A Similarity-Based Implementation of the Schaake Shuffle

ROMAN SCHEFZIK

Heidelberg Institute for Theoretical Studies, Heidelberg, Germany

(Manuscript received 16 June 2015, in final form 1 February 2016)

ABSTRACT

Contemporary weather forecasts are typically based on ensemble prediction systems, which consist of multiple runs of numerical weather prediction models that vary with respect to the initial conditions and/or the parameterization of the atmosphere. Ensemble forecasts are frequently biased and show dispersion errors and thus need to be statistically postprocessed. However, current postprocessing approaches are often univariate and apply to a single weather quantity at a single location and for a single prediction horizon only, thereby failing to account for potentially crucial dependence structures. Nonparametric multivariate postprocessing methods based on empirical copulas, such as ensemble copula coupling or the Schaake shuffle, can address this shortcoming. A specific implementation of the Schaake shuffle, called the SimSchaake approach, is introduced. The SimSchaake method aggregates univariately postprocessed ensemble forecasts using dependence patterns from past observations. Specifically, the observations are taken from historical dates at which the ensemble forecasts resembled the current ensemble prediction with respect to a specific similarity criterion. The SimSchaake ensemble outperforms all reference ensembles in an application to ensemble forecasts for 2-m temperature from the European Centre for Medium-Range Weather Forecasts.

1. Introduction

Contemporary weather forecasts are typically constructed from ensemble prediction systems, which have run operationally since the early 1990s. An ensemble consists of multiple runs of numerical weather prediction models, which vary with respect to the initial conditions and/or the parameterization of the atmosphere. Consequently, ensembles take account of the two major sources of uncertainty (Palmer 2002; Gneiting and Raftery 2005; Leutbecher and Palmer 2008). Ensemble forecasts are frequently biased and show dispersion errors (Hamill and Colucci 1997). Thus, they require statistical postprocessing to realize their full capability.

During the last decade, several ensemble postprocessing methods have been developed. Examples are (variants of) the ensemble model output statistics (EMOS; Gneiting et al. 2005, among others) approach, which is also known as nonhomogeneous regression, or Bayesian model averaging (BMA; Raftery et al. 2005, among others). However, the original EMOS and BMA approaches, as well as other postprocessing methods, only apply to a single weather variable at a single location and for a single prediction horizon. Thus, they fail to account for spatial, temporal, or intervariable dependence structures, which are crucial in many applications such as flood warning (Schaeke et al. 2010), winter road maintenance (Berrocal et al. 2010), or the handling of renewable energy sources (Pinson 2013).

In recent years, there has been a keen interest in the development of multivariate postprocessing methods being able to address this shortcoming, and considerable effort has been invested to that end. For instance, variants and modifications of EMOS and BMA, respectively, that can handle spatial (Berrocal et al. 2007, 2008; Feldmann et al. 2015) or intervariable (Schuhen et al. 2012; Möller et al. 2013; Sloughter et al. 2013; Baran and Möller 2015) dependencies are available.

Furthermore, there are methods for capturing temporal dependencies of consecutive lead times in postprocessed predictive distributions (Pinson et al. 2009; Schötz and Hense 2011). All of these multivariate postprocessing methods are parametric and work well in rather low dimensions and in settings in which the involved
correlation matrix can be taken to be highly structured. However, they model one type of dependence (spatial, intervariable, or temporal) only and appear to be inadequate when considering high-dimensional situations in which no particular structure can be exploited. These issues can be addressed using nonparametric techniques based on the use of empirical copulas (Schefzik 2015a). Following Wilks (2015), an empirical copula (Deheuvels 1979; Rüschendorf 2009) can be interpreted as a dependence template induced by a specific discrete multivariate dataset. It can be employed to transfer a particular dependence pattern to samples that are drawn independently from a collection of univariate marginal distributions (Wilks 2015). For example, in the ensemble copula coupling (ECC) approach of Schefzik et al. (2013), such an empirical copula is derived from the unprocessed ensemble forecast and then applied to samples from univariate postprocessed predictive distributions, which can be gained via the standard EMOS or BMA approaches. This is equivalent to ordering these samples according to the rank-dependence structure of the raw ensemble, thereby capturing the spatial, temporal, and intervariable flow dependence (Schefzik et al. 2013). Proceeding in a similar manner, the Schaake shuffle (Clark et al. 2004) employs an ordering based on past observations from a historical data archive. Consequently, the corresponding empirical copula in the Schaake shuffle is induced by an observational database rather than by an ensemble forecast. However, the standard Schaake shuffle fails to condition the multivariate dependence pattern on current or predicted atmospheric states. To address this shortcoming, Clark et al. (2004, p. 260) proposed to develop an extension thereof, driven by the idea

“to preferentially select dates from the historical record that resemble forecasted atmospheric conditions and use the spatial correlation structure from this subset of dates to reconstruct the spatial variability for a specific forecast.”

Inspired by this suggestion, a specific implementation of the Schaake shuffle, referred to as the SimSchaake approach, is introduced in this paper. Essentially, the SimSchaake method proceeds like the Schaake shuffle, but the observations determining the dependence structure are taken from historical dates at which the ensemble forecasts resembled the current ensemble prediction with respect to a specific similarity criterion. Hence, the dependence template in the SimSchaake approach is constructed by using an analog ensemble concept in the spirit of Hamill and Whitaker (2006) or Delle Monache et al. (2011).

The remainder of the paper is organized as follows. In section 2, we first discuss the general setting of empirical copula-based ensemble postprocessing and review ECC and the Schaake shuffle as reference examples. Then, we develop the SimSchaake approach. In section 3, this new method is evaluated and compared to the reference methods in a case study. The paper closes with a discussion in section 4.

2. Empirical copula-based ensemble postprocessing methods

Copulas are valuable and established tools for the modeling of stochastic dependence (Nelsen 2006; Joe 2014). They have been successfully employed in numerous application areas. A copula is an $L$-variate cumulative distribution function (CDF) with standard uniform univariate marginal CDFs, where $L \in \mathbb{N}$ and $L \geq 2$. As is manifested in the famous Sklar’s theorem (Sklar 1959), a copula $C$ links a multivariate CDF $H$ to its univariate marginal CDFs $F_1, \ldots, F_L$ via the decomposition

$$H(u_1, \ldots, u_L) = C(F_1(u_1), \ldots, F_L(u_L))$$

for $u_1, \ldots, u_L \in \mathbb{R}$. In a multivariate postprocessing setting, the sought multivariate CDF $H$ can thus be constructed by specifying both the univariate marginal CDFs $F_1, \ldots, F_L$ and the copula $C$ modeling the dependence. The CDFs $F_1, \ldots, F_L$ can be obtained by common univariate postprocessing for each location, weather variable, and look-ahead time individually, for instance performed via EMOS or BMA. For the choice of $C$, a prominent example is the Gaussian copula, which has been applied to a wide range of problems in climatology, meteorology, and hydrology (Genest and Favre 2007; Schölzel and Friederichs 2008; Pinson 2012; Schuhen et al. 2012; Mörller et al. 2013).

In this paper, we focus on the case in which $F_1, \ldots, F_L$ are the empirical CDFs given by samples from univariate postprocessed CDFs and $C$ is taken to be an empirical copula (Deheuvels 1979; Rüschendorf 2009). According to Wilks (2015), an empirical copula can be considered a multivariate dependence template derived from a specific discrete dataset. To describe this formally, let $z := \{(z_1^1, \ldots, z_N^1), \ldots, (z_1^L, \ldots, z_N^L)\}$ be a dataset consisting of $L$ tuples of size $N$ with entries in $\mathbb{R}$. Moreover, let $\text{rank}(z_n^\ell)$ denote the rank of $z_n^\ell$ in $\{z_1^\ell, \ldots, z_N^\ell\}$ for $n \in \{1, \ldots, N\}$ and $\ell \in \{1, \ldots, L\}$, assuming for simplicity that there are no ties. Then, the empirical copula $E_N$ induced by the dataset $z$ is given by

$$E_N\left(\frac{i_1}{N}, \ldots, \frac{i_N}{N}\right) := \frac{1}{N} \sum_{n=1}^{N} \mathbb{1}_{\{\text{rank}(z_n^\ell) \leq i_1, \ldots, \text{rank}(z_n^\ell) \leq i_L\}}$$

$$= \frac{1}{N} \sum_{n=1}^{N} \prod_{\ell=1}^{L} \mathbb{1}_{\{\text{rank}(z_n^\ell) \leq i_\ell\}} \quad (1)$$

Unauthenticated | Downloaded 12/12/20 08:56 PM UTC
for integers $0 \leq i_1, \ldots, i_L \leq N$, with $1_A$ denoting the indicator function whose value is 1 if the event $A$ occurs, and zero otherwise.

In this section, we review ensemble copula coupling (Schefzik et al. 2013) and the Schaake shuffle (Clark et al. 2004) as reference methods within the general framework of empirical copula-based multivariate ensemble postprocessing (Schefzik 2015a). In addition, we develop the SimSchaake method as a specific implementation of the Schaake shuffle. While ECC and the Schaake shuffle have been used as a benchmark in several papers (Schefzik et al. 2007; Voisin et al. 2011; Scheuerer and Hamill 2015; Vrac and Friederichs 2015; Wilks 2015), the SimSchaake method is new.

To allow for a formal description of the approaches, let us first define some notation that will be used throughout the whole section. Let $t \in \{1, \ldots, L\}$ be a weather variable, $j \in \{1, \ldots, J\}$ a location, and $k \in \{1, \ldots, K\}$ a look-ahead time. For simplicity, let $t^e := (i, j, k)$ and $t^p := (i, j)$ denote the corresponding multi-indices, and let $L := I \times J \times K$ and $L^e := I \times J$, respectively. Moreover, let $M$ denote the number of raw ensemble members, and $N$ the desired number of members the postprocessed ensemble shall consist of.

### a. General framework for empirical copula-based multivariate ensemble postprocessing

In general, multivariate empirical copula-based ensemble methods for postprocessing a raw ensemble forecast $x := \{(x^1_{i_1}, \ldots, x^M_{i_1}), \ldots, (x^1_{i_L}, \ldots, x^M_{i_L})\}$ initialized at a specific date $t_0$ proceed according to the following scheme.

1. To implement dependence structures, derive an empirical copula $E_N$ from a suitable dataset $z := \{(z^1_{x^1}, \ldots, z^M_{x_1}), \ldots, (z^1_{x_L}, \ldots, z^M_{x_L})\}$ via (1). Equivalently, derive the univariate order statistics $z^p_{(n)} \leq \cdots \leq z^p_{(N)}$ for $n \in \{1, \ldots, N\}$, and use their ranks.

2. For each margin $t^e$, perform univariate postprocessing of the raw ensemble forecast $x^t_{t^e}$ based on past forecasts and observations as training data, and obtain a postprocessed predictive CDF $F_t$ in each case.

3. Draw a sample $\tilde{x}^t_1, \ldots, \tilde{x}^t_N$ from each marginal CDF $F_t$. This can be most conveniently done by taking equidistant quantiles of $F_t$ of the form

\[
\tilde{x}^t_1 := F_t^{-1} \left( \frac{1}{N+1} \right), \ldots, \tilde{x}^t_N := F_t^{-1} \left( \frac{N}{N+1} \right),
\]

Alternatively, one could for instance work with a random sample of the form

\[
\tilde{x}^t_1 := F_t^{-1}(v_1), \ldots, \tilde{x}^t_N := F_t^{-1}(v_N),
\]

where $v_1, \ldots, v_N$ are independent standard uniform random variates.

4. Apply the empirical copula $E_N$ to the samples from step 3.

Equivalently, arrange these samples with respect to the rank-dependence structure of the chosen dataset $z$ in step 1. Using the permutation $\pi_z$, the final postprocessed ensemble $\tilde{x}^t_1, \ldots, \tilde{x}^t_N$ for each margin $t$ is thus given by

\[
\tilde{x}^t_1 := \tilde{x}^t_{(\pi(z))}, \ldots, \tilde{x}^t_N := \tilde{x}^t_{(\pi(N))}.
\]

### b. Reference methods

Now we review ECC (Schefzik et al. 2013) and the Schaake shuffle (Clark et al. 2004) as empirical copula-based postprocessing techniques within the scheme described above.

#### 1) Ensemble Copula Coupling

In the ECC approach (Schefzik et al. 2013), the dataset specifying the dependence structure is given by the raw ensemble forecast; that is, we have $z = x = \{(x^1_{i_1}, \ldots, x^M_{i_1}), \ldots, (x^1_{i_L}, \ldots, x^M_{i_L})\}$ in the scheme from section 2a. Depending on the chosen sampling procedure in step 3, one can distinguish between the ECC-Q and the ECC-R ensembles (Schefzik et al. 2013), referring to the use of equidistant quantiles according to (2) and random samples according to (3), respectively. The size of the ECC-Q ensemble is restricted to equal that of the unprocessed ensemble; that is, $N = M$. In contrast, the ECC-R concept principally allows postprocessed ensemble sizes that can be any integer multiple of the raw ensemble size (Wilks 2015). For convenience and theoretical reasons (Schefzik et al. 2013, and references therein), we will restrict our attention to the ECC-Q ensemble in what follows, where the suffix “-Q” will be omitted.

The ECC procedure operates under a perfect model assumption, implicitly assuming that the ensemble is capable of representing actual spatial, intervariable, and temporal dependence structures adequately. This may or may not be expected, but surely does not hold each and every day. Moreover, ECC only applies to ensembles whose members can be considered exchangeable, that is, statistically indistinguishable.

Schefzik et al. (2013) and Schefzik (2015b) show that ECC provides an overarching frame for seemingly unrelated approaches scattered throughout the literature such as those of Pinson (2012), Roulin and Vannitsem (2012),...
Flowerdew (2014), or Van Schaeybroeck and Vannitsem (2015), to name just a few. It has been used as a reference technique in the recent papers by Feldmann et al. (2015), Scheuerer and Hamill (2015), and Wilks (2015).

2) THE SCHAAKE SHUFFLE

In contrast to ECC, the dataset determining the dependence pattern in the Schaake shuffle (Clark et al. 2004) is not given by the raw ensemble forecasts, but by past observations $y = \{(y_1^1, \ldots, y_N^1), \ldots, (y_1^L, \ldots, y_N^L)\}$ taken from $N$ different dates within a historical archive. That is, observations from the same $N$ dates are employed for all locations and weather variables throughout the procedure of the Schaake shuffle for a fixed forecast instance. Hence, we have $z = y$ in the scheme from section 2a. In particular, $N$ does not need to equal the raw ensemble size $M$.

In the original implementation of the Schaake shuffle by Clark et al. (2004), the corresponding $N$ dates, from which the observations are taken, are chosen from all years in the historical record, except for the year of the forecast of interest, and lie within 7 days before and after the verification date $t$, regardless of the year. A more general implementation of the Schaake shuffle may use observations from arbitrary (or randomly selected) past dates across the whole historical record. We will refer to this procedure as the Random Schaake method in the case study in section 3.

The Schaake shuffle has been employed successfully in numerous applications (Schaake et al. 2007; Voisin et al. 2010, 2011; Robertson et al. 2013; Verkade et al. 2013; Vrac and Friederichs 2015; Wilks 2015).

c. The SimSchaake approach as a similarity-based implementation of the Schaake shuffle

As we have seen, the Schaake shuffle generates a postprocessed ensemble inheriting the rank dependence structure from historical observations. However, its standard implementations fail to condition the multivariate dependence pattern on current or predicted atmospheric states. Inspired by the quote of Clark et al. (2004) mentioned in section 1, we address this challenge by linking the Schaake shuffle to analog methods. In such analog approaches, one seeks forecasts in an archive of past data that are similar to the current one; these are referred to as analogs. The basic idea is that if a forecast similar to the current one can be detected in the historical data archive, the corresponding observed state of the atmosphere is supposed to be similar to the state to be predicted. Information about the forecast error of the current prediction can then be inferred from the errors of the analogs, for which there are realizing observations (Hamill and Whitaker 2006; Delle Monache et al. 2011; Messner and Mayr 2011). Analog-based techniques have become popular and important, as for instance witnessed by the papers of Bannayan and Hoogenboom (2008), Klausner et al. (2009), Hall et al. (2010), Delle Monache et al. (2011), Messner and Mayr (2011), and Delle Monache et al. (2013), as well as by more recent research (Alessandrini et al. 2015; Junk et al. 2015; Lerch and Baran 2015; Vanvye et al. 2015).

Within their context, the question of the choice of appropriate similarity criteria in a nearest-neighbor sense arises, with the papers above providing some proposals. The following new approach, which will be referred to as the SimSchaake method in what follows, combines the idea of searching for similar ensembles and the Schaake shuffle. Like the Schaake shuffle, it can be applied to any ensemble, regardless of whether it consists of exchangeable or nonexchangeable members, and the size $N$ of the final postprocessed ensemble is not restricted to equal the raw ensemble size $M$. Speaking in terms of the scheme from section 2a, the SimSchaake method uses the observations corresponding to specific analogs to determine the dependence template $z$ in the first step. However, the analogs and their respective observations are not necessarily employed as training data for parameter estimation in the univariate post-processing in the second step. To describe the SimSchaake approach formally in detail, let $\Lambda$ be the length of the training period required for the univariate ensemble post-processing, $t_0$ the initialization date of the ensemble forecast, and $t$ the verification date. Further, let $D$ be the number of dates in the past of $t_0$ for which ensemble forecast and observation data are available. For the feasibility of the SimSchaake approach, it is required that ensemble forecast and observation data are available for at least $\max\{N, \Lambda\}$ dates in the past of $t_0$; that is, $D \geq \max\{N, \Lambda\}$.

Considering the prediction horizon to be fixed, the SimSchaake approach then proceeds according to the empirical copula-based postprocessing as described in section 2a, where the dataset $z$ in step 1 is derived as follows.

(i) For a fixed margin $\ell^*$, let $x^{\ell,r} := (x_1^{\ell,r}, \ldots, x_M^{\ell,r})$ denote the (possibly standardized) $M$-member raw ensemble forecast valid on date $t$. Further, let

$$x^\tau := (x_1^{\tau,r}, \ldots, x_M^{\tau,r}) = ((x_1^{\tau,r}, \ldots, x_M^{\tau,r}), \ldots, (x_1^{\tau,r}, \ldots, x_M^{\tau,r}))$$

denote the corresponding $(L^* \times M)$ tuple consisting of the $M$-member ensemble forecasts of all $L^*$ margins, that is, combinations of weather variable and location. If weather variables with distinct units or magnitudes are involved, the components of $x^{\ell,r}$ should be standardized for each multi-index $\ell^*$. 

Unauthenticated | Downloaded 12/12/20 08:56 PM UTC
(ii) For each date \( t_d \) in the past of the initialization date \( t_0 \), where \( d \in \{1, \ldots, D\} \), compute a suitable fixed similarity criterion,

\[
\Delta^\ell; \mathbb{R}^{L_x \times M} \times \mathbb{R}^{L_x \times M} \to [0, \infty), (\mathbf{x}', \mathbf{x}'') \mapsto \Delta^\ell(\mathbf{x}', \mathbf{x}''),
\]

between the actual forecast \( \mathbf{x}' \) valid on the verification date \( t \) and the forecast \( \mathbf{x}'' \) valid on the date \( t_d \).

The similarity criterion \( \Delta^\ell \) is taken to be negatively oriented; that is, the lower the value of \( \Delta^\ell \), the more similar the ensemble forecasts. A similarity criterion value of exactly zero indicates that the corresponding ensemble forecasts are identical.

(iii) Choose those \( N \) dates \( \tau_1, \ldots, \tau_N \in \{t_1, \ldots, t_D\} \) for which the data are most similar to those for the date \( t \) in the sense that the corresponding values of \( \Delta^\ell_n \) for \( n \in \{1, \ldots, N\} \) are the smallest among the values of \( \Delta^\ell_d \) for \( d \in \{1, \ldots, D\} \). Note that the information of all multi-indices \( \ell^\alpha \) simultaneously is employed to determine the dates \( \tau_1, \ldots, \tau_N \).

(iv) For each margin \( \ell^\alpha \), let \( y_{n,\tau}^{\ell^\alpha} \) denote the corresponding \( N \) historical verifying observations valid on the dates \( \tau_1, \ldots, \tau_N \) determined in step iii. For simplicity, write \( y_{n}^{\ell^\alpha} := y_{n,\tau}^{\ell^\alpha} \) for \( n \in \{1, \ldots, N\} \) and build the data vector \( y^{\ell^\alpha} := (y_{1}^{\ell^\alpha}, \ldots, y_{N}^{\ell^\alpha}) \) for the verification date \( t \). The dataset \( \mathbf{z} \) in step 1 from the scheme in section 2a is then obtained by aggregating the historical observation databases of all margins \( \ell^\alpha \); that is, \( \mathbf{z} = (y^{\ell^\alpha}, \ldots, y^{\ell^\alpha}) \). From this template \( \mathbf{z} \), the empirical copula related to the SimSchaake approach is derived via (1).

Having obtained the dataset \( \mathbf{z} \) and the respective empirical copula according to the above procedure, steps 2–4 from the scheme in section 2a are performed in order to generate the final postprocessed SimSchaake ensemble. An outline of the SimSchaake approach is given in Fig. 1.

An appropriate choice of the similarity criterion \( \Delta^\ell_d \) in step ii, which is then consistently used throughout the whole SimSchaake approach, is crucial. We here consider

\[
\Delta^\ell(\mathbf{x}', \mathbf{x}'') := \sqrt{\frac{1}{L_x} \sum_{t=1}^{L_x} (\mathbf{x}_{\tau}^{\ell^\alpha} - \mathbf{x}_t^{\ell^\alpha})^2} + \frac{1}{L_x} \sum_{t=1}^{L_x} (s_{\tau}^{\ell^\alpha} - s_t^{\ell^\alpha})^2,
\]

(4)

where

\[
\mathbf{x}_{\tau}^{\ell^\alpha} := \frac{1}{M} \sum_{m=1}^{M} x_{m,\tau}^{\ell^\alpha} \quad \text{and} \quad s_{\tau}^{\ell^\alpha} := \sqrt{\frac{1}{M} \sum_{m=1}^{M} (x_{m,\tau}^{\ell^\alpha} - \mathbf{x}_{\tau}^{\ell^\alpha})^2},
\]

denote the empirical mean and standard deviation, respectively, of the ensemble forecast \( \mathbf{x}_{\tau}^{\ell^\alpha} \) for the fixed multi-index \( \ell^\alpha \) at date \( \tau \). As it does not depend on how the ensemble members are labeled, the similarity criterion in (4) can be applied to ensembles consisting of exchangeable members. In particular, it is suitable for the temperature predictions from the European Centre for Medium-Range Weather Forecasts (ECMWF).
used in the case study in the next section. While the use of the empirical mean to some degree accounts for seasonal aspects when considering temperature forecasts only, the empirical standard deviations reflect the uncertainties within the ensemble forecasts. Thus, these issues are addressed by (4) when comparing two ensemble forecasts. Alternative proposals for similarity criteria can be found in the references mentioned before.

It is possible to transform the values of a similarity criterion $\Delta^\nu$ from $[0, \infty)$ to $(0, 1]$ by employing the standardization $\Delta^\nu := \exp(-\Delta^\nu)$. With respect to $\Delta^\nu$, similarity values near to 1 indicate a very high similarity between $\mathbf{x}'$ and $\mathbf{x}^\nu$, while similarity values near to 0 point at no similarity. Accordingly, if using $\Delta^\nu$, we then have to choose the dates $\tau_1, \ldots, \tau_N$ corresponding to the highest, and not to the lowest, values of $\Delta^\nu$ in step iii.

As mentioned, the SimSchaake approach addresses two shortcomings in the standard ECC method considered in this paper using equidistant quantiles according to (2). First, it can also be applied to ensembles consisting of nonexchangeable members, as the reordering is not based on the ensemble forecasts, but on observations. Second, with the SimSchaake technique we can principally create ensembles of arbitrary size, as long as there are a sufficient number of historical observations in the past. The postprocessed ensemble is thus not restricted to have the same number of members as the raw ensemble.

In Fig. 2, an illustration of the three empirical copula-based postprocessing methods presented in this section—ECC in the first row, the Schaake shuffle in the second row, and the SimSchaake method in the third row—is given within the context of spatial modeling. We consider 24-h-ahead 2-m temperature forecasts (in degrees Celsius) at Vienna, Austria, and Bratislava, Slovakia, valid on the hot summer day at 1200 UTC 9 July 2011. In the subfigures, a raw or postprocessed ensemble forecast is indicated by the dots, the verifying observation by the cross, and past observations from a historical database by the circles. As the ECMWF raw ensemble used here is of size $M = 50$, the illustrations are for convenience based on $N = M = 50$ member ensembles for both the ECC and the Schaake shuffle-based methods. The first column in Fig. 2 shows the corresponding database that is used to determine the dependence structure and the empirical copula, respectively, of the postprocessing approach: the ECMWF raw ensemble in the case of ECC, past observations on random dates for the Random Schaake method, and past observations on specific dates selected according to the similarity criterion in (4) in the case of the SimSchaake approach. The second column in Fig. 2 shows the same individually postprocessed ensemble forecast three times. This is generated by randomly pairing the equidistant quantiles (2) from the predictive CDFs obtained by univariate EMOS postprocessing (Gneiting et al. 2005) at Vienna and Bratislava separately. Such an ensemble forecast does not take account of the pronounced positive spatial correlation structure. In the third column, the final postprocessed ensemble forecasts are shown, obtained by applying the corresponding empirical copula to the samples derived from the individual EMOS postprocessing. These empirical copula-based postprocessed ensembles have the same marginal distributions as the individual EMOS postprocessed ensemble, as witnessed by the respective histograms at the top and to the right of each subfigure. Additionally, they exhibit the same spatial correlation pattern as the underlying database specified in the first column, thus respecting dependencies.

3. Case study

a. Setting

In our case study, we employ predictions of the ECMWF core ensemble (ECMWF Directorate 2012), whose $M = 50$ members can be considered exchangeable. We focus on 24-h-ahead 2-m temperature forecasts jointly at Vienna, Bratislava, and Budapest, Hungary, and consequently on spatial dependencies only. For each location, the ground truth is given by the corresponding synoptic observations (SYNOP). The approximate distance from Vienna to Bratislava is 50 km, from Bratislava to Budapest 170 km, and from Vienna to Budapest 210 km. There are pronounced positive pairwise correlations between the observations at the three different locations, which are stronger the closer the respective stations are. These correlation patterns are basically reflected well in the ensemble forecasts.

We consider those 3985 test days during the period from 1 January 2003 to 31 December 2013 for which all required forecast and observation data are available at all the three stations. In our case study, all data are valid at 1200 UTC. Univariate postprocessing is performed via EMOS (Gneiting et al. 2005) using the R package ensembleMOS (Yuen et al. 2013; R Core Team 2014), employing a rolling window consisting of the last $\Lambda = 50$ days before the verification day as the training period. For each marginal EMOS postprocessed predictive CDF $F_\tau$, we follow the quantization (2) and take the $N$ equidistant ($n/(N + 1)$) quantiles, where $n = 1, \ldots, N$, as samples, focusing on the cases of $N = M = 50$ and $N = 80$, respectively, here.

For these desired ensemble sizes, we assess and compare the predictive performances of
(a) Database for Empirical Copula (Dependence Model)  
(b) Individually EMOS Postprocessed Ensemble  
(c) EMOS and Empirical Copula-Based Postprocessed Ensemble

- the ECMWF raw ensemble;
- the Individual EMOS ensemble, which assumes independence;
- the EMOS-ECC ensemble, which assumes a dependence structure according to that in the raw ensemble;
- the EMOS-Random Schaake ensemble, which assumes a dependence structure according to randomly selected historical observation data; and
- the EMOS-SimSchaake ensemble, which assumes a dependence structure according to specific historical observations valid on dates for which the ensemble forecast resembled the current one with respect to similarity criterion (4).

Obviously, results for the raw and the EMOS-ECC ensemble can only be reported for the case of \( N = M = 50 \), whereas the other ensembles are additionally evaluated for the final ensemble size of \( N = 80 \). For the two approaches employing the Schaake shuffle notion, the past dates from which the corresponding verifying observations
are taken are searched for among all available historical data, where ensemble forecast and observation data are available from 1 January 2002 to 31 December 2013. Hence, the database used for the Schaake shuffle–based methods grows over time. Recall that the EMOS-Random Schaake method only randomly selects those past dates, whereas the EMOS-SimSchaake approach chooses them based on the ensemble similarity criterion (4).

b. Evaluation tools

A probabilistic forecast distribution or an ensemble forecast, respectively, should be as sharp as possible, subject to calibration, which refers to statistical consistency between the forecasts and the observation (Gneiting et al. 2007). To assess the predictive performances of our different ensembles, several verification tools are available (Wilks 2011).

In univariate settings, calibration can be checked via the verification rank histogram (Anderson 1996; Talagrand et al. 1997; Hamill 2001). As we focus on the evaluation of multivariate quantities in this paper, ensemble calibration in our case study is checked via the multivariate (Gneiting et al. 2008), band depth, and average rank histograms (Thorarinsdottir et al. 2016). When an ensemble forecast is calibrated, the multivariate, band depth or average rank, respectively, is uniformly distributed. Calibration thus be diagnosed by compositing over forecast cases, plotting the corresponding multivariate, band depth, or average rank histogram, respectively, and checking for deviations from uniformity, that is, flatness of the histogram. For an interpretation of the different shapes a rank histogram for multivariate quantities can exhibit, see Gneiting et al. (2008) and Thorarinsdottir et al. (2016).

The overall forecast skill can be assessed via proper scoring rules (Gneiting and Raftery 2007), which are able to judge calibration and sharpness simultaneously and are taken to be negatively oriented here; that is, the lower the score, the better the predictive performance. A widely used proper scoring rule for univariate quantities is the continuous ranked probability score (CRPS; Matheson and Winkler 1976; Gneiting and Raftery 2007).

In this paper, we employ the energy score (ES; Gneiting et al. 2008), which is the analog of the CRPS for multivariate quantities. For an $N$-member ensemble forecast, the ES is computed as

$$\text{ES} = \frac{1}{N} \sum_{n=1}^{N} \| \mathbf{x}_n - \mathbf{y} \| - \frac{1}{2N^2} \sum_{\nu=1}^{N} \sum_{n=1}^{N} \| \mathbf{x}_\nu - \mathbf{x}_n \|,$$

with $\| \cdot \|$ denoting the Euclidean norm. As the ES reveals weaknesses in detecting misspecifications in the correlation structure (Pinson and Tastu 2013; Scheuerer and Hamill 2015), we additionally consider the variogram score (VS; Scheuerer and Hamill 2015) to address this, which is given by

$$\text{VS} = \sum_{\ell=1}^{L} \sum_{\lambda=1}^{L} w_{\ell \lambda} \left( \frac{1}{\text{dist}(\ell, \lambda)} \right) \left( \frac{1}{N} \sum_{n=1}^{N} \| \mathbf{x}_\ell - \mathbf{x}_n \| \right)^2,$$

where the $w_{\ell \lambda}$'s are (optional) nonnegative weights. For the case study in this paper, in which we focus on spatial dependencies, we follow the suggestion of Scheuerer and Hamill (2015) and let the weights be proportional to the inverse spatial distances between the corresponding locations. That is, we choose

$$w_{\ell \lambda} := \frac{1}{\text{dist}(\ell, \lambda)} \sum_{\ell=1}^{L} \frac{1}{\text{dist}(\ell, \lambda)}$$

for $\ell \neq \lambda$ and $w_{\ell \ell} := 0$ for $\ell = \lambda$, with $\text{dist}(\ell, \lambda)$ denoting the spatial distance between location $\ell$ and location $\lambda$, where all distances have to be measured in the same unit. For the specific implementation in our $L = 3$-dimensional setting here, we employ the distances between Vienna, Bratislava, and Budapest as mentioned in the preceding subsection.

In our case study, average scores over all forecast cases within our specific test period are reported.

c. Results

As we focus on the multivariate setting in this paper, we do not explicitly show the results for the univariate EMOS postprocessing at the three stations individually here. In a nutshell, the EMOS postprocessed ensemble forecasts exhibit a better predictive performance than do the unprocessed raw ensemble predictions, in that they are better calibrated and have smaller CRPS values. In Table 1, the average ES and VS as overall performance measures are shown. The results for the Individual EMOS and the EMOS-Random Schaake ensemble are averaged over 100 runs for each forecast instance, in order to account for randomizations. Precisely, for the Individual EMOS ensemble, the results for 100 different aggregations (i.e., assignments of the

Unauthenticated | Downloaded 12/12/20 08:56 PM UTC
When comparing the three empirical copula-based postprocessing methods considering different correlation structures (Thorarinsdottir et al. 2016), which is plausible, as this approach does not account for dependencies. With respect to the VS, the Individual EMOS ensemble performs worse than the other postprocessed ensembles, which assume specific correlation structures. In terms of the ES, the EMOS-ECC and the EMOS-SimSchaake ensembles outperform the Individual EMOS ensemble, while the distinctions are less pronounced than for the VS. This may be due to the discrimination inability of the ES with respect to correlations between different locations (Pinson and Tastu 2013; Scheuerer and Hamill 2015), as hinted at in the preceding subsection.

Comparing the three empirical copula-based postprocessing methods taking account of dependence patterns, the EMOS-ECC ensemble is well calibrated, apart from a slight overpopulation of the lowest ranks in all rank histograms. In contrast, the calibration of the EMOS-Random Schaake ensemble is not that good, as witnessed by the inverse U-shaped band depth and average rank histograms, indicating an overestimation of the correlation structures (Thorarinsdottir et al. 2016). Moreover, the multivariate rank histogram of the EMOS-Random Schaake ensemble is skewed with too many low ranks. The EMOS-SimSchaake ensemble is calibrated best, with the band depth and average rank histograms being close to uniform and an essentially flat multivariate rank histogram with a slight overpopulation of the lowest ranks only. The ranking of the three ensembles allowing for dependencies in terms of calibration is also reflected in the scores. Both for the ES and the VS, the EMOS-SimSchaake ensemble performs best, followed by the EMOS-ECC ensemble and finally the EMOS-Random Schaake ensemble.

The results and conclusions on calibration described above continue to hold analogously for the extended Individual EMOS, EMOS-Random Schaake, and EMOS-SimSchaake ensembles comprising \( N = 80 \) members. Similarly, the ranking of the predictive skill of these three extended postprocessed ensembles remains unchanged with respect to the ES and VS, respectively, compared to the case of \( N = M = 50 \), with the EMOS-SimSchaake ensemble still performing by far the best. The ES and VS values of the \( N = 80 \)-member ensembles are very similar to those of their counterparts in the case of \( N = M = 50 \). Although an extension of the ensemble size is generally useful, there is no pronounced need in our case here to increase the ensemble size of \( N = M = 50 \), which appears to already be reasonably large.

In a nutshell, the EMOS-SimSchaake ensemble based on the new method introduced in this paper performs best among all reference ensembles, both with respect to calibration and in terms of scores. In particular, the EMOS-SimSchaake ensemble outperforms the EMOS-Random Schaake ensemble. Hence, there appears to be a clear benefit in using the specific past dates on which the ensemble forecasts resembled the current one to create the historical observation database modeling the dependencies, rather than picking these dates randomly. The EMOS-SimSchaake ensemble also outperforms

| \( N \) | \( M \) | ES  | VS  |
|-------|-------|-----|-----|
| 50    | 50    | 2.241 | 0.333 |
| 80    |       | 1.976 | 0.323 |
|       | 1.957 | 0.270 |
|       | 1.998 | 0.300 |
|       | 1.952 | 0.265 |
|       | 1.971 | 0.327 |
|       | 1.996 | 0.300 |
|       | 1.947 | 0.266 |

| \( M = 50 \) | \( N = 50 \) | ECMWF raw ensemble | Individual EMOS ensemble | EMOS-ECC ensemble | EMOS-Random Schaake ensemble | EMOS-SimSchaake ensemble | EMOS-Random Schaake ensemble | EMOS-SimSchaake ensemble |
|-------------|-------------|-------------------|-------------------------|-----------------|--------------------------|------------------------|--------------------------|------------------------|
| 50          | 50          | 2.241             | 1.976                   | 1.957           | 1.998                    | 1.952                  | 1.971                    | 1.996                  | 1.947                  |
the EMOS-ECC ensemble. Finally, it is noteworthy that
the EMOS-ECC ensemble outperforms both the $N = M = 50$ member and the $N = 80$ member EMOS-Random Schaake ensembles in our case study here. This is to some extent contrary to the results from the related study of Wilks (2015) favoring the Schaake shuffle.

4. Discussion

We have discussed and compared empirical copula-based ensemble postprocessing methods that are able to account for dependencies. While ECC and the Schaake shuffle have been reviewed within a general framework, the SimSchaake scheme has been newly developed in this paper as a multivariate postprocessing tool. Essentially, the SimSchaake procedure aggregates samples from univariate postprocessed distributions, where the underlying dependence structure and the involved empirical copula, respectively, are derived from historical observations at dates in the past that showed a similar ensemble forecast to the current one. In our case study, the SimSchaake ensemble has performed the best overall and better than the ECC ensemble, while having the benefit of a broader applicability, in that it can also be employed on ensembles comprising nonexchangeable members and is not restricted to having the same size as the raw ensemble.

In the SimSchaake method, observations corresponding to specific analogs are employed to determine a dependence template, but not necessarily as training data for the univariate postprocessing. In contrast, the recent methods of Junk et al. (2015) and Lerch and Baran (2015)
include analogs in the training data used for parameter estimation in univariate EMOS postprocessing. In a future work, both concepts might be combined in an overarching approach that employs analogs both for the univariate postprocessing and the specification of the dependence structure.

The SimSchaake method depends on the design of a suitable similarity criterion, where the choice of (4) has proven to be useful, yielding good results. Like (4), many existing similarity criteria are also based on root-mean-square errors and thus might be expected to basically yield similar results. Exemplarily, replacing the ensemble mean as a summary statistic in (4) by the ensemble median essentially leads to the same results as obtained in the case study. Alternatively, similarity criteria can also be based on ranks (Hamill and Whitaker 2006; Messner and Mayr 2011). An initial test implementation of the EMOS-SimSchaake method using a rank-based similarity criteria did not perform better than the EMOS-SimSchaake ensemble based on (4). As a future work, more sophisticated similarity criteria, which might improve the predictive performance of the SimSchaake ensemble, could be designed, perhaps including a suitable weighting function (or monotone transformations thereof) that accounts for seasonal aspects. Moreover, the similarity criterion in (4) is tailored to ensembles consisting of exchangeable members such as the ECMWF ensemble in our case study. For ensembles comprising nonexchangeable members, other criteria might provide more reasonable and better choices.

Another issue with regard to the SimSchaake approach involves the choice of the length of the historical dataset in which the analogs are searched for. In general, it is recommended to use as much past data as may be available (Delle Monache et al. 2011, 2013). Particularly in settings dealing with rare predictions or extreme events, it might be necessary to look far back in the past to find suitable analogs. Because of computational aspects or the lack of data availability, one might be interested in finding a minimal size of the past dataset after which further increases do not change the predictive performance of the SimSchaake method very much. However, as such an analysis is expected to depend strongly on the weather variable and the specific setting of interest, it appears to be hard to give any general guidance with respect to this at the current stage. Further research in this direction is desirable.

A drawback of the standard ECC postprocessed ensemble employed here is that it is constrained to have the same size as the unprocessed ensemble, while the Schaake shuffle and the SimSchaake ensembles are not. However, as already hinted at, there have been first attempts to design ECC ensembles having more members than the raw ensemble (Schefzik 2015b; Wilks 2015). Similar to the work in Wilks (2015), an issue for future work may be to design and conduct a further case study including these approaches to allow for a broader comparison of ECC and Schaake shuffle–based concepts.

Acknowledgments. I gratefully acknowledge support by the Klaus Tschira Foundation and the VolkswagenStiftung under the project “Mesoscale Weather Extremes: Theory, Spatial Modeling and Prediction.” Initial work on this paper was done during my time as a Ph.D. student at Heidelberg University, funded by Deutsche Forschungsgemeinschaft through Research Training Group (RTG) 1953. I thank Tilmann Gneiting,
Stephan Hemri, Sebastian Lerch, and Martin Leutbecher for valuable comments, suggestions, and discussions. The forecast data used in the case study have been made available by the European Centre for Medium-Range Weather Forecasts. I thank Stephan Hemri for help with the data and Michael Scheuerer and Thordis Thorarinsdottir for R code. Moreover, I gratefully acknowledge the comments and suggestions of two anonymous reviewers.

REFERENCES

Alessandrini, S., L. Delle Monache, S. Sperati, and J. N. Nissen, 2015: A novel application of an analog ensemble for short-term wind power forecasting. Renew. Energy, 76, 768–781, doi:10.1016/j.renene.2014.11.061.

Anderson, J. L., 1996: A method for producing and evaluating probabilistic forecasts from ensemble model integrations. J. Climatol., 9, 1518–1530, doi:10.1175/1520-0442(1996)009<1518:AMFPAE>2.0.CO;2.

Bannayan, M., and G. Hoogenboom, 2008: Predicting realizations of daily weather data for climate forecasts using the non-parametric nearest-neighbour re-sampling technique. Int. J. Climatol., 28, 1357–1368, doi:10.1002/joc.1637.

Baran, S., and A. Möller, 2015: Joint probabilistic forecasting of wind speed and temperature using Bayesian model averaging. Environmetrics, 26, 120–132, doi:10.1002/env.2316.

Berrocal, V. J., A. E. Raftery, and T. Gneiting, 2007: Combining spatial statistical and ensemble information in probabilistic weather forecasts. Mon. Wea. Rev., 135, 1386–1402, doi:10.1175/MWR3341.1.

—, —, and —, 2008: Probabilistic quantitative precipitation field forecasting using a two-stage spatial model. Ann. Appl. Stat., 2, 1170–1193, doi:10.1214/08-AOAS203.

—, —, and —, 2010: Probabilistic weather forecasting for winter road maintenance. J. Amer. Stat. Assoc., 105, 522–537, doi:10.1198/jasa.2009.ap07184.

Clark, M. P., S. Gangopadhyay, L. E. Hay, B. Rajagopalan, and R. L. Wilby, 2004: The Schaake shuffle: A method for re-constructing space-time variability in forecasted precipitation and temperature fields. J. Hydrometeorol., 5, 243–262, doi:10.1175/1525-7541(2004)005<0243:TSSSAMF>2.0.CO;2.

Deheuvels, P., 1979: La fonction de dépendance empirique et ses propriétés. Un test non paramétrique d’indépendance. Acad. Roy. Belg. Bull. Cl. Sci., 65, 274–292.

Delle Monache, L., T. Nipen, Y. Liu, G. Roux, and R. Stull, 2011: Kalman filter and analog schemes to postprocess numerical weather predictions. Mon. Wea. Rev., 139, 3554–3570, doi:10.1175/2011MWR3653.1.

—, —, A. Eckel, D. L. Rife, B. Nagarajan, and K. Searight, 2013: Probabilistic weather prediction with an analog ensemble. Mon. Wea. Rev., 141, 3498–3516, doi:10.1175/MWR-D-12-00281.1.

ECMWF Directorate, 2012: Describing ECMWF’s forecasts and forecasting system. ECMWF Newsletter, No. 133, ECMWF, Reading, United Kingdom, 11–13.

Feldmann, K., M. Scheuerer, and T. L. Thorarinsdottir, 2015: Spatial postprocessing of ensemble forecasts for temperature using nonhomogeneous Gaussian regression. Mon. Wea. Rev., 143, 955–971, doi:10.1175/MWR-D-14-00210.1.

Flowerdew, J., 2014: Calibrating ensemble reliability whilst preserving spatial structure. Tellus, 66A, 22662, doi:10.3402/tellusa.v66.22662.

Genest, C., and A.-C. Favre, 2007: Everything you always wanted to know about copula modeling but were afraid to ask. J. Hydrol. Eng., 12, 347–368, doi:10.1061/(ASCE)1084-0699(2007)12:4(347).

Gneiting, T., and A. E. Raftery, 2005: Weather forecasting with ensemble methods. Science, 310, 248–249, doi:10.1126/science.1115255.

—, —, 2007: Strictly proper scoring rules, prediction, and estimation. J. Amer. Stat. Assoc., 102, 359–378, doi:10.1198/016214507000001437.

—, —. A. H. Westveld, and T. Goldman, 2005: Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation. Mon. Wea. Rev., 133, 1098–1118, doi:10.1175/MWR2904.1.

—, —. F. Balabdaoui, and A. E. Raftery, 2007: Probabilistic forecasts, calibration and sharpness. J. Roy. Stat. Soc., 69B, 243–268, doi:10.1111/j.1467-9868.2007.00587.x.

—, —. L. I. Stanberry, E. P. Grimit, L. Held, and N. A. Johnson, 2008: Assessing probabilistic forecasts of multivariate quantities, with applications to ensemble predictions of surface winds (with discussion and rejoinder). Test, 17, 211–264, doi:10.1007/s11749-008-0114-x.

Hall, T. J., R. N. Thessin, G. J. Bloy, and C. N. Mutcher, 2010: Analog sky condition forecasting based on a k-nm algorithm. Weather Forecasting, 25, 1463–1478, doi:10.1175/2010WAF222372.1.

Hamill, T. M., 2001: Interpretation of rank histograms for verifying ensemble forecasts. Mon. Wea. Rev., 129, 550–560, doi:10.1175/1520-0493(2001)129<0550:IRHFOR>2.0.CO;2.

—, and J. Colucci, 1997: Verification of Eta–RSM short-range ensemble forecasts. Mon. Wea. Rev., 125, 1312–1327, doi:10.1175/1520-0493(1997)125<1312:VOERET>2.0.CO;2.

—, and J. S. Whitaker, 2008: Probabilistic quantitative precipitation forecasts based on reforecast analogs: Theory and application. Mon. Wea. Rev., 134, 3209–3229, doi:10.1175/ MWR3237.1.

Joe, H., 2014: Dependence Modeling with Copulas. CRC Press, 480 pp.

Junk, C., L. Delle Monache, and S. Alessandrini, 2015: Analog-based ensemble model output statistics. Mon. Wea. Rev., 143, 2909–2917, doi:10.1175/MWR-D-15-0095.1.

Klausern, Z., H. Kaplan, and E. Fattal, 2009: The similar days method for predicting near surface wind vectors. Meteor. Appl., 16, 569–579, doi:10.1002/met.158.

Lerch, S., and S. Baran, 2015: Similarity-based semi-local estimation of EMOS models. arXiv.org, 31 pp. [Available online at http://arxiv.org/abs/1509.03521.]

Leutbecher, M., and T. N. Palmer, 2008: Ensemble forecasting. J. Comput. Phys., 227, 3515–3539, doi:10.1016/j.jcp.2007.02.014.

Matheson, J. E., and R. L. Winkler, 1976: Scoring rules for continuous probability distributions. Manage. Sci., 22, 1087–1096, doi:10.1287/mnsc.22.10.1087.

Messner, J. W., and G. J. Mayr, 2011: Probabilistic forecasts using analogs in the idealized Lorenz96 setting. Mon. Wea. Rev., 139, 1960–1971, doi:10.1175/2010MWR3542.1.

Möller, A., A. Lenkoski, and T. L. Thorarinsdottir, 2013: Multivariate probabilistic forecasting using ensemble Bayesian model averaging and copulas. Quart. J. Roy. Meteor. Soc., 139, 982–991, doi:10.1002/qj.2009.

Nelsen, R. B., 2006: An Introduction to Copulas. 2nd ed. Springer, 272 pp.

Palmer, T. N., 2002: The economic value of ensemble forecasts as a tool for risk assessment: From days to decades. Quart. J. Roy. Meteor. Soc., 128, 747–774, doi:10.1256/0035900021643593.
Pinson, P., 2012: Adaptive calibration of \((u, v)\)-wind ensemble forecasts. *Quart. J. Roy. Meteor. Soc.*, 138, 1273–1284, doi:10.1002/qj.1873.

——, 2013: Wind energy: Forecasting challenges for its operational management. *Stat. Sci.*, 28, 564–585, doi:10.1214/13-STS445.

——, and J. Tastu, 2013: Discrimination ability of the Energy score. Technical University of Denmark Tech. Rep., 16 pp. [Available online at http://orbit.dtu.dk/files/56966842/tr13_15_Pinson_Tastu.pdf.]

——, H. Madsen, H. A. Nielsen, G. Papaefthymiou, and B. Klöckl, 2009: From probabilistic forecasts to statistical scenarios of short-term wind power production. *Wind Energy*, 12, 51–62, doi:10.1002/we.284.

Raftery, A. E., T. Gneiting, F. Balabdaoui, and M. Polakowski, 2005: Using Bayesian model averaging to calibrate forecast ensembles. *Mon. Wea. Rev.*, 133, 1155–1174, doi:10.1175/MWR2906.1.

R Core Team, 2014: R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. [Available online at http://www.R-project.org.]

Robertson, D. E., D. L. Shrestha, and Q. J. Wang, 2013: Post-processing rainfall forecasts from numerical weather prediction models for short-term streamflow forecasting. *Hydrol. Earth Syst. Sci.*, 17, 3587–3603, doi:10.5194/hess-17-3587-2013.

Roulin, E., and S. Vannitsem, 2012: Post-processing of ensemble precipitation predictions with extended logistic regression based on hindcasts. *Mon. Wea. Rev.*, 140, 874–888, doi:10.1175/MWR-D-11-00062.1.

Rüschendorf, L., 2009: On the distributional transform, Sklar’s theorem, and the empirical copula process. *J. Stat. Plann. Inference*, 139, 3921–3927, doi:10.1016/j.jspi.2009.05.030.

Schaake, J. C., and Coauthors, 2007: Precipitation and temperature ensemble forecasts from single-valued forecasts. *Hydrol. Earth Syst. Sci.*, 4, 655–717, doi:10.5194/hessd-4-655-2007.

——, and Coauthors, 2010: Summary of recommendations of the first Workshop on Postprocessing and Downscaling Atmospheric Forecasts for Hydrologic Applications held at Météo-France, Toulouse, France, 15–18 June 2009. *Atmos. Sci. Lett.*, 11, 59–63, doi:10.1002/asl.267.

Scheffzik, R., 2015a: Multivariate discrete copulas, with applications in probabilistic weather forecasting. *Ann. I.S.U.P. Publ. Inst. Stat. Univ. Paris*, 5, 87–116.

——, 2015b: Physically coherent probabilistic weather forecasts using multivariate discrete copula-based ensemble post-processing methods. Ph.D. thesis, Heidelberg University, 204 pp. [Available online at http://archiv.ub.uni-heidelberg.de/volltextserver/18028/]

——, T. L. Thorarinsdottir, and T. Gneiting, 2013: Uncertainty quantification in complex simulation models using ensemble copula coupling. *Stat. Sci.*, 28, 616–640, doi:10.1214/13-STST443.

Scheuerer, M., and T. M. Hamill, 2015: Variogram-based proper scoring rules for probabilistic forecasts of multivariate quantities. *Mon. Wea. Rev.*, 143, 1321–1334, doi:10.1175/MWR-D-14-00269.1.

Schölzel, C., and P. Friederichs, 2008: Multivariate non-normally distributed random variables in climate research—Introduction to the copula approach. *Nonlinear Processes Geophys.*, 15, 761–772, doi:10.5194/npg-15-761-2008.

——, and A. Hense, 2011: Probabilistic assessment of regional climate change in southwest Germany by ensemble dressing. *Climate Dyn.*, 36, 2003–2014, doi:10.1007/s00382-010-0815-1.

Schuhen, N., T. L. Thorarinsdottir, and T. Gneiting, 2012: Ensemble model output statistics for wind vectors. *Mon. Wea. Rev.*, 140, 3204–3219, doi:10.1175/MWR-D-12-00028.1.

Sklar, A., 1959: Fonctions de répartition à n dimensions et leurs marges. *Publ. Inst. Stat. Univ. Paris*, 8, 229–231.

Slaughter, J. M., T. Gneiting, and A. E. Raftery, 2013: Probabilistic wind vector forecasting using ensembles and Bayesian model averaging. *Mon. Wea. Rev.*, 141, 2107–2119, doi:10.1175/MWR-D-12-00002.1.

Talagrand, O., R. Vautard, and B. Strauss, 1997: Evaluation of probabilistic prediction systems. *Proc. Workshop on Predictability*, Reading, United Kingdom, ECMWF, 1–25.

Thorarinsdottir, T. L., M. Scheuerer, and C. Heinz, 2016: Assessing the calibration of high-dimensional ensemble forecasts using rank histograms. *J. Comput. Graph. Stat.*, 25, 105–122, doi:10.1080/10618600.2014.977447.

Van Schaeybroeck, B., and S. Vannitsem, 2015: Ensemble post-processing using member-by-member approaches: Theoretical aspects. *Quart. J. Roy. Meteor. Soc.*, 141, 807–818, doi:10.1002/qj.2397.

Vanyvye, E., L. Delle Monache, A. J. Monaghan, and J. O. Pinto, 2015: Wind resource estimates with an analog ensemble approach. *Renew. Energy*, 74, 761–773, doi:10.1016/j.renene.2014.08.060.

Verkade, J. S., J. D. Brown, P. Reggiani, and A. H. Weerts, 2013: Post-processing ECMWF precipitation and temperature ensemble reforecasts for operational hydrologic forecasting at various spatial scales. *J. Hydrol.*, 501, 73–91, doi:10.1016/j.jhydrol.2013.07.039.

Voisin, N., J. C. Schaake, and D. P. Lettenmeier, 2010: Calibration and downscaling methods for quantitative ensemble precipitation forecasts. *Wea. Forecasting*, 25, 1603–1627, doi:10.1175/2010WAF2222367.1.

——, F. Pappenberger, D. P. Lettenmeier, R. Buizza, and J. C. Schaake, 2011: Application of a medium-range global hydrologic probabilistic forecast scheme to the Ohio River basin. *Wea. Forecasting*, 26, 425–446, doi:10.1175/WAF-D-10-0032.1.

Vrac, M., and P. Friederichs, 2015: Multivariate–intervariable, spatial, and temporal—Bias correction. *J. Climate*, 28, 218–237, doi:10.1175/JCLI-D-14-00059.1.

Wilks, D. S., 2011: *Statistical Methods in the Atmospheric Sciences*. 3rd ed. Elsevier, 676 pp.

——, 2015: Multivariate ensemble Model Output Statistics using empirical copulas. *Quart. J. Roy. Meteor. Soc.*, 141, 945–952, doi:10.1002/qj.2414.

Yuen, R. A., T. Gneiting, T. L. Thorarinsdottir, and C. Fraley, 2013: ensembleMOS: Ensemble model output statistics. R package version 0.7. [Available online at http://CRAN.R-project.org/package=ensembleMOS.]