997.
- Sklar, K.: *Fonctions de repartition à n dimensions et leurs marges*, Publications de l'Institut de Statistique de L'Université de Paris, 8, 229–231, 1959.
- van den Berg, M. J., Vandenberghe, S., De Baets, B., and Verhoest, N. E. C.: Copula-based downscaling of spatial rainfall: a proof of concept, *Hydrol. Earth Syst. Sci.*, 15, 1445–1457, doi:10.5194/hess-15-1445-2011, 2011.
- Villarini, G., Serinaldi, F., and Krajewski, W. F.: Modeling radar-rainfall estimation uncertainties using parametric and non-parametric approaches, *Adv. Water Resour.*, 31, 1674–1686, 2008.
- Wang, W., Van Gelder, P. H. A. J. M., Vrijling, J. K., and Ma, J.: Testing and modelling autoregressive conditional heteroscedasticity of streamflow processes, *Nonlinear Proc. Geophy.*, 12, 55–66, 2005.
- Zhang, L. S. R. and Singh, V.: Bivariate rainfall frequency distributions using archimedean copulas, *J. Hydrol.*, 32, 93–109, 2008.