Generating Synthetic Rainfall Fields by R-vine Copulas Applied to Seamless Probabilistic Predictions

Peter Schaumann¹, Martin Rempel², Ulrich Blahak² and Volker Schmidt¹

¹Institute of Stochastics, Ulm University, Ulm, Germany ²Deutscher Wetterdienst, Offenbach, Germany

Abstract

Many post-processing methods improve forecasts at individual locations but remove their correlation structure. However, this information is essential for forecasting larger-scale events, such as the total precipitation amount over areas like river catchments, which are relevant for weather warnings and flood predictions. We propose a method to reintroduce spatial correlation into a post-processed forecast using an R-vine copula fitted to historical observations. The method rearranges predictions at individual locations and ensures that they still exhibit the post-processed marginal distributions. It works similarly to well-known approaches, like the "Schaake shuffle" and "ensemble copula coupling". However, compared to these methods, which rely on a ranking with no ties at each considered location in their source for spatial correlation, the copula serves as a measure of how well a given arrangement compares with the observed historical distribution. Therefore, no close relationship is required between the post-processed marginal distributions and the spatial correlation source. This is advantageous for post-processed seamless forecasts in two ways. First, meteorological parameters such as the precipitation amount, whose distribution has an atom at zero, have rankings with ties. Second, seamless forecasts represent an optimal combination of their input forecasts and may spatially shifted from them at scales larger than the areas considered herein, leading to non-reasonable spatial correlation sources for the well-known methods. Our results indicate that the calibration of the combination model carries over to the output of the proposed model, i.e., the evaluation of area predictions shows a similar improvement in forecast quality as the predictions for individual locations. Additionally, the spatial correlation of the forecast is evaluated with the help of objectbased metrics, for which the proposed model also shows an improvement compared to both input forecasts.

1 Introduction

In operational weather forecasting, forecasters are supported by various forecast models to issue targeted warnings of potentially hazardous weather phenomena. The longer such warnings are accurate and reliable, the more time decisionmakers in hydrological and civil protection have to diminish possible harm to life and property. Usually, these warnings rely on nowcasting systems and numerical weather prediction (NWP). Both give valuable guidance, however, for different lead times (Hess, 2020; Ruti et al., 2020). Thus, both systems represent different sources of predictability that can be combined to give the best forecast at any time. The so-called seamless prediction is part of a whole value cycle as described in Ruti et al. (2020), reaching from information generation toward the outcomes and values.

The project Seamless Integrated Forecasting System (SINFONY) of Deutscher Wetterdienst (DWD) focuses on the seamless prediction of precipitation within the short-term range up to 12 h ahead. As a first step to achieve the combination, both forecasting techniques—nowcasting and NWP—are enhanced individually such that the gap between both of them is narrowed in terms of a verification metric. Then, based on the improved forecast systems, developing and implementing custom combination methods leads to a unique user-oriented forecast, including the best information for both individual forecasts. Within this project, among other techniques, we have developed an approach to seamlessly combine 6 h forecasts of the observation-based precipitation nowcasting scheme "STEPS-DWD" and the NWP model "ICON-D2-EPS". Both models provide ensemble forecasts and they are subsequently referred to as input forecasts.

The main component of the combination approach mentioned above is a model we proposed in Rempel et al. (2022). It is based on neural networks for combining and post-processing two ensemble forecasts (briefly called C^3 -model in the following, where C^3 stands for *combined*, *calibrated*, *consistent*). The output of the C^3 -model consists of calibrated probabilities for the exceedance of a set of precipitation thresholds at each grid point of a regular horizontal grid, see also Schaumann et al. (2021). Forecasts of the C^3 -model could provide a data basis for future customized warnings. However, a drawback of the model is the loss of any spatial correlation of the input forecasts during the calibration process.

This loss is a well-known problem often discussed in the literature about post-processing of ensemble forecasts (Bellier et al., 2018; Jobst et al., 2023; Möller et al., 2013; Wu et al., 2018). To overcome this drawback, several approaches have been developed that reintroduce spatial information to post-processed marginal distributions. Typically, these approaches are based on the "Schaake shuffle" (Clark et al., 2004) or on "ensemble copula coupling" (Schefzik et al., 2013). In these methods, sample values are drawn from the post-processed marginal distributions for individual locations and rearranged into synthetic ensembles according to the rank correlation of one provided ensemble forecast or a set of past observations. The resulting synthetic ensemble follows the post-processed marginal distributions while preserving the spatial correlations of the raw (i.e. unprocessed) data.

On the one hand, a close relationship between the source of the spatial correlation (e.g. one ensemble forecast or a set of past observations) and the post-processed marginal distributions is required to obtain suitable spatial correlations for these rearrangement methods (Schefzik, 2016). On the other hand, the ranks should be unambiguous to achieve a realistic result. These conditions are fulfilled in common applications of the "Schaake shuffle" or "ensemble copula coupling", where only one post-processed ensemble forecast is rearranged (Lerch et al., 2020; Lakatos et al., 2023, e.g.).

However, the conditions mentioned above are no longer met for seamless

probabilistic forecasts for precipitation provided by the C^3 -model (Rempel et al., 2022). The output of this model represents an optimal combination of predictions of two ensemble prediction systems: the precipitation nowcasting scheme "STEPS-DWD" (Reinoso-Rondinel et al., 2022) and the short-range high-resolution NWP-model "ICON-D2-EPS" (Zängl et al., 2015). Furthermore, precipitation forecasts have an atom at 0 mm, whereby the ranks are ambiguous and, due to thr blending of the two sometimes completely different ensemble forecasts, the output of the C³-model is a mixture of the distributions of the input forecasts. Being a mixture of both input forecasts, it may be spatially shifted compared to each input model at scales larger than the considered area for spatial rearranging. Thus, no corresponding ensemble forecasts exist that could be used to transfer its spatial correlation structure.

Therefore, in the present paper, we propose a multivariate stochastic model for generating synthetic ensemble members based on the idea of the "Schaake shuffle". However, in our approach, the spatial correlations are not derived from a given ensemble nor directly from historical observations but are modelled through an R-vine copula. Vine copulas are a powerful tool for parametric modelling of multivariate probability distributions (Joe and Kurowicka, 2011) and offer the flexibility to describe their tail behaviour adequately (Czado and Nagler, 2021). Thus, the need for having an ensemble forecast available in advance as a source of spatial correlations is circumvented. Furthermore, the spatial correlation structure can be reconstructed in any area without a readjustment of the model components for the new area considered. As a result, we offer the possibility to provide forecast information on user-oriented customized areas, which can be a basis for flexible warnings to end-users. Thus, with additional impact data, we can support the transition process of the current operational warning system towards an impact-oriented one (Kaltenberger et al., 2020; Kox et al., 2018; Potter et al., 2021). Finally, we estimate the distribution of the total precipitation amount within given regions (e.g. river catchments or municipal areas) using synthetic ensemble members drawn from the multivariate stochastic model.

Vine copulas are recently used in the field of energy meteorology to model the spatial interrelation of errors in probabilistic forecasts to assess the uncertainties in the power generation of photovoltaic systems (Aigner et al., 2023; Schinke-Nendza et al., 2021), or for the inclusion of spatial dependencies between multiple wind farms in wind energy scenarios (Li et al., 2022; Tu et al., 2023). Another field of application is within so-called weather generators. Since high-resolution time series of atmospheric variables are only available for a limited period (e.g. 30 years), it is necessary to extend these time series when assessing extreme events (Van de Velde et al., 2023) or designing hydrological applications (Callau Poduje and Haberlandt, 2018). In Brunner et al. (2019), several copula models have been evaluated to reproduce the spatial dependencies of gauging stations in a river catchment. Furthermore, in Erhardt et al. (2015), a spatial model based on an R-vine copula has been introduced to predict time series of the daily mean temperature at unobserved locations, and, in Nazeri Tahroudi et al. (2022), various C-, D-, and R-vine copulas have been examined to estimate rainfall deficiency structures in an Iranian river basin.

It should be noted that in the meteorological literature, the term "spatial correlation" sometimes refers to the spatial correlation of forecast uncertainty, which is then called "spatial error correlation", see, e.g. Feldmann et al. (2015), as opposed to the spatial correlation of the precipitation amount itself. The reason for this distinction is that the (actually observed) precipitation amount can be mathematically modelled either as a deterministic quantity or as a random variable.

The present paper is structured as follows. First, the utilized datasets are introduced in Section 2. This is followed in Section 3 by a description of the adjustment of the C^3 -model by a quantile regression, the generation of synthetic ensemble members, and the prediction of the total precipitation amount in a given area. Then, Section 4 discusses the proposed model's implementation details. Afterwards, the validation results are presented in Section 5. Finally, conclusions are drawn in Section 6.

2 Data

For adding spatial correlations to the C^3 -model forecasts, we use the same dataset as in Rempel et al. (2022). That set consists of, on the one hand, precipitation extrapolations of "STEPS-DWD", a nowcasting scheme. On the other hand, forecasts of an experimental version of "ICON-D2-EPS" are used, a high-resolution short-term NWP model. Both systems provide ensemble forecasts. In order to keep the present paper largely self-contained, we provide a brief description of the dataset, which covers three time periods from the years 2016, 2019 and 2020 (2016-05-26 - 2016-06-26, 2019-06-01 - 2019-06-23, 2020-06-03 - 2020-07-16).

2.1 STEPS-DWD

"STEPS-DWD" has been developed within SINFONY as an adaption of "STEPS" (Short-Term Ensemble Prediction System), see Bowler et al. (2006); Foresti et al. (2016); Seed (2003); Seed et al. (2013), where DWD's radar network provides radar reflectivities based on which precipitation rates are estimated by a hydrometeor-dependent Z-R relation optimised for the radar stations utilised by DWD (Steinert et al., 2021).

For the present study, "STEPS-DWD" is configured to consist of a cascade of first-order autoregressive processes on twelve spatial scales and to apply a new localisation approach (Pulkkinen et al., 2020; Reinoso-Rondinel et al., 2022) for the estimation of the autoregressive parameters on each individual scale. The spatially correlated noise field is estimated globally but is imprinted only in regions with precipitation due to the localised autoregressive parameters.

We use further 30-member "STEPS-DWD" extrapolations that are generated every 30 minutes running 6 h ahead on a $1 \times 1 \text{ km}^2$ grid with a temporal resolution of 5 minutes. For our purposes, all 5-minute forecasts within a given hour are aggregated into one hourly forecast. In the same way, radar observations are aggregated to obtain the ground truth for the C^3 -model and the estimated precipitation amounts for validating the synthetic rainfall fields.

2.2 ICON-D2-EPS

Furthermore, we use forecasts of an experimental version of "ICON-D2-EPS" (Zängl et al., 2015) that runs in limited area mode (LAM) with a horizontal grid spacing of $\Delta x \approx 2.2$ km. The domain of the 20-members ensemble is centred on Central Europe. In addition to conventional observation data and MODE-S aircraft measurements, radar reflectivities and radial winds of the 3D volume scans are assimilated by DWD's kilometre-scale ensemble data assimilation system KENDA. It implements a localised ensemble transform Kalman filter (Bick et al., 2016; Schraff et al., 2016). Only the first 20 members serve as initial conditions for the forecasts, while 40 members are used for the assimilation. "ICON-EU ensemble" forecasts (larger trans-European domain, grid spacing 6.5 km, parameterised deep convection) provide lateral and upper boundary conditions. The operational conventional one-moment cloud microphysics scheme is used.

Our experimental setting of "ICON-D2-EPS" generates hourly forecasts running 12 h ahead. The native ICON output is provided on an irregular triangular grid. Hence, forecasts of both—"STEPS-DWD" and "ICON-D2-EPS"—as well as the observations are interpolated onto a common regular $2.2 \times 2.2 \text{ km}^2$ grid that was established for the former operational NWP model COSMO-D2.

$2.3 C^3$ -model

The so-called C^3 -model further processes the previously presented "STEPS-DWD" and "ICON-D2-EPS" forecast datasets. It consists of a neural network with a feed-forward architecture featuring multiple hidden layers trained to merge precipitation forecasts into one combined probabilistic prediction. The main objective of this model is to learn and correct the biases within both input forecasts and to generate a calibrated seamless transition between the input forecast models.

In more detail, the C^3 -model takes as input the probabilities for exceeding several precipitation thresholds estimated by both input forecast models at a set of grid points. Its output, conversely, also comprises either threshold exceedance probabilities or a set of quantiles of the probability distribution for each grid point. The output forecast exhibits improved validation scores compared to both input forecasts across all considered lead times. It serves as the input for the copula method presented in the next section. For more details on the C^3 -model, see Schaumann et al. (2021).

3 Methods

The modelling approach proposed in this paper consists of several parts. First, the C³-model is used to predict calibrated quantiles of the precipitation amount distribution for each location under consideration. However, these pointwise predictions do not take spatial correlation into account. Therefore, in the next step, an R-vine copula is fitted to precipitation observation data for modelling the multivariate precipitation distribution at multiple locations. The R-vine copula model has been selected from the vine copula models occurring in the literature because it is the most general one with no restrictions regarding the vine structure. In other words, any valid D-vine and C-vine model is also a valid R-vine model. Finally, the copula model is used with a hill-climbing algorithm to order the predicted quantiles at each considered location into synthetic ensemble members. This step is conceptually similar to "Schaake shuffle" and "ensemble copula coupling" in that spatial correlation is imposed on samples of marginal distributions by rearranging the samples. However, for rearranging the predicted precipitation values, the fitted copula serves as a measure of how well a given arrangement compares with the observed distribution of precipitation in the historical data, where no close relationship is required between the postprocessed marginal distributions and the spatial correlation source. This is an advantage in comparison to "Schaake shuffle" and "ensemble copula coupling", which rely on the existence of a ranking with no ties at each considered location in their source for spatial correlations. The ensemble members obtained by our modelling approach exhibit not only the calibrated marginal distributions predicted by the C³-model but also the spatial correlation provided by the copula model. We refer to Section 4 below regarding various implementation aspects.

3.1 Quantile regression

Recall that the output of the C^3 -model, as proposed in Rempel et al. (2022) and Schaumann et al. (2021), consists of probabilities for the exceedance of an appropriately selected family of thresholds. For reliable fitting of probability densities to these datasets of exceedance probabilities, a relatively large number of thresholds would be required to cover the entire spectrum of possible precipitation amounts. In addition, picking a suitable family of parametric densities is not trivial to model the distribution predicted by the C^3 -model. Thus, we switch from considering threshold exceedance probabilities to a quantile regression approach to limit the number of necessary data points generated by the C³-model. For this, we replace the softmax layer of the C³-model with a dense layer with a linear activation function and replace the "categorical cross-entropy" loss function with "pinball" loss functions, see Steinwart and Christmann (2011) for details. This allows us to predict a set of evenly spaced quantiles on the [0, 1]interval. By sampling values directly from this set of quantiles, we obtain data which can be seen as realizations of the precipitation amount distribution predicted by the C³-model without the necessity to fit a suitable probability density function.

3.2 Sampling of predicted precipitation amounts for single locations

For some integer, m > 1, let $V = \{v_1, \ldots, v_m\} \subset \mathbb{R}^2$ denote a set of locations, for each location v_i of which the C³-model produces a calibrated distribution of precipitation amounts, as described above. For some sample size N > 0 and for each i = 1, ..., m, we draw N sample values $x_i^1, ..., x_i^N > 0$ for a specified hour $t > t_c$ from a certain random variable X_i , whose probability distribution describes the predicted precipitation amount at location v_i . It is important to note that the sample size N does not necessarily have to be the same as the number of ensemble members in either of the input forecasts. If resampling values from the predictions of the C^3 -model is allowed, i.e., bootstrapping, then the value of N can be any positive integer. It should be noted that drawing more bootstrapped samples from the number of available quantiles does not improve the predicted marginal distribution for a specific grid point. However, increasing the number of samples can enhance the model's ability to represent the overall multivariate distribution of precipitation within the considered area by generating additional synthetic ensemble members. However, if resampling is not allowed, then the value of N is constrained by the number of quantiles generated by the C^{3} -model. Nonetheless, it should be noted that the C^{3} -model can produce an arbitrary number of quantiles if trained to do so. Here and in the following, $t_c \in \mathbb{R}$ refers to the present point in time dividing the datasets into past and future, i.e., the data available for training and validation, respectively. Note that these values are independently sampled for each individual location, i.e., the random vector (X_1, \ldots, X_m) has independent components because the C³model predicts the univariate (marginal) distributions of precipitation amounts without taking spatial correlations into account. The sample values x_i^1, \ldots, x_i^N are used later on as building blocks for the construction of synthetic ensemble members, which follow the distribution of some m-dimensional random vector (X'_1,\ldots,X'_m) that describes the precipitation amounts predicted at locations $v_1, \ldots, v_m \in V$ for the same hour t, i.e., the (univariate) distributions of X'_i and X_i coincide for each $i = 1, \ldots, m$, where the spatial correlations observed in historical precipitation data are taken into consideration in (X'_1, \ldots, X'_m) .

3.3 Modeling spatial dependencies by measured precipitation amounts

This section briefly introduces copula theory and explains how copula models can capture spatial dependencies based on measured precipitation amounts. Copulas are a statistical tool to model multivariate (non-Gaussian) probability distributions. With the help of Sklar's fundamental theorem of copula theory (see Joe and Kurowicka (2011)) the cumulative distribution function of a multivariate probability distribution can be decomposed into several distinct functions, i.e., a univariate (marginal) cumulative distribution function for each marginal distribution and a so-called copula which models the underlying correlation structure. By decomposing the joint CDF into several components, each component can be modelled separately. In other words, copulas allow for modelling a random vector's correlation structure independently of its marginal distributions and, therefore, without imposing restrictions on these marginal distributions. Since copulas do not require the multivariate distribution to be Gaussian, they are popular in areas of research where non-Gaussian phenomena such as precipitation amounts are investigated.

In the present paper, we are interested in the multivariate distribution of m-dimensional vectors of precipitation amounts. More specifically, we want to model the multivariate density for precipitation amounts at several locations. However, the implementation explained in Section 4 below only allows for continuous or discrete distributions but not for mixtures of both distribution types. Since the distribution of precipitation amounts has an atom at 0 mm, it is a mixture distribution, which necessitates that we consider discretized marginal distributions instead. Furthermore, it should be noted that the method could be applied without discretization for meteorological quantities with continuous distributions such as temperature. Therefore, both the continuous and the discrete cases will be discussed in the following.

To determine the joint distribution of (historical) precipitation amounts at a set of locations $v_1, \ldots, v_m \in V$ modelled by an *m*-dimensional random vector (Y_1, \ldots, Y_m) , we use historical datasets $(y_{1,t_1}, \ldots, y_{m,t_1}), \ldots, (y_{1,t_k}, \ldots, y_{m,t_k})$, which are considered as realizations of (Y_1, \ldots, Y_m) being available from measurements of precipitation amounts for certain past times $t_1, \ldots, t_k \leq t_c$. Recall that $t_c \in \mathbb{R}$ refers to the present point in time dividing the datasets into past and future. Moreover, we use Sklar's representation formula (Joe and Kurowicka, 2011) for the joint cumulative distribution function $G : \mathbb{R}^m \to [0, 1]$ of (Y_1, \ldots, Y_m) , which states that

$$G(y_1, \dots, y_m) = C(G_1(y_1), \dots, G_m(y_m)) \quad \text{for all } y_1, \dots, y_m \in \mathbb{R}, \quad (1)$$

where $G_i : \mathbb{R} \to [0,1]$ denotes the univariate cumulative distribution function of Y_i for $i = 1, \ldots, m$, and $C : [0,1]^m \to [0,1]$ is an *m*-variate copula. More specifically, $C : [0,1]^m \to [0,1]$ is the restriction of the *m*-variate cumulative distribution function of an *m*-dimensional random vector to the cube $[0,1]^m$ such that its components are uniformly distributed on the unit interval [0,1]for each $i = 1, \ldots, m$. Thus, to determine G, it is sufficient to determine the univariate cumulative distribution functions G_1, \ldots, G_m and the copula C. In the following, the vector (U_1, \ldots, U_m) refers to the transformed random variables $U_i = G_i(Y_i)$ for $i = 1, \ldots, m$. Analogously, realizations of U_i are denoted as $u_i = G_i(y_i)$. Since each marginal random variable Y_i is transformed by its own CDF G_i , the corresponding transformed random variable U_i is standard uniformly distributed.

Note that Sklar's representation formula given in Equation (1) can be specified if the joint distribution of (Y_1, \ldots, Y_m) is (purely) continuous or discrete. In the discrete case, it is sufficient to determine the values of the copula C: $[0,1]^m \rightarrow [0,1]$ for the joint support $R_{U_1} \times \ldots \times R_{U_m} \subset [0,1]^m$ of the vector (U_1,\ldots,U_m) , where R_{U_i} denotes the support of U_i for $i = 1,\ldots,m$, instead of considering the values of C for the entire cube $[0,1]^m$. In this paper, we are mainly interested in the joint probability density (or, in the discrete case, the probability mass function) $g : \mathbb{R}^m \to [0, \infty)$ of (Y_1, \ldots, Y_m) , instead of determining the joint cumulative distribution function $G : \mathbb{R}^m \to [0, 1]$ of (Y_1, \ldots, Y_m) . For this, the following representation formulas for g can be derived from Equation (1), see (Joe and Kurowicka, 2011): In the continuous case, the density g of (Y_1, \ldots, Y_n) is given by

$$g(y_1,\ldots,y_m) = c(G_1(y_1),\ldots,G_m(y_m))\prod_{i=1}^m g_i(y_i) \quad \text{for all } y_1,\ldots,y_m \in \mathbb{R},$$

where $g_i : \mathbb{R} \to [0,\infty)$ denotes the density of Y_i for $i = 1, \ldots, m$, and $c : [0,1]^m \to [0,\infty)$ is the density of C. In the discrete case,

the probability mass function $g: \mathbb{R}^m \to [0,1]$ of (Y_1, \ldots, Y_m) is given by

$$g(y_1, \dots, y_m) = c(G_1(y_1), \dots, G_m(y_m)) \quad \text{for all } y_1 \in R_{Y_1}, \dots, y_m \in R_{Y_m},$$
(3)

where R_{Y_i} denotes the support of Y_i for i = 1, ..., m. Here, $c : R_{U_1} \times ... \times R_{U_m} \to [0, \infty)$ is the probability mass function of the transformed random vector $(U_1, ..., U_m)$ which is given by

$$c(u_1, \dots, u_m) = \sum_{j_1 \in \{0,1\}} \dots \sum_{j_m \in \{0,1\}} (-1)^{j_1 + \dots + j_m} C(u_1^{(j_1)}, \dots, u_m^{(j_m)})$$
(4)

for all $(u_1, \ldots, u_m) \in R_{U_1} \times \ldots \times R_{U_m}$, where $u_i^{(0)} = u_i$ and $u_i^{(1)} = \lim_{x \uparrow G_i^{-1}(u_i)} G_i(x)$ for $i = 1, \ldots, m$.

Thus, in order to obtain g for a dataset $(y_{1,t_1}, \ldots, y_{m,t_1}), \ldots, (y_{1,t_k}, \ldots, y_{m,t_k}) \in \mathbb{R}^m$, we first determine the univariate cumulative distribution functions $G_1, \ldots, G_m : \mathbb{R} \to [0,1]$ for each location in $v_1, \ldots, v_m \in V$. Then, in a second step, the dataset is transformed into standard uniformly distributed samples

 $(u_{1,t_1},\ldots,u_{m,t_1}),\ldots,(u_{1,t_k},\ldots,u_{m,t_k}) \in [0,1]^m$ based on which a copula model can be fitted. The literature uses many different types of parametric copula families to model multivariate distributions. However, many of these families are limited to the two-dimensional case and involve only a small number of parameters, making it difficult to accurately capture the complexity of multivariate distributions.

For this reason, we utilize so-called R-vine copulas to model the copula C of the multivariate distribution of precipitation amounts, which are more suitable for high-dimensional random vectors. More specifically, vine copulas decompose a multivariate distribution into many bivariate distributions, which are easier to model. In addition, the number of bivariate copulas within the vine copula increases quadratically with the number of dimensions of the random vector, i.e., vine copulas are less restrictive than other parametric copula families with regard to the number of degrees of freedom. There are different vine copulas, with the R-vine copula being the most general type that includes all others. Other common types include the D-vine and C-vine copulas, which make additional assumptions that restrict the flexibility of the vine structure. Therefore, this paper only considers R-vine copulas.

In the following, we briefly explain the components of a vine copula structure, using Figure 1 as an example. This figure shows the structure of an R-vine copula for a 5-dimensional random vector. A vine copula comprises several trees, whereby the edges of a tree correspond to the vertices of the subsequent tree, as indicated by the vertical dotted lines. The vertices of the first tree, displayed at the bottom of Figure 1, correspond to the marginal distributions of the considered random vector. The vertices of the second tree, located above the first, correspond to bivariate copulas that model the joint distributions of the two marginal distributions connected by the corresponding edge. The vertices in subsequent trees correspond to bivariate copulas that model conditional joint distributions for two dimensions each (indicated by the two numbers on the left within each box), conditioned on one or more other dimensions (indicated by the numbers on the right within each box). The arrangement of edges in each tree is selected to capture most of the correlation within the multivariate random vector in the earlier trees. The reasoning is that the fitted copulas become less precise with each additional tree, and maximizing the correlation modelled in the earlier trees improves the accuracy of the vine copula as a whole.

When constructing a vine copula, the trees are built step by step, starting with the first. The first tree involves fitting the marginal distributions from a dataset of multivariate samples of the considered multivariate random vector with the help of common approaches such as parametric probability distributions, kernel density estimators or empirical CDFs. These fitted marginal distributions are then used to transform the samples within the dataset into standard uniform samples, which are used to fit the bivariate copulas in the second tree. Either parametric copula families or a kernel density estimator are used for this. The previous tree's elements transform the dataset into the required samples to fit the copulas in each subsequent tree.

While it is outside of the scope of this present paper to discuss vine copulas in more detail, it should be mentioned that the vine structure is specifically built in such a way that once the bivariate copulas within the vine structure are fitted, they can be utilized to obtain the vine copulas CDF or PDF or to generate samples. For a more detailed description regarding the fitting procedure of Rvine copulas to data, we refer to Joe and Kurowicka (2011), see also Aigner et al. (2023).



Figure 1: Exemplary structure of an R-vine copula for a 5-dimensional distribution. Each vertex in the bottom tree represents a marginal distribution, while the vertices in the second tree represent joint bivariate distributions. In the third and following trees, each vertex represents a joint conditional distribution for two components conditioned on one or more other components. Once fitted to data, these bivariate copulas can generate samples from the fitted copula or obtain functions such as its density. Note that this example was previously depicted in Aigner et al. (2023).

3.4 Generation of synthetic ensemble members

From now on, we assume that all random variables, i.e., $X_1, \ldots, X_m, Y_1, \ldots, Y_m$, and X'_1, \ldots, X'_m , as well as the transformed random variables U_1, \ldots, U_m , considered in the following, have discrete distributions.

We show how the samples (x_i^1, \ldots, x_i^N) for $i = 1, \ldots, m$, which have been independently drawn from the components of the random vector (X_1, \ldots, X_m) as stated above, can be rearranged into N synthetic ensemble members $(x_1^{(j)}, \ldots, x_m^{(j)})$ for $j \in \{1, \ldots, N\}$, such that these ensemble members can be considered as realizations of some *m*-dimensional random vector (X'_1, \ldots, X'_m) , whose univariate marginal distributions coincide with those of (X_1, \ldots, X_m) and, in addition, which exhibits the spatial correlations observed in historical precipitation data.

For this purpose, i.e., to rearrange the samples $\{(x_i^1, \ldots, x_i^N), i = 1, \ldots, m\}$ into N synthetic ensemble members $\{(x_1^{(j)}, \ldots, x_m^{(j)}), j = 1, \ldots, N\}$, we determine a permutation $\sigma_i : \{1, \ldots, N\} \rightarrow \{1, \ldots, N\}$ for each $i = 1, \ldots, m$ with $(x_i^{\sigma_i^{-1}(1)}, \ldots, x_i^{\sigma_i^{-1}(N)}) = (x_i^{(1)}, \ldots, x_i^{(N)})$ such that the sample of ensemble members $\{(x_1^{(j)}, \ldots, x_m^{(j)}), j = 1, \ldots, N\}$ exhibits the spatial correlation of precipitation amounts modeled by the copula $C : R_{U_1} \times \ldots \times R_{U_M} \rightarrow [0, 1]$ introduced above. For this, we define the likelihood $L(\sigma)$ for a set of permutations $\sigma = \{\sigma_i, i = 1, ..., m\}$ as

$$L(\sigma) = \prod_{j=1}^{N} g(x_1^{(j)}, \dots, x_m^{(j)}),$$
(5)

where $g: R_{Y_1} \times \ldots \times R_{Y_m} \to [0, 1]$ is the probability mass function of the discrete random vector (Y_1, \ldots, Y_m) fitted to historical precipitation data.

Thus, a set of permutations σ , which maximizes the likelihood $L(\sigma)$, rearranges the samples $\{(x_1^1, \ldots, x_i^N), i = 1, \ldots, m\}$ into N synthetic ensemble members $\{(x_1^{(j)}, \ldots, x_m^{(j)}), j = 1, \ldots, N\}$ such that they match the spatial correlation of precipitation amounts in the best possible way.

3.5 Predicting the total precipitation amount in a given area

Recall that the *i*-th component X_i of the random vector (X_1, \ldots, X_m) introduced above has the same (univariate) distribution as the *i*-th component X'_i of (X'_1, \ldots, X'_m) for each $i \in \{1, \ldots, m\}$. However, in addition to this, the joint distribution of the random vector (X'_1, \ldots, X'_m) obtained by maximizing the likelihood $L(\sigma)$ given in Equation (5), also captures the spatial correlation of precipitation amounts. Thus, for any subset $S \subset \{1, \ldots, m\}$, the random sum $\sum_{i \in S} X'_i$ can be considered as an appropriate prediction model for the total precipitation amount in an area which is represented by the set $\{v_i, i \in S\} \subset V$ of locations. In particular, the probability $\mathbb{P}(\sum_{i \in S} X'_i \geq z)$ that the total precipitation amount $\sum_{i \in S} X'_i$ exceeds a certain (critical) threshold z > 0 can be estimated by the relative frequency $\#\{j: 1 \leq j \leq N, \sum_{v_i \in S} x'_i > z\}/N$, where # denotes cardinality.

4 Implementation of the copula-based model

We now discuss some implementation details of the copula-based model introduced in Section 3. First, we explain for which areas the proposed model generates synthetic ensembles, which are validated in Section 5 below. Next, the procedure for fitting an R-vine copula to historical data is explained in detail. Then, a hill-climbing algorithm is presented for optimizing the likelihood function $L(\sigma)$ introduced in Section 3. Finally, we discuss the validation scheme utilized for evaluating the results stated in Section 5.

Note that so far in this paper, the term "location" has been used because, in general, the model introduced in Section 3 does not assume that the underlying data is given on a regular grid, i.e., the set $V = \{v_1, \ldots, v_m\} \subset \mathbb{R}^2$ introduced in Section 3 can be arbitrarily shaped. However, in our case, the data described in Section 2 is arranged on a regular grid, and, therefore, we will refer to grid points (or grid boxes) in the following instead of calling them locations. Moreover, instead of considering one single set V of grid points, in the following, we apply

the copula-based model introduced in Section 3 to several sets of grid points simultaneously, i.e., we assume that the historical observations of precipitation amounts are statistically invariant in space and time. This means, in particular, that we fit one single R-vine copula to multiple precipitation observations from across the entire considered period and from non-overlapping subsets of some sampling window $W \subset \mathbb{R}^2$, respectively.

4.1 Areas, where synthetic ensembles are generated

In order to evaluate the copula-based model proposed in the present paper, it is applied to the combined forecast produced by the C³-model, which receives forecasts of "STEPS-DWD" and "ICON-D2-EPS" as input and predicts calibrated marginal distributions for all considered grid points without taking spatial correlations into account. This combined forecast is available for a sampling window $W \subset \mathbb{R}^2$, consisting of 350×450 grid points and a rectangular subset of the regular COSMO-D2 grid, enclosing Germany and parts of neighbouring countries.

However, due to computational complexity, applying the copula-based model to the whole sampling window W in a single step is impossible. So instead, for each hour t available in the dataset, which belongs to the time intervals listed in the introduction of Section 2, we successively choose $n_t > 1$ non-overlapping quadratic subsets $V_1^t, V_2^t, \ldots, V_{n_t}^t \subset W$, consisting of 9×9 grid points each and being positioned at random within W, until no further non-overlapping 9×9 subset can be found in W.

4.2 Fitting R-vine copulas to historical observations

For a given hour t_c representing the "current time", we fit an R-vine copula to the historical observations made within the areas $V_1^t, V_2^t, \ldots, V_{n_t}^t \subset W$ and for the k past hours $t \in \{t_1, \ldots, t_k\}$ introduced in Section 3 with $t \leq t_c$, i.e., the Rvine copula is fitted to the vector data of precipitation amounts observed within the sets of 9×9 grid points described above, without taking local peculiarities into account. In other words, since the areas $V_1^t, V_2^t, \ldots, V_{n_t}^t$ are selected at random from all parts of the sampling window W, the fitted R-vine copula does not model local correlations which might be specific to a particular 9×9 area. This has the advantage that rare events, which might occur only at a few grid points within the entire dataset, do not influence how spatial correlation is modelled globally in W.

Note that the dataset on which the R-vine copula is fitted consists of observations for different past hours $t \in \{t_1, \ldots, t_k\}$ and different 9×9 areas $V_1^t, \ldots, V_{n_t}^t \subset W$. However, for fitting the copula, this dataset is considered as realizations of one single random vector (Y_1, \ldots, Y_m) for one single (abstract) 9×9 area $V = \{v_1, \ldots, v_m\}$ with m = 81, as introduced in Section 3. Therefore, in the following, $V = \{v_1, \ldots, v_m\}$ does not refer to one specific area within W, but to an unspecified quadratic set of m = 81 grid points and their relative positions to each other for which the random vector (Y_1, \ldots, Y_m) is defined. To fit the R-vine copula, the library pyvinecopulib Vinecopulib (2023) is used, where pyvinecopulib requires that the univariate (marginal) distributions of Y_i for $i \in \{1, \ldots, m\}$ fitted to the historical precipitation data are either continuous or discrete. Thus, since the distribution of precipitation amounts has an atom at 0 mm, we consider a discretized marginal distribution for each $i \in \{1, \ldots, m\}$, represented by its cumulative distribution function $G_i : R_{Y_i} \to [0, 1]$. Next, using G_i , we transform the precipitation amount y_i observed at $v_i \in V$ to obtain $u_i = G_i(y_i) \in [0, 1]$.

The R-vine copula is now fitted to the vectors (u_1, \ldots, u_m) of transformed precipitation amounts, where the fact is used that an R-vine copula is built by a set of bivariate copulas. Thus, in the fitting process, pyvinecopulib determines the most suitable copula family for each of these bivariate copulas with the help of the Bayesian information criterion (BIC). For the results derived in this paper, the bivariate copulas are chosen from all available parametric copula families provided by pyvinecopulib (independent, Gaussian, Student, Clayton, Gumbel, Frank, Joe, BB1, BB6, BB7, BB8). However, to reduce computational costs, we truncate the R-vine copula considering only five trees within the vine structure, i.e., for each of the m = 81 arguments of the R-vine copula, the correlation structure with five other suitably chosen arguments is considered. For more details regarding truncated vine copulas and vine copulas in general, we refer to Joe and Kurowicka (2011). Exemplary samples from the fitted R-vine copula can be compared to a set of randomly chosen observations in Figure 9 in the Appendix.

4.3 Generating samples of predicted precipitation amounts

In the following, we describe how to generate a sample (x_i^1, \ldots, x_i^N) of predicted precipitation amounts for an hour $t > t_c$ and for each of the m = 81 grid points of each 9×9 area $V_1^t, \ldots, V_{n_t}^t \subset W$, where $i \in \{1, \ldots, m\}$. As described in Section 3, the C³-model has been modified to generate quan-

tiles of the predicted distributions of precipitation amounts for each grid point in W. For our purpose, the C^3 -model has been trained to generate vectors $q = (q_1, \ldots, q_{100})$ of 100 α -quantiles, where the values of α are evenly spaced between 0.0001 and 0.9999. Then, for each grid point in $V_1^t \cup \ldots \cup V_{n_*}^t$, we select N = 20 values at random among the components of the corresponding vector $q = (q_1, \ldots, q_{100})$. To ensure that these values are spread out across the entire support of the predicted precipitation distribution, the components of $q = (q_1, \ldots, q_{100})$ are divided into N = 20 consecutive groups $(q_1, \ldots, q_5), (q_6, \ldots, q_{10}), \ldots, (q_{91}, \ldots, q_{95}), (q_{96}, \ldots, q_{100})$, each consisting of five quantiles from which one quantile is selected at random. Note that this procedure is similar to the stratified sampling approach discussed in Hu et al. (2016). The stratified sampling approach guarantees that the drawn samples cover the entire density for each grid point. This approach also ensures that the samples for grid points with similar densities have similar values. This prevents scenarios where, e.g. two neighbouring grid points have similar densities but vastly different samples because one grid point received all samples from the upper tail and the other from the lower tail. In the following, the matrix of the N quantiles drawn from q for each of the m = 81 grid points within a given 9×9 area will be denoted by $x = (x_i^j, i \in \{1, \ldots, m\}, j \in \{1, \ldots, N\}) \in \mathbb{R}^{m \times N}$.

4.4 Hill climbing algorithm for ensemble generation

To find a set of permutations $\sigma = \{\sigma_i, i = 1, ..., m\}$ that maximizes the likelihood $L(\sigma)$ introduced in Equation (5), a hill climbing algorithm (Skiena, 1998) is applied. This algorithm starts with a set of random permutations $\sigma = \{\sigma_i, i = 1, ..., m\}$, i.e., for each grid point $i \in \{1, ..., m\}$ the quantiles in $(x_i^1, ..., x_i^N)$ are rearranged at random. These arrangements are iteratively changed such that the value of the evaluation function $L(\sigma)$ increases with each step, where the algorithm is structured as follows, see also Algorithm 1 in the Appendix. :

- 1. The inputs of the algorithm are the matrix $x = (x_i^j, i \in \{1, ..., m\}, j \in \{1, ..., N\}) \in \mathbb{R}^{m \times N}$, which contains N predictions for each of the m grid points, and the *m*-variate density function $g : R_Y \to [0, \infty)$ introduced in Section 3.
- 2. As already mentioned above, the algorithm starts with a set of random permutations $\sigma = \{\sigma_i, i = 1, ..., m\}$, i.e., there is no (spatial) correlation between the rows $(x_i^1, ..., x_i^N)$ of x after applying the set σ of random permutations.
- 3. Iterate over each synthetic ensemble member (x¹₁,...,x^k_m) for j ∈ {1,...,N}:
 (a) Find k ∈ {j,...,N} such that g(x^j₁,...,x^k_p,...,x^j_m) is maximized for each grid point p ∈ {1,...,m}. If k ≠ j switch x^j_p and x^k_p in order to improve g(x^j). (b) Repeat step (a) until no values in x have been switched.
- 4. Return x. Note that at this step of the algorithm, the variable x contains the ordered values that were referred to as $\{(x_1^{(j)}, \ldots, x_m^{(j)}), j = 1, \ldots, N\}$.

The idea of reordering samples drawn from marginal distributions in order to obtain realistic ensemble members is also used in the approaches of the "Schaake shuffle" (Clark et al., 2004) and "ensemble copula coupling" (Wilks, 2006). However, in those approaches, the permutations for each grid point are provided either by a given set of ensemble members or directly by a set of historical observations. In contrast, the approach considered in the present paper is based on a fitted copula model.

For the results stated in this paper, N is chosen to be 20; however, the last member is omitted since the number of values from which the algorithm can choose at a grid point decreases as $j \in \{1, ..., N\}$ increases, the quality of the synthetic members, therefore, decreases for higher values of j. The decrease in quality is depicted in Figure 2, where it is shown that this effect mainly affects the last member j = N.

4.5 Validation scheme

The copula model proposed in this paper is validated with the help of a rollingorigin scheme (Armstrong and Grohman, 1972). In this scheme, the available dataset is split by the "current time" t_c , into a "past" and a "future". Throughout the validation, the "current time" t_c is incrementally shifted from the start to the end of the dataset in chronological order. In each step, the model is updated on the "past" data while the "future" data is used to validate model predictions. This approach is especially suitable for datasets with a time axis and also because it closely simulates operational conditions.

Furthermore, in each step, the model is applied to randomly chosen, nonoverlapping areas of size 9×9 , as described at the beginning of this section. Each application results in 19 synthetic ensemble members. Based on these synthetic ensemble members, the probability is estimated that the total precipitation amount within the area exceeds a given threshold. The resulting validation scores are discussed in the following section.



Figure 2: Relative quality of all synthetic ensemble members $(x_1^{(i)}, \ldots, x_m^{(i)})$ evaluated in the dataset (~ 50000 in total) compared to the first member $(x_1^{(1)}, \ldots, x_m^{(1)})$ within the same 9×9 area for the lead time +3 h. The relative quality is expressed by the distribution of the quotient $\log(g(x_1^{(i)}, \ldots, x_m^{(i)}))/\log(g(x_1^{(1)}, \ldots, x_m^{(1)}))$ for $i = 2, \ldots, 20$ with 90% winsorization. Note that the distributions of the relative qualities looks essentially the same for other lead times. In other words, generating synthetic ensemble members one after another reduces the number of values to choose from within each iteration of the hill-climbing algorithm. This effect is most noticeable for the last member for which only one value per grid point is left, i.e., the last ensemble member contains all values not used in previous synthetic members. This figure illustrates the quality drop-off that occurs for ensemble members for which the hill climbing algorithm has fewer values to choose from.

5 Validation of synthetic rainfall fields

We want to assess how well the copula-based technique proposed in this paper, referred to as "COPULA" in the following, models realistic spatial dependencies and to which extent the original requirements of a consistent and calibrated combination forecast are retained. To achieve this, we compare synthetic rainfall fields generated by "COPULA" with three groups of benchmarks. First, we utilize both input forecasts, "ICON" and "STEPS". Since both serve as input for the C^3 -model, forecasts based on the output of the C^3 -model are expected to be at least as good as either input forecast.

Secondly, we examine extreme cases of spatial correlation. The minimum spatial correlation is realized by randomly arranging the values drawn from the C^3 -model at each grid point as described in Section 4. Conversely, the maximum possible spatial correlation is obtained when the sample values are sorted in ascending order; consequently, the first synthetic ensemble member comprises the lowest values at each grid point, followed by the second lowest ones, and so forth until the last ensemble member containing the highest values for every grid point. However, we only discuss the sorted arrangement since the random arrangement reveals poor results across all considered metrics. The sorted arrangement is named "C³ sorted" in this context.

Lastly, we consider post-processed values drawn from the C³-model arranged according to the ranks of one input ensemble forecast each using the "Schaake shuffle" to reintroduce a dependence structure. These configurations are called "C³+ICON" and "C³+STEPS". Here, the "Schaake shuffle" is applied to the C³-model output and the forecasts of "ICON-D2-EPS" and "STEPS-DWD", respectively, are used as source of spatial correlation. Due to the atom at 0 mm in the distribution of precipitation amounts, there may not always be a fully specified ranking provided by the considered ensemble forecast. In such cases where ties occur within the ranking, they are resolved randomly. Furthermore, it should be noted that "STEPS" consists of 30 ensemble members, while we draw only 20 values for each grid point from the distributions provided by the C³-model. Consequently, to apply the "Schaake shuffle", out of these 30 ensemble members from "STEPS", 20 are selected randomly as a source of spatial correlation.

For the evaluation of the core requirements, we first computed traditional metrics such as bias, Brier skill score and reliability for approximately 50000 (sub-) regions $V \subset W$ that we have been drawn at random for each considered hour within our three months period and for all lead times, see Section 2. Thus, the dataset consists of 1761 forecast hours for each lead time, where the average number of evaluated sub-regions per forecast hour is equal to 28, 4. These traditional metrics mentioned above have been computed for threshold exceedance probabilities to get easily interpretable information about the systematic model error (bias), the forecast quality regarding both model and forecast error (Brier skill score), and the conditional frequency bias (reliability).

Moreover, to get a better insight into the spatial structure of the synthetic rainfall fields and their realistic appearance, we identified objects as sets of connected grid points, in which precipitation occurs, by a standard segmentation method using 4-connectivity and, furthermore, evaluated the parameters of a fitted exponential variogram model. Based on the identified objects, we computed their number, overall area, scaled volume, and weighted centre distance. Note that the latter two metrics were initially introduced in Wernli et al. (2008). In addition, we assessed the model performance by the aggregation metrics D_0 and D_1 , which were introduced in Tobin et al. (2012). For a discussion of these



object-based metrics, see also Rempel et al. (2017).

Figure 3: Observed precipitation amounts and realizations of 1 h forecasts of the six considered models ("ICON", "STEPS", "COPULA", "C³ sorted", "C³+ICON", "C³+STEPS") for a randomly chosen subset of 9×9 grid points. The colour scale is given in millimetres. Precipitation amounts above the upper end of the colour scale are shown in black.

5.1 Visual Inspection

Figure 3 depicts an exemplary observation at the top for a randomly chosen sub-region $V \subset W$ and forecast hour, together with the ensemble members of the NWP forecast ("ICON") and the precipitation nowcasting ("STEPS"). Note that "ICON" provides 20 ensemble members, while "STEPS" has 30. The synthetic ensemble members resulting from the copula model ("COPULA") with restored spatial dependencies are illustrated as a fourth category. As discussed in the previous section and illustrated in Figure 2, the quality of the last synthetic ensemble member generated by the hill climbing algorithm differs vastly from previous iterations and is therefore omitted. Thus, "COPULA" shows only 19 ensemble members. Finally, the three benchmarks "C³ sorted", "C³+ICON" and "C³+STEPS" are shown. Recall that "C³+ICON" and "C³+STEPS" refer to the "Schaake shuffle" applied to the output of the C³-model, where each input forecast is used as the source for spatial correlation, respectively. For each ensemble, the members illustrated in Figure 3 are sorted in ascending order of the total amount of precipitation within the considered area.

It can be seen that each ensemble forecast exhibits different spatial correlation patterns. "ICON" ensemble members appear relatively smooth, while the ensemble members of "STEPS" show a high level of variation between neighbouring grid points. This is because the effective resolution of "ICON" is lower than the resolution of the considered grid, while the native grid of "STEPS" has an even higher resolution, as discussed in Section 2. When considering the ensemble members of "COPULA", we see that the proposed model reintroduced spatial features, i.e., values at neighbouring grid points are more similar compared to grid points further apart. When comparing the "COPULA" ensemble and " C^3 sorted", we see that the sorted version is much smoother since it is the arrangement with the highest possible correlation between grid points. Additionally, it is noticeable that both forecasts resulting from the "Schaake shuffle" display a comparable spatial structure to their corresponding input forecasts; specifically, "C³+ICON" appears smoother while "C³+STEPS" exhibits a more contoured field. Furthermore, similar structures, such as the diagonal lines observed in "ICON" and " C^3 +ICON", are evident. Note that the precipitation amounts for the ensemble members generated by "COPULA", "C³ sorted", " C^3 +ICON" and " C^3 +STEPS" are provided by the C^3 -model and differ only in their arrangement to synthetic ensemble members.

5.2 Traditional metrics

The results depicted in Figure 4 for bias (left), Brier skill score (centre), and reliability (right) for thresholds from 0.62 mm up to 3.7 mm are shown for "COP-ULA" (green), "C³ sorted" (light green), "C³+ICON" (cyan), "C³+STEPS" (red), "ICON" (blue) and "STEPS" (orange). For lower threshold values of 0.62 mm and 1.23 mm, the systematic error in area probabilities derived from "COPULA" and "C³ sorted" is higher than for the other considered model outputs. Conversely, at higher thresholds of 2.47 mm and 3.70 mm, "COPULA" and "C³ sorted" demonstrate improved bias compared to all other models.

Regarding the Brier skill score depicted in the centre column of Figure 4, all models based on samples derived from the C^3 -model exhibit comparable forecast quality with a smooth transition from "STEPS" at +1h to "ICON" +4h. However, enhancements over "ICON" are negligible for lower thresholds at later lead times.

The reliability depicted in the right column of Figure 4 indicates that mod-

els based on samples obtained from the C³-model exhibit improved reliability compared to both input forecasts. Specifically, "C³+ICON" is the most reliable for a threshold value of 0.62 mm; however, "COPULA" and "C³ sorted" appear to improve as thresholds increase.

Comparing these results for area probabilities with those from the grid point perspective in Rempel et al. (2022), it can be seen that there is no large decline in the results based on a traditional forecast verification. That indicates that our core requirements of a calibrated and consistent combination are still fulfilled and that the calibration of marginal distributions also improves the prediction of area probabilities.



Figure 4: Bias, Brier skill score and reliability for threshold exceedance probabilities predicted by the models "COPULA" (green), "C³ sorted" (light green), "C³+ICON" (cyan), "C³+STEPS" (red), "ICON" (blue) and "STEPS" (orange).

To further demonstrate the ensemble calibration of area probabilities, Figure 5 depicts reliability diagrams for every forecast lead time for a threshold of 2.47 mm. These diagrams allow for a more detailed insight into the reliability as the scores depicted in Figure 4. Here, the curves for "COPULA" and "C³ sorted" are close to the leading diagonal, suggesting that the calibration performed by the C³-model (Rempel et al., 2022) also enhances the calibration of area forecasts. Conversely, both models resulting from the "Schaake shuffle" exhibit less calibrated reliability diagrams; however, they still demonstrate improved performance compared to their respective input forecasts. Furthermore, the occurrence of probabilities close to 1 decreases with increasing forecast uncertainty, i.e., for increasing thresholds and longer lead times. This decrease is indicated by an increasing truncation of the curves. Compared with this, the input forecasts of "ICON" and "STEPS" reveal a lead-time invariant and increasing overforecast, respectively.



Figure 5: Reliability (top) and frequency diagrams (bottom) for threshold exceedance probabilities (2.47 mm) predicted by the models "COPULA" (green), "C³ sorted" (light green), "C³+ICON" (cyan), "C³+STEPS" (red), "ICON" (blue) and "STEPS" (orange).

We want to assess not only the forecast quality of exceedance probabilities in terms of bias, Brier skill score, and reliability, but also the statistical behavior of predicted precipitation amounts. To achieve this, probability integral transform (PIT) diagrams (Czado et al., 2009) for areal precipitation amounts predicted by the three models "STEPS", "ICON", "COPULA", and the considered benchmarks "C³ sorted", "C³+ICON", and "C³+STEPS" for various lead times are presented in Figure 6. In such diagrams, the horizontal axes are defined as the normalized value ranges of the forecasts, whereas the numbers of bins are equal to the ensemble sizes. The actually occurred precipitation amounts are sorted into the respective ensemble member's bin with an equal or higher rainfall amount forecast. Thus, events in the first (last) bin represent observed rainfall amounts below (above) the forecast value range.

Note that each bin of a PIT diagram corresponds to an interval between two ensemble members for a given metric. In cases where the considered metric is identical for two or more ensemble members, the corresponding intervals have length zero and observations with the same metric cannot be unambiguously assigned to one bin. When considering metrics like precipitation amounts, this is a common occurrence due to the atom in the distribution at 0 mm. To mend this, we use the approach presented in Equation (1) in Czado et al. (2009), where such observations are randomly assigned to one of the bins in question.

For both raw input ensemble forecasts, many observed precipitation amounts are in the first and last bin, respectively, revealing an underdispersive behaviour, i.e., the ensemble does not cover the whole range of observations. The overestimation at the lower percentile might be induced by non-precipitation cases but also by an overforecast of the precipitation amount, the latter, especially for "ICON" forecasts. However, it should be noted that this result may be sensitive to the QPE that we have used as observation. On the other hand, "STEPS" shows deficiencies with higher precipitation amounts. This may be induced by the cascade of autoregressive processes that reduces the maxima of intensity and cannot cover the range of growth/decay processes of precipitation.

With precipitation sums based on synthetic rainfall fields generated by "COP-ULA", we can at least cover the observed range of values at the upper bound. However, many cases with less precipitation are not covered. For this, also non-precipitation cases may play a role. Since at least a few values drawn from the upper tail of the probability distributions predicted by the C³-model are always positive, the total precipitation amount within synthetic ensemble members is often positive and not zero. This could be underpinned by the sorted arrangement depicted in the right column of Figure 6. In these diagrams, the sorted ensemble exhibits no outliers at both bounds due to a larger range of predicted area precipitation sums compared to "COPULA". This more extensive range is achieved because the sorted ensemble is the arrangement with the lowest and highest possible precipitation amount within one ensemble member each. However, the frequency of larger precipitation amounts is higher in the sorted arrangement, since most of the observed rainfall is below the median. Thus, the sorted arrangement achieves the best PIT diagrams compared to the other three models.

The "Schaake shuffle" applied to values drawn from the C^3 -model is illustrated in the two right columns of Figure 6 (a) and (b). In the unconditional case depicted in Figure 6a), the number of events where areal precipitation is overestimated is clearly higher for forecasts with the "Schaake shuffle" than for "COPULA" and "C³ sorted". This suggests a spatial mismatch between the post-processed precipitation distributions of the C³-model and the input forecasts' spatial dependence structures. It also suggests that the copula approach exhibits an improved dependence structure as it models spatial dependencies more accurately. This is further supported by the conditional case in Figure 6b), where at least one grid point exhibits precipitation compared to both "COPULA" and "C³ sorted". At the same time, "C³+STEPS" displays the underdispersive behavior characteristic of the original "STEPS" forecast. Additionally, one can observe a lead-time-dependent shift towards higher ensemble members in all considered models.



Figure 6: Probability integral transform (PIT) diagrams for area precipitation amounts predicted by the three models "STEPS", "ICON", "COPULA", and the considered benchmarks "C³ sorted", "C³+ICON", and "C³+STEPS" for lead times of +1 h (top), +3 h (centre), and +6 h (bottom). Here, the values observed in different hours are assigned to bins in the histogram, whereby the borders of the bins are given by the values predicted by the ensemble members of the corresponding ensemble forecast, i.e., the binned events indicate whether the observations are below or equal to the prediction of the respective ensemble member. Thus, events in the first (last) bin represent observations below (above) the lowest (highest) forecast. The orange horizontal line is drawn at y = 1and marks the bar height required for the ideal PIT diagram, i.e., the uniform distribution. Note that the scales of the y-axes differ between the PIT diagrams.

5.3 Object-based metrics and estimated variogram models

To highlight the benefits of the restoration of spatial dependencies and to compare the resulting precipitation structures in each ensemble member with that of the radar observation, we consider PIT diagrams of various object-based metrics in Figure 7. As an object, we define a contiguous area in the sense of a 4-connectivity where grid points exceed an hourly precipitation sum of 0.1 mm. The metrics consist of the total area of all objects and their number. Furthermore, we consider two metrics of the SAL-index (Structure, Amplitude, Location; (Wernli et al., 2008)). First, the scaled volume provides information about the average shape of identified objects. It is given by the precipitation mass of an object normalized by its maximum and additionally weighted with this mass. Here, the precipitation mass of an object is defined as the sum of the rainfall amount at each associated grid point. Second, the weighted centre distance describes the average distance between single objects and the total centre of mass and provides information about the aggregation of precipitation. The averaging is also based on the precipitation mass to favour larger objects. In addition, we assess D_0 and D_1 , which are components of SCAI (Simple Convective Aggregation Index; (Tobin et al., 2012)). The metrics D_0 and D_1 represent the geometric and arithmetic mean of the distances for all possible pairs of objects.

The evaluation results for the three models "STEPS", "ICON", "COPULA",

and the considered benchmarks "C³ sorted", "C³+ICON", and "C³+STEPS", are divided into two groups. First, we consider results based on the whole dataset, depicted in the six left columns of Figure 7. Second, instances without any observed precipitation are removed for the six columns on the right. This implies that at least at one grid point within the 9×9 sub-region $V \subset W$, the observed hourly rainfall must be $\geq 0.1 \,\mathrm{mm}$. "ICON" and "STEPS" reveal a tendency to underestimate the respective values in all metrics since there is a peak of observations in the last bin. This peak is even more distinct for the conditional case on the right, indicating many cases in which no precipitation is forecast or observed. The peak may be induced by cases where no precipitation is forecast at all. One should be aware that we statistically evaluated subregions with an edge length of $\approx 20 \,\mathrm{km}$ so that forecast errors in location (e.g. spatial shifts in forecast precipitation) strongly influence the depicted results. In some cases, "ICON" and "STEPS" overestimate the area, whereby "STEPS" further overestimates the weighted centre distance. This may be attributable to situations where only one smaller object is identified in the observation. Since precipitation fields in "ICON" are smoother, one can assume that, in general, the number of objects is smaller. Therefore, if only one object is detected, the centre of masses is "equal to itself", and the weighted centre distance is zero. Compared to both input forecasts, models based on the output of the C³-model exhibit fewer outliers. Among these, the "COPULA" model produces almost perfect PIT diagrams followed by the "C³ sorted", consistently displaying a small peak in the upper end of the PIT diagram. While both benchmarks resulting from the "Schaake shuffle" show a downward slope accompanied by some outliers, with "C³+ICON" consistently outperforming "C³+STEPS". This slope underpins the lower spatial correlation structure compared to the "COP-ULA" results, since the frequency to have objects identified even in the first ensemble members is increased. Therefore, the observed null event is binned on one of the first members. In the conditional case on the right, the PIT diagrams are less homogeneous among the considered metrics; however, for all metrics, "COPULA" or "C³+ICON" achieves the best calibration.



Figure 7: PIT diagrams as in Figure 6 but for various object-based metrics. Topdown are the area and the number of objects, scaled volume and weighted centre distance of SAL, D_0 and D_1 of SCAI. Results are shown for "STEPS", "ICON", "COPULA", and the considered benchmarks "C³ sorted", "C³+ICON", and "C³+STEPS" for +1 h lead time. Results for other lead times are omitted since they are essentially identical. The six columns on the left-hand side show the results for the whole dataset, while the six columns on the right-hand side are restricted to instances where precipitation was observed within the considered 9×9 sub-region $V \subset W$. The orange horizontal line is drawn at y = 1 and marks the bar height required for the ideal PIT diagram, i.e., the uniform distribution. Note that the scales of the y-axes differ between the PIT diagrams.

The forecasts of "C³ sorted" also show an underestimation of the metrics. However, at least a few ensemble members are above the observed values since the frequency in bins of the upper percentiles is also increased. Furthermore, at least the maximum value of the "C³ sorted" ensemble forecast seems to reproduce the observed values since the maximum peak is in the second last bin. Compared to the raw drawn values, the forecasts of "COPULA" reveal a nearly uniform distribution for the whole dataset. For the conditional case, the distribution of observed values is still skewed. However, the frequency begins to increase at lower percentiles. This indicates that the forecasts of copula exhibit a more realistic representation of spatial structures in terms of such a set of object-based metrics.

As an additional technique to assess the spatial structure of synthetic rainfall fields, we fit an exponential variogram model (e.g. (Journel and Huijbregts, 1978)) and compare the estimated fitting parameters. These parameters, namely, nugget, sill, and effective range, are shown in Figure 8 for lead times of +1 h, +3 h, and +6 h. Again, we consider the whole dataset (six columns on

the left-hand side) and the conditional case (six columns on the right-hand side). Nugget describes the portion of non-spatial variance, whereas sill represents the limit of the variogram. The effective range is the distance where 95% of the sill is exceeded.

For the unconditional case, "ICON" and "STEPS" forecasts display an underdispersive behavior in all three metrics. In contrast, for models based on the C^3 -model output, " C^3 sorted" consistently achieves the most calibrated PIT diagrams followed by "COPULA", which exhibits a downward slope throughout. Notably, both "Schaake shuffle" benchmarks significantly underperform all other models. However, uncertainties in fitting the exponential variogram model could be a reason for this overestimation.

In the conditional case, variogram parameter-based metrics exhibit higher variability than object-based metrics. It can be seen that models based on the C^3 -model consistently outperform both input forecasts; however, none consistently perform best for all considered metrics and lead times.



Figure 8: PIT diagrams as in Figure 6 and Figure 7 but for various variogram parameters of an exponential variogram model, which are the nugget, effective range, and sill. Results are shown for "STEPS", "ICON", "COPULA", and the considered benchmarks "C³ sorted", "C³+ICON", and "C³+STEPS" for the lead times of +1 h, +3 h, and +6 h. The six columns on the left show the results for the whole dataset, while the six columns on the right are restricted to instances where precipitation was observed within the considered 9×9 sub-region V. The orange horizontal line is drawn at y = 1 and marks the bar height required for the ideal PIT diagram, i.e., the uniform distribution. Note that the scales of the y-axes differ between the PIT diagrams.

6 Conclusion

Many post-processing techniques for weather forecasts apply statistical corrections to a forecast individually for each location considered. Thus, the inherent spatial correlation of the input forecast is lost, and the statistically postprocessed output includes only marginal distributions. However, the spatial correlation between locations is required to predict quantities in larger areas, e.g., the amount of rainfall within a river basin. This necessitates the development of models that reintroduce the spatial correlation into such post-processed forecasts.

Several methods have been developed for this purpose, and they are usually based on rearrangement methods such as the "Schaake shuffle" (Clark et al., 2004) or "ensemble copula coupling" (Schefzik et al., 2013). However, for these techniques, a close relationship between the source of the spatial correlation and the post-processed marginal distributions is required. Furthermore, the ranks should be unambiguous to obtain realistic results. These conditions are fulfilled in applications where only one post-processed ensemble forecast is rearranged. Here, the spatial information is derived directly from a suitably selected set of historical observations or the origin ensemble forecast.

The conditions of a close relationship and unambiguous ranks are no longer met for seamless probabilistic precipitation forecasts as provided by the C^{3} model (Rempel et al., 2022). The output of this model represents an optimal combination of forecasts of the two ensemble prediction systems: "STEPS-DWD", a precipitation nowcasting scheme, and "ICON-D2-EPS", a short-range high-resolution NWP-model. Therefore, the output of the C³-model reveals a mixture of the distributions of the input forecasts. Furthermore, it is worth noting that the C³-model's output combines both input forecasts. As a result, the output may have discrepancies, such as spatial shifts, when compared to each input forecast separately. These shifts can be greater than the sub-regions considered in this study. Therefore, the input ensemble forecasts may not be appropriate for transferring the spatial correlation structure as a close relationship is not guaranteed.

To overcome these limitations, in the present paper, we propose and apply a model that uses an R-vine copula fitted to historical precipitation observations to model the joint distribution of precipitation amounts at adjacent locations. It does not require such a close relationship and can be applied in cases, such as the combination model C^3 , where the marginal output distributions differ significantly from each input ensemble forecast. In addition, inferring spatial correlations from a set of historical observations or another ensemble forecast requires a ranking without ties of the predicted events for each location, which is not always given for meteorological quantities such as precipitation, whose probability distribution has an atom at 0mm. Deriving spatial correlations from a joint probability distribution avoids this problem.

In order to evaluate the performance of the proposed model, we considered several validation metrics such as bias, Brier skill score and reliability diagrams for precipitation amounts within sub-regions of 9×9 grid points. The results were compared to both input forecasts as well as three benchmarks. The first benchmark is the sorted C³-model output having a maximum in spatial correlation. The other two are the "Schaake shuffle" applied to the C³-model forecasts using "STEPS-DWD" and "ICON-D2-EPS", respectively, as source for the spatial correlation structure. As a result, it could be shown that the calibration of the marginal distributions by the previously applied combination model carries over to all considered C³-based models, which exhibit a noticeable enhancement in forecast quality compared to both input forecasts. Although the "Schaake shuffle" benchmarks appear to perform better for lower precipitation thresholds,

the Copula model and "C³ sorted" achieve superior aggregated validation scores for higher thresholds.

The proposed method has been evaluated on a dataset that includes several parts from different years. Furthermore, observations for randomly chosen areas from the entire available sampling window were used to train the R-vine copula model. We assume that the historical observations of precipitation amounts are statistically invariant in space and time. Therefore, the R-vine copula models the multivariate climatological precipitation distribution, ensuring that the presented results are less likely to be outliers based on the choice of the considered time frame or locality and, hence, more robust.

Furthermore, the spatial correlation is directly evaluated using object-based metrics such as the number of connected area components, distance metrics between precipitation clusters, and variogram parameters. The object-based metrics indicate that the output from the proposed copula model exhibits the most realistic precipitation patterns compared to all other forecast models considered in this paper. Concerning variogram-based metrics, "C³ sorted" demonstrated superior performance. Even if the benchmarks employing the "Schaake shuffle" are, on average, comparable with "COPULA" in terms of the traditional gridpoint-wise evaluation metrics, they did not achieve the same level of calibration and clearly underperform all models in terms of variogram-based metrics.

The present paper evaluates the performance of the copula model for areas consisting of 9×9 grid points on real-world data. In addition, similar to the approach presented in Lerch et al. (2020), further investigation into the behaviour of our proposed approach concerning its spatial dependence structure could be carried out through an analysis of synthetic scenarios. Furthermore, to apply the model to more general cases, such as river catchments, the model requires some modifications, which will be investigated in a forthcoming study. Such a modification could be the extension of the R-vine copula model to higher dimensions in order to include more grid points. This, however, would lead to increased algorithmic complexity and might not be directly feasible, depending on the size of the area. To take this into account, another modification could be the adaption of the hill climbing algorithm such that the existing copula for 9×9 areas is used to determine the permutation of values at a given grid point, i.e., for determining a permutations σ_i only the local neighbourhood is considered. Besides considering more general areas, including additional information into the R-vine copula fitting procedure would be interesting, e.g., the local orography or the distinction of convective and stratiform precipitation patterns, which could further improve the reconstruction of spatial correlation.

7 Statements

Data availability statement: The data used in this study is experimental and is not yet available for public access. New forecasts generated by a further improved NWP configuration named "ICON-D2-RUC" as well as forecasts of "STEPS-DWD" will be publicly accessible at *opendata.dwd.de* by the end of 2024.

Funding statement: The authors did not receive any funding for this article's research, publication, or authorship.

8 Appendix

Algorithm 1 Synthetic Ensemble Optimization with Hill Climbing

1: procedure HILL_CLIMBING $(x, g) \triangleright x \in \mathbb{R}^{m \times N}$ contains N predictions for m grid points $\triangleright q$ is the density function fitted in Section 3 2: 3: for $j \in \{1, ..., N\}$ do \triangleright Iterate over all synthetic ensemble members repeat 4: $c \gets \texttt{False}$ 5:for $p \in \{1, \ldots, m\}$ do \triangleright Iterate over all grid points (in random 6: order) $k \leftarrow \arg \max_{k \in \{j, \dots, N\}} g(x_1^j, \dots, x_p^k, \dots, x_m^j)$ \triangleright Find best 7: prediction for pif k > j then \triangleright true, if switching x_p^k with x_p^j improves $g(x^j)$ 8: $x_p^j \leftrightarrow x_p^k \ \triangleright$ Predictions in members j and k switch places 9: $c \leftarrow \text{True} \qquad \triangleright c = \text{True}$ indicates that a change has been 10:made to xend if 11:12:end for ▷ Stop if no improvement has been made for any grid 13:until $\neg c$ point in x^j end for 14: $\triangleright x$ contains the ordered values referred to as return x15: $\{(x_1^{(j)}, \dots, x_m^{(j)}), j = 1, \dots, N\}$ 16: end procedure



Figure 9: Left: Randomly chosen precipitation observations. Right: Random samples drawn from the R-vine copula model. Notes: Observations and samples without any precipitation were discarded. High values above the color scale are shown in black. Note that the copula models the general distribution observed in the historical observations irrespective of their location or time, i.e., the samples from the copula have no corresponding counterpart within the dataset of observations.

References

- Aigner, K.-M., Schaumann, P., Loeper, F. v., Martin, A., Schmidt, V., and Liers, F. (2023). Robust dc optimal power flow with modeling of solar power supply uncertainty via r-vine copulas. *Optimization and Engineering*, 24:1951–1982.
- Armstrong, J. S. and Grohman, M. C. (1972). A comparative study of methods for long-range market forecasting. *Management Science*, 19:211–221.
- Bellier, J., Zin, I., and Bontron, G. (2018). Generating coherent ensemble forecasts after hydrological postprocessing: Adaptations of ECC-based methods. *Water Resources Research*, 54:5741–5762.
- Bick, T., Simmer, C., Trömel, S., Wapler, K., Hendricks Franssen, H.-J., Stephan, K., Blahak, U., Schraff, C., Reich, H., Zeng, Y., and Potthast, R. (2016). Assimilation of 3D radar reflectivities with an ensemble Kalman filter on the convective scale. *Quarterly Journal of the Royal Meteorological Society*, 142:1490–1504.
- Bowler, N. E., Pierce, C. E., and Seed, A. W. (2006). STEPS: A probabilistic precipitation forecasting scheme which merges an extrapolation nowcast with downscaled NWP. *Quarterly Journal of the Royal Meteorological Society*, 132:2127–2155.

- Brunner, M. I., Furrer, R., and Favre, A.-C. (2019). Modeling the spatial dependence of floods using the Fisher copula. *Hydrology and Earth System Sciences*, 23:107–124.
- Callau Poduje, A. C. and Haberlandt, U. (2018). Spatio-temporal synthesis of continuous precipitation series using vine copulas. *Water*, 10:862.
- Clark, M., Gangopadhyay, S., Hay, L., Rajagopalan, B., and Wilby, R. (2004). The Schaake shuffle: A method for reconstructing space-time variability in forecasted precipitation and temperature fields. *Journal of Hydrometeorology*, 5:243–262.
- Czado, C., Gneiting, T., and Held, L. (2009). Predictive model assessment for count data. *Biometrics*, 65:1254–1261.
- Czado, C. and Nagler, T. (2021). Vine copula based modeling. Annual Review of Statistics and Its Application, 9:453–477.
- Erhardt, T. M., Czado, C., and Schepsmeier, U. (2015). R-vine models for spatial time series with an application to daily mean temperature. *Biometrics*, 71:323–332.
- Feldmann, K., Scheuerer, M., and Thorarinsdottir, T. L. (2015). Spatial postprocessing of ensemble forecasts for temperature using nonhomogeneous Gaussian regression. *Monthly Weather Review*, 143:955–971.
- Foresti, L., Reyniers, M., Seed, A., and Delobbe, L. (2016). Development and verification of a real-time stochastic precipitation nowcasting system for urban hydrology in Belgium. *Hydrology and Earth System Sciences*, 20:505.
- Hess, R. (2020). Statistical postprocessing of ensemble forecasts for severe weather at Deutscher Wetterdienst. Nonlinear Processes in Geophysics, 27:473–487.
- Hu, Y., Schmeits, M. J., Jan van Andel, S., Verkade, J. S., Xu, M., Solomatine, D. P., and Liang, Z. (2016). A stratified sampling approach for improved sampling from a calibrated ensemble forecast distribution. *Journal of Hydrometeorology*, 17:2405–2417.
- Jobst, D., Möller, A., and Groß, J. (2023). D-vine-copula-based postprocessing of wind speed ensemble forecasts. *Quarterly Journal of the Royal Meteoro*logical Society, 149:2575–2597.
- Joe, H. and Kurowicka, D. (2011). Dependence Modeling: Vine Copula Handbook. World Scientific.
- Journel, A. G. and Huijbregts, C. J. (1978). *Mining Geostatistics*. Academic Press.

- Kaltenberger, R., Schaffhauser, A., and Staudinger, M. (2020). "What the weather will do"–Results of a survey on impact-oriented and impact-based warnings in European NMHSs. Advances in Science and Research, 17:29–38.
- Kox, T., Kempf, H., Lüder, C., Hagedorn, R., and Gerhold, L. (2018). Towards user-orientated weather warnings. *International Journal of Disaster Risk Reduction*, 30:74–80.
- Lakatos, M., Lerch, S., Hemri, S., and Baran, S. (2023). Comparison of multivariate post-processing methods using global ECMWF ensemble forecasts. *Quarterly Journal of the Royal Meteorological Society*, 149:856–877.
- Lerch, S., Baran, S., Möller, A., Groß, J., Schefzik, R., Hemri, S., and Graeter, M. (2020). Simulation-Based Comparison of Multivariate Ensemble Post-Processing Methods. *Nonlinear Processes in Geophysics*, 27:349–371.
- Li, Y., Peng, X., and Zhang, Y. (2022). Forecasting methods for wind power scenarios of multiple wind farms based on spatio-temporal dependency structure. *Renewable Energy*, 201:950–960.
- Möller, A., Lenkoski, A., and Thorarinsdottir, T. L. (2013). Multivariate probabilistic forecasting using ensemble bayesian model averaging and copulas. *Quarterly Journal of the Royal Meteorological Society*, 139:982–991.
- Nazeri Tahroudi, M., Ramezani, Y., de Michele, C., and Mirabbasi, R. (2022). Multivariate analysis of rainfall and its deficiency signatures using vine copulas. *International Journal of Climatology*, 42:2005–2018.
- Potter, S., Harrison, S., and Kreft, P. (2021). The benefits and challenges of implementing impact-based severe weather warning systems: Perspectives of weather, flood, and emergency management personnel. Weather, Climate, and Society, 13:303–314.
- Pulkkinen, S., Chandrasekar, V., von Lerber, A., and Harri, A.-M. (2020). Nowcasting of convective rainfall using volumetric radar observations. *IEEE Transactions on Geoscience and Remote Sensing*, 58:7845–7859.
- Reinoso-Rondinel, R., Rempel, M., Schultze, M., and Trömel, S. (2022). Nationwide radar-based precipitation nowcasting - A localization filtering approach and its application for Germany. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15:1670–1691.
- Rempel, M., Schaumann, P., Hess, R., Schmidt, V., and Blahak, U. (2022). Adaptive blending of probabilistic precipitation forecasts with emphasis on calibration and temporal forecast consistency. *Artificial Intelligence for the Earth Systems*, 1:e20020.
- Rempel, M., Senf, F., and Deneke, H. (2017). Object-based metrics for forecast verification of convective development with geostationary satellite data. *Monthly Weather Review*, 145:3161–3178.

- Ruti, P. M., Tarasova, O., Keller, J. H., Carmichael, G., Hov, Ø., Jones, S. C., Terblanche, D., Anderson-Lefale, C., Barros, A. P., Bauer, P., Bouchet, V., Brasseur, G., Brunet, G., DeCola, P., Dike, V., Kane, M. D., Gan, C., Gurney, K. R., Hamburg, S., Hazeleger, W., Jean, M., Johnston, D., Lewis, A., Li, P., Liang, X., Lucarini, V., Lynch, A., Manaenkova, E., Jae-Cheol, N., Ohtake, S., Pinardi, N., Polcher, J., Ritchie, E., Sakya, A. E., Saulo, C., Singhee, A., Sopaheluwakan, A., Steiner, A., Thorpe, A., and Yamaji, M. (2020). Advancing research for seamless Earth system prediction. *Bulletin of the American Meteorological Society*, 101:E23–E35.
- Schaumann, P., Hess, R., Rempel, M., Blahak, U., and Schmidt, V. (2021). A calibrated and consistent combination of probabilistic forecasts for the exceedance of several precipitation thresholds using neural networks. *Weather* and Forecasting, 36:1076–1096.
- Schefzik, R. (2016). A similarity-based implementation of the Schaake shuffle. Monthly Weather Review, 144:1909–1921.
- Schefzik, R., Thorarinsdottir, T. L., and Gneiting, T. (2013). Uncertainty quantification in complex simulation models using ensemble copula coupling. *Statistical Science*, 28:616–640.
- Schinke-Nendza, A., von Loeper, F., Osinski, P., Schaumann, P., Schmidt, V., and Weber, C. (2021). Probabilistic forecasting of photovoltaic power supply—A hybrid approach using D-vine copulas to model spatial dependencies. *Applied Energy*, 304:117599.
- Schraff, C., Reich, H., Rhodin, A., Schomburg, A., Stephan, K., Perianez, A., and Potthast, R. (2016). Kilometre-scale ensemble data assimilation for the COSMO model (KENDA). Quarterly Journal of the Royal Meteorological Society, 142:1453–1472.
- Seed, A. W. (2003). A dynamic and spatial scaling approach to advection forecasting. *Journal of Applied Meteorology*, 42:381–388.
- Seed, A. W., Pierce, C. E., and Norman, K. (2013). Formulation and evaluation of a scale decomposition-based stochastic precipitation nowcast scheme. *Water Resources Research*, 49:6624–6641.
- Skiena, S. S. (1998). The Algorithm Design Manual. Springer.
- Steinert, J., Tracksdorf, P., and Heizenreder, D. (2021). Hymec: Surface precipitation type estimation at the German Weather Service. Weather and Forecasting, 36:1611–1627.
- Steinwart, I. and Christmann, A. (2011). Estimating conditional quantiles with the help of the pinball loss. *Bernoulli*, 17:211–225.

- Tobin, I., Bony, S., and Roca, R. (2012). Observational evidence for relationships between the degree of aggregation of deep convection, water vapor, surface fluxes, and radiation. *Journal of Climate*, 25:6885–6904.
- Tu, Q., Miao, S., Yao, F., Yang, W., Lin, Y., and Zheng, Z. (2023). An improved wind power uncertainty model for day-ahead robust scheduling considering spatio-temporal correlations of multiple wind farms. *International Journal of Electrical Power & Energy Systems*, 145:108674.
- Van de Velde, J., Demuzere, M., De Baets, B., and Verhoest, N. (2023). Future multivariate weather generation by combining Bartlett-Lewis and vine copula models. *Hydrological Sciences Journal*, 68:1–15.
- Vinecopulib (2023). Vinecopulib/pyvinecopulib: A python library for vine copula models. https://github.com/vinecopulib/pyvinecopulib. Accessed: 2023-04-20.
- Wernli, H., Paulat, M., Hagen, M., and Frei, C. (2008). SAL—A novel quality measure for the verification of quantitative precipitation forecasts. *Monthly Weather Review*, 136:4470–4487.
- Wilks, D. S. (2006). *Statistical Methods in the Atmospheric Sciences*. Academic Press.
- Wu, L., Zhang, Y., Adams, T., Lee, H., Liu, Y., and Schaake, J. (2018). Comparative evaluation of three Schaake shuffle schemes in postprocessing GEFS precipitation ensemble forecasts. *Journal of Hydrometeorology*, 19:575–598.
- Zängl, G., Reinert, D., Rípodas, P., and Baldauf, M. (2015). The ICON (ICOsahedral Non-hydrostatic) modelling framework of DWD and MPI-M: Description of the non-hydrostatic dynamical core. *Quarterly Journal of the Royal Meteorological Society*, 141:563–579.