Copula-based modeling of dependence structure in geodesy and GNSS applications: case study for zenith tropospheric delay in complex terrain

Roya Mousavian, Christof Lorenz, Masoud Mashhadi Hossainali, Benjamin Fersch & Harald Kunstmann

GPS Solutions
The Journal of Global Navigation Satellite Systems

ISSN 1080-5370
Volume 25
Number 1

GPS Solut (2021) 25:1-17
DOI 10.1007/s10291-020-01044-4

Springer

GPS Solutions (2021) 25:12
https://doi.org/10.1007/s10291-020-01044-4

ORIGINAL ARTICLE

Copula-based modeling of dependence structure in geodesy and GNSS applications: case study for zenith tropospheric delay in complex terrain

Roya Mousavian[1,2] · Christof Lorenz[2] · Masoud Mashhadi Hossainali[1] · Benjamin Fersch[2] · Harald Kunstmann[2,3]

Received: 9 October 2019 / Accepted: 9 October 2020

Abstract

Modeling and understanding the statistical relationships between geophysical quantities is a crucial prerequisite for many geodetic applications. While these relationships can depend on multiple variables and their interactions, commonly used scalar methods like the (cross) correlation are only able to describe linear dependencies. However, particularly in regions with complex terrain, the statistical relationships between variables can be highly nonlinear and spatially heterogeneous. Therefore, we introduce Copula-based approaches for modeling and analyzing the full dependence structure. We give an introduction to Copula theory, including five of the most widely used models, namely the Frank, Clayton, Ali-Mikhail-Haq, Gumbel and Gaussian Copula, and use this approach for analyzing zenith tropospheric delays (ZTDs). We apply modeled ZTDs from the Weather and Research Forecasting (WRF) model and estimated ZTDs through the processing of Global Navigation Satellite System (GNSS) data and evaluate the pixel-wise dependence structures of ZTDs over a study area with complex terrain in Central Europe. The results show asymmetry and nonlinearity in the statistical relationships, which justifies the application of Copula-based approaches compared to, e.g., scalar measures. We apply a Copula-based correction for generating GNSS-like ZTDs from purely WRF-derived estimates. Particularly the corrected time series in the alpine regions show improved Nash–Sutcliffe efficiency values when compared against GNSS-based ZTDs. The proposed approach is therefore highly suitable for analyzing statistical relationships and correcting model-based quantities, especially in complex terrain, and when the statistical relationships of the analyzed variables are unknown.

Keywords Zenith tropospheric delay · Copulas · GNSS · Dissimilarity measures · Atmospheric modeling · Correlation

Introduction

Knowing and modeling the dependence structure of different variables, including the statistical distributions of and relationships between geophysical quantities, is mandatory for a wide range of applications in geosciences. For example, the dependence between different observations (Tiberius and Borre 2000) and satellite-based errors (Heng et al. 2011) is required for the processing of GPS data. The normal or Gaussian distribution is the most known and widely used distribution, and many applications in geosciences require variables to be normally distributed. Furthermore, a multivariate normal distribution is defined only by the mean and covariance matrix (including the variance of the variables and Pearson's correlation between them) and, therefore, only allows for the description of linear dependencies. However, in many cases, the assumptions of normally distributed data and variables, as well as linear dependencies, are not valid. For instance, the distribution of daily precipitation is non-normal (Ye et al. 2018), geodynamic and geologic systems for earthquake prediction are extremely nonlinear (Koronovskii and Naimark 2012), and the tropospheric and ionospheric effects on GNSS signals, which are crucial quantities for precise point positioning (Mendez Astudillo

✉ Roya Mousavian
r_mousavian@yahoo.com; roya_mousavian@mail.kntu.ac.ir

[1] Faculty of Geodesy and Geomatics Engineering, K. N. Toosi University of Technology, No. 1346, ValiAsr Street, Mirdamad Cross, Tehran, Iran

[2] Institute of Meteorology and Climate Research (IMK-IFU), Campus Alpin, Karlsruhe Institute of Technology, Kreuzeckbahnstr. 19, 82467 Garmisch-Partenkirchen, Germany

[3] Institute of Geography, University of Augsburg, Alter Postweg 118, 86159 Augsburg, Germany

Published online: 02 November 2020

et al. 2018) and also the fundamental basis of GNSS tomography, are nonlinear and highly spatiotemporally heterogeneous (Norberg et al. 2015; Sun et al. 2017; Turel and Arikan 2010; Wilgan et al. 2017).

While ionospheric effects have strong periodic characteristics that can be approximated with models, the description of tropospheric effects is far more complicated due to their strong relationship with weather and climate processes. Zenith tropospheric delays (ZTDs) are used for improving weather forecasts by constructing and assimilating water vapor map (Chen et al. 2018) in numerical weather prediction (NWP) models (Gendt et al. 2004; Giannaros et al. 2020; Rohm et al. 2019; Singh et al. 2019). On the other hand, information from NWP models is widely used in geodesy and geodetic applications, e.g., GPS processing (Nordman et al. 2007), GPS kinematic positioning via ray tracing in a three-dimensional grid (Nievinski et al. 2005), and improving the estimated GNSS ZTDs (Dousa et al. 2018; Zus et al. 2019). Due to the importance of estimating ZTDs, different methods and processing strategies have been investigated, e. g., using a correction model to enhance evaluating tropospheric delays (Jgouta et al. 2016; Dousa et al. 2018), and assessing of tropospheric effects by least-squares collocation of meteorological and GNSS data (Wilgan et al. 2017). However, particularly the model estimates over complex terrain, like the Alpine region, might suffer from severe model biases, which have to be considered for GNSS applications. Thus, such correction requires robust knowledge about the statistical characteristics and relationships of the estimated tropospheric delays, which can be highly complex and nonlinear.

For describing the dependency structures of such complex quantities, we need a flexible technique that (a) allows for any univariate marginal distribution, (b) can be applied with a reasonable amount of data and (c) provides information about the full dependence structure including both linear and nonlinear relationships.

In this study, we propose a Copula-based analysis, which is not only limited to single dependence measures like the Pearson's correlation coefficient, but also allows for the rank-dependent evaluation of statistical relationships (e.g., between high or low values). Due to their flexibility, Copulas have been widely used in different disciplines such as economics and finance (Ang and Chen 2002; Longin and Solnik 2002), hydrology (Bardossy and Li 2008), meteorology (Vogl et al. 2012) and remote sensing (Brunel and Pieczynski 2005; Lorenz et al. 2018). However, using Copulas in geodesy is generally new and currently, according to our knowledge, is limited to a single study about polar motion prediction (Modiri et al. 2018).

We apply Copulas for modeling and analyzing the dependence structure of ZTDs from an atmospheric model. As such an analysis is not feasible for large-scale applications, which involve evaluations across a large variety of locations, we further employ different Copula-based scalar dissimilarity measures. These allow for describing the spatiotemporal relationship, including linear and nonlinear dependencies, particularly over complex terrain and therefore provide a powerful and sophisticated alternative to, e.g., the widely used correlation coefficients.

As the estimated ZTDs from postprocessing of the GNSS data are assumed to be the most precise estimates with a precision of 2–4 mm (Byram et al. 2011), we further analyze the potential of a Copula-based conditional random sampling approach to derive GNSS-like ZTDs from modeled ZTDs. This method finally allows for the correction of WRF-based ZTDs as well as the filling of gaps in GNSS ZTD time series or their extension.

The main objectives are hence (a) to provide a simple workflow for computing and evaluating Copula-based dependence structures, (b) to implement Copula-based dissimilarity measures and (c) to give a real-life example of the previously derived measures and analyses using ZTD and modeling Copula-based ZTDs in GPS stations by improving the WRF-based ZTDs.

Methodology

Understanding the full dependence structure of variables and parameters is mandatory for precise positioning, tropospheric modeling and other application of GNSS data in geosciences. However, e.g., Pearson's correlation as a well-known measure for modeling dependency does not show all statistical relationships (i.e., linear and nonlinear) between variables. Therefore, we apply a Copula-based approach to fully describe the dependence structure and have a great flexibility to model nonlinear dependencies between variables.

Copula theory

Copulas are mathematical functions that link multivariate distributions to their one-dimensional marginal distributions: (Nelsen 2006; Sklar 1959):

$$F(x_1, x_2, \dots, x_n) = C(F_1(x_1), F_2(x_2), \dots, F_n(x_n)) \tag{1}$$

where $F, C, x_1, x_2, \dots, x_n \in \mathbb{R}$ and $F_1, F_2, \dots, F_n$ are the multivariate distribution, Copula, random variables and their marginals, respectively. Each n-dimensional Copula is hence a function which transforms data from the unit n-cube $[0, 1]^n$ to the unit interval $[0, 1]$ and satisfies the characteristics of cumulative distribution functions (CDFs). The multivariate Copula density is then given through:

$$c(F_1(x_1), F_2(x_2), \dots, F_n(x_n)) = \frac{\partial^n C(F_1(x_1), F_2(x_2), \dots, F_n(x_n))}{\partial F_1(x_1) \dots \partial F_n(x_n)} \tag{2}$$

Different Copulas can be used to model different dependency structures between random variables. In general, we can distinguish between implicit and explicit Copulas. The latter, which also contains the widely used family of Archimedean Copulas, can be formulated in a closed mathematical form. For simplicity, we focus on bivariate applications in this study.

Implicit Copulas

Implicit Copulas are implied by well-known multivariate distributions and do not have a closed mathematical form. Their Copula density c is given by

$$c(F_1(x_1), F_2(x_2)) = \frac{f(x_1, x_2)}{f_1(x_1) \cdot f_2(x_2)} \tag{3}$$

where f, $u_i = F_i, i = 1, 2$ and $f_i, i = 1, 2$ are the density of their joint distribution function, marginal CDFs and PDFs of the random variables x_1 and x_2, respectively. From this category, the Gaussian Copula is the most familiar Copula as it is derived from the standardized multivariate Gaussian distribution $\Phi_R(0, \mathbf{R})$ with zero mean and correlation matrix $\mathbf{R}$ (Mikosch 2006):

$$C_{\mathbf{R}}^{\mathrm{Ga}}(\mathbf{u}) = \Phi_{\mathbf{R}}(\Phi^{-1}(u_1), \Phi^{-1}(u_2)) \tag{4}$$

Here, $\Phi^{-1}(.)$ is the inverse of the standard univariate normal distribution function. The density of the bivariate Gaussian Copula is given by (Arbenz 2013)

$$c(\mathbf{u}) = \frac{1}{\sqrt{\det(\mathbf{R})}} \exp(-1/2(\mathbf{X}^{\mathrm{T}}(\mathbf{R}^{-1} - \mathbf{I})\mathbf{X})) \tag{5}$$

where $\mathbf{X} = (\Phi^{-1}(u_1), \Phi^{-1}(u_2))^{\mathrm{T}}$ and $\mathbf{I}$ is the identity matrix.

Explicit Copulas

Explicit Copulas can be expressed in closed mathematical forms. The most important type of these Copulas is the family of Archimedean Copulas, which are expressed as (Embrechts et al. 2003; Nelsen 2006):

$$C_\phi(u_1, u_2) = \phi^{-1}\{\phi(u_1) + \phi(u_2)\}, \quad \mathbf{u} \in [0, 1] \tag{6}$$

Here, $u_i, i = 1, 2$ are the marginals of the random variables $x_i, i = 1, 2$ and $\phi : [0, 1] \to [0, \infty)$ is the generator function of the Copula C_ϕ. This generator contains all the information on the dependence structure between the random variables, and we can, therefore, reduce the evaluation of a multivariate Copula to a single function. Moreover, there is a relationship between the generator function as a function of the Copula parameter and the rank correlation coefficient Kendall's τ as:

$$\tau = 1 + 4 \int_0^1 \frac{\phi(t)}{\phi'(t)} \mathrm{d}t \tag{7}$$

where $\phi(t)$ and $\phi'(t)$ are the generator function and its first derivative, respectively (Table 1). Hence, we can also interpret the Copula parameter as a measure for the strength of dependence. Table 1 gives the closed forms for four of the most widely used Archimedean Copulas, their generator functions, the range of their parameters (θ) and their Kendall's τ.

Figure 1 shows bivariate Copula densities from the Gaussian Copula and the Archimedean families in Table 1 with three different Copula parameters (θ). Since the range of the Copula parameters is not the same in all families (see Table 1), we used different parameters and colorbars for each Copula in order to better identify the different dependency structures. In these figures, horizontal and vertical axes are uniform variables on [0, 1], which are indicated by u_1 and u_2, respectively. It should be noted that the colored surface in each plot shows the strength of the Copula density ($c(u_1, u_2)$). As the Copula density from different families cannot be directly compared due to different ranges of values, we add separate colorbars for each of the depicted Copulas. Moreover, it is obvious that u_1, u_2 and $c(u_1, u_2)$ are unitless. One key characteristic is that the Gumbel, Clayton and Ali-Mikhail-Haq (AMH)-Copulas are asymmetric. This means that these families are suitable for modeling either upper or lower tail dependence. In particular, the Gumbel Copula shows significant upper tail dependence (i.e., more correlated in large values), while both the Clayton and the AMH show lower tail dependence (i.e., more correlated in small values). Both the Gaussian and Frank Copula are symmetric.

Table 1 Overview of four of the most common Copula families, including their generator functions, the range of their Copula parameter θ and their rank correlation Kendall's τ

Family	ϕ (t)	θ	$C_\phi(u_1, u_2)$	τ
Clayton	$\frac{t^{-\theta}-1}{\theta}$	$(0, \infty)$	$(u_1^{-\theta} + u_2^{-\theta} - 1)^{-\frac{1}{\theta}}$	$\frac{\theta}{\theta+2}$
Gumbel	$(-\log(t))^\theta$	$[1, \infty)$	$\exp\left\{-[(-\log(u_1))^\theta + (-\log(u_2))^\theta]^{\frac{1}{\theta}}\right\}$	$1 - \frac{1}{\theta}$
Frank	$\log\left(\frac{e^{\theta t}-1}{e^\theta-1}\right)$	$\mathbb{R}\backslash\{0\}$	$-\frac{1}{\theta}\log\left\{1 + \frac{(e^{-\theta u_1}-1)(e^{-\theta u_2}-1)}{e^{-\theta}-1}\right\}$	$1 + \frac{4[D_1(\theta)-1]}{\theta}$ $D_1(\theta) = \frac{1}{\theta}\int_0^\theta \frac{t}{\exp(t)-1} dt$
AMH	$\frac{\ln(1-\theta(1-t))}{t}$	$[-1, 1]$	$\frac{u_1 u_2}{1-\theta(1-u_1)(1-u_2)}$	$1 - \frac{2(\theta+(1-\theta)^2\log(1-\theta))}{\theta}$

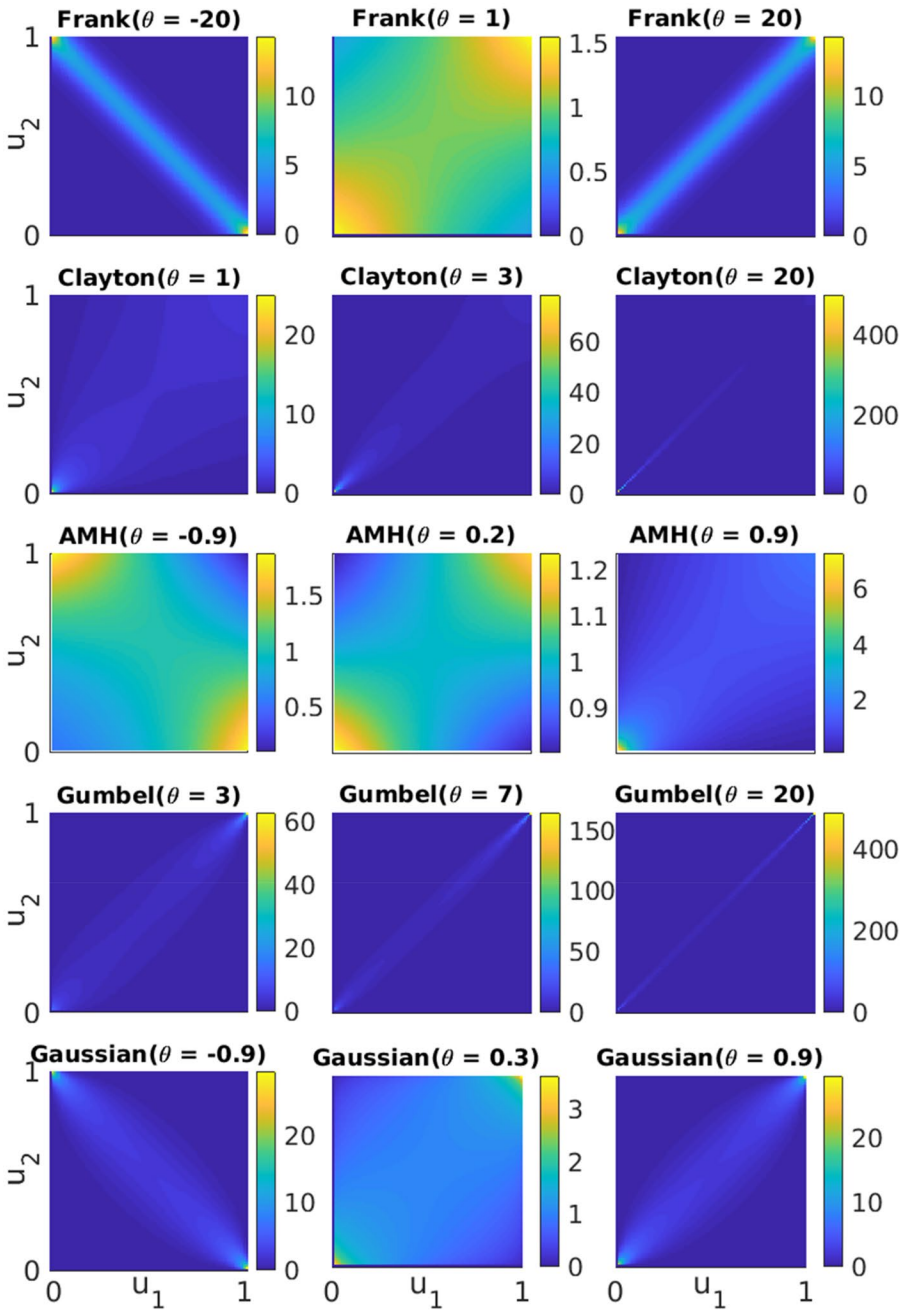

Fig. 1 Copula densities for the bivariate sample (u_1, u_2) with different parameters for Copulas

Copula-based modeling of bivariate dependence structures

The workflow for the Copula-based dependence structure modeling requires several steps, which are shown in Fig. 2 and discussed in the next subsections. In this figure, those elements which are linked together by solid arrows show the required steps for the Copula-based modeling of a data set. Moreover, the dashed arrow pointed to the Copula-based dissimilarity measures, which can be estimated by empirical Copulas and empirical Copula densities. Other elements indicate the applied statistical methods to perform the linked steps by a dashed line to them.

Estimation of univariate marginal distributions

The first step is the estimation of the marginal distributions and the transformation of the input data to the unit interval ($[0, 1]$) or rank space. For this, depending on the available a priori information, we can use parametric, semi-parametric and nonparametric methods (Charpentier et al. 2007; Choros et al. 2010; Laux et al. 2011; Mao

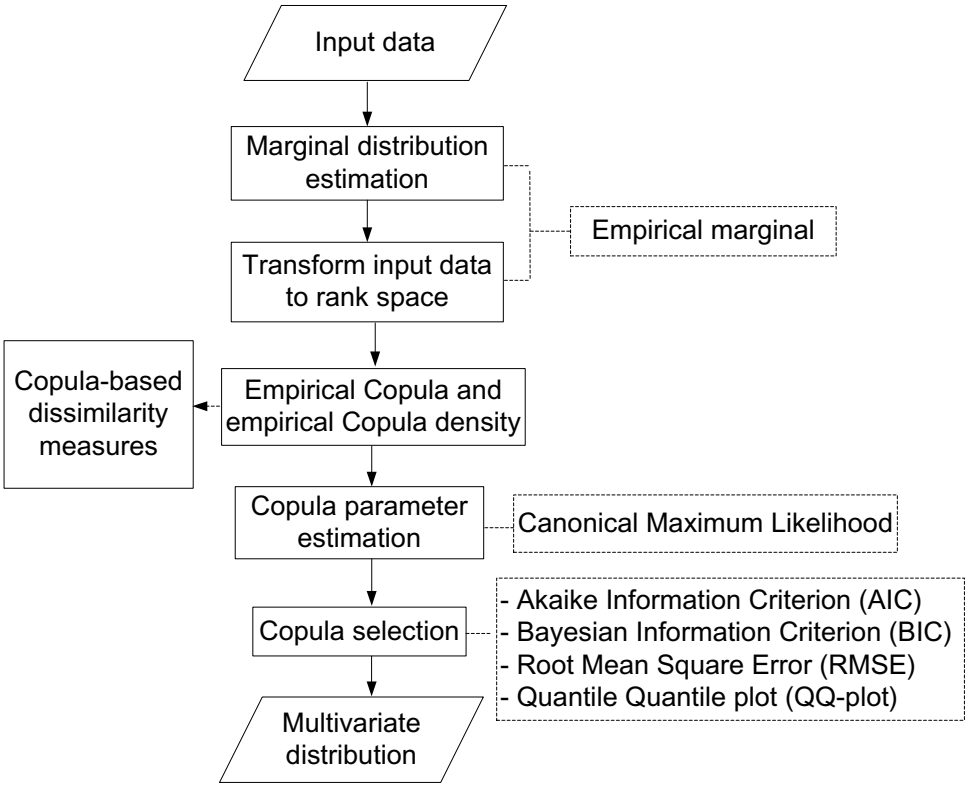

Fig. 2 Procedure of Copula-based dependence structure modeling

et al. 2015). Choosing an appropriate method depends on the sample size, the distribution of values within that sample and several other factors. If no a priori information about a random variable X and its distribution is available and we have a sufficient number of samples T, the empirical distribution is usually a good approximation of the true underlying distribution (Charpentier et al. 2007):

$$u_i = \hat{F}_i(x) = \frac{1}{T+1} \sum_{t=1}^{T} I(X_t \leq x), \quad i = 1, 2 \tag{8}$$

where $I(.)$ denotes the indicator function and is defined as

$$I_A(x) := \begin{cases} 1 & x \in A \\ 0 & x \notin A \end{cases} \tag{9}$$

Here, A is any nonempty subset of set X.

The empirical distribution has some considerable limitations. If, e.g., new data become available, which is beyond the limits of the old sample, extrapolation techniques have to be applied. Furthermore, depending on the discretization and the value range of the data, the empirical distribution might contain steep steps, which complicate, e.g., the interpolation of new values. Therefore, using an empirical distribution always requires a close and critical look at the results.

Estimation of Copula parameters

The next step is the estimation of the Copula parameters. An overview of several estimation methods is presented by Kim et al. (2007). They concluded that when the marginals are unknown, using the semi-parametric pseudo-maximum likelihood (PML) (Genest et al. 1995) or canonical maximum likelihood (CML) (Cherubini et al. 2004) estimation, outperforms fully parametric methods like maximum likelihood (ML) and inference function for margins (IFM) (Joe 1997). Hence, we also focus on the CML estimation in this study.

CML consists of transforming the sample data $\{x_{1t}, x_{2t}\}_{t=1}^{T}$ to uniform variables $\{u_{1t}, u_{2t}\}_{t=1}^{T}$ using empirical marginal distributions and then computing the corresponding Copula parameter using a maximum likelihood estimate:

$$\hat{\theta} = \mathrm{ArgMax}_{\theta} \sum_{t=1}^{T} \ln c(\hat{F}_1(x_{1t}), \hat{F}_2(x_{2t}); \theta) \tag{10}$$

where $c(u_{1t}, u_{2t}; \theta)$ is the theoretical copula density with its parameters θ(see (2) and (5) and Table 1) and ArgMax_{θ} refers to the θ values at which the summation of the logarithm of the Copula density is maximized.

Estimation of the empirical Copula CDF

The empirical Copula was introduced as an empirical dependence function (Deheuvels 1979a; Nelsen 2006). It has been shown that the empirical Copula converges uniformly to the underlying Copula (Deheuvels 1979b, 1981). We can compute this empirical Copula CDF through

$$\hat{C}(u_{1j}, u_{2j}) = \frac{1}{T}\sum_{t=1}^{T} I(\hat{F}_1(x_{1t}) \le u_{1j}, \hat{F}_2(x_{2t}) \le u_{2j}), \quad j = 1, 2, \ldots, T \tag{11}$$

where $\hat{F}_i(x_i), i = 1, 2$ are the univariate marginal distributions of the random variables x_1 and x_2 with the sample size T and $I(.)$ is again the indicator function.

Estimation of the empirical Copula PDF

Similar to the empirical Copula CDF, we can estimate an empirical approximation of the Copula PDF from the data through nonparametric methods (Charpentier et al. 2007; Mao et al. 2015). Therefore, we discretize the interval from 0 and 1 to a regular $k \times k$ grid based on the sample size. Then, the empirical Copula PDF is calculated by (Bardossy 2006; Nelsen 2006)

$$\hat{c}\left(\frac{2r-1}{2k}, \frac{2s-1}{2k}\right) = \frac{k^2}{T} q_{rs} \tag{12}$$

where T and q_{rs} are sample size and the empirical frequency corresponding to the grid coordinates (r, s) with $\{r, s = 1, 2, \ldots, k\}$, respectively. For a given pair, q_{rs} is given as

$$q_{rs} = \left| \left\{ \frac{r-1}{k} < u_1 < \frac{r}{k} \text{ and } \frac{s-1}{k} < u_2 < \frac{s}{k} \right\} \right| \tag{13}$$

where u_1 and u_2 are the transferred variables to rank space and $|\cdot|$ denotes the cardinality (the number of elements of a set). To ensure that the dependence between random variables is accurately represented, it is recommended that the sample size should be large enough, i.e., $T > 5k^2$ (Samaniego et al. 2010).

Similar to the estimation of univariate empirical distributions, the empirical Copulas have limitations. In particular, the sample size plays a significant role during the estimation. For computing the densities at discrete locations, we have to define a two-dimensional grid that matches the sample size. Thus, for small sample sizes, we have to use a coarse grid with only a few values between 0 and 1 for r and s. If the discretization is too fine, the estimated values for q_{rs} are based on too few values and the resulting empirical Copula does not show a continuous surface, but rather some scattered peaks around many zero values. Therefore, similar to the empirical marginals, the estimation of empirical Copulas has to be done with care, as they provide the basis for all further steps.

Identification of the best Copula

There is a wide range of methods for identifying the most suitable Copula. Except for the Bayesian method (Huard et al. 2006), the selection approaches usually require to estimate the Copula parameters in advance. These methods include likelihood-based criteria like the Akaike information criterion (AIC) (Aho et al. 2014; Akaike 1974), Bayesian information criterion (BIC) (Schwarz 1978), minimum distance estimation (Durrleman et al. 2000; Wolfowitz 1957), the root-mean-square error (Vandenberghe et al. 2010) and the graphical identification using a quantile quantile plot (QQ plot). The latter is particularly suitable for bivariate applications (Genest and Favre 2007). According to Fang et al. (2014) and Topcu (2016), AIC and BIC outperform other goodness of fit (GOF) criteria for bivariate Archimedean Copulas. However, for the sake of completeness, we give a brief overview of all previously mentioned methods.

The AIC and BIC for bivariate Copulas are defined as

$$\mathrm{AIC} = -2\sum_{t=1}^{T} \ln[c(u_{1t}, u_{2t}); \theta] + 2k \tag{14}$$

$$\mathrm{BIC} = -2\sum_{t=1}^{T} \ln[c(u_{1t}, u_{2t}); \theta] + \ln(Tk) \tag{15}$$

where $c(u_{1t}, u_{2t})$ is the Copula density, T is the sample size, and k is the number of parameters in the desired Copula. Using the previously derived Copula parameter, we can test different theoretical Copulas and compare the resulting AIC and BIC values for identifying the most suitable Copula. Lower values of AIC and BIC indicate a better fit to the data.

The root-mean-square error (RMSE) as a measure to select an appropriate Copula is calculated as (Vandenberghe et al. 2010):

$$\mathrm{RMSE} = \sqrt{\frac{1}{T}\sum_{t=1}^{T} (C(u_{1t}, u_{2t}) - \hat{C}(u_{1t}, u_{2t}))^2} \tag{16}$$

where T is sample size, and $C(u_{1t}, u_{2t})$ and $\hat{C}(u_{1t}, u_{2t})$ are values from the theoretical and empirical Copula, respectively. Similar to the AIC- and BIC-based evaluation, we compute the RMSE for different Copulas and select the family with the lowest RMSE.

The quantile quantile (QQ) plot is a graphical method for comparing two probability distributions by plotting their quantiles against each other (Wilk and Gnanadesikan 1968). If the two distributions are similar, the points in the QQ plot

will approximately lie on the $y=x$ line. To apply this method for identifying a suitable Copula, we draw empirical against theoretical Copula quantiles (Genest and Favre 2007) and compare the distance from the $y=x$ line for each family. The family with the closest distance is selected as the most suitable Copula. Besides the visual inspection, we can also use statistical tests like the Cramer–von Mises or Kolmogorov–Smirnov test (Darling 1957) for calculating the distance between the empirical and theoretical quantiles.

Conditional sampling

After identifying the most suitable Copula, we can generate conditional random samples using the conditional Copula (Trivedi and Zimmer 2005):

$$C\left(u_2|u_1=v\right)=\frac{\partial C(u_1,u_2)}{\partial u_1} \tag{17}$$

The conditional Copula is simply the uniform CDF for the variable u_2 when $u_1=v$. The conditional random samples can then transform back to the data space using the inverse marginal CDF:

$$u_i=F_i\left(x_i\right)\Leftrightarrow x_i=F_i^{-1}\left(u_i\right),\quad i=1,2 \tag{18}$$

where F_i, x_i and u_i, $i=1,2$ are marginal CDFs, variables in data space and variables in rank space, respectively. Moreover, F_i^{-1} is the generalized inverse of F_i which is defined by $F^{-1}(t):=\text{infimum}\{y|F(y)\geq t\}$ for $t\in\text{Ran}\,F_i$ and $y\in\mathbb{R}$.

Copula-based dissimilarity measures

Dissimilarity measures are numerical measures of how different two data sets or variables are, and they are used to quantify independency (or, vice versa, dependency) between variables. Various measures have been formulated throughout the years, each with its own strengths and weaknesses (Goshtasby 2012). In this study, we apply three Copula-based dissimilarity measures as λ^1, λ^2 and λ^3 which are based on the empirical Copula (Samaniego et al. 2010) for analyzing the spatiotemporal relationship between geophysical quantities.

The first and simplest dissimilarity measure λ_{ij}^1 for each two variables u_i and u_j in a set with n pairs of variables is defined as the probability of upper and lower tail dependence:

$$\lambda_{ij}^1=(P-L_{ij})+\frac{|U_{ij}-L_{ij}|}{U_{ij}+L_{ij}} \tag{19}$$

where P is a given probability threshold, which separates the tails from the rest of the distribution. Here, we set P to 0.2. U_{ij} and L_{ij} are the upper and the lower tail probability of the empirical Copula density, respectively:

$$L_{ij}=\int_0^P\int_0^P\hat{c}(u_i,u_j)\mathrm{d}u_i\mathrm{d}u_j \tag{20}$$

$$U_{ij}=\int_{1-P}^1\int_{1-P}^1\hat{c}(u_i,u_j)\mathrm{d}u_i\mathrm{d}u_j \tag{21}$$

where $\hat{c}$ denotes the empirical Copula densities. As these densities are given on a regular grid, the integrals can be transformed into sums.

This dissimilarity measure consists of two terms: The first term estimates how far the empirical Copula density is from a perfect lower corner dependence, and the second one describes the level of asymmetry of the upper and lower tails about the axis $u_j=1-u_i$. If the empirical Copula density between two variables is symmetric, λ^1 would be obviously equal to zero.

The second measure does not require any pre-defined parameters and is given as

$$\lambda_{ij}^2=(1-r_{ij})+\xi\left|A_{ij}\right| \tag{22}$$

where ξ is a scaling factor which is selected so that $\sup(1-r_{ij})\approx\sup\left|A_{ij}\right|$. r_{ij} and A_{ij} are Spearman's rank correlation, which is computed from the empirical Copula (see (23)), and the degree of asymmetry of the empirical Copula density, respectively. They are calculated according to

$$r_{ij}=12\int_0^1\int_0^1(\hat{C}(u_i,u_j)-u_iu_j)\mathrm{d}u_i\mathrm{d}u_j \tag{23}$$

$$A_{ij}=\int_0^1\int_0^1\left[\left(u_i-\frac{1}{2}\right)^2\left(u_j-\frac{1}{2}\right)+\left(u_i-\frac{1}{2}\right)\left(u_j-\frac{1}{2}\right)^2\right]\hat{c}(u_i,u_j)\mathrm{d}u_i\mathrm{d}u_j \tag{24}$$

Therefore, the first term of this measure indicates the statistical dependence between variables and the second term describes the dissimilarity due to the asymmetry of their empirical Copulas.

The third dissimilarity measure consists of the asymmetry of the Copula density (i.e., A_{ij}, see (24)) and the Kolmogorov–Smirnov statistic (M_{ij}) and is estimated through

$$\lambda_{ij}^{3} = M_{ij} + \xi\left|A_{ij}\right| \tag{25}$$

where

$$M_{ij} = \sup\left|H\left(\Delta x_i(t)\right) - H(\Delta x_j(t))\right| \tag{26}$$

Here, t is time in the time series and H is the distribution function of the lag 1 differences of each variable's time series which are computed as

$$\Delta x(t) = x(t) - x(t-1) \tag{27}$$

In contrast to λ^1 which is sensitive to the selection of P, λ^2 and λ^3 are free of any pre-defined parameters and more general measures. Furthermore, Samaniego et al. 2010 conclude that λ^1 has only moderate sensitivity with respect to extreme values and the asymmetry of the Copula density. Hence, λ^2 and λ^3 should be preferred for variables that are strongly correlated, particularly in the tails of their distributions.

Data

Transmitted radio signals from satellites at frequencies above 30 MHz are affected by fluctuations of pressure, temperature and humidity in the troposphere (Hitney et al. 1985). These fluctuations cause a delay in receiving signals at ground stations, which need to be considered for all GNSS-based applications. If this delay is evaluated in zenith direction, it is referred to as ZTD. Tropospheric delay can either be calculated during the processing of GNSS data or approximated through atmospheric models. In this study, we use the estimated ZTDs from both GNSS observations and the atmospheric Weather Research and Forecasting (WRF) (Skamarock et al. 2008) model in high spatiotemporal resolution (Fersch et al. 2020).

The GNSS-based ZTDs are provided by the German Research Center for Geosciences (GFZ). They are estimated every 15 min during the processing of the data at each GPS receiver of a regional network with approximately 300 stations across Germany (Gendt et al. 2004). Most stations belong to the network Satellite Positioning Service of the German Land Surveying Agencies (SAPOS), and some of them are from the German Federal Agency for Cartography and Geodesy (BKG). In this study, we use observations from 22 GPS stations that are well distributed across the study domain (Fig. 3).

We further use simulated atmospheric pressure, humidity and temperature in high spatiotemporal resolution from the WRF model. The formulas for computing ZTDs from these meteorological variables are given in "Appendix." The WRF model was run for Central Europe during the period of April to October 2016. We focus on a sub-domain that is dominated by complex terrain including the Alps in the South and the Bavarian Forest in the North-East (see Fig. 3). The model was driven with 6-hourly ERA Interim (Dee et al. 2011) reanalysis data, which has a spatial resolution of 0.75° and 37 pressure levels from 1000 to 1 hPa. The spatial resolution of WRF was set to 3×3 km, resulting in 315 grid cells in West–East and 280 grid cells in South–North direction. We further used 50 vertical layers from 1000 to 10 hPa to get a high-resolution representation of the vertical structure of the atmosphere.

Experimental setup

To demonstrate the general workflow of the Copula-based analysis for WRF-derived ZTDs, we select representative pixels across our study domain (locations A, B and C in Fig. 3) for which we derive the statistical relationships with all other pixels. We first estimate the univariate marginal distributions from hourly ZTDs for all WRF pixels ($N=109\times74=8066$ pixels) for the transformation from the data to the rank space. The transformed data are then used for estimating the empirical Copula CDF and PDF between the three target pixels and all other pixels across the study domain. Finally, we compute the corresponding Copula parameter for all five Copulas and identify the most suitable model by comparing their AIC and BIC measures. For evaluating the spatiotemporal dependency of the ZTDs, we compute the three dissimilarity measures between the target pixels and all other pixels.

Besides this purely WRF-based dissimilarity and dependency analysis, we also investigate the dependency between WRF- and GNSS-derived ZTDs. We use the estimated ZTDs of 22 GPS stations (Fig. 3) and WRF-based ZTD time series from the closest pixels to each of the stations. To reveal the statistical relationship between WRF- and GNSS-based ZTDs independent of climatic cycles (which would artificially increase the dependency between the variables), we first remove the monthly and daily cycle from the signals. Using these anomaly pairs, we again transform the WRF- and GNSS-based time series to the rank space and identify the most suitable Copula. This calibration procedure is carried out using the first half of the study period from mid-April to mid-June 2016. Then, we condition the Copula on the WRF-derived ZTDs, which finally allows us to generate GNSS-like conditional

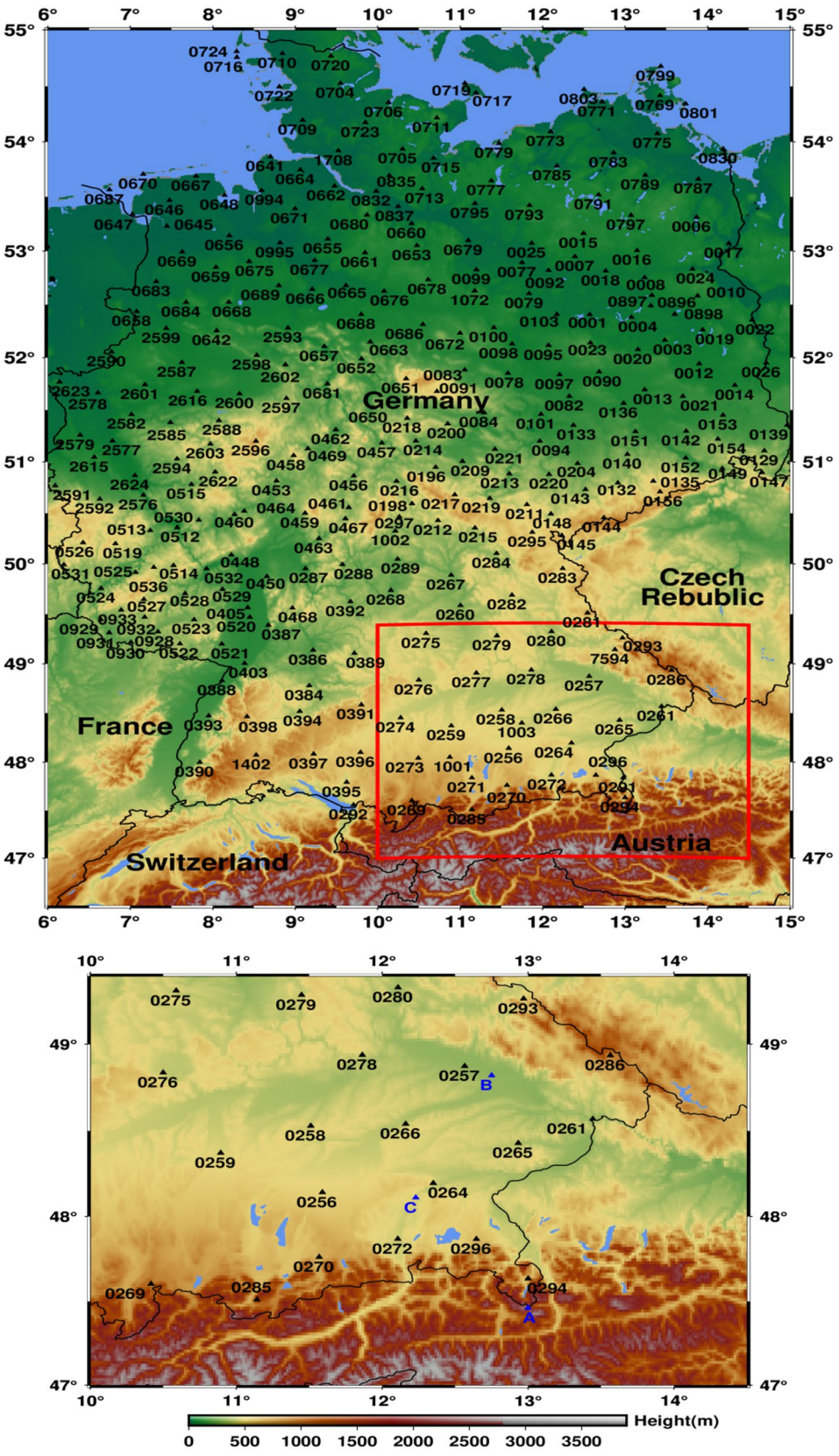

Fig. 3 Location of the study domain together with GPS stations of SAPOS (top) and studied GPS stations (bottom). Topography is shown based on GTOPO 30 digital elevation model

random samples from the WRF estimates (see (17) and (18)). Using the second half of the study period, we generate 500 conditional random samples for each time step from the WRF-based ZTDs and use the inverse marginal distribution from the GNSS-based ZTDs for transforming the conditional random samples back to the data space. From this set, we then compute the median, which is the final Copula-based GNSS-like ZTD. To analyze the performance between this Copula and purely WRF-based ZTDs, we compare the two time series at all 22 stations using the Nash–Sutcliffe efficiency (NSE)

$$\mathrm{NSE} = 1 - \frac{\sum_{t=1}^{T} (X_t - \hat{X}_t)^2}{\sum_{t=1}^{T} (X_t - \overline{X}_t)^2} \tag{28}$$

where $\hat{X}_t$ are the estimated ZTDs from WRF before and after the Copula-based correction, X_t are the calculated ZTD from the GNSS data, T is the length of the time series, and $\overline{X}_t$ denotes the mean value of X_t, which is in our case set to zero as we are using ZTD anomalies. The higher NSE demonstrates better modeling with respect to GNSS-based ZTD.

Results and discussion

As the main goal of this study is the investigation of Copula-based approaches to improve the WRF-derived ZTDs, we firstly use the Copula workflow to evaluate the dependence structure between these time series across the region. Then,

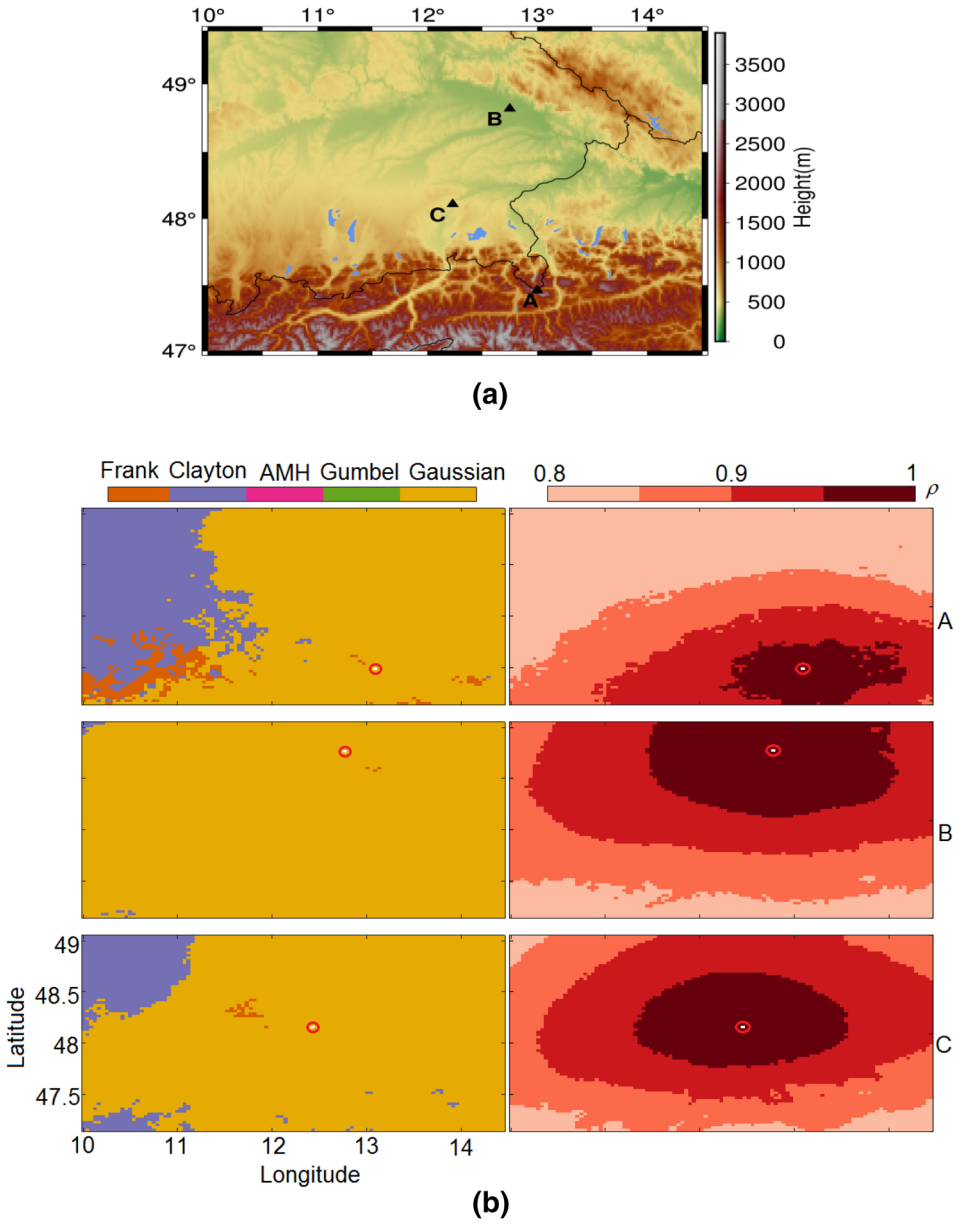

Fig. 4 **a** Topography and location of the target pixels. **b** Copula-based dependence structure (left column) and Pearson's correlation (right column) between WRF-derived ZTDs for pixel pairs with target pixels A (top), B (middle) and C (bottom), respectively. The location of the target pixels is highlighted by red circles

we apply a Copula-based random sampling technique, which allows estimating conditional GPS-like ZTDs from the WRF model.

Modeling Copula-based dependence between WRF-derived ZTDs

The most likely Copula for each of the three target pixels, and all other pixels in the study domain, is shown in Fig. 4. For the target pixel B (which is located in flat terrain), the Gaussian Copula is selected as the most suitable Copula across 99.7% of the domain. The target pixels A and C, which are both located in more complex terrain, show different behavior. While the Gaussian Copula dominates the study region, the asymmetric Clayton Copula is identified for at least 25.8% (A) and 11.2% (C) of the pixel pairs, especially in the western part of the domain. One reason why the Gaussian Copula is especially suitable for describing the statistical relationships is due to the strong linear correlations.

The right column of Fig. 4b shows the correlation coefficients between the three target pixels and the other pixels. In these plots, even at the boundaries of the domain, which can be up to 300 km away from the target pixel, there are still correlations of 0.8 and higher. In general, the correlation coefficients indicate circular patterns of decreasing values with increasing distance from the target pixels and do not show a significant pattern across different parts of study domain. However, due to the impact of altitude and complex terrain on humidity, temperature and pressure (which are used to estimate the WRF-based ZTDs), it should be expected that this topography also has some impact on the statistical relationships between the ZTD time series.

Therefore, we now use the Copula-based dissimilarity measures to check whether they allow for a more sensible evaluation of the statistical relationships by including nonlinear dependencies. Figure 5b shows λ^1,λ^2 and λ^3 for the target pixels A, B and C. Similar to Fig. 4b, the values of the dissimilarity measures increase with distance from the target pixels. But this decrease of dependency is not isotropic as there is a strong impact of topography and altitude on the dependence between pixels. For example, the spatial patterns of stronger dependency around pixel A in the alpine area are much narrower compared to those around pixels C and B in more flat terrains. Furthermore, the mountain range in

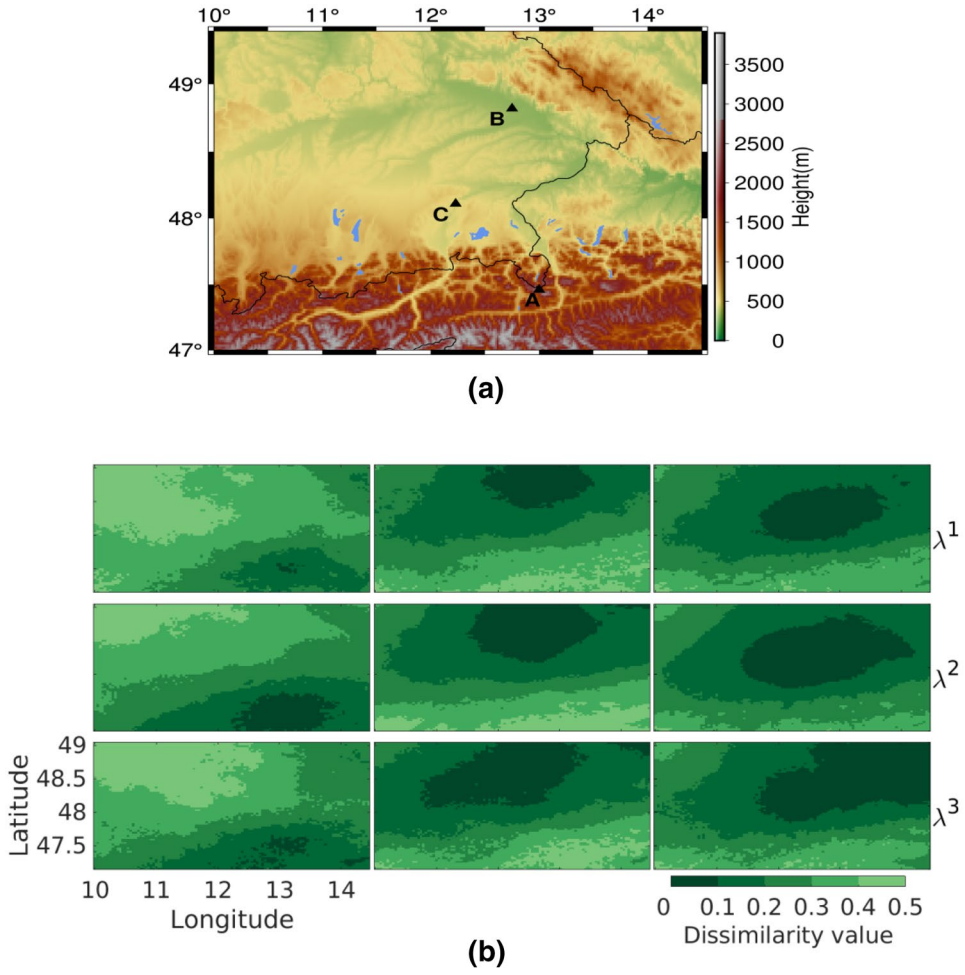

Fig. 5 **a** Topography and location of the target pixels. **b** Spatial representation of the dissimilarity measures λ^1(upper row),λ^2(middle row) and λ^3 (bottom row) for pixels A (left column), B (center column) and C (right column). Due to their definition, lower values indicate stronger statistical relationships

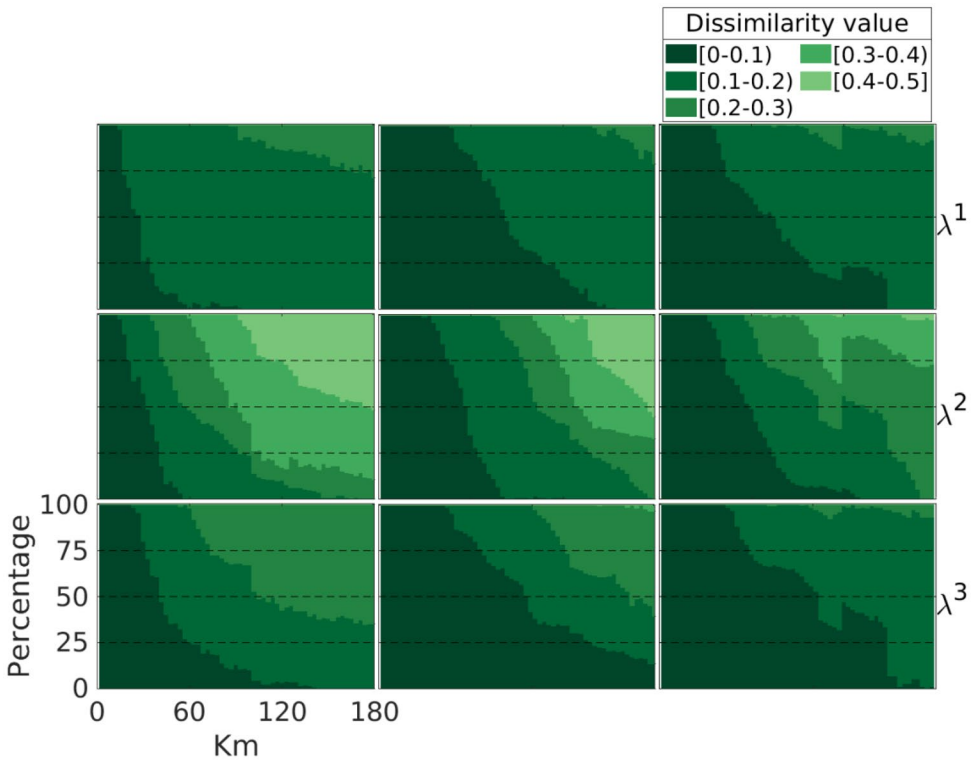

Fig. 6 Distance-based evaluation of λ^1 (upper row), λ^2(middle row) and λ^3(bottom row) for target pixels A (left column), B (middle column) and C (right column)

the southern part of the domain can be clearly identified in the dissimilarity measures from pixels B and C, where the highest values, i.e., little dependency, are found.

A different way to look at the spatial dependency is shown in Fig. 6, where the percentage of pixels within a certain dissimilarity category is plotted against the distance from the target pixel. Such analysis is similar to the widely used spatial variogram, with which we can investigate the spatial variability of geophysical variables. An extension to the full domain including all possible combinations of pixel pairs (not only those with the three target pixels) would finally allow for a Copula-based spatial variogram.

Again, the figure clearly shows the relationship between the spatial dependence and the distance from the target pixels. Pixel A has a much narrower radius of higher dissimilarity values compared to pixels B and C. For λ^1 and λ^3 the highest values can only be observed within a radius of less than 60 km around pixel A. Pixel C, which is located in the foothills of the Alps, shows a slightly larger radius with low dissimilarity values. Finally, Pixel B has a radius of around 80 km, where values between 0 and 0.1 can be observed.

For all target pixels, λ^2 shows a sharper increase of dissimilarity with increasing distance. It should be noted that λ^1 does not allow for analyzing the degree of asymmetry in the empirical Copula as both lower and upper tails are treated in the same way. However, as we have identified the asymmetric Clayton Copula as the most suitable model in some parts of the study domain, dissimilarity measures accounting for this asymmetry like λ^2 or λ^3 should be preferred. Moreover, both measures require a pre-analysis of all possible values of rank correlations and the Copula asymmetry within the study domain for computing the scaling parameter ξ. Therefore, the estimated measure is relative and does not allow for a simple extension to other areas.

Hence, the final choice of a suitable dependence measure depends on the application (e. g. if new data should be included in the future) and the level of asymmetry of the Copula. For linear relationships and symmetric Copulas, it is therefore suggested to use λ^1 as its computation is simple and straightforward. For other cases λ^2 or λ^3 is preferred. While λ^2 only accounts for relationships between the same time steps, λ^3 also considers time-lagged dependencies and can be assumed as the most comprehensive dissimilarity measure.

Copula-based bias correction and gap filling

Finally, we want to assess the performance of the proposed workflow for improving WRF-based ZTDs by applying a Copula-based correction. We use NSE (28) for this assessment. Figure 7 shows the NSE values for each of the 22 stations. The performance measures for the raw WRF- and Copula-based ZTDs are shown in blue and red, respectively. Over most stations, the performance of the WRF-based ZTDs can be significantly improved using the Copula-based

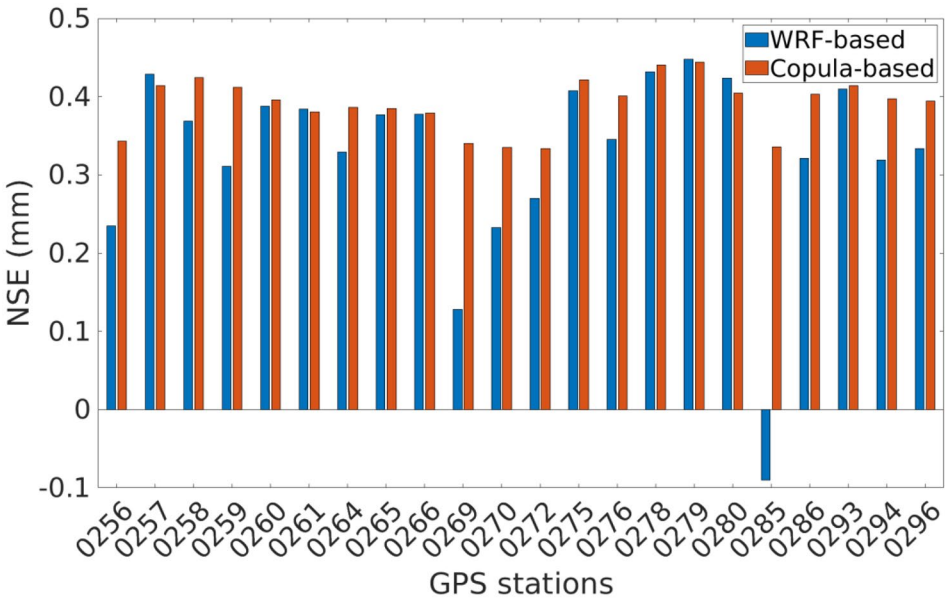

Fig. 7 NSE between purely WRF (blue) and Copula (red) based against GPS-derived ZTDs. Better performance is indicated by higher NSE values

correction. In particular, over the stations that are located in the Alps (269, 270, 272, 285, 294, 296), the NSE values of the Copula-based ZTDs are much higher compared to the purely WRF-based estimates. This can be explained by the fact that we assume higher biases of the modeled atmospheric quantities, especially in complex terrain and high altitudes. The stations where the Copula-based NSE values are similar or slightly worse compared to the purely WRF-derived estimates (257, 261, 266, 278, 279, 280, 293) are located in low-altitude and flat areas. Hence, in those regions, it can be assumed that the pure WRF estimates are already acceptable and that the dependency between the WRF- and GNSS-based anomalies does not allow for a significant improvement of the WRF ZTDs toward the station data. Nevertheless, as most stations show higher NSE values after applying the Copula-based correction, it can be concluded that the proposed approach is generally capable of improving WRF-derived ZTDs across the study domain.

Summary and conclusion

In this paper, we provide a workflow for the modeling of dependence structures using Copulas. While we apply the proposed methods to ZTDs, it can be easily used in other GNSS and geodesy applications when, e.g., marginal distributions are not known or if the analysis of multivariate dependency structures is required. Compared to approaches that require, e.g., normally distributed data, this framework allows for a much more flexible and general description of statistical relationships with arbitrary distributions including dedicated measures to analyze, e.g., tail dependence.

We use the proposed workflow to investigate the dependence structures of modeled and observed ZTDs across a study domain in Central Europe. Due to the complex topography in this domain, it is assumed that these dependency structures are not only linear but might contain strong nonlinearity as well.

First, modeled humidity, pressure and temperature from WRF are used for calculating ZTDs in the study area during the period of April to October 2016. Three pixels are selected as focus points representing different altitudes from flat to mountainous terrain. For each of these pixels, we identify the most suitable Copula between the target pixels and all other pixels in the study domain. This allows for a description of the statistical relationships of ZTDs across the whole region. While we use five different Copulas (Gaussian, Gumbel, Clayton, Frank and AMH Copula), the Gaussian Copula is identified as the most suitable model across most parts of the domain. However, in several regions, the Clayton Copula, which allows for an asymmetric tail dependence, is more appropriate. This indicates that for large-scale applications, where we cannot manually analyze every single location, we have to use approaches that also take more complex (e.g., nonlinear) relationships into account to capture the full dependency between variables.

We also discuss three Copula-based dissimilarity measures, which are used to analyze the spatial variability of ZTDs. We compute the measures between three target pixels and all other pixels in the study domain to investigate the relationship between statistical dependency and distance. All three measures show more spatial details compared to, e.g., Pearson's correlation. For example, the dissimilarity patterns clearly reveal a significant difference between the flat areas in the center of the domain and the alpine region in the

South. Therefore, such Copula-based dissimilarity measures are highly suitable to describe more complex (or previously unknown) statistical relationships.

We further apply a Copula-based correction approach for generating GNSS-like ZTDs from purely WRF-derived estimates. While the performance over some stations is similar or even slightly worse after correction, most corrected time series show significantly improved NSE values when compared against the GNSS-based ZTDs. In particular, the corrected ZTDs in the alpine regions in the South of the study domain show a much better agreement with the station data. Hence, this Copula-based approach is suitable for, e.g., filling gaps or extending discontinued GNSS measurements using an atmospheric model or correcting modeled atmospheric quantities using a set of station observations.

As a conclusion, our study provides tools and measures for modeling, analyzing and applying complex and nonlinear statistical relationships not only between ZTDs, but between geophysical variables in general. It hence is an innovative approach when working with new or unknown data or if extreme values are of major importance. All in all, it provides a statistical basis for the development of future methods in fields of data fusion or assimilation, bias correction, gap filling and many other applications in geodesy and geosciences.

Acknowledgements This study was funded by scholarships from K. N. Toosi University of Technology (Tehran, Iran) and the Karlsruhe Institute of Technology—Institute of Meteorology and Climate Research, Atmospheric Environmental Research (Garmisch-Partenkirchen, Germany). It was enabled additionally by funds from the German Research Foundation (DFG-ATMOWATER, KU 2090/10) and the German Ministry of Science and Education (BMBF)-funded GROW-SaWaM project. We would further like to thank the German Research Center for Geosciences (GFZ) for providing the GNSS data that were used for this work.

Appendix: The estimation of zenith tropospheric delay

Zenith tropospheric delays (ZTDs) can be expressed in two components:

$$\mathrm{ZTD} = \mathrm{ZHD} + \mathrm{ZWD} \tag{29}$$

where ZHD and ZWD are dry or zenith hydrostatic and wet delays considered from the surface to the top level in the troposphere, respectively. They are computed by (Ghoddousi-Fard 2009)

$$\mathrm{ZHD} = 10^{-6} \sum_{i=\mathrm{surface}}^{\text{last level}} \left(k_1 R_\mathrm{d} \left(\frac{p_i - e_i}{R_\mathrm{d} T_i} + \frac{e_i}{R_\mathrm{w} T_i} \right) \right) \mathrm{d}h_i \tag{30}$$

$$\mathrm{ZWD} = 10^{-6} \sum_{i=\mathrm{surface}}^{\text{Last level}} \left(k_2' \frac{e_i}{T_i} + k_3 \frac{e_i}{T_i^2} \right) \mathrm{d}h_i \tag{31}$$

In these equations, T_i is the temperature in Kelvin, p_i is the total pressure in hPa, and $R_\mathrm{d} = 287$ J/kg k and $R_\mathrm{w} = 461$ J/kg k are gas constants for dry air and water vapor, respectively. Moreover, the index i refers to the level and k_1, k_2' and k_3 are empirically determined constants. Various researchers have proposed different values for these parameters. For example, according to (Bevis et al. 1994):

$$\begin{aligned} k_1 &= 77.60 \pm 0.05\ \mathrm{Kh\,Pa^{-1}} \\ k_2' &= 22.2 \pm 2.2\ \mathrm{K\,h\,Pa^{-1}} \\ k_3 &= 3.739 \pm 0.012 \times 10^5\ \mathrm{K^2\,h\,Pa^{-1}} \end{aligned} \tag{32}$$

Finally, e_i is the water vapor pressure in hPa and is computed using the following formula:

$$e_i = \frac{q_i p_i}{\frac{M_\mathrm{w}}{M_\mathrm{d}} + \left(1 - \frac{M_\mathrm{w}}{M_\mathrm{d}}\right) q_i} \tag{33}$$

where q_i is specific humidity in kg/kg and M_w and M_d are the molar weight of wet and dry air, respectively. These parameters, together with the gas constants, fulfill in the following relation:

$$\frac{M_\mathrm{w}}{M_\mathrm{d}} = \frac{R_\mathrm{w}}{R_\mathrm{d}} = 0.62197 \simeq 0.622 \tag{34}$$

Since there is no bending effect in the zenith direction, the distance traveled by the ray in each layer can be considered the same as user-defined values (integration step size). Hence, ZTD can be calculated simply by the following summation (Ghoddousi-Fard 2009):

$$\mathrm{ZTD} = 10^{-6} \sum_{i=\mathrm{surface}}^{\text{last level}} \left(\left(k_1 R_\mathrm{d} \left(\frac{p_i - e_i}{R_\mathrm{d} T_i} + \frac{e_i}{R_\mathrm{w} T_i} \right) \right) + \left(k_2' \frac{e_i}{T_i} + k_3 \frac{e_i}{T_i^2} \right) \right) \times \mathrm{d}h_i \tag{35}$$

where $\mathrm{d}h_\mathrm{i}$ is the integration step size at level i, and other parameters have already been introduced in (30) to (34). Using (35) one calculates the total amount of ZTD through the troposphere in each position.

References

Aho K, Derryberry D, Peterson T (2014) Model selection for ecologists: the worldviews of AIC and BIC. Ecology 95(3):631–636

Akaike H (1974) A new look at the statistical model identification. IEEE Trans Autom Control 19(6):716–723. https://doi.org/10.1109/TAC.1974.1100705

Ang A, Chen J (2002) Asymmetric correlations of equity portfolios. J Financ Econ 63(3):443–494

Arbenz P (2013) Bayesian Copulae distributions, with application to operational risk management—some comments. Methodol Comput Appl Probab 15(1):105–108. https://doi.org/10.1007/s11009-011-9224-0

Bardossy A (2006) Copula based geostatistical models for groundwater quality parameters. Water Resour Res. https://doi.org/10.1029/2005WR004754

Bardossy A, Li J (2008) Geostatistical interpolation using Copulas. Water Resour Res. https://doi.org/10.1029/2007WR006115

Bevis M, Businger S, Chiswell S, Herring T, Anthes RA, Rocken C, Ware RH (1994) GPS meteorology: mapping zenith wet delays onto precipitable water. J Appl Meteorol 33(3):379–386. https://doi.org/10.1175/1520-0450(1994)033%3c0379:GMMZWD%3e2.0.CO;2

Brunel N, Pieczynski W (2005) Unsupervised signal restoration using hidden Markov chains with Copulas. Signal Process 85(12):2304–2315. https://doi.org/10.1016/j.sigpro.2005.01.018

Byram S, Hackman C, Tracey J (2011) Computation of a high precision GPS-based troposphere product by the USNO. In: Proceedings ION GNSS 2011, September 20-23. Institute of Navigation, Portland, OR, pp 572–578

Charpentier A, Fermanian J-D, Scaillet O (2007) The estimation of Copulas: theory and practice. In: Rank J (ed) Copulas: from theory to application in finance. Risk Books, London, pp 35–64

Chen B, Dai W, Liu Z, Wu L, Kuang C, Ao M (2018) Constructing a precipitable water vapor map from regional GNSS network observations without collocated meteorological data for weather forecasting. Atmos Meas Tech 11(9):5153–5166. https://doi.org/10.5194/amt-11-5153-2018

Cherubini U, Luciano E, Vecchiato W (2004) Copula methods in finance. Wiley, England

Choros B, Ibragimov R, Permiakova E (2010) Copula estimation. In: Jaworski P, Durante F, Härdle WK, Rychlik T (eds) Copula theory and its applications, vol 198. Springer, Berlin, pp 77–91. https://doi.org/10.1007/978-3-642-12465-5_3

Darling DA (1957) The Kolmogorov–Smirnov, Cramer–von Mises tests. Ann Math Stat 28(4):823–838. https://doi.org/10.1214/aoms/1177706788

Dee D et al (2011) The ERA-Interim reanalysis: configuration and performance of the data assimilation system. Q J R Meteorol Soc 137(656):553–597. https://doi.org/10.1002/qj.828

Deheuvels P (1979a) La fonction de dépendence empirique et ses propiétés. Un test non paramétrique d'indépendence. Acad R Belg Bull Cl Sci 65(5):274–292

Deheuvels P (1979b) Non parametric tests of independence. In: Raoult J-P (ed) Statistique non Paramétrique Asymptotique. Springer, Berlin. https://doi.org/10.1007/BFB0097426

Deheuvels P (1981) An asymptotic decomposition for multivariate distribution-free tests of independence. J Multivar Anal 11(1):102–113. https://doi.org/10.1016/0047-259X(81)90136-6

Dousa J, Elias M, Vaclavovic P, Eben K, Krc P (2018) A two-stage tropospheric correction model combining data from GNSS and numerical weather model. GPS Solut 22(3):274–292. https://doi.org/10.1007/s10291-018-0742-x

Durrleman V, Nikeghbali A, Roncalli T (2000) Which Copula is the right one? SSRN Electron J. https://doi.org/10.2139/ssrn.1032545

Embrechts P, Lindskog F, McNeil A (2003) Modelling dependence with copulas and applications to risk management. In: Rachev S (ed) Handbook of heavy tailed distributions in finance. Elsevier, Amsterdam

Fang Y, Madsen L, Liu L (2014) Comparison of two methods to check Copula fitting. IAENG Int J Appl Math 44(1):53–61

Fersch B, Senatore A, Adler B, Arnault J, Mauder M, Schneider K, Völksch I, Kunstmann H (2020) High-resolution fully coupled atmospheric–hydrological modeling: a cross-compartment regional water and energy cycle evaluation. Hydrol Earth Syst Sci 24:2457–2481. https://doi.org/10.5194/hess-24-2457-2020

Gendt G, Dick G, Reigber C, Tommasini M, Liu Y, Romatschi M (2004) Near real time GPS water vapor monitoring for numerical weather prediction in Germany. J Meteorol Soc Jpn 82(1B):361–370

Genest C, Favre AC (2007) Everything you always wanted to know about Copula modeling but were afraid to ask. J Hydrol Eng 12(4):347–368

Genest C, Ghoudi K, Rivest LP (1995) A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. Biometrika 82(3):543–552

Giannaros C, Kotroni V, Lagouvardos K, Giannaros TM, Pikridas C (2020) Assessing the impact of GNSS ZTD data assimilation into the WRF modeling system during high-impact rainfall events over Greece. Remote Sens. https://doi.org/10.3390/rs12030383

Ghoddousi-Fard R (2009) Modelling tropospheric gradients and parameters from NWP models: Effects on GPS estimates. Dissertation, University of New Brunswick

Goshtasby AA (2012) Similarity and dissimilarity measures. In: Image registration. Advances in computer vision and pattern recognition. Springer, London

Heng L, Gao GX, Walter T, Enge P (2011) Statistical characterization of GPS signal-in-space errors. In: Proceedings ION ITM 2011, January 24–26. Institute of Navigation, Cambridge, MA, pp 312–319

Hitney HV, Richter JH, Pappert RA, Anderson KD, Baungartner GB (1985) Tropospheric radio propagation assessment. Proc IEEE 73(2):265–283

Huard D, Evin G, Favre AC (2006) Bayesian Copula selection. Comput Stat Data Anal 51(2):809–822. https://doi.org/10.1016/j.csda.2005.08.010

Jgouta M, Nsiri B, Marrakh R (2016) Usage of a correction model to enhance the evaluation of the zenith tropospheric delay. Int J Appl Eng Res 11:4648–4654

Joe H (1997) Multivariate models and dependence concepts. Chapman & Hall, London

Kim G, Silvapulle MJ, Silvapulle P (2007) Comparison of semiparametric and parametric methods for estimating Copulas. Comput Stat Data Anal 51(6):2836–2850

Koronovskii NV, Naimark AA (2012) The Unpredictability of earthquakes as the fundamental result of the nonlinearity of geodynamic systems. Mosc Univ Geol Bull 67(6):323–331. https://doi.org/10.3103/S0145875212060026

Laux P, Vogl S, Qiu W, Knoche HR, Kunstmann H (2011) Copula-based statistical refinement of precipitation in RCM simulations over complex terrain. Hydrol Earth Syst Sci 15(7):2401–2419. https://doi.org/10.5194/hess-15-2401-2011

Longin F, Solnik B (2002) Extreme correlation of international equity markets. J Finance 56(2):649–676. https://doi.org/10.1111/0022-1082.00340

Lorenz C, Montzka C, Jagdhuber T, Laux P, Kunstmann H (2018) Long-term and high-resolution global time series of brightness

temperature from Copula-based fusion of SMAP enhanced and SMOS data. Remote Sens. https://doi.org/10.3390/rs10111842
Mao G, Vogl S, Laux P, Wagner S, Kunstmann H (2015) Stochastic bias correction of dynamically downscaled precipitation fields for Germany through Copula-based integration of gridded observation data. Hydrol Earth Syst Sci 19(4):1787–1806. https://doi.org/10.5194/hess-19-1787-2015
Mendez Astudillo J, Lau L, Tang YT, Moore T (2018) Analysing the zenith tropospheric delay estimates in on-line precise point positioning (PPP) services and PPP software packages. Sensors. https://doi.org/10.3390/s18020580
Mikosch T (2006) Copulas: tales and facts. Extremes 9(1):3–20. https://doi.org/10.1007/s10687-006-0015-x
Modiri S, Belda S, Heinkelmann R, Hoseini M, Ferrándiz JM, Schuh H (2018) Polar motion prediction using the combination of SSA and Copula-based analysis. Earth Planets Space. https://doi.org/10.1186/s40623-018-0888-3
Nelsen RB (2006) An introduction to Copulas. Springer series in statistics, 2nd edn. Springer, Berlin
Nievinski F, Cove K, Santos M, Wells1 D, Kingdon R (2005) Range-extended GPS kinematic positioning using numerical weather prediction model. In: Proceeding of ION AM. Institute of Navigation, Cambridge, MA, June 27–29, pp 902–911
Norberg J, Roininen L, Vierinen J, Amm O, McKay-Bukowski D, Lehtinen M (2015) Ionospheric tomography in Bayesian framework with Gaussian Markov random field priors. Radio Sci 50(2):138–152. https://doi.org/10.1002/2014RS005431
Nordman M, Eresmaa RI, Poutanen M, Järvinen HJ, Koivula H, Luntama J-P (2007) Using numerical weather prediction model derived tropospheric slant delays in GPS processing: a case study. Geophysica 43(1–2):49–57
Rohm W, Guzikowski J, Wilgan K, Kryza M (2019) 4DVAR assimilation of GNSS zenith path delays and precipitable water into a numerical weather prediction model WRF. Atmos Meas Tech 12(1):345–361. https://doi.org/10.5194/amt-12-345-2019
Samaniego L, Bardossy A, Rohini K (2010) Streamflow prediction in ungauged catchments using copula-based dissimilarity measures. Water Resour Res. https://doi.org/10.1029/2008WR007695
Schwarz G (1978) Estimating the dimension of a model. Ann Stat 6(2):461–464. https://doi.org/10.1214/aos/1176344136
Singh R, Ojha SP, Puviarasan N, Singh V (2019) Impact of GNSS signal delay assimilation on short range weather forecasts over the Indian region. JGR Atmos. https://doi.org/10.1029/2019JD030866
Skamarock WC, Klemp JB, Dudhia J, Gill DO, Barker D, Duda MG, Powers JG (2008) A description of the advanced research WRF version 3. NCAR technical note NCAR/TN-475+STR. National Center for Atmospheric Research, Boulder, Colorado, USA. https://doi.org/10.5065/D68S4MVH
Sklar A (1959) Fonctions de repartition a n dimensions et leurs marges. de l'Institut de Statistique de l'Universite de Paris 8:229–231
Sun J, Wu Z, Yin Z, Ma B (2017) A simplified GNSS tropospheric delay model based on the nonlinear hypothesis. GPS Solut 21(4):1735–1745. https://doi.org/10.1007/s10291-017-0644-3
Tiberius CCJM, Borre K (2000) Are GPS data normally distributed. In: Schwarz KP (ed) Geodesy beyond 2000. International association of geodesy symposia, vol 121. Springer, Berlin
Topcu C (2016) Comparison of some selection criteria for selecting bivariate Archimedean Copulas. AKU-J Sci Eng 16:250–255. https://doi.org/10.5578/fmbd.27971
Trivedi PK, Zimmer DM (2005) Copula modeling: an introduction for practitioners. Found Trends Econom 1(1):1–111
Turel N, Arikan F (2010) Probability density function estimation for characterizing hourly variability of ionospheric total electron content. Radio Sci. https://doi.org/10.1029/2009RS004345
Vandenberghe S, Verhoest NEC, De Baets B (2010) Fitting bivariate Copulas to the dependence structure between storm characteristics: a detailed analysis based on 105 year 10 min rainfall. Water Resour Res. https://doi.org/10.1029/2009WR007857
Vogl S, Laux P, Qiu W, Mao G, Kunstmann H (2012) Copula-based assimilation of radar and gauge information to derive bias-corrected precipitation fields. Hydrol Earth Syst Sci 16(7):2311–2328. https://doi.org/10.5194/hess-16-2311-2012
Wilgan K, Hurter F, Geiger H, Rohm R, Bosy J (2017) Tropospheric refractivity and zenith path delays from least-squares collocation of meteorological and GNSS data. J Geod 91(2):117–134. https://doi.org/10.1007/s00190-016-0942-5
Wilk MB, Gnanadesikan R (1968) Probability plotting methods for the analysis of data. Biometrika 55(1):1–17. https://doi.org/10.1093/biomet/55.1.1
Wolfowitz J (1957) The minimum distance method. Ann Math Stat 28(1):75–87
Ye L, Hanson LS, Ding P, Wang D, Vogel RM (2018) The probability distribution of daily precipitation at the point and catchment scales in the United States. Hydrol Earth Syst Sci 22(12):6519–6531. https://doi.org/10.5194/hess-22-6519-2018
Zus F, Douša J, Kacmarík M, Vaclavovic P, Balidakis K, Dick G, Wickert J (2019) Improving GNSS zenith wet delay interpolation by utilizing tropospheric gradients: experiments with a dense station network in central Europe in the warm season. Remote Sens. https://doi.org/10.3390/rs11060674

Roya Mousavian received her M.Sc. degree in Geodesy at K. N. Toosi University of Technology in 2013, and she is a Ph.D. candidate of Geodesy there. She did a research visit at the Institute of Meteorology and Climate Research (IMK-IFU) of Karlsruhe Institute of Technology (KIT) in 2018 and 2019. Her research interests are GNSS meteorology, satellite geodesy and slow earthquakes.

Christof Lorenz is a postdoctoral researcher at the Campus Alpin of the Karlsruhe Institute of Technology. He studied Geodesy and Geoinformatics at the University of Stuttgart and did his Ph.D. in the field of climate and environmental sciences at the University of Augsburg. His current research focuses on the development of empirical–statistical data fusion approaches, especially for seasonal hydrometeorological predictions and remote sensing information.

Masoud Mashhadi Hossainali is an associate professor in Geodesy and dean of faculty of Geodesy and Geomatics Engineering at K. N. Toosi University of Technology. He received his Ph.D. degree at Darmstadt University of Technology. His current research interests are GNSS meteorology, positioning and navigation, deformation monitoring and constellation design for various space mission.

Benjamin Fersch received the Diploma degree in hydrology from the Albert-Ludwigs University, Freiburg, and the Ph.D. degree from the University of Augsburg in 2007 and 2011, respectively. Since 2007 he is affiliated with the Karlsruhe Institute of Technology, Institute of Meteorology and Climate Research (IMK-IFU). In 2011 and 2013 he joined the National Center for Atmospheric Research (NCAR) Boulder as a visiting scientist. His research emphasis is on soil moisture observation and fully coupled hydrological—atmospheric local area modeling.

Harald Kunstmann is a professor and chair for regional climate and hydrology at the University of Augsburg in joint appointment with Karlsruhe Institute of Technology (Campus Alpin). His current research focuses on observation and modeling of the coupled atmospheric-terrestrial water cycle for different regions worldwide.