Ruprecht-Karls-Universität Heidelberg

Fakultät für Mathematik und Informatik

Ensemble Copula Coupling

Diplomarbeit

von

Roman Schefzik

Betreuer: Prof. Dr. Tilmann Gneiting
Dr. Thordis Thorarinsdottir

März 2011

Abstract

In modern meteorology, weather forecasts are often constructed from ensemble prediction systems which consist of multiple runs of dynamical numerical weather prediction models differing in the initial conditions and/or in the details of the parameterised numerical representation of the atmosphere. Statistical postprocessing of the ensemble forecasts is often required to realise their full potential in the sense that biases and dispersion errors need to be addressed. Several ensemble postprocessing methods have been proposed, yet state-of-the-art approaches only apply to a single weather quantity, at a single location and for a single prediction horizon. In many applications, however, there is a critical need to account for spatial, temporal and cross-variate dependence structures within the ensemble. To address this, we propose a novel tool called ensemble copula coupling (ECC), a copula-based multivariate statistical postprocessing technique. ECC is a two-step procedure in which existing univariate postprocessing methods are employed in a first step to obtain calibrated and sharp forecasts for each location, variable and look-ahead time separately. In a second step, the resulting univariate distributions are aggregated in a discrete copula approach, where the multivariate rank dependence pattern is inherited from the unprocessed raw ensemble, thereby capturing the flow dependency. In this thesis, we present the ECC technique, study its relationships to discrete copulas and apply it at major airports in Germany using the 50-member European Centre for Medium-Range Weather Forecasts (ECMWF) ensemble.

Zusammenfassung

In der modernen Meteorologie werden Wettervorhersagen oft anhand von Ensemblevorhersagesystemen erstellt, welche aus mehreren Durchläufen bestehen, die sich hinsichtlich der Anfangsbedingungen und/oder der Details der parametrisierten numerischen Darstellung der Atmosphäre unterscheiden. Oft benötigt man eine statistische Aufbereitung der Ensemblevorhersagen, um deren volle Leistungsfähigkeit in dem Sinn, dass systematische und Dispersionsfehler berücksichtigt werden müssen, zu erfassen. Es sind bereits einige Nachbearbeitungsmethoden für Ensembles veröffentlicht worden, jedoch sind die gegenwärtigen Ansätze nur für einzelne Wettergrößen, an einzelnen Stationen und für einzelne Vorhersagehorizonte anwendbar. In vielen Anwendungen ist es aber von entscheidender Bedeutung, Abhängigkeitsstrukturen in Raum, Zeit und unter den Wettergrößen innerhalb des Ensembles zu berücksichtigen. Um dies zu bewerkstelligen, präsentieren wir ein neuartiges Werkzeug namens Ensemble Copula Coupling (ECC), eine copulabasierte multivariate statistische Aufbereitungstechnik. ECC ist ein zweistufiges Verfahren, in welchem zunächst bestehende univariate Nachbearbeitungsmethoden verwendet werden, um kalibrierte und scharfe Vorhersagen für jeden Ort, jede Variable und jeden Vorhersagehorizont getrennt zu erzielen. In einem zweiten Schritt werden die erhaltenen univariaten Verteilungen in einem diskreten Copulaansatz zusammengeführt, wobei das multivariate Rangabhängigkeitsmuster vom ursprünglichen Ensemble geerbt wird, wodurch die Datenabhängigkeit erfasst wird. In dieser Arbeit stellen wir die ECC-Methode vor, untersuchen ihre Verbindung zu diskreten Copulas und wenden sie auf Hauptflughäfen in Deutschland an, wobei wir das 50 Mitglieder umfassende Ensemble des Europäischen Zentrums für mittelfristige Wettervorhersage (EZMW) benutzen.

Contents

Chapter 1

Introduction and motivation

It has always been an ambitious goal in meteorology to provide accurate forecasts for a variety of weather quantities like temperature, precipitation or wind speed, for instance. This is very interesting and important for numerous reasons such as

- energy generation through wind power plants or solar technology,
- organisation of one's leisure time activities including travel, sports or sailing,
- warnings about natural catastrophes like storms, droughts or inundations and
- economic and financial purposes, consider for example weather derivatives,

to name just a few.

In practice, there are various approaches to achieve weather predictions. Until the early 1990s, weather forecasting was regarded as a purely deterministic issue in the sense that carefully developed dynamical and numerical weather prediction models were implemented in order to get deterministic forecasts of future atmospheric states. Thereby, the idea is that for a certain set of "best" input data, one "best" weather forecast is produced. Although those methods are still used today, a radical change in the development of weather forecasting has taken place during the last twenty years. Weather prediction has been transformed by the implementation of so-called ensemble forecasts or ensemble prediction systems.
An ensemble forecast consists of multiple runs, typically between five and one hundred, of dynamical numerical weather forecast models differing in the initial conditions and/or the details of the parameterised numerical representation of the atmosphere, compare [12]. However, the fact that model biases and dispersion errors need to be addressed requires of statistical postprocessing for the model output of an ensemble forecast.

Together with statistical postprocessing, ensembles lead to so-called probabilistic forecasts, which means that we obtain a predictive probability distribution over future weather quantities. Fortunately, statistically postprocessed ensemble forecasts turn out

to be much better calibrated than the raw ensemble forecasts. Besides sharpness, calibration is an essential criterion for a "good" probabilistic forecast, see [14]. In many applications, not necessarily in meteorology, probabilistic forecasts form an improvement compared to purely deterministic point forecasts, like for example in weather-risk finance or disease modelling, see [42]. A non-technical introduction to probabilistic forecasting with applications can be found in [17].
During the past few years, several statistical postprocessing methods have been developed for and applied to several weather variables such as temperature, precipitation, sea level pressure, wind speed, wind direction or wind vectors. Common state-of-the-art approaches are

- Bayesian model averaging (BMA), compare [46], [54], [55] and [1],
- ensemble model output statistics (EMOS), see [13] and [57],
- logistic regression, compare [18], [19], [60] and [61], and
- quantile regression, see [4].

These techniques provide calibrated univariate probabilistic forecasts. Unfortunately, they can only be used for a single weather quantity at a single location and for a single look-ahead time. However, for many applications it is exceedingly important that postprocessed forecast fields have physically coherent multivariate dependence structures across space, time and variables. In the field of air traffic management, for example, it would be very helpful if there was a tool ensuring physical consistency of statistically postprocessed forecasts simultaneously at several airport stations, for several prediction horizons and for several weather quantities like precipitation and wind speed. Besides air traffic control, we can think of flood management and ship routeing as further applications. Thereby, we should notice that, in reality, we have to face an extremely high dimensionality in our challenge. For example, if we consider a probabilistic weather forecast on a three-dimensional grid over Europe, we might have to work with some

$$\underbrace{500 \times 500}_{\text{horizontal grid}} \times \underbrace{30}_{\text{vertical grid}} \times \underbrace{72}_{\text{look-ahead times}} \times \underbrace{10}_{\text{weather variables}} = 9\,000\,000$$

univariate distributions or individual forecasts, respectively.
There already exist some techniques handling spatial dependence structures, compare [10], [2] and [3]. However, these methods cannot be used to describe intervariable or temporal dependencies.

To solve the problem exposed above, a novel approach called ensemble copula coupling (ECC) is introduced in this thesis.
For the first time, a part of the notion of ECC appeared in the context of the design of the Bayesian processor of ensemble (BPE) proposed by Krzysztofowicz and Toth, compare [30], and Bremnes, see [5]. In this thesis, we develop this very useful idea and connect it to the field of copulas, which is presented in a general frame by Nelsen in [40],

and discrete copulas as discussed by Kolesárová et al. in [24], in particular.
Roughly speaking, ECC proceeds in two steps as follows: First, we use the well-known and established ensemble postprocessing methods previously mentioned to generate calibrated and sharp individual univariate probabilistic forecasts for each location, prediction horizon and weather quantity. Then, we employ our novel ECC concept to connect the obtained univariate distributions in a copula approach. Thereby, the multivariate dependencies are inherited from the unprocessed raw ensemble, and the flow dependency is captured.

The thesis at hand is organised as follows.
In Chapter 2, we outline the theoretical and methodological background of the ECC technique. First, we expose a general overview of the topical field of copulas, which is fast-paced and has many applications in the context of modelling dependencies. Then, we discuss the state-of-the-art ensemble postprocessing methods named above which lead to univariate probabilistic forecast distributions.
We then carry on with Chapter 3 starting with a description of the origins of ECC by studying the BPE, followed by initial examples and the theoretical development of our novel ECC method, which forms the key part of the thesis. Moreover, we show that ECC — as suggested by the name — can indeed be interpreted as a copula approach by pointing out its relationships to discrete copulas.
In Chapter 4, we present several tools to assess the predictive performance of probabilistic forecasting methods for both univariate and multivariate quantities as suggested in [16]. Some of them are employed to evaluate the ECC technique later on.
We continue with practical applications of the novel ECC approach in Chapter 5. ECC is applied to data provided by the European Centre for Medium-Range Weather Forecasts (ECMWF) in a case study. More precisely, ECC is employed for statistical postprocessing of ensemble forecasts based on the ECMWF ensemble consisting of 50 members for several weather quantities at the three locations Berlin, Hamburg and Frankfurt during a test period in the year 2010. Moreover, as mentioned before, we employ verification methods for multivariate quantities to assess the predictive performance of the ECC forecast fields and compare to other reference forecasts like the raw ensemble.
Finally, in Chapter 6, we give both a summary and a discussion of the results in this thesis. Furthermore, we provide an overview of topics, challenges and subjects of current research related to ECC and give an outlook for possible future work.

Chapter 2

Theoretical and methodological background

In this chapter, we develop some theoretical and methodological background facts connected to the ECC method, which will be introduced in Chapter 3 later on. First, we study copulas in general and collect the most important facts about them. Afterwards, we describe several ensemble postprocessing techniques that are employed to obtain calibrated and sharp univariate probabilistic forecasts for each weather quantity, location and look-ahead time separately, namely those mentioned in the introduction in the preceding chapter.

2.1 Copulas

It has always been an important and interesting task for statisticians to study the relationship between a multivariate distribution function and its lower dimensional marginals. After several mathematicians like Fréchet, Dall'Aglio and Hoeffding had worked on this topic by examining the bivariate and trivariate distribution functions with given univariate marginals, Sklar solved the problem for the univariate marginals case in 1959 by introducing a new class of functions which he called "copulas". Between 1959 and 1976, copulas mainly appeared in the development of the theory of so-called probabilistic metric spaces, especially in the study of binary operations on the space of probability distribution functions. Facts about the historical development of probabilistic metric spaces and information about this field in general can be found in [51]. Later, copulas were employed to define nonparametric measures of dependence between random variables. Since then, the concept of copula began to play a decisive role in solving problems in probability theory and mathematical statistics, particularly in questions about dependencies, given marginals and functions of random variables which are invariant under monotone transformations. Moreover, copulas were used in simulation studies, for example for generating a sample from a specified joint distribution. All those and also further annotations about the history of copulas can be seen in "An Introduction to Copulas" written by Nelsen, see [40]. This textbook is the most important summary

about the field of copulas, and this section essentially is based on it. Other references are explicitly stated.

2.1.1 Preliminaries

Before giving a rigorous mathematical definition of a copula, we introduce some notation as well as some basic definitions and facts.

For any $n \in \mathbb{N} = \{1, 2, 3, ...\}$, we let $\overline{\mathbb{R}}^n$ denote the extended n-space $\underbrace{\overline{\mathbb{R}} \times \cdots \times \overline{\mathbb{R}}}_{n \text{ times}}$, where $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$.
Furthermore, we use vector notation for points in $\overline{\mathbb{R}}^n$, for example $\mathbf{a} = (a_1, ..., a_n)$ or $\mathbf{b} = (b_1, ..., b_n)$, and write $\mathbf{a} \leq \mathbf{b}$ ($\mathbf{a} < \mathbf{b}$) if $a_k \leq b_k$ ($a_k < b_k$) for all $k \in \{1, ..., n\}$.
For $\mathbf{a} \leq \mathbf{b}$ let $[\mathbf{a}, \mathbf{b}]$ be the n-box $B = [a_1, b_1] \times \cdots \times [a_n, b_n]$, that is the Cartesian product of n closed intervals.
The vertices of an n-box B are the points $\mathbf{c} = (c_1, ..., c_n)$, where each c_k is equal to either a_k or b_k for $k \in \{1, ..., n\}$.
The unit n-cube $\mathbb{I}^n$ is the product $\underbrace{\mathbb{I} \times \cdots \times \mathbb{I}}_{n \text{ times}}$, where $\mathbb{I} = [0, 1]$ is the unit interval.
An n-place real function H is a function whose domain $\mathrm{Dom}(H)$ is a subset of $\overline{\mathbb{R}}^n$ and whose range $\mathrm{Ran}(H)$ is a subset of $\mathbb{R}$.

Example 2.1: For $n = 2$ we have the rectangle $B = [x_1, x_2] \times [y_1, y_2]$ in $\overline{\mathbb{R}}^2$ as a "2-box" with vertices $(x_1, y_1), (x_1, y_2), (x_2, y_1)$ and (x_2, y_2). Moreover, the unit "2-cube" is $\mathbb{I}^2$, the unit square. A 2-place real function H is a function with $\mathrm{Dom}(H) \subseteq \overline{\mathbb{R}}^2$ and $\mathrm{Ran}(H) \subseteq \mathbb{R}$.

Definition 2.2: Let $S_1, ..., S_n$ be nonempty subsets of $\overline{\mathbb{R}}$, let H be an n-place real function with $\mathrm{Dom}(H) = S_1 \times \cdots \times S_n$ and let $B = [\mathbf{a}, \mathbf{b}]$ be an n-box all of whose vertices are in $\mathrm{Dom}(H)$.
Then $V_H(B) = \sum_{\mathbf{c}} \mathrm{sgn}(\mathbf{c}) H(\mathbf{c})$ is called the H-volume of B, where the sum extends over all vertices $\mathbf{c}$ of B and

$$\mathrm{sgn}(\mathbf{c}) = \begin{cases} 1 & \text{if } c_k = a_k \text{ for an even number of } k\text{'s} \\ -1 & \text{if } c_k = a_k \text{ for an odd number of } k\text{'s,} \end{cases}$$

with $k \in \{1, ..., n\}$.

Remark 2.3: Equivalently, $V_H(B)$ is the n-th order difference of H on B, that is,

$$V_H(B) = \Delta_{\mathbf{a}}^{\mathbf{b}} H(\mathbf{t}) = \Delta_{a_n}^{b_n} \Delta_{a_{n-1}}^{b_{n-1}} ... \Delta_{a_1}^{b_1} H(\mathbf{t}),$$

where

$$\Delta_{a_k}^{b_k} H(\mathbf{t}) = H(t_1, ..., t_{k-1}, b_k, t_{k+1}, ..., t_n) - H(t_1, ..., t_{k-1}, a_k, t_{k+1}, ..., t_n).$$

Example 2.4:

1. For a 2-place real function H with domain $\overline{\mathbb{R}}^2$ and the rectangle $B = [x_1, x_2] \times [y_1, y_2]$, the H-volume of B is

$$V_H(B) = H(x_2, y_2) - H(x_2, y_1) - H(x_1, y_2) + H(x_1, y_1).$$

We note that for $\Delta_{x_1}^{x_2} H(x, y) = H(x_2, y) - H(x_1, y)$ and $\Delta_{y_1}^{y_2} H(x, y) = H(x, y_2) - H(x, y_1)$ we get

$$\begin{aligned} V_H(B) &= \Delta_{y_1}^{y_2} \Delta_{x_1}^{x_2} H(x, y) \\ &= \Delta_{y_1}^{y_2} (H(x_2, y) - H(x_1, y)) \\ &= H(x_2, y_2) - H(x_2, y_1) - H(x_1, y_2) + H(x_1, y_1). \end{aligned}$$

2. For a 3-place real function H with domain $\overline{\mathbb{R}}^3$ and the 3-box $B = [x_1, x_2] \times [y_1, y_2] \times [z_1, z_2]$, the H-volume of B is

$$\begin{aligned} V_H(B) &= H(x_2, y_2, z_2) - H(x_2, y_2, z_1) - H(x_2, y_1, z_2) - H(x_1, y_2, z_2) \\ &\quad + H(x_2, y_1, z_1) + H(x_1, y_2, z_1) + H(x_1, y_1, z_2) - H(x_1, y_1, z_1). \end{aligned}$$

Definition 2.5:

1. An n-place real function H is called n-increasing if $V_H(B) \geq 0$ for all n-boxes B whose vertices lie in $\mathrm{Dom}(H)$.

2. Suppose that $\mathrm{Dom}(H) = S_1 \times \cdots \times S_n$, where each S_k, $k \in \{1, ..., n\}$, has a least element a_k. Then, H is grounded if $H(\mathbf{t}) = 0$ for all $\mathbf{t} \in \mathrm{Dom}(H)$ such that $t_k = a_k$ for at least one k.
   H has margins if each S_k is nonempty and has a greatest element b_k.
   The one-dimensional margins of H are the functions H_k given by $\mathrm{Dom}(H_k) = S_k$ and $H_k(x) = H(b_1, ..., b_{k-1}, x, b_{k+1}, ..., b_n)$ for all $x \in S_k$.
   Note that we can define higher dimensional margins by fixing fewer places in H.

Example 2.6: Let H be the function with $\mathrm{Dom}(H) = [-1, 1] \times [0, \infty]$ given by

$$H(x, y) = \frac{(x+1)(e^y - 1)}{x + 2e^y - 1}.$$

Since $H(x, 0) = 0$ and $H(-1, y) = 0$, H is grounded. Moreover, H has the one-dimensional margins

$$\begin{aligned} H_1(x) &= H(x, \infty) = \lim_{y \to \infty} \frac{(x+1)(e^y - 1)}{x + 2e^y - 1} = \lim_{y \to \infty} \frac{xe^y - x + e^y - 1}{x + 2e^y - 1} \\ &= \lim_{y \to \infty} \frac{xe^y}{x + 2e^y - 1} - 0 + \lim_{y \to \infty} \frac{e^y}{x + 2e^y - 1} - 0 \overset{\text{L'Hospital}}{=} \lim_{y \to \infty} \frac{xe^y}{2e^y} + \lim_{y \to \infty} \frac{e^y}{2e^y} \\ &= \frac{x}{2} + \frac{1}{2} = \frac{x+1}{2} \end{aligned}$$

and

$$H_2(y) = H(1, y) = \frac{2(e^y - 1)}{2e^y} = 1 - \frac{1}{e^y} = 1 - e^{-y}.$$

With immediate effect, one-dimensional margins are called "margins", and for $k \geq 2$, we write "k-margins" for k-dimensional margins.

Definition 2.7: H is called non-decreasing in each argument if $(t_1, ..., t_{k-1}, x, t_{k+1}, ..., t_n), (t_1, ..., t_{k-1}, y, t_{k+1}, ..., t_n) \in \text{Dom}(H)$ and $x < y$ implies that $H(t_1, ..., t_{k-1}, x, t_{k+1}, ..., t_n) \leq H(t_1, ..., t_{k-1}, y, t_{k+1}, ..., t_n)$.

Lemma 2.8: If $S_1, ..., S_n$ are nonempty subsets of $\overline{\mathbb{R}}$ and if H is a grounded n-increasing function with Dom (H)=$S_1 \times \cdots \times S_n$, then H is nondecreasing in each argument.

Proof: This follows directly from Definition 2.2, Remark 2.3 and Definition 2.5.1. □

Lemma 2.9: Let $S_1, ..., S_n$ be nonempty subsets of $\overline{\mathbb{R}}$, let H be a grounded, n-increasing function with margins with domain $\text{Dom}(H)$=$S_1 \times \cdots \times S_n$, and let $\mathbf{x} = (x_1, ..., x_n)$ and $\mathbf{y} = (y_1, ..., y_n)$ be any points in $S_1 \times ... \times S_n$. Then,

$$|H(\mathbf{x}) - H(\mathbf{y})| \leq \sum_{k=1}^{n} |H_k(x_k) - H_k(y_k)|.$$

Proof: See [51]. □

2.1.2 Definition and examples

Due to the preliminaries in Subsection 2.1.1 we are now able to give a definition of the term "copula". Roughly speaking, copulas are functions that join or couple multivariate distribution functions to their one-dimensional marginal distribution functions. We can also say that a copula is a distribution function $C : \mathbb{I}^n \to \mathbb{I}$ whose one-dimensional margins are uniform on $\mathbb{I}$, that is to say $C(1, ..., 1, \underbrace{u_j}_{\text{position } j}, 1, ..., 1) = u_j$ for all $j \in \{1, ..., n\}$. Nevertheless, we still need a mathematically rigorous definition of copulas, which is given as follows.

Definition 2.10:

1. An n-dimensional subcopula, or also an n-subcopula, is a function $\tilde{C}$ with the following properties:
   - Dom $(\tilde{C})$=$S_1 \times \cdots \times S_n$, where each S_k, $k \in \{1, ..., n\}$, is a subset of $\mathbb{I}$ containing 0 and 1,
   - $\tilde{C}$ is grounded and n-increasing and

- $\tilde{C}$ has one-dimensional margins $\tilde{C}_k$, where $k \in \{1, ..., n\}$, which satisfy $\tilde{C}_k(u) = u$ for all $u \in S_k$.

We remark that $0 \leq \tilde{C}(\mathbf{u}) \leq 1$ for all $\mathbf{u} \in \mathrm{Dom}(\tilde{C})$ implies that $\mathrm{Ran}(\tilde{C}) \subseteq \mathbb{I}$.

2. An n-dimensional copula, or also an n-copula, is an n-subcopula whose domain is $\mathbb{I}^n$.

Remark 2.11: Equivalently, an n-copula is a function $C : \mathbb{I}^n \to \mathbb{I}$ such that

- $C(\mathbf{u}) = 0$ if at least one coordinate of $\mathbf{u}$ is 0 and $C(\mathbf{u}) = u_k$ if all coordinates of $\mathbf{u}$ are 1 except u_k for all $\mathbf{u} \in \mathbb{I}^n$, where $k \in \{1, ..., n\}$.
- $V_C([\mathbf{a}, \mathbf{b}]) \geq 0$ for all $\mathbf{a}, \mathbf{b} \in \mathbb{I}^n$ with $\mathbf{a} \leq \mathbf{b}$.

Remark 2.12: It can be shown that for any n-copula C, where $n \geq 3$, each k-margin of C is a k-copula, where $2 \leq k < n$.

We now give several examples for copulas, which can be found in [40] and [44].

Example 2.13:

1. Let C be a function defined by $C(u, v, w) = w \cdot \min\{u, v\}$.

   - By the definition of C it can be easily seen that $C(u, v, w) = 0$ if at least one coordinate, namely u, v or w, is 0 and that
   $$C(u, v, w) = \begin{cases} u & \text{if } v, w = 1, \\ v & \text{if } u, w = 1, \\ w & \text{if } u, v = 1, \end{cases}$$
   for every $u, v, w \in \mathbb{I}$.
   - The C-volume of the 3-box $B = [a_1, b_1] \times [a_2, b_2] \times [a_3, b_3]$, where $a_k, b_k \in \mathbb{I}$, $a_k \leq b_k$ and $k = 1, 2, 3$, is
   $V_C(B) = \Delta_{a_3}^{b_3} \Delta_{a_2}^{b_2} \Delta_{a_1}^{b_1} C(u, v, w) = \underbrace{(b_3 - a_3)}_{\geq 0 (b_3 \geq a_3)} \underbrace{\Delta_{a_2}^{b_2} \Delta_{a_1}^{b_1} \min\{u, v\}}_{\geq 0} \geq 0.$

   Therefore, according to Remark 2.11, C is indeed a 3-copula.
   The 2-margins of C are
   $C_{1,2}(u, v) = C(u, v, 1) = \min\{u, v\}$,
   $C_{1,3}(u, w) = C(u, 1, w) = w \cdot \min\{u, 1\}$ and
   $C_{2,3}(v, w) = C(1, v, w) = w \cdot \min\{1, v\}$.
   These are all 2-copulas, compare Remark 2.12.

2. $\Pi^n(\mathbf{u}) = \prod_{i=1}^{n} u_i$ is an n-copula, the so-called product copula.

3. For $\alpha, \beta \in \mathbb{I}$, the function $C_{\alpha,\beta}(u, v) = \min\{u^{1-\alpha} v, u v^{1-\beta}\}$ is a two-parameter family of 2-copulas, the so-called Marshall-Olkin family.

4. For $\theta \in [-1, 1]$, the function $C_\theta(u, v) = uv + \theta uv(1-u)(1-v)$ is a one-parameter family of 2-copulas, the so-called Farlie-Gumbel-Morgenstern family.

2.1.3 Sklar's theorem and other properties

Since we now have a rigorous definition of copulas, we want to collect several essential properties about them, including Sklar's theorem, the most important one.

Uniform continuity and partial derivatives

We start with the uniform continuity of n-subcopulas and hence n-copulas.

Theorem 2.14: For every n-subcopula $\tilde{C}$ and every $\mathbf{u}, \mathbf{v} \in \text{Dom}\ (\tilde{C})$, the inequality

$$|\tilde{C}(\mathbf{v}) - \tilde{C}(\mathbf{u})| \leq \sum_{k=1}^{n} |v_k - u_k|$$

holds. Hence, $\tilde{C}$ is uniformly continuous on its domain.

Proof: This follows immediately from Lemma 2.9. □

We continue with a theorem about the partial derivatives of n-copulas.

Theorem 2.15: Let C be an n-copula. Then,

1. $\frac{\partial C(\mathbf{u})}{\partial u_i}$ exists for all $\mathbf{u} \in \mathbb{I}^n$, $i \in \{1, ..., n\}$, almost everywhere on $\mathbb{I}$, in the sense of Lebesgue measure,
2. $0 \leq \frac{\partial C(\mathbf{u})}{\partial u_i} \leq 1$ for all $i \in \{1, ..., n\}$ and
3. the functions $C_{u_i}(\mathbf{u}) = \frac{\partial C(\mathbf{u})}{\partial u_i}$ are non-decreasing for all $u_i \neq u_j$, $j \in \{1, ..., n\} \setminus \{i\}$, almost everywhere on $\mathbb{I}$.

Proof: A proof for the bivariate case can be found in [40]. Since all relevant results for the proof also hold for n dimensions, where $n \geq 3$, it is straightforward to verify the n-dimensional analogon, compare also Section 2.2 in [49]. □

Sklar's theorem

Before stating the central theorem of this section, we need the following definitions and remarks

Definition 2.16: A function H with Dom $(H)=\overline{\mathbb{R}}^n$ such that

1. H is n-increasing and

2. $H(\mathbf{t}) = 0$ for all $\mathbf{t} \in \overline{\mathbb{R}}^n$ such that $t_k = -\infty$ for at least one $k \in \{1, ..., n\}$ and $H(\infty, ..., \infty) = 1$

is called an n-dimensional distribution function.

Remark 2.17: If H is grounded and Dom $(H)=\overline{\mathbb{R}}^n$, then, according to Lemma 2.8, the one-dimensional margins of an n-dimensional distribution function are distribution functions. They are denoted by $F_1, ..., F_n$.

Definition 2.18: Let F be a distribution function. A function $F^{(-1)}$ with domain $\mathbb{I}$ such that

1. $t \in \text{Ran}(F)$ implies that $F^{(-1)}(t) = x \in \overline{\mathbb{R}}$ with $F(x) = t$, that is, $F(F^{(-1)}(t)) = t$ for all $t \in \text{Ran}\ (F)$, and
2. $t \notin \text{Ran}(F)$ implies that $F^{(-1)}(t) = \inf\{x|F(x) \geq t\} = \sup\{x|F(x) \leq t\}$

is called a quasi-inverse of F.

Remark 2.19: If F is strictly increasing, there is a single quasi-inverse, which is the ordinary inverse denoted by F^{-1}.

We are now ready to formulate Sklar's theorem, which explains the role copulas play in the relationship between multivariate distribution functions and their univariate margins. Most of the applications in the field of copulas are based on this result whose tremendous importance is therefore obvious.

Theorem 2.20: (Sklar)
Let H be an n-dimensional distribution function with margins $F_1, ..., F_n$.
Then, there is an n-copula C such that

$$H(x_1, ..., x_n) = C(F_1(x_1), ..., F_n(x_n)) \text{ for all } \mathbf{x} \in \overline{\mathbb{R}}^n. \tag{2.1}$$

If $F_1, ..., F_n$ are all continuous, then C is unique. Otherwise, C is uniquely determined on $\text{Ran}(F_1) \times \cdots \times \text{Ran}(F_n)$.
Conversely, if C is an n-copula and $F_1, ..., F_n$ are distribution functions, then the function H defined by (2.1) is an n-dimensional distribution function with margins $F_1, ..., F_n$.

Proof: For the case of $n = 2$, a complete proof can be found in Section 2.3. in [40]. The proof of the n-dimensional case, where $n \geq 3$, proceeds in a similar way by showing the n-dimensional versions of Lemma 2.3.4 and Lemma 2.3.5, the so-called "extension lemma", in [40]. □

Corollary 2.21: Let $H, C, F_1, ..., F_n$ be as in Theorem 2.20, and let $F_1^{(-1)}, ..., F_n^{(-1)}$ be quasi-inverses of $F_1, ..., F_n$, respectively. Then,

$$C(u_1, ..., u_n) = H(F_1^{(-1)}(u_1), ..., F_n^{(-1)}(u_n))$$

for all $\mathbf{u} \in \mathbb{I}^n$.

We note that no probabilistic elements like for example random variables are mentioned in the definitions and results up to now. In particular, the definition of distribution functions is non-probabilistic.
In a probabilistic sense we say that F is the distribution function of the random variable X if $F_X(x) = \mathbb{P}[X \leq x]$ for all $x \in \overline{\mathbb{R}}$ and we define F to be right-continuous. For n random variables $X_1, ..., X_n$ we also have the common definition for a joint distribution function F, namely $F_{\mathbf{X}}(\mathbf{x}) = \mathbb{P}[X_1 \leq x_1, ..., X_n \leq x_n]$ for all $\mathbf{x} \in \overline{\mathbb{R}}^n$.
All results obtained before also hold for the probabilistic definition of distribution functions, regardless of any additional restrictions that may be imposed. We always assume that all random variables can be defined on a common probability space $(\Omega, \mathcal{A}, P)$.
In particular, there exists a version of Sklar's theorem for random variables.
Before stating this theorem, we recall the following well-known fact from probability theory.

Theorem 2.22: Let X be a random variable with values in $\mathbb{R}$ and continuous distribution function F. Then $U := F(X) \sim \mathcal{U}_{\mathbb{I}}$, where $\mathcal{U}_{\mathbb{I}}$ denotes the uniform distribution on $\mathbb{I}$.

We now present a version of Sklar's theorem in terms of random variables.

Theorem 2.23: (Sklar's theorem for random variables)
Let $X_1, ..., X_n$ be random variables defined on a common probability space with distribution functions $F_1, ..., F_n$, respectively, and joint distribution function H.
Then, there exists an n-copula C such that

$$H(x_1, ..., x_n) = C(F_1(x_1), ..., F_n(x_n)) \text{ for all } \mathbf{x} \in \overline{\mathbb{R}}^n. \tag{2.2}$$

If $F_1, ..., F_n$ are all continuous, C is unique. Otherwise, C is uniquely determined on $\text{Ran}(F_1) \times \cdots \times \text{Ran}(F_n)$.

Corollary 2.24: Let $H, C, F_1, ..., F_n$ be as in Theorem 2.23, and let $F_1^{(-1)}, ..., F_n^{(-1)}$ be quasi-inverses of $F_1, ..., F_n$, respectively.
Then,

$$C(u_1, ..., u_n) = H(F_1^{(-1)}(u_1), ..., F_n^{(-1)}(u_n))$$

for all $\mathbf{u} \in \mathbb{I}^n$.

For reasons of simplicity, we assume that $F_1, ..., F_n$ are continuous and differentiable distribution functions. Then, according to Sklar's theorem, C is unique and can also be expressed by

$$C(u_1, ..., u_n) = \int_0^{u_1} \dots \int_0^{u_n} c(v_1, ..., v_n) dv_1 \dots dv_n,$$

where $u_j = F_j(x_j)$ and c is the corresponding copula density. Hence, for continuous random variables $X_1, ..., X_n$ with distribution function
$H(x_1, ..., x_n) = C(F_1(x_1), ..., F_n(x_n))$ we have the joint probability density

$$f(x_1, ..., x_n) = c(F_1(x_1), ..., F_n(x_n)) \cdot f_1(x_1) \cdot ... \cdot f_n(x_n),$$

where

$$c(F_1(x_1), ..., F_n(x_n)) = \frac{\partial}{\partial F_1(x_1)} ... \frac{\partial}{\partial F_n(x_n)} C(F_1(x_1), ..., F_n(x_n))$$

and $f_j(x_j) = \frac{dF_j(x_j)}{dx_j}$ for all $j \in \{1, ..., n\}$, compare Section 3 in [50].

Fréchet-Hoeffding bounds

We now search for both a lower and an upper bound for every copula C. As we will see, the functions

$$M^n(\mathbf{u}) = \min\{u_1, ..., u_n\} \text{ and } W^n(\mathbf{u}) = \max\{u_1 + ... + u_n - n + 1, 0\}$$

play an important role in finding the answer to that question.

Theorem 2.25: (Fréchet-Hoeffding bounds inequality, copula version)
If $\tilde{C}$ is any n-subcopula, then

$$W^n(\mathbf{u}) \leq \tilde{C}(\mathbf{u}) \leq M^n(\mathbf{u})$$

for all $\mathbf{u} \in \text{Dom}(\tilde{C})$.

Proof: This follows directly from Lemmas 2.8 and 2.9. □

Remark 2.26:

1. W^n is called the Fréchet-Hoeffding lower bound, M^n is called the Fréchet-Hoeffding upper bound.
2. The function M^n is a copula for all $n \geq 2$.
3. The function W^n is a copula for $n = 2$, but never for $n \geq 3$.

The left-hand inequality in Theorem 2.25 is "best-possible" in the following sense.

Theorem 2.27: For any $n \geq 3$ and any $\mathbf{u} \in \mathbb{I}^n$, there is an n-copula $C = C(\mathbf{u})$ such that $C(\mathbf{u}) = W^n(\mathbf{u})$.

Proof: Compare the proof of Theorem 2.10.13 in [40]. □

An interesting property of the Fréchet-Hoeffding upper bound M^n is the following one.

Theorem 2.28: For $n \geq 2$, each of the continuous random variables $X_1, ..., X_n$ is almost surely a strictly increasing function of any of the others if and only if the n-copula of $X_1, ..., X_n$ is M^n.

For further properties of the Fréchet-Hoeffding bounds for joint distribution functions in the case of $n = 2$, we refer to Section 2.5 in [40].

Independence of random variables

We have already defined the product copula in Example 2.13.2., namely $\Pi^n(\mathbf{u}) = \prod_{i=1}^n u_i$. The following theorem provides a characterisation of independent random variables via the product copula.

Theorem 2.29: For $n \geq 2$, let $X_1, ..., X_n$ be continuous random variables with distribution functions $F_1, ..., F_n$, respectively, and joint distribution function H.
Then, $X_1, ..., X_n$ are independent if and only if the n-copula of $X_1, ..., X_n$ is Π^n.

Proof: This follows from Sklar's theorem for random variables, see Theorem 2.23, and the fact that $X_1, ..., X_n$ are independent if and only if $H(x_1, ..., x_n) = F_1(x_1) \cdot ... \cdot F_n(x_n)$ for all $x_k \in \overline{\mathbb{R}}$, where $k \in \{1, ..., n\}$. □

Monotone transformations

A reason for the use of copulas in the study of nonparametric statistics is that, for strictly monotone transformations of the random variables, copulas are either invariant or change in predictable ways. This is expressed in the following theorem in which we limit ourselves to the case of $n = 2$. Remember that for a continuous distribution function of a random variable X and a strictly monotone function α whose domain contains $\mathrm{Ran}(X)$ the distribution function of the random variable $\alpha(X)$ is continuous as well.

Theorem 2.30: (monotone transformations)
Let X_1 and X_2 be continuous random variables with copula $C_{X_1X_2}$ and let α and β be strictly monotone functions on $\mathrm{Ran}(X_1)$ and $\mathrm{Ran}(X_2)$, respectively.

1. If α and β are strictly increasing, then
   $C_{\alpha(X_1)\beta(X_2)}(x_1, x_2) = C_{X_1X2}(x_1, x_2)$.
2. If α is strictly increasing and β is strictly decreasing, then
   $C_{\alpha(X_1)\beta(X_2)}(x_1, x_2) = x_1 - C_{X_1X2}(x_1, 1 - x_2)$.
3. If α is strictly decreasing and β is strictly increasing, then
   $C_{\alpha(X_1)\beta(X_2)}(x_1, x_2) = x_2 - C_{X_1X2}(1 - x_1, x_2)$.
4. If α and β are strictly decreasing, then
   $C_{\alpha(X_1)\beta(X_2)}(x_1, x_2) = x_1 + x_2 - 1 - C_{X_1X2}(1 - x_1, 1 - x_2)$.

Remark 2.31: In Theorems 2.6 and 2.7 proposed in [9], the generalisation of the monotone transformations to the more complex multivariate case is exposed and proved. Hence, the two-dimensional statements we consider here follow immediately as a special case.

2.1.4 Construction of copulas

In this subsection, we deal with the methods of constructing copulas explicitly.
If we have a collection of copulas, then, according to Sklar's theorem, we have one of bi- or multivariate distributions with whatever marginal distribution we desire. This fact is very useful in the fields of modelling and simulation.
We start with the presentation of three general methods to construct bivariate copulas, that is, we consider the case of $n = 2$, and then continue with the construction of multivariate n-copulas, where $n \geq 3$.

Construction of bivariate copulas

Essentially, there are three general approaches to construct bivariate copulas.

1. Inversion method

The inversion method is based on Sklar's theorem and the corresponding corollary. Copulas are produced directly from joint distributions. Given a bivariate distribution function $H(u, v)$ with continuous margins $F(u)$ and $G(v)$, inverting according to the corollary of Sklar's theorem leads to the copula $C(u, v) = H(F^{(-1)}(u), G^{(-1)}(v))$. With this copula C, we can construct new bivariate distributions $\tilde{H}$ with arbitrary margins, say $\tilde{F}$ and $\tilde{G}$. Due to Sklar's theorem, we have $\tilde{H}(x, y) = C(\tilde{F}(x), \tilde{G}(y))$.
We now give two explicit examples for the inversion method to construct copulas.

Example 2.32: It can be verified that the function

$$H(x,y) = \begin{cases} \frac{(x+1)(e^y-1)}{x+2e^y-1} & \text{for } (x,y) \in [-1,1] \times [0,\infty], \\ 1 - e^{-y} & \text{for } (x,y) \in (1,\infty) \times [0,\infty], \\ 0 & \text{elsewhere,} \end{cases}$$

is indeed a distribution function. Its margins F and G are given by

$$F(x) = \begin{cases} 0 & \text{for } x < -1, \\ \frac{x+1}{2} & \text{for } x \in [-1,1], \\ 1 & \text{for } x > 1, \end{cases}$$

and

$$G(y) = \begin{cases} 0 & \text{for } y < 0, \\ 1 - e^{-y} & \text{for } y \geq 0. \end{cases}$$

The quasi-inverses of F and G are

$$F^{(-1)}(u) = \inf\{x|F(x) \geq u\} = \inf\left\{x \,\middle|\, \frac{x+1}{2} \geq u\right\} = \inf\{x|x \geq 2u-1\} = 2u-1$$

and

$$\begin{aligned} G^{(-1)}(v) &= \inf\{y|G(y) \geq v\} = \inf\{y|1-e^{-y} \geq v\} = \inf\{y|e^y(1-v) \geq 1\} \\ &= \inf\left\{y \,\middle|\, e^y \geq \frac{1}{1-v}\right\} = \inf\left\{y \,\middle|\, y \geq \log\left(\frac{1}{1-v}\right)\right\} = \log\left(\frac{1}{1-v}\right) \\ &= -\log(1-v) \end{aligned}$$

for $u, v \in \mathbb{I}$. Since $\mathrm{Ran}(F) = \mathrm{Ran}\ (G) = \mathbb{I}$, we obtain the copula

$$\begin{aligned} C(u,v) &= \frac{(2u-1+1)(e^{-\log(1-v)}-1)}{2u-1+2e^{-\log(1-v)}-1} = \frac{2u\left(\frac{1}{1-v}-1\right)}{2u-2+\frac{2}{1-v}} = \frac{u\left(\frac{1}{1-v}-1\right)}{u-1+\frac{1}{1-v}} \\ &= \frac{\frac{uv}{1-v}}{\frac{(u-1)(1-v)+1}{1-v}} = \frac{uv(1-v)}{(1-v)(u-uv+v-1+1)} = \frac{uv}{u+v-uv}. \end{aligned}$$

Example 2.33: Gumbel's bivariate exponential function is given by the joint distribution function

$$H_\theta(x,y) = \begin{cases} 1-e^{-x}-e^{-y}+e^{-(x+y+\theta xy)} & \text{for } x,y \geq 0, \\ 0 & \text{otherwise}, \end{cases}$$

where $\theta \in \mathbb{I}$.
The margins F and G are exponentials that have quasi-inverses $F^{(-1)}(u) = -\log(1-u)$ and $G^{(-1)}(v) = -\log(1-v)$ for $u, v \in \mathbb{I}$. Hence, we obtain the copula

$$\begin{aligned} C_\theta(u,v) &= 1-e^{-(-\log(1-u))}-e^{-(-\log(1-v))}+e^{-(-\log(1-u)-\log(1-v)+\theta\log(1-u)\log(1-v))} \\ &= 1-(1-u)-(1-v)+(1-u)(1-v)e^{-\theta\log(1-u)\log(1-v)} \\ &= u+v-1+(1-u)(1-v)e^{-\theta\log(1-u)\log(1-v)}. \end{aligned}$$

2. Geometric methods

With geometric methods we are able to construct singular copulas whose support lies in a specified set or copulas with sections given by simple functions like polynomials, for instance. These techniques follow the original conception of copulas in Definition 2.10. In particular, no distribution functions or random variables are required for the construction.
There are three well-known types of copulas that are based upon geometric construction procedures, namely

- ordinal sums,

- shuffles of M and
- convex sums.

For details, see [40], Chapter 3.2.

3. Algebraic method

In this approach, copulas are constructed from relationships involving the bivariate and marginal distribution function. The algebraic relationship could be a ratio, for instance. For details, we refer to [40], Chapter 3.3.

Construction of multivariate n-copulas

The construction of multivariate n-copulas, where $n \geq 3$, is a very difficult challenge and there are often no n-dimensional analogues of the procedures previously described for the case of $n = 2$.
For the following statements, we recall the definitions of the functions M^n, W^n and Π^n, that is
$M^n(\mathbf{u}) = \min\{u_1, ..., u_n\}$,
$W^n(\mathbf{u}) = \max\{u_1 + ... + u_n - n + 1, 0\}$ and
$\Pi^n(\mathbf{u}) = \prod_{i=1}^{n} u_i$,
where $n \geq 2$ and $\mathbf{u} = (u_1, ..., u_n)$. In the case of $n = 2$, we simply write M, W and Π instead of M^2, W^2 and Π^2 with immediate effect.
We remember that 2-copulas join one-dimensional distribution functions to form bivariate distribution functions. A first naive idea might be that 2-copulas couple other 2-copulas.

Example 2.34: We define C via $C(u, v, w) = \Pi(M(u, v), w) = w \cdot \min\{u, v\}$. It can be shown that C is a 3-copula.

However, the following counterexample clarifies that our naive technique can fail.

Example 2.35: Let C be $C(u, v, w) = W(W(u, v), w) = \max\{u + v + w - 2, 0\}$. Hence, $C = W^3$, but it can be proved that W^3 is not a copula. Each of the 2-margins of W^3 is W and in a set of three random variables X, Y and Z it is impossible for each random variable to be almost surely a decreasing function of each of the remaining two.

Definition 2.36: C_1 is directly compatible with C_2 if C_1 and C_2 are 2-copulas such that $C_2(C_1(u, v), w)$ is a copula.

There are several criteria for direct compatibility if C_1 or C_2 are M, W or Π.

Theorem 2.37:

1. Every 2-copula is directly compatible with Π.
2. The only 2-copula directly compatible with M is M.
3. The only 2-copula directly compatible with W is M.
4. M is directly compatible with every 2-copula C.
5. W is directly compatible only with Π.
6. Π is directly compatible with a 2-copula C if and only if the function $u \mapsto V_C([uv_1, uv_2] \times [w_1, w_2])$ is non-decreasing on $\mathbb{I}$ for all $v_1, v_2, w_1, w_2 \in \mathbb{I}$ with $v_1 \leq v_2$ and $w_1 \leq w_2$.

Proof: See [43]. □

Endowing a 2-copula with a multivariate margin often succeeds in the special class of Archimedean copulas, compare Chapter 4 in [40], which we discuss in the next subsection.
Sklar's theorem shows that if C is a 2-copula and F and G are univariate distribution functions, then $C(F(x), G(y))$ is a two-dimensional distribution function.
The question now is whether an extension is possible, namely whether F and G can be replaced by multivariate distribution functions. In other words, we search for copulas satisfying the following implication: If $F(\mathbf{x})$ is an m-dimensional distribution function and $G(\mathbf{y})$ is an n-dimensional distribution function, then $C(F(\mathbf{x}), G(\mathbf{y}))$ is an $(m+n)$-dimensional distribution function for all m, n with $m + n \geq 3$.
An answer to that issue is provided by the next theorem.

Theorem 2.38: Let m and n be positive integers with $m + n \geq 3$ and suppose that C is a 2-copula such that $H(\mathbf{x}, \mathbf{y}) = C(F(\mathbf{x}), G(\mathbf{y}))$ is an $(m+n)$-dimensional distribution function with margins $H(\mathbf{x}, \infty) = F(\mathbf{x})$ and $H(\infty, \mathbf{y}) = G(\mathbf{y})$ for all m-dimensional distribution functions $F(\mathbf{x})$ and n-dimensional distribution functions $G(\mathbf{y})$.
Then, $C = \Pi$.

Proof: See [11]. □

Another important result related to the preceding theorem is the following one.

Theorem 2.39: Let m and n be integers with $n, m \geq 2$, and let C_1 be an m-copula and C_2 be an n-copula.

1. Define $C : \mathbb{I}^{m+n} \to \mathbb{I}$ via
   $C(x_1, ..., x_{m+n}) = M(C_1(x_1, ..., x_m), C_2(x_{m+1}, ..., x_{m+n}))$.
   Then, C is an $(m+n)$-copula if and only if $C_1 = M^m$ and $C_2 = M^n$.
2. Let C', C'' and C''' be functions given by
   $C'(x_1, ..., x_{m+1}) = \Pi(C_1(x_1, ..., x_m), x_{m+1})$,

$C''(x_1, ..., x_{n+1}) = \Pi(x_1, C(x_2, ..., x_{n+1}))$ and
$C'''(x_1, ..., x_{m+n}) = \Pi(C_1(x_1, ..., x_m), C_2(x_{m+1}, ..., x_{m+n}))$.
Then,
C' is always an $(m+1)$-copula,
C'' is always an $(n+1)$-copula and
C''' is always an $(m+n)$-copula.

Proof: See [51]. □

Example 2.40: We consider the Farlie-Gumbel-Morgenstern (FGM) n-copulas.
The FGM family $C_\theta(u, v) = uv + \theta uv(1-u)(1-v)$ with $\theta \in [-1, 1]$, see Example 2.13.4 in this thesis, has an extension to a $(2^n - n - 1)$-parameter family of n-copulas, where $n \geq 3$, by

$$C(\mathbf{u}) = u_1...u_n \left[1 + \sum_{k=2}^{n} \sum_{1 \leq j_1 < ... < j_k \leq n} \theta_{j_1...j_k}(1-u_{j_1})...(1-u_{j_k}) \right].$$

Each copula of this family is absolutely continuous and has density

$$\frac{\partial^n C(\mathbf{u})}{\partial u_1 ... \partial u_n} = 1 + \sum_{k=2}^{n} \sum_{1 \leq j_1 < ... < j_k \leq n} \theta_{j_1...j_k}(1-2u_{j_1})...(1-2u_{j_k}).$$

Since $C(\mathbf{u})$ is quadratic in each variable, $\frac{\partial^n C(\mathbf{u})}{\partial u_1 ... \partial u_n}$ is linear in each variable. Hence, the density is greater than or equal to zero on $\mathbb{I}^n$ if and only if the density is greater than or equal zero at each of the 2^n vertices of $\mathbb{I}^n$. Therefore, we obtain 2^n constraints for the parameters, namely

$$1 + \sum_{k=2}^{n} \sum_{1 \leq j_1 < ... < j_k \leq n} \epsilon_{j_1}...\epsilon_{j_k}\theta_{j_1...j_k} \geq 0,$$

where $\epsilon_{j_1}, ..., \epsilon_{j_k} \in \{-1, 1\}$.
This means that each parameter has to satisfy $|\theta| \leq 1$.
We note that each k-margin, where $2 \leq k < n$, of an FGM-copula is an FGM k-copula.

Remark 2.41: As mentioned in the beginning, the construction of multivariate copulas is an exceedingly difficult task because the methods are not obvious in many cases. There are still many open questions about this topic.

2.1.5 Classification of copulas

We can group copula functions or copula densities into families. Two of the most important classes are elliptical copulas and Archimedean copulas, which are discussed in this subsection.

Elliptical copulas

At first, we consider elliptical distributions which are defined as follows in [9].

Definition 2.42: Let $\mathbf{X}$ be an n-dimensional random vector, $\boldsymbol{\mu} \in \mathbb{R}^n$ and Σ a non-negative definite, symmetric $(n \times n)$-matrix. Moreover, let the characteristic function $\varphi_{\boldsymbol{X}-\boldsymbol{\mu}}(\mathbf{t})$ of $\boldsymbol{X} - \boldsymbol{\mu}$ be a function of the quadratic form $\mathbf{t}^t\Sigma\mathbf{t}$, that is $\varphi_{\boldsymbol{X}-\boldsymbol{\mu}}(\mathbf{t}) = \phi(\mathbf{t}^t\Sigma\mathbf{t})$. Then, $\mathbf{X}$ has an elliptical distribution with parameters $\boldsymbol{\mu}, \Sigma$ and ϕ, and we write $\mathbf{X} \sim E_n(\boldsymbol{\mu}, \Sigma, \phi)$.

According to [50], elliptical copulas are the copulas of elliptical distributions. They extend to arbitrary dimensions and are comparably rich in parameters. Moreover, they are restricted to radial symmetry. This represents a strong limitation with respect to tail dependence. Hence, elliptical copulas might not be suitable to model precipitation, for example, if we consider practical applications in meteorology. The most important members of the class of elliptical copulas are the normal or Gaussian copula and the Student-t copula, which are presented in the following.

Normal copula

We follow [50] and assume that $\mathbf{X} = (X_1, ..., X_n)$ is a random vector whose components have known marginal cumulative distribution functions (CDF) $F_1, ..., F_n$.
Since $U_j = F_j(X_j) \sim \mathcal{U}_{\mathbb{I}}$ for $j \in \{1, ..., n\}$, each X_j can be transformed to a random variable $Z_j := \Phi^{-1}(F_j(X_j)) \sim \mathcal{N}(0, 1)$ having a standard normal distribution, where Φ^{-1} is the inverse function of the CDF of the $\mathcal{N}(0, 1)$-distribution. Furthermore, let $\mathbf{Z} = (Z_1, ..., Z_n)$ follow a multivariate standard normal distribution $\mathcal{N}_n(\mathbf{0}, \Sigma)$ with corresponding probability density function (PDF) h and CDF H, where Σ denotes the covariance matrix.
Then, according to Corollary 2.24, the PDF of $\mathbf{X}$, which is the copula in our situation here, is given by

$$C(u_1, ..., u_n) = H(\Phi^{-1}(u_1), ..., \Phi^{-1}(u_n)).$$

Hence, the normal copula density is

$$\begin{aligned} c(u_1, ..., u_n) &= \frac{\partial}{\partial u_1} \cdots \frac{\partial}{\partial u_n} C(u_1, ..., u_n) \\ &= \frac{h(\Phi^{-1}(u_1), ..., \Phi^{-1}(u_n))}{\prod_{j=1}^{n} \varphi(\Phi^{-1}(u_j))}, \end{aligned}$$

where φ denotes the PDF of the $\mathcal{N}(0, 1)$-distribution.
In the bivariate case, the copula expression can be written as

$$C(u, v) = \int_{-\infty}^{\Phi^{-1}(u)} \int_{-\infty}^{\Phi^{-1}(v)} \frac{1}{2\pi\sqrt{1 - R_{12}^2}} \exp\left(-\frac{s^2 - 2R_{12}st + t^2}{2(1 - R_{12}^2)}\right) ds\, dt\,,$$

where R_{12} is the usual linear correlation coefficient of the corresponding bivariate normal distribution, compare [9].
The normal copula can be employed in a range of applications. The estimation procedures are well-known and hence quality is controllable. However, the normal copula assumes linear relationships among the transformed random variables with rather small dependence for large deviations. Moreover, normal copulas do not necessarily exist in closed form expressions.

Student-t copula

Again, we follow [50] and define the t-copula via a multivariate extension of the t-distribution, namely

$$C(u_1, ..., u_n) = H_{t(\nu,\Sigma)}(F^{-1}_{t(\nu)}(u_1), ..., F^{-1}_{t(\nu)}(u_n)),$$

where $F_{t(\nu)}$ denotes the CDF of a univariate t-distribution with ν degrees of freedom and positive definite dispersion matrix Σ, which leads to $1 + \frac{n(n-1)}{2}$ parameters.
In the bivariate case, the copula expression is

$$C(u, v) = \int_{-\infty}^{F^{-1}_{t(\nu)}(u)} \int_{-\infty}^{F^{-1}_{t(\nu)}(v)} \frac{1}{2\pi\sqrt{1 - R_{12}^2}} \left(1 + \frac{s^2 - 2R_{12}st + t^2}{\nu(1 - R_{12}^2)}\right)^{-\frac{\nu+2}{2}} ds\, dt\,,$$

where R_{12} is the usual linear correlation coefficient of the corresponding bivariate t_ν-distribution for $\nu > 2$, compare [9].
We note that there are various definitions of a multivariate t-distribution and consequently of t-copulas.
The t-copula offers some flexibility as far as covariance structure and tail dependence are concerned. In this context, we can interpret tail dependence as the conditional probability of observing an extreme in one component given the other being in an extreme state.

Archimedean copulas

Archimedean copulas show some advantages, that are responsible for the fact that they are used in a wide range of applications.
Archimedean copulas allow for a wider variety of dependence structures, are able to model different upper and lower tail dependence, can be constructed easily, have nice properties and many copulas belong to this class, see [40]. According to [50], they are mostly employed in the bivariate case and thus we restrict to this situation at first. Later, we generalise the concept to the multivariate case.

We start with the definition and some properties of the so-called pseudo-inverse, which plays a central role in the construction of Archimedean copulas.

Definition 2.43: Let $\varphi : \mathbb{I} \to [0, \infty]$ be a continuous, strictly decreasing function with $\varphi(1) = 0$.
Then the pseudo-inverse of φ, denoted by $\varphi^{[-1]}$, satisfies $\mathrm{Dom}(\varphi^{[-1]}) = [0, \infty]$ and $\mathrm{Ran}(\varphi^{[-1]}) = \mathbb{I}$ and is given by

$$\varphi^{[-1]}(t) = \begin{cases} \varphi^{-1}(t) & \text{for } 0 \leq t \leq \varphi(0), \\ 0 & \text{for } \varphi(0) \leq t \leq \infty. \end{cases}$$

Remark 2.44:

- $\varphi^{[-1]}$ is continuous and non-increasing on $[0, \infty]$ and strictly decreasing on $[0, \varphi(0)]$.
- $\varphi^{[-1]}(\varphi(u)) = u$ on $\mathbb{I}$ and

$$\varphi(\varphi^{[-1]}(t)) = \left\{ \begin{array}{cc} t & \text{for } 0 \leq t \leq \varphi(0) \\ \varphi(0) & \text{for } \varphi(0) \leq t \leq \infty \end{array} \right\} = \min\{t, \varphi(0)\}$$

- $\varphi(0) = \infty$ implies that $\varphi^{[-1]} = \varphi^{-1}$.

Now we are ready to define Archimedean copulas.

Definition 2.45: Copulas of the form

$$C : \mathbb{I}^2 \to \mathbb{I},\ C(u, v) = \varphi^{[-1]}(\varphi(u) + \varphi(v)), \tag{2.3}$$

are called Archimedean copulas.
The function φ is called the generator of the Archimedean copula.
If $\varphi(0) = \infty$, φ is called a strict generator. Then, $\varphi^{[-1]} = \varphi^{-1}$, and C is called a strict Archimedean copula.

Remark 2.46: It can be shown that C as defined by (2.3) satisfies the copula boundary conditions and that C is 2-increasing if and only if $u_1 \leq u_2$ and $C(u_2, v) - C(u_1, v) \leq u_2 - u_1$, compare [40]. Furthermore, and this is very important, the function C as defined in (2.3) is a copula if and only if φ is convex.

Example 2.47:

1. We consider $\varphi(t) = \log(t)$ for $t \in \mathbb{I}$. Since $\varphi(0) = \infty$, φ is strict. Thus, $\varphi^{[-1]}(t) = \varphi^{-1}(t) = \exp(-t)$, and $C(u, v) = \exp(-[(-\log u) + (-\log v)]) = uv = \Pi(u, v)$. Hence, Π is a strict Archimedean copula.

2. We consider $\varphi(t) = 1 - t$ for $t \in \mathbb{I}$. Then,

$$\varphi^{[-1]}(t) = \left\{ \begin{array}{cc} 1 - t & \text{for } t \in \mathbb{I} \\ 0 & \text{for } t > 1 \end{array} \right\} = \max\{1 - t, 0\}.$$

   Hence, $C(u, v) = \max\{u + v - 1, 0\} = W(u, v)$, and W is an Archimedean copula.

3. An overview of Archimedean copulas and their corresponding generators can be found in [40]. An excerpt is given in Table 2.1.

Copula	$C_\theta(u,v)$	$\varphi_\theta(t)$	$\theta \in$	strict?
Clayton	$\max((u^{-\theta}+v^{-\theta}-1)^{-1/\theta},0)$	$\frac{1}{\theta}(t^{-\theta}-1)$	$[-1,\infty)\setminus\{0\}$	for $\theta \geq 0$
Frank	$-\frac{1}{\theta}\log(1+\frac{(e^{-\theta u}-1)(e^{-\theta v}-1)}{e^{-\theta}-1})$	$-\log\frac{e^{-\theta t}-1}{e^{-\theta}-1}$	$(-\infty,\infty)\setminus\{0\}$	yes
Gumbel	$\exp(-[(-\log u)^\theta+(-\log v)^\theta]^{1/\theta})$	$(-\log t)^\theta$	$[1,\infty)$	yes
Al-Mikhail-Haq	$\frac{uv}{1-\theta(1-u)(1-v)}$	$\log\frac{1-\theta(1-t)}{t}$	$[-1,1)$	yes

Table 2.1: Some Archimedean copulas

Remark 2.48: According to the preceding facts, for the construction of Archimedean copulas we only need functions φ serving as generators and possessing the necessary properties previously mentioned, therefore $\varphi : \mathbb{I} \to [0,\infty]$ should satisfy $\varphi(1) = 0$ and should be continuous, strictly decreasing and convex. Then, an Archimedean copula can be constructed via (2.3).

We now extend the Archimedean copula concept to the multivariate case and set

$$C^n(\mathbf{u}) = \varphi^{[-1]}(\varphi(u_1)+\varphi(u_2)+...+\varphi(u_n)). \tag{2.4}$$

The functions C^n are the serial iterates of the Archimedean 2-copula generated by φ, that is, if we set

$$C^2(u_1,u_2) = C(u_1,u_2) = \varphi^{[-1]}(\varphi(u_1)+\varphi(u_2)),$$

then we get

$$C^n(u_1,...,u_n) = C(C^{n-1}(u_1,...,u_{n-1}),u_n) \text{ for all } n \geq 3,$$

see [51].
However, this method of composing copulas fails in general. For example, if we set $\varphi(t) = 1-t$ in (2.4), we obtain W^n, but that fails to be a copula for any $n \geq 2$, as previously mentioned.
We have already stated properties of φ such that C^2 is a copula and now search for similar characteristics in the case of $n \geq 3$.

Definition 2.49: A function $g(t)$ is called completely monotone on an interval $J \subseteq \mathbb{R}$ if it is continuous in J and satisfies $(-1)^k \frac{d^k}{dt^k} g(t) \geq 0$ for all $t \in \text{int}(J)$ and $k \in \{0,1,2,...\}$, where $\text{int}(J)$ denotes the interior of J.

Theorem 2.50: Let $\varphi : \mathbb{I} \to [0,\infty]$ be a continuous, strictly decreasing function with $\varphi(0) = \infty$ and $\varphi(1) = 0$ and let φ^{-1} denote the inverse of φ. If $C^n : \mathbb{I}^n \to \mathbb{I}$ is given by

(2.4), then C^n is an n-copula for all $n \geq 2$ if and only if φ^{-1} is completely monotone on $[0, \infty)$.

Proof: See [23]. □

A recent result proposed by McNeil and Nešlehová in [36], which is structurally similar to Theorem 2.50, uses the concept of n-monotonicity to characterise n-dimensional Archimedean copulas.

Definition 2.51: A real function $g(t)$ is called n-monotone in (a, b), where $a, b \in \overline{\mathbb{R}}$ and $n \geq 2$, if it is differentiable there up to order $n-2$ and if the derivatives satisfy $(-1)^k \frac{d^k}{dt^k} g(t) \geq 0$ for $k \in \{0, 1, ..., n-2\}$ for any $t \in (a, b)$ and further if $(-1)^{n-2} \frac{d^{n-2}}{dt^{n-2}} g(t)$ is non-increasing and convex in (a, b).
For $n = 1$, g is called 1-monotone if it is non-negative and non-increasing there.

It is possible to extend Definition 2.51 to functions on not necessarily open intervals as follows.

Definition 2.52: A real function g on an interval $J \subseteq \overline{\mathbb{R}}$ is n-monotone on J, where $n \in \mathbb{N}$, if it is continuous there and if g restricted to $\text{int}(J)$ is n-monotone on $\text{int}(J)$.

The following important Theorem 2.53 shows the relationship between n-monotonicity and n-dimensional Archimedean copulas and can be found in [36].
Note that in [36], Archimedean copulas are defined in terms of φ rather than φ^{-1}, and the definition of generators differs analogously. However, we stick to the definition introduced before, and thus, the following results taken from [36] are rewritten in our terminology.

Theorem 2.53: Let φ be a generator. Then,

$$C^n : \mathbb{I}^n \to \mathbb{I},\ C^n(\mathbf{u}) = \varphi^{-1}(\varphi(u_1) + ... + \varphi(u_n)),$$

where $\mathbf{u} = (u_1, ..., u_n) \in \mathbb{I}^n$, is an n-dimensional copula if and only if φ^{-1} is n-monotone on $[0, \infty)$.

Proof: See [36]. □

Corollary 2.54: Let φ be a generator whose inverse φ^{-1} has derivatives up to order n on $(0, \infty)$. Then, φ generates an Archimedean copula if and only if $(-1)^k \frac{d^k}{dt^k} \varphi^{-1}(t) \geq 0$ for $k \in \{1, ..., n\}$.

Remark 2.55: Evidently, generators with n-monotone inverses do not necessarily generate Archimedean copulas in dimensions higher than n. Hence, n-monotonicity of a function ψ does not imply that ψ is k-monotone, where $k > n$.

However, on the other hand, it can be easily seen that an n-monotone function, where $n \geq 2$, is also k-monotone for any $k \in \{1, ..., n\}$.

We close this subsection with several examples for n-dimensional Archimedean copulas. In particular, we consider the Clayton, Frank and Gumbel families, which have n-dimensional extensions for certain values of θ.

Example 2.56: Let $\varphi_\theta(t) = t^{-\theta}$ for $\theta > 0$. This φ generates the subfamily of the bivariate Clayton family, whose generators are strict. We get $\varphi_\theta^{-1}(t) = (1+t)^{-1/\theta}$, and it can be verified that φ_θ^{-1} is completely monotone on $[0, \infty)$. Hence, the Clayton family of n-copulas for $\theta > 0$ and any $n \geq 2$ can be expressed as

$$C_\theta^n(\mathbf{u}) = (u_1^{-\theta} + ... + u_n^{-\theta} - n + 1)^{-1/\theta}.$$

Example 2.57: We consider $\varphi_\theta(t) = -\log\left(\frac{e^{-\theta t}-1}{e^{-\theta}-1}\right)$, the generator of the bivariate Frank family, whose inverse is given by $\varphi_\theta^{-1}(t) = -\frac{1}{\theta}\log(1-(1-e^{-\theta})e^{-t})$. Let $\theta \in (0, \infty)$. Then, $f(x) = -\frac{1}{\theta}\log(1-x)$ is absolutely monotone for $x \in (0,1)$, and $g(t) = (1-e^{-\theta})e^{-t}$ is completely monotone for $t \in [0, \infty)$. Hence, φ_θ^{-1} is completely monotone on $[0, \infty)$. Therefore, a generalisation to the Frank family of n-copulas for $\theta > 0$ and any $n \geq 2$ is given by

$$C_\theta^n(\mathbf{u}) = -\frac{1}{\theta}\left(1 + \frac{(e^{-\theta u_1}-1)...(e^{-\theta u_n}-1)}{(e^{-\theta}-1)^{n-1}}\right).$$

We note that for $\theta < 0$, φ_θ^{-1} is not completely monotone.

Example 2.58: We consider $\varphi_\theta(t) = (-\log t)^\theta$ for $\theta \geq 1$, the generator of the bivariate Gumbel family. Its inverse is given by $\varphi_\theta^{-1}(t) = \exp(-t^{1/\theta})$. Since e^{-x} is completely monotone and $t^{1/\theta}$ is a positive function that has a completely monotone derivative, $\varphi_\theta^{-1}(t)$ is completely monotone. Hence, a generalisation to the Gumbel family of n-copulas for $\theta \geq 1$ and any $n \geq 2$ is provided by

$$C_\theta^n(\mathbf{u}) = \exp(-[(-\log u_1)^\theta + ... + (-\log u_n)^\theta]^{1/\theta}).$$

2.1.6 Copulas and dependence

In this subsection, we discuss several aspects of dependence and dependence measures with respect to copulas. Thereby, we focus on the bivariate case, but comment on extensions to the multivariate case, too.

Perfect dependence

We recall the Fréchet-Hoeffding inequality for an n-copula, namely

$$W^n(u_1, ..., u_n) \leq C(u_1, ..., u_n) \leq M^n(u_1, ..., u_n),$$

and the fact that the lower and the upper bounds are copulas themselves in the case of $n = 2$. Moreover, it can be shown that W and M are the bivariate distribution functions of $(U, 1-U)$ and (U, U), respectively, where $U \sim \mathcal{U}_{\mathbb{I}}$. In this situation, W describes perfect negative and M perfect positive dependence, compare [9].

Theorem 2.59: Let (X, Y) have one of the copulas W or M. Then, there are two monotone functions $\alpha, \beta : \mathbb{R} \to \mathbb{R}$ and a random variable Z with $(X, Y) =^d (\alpha(Z), \beta(Z))$. In this connection, α is increasing and β is decreasing in the case of W and both α and β are increasing in the case of M. The converse holds, too.

Proof: See [8]. □

Concordance

Definition 2.60: Let X and Y be continuous random variables and (x, y) and $(\tilde{x}, \tilde{y})$ two observations from (X, Y). Then, (x, y) and $(\tilde{x}, \tilde{y})$ are concordant if $(x - \tilde{x}) \cdot (y - \tilde{y}) > 0$ and discordant if $(x - \tilde{x}) \cdot (y - \tilde{y}) < 0$.

Theorem 2.61: Let X and Y be continuous random variables and (X, Y) and $(\tilde{X}, \tilde{Y})$ be independent with joint distribution functions H and $\tilde{H}$, respectively, with common margins F of X and $\tilde{X}$ and G of Y and $\tilde{Y}$. Moreover, let C and $\tilde{C}$ be the copulas of (X, Y) and $(\tilde{X}, \tilde{Y})$, respectively, such that $H(x, y) = C(F(x), G(y))$ and $\tilde{H}(x, y) = \tilde{C}(F(x), G(y))$, according to Sklar's theorem. We define the concordance function Q by

$$Q := \mathbb{P}[(X - \tilde{X}) \cdot (Y - \tilde{Y}) > 0] - \mathbb{P}[(X - \tilde{X}) \cdot (Y - \tilde{Y}) < 0]. \tag{2.5}$$

Then,

$$Q = Q(C, \tilde{C}) = 4 \cdot \int_0^1 \int_0^1 \tilde{C}(u, v) dC(u, v) - 1.$$

Proof: See [40]. □

Corollary 2.62: Q is symmetric, therefore $Q(C, \tilde{C}) = Q(\tilde{C}, C)$.

Measures of concordance: Kendall's tau and Spearman's rho

Definition 2.63: A measure $\kappa \in \mathbb{R}$ of dependence between two continuous random variables X and Y with copula C is a measure of concordance if the following properties, that can be found in [47], are satisfied:

1. κ is defined for every pair (X, Y) of continuous random variables.
2. $-1 \leq \kappa_{X,Y} \leq 1$, $\kappa_{X,X} = 1$ and $\kappa_{X,-X} = -1$.
3. $\kappa_{X,Y} = \kappa_{Y,X}$.

4. If X and Y are independent, then $\kappa_{X,Y} = \kappa_\Pi = 0$.

5. $\kappa_{-X,Y} = \kappa_{X,-Y} = -\kappa_{X,Y}$.

6. If C and $\tilde{C}$ are copulas such that C is less concordant than $\tilde{C}$, that is $C(u,v) \leq \tilde{C}(u,v)$ for all $(u,v) \in \mathbb{I}^2$, then $\kappa_C \leq \kappa_{\tilde{C}}$.

7. If $\{X_n, Y_n\}$ is a sequence of continuous random variables with copulas C_n, and if $\{C_n\}$ converges pointwise to C, then $\lim_{n\to\infty} \kappa_{C_n} = \kappa_C$.

In what follows, we present two measures of concordance, namely Kendall's tau and Spearman's rho. In [40] it is shown that those measures defined below satisfy the properties in Definition 2.63. Hence, they are indeed measures of concordance.

Definition 2.64: Let (X,Y) be a random vector.

1. Let $(\tilde{X},\tilde{Y})$ denote an independent copy of (X,Y). Then, Kendall's tau for (X,Y) is given by

$$\begin{aligned}\tau(X,Y) &= \mathbb{P}[(X-\tilde{X})\cdot(Y-\tilde{Y})>0] - \mathbb{P}[(X-\tilde{X})\cdot(Y-\tilde{Y})<0] \\ &= \text{probability of concordance} - \text{probability of discordance}\,.\end{aligned}$$

2. Let (X,Y), $(\tilde{X},\tilde{Y})$ and $(\hat{X},\hat{Y})$ be independent copies. Then, Spearman's rho is defined as

$$\rho(X,Y) = 3\cdot(\mathbb{P}[(X-\tilde{X})\cdot(Y-\hat{Y})>0] - \mathbb{P}[(X-\tilde{X})\cdot(Y-\hat{Y})<0]).$$

For non-elliptical distributions, the usual linear correlation coefficient is inappropriate. In this situation, Kendall's tau and Spearman's rho provide good alternatives. We now study several properties of and relationships between Kendall's tau and Spearman's rho.

Theorem 2.65: For a vector (X,Y) of continuous random variables with copula C, Kendall's tau is given by

$$\tau(X,Y) = Q(C,C) = 4\cdot\int_0^1\int_0^1 C(u,v)dC(u,v) - 1.$$

Proof: See [40]. □

In many cases, the double integral in Theorem 2.65 cannot be calculated easily. However, for Archimedean copulas there exists another expression for Kendall's tau depending on the generator and its derivative.

Theorem 2.66: For random variables X and Y with an Archimedean copula C generated by φ, Kendall's tau is given by

$$\tau = 1 + 4 \cdot \int_0^1 \frac{\varphi(t)}{\varphi'(t)} dt.$$

Proof: See [40] or [9]. □

Example 2.67: We consider the Gumbel family whose generator is $\varphi(t) = (-\log t)^\theta$ for $\theta \geq 1$. Then,

$$\varphi'(t) = \theta(-\log t)^{\theta-1} \cdot (-\frac{1}{\theta}) = -\frac{t}{\theta}(-\log t)^{\theta-1},$$

and hence

$$\frac{\varphi(t)}{\varphi'(t)} = \frac{(-\log t)^\theta}{-\frac{t}{\theta}(-\log t)^{\theta-1}} = -\frac{t}{\theta}(-\log t) = \frac{t \log t}{\theta}.$$

According to Theorem 2.66, we obtain

$$\begin{aligned} \tau &= 1 + 4 \int_0^1 \frac{t \log t}{\theta} dt = 1 + \frac{4}{\theta} \int_0^1 t \log t dt = 1 + \frac{4}{\theta} \left(\left[\frac{1}{2} t^2 \log t \right]_0^1 - \int_0^1 \frac{1}{2} t^2 \frac{1}{t} dt \right) \\ &= 1 + \frac{4}{\theta} \left(0 - 0 - \left[\frac{1}{4} t^2 \right]_0^1 \right) = 1 + \frac{4}{\theta} \cdot \left(-\frac{1}{4} \right) = 1 - \frac{1}{\theta}. \end{aligned}$$

Theorem 2.68: For a vector (X, Y) of continuous random variables with copula C, Spearman's rho is given by

$$\rho_C = 3Q(C, \Pi) = 12 \cdot \int_0^1 \int_0^1 u\, v\, dC(u, v) - 3 = 12 \cdot \int_0^1 \int_0^1 C(u, v)\, du\, dv - 3.$$

Proof: See [40] and Section 3 in [48], respectively. □

Remark 2.69: According to [40], Spearman's rho is often called the "grade" correlation coefficient. Grades are the population analogs of ranks: If x and y are observations from two random variables X and Y, respectively, with corresponding distribution functions F and G, then $u = F(x)$ and $v = G(y)$ are the grades of x and y. Observe that the grades u and v are observations from the random variables $U = F(X)$ and $V = G(Y)$, which are uniformly distributed on $(0, 1)$ and whose joint distribution function is C.
From elementary probability theory, it is well-known that U and V each have mean $\frac{1}{2}$ and variance $\frac{1}{12}$ in that case. Hence, the expression for ρ_C in Theorem 2.68 can be

written as

$$\begin{aligned}\rho_C &= 12\int_0^1\int_0^1 uv\,dC(u,v) - 3 = 12\,\mathbb{E}[UV] - 3 = \frac{\mathbb{E}[UV] - \frac{1}{4}}{\frac{1}{12}} \\ &= \frac{\mathbb{E}[UV] - \frac{1}{2}\cdot\frac{1}{2}}{\sqrt{\frac{1}{12}}\cdot\sqrt{\frac{1}{12}}} = \frac{\mathbb{E}[UV] - \mathbb{E}[U]\,\mathbb{E}[V]}{\sqrt{\mathrm{Var}(U)}\,\sqrt{\mathrm{Var}(V)}}.\end{aligned}$$

As we see, Spearman's rho for a pair of continuous random variables X and Y is identical to Pearson's product-moment correlation coefficient for the grades of X and Y, that is, the random variables $U = F(X)$ and $V = G(Y)$.

The following result holds both for Kendall's tau and Spearman's rho.

Theorem 2.70: Let X and Y be continuous random variables with copula C and let κ be Kendall's tau or Spearman's rho. Then,

1. $\kappa(X,Y) = 1$ if and only if $C = M$ and
2. $\kappa(X,Y) = -1$ if and only if $C = W$.

Proof: Compare [8]. □

The following results reflect the relationships between Kendall's tau and Spearman's rho.

Theorem 2.71: Let X and Y be continuous random variables, τ denote Kendall's tau and ρ denote Spearman's rho. Then,

1. $-1 \leq 3\tau - 2\rho \leq 1$ and
2. $\frac{1+\rho}{2} \geq (\frac{1+\tau}{2})^2$ and $\frac{1-\rho}{2} \geq (\frac{1-\tau}{2})^2$.

Proof: See [40]. □

Corollary 2.72: $\frac{3\tau-1}{2} \leq \rho \leq \frac{1+2\tau-\tau^2}{2}$ for $\tau \geq 0$ and $\frac{\tau^2+2\tau-1}{2} \leq \rho \leq \frac{1+3\tau}{2}$ for $\tau \geq 0$.

Proof: We obtain the claims by combining Theorems 2.70.1 and 2.70.2, compare [40]. □

Remark 2.73: Other interesting scalar measures of dependence are given by Schweizer and Wolff's sigma and Gini's gamma, compare [40]. While Gini's gamma is indeed a concordance measure, Schweizer and Wolff's sigma is not.

After having discussed both the notion of concordance and concordance measures in the bivariate case, we now focus on extensions to the multivariate situation by generalising the concept of concordance and presenting a multivariate version of Spearman's rho.

Remark 2.74: For a generalisation of the notion of concordance and the concordance function Q defined in (2.5), we assume that $\mathbf{x}$ and $\mathbf{y}$ are two observations of a vector $\mathbf{X}$ consisting of continuous random variables.
According to [39], concordance generalises in the following sense: We say that $\mathbf{x}$ and $\mathbf{y}$ are concordant if (x_i, x_j) and (y_i, y_j) are concordant for all $i \neq j$. However, it is not possible to extend the notion of discordance to dimensions $n \geq 3$. Indeed, if (x_1, x_2) and (y_1, y_2) are discordant and (x_2, x_3) and (y_2, y_3) are discordant, then (x_1, x_3) and (y_1, y_3) have to be concordant. Hence, in the multivariate case, we consider the probability of concordance alone, rather than the difference of the probabilities of concordance and discordance.
Let $\mathbf{X}_1$ and $\mathbf{X}_2$ be independent vectors of continuous random variables with common univariate margins and n-copulas C_1 and C_2, respectively. We follow [39] and define

$$Q'_n := \mathbb{P}[\mathbf{X}_1 > \mathbf{X}_2] + \mathbb{P}[\mathbf{X}_1 < \mathbf{X}_2]$$

as the probability of concordance between $\mathbf{X}_1$ and $\mathbf{X}_2$. It can be shown that then,

$$Q'_n = Q_n(C_1, C_2) = \int_0^1 \ldots \int_0^1 C_2(\mathbf{u})\, dC_1(\mathbf{u}) + \int_0^1 \ldots \int_0^1 C_1(\mathbf{u})\, dC_2(\mathbf{u}).$$

A multivariate analog Q_n of the concordance function Q for continuous random vectors $\mathbf{X}_1$ and $\mathbf{X}_2$ with corresponding copulas C_1 and C_2 is given by

$$Q_n(C_1, C_2) := \frac{1}{2^{n-1} - 1} \left[2^{n-1} Q'_n(C_1, C_2) - 1 \right]. \tag{2.6}$$

For a detailed discussion of multivariate concordance in general, we refer to [20].
Many of the measures of concordance have multivariate versions. For instance, a multivariate population version of Spearman's rho for a continuous random vector $\mathbf{X}$ with copula C is given by

$$\begin{aligned} \rho_{n,C} &= \frac{(n+1)(2^{n-1} - 1)}{2^n - (n+1)}\, Q_n(C, \Pi^n) \\ &= \frac{n+1}{2^n - (n+1)} \left[2^{n-1} \left(\int_0^1 \ldots \int_0^1 C(\mathbf{u})\, d\Pi^n(\mathbf{u}) + \int_0^1 \ldots \int_0^1 \Pi^n(\mathbf{u})\, dC(\mathbf{u}) \right) - 1 \right], \end{aligned}$$

where $n \geq 2$, see [39]. Here, Q_n is defined as in (2.6).

Remark 2.75: Note that there are several multivariate versions for a measure of bivariate concordance. For Spearman's rho, we follow [48] and derive an extension which is different to that presented in the preceding remark in what follows.
Recall that, according to Theorem 2.68, Spearman's rho for a two-dimensional vector (X, Y) of continuous random variables with copula C can be written as

$$\rho_C = 12 \int_0^1 \int_0^1 C(u, v)\, du\, dv - 3.$$

Moreover, it can be shown that

$$\int_0^1 \int_0^1 \Pi(u,v)\,du\,dv = \int_0^1 \int_0^1 uv du\,dv = \frac{1}{4}$$

and

$$\int_0^1 \int_0^1 M(u,v)\,du\,dv = \int_0^1 \int_0^1 \min\{u,v\}\,du\,dv = \frac{1}{3}.$$

Hence,

$$\begin{aligned}
\rho_C &= 12 \int_0^1 \int_0^1 C(u,v)\,du\,dv - 3 = \frac{\int_0^1 \int_0^1 C(u,v)\,du\,dv}{\frac{1}{12}} - \frac{1}{\frac{4}{12}} \\
&= \frac{\int_0^1 \int_0^1 C(u,v)\,du\,dv - \frac{1}{4}}{\frac{1}{12}} = \frac{\int_0^1 \int_0^1 C(u,v)\,du\,dv - \frac{1}{4}}{\frac{1}{3} - \frac{1}{4}} \\
&= \frac{\int_0^1 \int_0^1 C(u,v)\,du\,dv - \int_0^1 \int_0^1 \Pi(u,v)\,du\,dv}{\int_0^1 \int_0^1 M(u,v)\,du\,dv - \int_0^1 \int_0^1 \Pi(u,v)\,du\,dv} \\
&= \frac{\int_0^1 \int_0^1 uv\,dC(u,v) - \int_0^1 \int_0^1 uv\,d\Pi(u,v)}{\int_0^1 \int_0^1 uv\,dM(u,v) - \int_0^1 \int_0^1 uv\,d\Pi(u,v)}.
\end{aligned}$$

Therefore, ρ_C can be seen as the normalised average distance between the copula C and the product copula Π.
Due to this representation of ρ_C and the identities

$$\int_0^1 ... \int_0^1 \Pi^n(\mathbf{u})\,d\mathbf{u} = \int_0^1 ... \int_0^1 u_1...u_n\,du_1...du_n = \frac{1}{2^n}$$

and

$$\int_0^1 ... \int_0^1 M^n(\mathbf{u})\,d\mathbf{u} = \int_0^1 ... \int_0^1 \min\{u_1,...,u_n\}\,du_1...du_n = \frac{1}{n+1},$$

whose proofs are straightforward, a canonical n-dimensional extension of ρ_C to $\rho_{n,C}$ for an n-copula C^n is given by

$$\begin{aligned}
\rho_{n,C} &= \frac{\int_0^1 \dots \int_0^1 C^n(\mathbf{u})\, d\mathbf{u} - \int_0^1 \dots \int_0^1 \Pi^n(\mathbf{u})\, d\mathbf{u}}{\int_0^1 \dots \int_0^1 M^n(\mathbf{u})\, d\mathbf{u} - \int_0^1 \dots \int_0^1 \Pi^n(\mathbf{u})\, d\mathbf{u}} \\
&= \frac{\int_0^1 \dots \int_0^1 C^n(\mathbf{u})\, d\mathbf{u} - \frac{1}{2^n}}{\frac{1}{n+1} - \frac{1}{2^n}} = \frac{\frac{2^n \int_0^1 \dots \int_0^1 C^n(\mathbf{u})\, d\mathbf{u}}{2^n}}{\frac{2^n-(n+1)}{(n+1)2^n}} \\
&= \frac{n+1}{2^n-(n+1)} \left\{ 2^n \int_0^1 \dots \int_0^1 C^n(\mathbf{u})\, d\mathbf{u} - 1 \right\}.
\end{aligned}$$

This multivariate version of Spearman's rho will play a role later on.

Tail dependence

According to [9], tail dependence reflects the amount of dependence in the upper-right-quadrant tail or the lower-left-quadrant tail if we consider a bivariate distribution. Moreover, it is important for the study of dependence between extreme values.

Definition 2.76: Let X and Y be continuous random variables with marginal distribution functions F and G, respectively. The coefficient λ_U of upper tail dependence is defined by

$$\lambda_U = \lim_{u \nearrow 1} \mathbb{P}[Y > G^{-1}(u) | X > F^{-1}(u)]$$

if the limit $\lambda_U \in [0,1]$ exists.
In the case of $\lambda_U \in (0,1]$, X and Y are asymptotically dependent in the upper tail. In the case of $\lambda_U = 0$, X and Y are asymptotically independent in the upper tail.

An alternative and equivalent definition for continuous random variables is the following one.

Definition 2.77: If the limit

$$\lambda_U = \lim_{u \nearrow 1} \left(\frac{1 - 2u + C(u,u)}{1-u} \right)$$

exists for a bivariate copula C, then C has upper tail dependence if $\lambda_U \in (0,1]$ and upper tail independence if $\lambda_U = 0$.

Remark 2.78: For copulas that have no simple closed form, like the Gaussian copula, an alternative formula for λ_U is more suitable. We recall Theorem 2.22 and remark

that for a pair of random variables (U, V), where $U, V \sim \mathcal{U}_{\mathbb{I}}$, with copula C, λ_U can be determined as follows:

$$\lambda_U = \lim_{u \nearrow 1} \{\mathbb{P}[V > u | U = u] + \mathbb{P}[U > u | V = u]\}.$$

If $C(u, v) = C(v, u)$, the expression simplifies to $\lambda_U = 2 \cdot \lim_{u \nearrow 1} \mathbb{P}[V > u | U = u]$.
For a derivation, we refer to [9].

In the case of Archimedean copulas, tail dependence can be written in terms of the generators.

Theorem 2.79: Let φ be a strict generator. If $\varphi^{-1\,\prime}(0)$ is finite, then $C(u, v) = \varphi^{-1}(\varphi(u) + \varphi(v))$ does not have upper tail dependence. If C has upper tail dependence, then $\varphi^{-1\,\prime}(0) = -\infty$ and the coefficient of upper tail dependence is given by

$$\lambda_U = 2 - 2 \cdot \lim_{s \searrow 0} \left(\frac{\varphi^{-1\,\prime}(2s)}{\varphi^{-1\,\prime}(s)} \right).$$

Proof: See [9]. □

Example 2.80: We determine λ_U for the bivariate Gumbel family of copulas given by

$$C_\theta(u, v) = \exp(-[(-\log u)^\theta + (-\log v)^\theta]^{1/\theta}),$$

where $\theta \geq 1$, in two different ways.

1. We have $\frac{1-2u+C(u,u)}{1-u} = \frac{1-2u+\exp(-[2(-\log u)^\theta]^{1/\theta})}{1-u} = \frac{1-2u+\exp(2^{1/\theta} \log u)}{1-u}$
   $= \frac{1-2u+\exp(\log u^{2^{1/\theta}})}{1-u} = \frac{1-2u+u^{2^{1/\theta}}}{1-u}$, and by employing Definition 2.77 we get

   $$\begin{aligned} \lambda_U &= \lim_{u \nearrow 1} \frac{1 - 2u + C(u, u)}{1 - u} = \lim_{u \nearrow 1} \frac{1 - 2u + u^{2^{1/\theta}}}{1 - u} = \lim_{u \nearrow 1} \frac{-2 + 2^{1/\theta} u^{2^{1/\theta} - 1}}{-1} \\ &= 2 - \lim_{u \nearrow 1} (2^{1/\theta} u^{2^{1/\theta} - 1}) = 2 - 2^{1/\theta}. \end{aligned}$$

   Hence, C_θ has upper tail dependence for $\theta > 1$.

2. The Gumbel copulas are strict Archimedean with generator $\varphi(t) = (-\log t)^\theta$. Thus, we have $\varphi^{-1}(s) = \exp(-s^{1/\theta})$, and $\varphi^{-1\,\prime}(s) = -s^{\frac{1}{\theta} - 1} \exp(-s^{1/\theta}) \cdot \frac{1}{\theta}$. Theorem 2.79 yields

   $$\begin{aligned} \lambda_U &= 2 - 2 \cdot \lim_{s \searrow 0} \frac{\varphi^{-1\,\prime}(2s)}{\varphi^{-1\,\prime}(s)} = 2 - 2 \cdot \lim_{s \searrow 0} \left(\frac{-(2s)^{1/\theta - 1} \exp(-(2s)^{1/\theta}) \cdot \frac{1}{\theta}}{-s^{1/\theta - 1} \exp(-s^{1/\theta}) \cdot \frac{1}{\theta}} \right) \\ &= 2 - 2 \cdot 2^{1/\theta - 1} \underbrace{\lim_{s \searrow 0} \frac{\exp(-(2s)^{1/\theta})}{\exp(-s^{1/\theta})}}_{=1} = 2 - 2^{1/\theta}, \end{aligned}$$

   the result from 1.

Copula	Upper Tail Dependence	Lower Tail Dependence
Product	−	−
Normal	−	−
Student-t	+	+
Clayton	−	+
Gumbel	+	−
Frank	−	−
Fréchet	+	+
Marshall-Olkin	+	−
Farlie-Gumbel-Morgenstern	−	−

Table 2.2: Ability of copulas to model upper or lower tail dependence

Similar to the upper tail dependence, we can define the lower tail dependence as follows.

Definition 2.81: If the limit

$$\lambda_L = \lim_{u \searrow 0} \frac{C(u,u)}{u}$$

exists, then C has lower tail dependence if $\lambda_L \in (0,1]$ and lower tail independence if $\lambda_L = 0$.

Remark 2.82: For copulas without a simple closed form, we use

$$\lambda_L = \lim_{u \searrow 0} \{\mathbb{P}[V < u | U = u] + \mathbb{P}[U < u | V = u]\}$$

as an alternative formula for λ_L,where (U,V) is a random vector with copula C, see [9]. If $C(u,v) = C(v,u)$, then λ_L simplifies to $\lambda_L = 2 \cdot \lim_{u \searrow 0} \mathbb{P}[V < u | U = u]$.

As in the case of λ_U, there is an alternative expression for λ_L if C is a strict Archimedean copula.

Theorem 2.83: Let φ be as in Theorem 2.79. Then, for the copula $C(u,v) = \varphi^{-1}(\varphi(u) + \varphi(v))$, we obtain

$$\lambda_L = 2 \cdot \lim_{s \to \infty} \frac{\varphi^{-1\,\prime}(2s)}{\varphi^{-1\,\prime}(s)}.$$

Proof: The proof is similar to that of Theorem 2.79. □

In Table 2.2, we employ the information given in [9] and [50] and summarise the ability of frequently used copulas to model upper or lower tail dependence, where + denotes adequacy and − denotes inadequacy of the copula for the corresponding modelling.

2.2 State-of-the-art ensemble postprocessing techniques

In this section, we summarise several ensemble postprocessing methods that apply to a single weather quantity at a single location and for a single look-ahead time only and lead to univariate probabilistic forecasts. To be precise, we discuss the approaches mentioned in the introduction in Chapter 1, namely Bayesian model averaging (BMA), ensemble model output statistics (EMOS), logistic regression and quantile regression.

2.2.1 Bayesian model averaging (BMA)

We discuss four different versions of the BMA approach that can be used depending on the weather quantity to be forecasted. The first variant deals with the prediction of temperature and sea level pressure, the second with precipitation forecasting, the third with wind speed forecasting and the fourth with the prediction of wind directions.

BMA for temperature and sea level pressure

First, we describe the BMA method to obtain statistically postprocessed, calibrated and sharp ensemble forecasts for temperature and sea level pressure as suggested in [46].
We take a look at an ensemble of M weather field forecasts and write y for the weather quantity we are interested in and $f_1, ..., f_M$ for the respective ensemble member forecasts. Moreover, let $g_m(y|f_m)$ denote the conditional probability density function (PDF) we associate with each ensemble member. This PDF can be loosely interpreted as the conditional PDF of y given the fact that member m is the best among the ensemble member predictions. The BMA predictive PDF p for y is

$$p(y|f_1, ..., f_M) = \sum_{m=1}^{M} w_m g_m(y|f_m), \tag{2.7}$$

where w_m denotes the probability that ensemble member m is the best. Since the w_m's are probabilities, they are non-negative and $\sum_{m=1}^{M} w_m = 1$.
Hence, the BMA predictive PDF is a weighted average of predictive PDFs associated with each individual ensemble member, and the weights reflect the member's relative skill or performance.
In [46], the approximation of the conditional PDF $g_m(y|f_m)$ for forecasting temperature and sea level pressure is given by a univariate normal density centered at a linearly bias-corrected forecast with a standard deviation assumed to be constant across ensemble members, that is

$$y|f_m \sim \mathcal{N}(a_m + b_m f_m, \sigma^2). \tag{2.8}$$

Then, the BMA predictive mean is the conditional expectation of y given the forecasts, namely

$$\mathbb{E}[y|f_1, ..., f_M] = \sum_{m=1}^{M} w_m(a_m + b_m f_m).$$

We estimate the weights w_m in (2.7) and the parameters a_m, b_m and σ in (2.8) from training data as follows:

- We use linear least squares regression to estimate a_m and b_m for each ensemble member separately.
- $w_1, ..., w_M$ and σ^2 are estimated simultaneously for all the M ensemble members by employing the Expectation Maximization (EM) algorithm to maximise the log-likelihood function in the Maximum Likelihood (ML) approach for the estimation of $w_1, ..., w_M$ and σ^2. Finally, the estimate of σ is refined by optimising the so-called continuous ranked probability score (CRPS), compare Section 4.3, for the training data, that is we employ minimum CRPS estimation as described more explicitly in the EMOS approach presented below. This is done by searching numerically over a range of values of σ centered at the ML estimate, where the other parameters are fixed.

In the following, we denote space and time by subscripts s and t, respectively. Consequently, f_{mst} denotes the mth forecast in the ensemble for place s and time t and y_{st} the corresponding verification.
The BMA predictive variance of y_{st} given the ensemble of forecasts can be expressed by

$$\mathrm{Var}(y_{st}|f_{1st}, ..., f_{Mst}) = \sum_{m=1}^{M} w_m((a_m + b_m f_{mst}) - \sum_{n=1}^{M} w_n(a_n + b_n f_{nst}))^2 + \sigma^2. \quad (2.9)$$

In other words, we have a decomposition of the BMA forecast variance into two components, namely

$$\text{Predictive Variance} = \text{Between-Forecast Variance} + \text{Within-Forecast Variance}. \quad (2.10)$$

The ensemble spread only captures the first component. Hence, we would expect to have a spread-error correlation because the predictive variance includes the spread as a component. But, in addition to this, employing the ensemble spread alone might underestimate uncertainty by ignoring the second term on the right-hand side of (2.9) or (2.10). So the decomposition explains the coexistence of a positive spread-error correlation and a lack of calibration in the sense of underdispersion, which can be observed in several ensembles.

BMA for precipitation

In the case of precipitation we have to modify the BMA approach presented above because the predictive distribution of precipitation is far from normal for two major reasons: It has a positive probability of being equal to zero, and if it is not zero, the predictive density is skewed. Therefore, we cannot employ normal distributions to approximate the conditional PDF as in the case of temperature or sea level pressure.
Instead, we model the conditional PDF $g_m(y|f_m)$ for a given ensemble member as a mixture of a point mass at zero and a gamma distribution as proposed by Sloughter et

al. in [54].
Our model for $g_m(y|f_m)$ is divided into two parts.

First, we specify the probability of precipitation as a function of the forecast f_m by using logistic regression with a power transformation of the forecast as a predictor variable and obtain

$$\operatorname{logit} \mathbb{P}[y = 0|f_m] = \log \frac{\mathbb{P}[y = 0|f_m]}{\mathbb{P}[y > 0|f_m]} = a_{0m} + a_{1m} f_m^{1/3} + a_{2m}\delta_m, \tag{2.11}$$

where the predictor δ_m is equal to 1 if $f_m = 0$ and equal to 0 otherwise, and $\mathbb{P}[y > 0|f_m]$ denotes the probability of non-zero precipitation given f_m if f_m is the best ensemble member forecast for that day.

Second, we specify the PDF of the precipitation amount given that it is not zero. We consider the gamma distribution with shape parameter α and scale parameter β having PDF

$$h(y) = \begin{cases} \frac{1}{\beta^\alpha \Gamma(\alpha)} y^{\alpha-1} \exp\left(-\frac{y}{\beta}\right) & \text{for } y > 0, \\ 0 & \text{otherwise,} \end{cases}$$

mean $\mu = \alpha\beta$ and variance $\sigma^2 = \alpha\beta^2$, where Γ denotes the well-known gamma function from analysis. Gamma distributions are appropriate to model precipitation amounts because they can fit skewed data and are flexible.
Rather than fitting the gamma distribution to the observed precipitation amounts themselves, we fit the gamma distribution to powers of the observed values. It turns out to be the best if we fit the gamma distribution to the cube root of the observed precipitation amounts.
For the data of Sloughter et al. in [54], the mean of the fitted gamma distribution of the cube root of precipitation is approximately linear as a function of the cube root of the forecast accumulation, and the variance of the fitted gamma distribution is approximately linear as a function of the forecast.
Thus, given that forecast f_m is the best, we obtain

$$g_m(y|f_m) = \mathbb{P}[y = 0|f_m]\mathbb{1}_{\{y=0\}} + \mathbb{P}[y > 0|f_m]h_m(y|f_m)\mathbb{1}_{\{y>0\}},$$

where y denotes the cube root of the precipitation accumulation, and $\mathbb{P}[y = 0|f_m]$ is specified by (2.11).
As the conditional PDF $h_m(y|f_m)$ of the cube root precipitation amount y given that it is positive we have

$$h_m(y|f_m) = \frac{1}{\beta_m^{\alpha_m}\Gamma(\alpha_m)} y^{\alpha_m - 1} \exp\left(-\frac{y}{\beta_m}\right).$$

Moreover, the mean $\mu_m = \alpha_m\beta_m$ and the variance $\sigma_m^2 = \alpha_m\beta_m^2$ of the gamma distribution can be expressed via

$$\mu_m = b_{0m} + b_{1m} f_m^{1/3}$$

and

$$\sigma_m^2 = c_{0m} + c_{1m} f_m, \tag{2.12}$$

where f_m is the original forecast.
Since it can be observed that c_{0m} and c_{1m} in (2.12) do not vary much from one model to another, Sloughter et al. consider them to be constant across all ensemble members and thus reduce the number of parameters. Thus, we get as final BMA model for the predictive PDF of y which is the cube root of precipitation accumulation here:

$$p(y|f_1, ..., f_M) = \sum_{m=1}^{M} w_m \{\mathbb{P}[y = 0|f_m]\mathbb{1}_{\{y=0\}} + \mathbb{P}[y > 0|f_m] h_m(y|f_m)\mathbb{1}_{\{y>0\}}\}. \tag{2.13}$$

In this connection,

- w_m can be interpreted as the posterior probability of ensemble member m being the best,
- f_m is the original forecast,
- logit $\mathbb{P}[y = 0|f_m] = a_{0m} + a_{1m} + f_m^{1/3} + a_{2m}\delta_m$,
  where δ_m equals 1 if $f_m = 0$ and 0 otherwise,
- $h_m(y|f_m) = \frac{1}{\beta_m^{\alpha_m}\Gamma(\alpha_m)} y^{\alpha_m - 1} \exp\left(-\frac{y}{\beta_m}\right)$ and
- the parameters $\alpha_m = \frac{\mu_m^2}{\sigma_m^2}$ and $\beta_m = \frac{\sigma_m^2}{\mu_m}$ of the gamma distribution depend on f_m via
  $\mu_m = \alpha_m\beta_m = b_{0m} + b_{1m} f_m^{1/3}$ and $\sigma_m^2 = \alpha_m\beta_m^2 = c_0 + c_1 f_m$.

We note that the results we obtain by (2.13) in terms of the cube root of the precipitation amount can be expressed easily in terms of the original amounts.

Finally, we briefly comment on the methods employed to estimate the parameters based on data from an adequate training period.

- a_{0m}, a_{1m} and a_{2m} are member specific and estimated by logistic regression with precipitation/no precipitation as dependent variable and $f_m^{1/3}$ and δ_m as predictor variables.
- b_{0m} and b_{1m} are member specific and estimated by linear regression, where the non-zero precipitation observations are cases, the cube root of the precipitation amount is the dependent variable and the cube root of the forecasted accumulations is the predictor variable.
- $w_1, ..., w_M, c_0$ and c_1 are estimated by the ML method and the EM algorithm, respectively, from the training data.

For a detailed description, we refer to [54].

BMA for wind speed

As a further variant of the BMA method, we study BMA for wind speed forecasting as suggested by Sloughter et al. in [55]. As far as wind speed is concerned, there are two major challenges statistical postprocessing approaches have to face.
First, the distribution of wind speed is extremely skewed.
Second, wind observations are usually reported to the nearest whole unit, and any wind speeds below unit one are recorded as zero. This is a much coarser discretisation than that employed for other weather quantities.
Hence, our extended BMA approach for wind speed forecasting has to take account of both the skewness of the predictive distributions and the discreteness of the observations. Mostly, we need not model the separate probability of wind speed being equal to zero. We create a BMA variant for wind speed by taking a gamma distribution to model the component distribution for a given ensemble member.
Wind speed distributions are often modelled by Weibull densities. This is generalised by using gamma distribution fits to power transformations of the observed wind speeds. Experimentation shows that it is appropriate to model the component PDFs of wind speed as untransformed gamma distributions instead of using any power transformation. The $\Gamma(\alpha, \beta)$-distribution has PDF

$$g(y) = \begin{cases} \frac{1}{\beta^\alpha \Gamma(\alpha)} y^{\alpha-1} \exp\left(-\frac{y}{\beta}\right) & \text{for } y > 0 \\ 0 & \text{otherwise} \end{cases}$$

with mean $\mu = \alpha\beta$ and variance $\sigma^2 = \alpha\beta^2$.
By an exploratory data analysis, Sloughter et al. find in [55] that the observed wind speed is approximately linear as a function of the forecasted wind speed, and the corresponding standard deviation is also approximately linear as a function of the forecast.
Thus, we have

$$g_m(y|f_m) = \frac{1}{\beta_m^{\alpha_m} \Gamma(\alpha_m)} y^{\alpha_m - 1} \exp\left(-\frac{y}{\beta_m}\right) \tag{2.14}$$

as a model for the component gamma PDF of wind speed. Furthermore, the mean $\mu_m = \alpha_m\beta_m$ and the standard deviation $\sigma_m = \sqrt{\alpha_m}\beta_m$ of the distribution can be expressed depending on the ensemble member forecast f_m via

$$\mu_m = b_{0m} + b_{1m} f_m \text{ and } \sigma_m = c_{0m} + c_{1m} f_m.$$

By restricting the standard deviation parameters to be constant across all ensemble members, we reduce the number of parameters and thus replace c_{0m} and c_{1m} by c_0 and c_1, respectively. So we have

$$p(y|f_1, ..., f_M) = \sum_{m=1}^{M} w_m g_m(y|f_m)$$

as final model for the predictive PDF of maximum wind speed y, where g_m is defined by (2.14).

Finally, we name the standard methods we employ for the estimation of the parameters, which is based on forecast-observation pairs from an adequate training period.

- b_{0m} and b_{1m} are estimated by linear regression.
- $w_1, ..., w_M, c_0$ and c_1 are estimated by the ML method and the EM algorithm, respectively, from the training data.

For other methods that are possible and details of the parameter estimation procedure, compare [55].

BMA for wind direction

Having precise forecasts of wind direction is important in a large number of applications, like air pollution management or aircraft and ship routeing, for instance.
However, in contrast to weather quantities as temperature, sea level pressure or wind speed, which are linear variables taking values on the real line $\mathbb{R}$, wind direction is an angular variable taking values on the circle. Hence, the BMA postprocessing methods discussed before become ineffective or inapplicable.
To circumvent this problem, Bao et al. propose an effective bias correction and a BMA version tailored to wind direction in [1], the paper which we refer to in what follows.

Since wind direction is an angular variable taking values on the circle, as mentioned before, we employ degrees to describe wind directions, where 0, 90, 180 and 270 degrees correspond to a northerly, easterly, southerly and westerly wind, respectively.
We define the angular distance or the circular absolute error between two directions f and y, where f, $y \in [0, 360)$, by

$$\mathrm{AE}_{\mathrm{circ}}(f, y) = \min\{|y - f|, 360 - |y - f|\},$$

which is non-negative and has a maximum of 180 degrees.
Sometimes, it can be helpful to identify a direction y with the point $\theta(y) = e^{i\pi\frac{90-y}{180}}$ on the unit circle in the complex plane. In this case, wind directions of 0, 90, 180 and 270 degrees correspond to $i, 1, -i$ and -1, respectively, where i denotes the imaginary unit, that is, $i^2 = -1$.

For the bias correction, which is an important step in weather forecasting, we follow [1] and employ a circular-circular regression technique. For this purpose, let f and y be the predicted and the observed wind direction, respectively, and $\theta(f)$ and $\theta(y)$ denote the associated points on the unit circle in the complex plane. Then, the regression equation is given by

$$\theta(y) = \beta_0 \cdot \frac{\theta(f) + \beta_1}{1 + \bar{\beta_1}\theta(f)}, \tag{2.15}$$

where $\beta_0 \in \mathbb{C}$ with $|\beta_0| = 1$, $\beta_1 \in \mathbb{C}$, and $\bar{\beta_1}$ is the complex conjugate of β_1.
The mapping from $\theta(f)$ to $\theta(y)$ is a Moebius transformation in the complex plane and

maps the unit circle to itself. Hence, we have a nice connection to the field of complex analysis.
The regression parameters β_0 and β_1 can be interpreted as follows: β_0 is a rotation parameter and β_1 stands for pulling a direction towards a fixed angle, namely the point $\frac{\beta_1}{|\beta_1|}$ on the unit circle, with the concentration about $\frac{\beta_1}{|\beta_1|}$ increasing as $|\beta_1|$ increases.
Both β_0 and β_1 in (2.15) are estimated from training data by minimising the sum of the circular distances between the fitted bias-corrected forecasts and the corresponding verifying directions as a function of β_0 and β_1.
Besides the circular-circular regression technique, there exist other approaches to correct systematic biases, for example the median-angle correction or the mean-angle correction. For details, see Section 2.a in [1].

To model angular data like wind direction, the so-called von Mises distribution is a natural choice. An angular variable has a von Mises distribution with mean direction μ and concentration parameter $\kappa \geq 0$ if it has density

$$g(y|\mu,\kappa) = \frac{1}{360} \cdot \frac{\exp(\kappa \cdot \cos((y-\mu)\frac{\pi}{180}))}{I_0(\kappa)}$$

on the circle, where I_0 denotes a modified Bessel function of the first kind and order zero. We note that the von Mises distribution becomes a uniform distribution on the circle as κ tends to zero.
We are now in the position to develop the BMA approach for wind directions as proposed in [1]. Let $f_1, ..., f_M$ be an ensemble of forecasts that are bias-corrected. Then, the BMA predictive PDF p is a mixture

$$p(y|f_1, ..., f_M) = \sum_{m=1}^{M} w_m g_m(v|f_m, \kappa_m),$$

where the components are von Mises distributions with mean direction f_m and concentration parameter κ_m, and the BMA weights $w_1, ..., w_M$ satisfy both $w_m \geq 0$ for $m \in \{1, ..., M\}$ and $\sum_{m=1}^{M} w_m = 1$.
A standard version of this model uses a common concentration parameter κ. This simplifies and stabilises estimation, does not necessarily deteriorate the predictive performance and leads to

$$p(y|f_1, ..., f_M) = \sum_{m=1}^{M} w_m g_m(y|f_m, \kappa).$$

The weights $w_1, ..., w_M$ and the parameter κ of the component PDFs are estimated by maximum likelihood and the EM algorithm, respectively, from training data. For a detailed description of the procedure, compare Section 2.c in [1].
Moreover, there exists a specification of the precedent BMA method with a uniform mixture component, which might be useful if all member forecasts turn out to be substantially different from the verifying direction, compare Section 2.c in [1].

Weather Quantity	Model Distribution / Density
temperature	Normal distribution
sea level pressure	Normal distribution
precipitation	Gamma distribution with point mass in zero
wind speed	Gamma distribution
wind direction	von Mises distribution

Table 2.3: BMA: Weather variables and component densities

In this case, we have the so-called BMA$^+$ predictive density

$$p(y|f_1, ..., f_M) = \sum_{m=1}^{M} w_m g_m(v|f_m, \kappa) + w_{M+1} u(v),$$

where $w_m \geq 0$ for all $m \in \{1, , ..., M+1\}$, $\sum_{m=1}^{M+1} w_m = 1$, and u is the density of a uniform distribution on the circle, that is a von Mises distribution with $\kappa = 0$.
It is straightforward to adapt the EM algorithm from the BMA to the BMA$^+$ technique in order to estimate the weights $w_1, ..., w_{M+1}$ and the parameter κ.

Summary

In a nutshell, a BMA predictive PDF p is given by a mixture of the form

$$p(y|f_1, ..., f_M) = \sum_{m=1}^{M} w_m g_m(y|f_m),$$

where $f_1, ..., f_M$ denote the forecasts of an ensemble of size M, y the weather variable of interest, w_m the weights satisfying both $w_m \geq 0$ for all $m \in \{1, ..., M\}$ and $\sum_{m=1}^{M} w_m = 1$ and g_m a modeling density depending on the weather quantity of interest. In Table 2.3, we give an overview of the densities employed for modelling the different weather variables in the BMA approach.
Instead of considering wind direction and wind speed, we can also study wind vectors consisting of a horizontal u- and a vertical v-component as an equivalent representation. For details about this option, we refer to Subsection 5.1.3 in the thesis on hand. The BMA predictive PDF for the u- and the v-component, respectively, of a wind vector can be modelled by a mixture of normal distributions as in the case of temperature and sea level pressure.

For the R language and environment for statistical computing and graphics, compare [45], there exists a package called `ensembleBMA` comprising most of the methods discussed in this subsection in order to compute statistically postprocessed ensemble forecasts according to the BMA approach. It is available online at `http://www.cran.r-project.org`. We work with this tool in the applications of the ECC technique, compare Chapter 5 of this thesis.

2.2.2 Ensemble model output statistics (EMOS)

We present three versions of the EMOS approach that can be employed depending on the form of the weather quantity, where we focus on temperature, sea level pressure and wind speed.
In addition, the development of a bivariate EMOS approach applying to wind vectors is a subject of current research, and an R package tentatively called `ensembleMOS` offering several EMOS techniques for computation is under construction. The current version of `ensembleMOS` is available online at `http://www.stat.lsa.umich.edu/~bobyuen/ensembleMOS`.

Standard EMOS for temperature and sea level pressure

Starting with the standard EMOS approach, which can be applied to continuous, real-valued weather quantities like temperature and sea level pressure, that can be assumed to be normally distributed as described in [13], we suppose that $X_1, ..., X_M$ is an ensemble of individually distinguishable forecasts for a univariate weather variable Y and write

$$Y = a + b_1 X_1 + ... + b_M X_M + \epsilon \tag{2.16}$$

as a multiple linear regression equation for Y in terms of $X_1, ..., X_M$, where ϵ denotes an error term averaging to zero, and $a, b_1, ..., b_M \in \mathbb{R}$.
Moreover, let S^2 denote the ensemble variance, that is $S^2 = \frac{1}{M} \sum_{m=1}^{M} (X_m - \bar{X})^2$, where $\bar{X} = \frac{1}{M} \sum_{m=1}^{M} X_m$.
The variance of the error term ϵ in (2.16) is modelled by a linear function of the ensemble spread, namely

$$\mathrm{Var}(\epsilon) = c + dS^2, \tag{2.17}$$

where $c, d \geq 0$.
The EMOS predictive PDF then is a normal distribution $\mathcal{N}(\mu, \sigma^2)$ with EMOS predictive mean μ and variance σ^2, where

- $\mu = a + b_1 X_1 + ... + b_M X_M$ is a weighted average of the ensemble member forecasts that is bias-corrected and
- $\sigma^2 = c + dS^2$ is a linear function of the ensemble variance.

We now discuss the meaning of the parameters.

- The regression coefficient a is a correction term for the bias.
- $b_1, ..., b_M$ are regression coefficients, which reflect both the overall performance of the ensemble member models over the training period, relative to the other members, and the correlations between the ensemble members.

- c and d are variance coefficients, that reflect the ensemble spread and the performance of the ensemble mean forecast.
  If all else is equal, larger values of d suggest a more pronounced spread-error relationship.
  In the case of independency between spread and error, we will estimate d as negligibly small.

Therefore, we can say that EMOS is robust because it adopts both to the presence and the absence of significant spread-error correlation.
The challenge is now to estimate the coefficients $a, b_1, ..., b_M, c$ and d from training data. Classically, this can be handled by ML estimation. However, there is a general estimation strategy including the ML method as a special case, namely the following one. We consider so-called scoring rules, which are functions that are used for the evaluation of forecast performances. For a more detailed description of them we refer to Section 4.3. After having selected a scoring rule being appropriate to the problem we are interested in, we express the score for the training data as a function of the coefficients. Then, we optimise that function with respect to the parameter values. Since we take scoring rules to be negatively orientated, that is, the smaller, the better, we have to minimise the training score.
For reasons of robustness, we choose the continuous ranked probability score (CRPS) as a scoring rule. The CRPS is defined via

$$\operatorname{crps}(F, y) = \int_{-\infty}^{\infty} (F(x) - \mathbb{1}_{\{x \geq y\}})^2 dx,$$

where F is the predictive CDF, y the verifying observation and $\mathbb{1}$ the indicator function, see also Section 4.3.
In our situation here, we have to find the parameters $a, b_1, ..., b_M, c$ and d minimising the CRPS value for the training data.
For a CDF of a normal distribution $\mathcal{N}(\mu, \sigma^2)$ it can be calculated that

$$\operatorname{crps}(\mathcal{N}(\mu, \sigma^2), y) = \sigma \left\{ \frac{y-\mu}{\sigma} \left[\Phi\left(\frac{y-\mu}{\sigma} \right) - 1 \right] + 2\varphi\left(\frac{y-\mu}{\sigma} \right) - \frac{1}{\sqrt{\pi}} \right\}, \tag{2.18}$$

where φ and Φ denote the PDF and the CDF, respectively, of a standard normal distribution $\mathcal{N}(0, 1)$.
Furthermore, we note that the average score is defined by

$$\text{CRPS} = \frac{1}{N} \sum_{n=1}^{N} \operatorname{crps}(F_n, y_n). \tag{2.19}$$

By using (2.18) and (2.19), the training CRPS can be written as

$$\Gamma(a; b_1, ..., b_M; c; d) = \frac{1}{N} \sum_{n=1}^{N} \sqrt{c + dS_n^2} \left\{ Z_n \left[2\Phi(Z_n) - 1 \right] + 2\varphi(Z_n) - \frac{1}{\sqrt{\pi}} \right\}, \tag{2.20}$$

where $Z_n = \frac{Y_n - (a + b_1 X_{n1} + ... + b_M X_{nM})}{\sqrt{c + dS^2}}$ denotes the standardised forecast error, $X_{n1}, ..., X_{nM}$ the nth ensemble forecast in the training set, S_n^2 its variance and Y_n the nth verification. The sum extends over all training data.
The coefficient values minimising (2.20) can be determined numerically by the Broyden-Fletcher-Goldfarb-Shanno algorithm, which is implemented in the R language and environment.
We constrain the parameters c and d to be non-negative, which is not an issue for c. To guarantee that $d \geq 0$, we set $d = \delta^2$, where $\delta \in \mathbb{R}$, and optimise over δ.
In addition, the algorithm requires initial values. Those based on past experience often lead to good solutions. Unfortunately, the solution might be sensitive to the initial values, and convergence to a global extremum is not ensured.

EMOS$^+$ method for temperature and sea level pressure

As suggested in [13], a modification of the standard EMOS approach described above is given by the so-called EMOS$^+$ method in which the regression coefficients $b_1, ..., b_M$ in equation (2.16) are constrained to be non-negative, that is $b_1, ..., b_M \geq 0$.
This restriction can be very useful because negative weights seem much harder to interpret in the context of ensemble forecasts.
If EMOS weights are exactly zero, they can be seen in terms of reduced ensembles.
To get the EMOS$^+$ model, we proceed according to the following scheme:

- We estimate the parameters of the standard EMOS model as previously described by minimising the CRPS value (2.18).
- If all estimated regression coefficients are non-negative, we employ the estimates of standard EMOS for the EMOS$^+$ model and are done.
- If one or more of the $b_1, ..., b_M$ are negative, those are set to zero and (2.18) is minimised under this enforcement.
- The ensemble variance is recomputed, where only the ensemble members remaining in the regression equation are used. We subsequently employ the recomputed ensemble spread.
- We iterate these steps until all estimated $b_1, ..., b_M$ are non-negative.

EMOS for wind speed

The two EMOS variants previously discussed cannot be applied directly to non-negative weather quantities like wind speed. Hence, we propose the following modification of the EMOS technique, which can be found in [57].
Let $X_1, ..., X_M$ be an ensemble of individually distinguishable point forecasts for a univariate, $\mathbb{R}_+^0$-valued, continuous quantity Y. To take the non-negativity of the predictand

into account, we use a truncated normal predictive distribution having a cutoff at zero, that is

$$\mathcal{N}^0(\mu, \sigma^2), \text{ where } \mu = a + b_1 X_1 + ... + b_M X_M \text{ and } \sigma^2 = c + dS^2. \tag{2.21}$$

We recall that S^2 is defined by $S^2 = \frac{1}{M} \sum_{m=1}^{M} (X_m - \bar{X})^2$, where $\bar{X} = \frac{1}{M} \sum_{m=1}^{M} X_m$.
In this connection, the interpretation of the parameters is the same as in the standard EMOS method.
We obtain the EMOS predictive density f for the future variable Y by

$$f(y) = \begin{cases} \frac{\frac{1}{\sigma}\varphi\left(\frac{y-\mu}{\sigma}\right)}{\Phi\left(\frac{\mu}{\sigma}\right)} & \text{for } y > 0, \\ 0 & \text{otherwise.} \end{cases}$$

To ensure that (2.21) gives a valid probability distribution, the coefficients c and d need to be non-negative. Moreover, we enforce $b_1, ..., b_M$ to be non-negative. Therefore, we write $b_1 = \beta_1^2, ..., b_M = \beta_M^2, c = \gamma^2$ and $d = \delta^2$, where $\beta_1, ..., \beta_M, \gamma, \delta \in \mathbb{R}$.
Similar to the procedure in the other EMOS approaches, we optimise a negatively orientated proper scoring rule like the CRPS, which we employ for reasons of robustness here, as a function of the parameter values, here $a, \beta_1, ..., \beta_M, \gamma$ and δ, on training data. For a truncated normal predictive distribution and an observation $y \geq 0$, we obtain

$$\begin{aligned} \text{crps}(\mathcal{N}^0(\mu, \sigma^2), y) = \sigma \Phi\left(\frac{\mu}{\sigma}\right)^{-2} \Bigg[& \frac{y-\mu}{\sigma} \Phi\left(\frac{\mu}{\sigma}\right) \left\{ 2\Phi\left(\frac{y-\mu}{\sigma}\right) + \Phi\left(\frac{\mu}{\sigma}\right) - 2 \right\} \\ & + \; 2\varphi\left(\frac{y-\mu}{\sigma}\right) \Phi\left(\frac{\mu}{\sigma}\right) - \frac{1}{\sqrt{\pi}} \Phi\left(\sqrt{2}\frac{\mu}{\sigma}\right) \Bigg]. \end{aligned}$$

To solve our problem, we find the values of $a, \beta_1, ..., \beta_M, \gamma$ and δ which minimise the expression

$$\frac{1}{N} \sum_{n=1}^{N} \text{crps}(\mathcal{N}^0(a + \beta_1^2 X_{n1} + ... + \beta_M^2 X_{nM}, \gamma^2 + \delta^2 S_n^2), Y_n),$$

where the sum extends over the forecast cases in the training set. This is done by employing the Broyden-Fletcher-Goldfarb-Shanno algorithm implemented in R.

2.2.3 Logistic regression

In probability forecasts we consider the predictand to be a binary variable rather than a measurable quantity like a weather variable. A non-linear regression technique being convenient for probability forecasting is given by logistic regression, see [61]. If p is the probability being forecast, a logistic regression has the form

$$p = \frac{\exp(f(\mathbf{x}))}{1 + \exp(f(\mathbf{x}))}. \tag{2.22}$$

In this connection, f denotes a linear function of the predictor variables $\mathbf{x} = (x_1, ..., x_M)$, that is to say

$$f(\mathbf{x}) = b_0 + b_1 x_1 + ... + b_M x_M.$$

We note that the logistic regression equation (2.22) leads to S-shaped prediction functions being strictly bounded on the unit interval due to $0 < p < 1$ and is linear on the logistic scale

$$\log\left(\frac{p}{1-p}\right) = f(\mathbf{x}). \tag{2.23}$$

Since, in training data, the predictand values are binary, that is they have values 0 or 1, the left-hand side of (2.23) is not defined. That implies that we cannot employ standard linear regression models in order to estimate the regression coefficients although (2.23) has a linear form. Hence, we estimate the parameters via an iterative ML procedure.
Logistic regression can be employed in the context of ensemble forecast postprocessing for continuous predictands like temperature or precipitation, compare [18], [19] and [60]. In those cases, the forecast probabilities pertain to the occurrence of the verification V above or below a fixed prediction threshold that corresponds to a particular data quantile q. So we have, for example, the two-predictor logistic regression

$$p = \mathbb{P}[V \leq q] = \frac{\exp(b_0 + b_1 \bar{x}_{\text{ens}} + b_2 \bar{x}_{\text{ens}} s_{\text{ens}})}{1 + \exp(b_0 + b_1 \bar{x}_{\text{ens}} + b_2 \bar{x}_{\text{ens}} s_{\text{ens}})} \tag{2.24}$$

as proposed in [60].
In this context, b_0, b_1 and b_2 denote fitted constants, $\bar{x}_{\text{ens}}$ the ensemble mean forecast and s_{ens} the ensemble spread, namely the standard deviation.
Instead of employing the product of ensemble mean and ensemble spread, another approach in [19] uses the ensemble spread alone as the second predictor. Thus, as far as EMOS is concerned, x_1 is generally the ensemble mean and x_2 may involve the ensemble standard deviation, either alone or multiplied by the ensemble mean, if ensemble spread provides significant predictive information. If the second predictor does not improve the forecast performance, we can set $b_2 = 0$ and employ the one-predictor version of (2.24) as discussed in [18].
At the beginning, logistic regression has been employed for MOS postprocessing in the sense that separate equations for selected predictand quantile thresholds were fitted. However, according to [61], this approach has the following shortcomings:

- The method can lead to nonsense results as illustrated in the following example. We look at probability forecasts for the lower tercile $q_{1/3}$ and the upper tercile $q_{2/3}$ of the climatological distribution of a predictand. Then, we would forecast $p_{1/3} = \mathbb{P}[V \leq q_{1/3}]$ and $p_{2/3} = \mathbb{P}[V \leq q_{2/3}]$ by $\log\left(\frac{p_{1/3}}{1-p_{1/3}}\right) = f_{1/3}(\mathbf{x})$ and $\log\left(\frac{p_{2/3}}{1-p_{2/3}}\right) = f_{2/3}(\mathbf{x})$, respectively, according to (2.23). Unless $f_{1/3}(\mathbf{x})$ and $f_{2/3}(\mathbf{x})$ are exactly parallel, that is to say they only differ with respect to their intercept parameters b_0, they will cross for some values of $\mathbf{x}$. Therefore, we obtain that $p_{1/3} > p_{2/3}$, which implies that $\mathbb{P}[q_{1/3} < V < q_{2/3}] < 0$. Since probabilities are non-negative, this is of course a nonsense result.

- We need some kind of interpolation to estimate the probabilities that correspond to threshold quantiles for which regressions have not been fitted.
- We have to estimate a lot of parameters to fit many prediction equations.

To avoid those deficits, we need a well-fitting regression that can be estimated simultaneously for all forecast quantiles.
The following approach to obtain that is proposed in [61]. We extend (2.22) and (2.23) to include a non-decreasing function $g(q)$ of the threshold quantile q and fuse the equations for individual quantiles to get a single equation belonging to any quantile. So we have

$$p(q) = \frac{\exp(f(\mathbf{x}) + g(q))}{1 + \exp(f(\mathbf{x}) + g(q))} \tag{2.25}$$

or

$$\log\left(\frac{p(q)}{1 - p(q)}\right) = f(\mathbf{x}) + g(q), \tag{2.26}$$

respectively.
Equation (2.26) specifies parallel functions of $\mathbf{x}$ whose intercepts $b_0^*(q)$ show a monotone increase with q, namely

$$\begin{aligned} \log\left(\frac{p(q)}{1 - p(q)}\right) &= \underbrace{b_0 + g(q)}_{=b_0^*(q)} + b_1 x_1 + ... + b_M x_M \\ &= b_0^*(q) + b_1 x_1 + ... + b_M x_M. \end{aligned}$$

In practice, the challenge is to find a functional form $g(q)$ such that (2.25) provides better forecasts than the single-quantile approach according to (2.22) or at least forecasts being as good as those.

2.2.4 Quantile regression

An alternative approach of probabilistic weather forecasting is given by quantile regression whose goal is to make forecasts in terms of quantiles. In [4], the article which we refer to in the following, that method is discussed for precipitation.
We informally define the pth quantile as the value where the probability of an observation less than this value is p. If we specify probabilities in percent, we refer to quantiles as percentiles.
Our approach employs means of regression methods with the advantage that distributional assumptions are not necessary, and any type of essential information can be included as predictors. For instance, ensembles are not required, but possible.
First, we note that precipitation shows a discrete-continuous probability distribution. A common procedure to address this problem is to estimate the probability of precipitation first and then model the precipitation amounts conditional on the occurrence of precipitation. However, these two steps can be connected by using results from probability theory.

Probability of precipitation

Let $y_1, ..., y_L$ be binary observations taking the value 1 if precipitation occurs and 0 if there is no precipitation and being realisations of the random variables $Y_1|\mathbf{x}_1, ..., Y_L|\mathbf{x}_L$ from Bernoulli distributions. In this connection, the vector variables $\mathbf{x}_1, ..., \mathbf{x}_L$ are predictors and contain the information available about the observations at a given time before they become known. Our model is

$$Y_l|\mathbf{x}_l \sim \text{Ber}(\pi_l)$$

and

$$\Phi^{-1}(\pi_l) = \alpha_0 + \sum_{n=1}^{N} \alpha_n x_{ln} \text{ for } l \in \{1, ..., L\}, \tag{2.27}$$

where π_l denotes the probability of precipitation for case l, x_{ln} the nth component of $\mathbf{x}_l$ and Φ^{-1} the inverse $\mathcal{N}(0,1)$-distribution function. The unknown parameters $\alpha_0, \alpha_1, ..., \alpha_J$ are to be estimated.

Precipitation amounts given occurrence of precipitation

In the case of precipitation amounts, we estimate directly the quantiles we are interested in by quantile regression. We assume that $r_1, ..., r_{L^*}$ are observed precipitation amounts of cases with observed precipitation above a given lower threshold. Moreover, let $\mathbf{z}_1, ..., \mathbf{z}_{L^*}$ be the respective predictor values, where $\mathbf{z}_l = (z_{l1}, ..., z_{lM})$.
For linear quantile functions

$$q_\theta(\mathbf{z}_l; \boldsymbol{\beta}) = \beta_0 + \sum_{m=1}^{M} \beta_m z_{lm},$$

we solve the minimisation problem

$$\arg\min_{\boldsymbol{\beta}} \sum_{l=1}^{L^*} \rho_\theta(r_l - q_\theta(\mathbf{z}_l; \boldsymbol{\beta})),$$

where ρ_θ is defined by

$$\rho_\theta(u) = \begin{cases} u\theta & \text{for } u \geq 0, \\ u(\theta - 1) & \text{otherwise,} \end{cases}$$

in order to get an estimate of the θ-quantile $q_\theta(z; \boldsymbol{\beta})$, where $\theta \in (0, 1)$.
We have to repeat the minimisation for each quantile if we are interested in several quantiles. Perhaps, we also have to put constraints on the β's if we want to avoid crossing quantiles.

However, the standard quantile regression might have the following two shortcomings:

- Restricting quantiles to be linear can be unrealistic, and it might be a problem to find suitable transformations of the predictors.

- It might cause trouble to put constraints on all the β's in order to avoid crossing quantiles.

Therefore, we employ the local quantile regression (LQR), that is we assume that the quantiles are only locally linear in the neighbourhood of a given predictor value $\mathbf{z}$.
We obtain an estimate for $q_\theta(\mathbf{z};\boldsymbol{\beta})$ for a given predictor value $\mathbf{z}$ by

$$\arg\min_{\boldsymbol{\beta}} \sum_{l=1}^{L^*} w_l \rho_\theta(r_l - q_\theta(\mathbf{z}_l;\boldsymbol{\beta})),$$

where

$$q_\theta(\mathbf{z}_l;\boldsymbol{\beta}) = \beta_0 + \sum_{m=1}^{M} \beta_m(z_{lm} - z_m),$$

and z_m denotes the mth component of $\mathbf{z}$.
In contrast to the first approach, the LQR technique

- gives a weight w_l to each item in the sum and
- centers the predictors around $\mathbf{z}$ such that the estimate of $q_\theta(\mathbf{z};\boldsymbol{\beta})$ is the estimate of β_0.

Hence, it suffices to put possible constraints on β_0, and we need not do that for all β's simultaneously. We choose

$$w_l = w\left(\frac{||\mathbf{z}_l - \mathbf{z}||}{h_\lambda(\mathbf{z})}\right) \text{ for } l \in \{1, ..., L^*\}$$

as weights, where

$$w(u) = \begin{cases} (1-u^3)^3 & \text{for } u \in [0,1), \\ 0 & \text{otherwise,} \end{cases}$$

which is one possibility among many others.
In this connection, $||.||$ is the Euclidean norm, $\lambda \in (0,1]$ a smoothing parameter and $h_\lambda(\mathbf{z})$ the distance from $\mathbf{z}$ to the λL^* nearest predictor value.
Thus, cases with predictor values close to $\mathbf{z}$ have greater impact on the fit than those farther away.
Furthermore, we remark that there should be a scaling of each predictor variable before computing the weights.

Forecasting

We make forecasts for the probability of precipitation by replacing $\alpha_0, \alpha_1, ..., \alpha_J$ in (2.27) with estimates and inserting values for the predictor. This does not take account of the uncertainties in the parameter estimates. However, those are negligible compared to the model assumptions in many cases.
To make forecasts in terms of quantiles, let R be a random variable for precipitation,

and q_p the quantile we are interested in, that is $\mathbb{P}[R \leq q_p] = p$. Moreover, let $\hat{\pi}$ denote the estimated probability of precipitation and c the threshold for precipitation.
To estimate q_p, we need the estimates of the correct quantile for precipitation amounts given that there is precipitation.
By using well-known facts from probability theory, we get

$$\begin{aligned}
\mathbb{P}[R \leq q_p | R > c] &= \frac{\mathbb{P}[c < R \leq q_p]}{\mathbb{P}[R > c]} = \frac{1 - \mathbb{P}[R \leq c] - \mathbb{P}[R > q_p]}{\mathbb{P}[R > c]} \\
&= \frac{1 - (1 - \hat{\pi}) - (1 - p)}{\hat{\pi}} = \frac{\hat{\pi} - (1 - p)}{\hat{\pi}} \\
&= 1 - \frac{1 - p}{\hat{\pi}}.
\end{aligned}$$

Hence, if we want to have a forecast of the p-quantile, we have to estimate the $(1 - \frac{1-p}{\hat{\pi}})$-quantile in the preceding considerations, which is possible only if $\hat{\pi}$ is larger than $1 - p$.

To get good forecasts, we have to select the predictors carefully. For more flexibility and a possible improvement of the forecasts one might allow the quantiles to depend on different predictors. If the information consists of an raw ensemble of forecasts, that could be useful.
It is straightforward to apply the method presented here to other weather quantities like wind speed or temperature.

Since quantile regression yields quantile forecasts at a given level only, the challenge is to ensure consistency in the sense that quantiles need to be constrained to be monotonically increasing.
In the case of logistic regression, which yields probability forecasts at a given threshold only, we have already seen in Subsection 2.2.3 how to solve the problem that threshold non-exceedance probabilities need to be monotonically increasing.
The development of an analogous technique for quantile regression is a subject of current research.

Chapter 3

The ensemble copula coupling (ECC) approach

Unfortunately, all the state-of-the-art techniques described in Section 2.2 are applicable to single weather quantities at a single location and for a single look-ahead time only, as previously mentioned.
On the contrary, the novel ensemble copula coupling (ECC) approach, which is introduced in this chapter, is able to cope with spatial, temporal and cross-variate dependencies.
We start with a description of the origins of the ECC notion, which are connected to the design of the Bayesian processor of ensemble (BPE). Then, we carry on with elementary and introductory examples to illustrate the basic idea of ECC. Afterwards, we expose the complete theoretical development of our novel ECC technique. Finally, we show that it is justified to consider ECC as a copula approach by studying its relationships to so-called discrete copulas, which are introduced in this connection.

3.1 Origins of ECC: The Bayesian processor of ensemble (BPE)

In this section, we study the so-called BPE, a theoretically-based method for probabilistic forecasting of weather variables. This approach includes a crude, but very useful idea, which we develop in our novel ECC method later on. In this sense, the BPE can be considered as a contribution to the ECC notion. The BPE is an extended technique and a generalisation of the Bayesian processor of output (BPO) introduced by Krzysztofowicz and Maranzano, see [25] and [26], which processes output data from a numerical weather prediction (NWP) model and optimally fuses them with climatic data in order to quantify uncertainty about a predictand. In other words, the BPO processes single values of multiple predictors into a posterior distribution function of a predictand.
In addition to this, according to [27] and [28], the achievement of the BPE is to process an ensemble generated by multiple integrations of an NWP model and optimally fuse it

with climatic data with the intention to quantify uncertainty about the predictand, to get a full probability distribution of each predictand and to adjust the ensemble. This means that the BPE technique can map a given raw ensemble into a posterior ensemble which is calibrated, provides maximal informativeness and preserves the spatial, temporal and intervariate dependence structure of the NWP output fields. The theoretical framework of the BPE approach is given by the well-known and famous Bayesian statistical theory. Bayes' theorem and the law of the total probability are suitable and useful tools for fusing information from different sources as well as for getting the posterior distribution function of a predictand.
Hence, the BPE method shows that Bayesian theory can be applied to complex forecasting challenges and contributes to an advanced understanding of the stochastic properties of meteorological ensembles.
In the following, we summarise the main aspects of works by Bremnes, compare [5], on the one hand, and Krzysztofowicz and Toth, see [27], [28], [29], [30] and [31], on the other hand, in which the BPE is described in detail.

3.1.1 BPE for precipitation: Bremnes' contribution

In [5], the BPE method is studied for two situations. First, a model for the probability of precipitation is presented. Then, a further one for the distribution of precipitation amount given that precipitation occurs is discussed.

Probability of precipitation

The BPE approach mainly consists of the idea to transform each variable to standard normal and carry on as if the joint distribution was multivariate normal. To be precise, Bayes' rule, which is well-known from probability theory, is used to decompose the estimation in simpler tasks. In our case here, this involves estimation of the distributions of the predictor for the event that precipitation occurs and that precipitation does not occur, respectively.
To continue with the theoretical details, let Y be a binary random variable defined via

$$Y = \begin{cases} 1 & \text{if precipitation occurs,} \\ 0 & \text{if precipitation does not occur,} \end{cases}$$

and let $f_0(\mathbf{x}^*)$ and $f_1(\mathbf{x}^*)$ be the respective densities of the predictor vector $\mathbf{x}^*$ of length M, where $M \in \mathbb{N}$ denotes the number of ensemble members, when observing precipitation or not. Then, the probability of precipitation π can be calculated by

$$\begin{aligned}
\pi &= \mathbb{P}[Y=1|\mathbf{x}^*] \\
&= \frac{\pi_c f_1(\mathbf{x}^*)}{(1-\pi_c)f_0(\mathbf{x}^*)+\pi_c f_1(\mathbf{x}^*)} \\
&= \left[\frac{(1-\pi_c)f_0(\mathbf{x}^*)+\pi_c f_1(\mathbf{x}^*)}{\pi_c f_1(\mathbf{x}^*)}\right]^{-1} \\
&= \left[1+\frac{1-\pi_c}{\pi_c}\cdot\frac{f_0(\mathbf{x}^*)}{f_1(\mathbf{x}^*)}\right]^{-1}, \qquad (3.1)
\end{aligned}$$

where $\pi_c = \mathbb{P}[Y=1]$ denotes the climatological probability of precipitation, and the second equality follows from Bayes' rule, as well as from the law of the total probability. For multivariate Gaussian densities $f_0(\mathbf{x}^*)$ and $f_1(\mathbf{x}^*)$ with mean vector $\mathbf{0} = (0,...,0)$ and invertible correlation matrices Σ_0 and Σ_1, we obtain by inserting in (3.1) that

$$\begin{aligned}
\pi &= \left[1+\frac{1-\pi_c}{\pi_c}\,\frac{\frac{1}{(2\pi)^{n/2}\sqrt{\det(\Sigma_0)}}\exp\left(-\frac{1}{2}\mathbf{x}^{*t}\Sigma_0^{-1}\mathbf{x}^*\right)}{\frac{1}{(2\pi)^{n/2}\sqrt{\det(\Sigma_1)}}\exp\left(-\frac{1}{2}\mathbf{x}^{*t}\Sigma_1^{-1}\mathbf{x}^*\right)}\right]^{-1} \\
&= \left[1+\frac{1-\pi_c}{\pi_c}\sqrt{\frac{\det(\Sigma_1)}{\det(\Sigma_0)}}\exp\left(-\frac{1}{2}\left(\mathbf{x}^{*t}\Sigma_0^{-1}\mathbf{x}^*-\mathbf{x}^{*t}\Sigma_1^{-1}\mathbf{x}^*\right)\right)\right]^{-1}. \qquad (3.2)
\end{aligned}$$

However, forecasts based on output data from atmospheric models are not normally distributed.

Hence, we have to transform the original predictors $\mathbf{x}$. While a single multivariate transformation of $\mathbf{x}$ to $\mathbf{x}^*$ cannot be found easily, that of each variable is possible by means of the normal quantile transform as suggested in [25]. In [5], another technique called local quantile-to-quantile transformation (LQQT) is employed, in which quantiles in the standard normal distribution instead of the observations are used.

As a side note, we outline the technical details of the LQQT approach, which operates on each ensemble member individually and whose aim is to remove biases by evolving transformations such that the climatology or marginal distribution of the adjusted forecasts and of the observations is the same. Roughly speaking, this is achieved by ordering historical predictions and observations separately and estimating the relationship between them. Then, we create a new adjusted ensemble by applying the fitted relation to a new ensemble forecast.

To describe the procedure theoretically, let F_Y and F_X be the cumulative distribution functions for observations and forecasts, respectively. From probability theory, it is well-known that the random variable $F_Y^{-1}(F_X(X))$ has the same distribution as the observations, namely F_Y. However, this result only holds for variables which have a continuous distribution, and therefore it cannot be used directly for daily precipitation data. Furthermore, both distribution functions are unknown, and we are forced to estimate them. Our estimation of the transformation is based upon the relationship between the

ordered samples of the two quantities. Then, we make adequate adjustments to deal with no precipitation events.
We assume that the size of our training sample is N and that $y_{(1)} \leq ... \leq y_{(N)}$ and $x_{(1)} \leq ... \leq x_{(N)}$ denote ordered observations and forecasts, respectively. Then, the estimation of the transformation is based upon the pairs $(x_{(n)}, y_{(n)})$, where $n \in \{1, ..., N\}$. Let n_0 be the number of pairs for which either the observation or the forecast is zero or below the lower threshold for precipitation occurrence, $t(.)$ the estimated transformation and r randomly chosen from $\{1, 2, ..., n_0\}$.
For a new forecast x_{new}, we obtain the adjusted forecast y_{new} via

$$y_{\text{new}}(x_{\text{new}}) = \begin{cases} t(x_{\text{new}}) & \text{for } x_{\text{new}} > 0 \\ y_{(r)} & \text{for } x_{\text{new}} = 0. \end{cases}$$

The transformation $t(.)$ can be estimated by local linear regression as follows. We assume that, in the neighbourhood of x_{new}, the relation can be expressed by a linear function $t(x) = \alpha_0 + \alpha_1(x - x_{\text{new}})$ whose coefficients α_0 and α_1 are estimated by minimising the weighted least square loss function, namely

$$\sum_{n=n_0+1}^{N} (y_{(n)} - t(x_{(n)}))^2 w(x_{(n)}, x_{\text{new}}). \tag{3.3}$$

In this connection, $w(.)$ represents an appropriate weight function [1] giving more weight to the forecast values close to x_{new} than to those further away. In accordance with the definition of $t(.)$, $t(x_{\text{new}})$ is the estimate of α_0. In practice, it often suffices to estimate the relation on a fine grid and get forecasts in between by linear interpolation, so that we need not solve the minimisation problem for every new predictor value as theoretically required. Ideally, $t(x_{(n_0)})$ should take the value 0, but this need not be obeyed by (3.3). However, this constraint usually poses no problems in practical applications, especially if large training samples are involved.

Precipitation amounts

In the BPE approach for continuous variables, like precipitation amount in our case here, we first have to transform the observed precipitation amounts and each predictor variable separately to the standard normal distribution. In this connection, we have to be conscious of the fact that all distributions are conditioned on that there will be precipitation.
Let Y^* and $\mathbf{X}^*$ be random variables for precipitation observations and forecasts for days with precipitation, respectively. In practice, Y^* and $\mathbf{X}^*$ are transformations of the

[1] For the experiments and evaluations in [5],
$w(x_m, x_{\text{new}}) = \begin{cases} (1 - (\frac{|x_m - x|}{d})^3)^3 & \text{if } |x_m - x_{\text{new}}| < d, \\ 0 & \text{otherwise,} \end{cases}$
is taken as a weight function, where the constant d is the degree of smoothness in the transformation.

original variables.
Moreover, we assume that

$$Y^* \sim \mathcal{N}(\mu_0, \sigma_0^2)$$

and

$$\mathbf{X}^* | Y^* = y \sim \mathcal{N}(\boldsymbol{\alpha} + \boldsymbol{\beta}\, \mathrm{y}, \mathbf{S}),$$

which means that multivariate linear regression can be employed to model the distribution of the predictor vector given the observation, see [21].
Since, in practice, all parameters are unknown, we employ plug-in estimates to ascertain the distribution of $Y^* | \mathbf{X}^* = \mathbf{x}$.
It is important to remember that, in reality, precipitation observations and predictions do not follow a normal distribution. Therefore, we have to use transformations of those to ensure the correspondence to reality of our model.
Again, the LQQT approach is applied in order to achieve a transformation for each variable separately by means of the normal quantile transform.
Finally, the distribution of $Y^* | \mathbf{X}^* = \mathbf{x}$ allows the derivation of the conditional distribution in original units.

Application of BPE forecasts in hydrological models

For deterministic hydrological runoff models, temporal simulations of precipitation and temperature are required as input, amongst other data. Calibration of ensembles using separate BPE models for each lead time can be problematic because the BPE technique provides fully specified probability distributions as an output, but the temporal dimension is ignored. If we simply sampled from each BPE model and randomly linked the samples in time, we would disregard the inherent temporal dependence structure in the raw ensemble forecasts.
This unfortunate situation can be improved by proceeding according to the following scheme, which can be found in Section 3.3 in [5]:

1. For each lead time, we compute as many quantiles from the BPE model as there are ensemble members. It is desirable that the quantiles are evenly distributed. If the ensemble comprises M members, then we could take the $\frac{1}{M+1}, \frac{2}{M+1}, ..., \frac{M}{M+1}$ quantiles, for instance.

2. We determine the rank for each raw ensemble member at every lead time.

3. For each ensemble member, we employ its ranks to select corresponding BPE quantiles and connect those in time.

The procedure described above is very useful, and a similar concept is employed in the development of our novel ECC approach, which is presented in the next sections.

3.1.2 BPE: Krzysztofowicz' and Toth's contribution

We now get a general idea of the BPE concept by Krzysztofowicz and Toth as described in [27], [28], [29] and [30]. The aim of the development of the BPE is to provide a statistical technique for probabilistic weather forecasting in the context of ensembles based on Bayesian statistical theory. In particular, the BPE method should fulfil requirements such as

- calibration of forecasts, that is reliability and unbiasedness,
- robustness of estimators,
- structural flexibility and ability to combine information from different sources and of different types, namely ensembles from multiple centers, high resolution forecasts and observations and climatic data, respectively,
- computational efficiency and
- appropriation of applicable versions for binary, multi-category and continuous predictands.

The main information to be fused by the BPE approach stem from two sources, namely from the NWP ensemble forecasts on the one hand and the climatic data or observations on the other hand.
For the description of the theoretical framework, let Y denote a predictand, $\mathbf{X} = (X_1, ..., X_M)$ an ensemble forecast and $\mathbf{\Xi} = (\Xi_1, ..., \Xi_N)$ sufficient statistics, where $\mathbf{\Xi} = T(\mathbf{X})$ and $N < M$. Furthermore, to sketch the BPE theory for continuous predictands, let $g(y)$ be the prior density modelling the climatic uncertainty and stemming from a long climatic sample of Y and $f(\boldsymbol{\xi}|y)$ be the conditional density function modelling stochastic dependence and based on a short joint sample of $(\mathbf{\Xi}, Y)$. Stochastic dependence is often modelled by a joint density function π. Here, we consider two continuous variates interpreted as a predictand Y and a predictor $\mathbf{\Xi}$ such that

$$\pi(y, \boldsymbol{\xi}) = \phi(y|\boldsymbol{\xi})\kappa(\boldsymbol{\xi}) = f(\boldsymbol{\xi}|y)g(y), \tag{3.4}$$

where ϕ and f are conditional density functions and κ and g marginal density functions. As similarly described in [31], the two factorisations of π in (3.4) are equivalent if they satisfy

1. the law of the total probability

$$\kappa(\boldsymbol{\xi}) = \int_{-\infty}^{\infty} f(\boldsymbol{\xi}|y)g(y)dy \tag{3.5}$$

and

2. Bayes' theorem

$$\phi(y|\boldsymbol{\xi}) = \frac{f(\boldsymbol{\xi}|y)g(y)}{\kappa(\boldsymbol{\xi})}. \tag{3.6}$$

These two steps are known as the BPE.
All in all, we have to model and estimate g and f first. Then, we derive κ and ϕ according to the two steps. This leads to a complete and consistent model of stochastic dependence which employs all available and given data.
We have to note that $f(\boldsymbol{\xi}|y)$ replaces $f(\mathbf{x}|y)$ in those calculations. This is correct according to the informativeness arguments discussed above. In real time, $\boldsymbol{\xi}$ is given, and we write $\phi(y)$ instead of $\phi(y|\boldsymbol{\xi})$.
In [31], it is described how to model and estimate, respectively, f and g in the case of a univariate predictor Ξ. By inserting the properties of f and g in (3.5), we get κ, which leads to ϕ via (3.6). This procedure might be applied to multivariate predictors $\boldsymbol{\Xi}$ as ensemble forecasts similarly, for example by using multiple linear regression instead of simple linear regression as in [31] in the context of parameter estimation.

At the end of this subsection, we summarise the main properties of the BPE approach. The method provides a theoretically-based optimal fusion of an ensemble forecast with climatic data. It revises a prior climatic distribution given an ensemble forecast based on the comparison of past forecasts with observations. The theoretic details of the BPE technique are always valid, and a framework for different modelling assumptions or estimation procedures is offered. Moreover, the BPE method is able to handle distributions of arbitrary form as well as non-linear and heteroscedastic dependency structures. It is a parametric approach being robust when the joint sample is small. As an output, the BPE provides the posterior distribution with corresponding density function and the probability of non-exceedance for each member, which is identical for all predictands.

Moreover, BPE produces a calibrated posterior ensemble by mapping each member into a posterior quantile via the inverse of the posterior distribution function, see Slide 6 in [30]. In this connection, note that given the posterior distribution function, any quantile can be calculated. This procedure is very similar to the notion proposed by Bremnes in [5] in the context of the application of BPE forecasts in hydrological models, which we discussed right at the end of Subsection 3.1.1.
It is illustrated for the case of precipitation amount in Figure 3.1, which is taken from Slide 7 in [30]. Thereby, departing from our previous notation, $\mathbf{Y}$ is employed to describe ensemble forecasts, W denotes a predictand—precipitation amount in our case here—, and $\mathbf{X} = T(\mathbf{Y})$ are sufficient statistics. In particular, $y_{(1)}, y_{(2)}$ and $y_{(3)}$ denote the ordered raw ensemble forecasts and $p_{(1)}, p_{(2)}$ and $p_{(3)}$ the corresponding probabilities of non-exceedance. The postprocessed BPE ensemble is then given by the quantiles $w_{(1)}, w_{(2)}$ and $w_{(3)}$, where $w_{(1)} = \Phi^{-1}(p_{(1)}), w_{(2)} = \Phi^{-1}(p_{(2)})$ and $w_{(3)} = \Phi^{-1}(p_{(3)})$, and Φ^{-1} denotes the inverse of the posterior distribution function.
The crude, but very useful notion by Krzysztofowicz and Toth, see [30], and Bremnes, compare [5], appearing in the context of the BPE and discussed at the end of this and

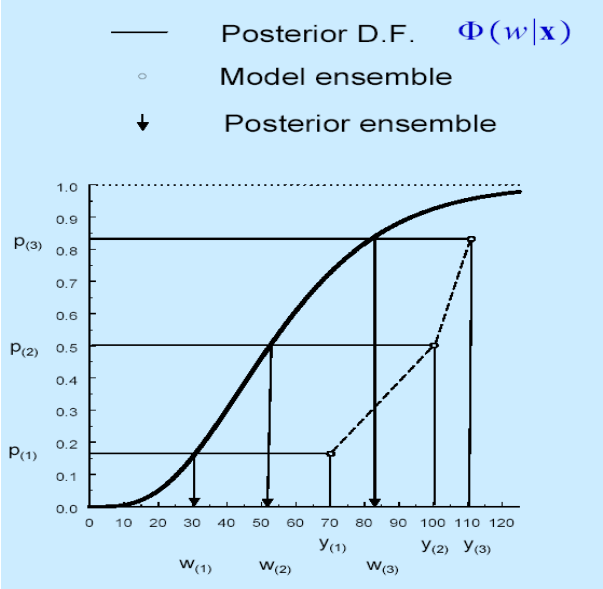

Figure 3.1: Construction of the calibrated BPE ensemble, compare Slide 7 in [30]. Take care of the different notation in comparison to that of the main text in this subsection.

the preceding subsection, respectively, will play a role in the development of the ECC approach, which we present in the next sections.

All in all, the BPE technique removes biases in the distribution, guarantees calibration of the adjusted ensemble, and spatial, temporal or cross-variate rank correlations in the raw ensemble are preserved. A scheme of the BPE approach is given in Figure 3.2, which summarises the main aspects of the BPE procedure discussed before and is based upon the information given in the references mentioned above.

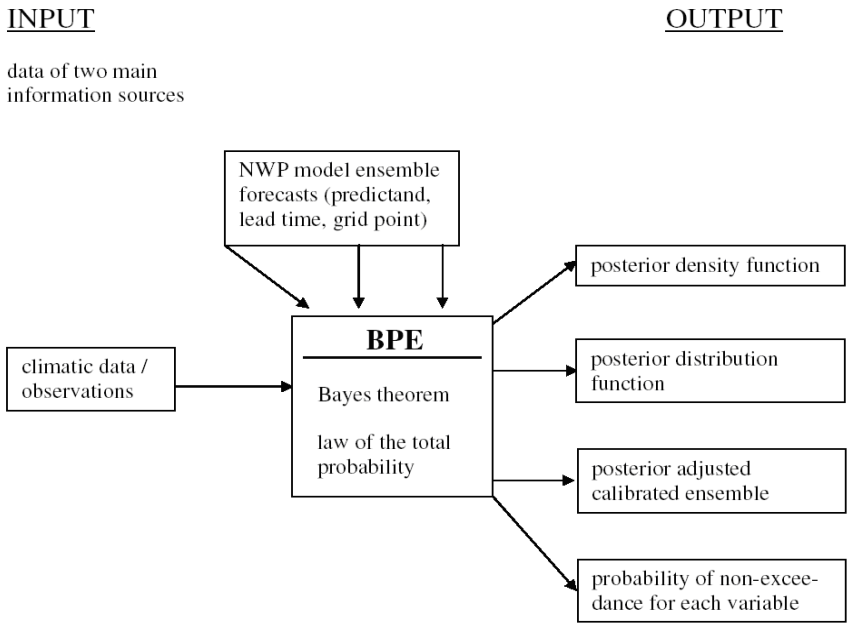

Figure 3.2: Scheme of the BPE approach

3.2 Basic idea and initial simulation examples of ECC

Before outlining its theoretical details, we describe briefly the general notion of the ECC approach and study first simulations to illustrate the potential of our novel tool.
The ECC method essentially proceeds in two steps:

1. We employ state-of-the-art ensemble postprocessing techniques as proposed in the preceding section in order to obtain calibrated and sharp univariate probabilistic forecasts for each weather variable, location and look-ahead time individually.

2. We aggregate the univariate distributions of the first step in a copula approach and ensure that the multivariate dependence structure is inherited from the original numerical weather prediction (NWP) ensemble, whence the flow dependency is captured.

To get a first impression of the idea of ECC, we consider 24 hours-ahead ensemble forecasts for temperature in Kelvin at the two locations Berlin and Hamburg, valid at 0000 Coordinated Universal Time (UTC) on 27 May 2010. The forecast data are from the European Centre for Medium-Range Weather Forecasts (ECMWF) ensemble consisting of $M = 50$ members that are exchangeable. This data set is employed in the practical part of this thesis later on, see Chapter 5.

We start with the unprocessed raw ensemble forecasts for temperature x_m in Berlin and

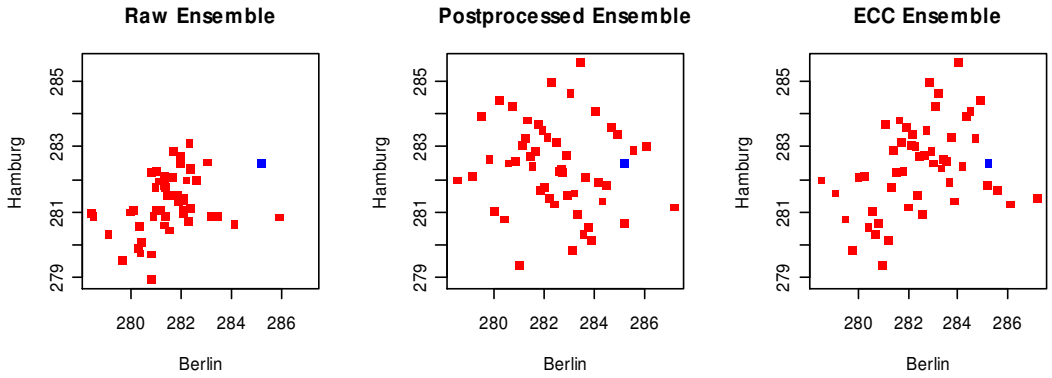

Figure 3.3: Raw ensemble, postprocessed ensemble and ECC ensemble for 24 hours-ahead forecasts for temperature in Kelvin at Berlin and Hamburg valid at 0000 UTC on 27 May 2010. The verifying observation is indicated by a blue dot.

y_m in Hamburg, where $m \in \{1, ..., 50\}$, given by the NWP models, that is

$$\underbrace{(\underbrace{x_1}_{\text{Berlin}}, \underbrace{y_1}_{\text{Hamburg}})}_{\text{ensemble member number 1}}, ..., \underbrace{(\underbrace{x_{50}}_{\text{Berlin}}, \underbrace{y_{50}}_{\text{Hamburg}})}_{\text{ensemble member number 50}}.$$

Then, we form the univariate order statistics $x_{\sigma(1)} < ... < x_{\sigma(50)}$ for Berlin and $y_{\tau(1)} < ... < y_{\tau(50)}$ for Hamburg, respectively, where σ and τ denote permutations of $\{1, ..., 50\}$. For ties, that is, if at least two ensemble forecasts have the same value, the corresponding ranks are randomly chosen from those that are possible depending on the respective situation and assigned then.
In the next step, the well-known univariate postprocessing techniques previously discussed provide predictive distributions, say F_x and F_y, for temperature at the two locations Berlin and Hamburg separately.
Afterwards, we take samples $\hat{x}_1, ..., \hat{x}_{50}$ and $\hat{y}_1, ..., \hat{y}_{50}$ from F_x and F_y, respectively.
In our case, F_x and F_y are obtained by BMA, compare Subsection 2.2.1, and the samples are just the $\frac{m}{50.5}$-quantiles, where $m \in \{1, ..., 50\}$. Details are explained later on in this thesis.
Under the assumption of independence, a first naive approach would take $(\hat{x}_1, \hat{y}_1), ...,$ $(\hat{x}_{50}, \hat{y}_{50})$ as the statistically postprocessed bivariate ensemble. This site-by-site postprocessed ensemble is bias-corrected and removes dispersion errors, but ignores dependencies, in the sense that the bivariate rank order information of the NWP raw ensemble is lost.
This shortcoming is restored by the ECC technique taking $(\hat{x}_{\sigma(1)}, \hat{y}_{\tau(1)}), ...,$ $(\hat{x}_{\sigma(50)}, \hat{y}_{\tau(50)})$ as the postprocessed bivariate ensemble, which ensures the inheritance of the bivariate dependence structure from the dynamic NWP ensemble. That is, we reorder both the $\hat{x}_m$-values following the order of the raw ensemble x_m and the $\hat{y}_m$-values following the order of the raw ensemble y_m to get a postprocessed bivariate forecast

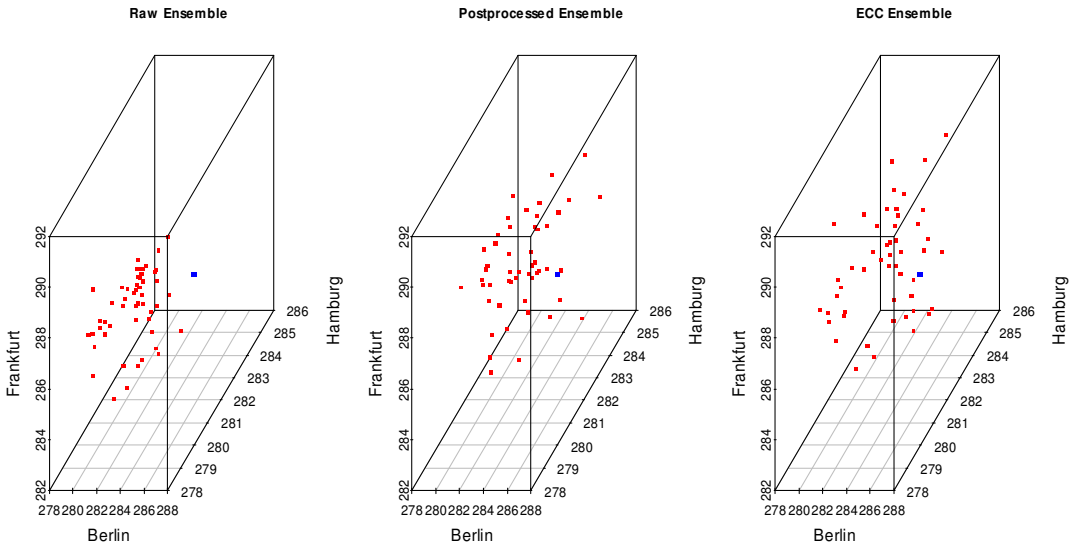

Figure 3.4: Raw ensemble, postprocessed ensemble and ECC ensemble for 24 hours-ahead forecasts for temperature in Kelvin at Berlin, Hamburg and Frankfurt valid at 0000 UTC on 27 May 2010. The verifying observation is indicated by a blue dot.

ensemble retaining the dynamic properties from the original physical model.
We remark that the site-by-site postprocessed and the ECC postprocessed ensemble forecasts possess the same univariate margins. Nevertheless, they obviously differ in the bivariate dependence structure to a great extent.

To underline our explanations, we consider the illustrations in Figure 3.3.
In the left panel, the unprocessed raw ensemble forecasts are shown.
The picture in the middle presents the naively site-by-site postprocessed ensemble forecasts. Here, we detect the loss of the bivariate rank order characteristics of the unprocessed forecasts from the left pattern, even though biases and dispersion errors have been corrected.
Finally, we have the ensemble forecasts postprocessed by the ECC method in the right panel, and see that ECC corrects biases and dispersion errors like the naively postprocessed ensemble does, but also takes account of the dependence structure given by the dynamic NWP ensemble.

It is straightforward to employ this idea for J locations, where $J \geq 3$. Another illustration for the case of $J = 3$ is given in Figure 3.4, in which we consider 24 hours-ahead raw ensemble forecasts of the 50-member ECMWF ensemble for temperature in Kelvin valid at 0000 UTC on 27 May 2010 at the three locations Berlin, Hamburg and Frankfurt and the corresponding postprocessed ensembles according to the procedure for the bivariate case in three-dimensional scatterplots. Again, it can be seen that ECC retains the rank

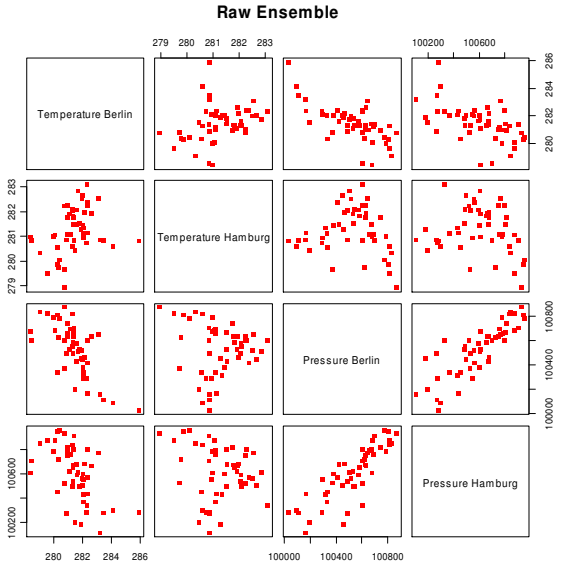

Figure 3.5: Raw ensemble for 24 hours-ahead forecasts for temperature in Kelvin and sea level pressure in Pascal at Berlin and Hamburg valid at 0000 UTC on 27 May 2010.

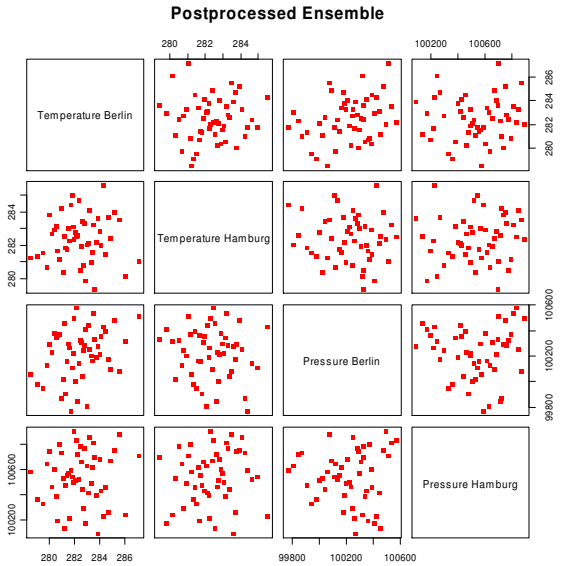

Figure 3.6: Postprocessed ensemble for 24 hours-ahead forecasts for temperature in Kelvin and sea level pressure in Pascal at Berlin and Hamburg valid at 0000 UTC on 27 May 2010.

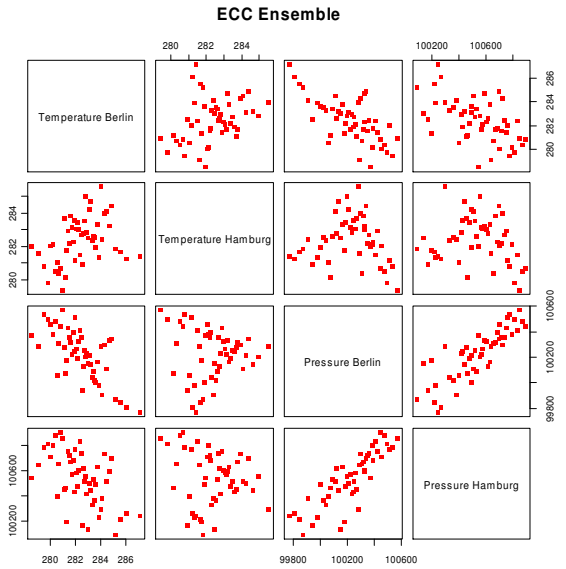

Figure 3.7: ECC ensemble for 24 hours-ahead forecasts for temperature in Kelvin and sea level pressure in Pascal at Berlin and Hamburg valid at 0000 UTC on 27 May 2010.

order information from the raw ensemble.

In a next step, we consider 24 hours-ahead raw ensemble forecasts and the corresponding postprocessed ensembles for two weather quantities at two locations. An example for our ECMWF ensemble of size 50 is given by the scatterplot matrices in Figures 3.5, 3.6 and 3.7, where we focus on forecasts for temperature in Kelvin and sea level pressure in Pascal at Berlin and Hamburg valid at 0000 UTC on 27 May 2010. We remark that the results obtained in the precedent examples are corroborated.

After having presented several examples and simulations to get an overview of the topic, we provide a complete theoretical description of the ECC method in the following section.

3.3 Theoretical development of the ECC method

In this section, we develop the ECC approach theoretically. Essentially, we generalise the idea presented in the preceding section to an arbitrary number of locations, weather quantities and prediction horizons.
For this purpose, let M denote the ensemble size, I the number of weather variables, J the number of locations and K the number of look-ahead times. Moreover, the M ensemble members are assumed to be exchangeable. Then, the ECC technique proceeds as follows:

1. We are given the raw ensemble, namely the separate forecasts $x_1, ..., x_M$ of all M ensemble members for each weather quantity i, location j and prediction horizon k individually, where $i \in \{1, ..., I\}, j \in \{1, ..., J\}$ and $k \in \{1, ..., K\}$. The raw ensemble is denoted by

$$x_1^{(i,j,k)}, ..., x_M^{(i,j,k)} \text{ or } (x_m^{(i,j,k)})_{m=1,...,M}, \text{ respectively,}$$

for each fixed $i \in \{1, ..., I\}, j \in \{1, ..., J\}$ and $k \in \{1, ..., K\}$, where m runs through all ensemble members from 1 to M.
For instance, $x_1^{(2,1,3)}, ..., x_{10}^{(2,1,3)}$ or $(x_m^{(2,1,3)})_{m=1,...,10}$, respectively, denotes the raw ensemble consisting of forecasts of 10 ensemble members for weather variable 2, location 1 and prediction horizon 3.
For reasons of simplicity and clarity, we write l for the vector consisting of the indices i, j and k, where $i \in \{1, ..., I\}, j \in \{1, ..., J\}$ and $k \in \{1, ..., K\}$, that is, we set $l := (i, j, k)$. Hence, if L denotes the number of all possible combinations of i, j and k, we have $l \in \{1, ..., L\}$ and obtain $(x_m^l)_{m=1,...,M}$ as the raw ensemble for each fixed l.

2. We determine the order statistics of the M ensemble member forecasts at each location, for each weather variable and for each prediction horizon separately. Hence, if there are no ties in the raw ensemble forecasts, we have

$$x^l_{\sigma_l(1)} < ... < x^l_{\sigma_l(M)} \text{ for each fixed } l,$$

where σ_l is a permutation of $\{1, ..., M\}$.
If there are ties in the raw ensemble forecasts, as is often the case when considering precipitation, for example, the corresponding ranks, which are natural numbers, and hence the corresponding permutation, are randomly chosen from those that are possible depending on the situation at hand.
To be precise, if there are T raw ensemble forecast values that are equal, that is, $x_{m_1} = ... = x_{m_T}$ for $m_1, ..., m_T \in \{1, ..., M\}$, we randomly choose the permutation σ_l from above which fixes the uniquely determined ranks of the raw ensemble forecasts without ties and randomly assigns the T ranks $t_1, ..., t_T$, where $t_1, ..., t_T$ are T successive natural numbers from $\{1, ..., M\}$, to the forecasts $x_{m_1}, ..., x_{m_T}$.
Note that there are $T!$ possibilities to assign T ranks to the corresponding ensemble members.
Hence, any natural number from 1 to M appears as a rank for one of the M ensemble forecasts.

3. By employing state-of-the-art statistical ensemble postprocessing methods, compare Section 2.2, we obtain calibrated and sharp univariate predictive distributions F_{x^l} for each fixed l, that is to say for each location, weather variable and look-ahead time individually.

4. Then, in each case, that is, for each l, we generate a sample of size M from the predictive distributions obtained in the preceding step and denote them by

$(\hat{x}^l_m)_{m=1,...,M}$ for each l.
For instance, those M values can be obtained by taking the equally spaced $\frac{m}{M+0.5}$-quantiles of the corresponding predictive distribution F_{x^l}, where m runs through all natural numbers from 1 to M, that is, the $\frac{1}{M+0.5}, \frac{2}{M+0.5}, ..., \frac{M}{M+0.5}$-quantiles.

5. Finally, the ECC approach takes $(\hat{x}^l_{\sigma_l(m)})_{m=1,...,M}$ for each l as the statistically postprocessed ensemble forecast. The ECC ensemble inherits the multivariate dependence structure from the raw ensemble, whose rank order information is retained.

Indeed, as previously mentioned, ECC employs elements discussed in the context of the BPE before, compare Section 3.1, and consists of two steps by firstly using ensemble postprocessing methods to get univariate predictive distributions for each location, weather variable and prediction horizon separately and then connecting them in a copula approach, while the multivariate dependence structure is inherited from the raw ensemble.
In the next section, we discuss to what extent ECC can be regarded as a copula technique by studying its relationships to this rapidly developing field.

3.4 Relationships of the ECC approach to copulas

We now outline the connections of our novel ECC approach described in the previous subsection to the field of copulas introduced in Section 2.1 and show that ECC — as suggested by the name — is indeed a copula technique. For this purpose, we have to introduce the concept of discrete copulas as proposed in [24] first.

3.4.1 Discrete copulas

The notion of discrete copulas can be seen as a discrete multivariate distribution function with uniform discrete univariate marginals. In [24], Kolesárová et al. consider the special case when the marginals coincide and correspond to the uniform probability distribution on a set of cardinality M implying that then the range of the distribution function is

$$I_M := \left\{0, \frac{1}{M}, \frac{2}{M}, ..., \frac{M-1}{M}, 1\right\} \subset \mathbb{I} = [0,1].$$

However, the discussion in [24] focuses on the bivariate case only. We generalise certain definitions and results by extending them to the multivariate situation required by the ECC approach. If such a generalisation is not possible or not obvious at all, we comment on the results in [24] for the bivariate case.
Before starting with the technical details, we want to emphasise the relevance of discrete copulas to the ECC method. As an initial example, we consider a bivariate random vector $H = (X, Y)$ for a population of size M. For instance, in terms of ECC, we can sample two-dimensional data corresponding to 24 hours-ahead forecasts for temperature at Berlin X and Hamburg Y for an ensemble consisting of M members. If we assume

that there are no ties among the forecast values, then both margins X and Y of the bivariate distribution of H are uniformly distributed with step $\frac{1}{M}$, and the distribution functions F and G, respectively, of the margins X and Y take values in I_M, that is, Ran(F)=Ran(G)=I_M. According to the discrete version of Sklar's theorem, which will be presented in Theorem 3.8 in Subsection 3.4.3 later on, there exists a unique discrete copula D on I_M such that $H(x,y) = D(F(x), G(y))$, that is, the bivariate distribution is connected to its univariate margins by the discrete copula D.
After this illustration, we switch to the general situation by considering random vectors $(X^1, ..., X^L)$ and note that even in case of ties, the range of the distribution functions of the X_l, where $l \in \{1, ..., L\}$, is not I_M, but a proper subset of I_M that includes 0 and 1. The main idea is that the joint distribution of $(X^1, ..., X^L)$ can be expressed in terms of the subcopula C_M of some copula C. Hence, discrete copulas contribute to a better understanding of the dependence structure of the multivariate distribution.
We continue with the theoretical development of the concept of discrete copulas and show the relationship to the setting of our ECC approach.

Discrete copulas are a special class of subcopulas D, compare Definition 2.10 in Subsection 2.1.2, defined on $\text{Dom}(D) = I_{M_1} \times \cdots \times I_{M_L}$, where $I_{M_l} := \left\{0, \frac{1}{M_l}, \frac{2}{M_l}, ..., \frac{M_l-1}{M_l}, 1\right\}$ is a subset of $\mathbb{I}$ containing 0 and 1, and $l \in \{1, ..., L\}$.

Definition 3.1: A function $D : I_{M_1} \times \cdots \times I_{M_L} \to \mathbb{I}$ with the following properties is called an L-dimensional discrete copula or a discrete L-copula on $I_{M_1} \times ... \times I_{M_L}$.

1. D is grounded, that is $D(\mathbf{u}) = 0$ for all $\mathbf{u} = (u_1, ..., u_L) \in \text{Dom}\ (D)$ such that $u_l = 0$ for at least one $l \in \{1, ..., L\}$.

2. D has one-dimensional margins D_l, that is, $D_l(x) = D(1, ..., 1, \underbrace{x}_{\text{position } l}, 1, ..., 1) = x$ for all $x \in I_{M_l}, l \in \{1, ..., L\}$. Hence, if all coordinates of $\mathbf{u} \in \text{Dom}(D)$ are 1 except u_l, then $D(\mathbf{u}) = u_l$.

3. D is L-increasing, that is $V_D([\mathbf{a}, \mathbf{b}]) \geq 0$ for every $\mathbf{a}, \mathbf{b} \in \text{Dom}(D)$ such that $\mathbf{a} \leq \mathbf{b}$, where $V_D([\mathbf{a}, \mathbf{b}]) = \Delta_{\mathbf{a}}^{\mathbf{b}} D(\mathbf{u}) = \Delta_{a_L}^{b_L} ... \Delta_{a_1}^{b_1} D(\mathbf{u})$ with

$$\Delta_{a_l}^{b_l} D(\mathbf{u}) = D(u_1, ..., u_{l-1}, b_l, u_{l+1}, ..., u_L) - D(u_1, ..., u_{l-1}, a_l, u_{l+1}, ..., u_L),$$

compare Remark 2.3 in Subsection 2.1.1.

In what follows, we consider the situation in our ECC approach and stick to the notation used in the previous section. Then, we have $M := M_1 = \cdots = M_L$ and thus $I_M = \left\{0, \frac{1}{M}, \cdots, \frac{M-1}{M}, 1\right\}$, where M denotes the number of ensemble members involved in the ECC method. Moreover, L can be interpreted as the number of all possible states of $l := (i, j, k)$, that is $l \in \{1, ..., L\}$, where we recall that $i \in \{1, ..., I\}$ denotes a weather quantity, $j \in \{1, ..., J\}$ a location and $k \in \{1, ..., K\}$ a prediction horizon among the I weather variables, J locations and K look-ahead times we consider.

In our situation, which we refer to with immediate effect, we define discrete copulas as follows.

Definition 3.2: A function $D : \underbrace{I_M \times \cdots \times I_M}_{L \text{ times}} \to \mathbb{I}$ is called a discrete (L-)copula on $I_M \times \cdots \times I_M$ if

- D is grounded,
- D has one-dimensional margins D_l satisfying $D_l(t) = t$ for all $t \in I_M$, where $l \in \{1, ..., L\}$, and
- D is L-increasing.

From now on, we write D_M instead of D in this case in order to underline that we deal with discrete copulas.

Remark 3.3: Since I_M is a subset of $\mathbb{I}$ containing 0 and 1, and by setting $S_k = I_M$ for all $k \in \{1, ..., n\}$ in Definition 2.10 in Subsection 2.1.2, we see that every discrete copula is indeed a subcopula by definition.

Remark 3.4: Every L-subcopula and hence every discrete L-copula can be extended to an L-copula, see [40]. This can be done by a multilinear interpolation of the subcopula to a copula. For the bivariate case of $L = 2$, a proof can be found in [40]. In the multivariate case of $L > 2$, the proof is more involved, compare [53].
Due to this possible extension, any discrete copula D_M on $I_M \times \cdots \times I_M$ is the restriction to $I_M \times \cdots \times I_M$ of some copula C, that is, we have $D_M = C|_{I_M \times \cdots \times I_M}$. In general, for a given discrete copula D_M, the copula C whose restriction is equal to D_M need not be unique.
Conversely, the restriction to $I_M \times \cdots \times I_M$ of a given copula C is a discrete copula.

3.4.2 Discretisation of copulas

Both the following definition and the subsequent results about the discretisation of copulas are versions for the multivariate case obtained by extending the corresponding concepts for the bivariate case in Theorems 1 and 2 in [24].

Definition 3.5: For a given copula C, its restriction D_M to $\underbrace{I_M \times \cdots \times I_M}_{L \text{ times}}$ is called the discretisation of order M of C and we have

$$D_M\left(\frac{i_1}{M}, ..., \frac{i_L}{M}\right) = C\left(\frac{i_1}{M}, ..., \frac{i_L}{M}\right),$$

where $i_l \in \{0, 1, ..., M\}$ for $l \in \{1, ..., L\}$.

Theorem 3.6: For a copula C and its discretisation D_M of order M, we have

$$\lim_{M\to\infty} D_M\left(\frac{[Mu_1]}{M},...,\frac{[Mu_L]}{M}\right) = \lim_{M\to\infty} C\left(\frac{[Mu_1]}{M},...,\frac{[Mu_L]}{M}\right) = C(u_1,...,u_L),$$

where $u_1,...,u_L \in \mathbb{I}$, and $[x]$ is the integer part of $x \in \mathbb{R}$.

Hence, every copula is the limit of the sequence of its discretisations. Conversely, the following theorem holds.

Theorem 3.7: Let $r \geq 2$ be a natural number and $\{D_{r^s} | s \in \mathbb{N}, s \geq 2\}$ be a sequence of discrete copulas satisfying the condition $D_{r^s}(\frac{ri_1}{r^s},...,\frac{ri_L}{r^s}) = D_{r^{s-1}}(\frac{i_1}{r^{s-1}},...,\frac{i_L}{r^{s-1}})$ for $0 \leq i_1,...,i_L \leq r^{s-1}$. Then, the limit of the sequence $\{D_{r^s}(\frac{[r^s u_1]}{r^s},...,\frac{[r^s u_L]}{r^s})\}$ exists at every point $(u_1,...,u_L) \in \mathbb{I}^L$ and the function

$$C : \mathbb{I}^L \to \mathbb{I},\ C(u_1,...,u_L) := \lim_{s\to\infty} D_{r^s}\left(\frac{[r^s u_1]}{r^s},...,\frac{[r^s u_L]}{r^s}\right)$$

is a copula.

3.4.3 Discrete version of Sklar's theorem

In Chapter 2, we have stated that Sklar's theorem, see Theorems 2.20 and 2.23, respectively, is a very important and fundamental result in the theory of copulas in general. Thus, it is natural to search for a corresponding theorem which is valid in the discrete setting we deal with in this section. In [35], Mayor et al. propose a discrete version of Sklar's theorem, which is presented in our terminology in the following result.

Theorem 3.8: Let F and G be distribution functions such that Ran $(F) \subseteq I_M$ and Ran $(G) \subseteq I_M$, where $M \geq 1$.
If D_M is a discrete copula on I_M, then

$$H(u,v) = D_M(F(u),G(v)) \text{ for all } (u,v) \in \bar{\mathbb{R}}^2$$

is a joint distribution function with $\mathrm{Ran}(H) \subseteq I_M$ having F and G as marginal distribution functions.
Conversely, if H is a joint distribution function with marginal distribution functions F and G and $\mathrm{Ran}(H) \subseteq I_M$, then there is a discrete copula D_M on I_M such that

$$H(u,v) = D_M(F(u),G(v)) \text{ for all } (u,v) \in \bar{\mathbb{R}}^2.$$

Moreover, this copula D_M is uniquely determined if and only if $\mathrm{Ran}(F)=\mathrm{Ran}(G)=I_M$.

Proof: See the proof of Theorem 1 in [35]. Note that the proof in [35] is given for the equivalent scale $I_M = \{0,...,M\}$. □

Furthermore, the following result holds, compare Corollary 2 in [35].

Corollary 3.9: Given F and G with Ran(F)=Ran(G)=I_M, there are exactly $M!$ joint distribution functions H with Ran(H)$\subseteq I_M$ that have F and G as their marginal distributions.

Remark 3.10: Following [35], Theorem 3.8 is formulated for the bivariate situation here. However, a generalisation of the discrete version of Sklar's theorem as stated above to the multivariate case is very likely to be valid, too. Nevertheless, a rigorous mathematical proof by extending the ideas and techniques used in the proof for the bivariate case in [35] to higher dimensions has to be worked out.

3.4.4 Irreducible discrete copulas

Discrete copulas defined on $\underbrace{I_M \times \cdots \times I_M}_{L \text{ times}}$ allow to describe all L-variate random vectors having discrete uniformly distributed marginals.
Note that $I_M \subseteq$ Ran(D_M) for any discrete copula D_M, compare [24].
A special class of discrete copulas is given by the so-called irreducible discrete copulas.

Definition 3.11: A discrete (L-)copula $D_M : \underbrace{I_M \times \cdots \times I_M}_{L \text{ times}} \to \mathbb{I}$ is called irreducible if it has minimal range, that is, Ran(D_M)= I_M. We denote the family of the irreducible discrete copulas on I_M by $\mathcal{B}_M$.

According to [37], the practical frame of irreducible discrete copulas is given by the following considerations. For a population of size M, we look at a random vector $\mathbf{H} := (X^1, ..., X^L)$. In terms of the ECC method, we can think of an ensemble consisting of M members and random variables $X^1, ..., X^L$ describing predictive distributions for a single weather variable at a single location and for a single look-ahead time. We can sample M times L-dimensional data corresponding to $X^1, ..., X^L$.
For reasons of simplicity, we assume that all sample values of $X^1, ..., X^L$ are different. Then, $X^1, ..., X^L$ are uniformly distributed and Ran(F_{X^1}) $= \cdots =$ Ran(F_{X^L}) $= I_M$, where F_{X^l} denotes the distribution function of X^l for $l \in \{1, ..., L\}$. Furthermore, we have Ran($F_{\mathbf{H}}$) $= I_M$ for the joint distribution function $F_{\mathbf{H}}$ of $\mathbf{H}$, and hence the subcopula $D_M : \underbrace{I_M \times \cdots \times I_M}_{L \text{ times}} \to I_M$, which describes exactly this situation, is an irreducible discrete copula.
The following theorem characterises irreducible discrete copulas in the bivariate case and can be found in [24].

Theorem 3.12: A function $D_M : I_M \times I_M \to \mathbb{I}$ is an irreducible discrete copula if

and only if there is a permutation σ of $\{1, ..., M\}$ such that

$$D_M\left(\frac{i}{M}, \frac{j}{M}\right) = \frac{1}{M}\sum_{r=1}^{j} \mathbb{1}_{\{1,...,i\}}(\sigma(r)) \text{ for } i, j \in \{0, ..., M\}. \tag{3.7}$$

Moreover, there are exactly $M!$ irreducible discrete copulas on I_M.

Proof: A proof for the equivalent scale $I_M = \{0, ..., M\}$ is given in [34], compare Proposition 6 and Corollary 1 in that paper. □

In the context discussed before, the discrete probability distribution of the random vector $\mathbf{H} := (X^1, X^2)$ can be expressed by a matrix $P_{\mathbf{H}} = (p_{ij}^{\mathbf{H}})_{i,j=1}^{M}$, where $p_{ij}^{\mathbf{H}} = \mathbb{P}[X^1 = x_{(i)}^1, X^2 = x_{(j)}^2]$, and $x_{(i)}^1$ and $x_{(j)}^2$ denote the corresponding order statistics from samples that describe values of X^1 and X^2, respectively, see [37]. We assume that all observed values of X^1 and X^2, respectively, are different and recall the definition of a permutation matrix, compare [34].
An $(M \times M)$-permutation matrix A is an $(M \times M)$-matrix $A = (a_{ij})_{i,j=1}^{M}$ such that there is a permutation σ of $\{1, ..., M\}$ such that $a_{ij} = \begin{cases} 1 & \text{if } i = \sigma(j) \\ 0 & \text{otherwise} \end{cases}$, which is equivalent to the fact that in each row and in each column of A all entries are equal to 0 except one which takes the value 1. Then, $p_{ij}^{\mathbf{H}} \in \{0, \frac{1}{M}\}$ for all i, j and $MP_{\mathbf{H}}$ is a permutation matrix characterised by a permutation $\sigma_{\mathbf{H}}$ of $\{1, ..., M\}$, $Mp_{ij}^{\mathbf{H}} = \mathbb{1}_{\sigma_{\mathbf{H}}(i)}(j)$.
The irreducible discrete copula D_M written as an $(M \times M)$-matrix $D = (d_{ij}^{\mathbf{H}})_{i,j=1}^{M}$ is the cumulative counterpart of $P_{\mathbf{H}}$, namely

$$d_{ij}^{\mathbf{H}} = \sum_{r \le i, s \le j} p_{rs} = \frac{1}{M}\sum_{s \le j} \mathbb{1}_{\{\sigma_{\mathbf{H}}(1),...,\sigma_{\mathbf{H}}(i)\}}(s) = \text{card}(\{\sigma_{\mathbf{H}}(1), ..., \sigma_{\mathbf{H}}(i)\} \cap \{1, ..., j\}).$$

Conversely, we have $p_{ij}^{\mathbf{H}} = d_{ij}^{\mathbf{H}} - d_{i-1,j}^{\mathbf{H}} - d_{i,j-1}^{\mathbf{H}} + d_{i-1,j-1}^{\mathbf{H}}$, where $d_{0j}^{\mathbf{H}} = d_{i0}^{\mathbf{H}} = 0$ by convention.
In a nutshell, the joint distribution function $F_{\mathbf{H}}$ is fully characterised by the marginal distribution functions F_{X^1} and F_{X^2} and the corresponding permutation $\sigma_{\mathbf{H}}$, which specifies the dependence structure of $\mathbf{H}$, see [37].

Example 3.13: In the style of the illustrations in [37], we discuss a specific example for the considerations above. We consider 24 hours-ahead forecasts for temperature at Berlin, Hamburg and Frankfurt described by random variables X^1, X^2 and X^3, respectively. We may think of statistically postprocessed ensemble forecasts obtained by BMA such that the setting is similar to that of the initial simulation examples in Section 3.2, for instance. However, for reasons of clarity, we confine ourselves to an ensemble of size 5 in this example here.
Let the forecast data be as in Table 3.1. The temperature values are given in Kelvin. The irreducible discrete copula $D_5^{\mathbf{H}} : I_5 \times I_5 \to I_5$ which describes the random vector $\mathbf{H} = (X^1, X^2)$, given in Table 3.1, can be expressed by a (5×5)-matrix $D := (d_{ij}^{\mathbf{H}})_{i,j=1}^{5}$

Ensemble Member	Temp. Berlin X^1 X^1-Order	Temp. Hamburg X^2 X^2-Order	Temp. Frankfurt X^3 X^3-Order
1	289.2	288.7	289.8
	1	4	4
2	290.4	289.5	290.1
	5	5	5
3	289.6	288.4	289.4
	3	2	1
4	289.4	288.3	289.5
	2	1	2
5	289.8	288.6	289.7
	4	3	3

Table 3.1: Data for Example 3.13

with $d_{ij}^{\mathbf{H}} = D_5^{\mathbf{H}}(\frac{i}{5}, \frac{j}{5})$.
Since copulas are grounded, that is, $C(0,x) = C(x,0) = 0$ for any (discrete) copula C, we need not describe the values $D_5^{\mathbf{H}}(0, \frac{i}{5}) = D_5^{\mathbf{H}}(\frac{i}{5}, 0) = 0$ for each $i \in \{0,1,...,5\}$. We obtain

$$P_{\mathbf{H}} = \begin{pmatrix} 0 & 0 & 0 & \frac{1}{5} & \frac{1}{5} \\ \frac{1}{5} & \frac{1}{5} & \frac{1}{5} & \frac{2}{5} & \frac{2}{5} \\ \frac{1}{5} & \frac{2}{5} & \frac{2}{5} & \frac{3}{5} & \frac{3}{5} \\ \frac{1}{5} & \frac{2}{5} & \frac{3}{5} & \frac{4}{5} & \frac{4}{5} \\ \frac{1}{5} & \frac{2}{5} & \frac{3}{5} & \frac{4}{5} & 1 \end{pmatrix}.$$

Observe that $d_{ij}^{\mathbf{H}} = \mathbb{P}[\mathbf{H} \in (-\infty, u) \times (-\infty, v)]$, where $u, v \in \mathbb{R}$, denote the values, for which $F_{X^1}(u) = \mathbb{P}[X^1 < u] = \frac{i}{5}$ and $F_{X^2}(v) = \mathbb{P}[X^2 < v] = \frac{j}{5}$, that is, $d_{ij}^{\mathbf{H}} = \mathbb{P}[X^1 < u, X^2 < v]$.
For instance, if we consider the case $i = j = 2$, then any $u \in (289.4, 289.6]$ and $v \in (288.4, 288.6]$ fit $F_{X^1}(u) = F_{X^2}(v) = \frac{2}{5}$. Furthermore,

$$d_{22}^{\mathbf{H}} = \mathbb{P}[X^1 < 289.6, X^2 < 288.6] = \mathbb{P}[\mathbf{H} \in \{(289.4, 288.3)\}] = \frac{1}{5}.$$

By employing the procedure proposed right before this example, we get

$$P_{\mathbf{H}} = \begin{pmatrix} 0 & 0 & 0 & \frac{1}{5} & 0 \\ \frac{1}{5} & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{5} & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{5} & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{5} \end{pmatrix},$$

where $P_{\mathbf{H}} = (p_{ij}^{\mathbf{H}})_{i,j=1}^5$ describes the discrete probability distribution of $\mathbf{H} = (X^1, X^2)$ and $p_{ij}^{\mathbf{H}} = \mathbb{P}[X^1 = x_{(i)}^1, X^2 = x_{(j)}^2]$. Thereby, $x_{(i)}^1$ and $x_{(j)}^2$ denote the corresponding order

statistics from samples that describe values of X^1 and X^2, respectively.
Hence, in our situation, we obtain the permutation matrix

$$5\,P_{\mathbf{H}} = \begin{pmatrix} 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix},$$

which is characterised by the permutation $\sigma_{\mathbf{H}} = (4, 1, 2, 3, 5)$ of $\{1, ..., 5\}$, that is,

$$\sigma_{\mathbf{H}} = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ \sigma_{\mathbf{H}}(1) & \sigma_{\mathbf{H}}(2) & \sigma_{\mathbf{H}}(3) & \sigma_{\mathbf{H}}(4) & \sigma_{\mathbf{H}}(5) \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 4 & 1 & 2 & 3 & 5 \end{pmatrix}.$$

To illustrate the connection of $\sigma_{\mathbf{H}}$ to the data given in Table 3.1, we enumerate the order of the forecasts for temperature in Berlin and Hamburg. The results can be found in Table 3.1, too. After having reordered the forecasts following the X^1-order, we see that $\sigma_{\mathbf{H}}$ corresponds to the record of X^2-order. Moreover, we recall Table 3.1 and observe that it is possible to compute the permutation $\sigma_{\mathbf{G}}$ describing the dependence structure of the random vector $G = (X^2, X^3)$ without knowing $D_5^{\mathbf{G}}$ or the corresponding matrix, see Table 3.1, from which we derive that $\sigma_{\mathbf{G}} = (2, 1, 3, 4, 5)$.

Remark 3.14: For ordinary bivariate copulas, a product structure can be defined, compare [6]. Its irreducible discrete counterpart can be specified by the composition of the corresponding permutations, see [34]. According to [37], the statistical interpretation is the following one: If $\sigma_{\mathbf{H}}$ characterises the dependence structure of the random vector $\mathbf{H} = (X^1, X^2)$, and $\sigma_{\mathbf{G}}$ the one of $\mathbf{G} = (X^3, X^2)$, then the dependence structure of $\mathbf{Q} = (X^1, X^3)$ is specified by $\sigma_{\mathbf{Q}} = \sigma_{\mathbf{H}} \circ \sigma_{\mathbf{G}} = \sigma_{(X^1,X^2)} \circ \sigma_{(X^2,X^3)}$.

Example 3.15: We continue Example 3.13 and stick to the data and notation employed therein. By applying Remark 3.14, we obtain that the permutation $\sigma_{\mathbf{Q}}$ describing the dependence structure of $\mathbf{Q} = (X^1, X^3)$ is given by

$$\sigma_{\mathbf{Q}} = \sigma_{\mathbf{H}} \circ \sigma_{\mathbf{G}} = (4, 1, 2, 3, 5) \circ (2, 1, 3, 4, 5) = (4, 2, 1, 3, 5).$$

To understand this result, we recall the ordered X^1-sample (289.2,289.4,289.6,289.8.290.4) and the ordered X^3-sample (289.4,289.5,289.7,290.1).
Since $\sigma_{\mathbf{Q}}(1) = 4$, $\mathbf{Q}$ should contain the pair (289.2,289.8) in its sample.
Analogously, we derive that due to $\sigma_{\mathbf{Q}}(2) = 2$, $\sigma_{\mathbf{Q}}(3) = 1$, $\sigma_{\mathbf{Q}}(4) = 3$ and $\sigma_{\mathbf{Q}}(5) = 5$, respectively, $\mathbf{Q}$ should also contain the pairs (289.4,289.5), (289.6,289.4), (289.8,289.7) and (290.4,290.1), respectively, in its sample.
Obviously, the pairs named above exactly correspond to the forecast data introduced in Table 3.1.

3.4.5 Interpretation of ECC as a copula approach

We now want to show the relationship between the facts about discrete copulas we collected and our ECC technique.
In the bivariate case, we know that each irreducible discrete copula related to a permutation σ describes the order statistics of x^1 and x^2 samples that are coupled together, that is, if $x^1_{(i)}$ is the j-th order statistic in the x^1 sample, then $x^2_{(i)}$ is the $\sigma(j)$-th order statistic in the x^2 sample, compare [37] or Section 4 in [38].
For the L-dimensional case, where $L > 2$, we employ the following very important and useful remark, which can be found in section 4 in [38], to demonstrate the connections between discrete copulas and ECC.

Remark 3.16: $D_M : \underbrace{I_M \times \cdots \times I_M}_{L\,\text{times}} \to I_M$ is an irreducible discrete copula if and only if there are permutations $\sigma_1, ..., \sigma_L$ of $\{1, ..., M\}$ such that the sample

$$(x^1_1, x^2_1, ..., x^L_1), (x^1_2, x^2_2, ..., x^L_2), ..., (x^1_M, x^2_M, ..., x^L_M) \text{ or — in short — } (x^l_m)^{l=1,...,L}_{m=1,...,M}$$

with distinct values on each fixed coordinate can be written as

$$(x^1_{(\sigma_1(1))}, x^2_{(\sigma_2(1))}, ..., x^L_{(\sigma_L(1))}), (x^1_{(\sigma_1(2))}, x^2_{(\sigma_2(2))}, ..., x^L_{(\sigma_L(2))}), ...,$$
$$(x^1_{(\sigma_1(M))}, x^2_{(\sigma_2(M))}, ..., x^L_{(\sigma_L(M))}) \text{ or — in short — } (x^l_{(\sigma_l(m))})^{l=1,...,L}_{m=1,...,M},$$

where $x^i_{(j)}$ is the j-th order statistic in the sample from the i-th coordinate.

We now comment on this result with a view to the ECC method.
The remark explains why it is justified to consider ECC as a copula approach because it establishes a correspondence of the ECC procedure to irreducible discrete copulas. The second part in Remark 3.16 describes exactly the steps to be done in the ECC approach if we identify the notation and the elements, respectively, as follows:

- $m \in \{1, ..., M\}$ denotes an ensemble member among the ensemble of size M.
- $l = (i, j, k)$ denotes a combination of $i \in \{1, ..., I\}$, $j \in \{1, ..., J\}$ and $k \in \{1, ..., K\}$, where i represents one of I weather variables, j one of J locations and k one of K prediction horizons.
  If L is the number of all possible combinations of i, j and k, we have $l \in \{1, ..., L\}$. Hence, L can also be interpreted as the number of univariate statistically postprocessed prediction distributions for a single location, a single weather quantity and a single look-ahead time in the ECC approach.
- The permutations $\sigma_1, ..., \sigma_L$ of $\{1, ..., M\}$ in terms of ECC are just those generating the order statistics of the raw ensemble.
- In the ECC method, the sample $(x^l_m)^{l=1,...,L}_{m=1,...,M}$ in Remark 3.16 is just the sample $(\hat{x}^l_m)^{l=1,...,L}_{m=1,...,M}$ of size M for fixed l from the univariate distribution F_{x^l} obtained by

state-of-the-art univariate postprocessing methods, using for example the $\frac{m}{M+0.5}$-quantiles, where $m \in \{1, ..., M\}$.

If Ξ^l for fixed $l \in \{1, ..., L\}$ denotes the discrete random variable that can take the values $(\hat{x}^l_m)_{m=1,...,M} = (\hat{x}^l_1, ..., \hat{x}^l_M)$ with corresponding distribution function G_{Ξ^l} taking values in I_M, then, according to Sklar's theorem, we have H defined by

$$H(u_1, ..., u_L) := D_M(G_{\Xi^1}(u_1), ..., G_{\Xi^L}(u_L)) \text{ for all } (u_1, ..., u_L) \in \bar{\mathbb{R}}^L$$

where D_M is the irreducible discrete copula on I_M from Remark 3.16, as a joint distribution function with $\text{Ran}(H) \subseteq I_M$ having $G_{\Xi^1}, ..., G_{\Xi^L}$ as marginal distribution functions.

Hence, according to Remark 3.16, ECC is clearly linked to irreducible discrete copulas, and there is indeed a justification to speak of ECC as a copula approach.

3.4.6 Several properties of discrete copulas in the bivariate case

Finally, for reasons of completeness, we discuss some interesting properties of discrete copulas in the bivariate situation only, compare [24]. They do not go with our general ECC approach, but of course they can be employed if ECC is applied in a bivariate case, for example for the statistical postprocessing of ensemble forecasts for a single weather quantity and a single prediction horizon at two locations.
Before starting, we recall the definition of a bistochastic matrix from probability theory.

Definition 3.17: A matrix $A = (a_{ij})_{i,j=1}^M$ is called bistochastic if all its entries are non-negative and the sum of the entries of every row and every column is equal to 1, that is, $a_{ij} \geq 0$ for all i, j and $\sum_{i=1}^M a_{im} = \sum_{j=1}^M a_{mj} = 1$ for all $m \in \{1, ..., M\}$.

The following theorem deals with the relationship of discrete copulas and bistochastic matrices.

Theorem 3.18: For a function $D_M : I_M \times I_M \to \mathbb{I}$, the following statements are equivalent:

1. D_M is a discrete copula.
2. There exists a bistochastic matrix $A = (a_{ij})_{i,j=1}^M$ such that

$$d_{ij}^{(M)} := D_M\left(\frac{i}{M}, \frac{j}{M}\right) = \frac{1}{M} \sum_{r=1}^{i} \sum_{s=1}^{j} a_{rs} \tag{3.8}$$

for $i, j \in \{0, 1, 2, ..., M\}$.

Proof: See the proof of Proposition 2 in [24]. □

Example 3.19:

1. The bistochastic matrix corresponding to the discrete copula $W_M^2 : I_M \times I_M \to \mathbb{I}$, $W_M^2\left(\frac{i}{M}, \frac{j}{M}\right) = \max\left\{\frac{i+j-M}{M}, 0\right\}$, where $i, j \in \{0, ..., M\}$, is given by

$$A(W_M^2) = \begin{pmatrix} 0 & 0 & \dots & 0 & 1 \\ 0 & 0 & \dots & 1 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & 0 & \dots & 0 & 0 \end{pmatrix}.$$

2. The bistochastic matrix corresponding to the discrete copula $\Pi_M^2 : I_M \times I_M \to \mathbb{I}$, $\Pi_M^2\left(\frac{i}{M}, \frac{j}{M}\right) = \frac{i}{M} \cdot \frac{j}{M} = \frac{ij}{M^2}$, where $i, j \in \{0, ..., M\}$, is given by

$$A(\Pi_M^2) = \begin{pmatrix} \frac{1}{M} & \frac{1}{M} & \cdots & \frac{1}{M} \\ \frac{1}{M} & \frac{1}{M} & \cdots & \frac{1}{M} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{1}{M} & \frac{1}{M} & \cdots & \frac{1}{M} \end{pmatrix}.$$

Hence, we have a one-to-one correspondence between the set of all copulas of order M and the set of all $(M \times M)$-matrices that are bistochastic.
It is obvious that (3.8) generalises (3.7) in the sense that irreducible discrete copulas are characterised by bistochastic matrices with values 0 and 1, namely $a_{ij} = \mathbb{1}_{\{i\}}(\sigma(j)) \in \{0, 1\}$.
As we have seen before, the product of irreducible discrete copulas is characterised by the composition of the corresponding permutations. We now extend this notion to all discrete copulas on I_M as proposed in [24].

Definition 3.20: Let $D_M^{(1)}$ and $D_M^{(2)}$ be discrete copulas both defined on $I_M \times I_M$ with values in $\mathbb{I}$ and let A_1 and A_2 be the corresponding $(M \times M)$-bistochastic matrices of $D_M^{(1)}$ and $D_M^{(2)}$, respectively. Then, the discrete copula D_M that corresponds to the bistochastic matrix $A = A_1 \cdot A_2$ is called the product of $D_M^{(1)}$ and $D_M^{(2)}$, and we write $D_M = D_M^{(1)} * D_M^{(2)}$.

Remark 3.21:

1. The product of bistochastic matrices is a bistochastic matrix again.

2. "$*$" is an associative binary operation, which is not commutative. Its neutral element is linked to the identity matrix corresponding to the discrete copula defined by $M_M^2\left(\frac{i}{M}, \frac{j}{M}\right) = \min\left\{\frac{i}{M}, \frac{j}{M}\right\}$, or equivalently $M_M^2 = M^2|_{I_M \times I_M}$.

Another useful result is that the product of discrete copulas commutes with convex sums.

Theorem 3.22: Let $C_M = \sum_{r=1}^{R} c_r C_M^{(r)}$ and $D_M = \sum_{s=1}^{S} d_s D_M^{(s)}$ be convex sums of the

discrete copulas $C_M^{(1)}, ..., C_M^{(R)}$ and $D_M^{(1)}, ..., D_M^{(S)}$ defined on $I_M \times I_M$; thus, we have $c_r \geq 0$, $d_s \geq 0$ and $\sum_{r=1}^{R} c_r = \sum_{s=1}^{S} d_s = 1$ for all $r \in \{1, ..., R\}$ and $s \in \{1, ..., S\}$. Then,

$$C_M * D_M = \sum_{r=1}^{R} \sum_{s=1}^{S} c_r d_s (C_M^{(r)} * D_M^{(s)}).$$

Proof: See the proof of Proposition 3 in [24]. □

Remark 3.23: The class $\mathcal{D}_M$ of all discrete copulas defined on $I_M \times I_M$ is a convex subclass of $\mathbb{I}^{I_M \times I_M}$. Furthermore, $\mathcal{D}_M$ is generated by the class $\mathcal{B}_M$ of irreducible discrete copulas defined on $I_M \times I_M$.

Theorem 3.24: $\mathcal{D}_M$ is the smallest convex set which contains $\mathcal{B}_M$, that is $\mathcal{D}_M$ is the convex closure of $\mathcal{B}_M$.

Proof: See the proof of Proposition 4 in [24]. □

Remark 3.25: According to [24], Theorem 3.24 is responsible for the fact that several dependence parameters of discrete bivariate distributions commuting with convex sums, such as Spearman's rho, can be computed in a simple way. In such cases, it suffices to know the values of these indices for irreducible discrete copulas only. Their computation is rather easy and we finally use the representation of discrete copulas as convex sums of irreducible discrete copulas.
In addition to this, since Theorems 3.22 and 3.24 hold, we can express the product of discrete copulas by the convex sum of products of irreducible discrete copulas and hence by the composition of permutations.

Remark 3.26: In this subsection, we focused on the presentation of concepts and results for discrete copulas in the bivariate case only. However, an exceedingly interesting question is if it would be possible to generalise the properties discussed before and extend them to the multivariate situation. In this context, the general concept of an n-stochastic matrix or a stochastic array $A = (a_{i_1 \ldots i_n})_{i_1, \ldots, i_n}$, for which $a_{i_1 \ldots i_n} \geq 0$ and $\sum_{i_l} a_{i_1 \ldots i_n} = 1$ for $i_1, ..., i_{l-1}, i_l, ..., i_n$ fixed and for all $l \in \{1, ..., n\}$, might be very useful.
A rigorous development of properties for discrete copulas in the multivariate case, which could be based upon the notion mentioned before, is an exciting challenge for additional future work. It would provide an advanced theoretical frame for our novel ECC method and is therefore of great interest. Hence, it is desirable to make further examinations in this direction.

3.4.7 Empirical copulas and Spearman's rho

Another relationship of copulas to our ECC approach is given in the context of Spearman's rho, a measure of concordance we have introduced in Subsection 2.1.6, where we

studied among other things population versions of Spearman's rho in terms of copulas. We recall that, in the bivariate case, for a vector (X^1, X^2) of continuous random variables with copula C, we have

$$\rho(X_1, X_2) = 12 \int_0^1 \int_0^1 C(u,v) du\, dv - 3 = 12 \int_0^1 \int_0^1 (C(u,v) - uv) du\, dv,$$

as a population version of Spearman's rho, compare Theorem 2.68, and a generalisation for the multivariate situation of $(X^1, ..., X^L)$ is given by

$$\rho_{n,C} = \frac{L+1}{2^L - (L+1)} \left\{ 2^L \int_0^1 ... \int_0^1 C^n(\mathbf{u}) d\mathbf{u} - 1 \right\}, \;\; L \geq 2, \;\; \mathbf{u} = (u_1, ..., u_L),$$

see Remark 2.75. The goal is now to study a sample version of Spearman's rho that can be applied to the ECC approach and is expressed in terms of copulas. For this purpose, we present a special copula, the so-called empirical copula, which will play an important role in what follows. We limit ourselves to the case of a sample of size M from a continuous multivariate distribution, that is, we assume that there are no ties.

Definition 3.27: Let $(x_m^l)_{m=1,...,M}^{l=1,...,L}$ denote a sample of size M from a multivariate continuous distribution. Then, the empirical copula E_M is defined by

$$E_M \left(\frac{i_1}{M}, ..., \frac{i_L}{M} \right) = \frac{\text{card}(\{(x^1, ..., x^L) \,\text{in the sample}\; | x^1 \leq x_{(i_1)}^1, ..., x^L \leq x_{(i_L)}^L\})}{M}$$

if $i_l \geq 1$ for $l \in \{1, ..., L\}$ and $E_M = 0$ if $i_l = 0$ for at least one $l \in \{1, ..., L\}$,
where $x_{(i_1)}^1, ..., x_{(i_L)}^L$ for $1 \leq i_1, ..., i_L \leq M$ are the marginal order statistics from the sample.

Remark 3.28: Definition 3.27 is a generalisation of the definition of an empirical copula in the bivariate case as proposed in [40]. Obviously, the empirical copula E_M is an irreducible discrete copula by definition because $\text{Ran}(E_M) = I_M$ and $\text{Dom}(E_M) = \underbrace{I_M \times \cdots \times I_M}_{L \,\text{times}}$.

Remark 3.29: An equivalent definition of empirical copulas is the following one which can be found in Section 4 in [48]. For a sample $(x_m^l)_{m=1,...,M}^{l=1,...,L}$ of size M from a multivariate continuous distribution, we set $U_{m,M}^l := \frac{1}{M} \text{rk}(x_m^l)$, where $\text{rk}(x_m^l)$ denotes the rank of x_m^l in $x_1^l, ..., x_M^l$. Then, we define the empirical copula by

$$E_M \left(\frac{i_1}{M}, ..., \frac{i_L}{M} \right) = \frac{1}{M} \sum_{m=1}^{M} \prod_{l=1}^{L} \mathbb{1}_{\{U_{m,M}^l \leq i_l\}},$$

where $0 \leq i_1, ..., i_L \leq M$.

For the bivariate case, we obtain the following sample version of Spearman's rho.

Theorem 3.30: Let E_M be the empirical copula for the sample $(x_m^l)_{m=1,...,M}^{l=1,2}$. Then, a sample version of Spearman's rho is given by

$$r_{\text{bi}} := \rho_{\text{sample, bi}} = \frac{12}{M^2-1} \sum_{i=1}^{M} \sum_{j=1}^{M} \left(E_M \left(\frac{i}{M}, \frac{j}{M} \right) - \frac{i}{M} \cdot \frac{j}{M} \right).$$

Proof: See [40]. □

Remark 3.31: According to [48], another sample version of Spearman's rho in terms of the considerations in Remark 3.29 is given by

$$\tilde{r}_{\text{bi}} := 1 - \frac{6M}{M^2-1} \sum_{m=1}^{M} (U_{m,M}^1 - U_{m,M}^2)^2$$

if there are no ties in the sample.

Finally, we consider the multivariate situation.

Remark 3.32: For the multivariate case, we obtain

$$\begin{aligned} r_{\text{mult}} := \rho_{\text{sample, mult}} &= \frac{L+1}{2^L-(L+1)} \left\{ 2^L - \underbrace{\int_0^1 ... \int_0^1}_{L \text{ times}} E_M(\mathbf{u}) d\mathbf{u} - 1 \right\} \\ &= \frac{L+1}{2^L-(L+1)} \left\{ \frac{2^L}{M} \sum_{m=1}^{M} \prod_{l=1}^{L} (1 - U_{m,M}^l) - 1 \right\}, \end{aligned}$$

where $\mathbf{u} = (\frac{i_1}{M}, ..., \frac{i_L}{M}) \in \underbrace{I_M \times \cdots \times I_M}_{L \text{ times}}$, as a sample version of Spearman's rho, compare [48].
We note that, for $L = 2$, r_{mult} differs slightly from the expression for $\tilde{r}_{\text{bi}}$ in Remark 3.31. However, according to [48], it can be verified that $r_{\text{mult}} \leq \tilde{r}_{\text{bi}}$ for $M \in \mathbb{N}$ and $\lim_{M \to \infty} \sqrt{M}(r_{\text{mult}} - \tilde{r}_{\text{bi}}) = 0$ almost surely. Hence, r_{mult} and $\tilde{r}_{\text{bi}}$ have the same asymptotic distribution in the bivariate case.

We recall that only continuous multivariate distributions were considered. For a sample version of Spearman's rho in the bivariate case which takes possible ties into account, we refer to Section 5 in [41].

As we have seen, Spearman's rho can be expressed in terms of empirical copulas, which are discrete copulas. Since Spearman's rho only depends on the empirical copula and the

ranks, respectively, and the ensemble size, but not on the observation or forecast values themselves, and ECC is just designed in such a manner that the ECC ensemble inherits the multivariate rank structure from the raw ensemble, Spearman's rho obviously has the same value for both the raw ensemble and the ECC postprocessed ensemble. Hence, we have another justification to consider ECC as a copula approach.

Chapter 4

Assessment methods for uni- and multivariate probabilistic forecasts

In this chapter, we discuss methods of assessing the predictive performance of probabilistic forecasts for both univariate and multivariate, vector-valued quantities as suggested by Gneiting et al. [16]. Such techniques are presented for both a discrete forecast ensemble and a continuous predictive distribution in form of a density forecast. However, the distinction between ensemble forecasts and density forecasts is artificial because, on the one hand, we can sample from a predictive density to get a forecast ensemble, and on the other hand, we might replace a forecast ensemble by a density estimate.
According to [14], the goal of probabilistic forecasting is that the predictive distributions show maximal sharpness, subject to calibration.
Moreover, so-called proper scoring rules provide summary measures that are employed to evaluate probabilistic forecasts, see for example [15].
In what follows, we outline the methods proposed in [16] for the assessment of calibration and sharpness for probabilistic forecasts of multivariate, L-dimensional quantities, as well as some proper scoring rules for this setting, and establish relationships to the univariate case of $L = 1$.
In so doing, we often focus on the case of small dimension, say $L = 2$, and we remark that whenever the dimension of the forecast vector is large, attention typically concentrates on low-dimensional functionals. In this case, the approaches described here can still be applied.
We limit ourselves to the case of quantities with values in $\mathbb{R}$ or $\mathbb{R}^L$, respectively. Assessment methods for angular variables like wind direction in the univariate case can be found in Section 2.d in [1].
Some of the methods below are employed for the evaluation of the ECC technique, which is discussed in Chapter 5.

Form of the VRH	Interpretation
uniform	calibrated ensemble
U-shaped	underdispersed ensembles, ensemble ranges are too narrow
hump- or inverse U-shaped	overdispersed ensemble
skewed	central tendencies are biased

Table 4.1: Interpretation of the form of a VRH

4.1 Assessing calibration

Calibration relates to the statistical consistency between the probabilistic forecasts and the observations. It is a joint property of both the predictive distributions and the vector-valued events that occur.
For a univariate predictand, calibration is checked by the so-called Talagrand diagram or verification rank histogram (VRH). For an ensemble forecast of size M, the ordered ensemble values partition the real line into $M + 1$ bins. We find the verification rank, which is the bin occupied by the verifying observation, namely a number between 1 and $M + 1$, repeat over a sizable amount of individual forecast cases and aggregate ranks. The Talagrand histogram or VRH is the histogram of these ranks.
In a calibrated ensemble, the verifying observation is equally likely to fall into any of the bins. Therefore, we evaluate calibration by checking for deviations from uniformity in the VRH.
To quantify this, we look at the discrepancy or reliability index

$$\Delta = \sum_{m=1}^{M+1} \left| f_m - \frac{1}{M+1} \right|, \tag{4.1}$$

where f_m denotes the observed relative frequency of rank m.
From a graphical point of view, the form of the VRH provides information about calibration, which is summarised in Table 4.1.

An analogue of the VRH for ensemble forecasts of a multivariate quantity in $\mathbb{R}^L$ is the multivariate rank histogram (MRH), which is introduced in the following.
Let M be the ensemble size and $\mathbf{x} = (x_1, ..., x_L) \in \mathbb{R}^L$ and $\mathbf{y} = (y_1, ..., y_L) \in \mathbb{R}^L$ be two vectors. We write $\mathbf{x} \leq \mathbf{y}$ if and only if $x_l \leq y_l$ for all $l \in \{1, ..., L\}$.

Let $\{\mathbf{x}_m \in \mathbb{R}^L | m = 1, ..., M\}$ be a given ensemble forecast and $\mathbf{x}_0 \in \mathbb{R}^L$ the corresponding verifying observation. Then, the MRH is constructed according to the following scheme:

1. Standardisation (not necessary, but often useful):
   A principal component transform to the pooled set $\{\mathbf{x}_m | m = 0, ..., M\}$ is used in order to get a standardised observation $\mathbf{x}_0^*$, as well as standardised ensemble member forecasts $\mathbf{x}_m^*$, where $m = 1, ..., M$.

2. Assignment of pre-ranks:
   For $m = 0, ..., M$, we find the pre-ranks $\rho_m = \sum_{n=0}^{M} \mathbb{1}_{\{\mathbf{x}_n^* \leq \mathbf{x}_m^*\}}$ of $\mathbf{x}_m^*$ among the union of the observation and the ensemble member forecasts, which might be standardised. We note that ρ_m is an integer between 1 and $M+1$ for each m.

3. Determination of the multivariate rank r:
   r is the rank of the observation pre-rank, with ties resolved at random. Let $s^<$ and $s^=$ be defined by $s^< = \sum_{m=0}^{M} \mathbb{1}_{\{\rho_m < \rho_0\}}$ and $s^= = \sum_{m=0}^{M} \mathbb{1}_{\{\rho_m = \rho_0\}}$. Then, r, which is an integer between 1 and $M+1$, is selected from a discrete uniform distribution on $\{s^< + 1, ..., s^< + s^=\}$.

4. Repetition over individual forecast cases

5. Plot of the resulting rank histogram

The multivariate rank is uniform if the ensemble members and the verifying observation are exchangeable and reduces to the previously discussed univariate verification rank in the case of $L = 1$.
From a graphical point of view, the MRH is a plot of the empirical frequency of the multivariate ranks. The interpretation of the form of the MRH is the same as for the VRH in dimension $L = 1$.

There exists another tool to check whether multivariate ensemble forecasts are calibrated which is similar to the MRH. It is called the minimum spanning tree rank histogram (MSTRH), is proposed in [56] and obtained by replacing the steps 2 and 3 in the construction of the MRH by 2* and 3* as follows:

2.* Computation of minimum spanning trees:
   We find the minimum spanning tree (MST) of $\{\mathbf{x}_n^* | n \in \{0, ..., M\} \setminus \{m\}\}$ and its length $l_m > 0$ for $m = 0, ..., M$.
   If we have a set of M points in $\mathbb{R}^L$, a spanning tree is a collection of $M-1$ edges such that all points are employed and that with the smallest length is the MST.

3.* Determination of the MST ranks r:
   r is the rank of l_0 within the pooled sample of MST lengths, with any ties resolved at random.
   If $s^<$ and $s^=$ are defined by $s^< = \sum_{m=0}^{M} \mathbb{1}_{\{l_m < l_0\}}$ and $s^= = \sum_{m=0}^{M} \mathbb{1}_{\{l_m = l_0\}}$, then r, which is a number between 1 and $M+1$, is selected from a uniform distribution on $\{s^< + 1, ..., s^< + s^=\}$.
   For a given ensemble forecast, the MST rank can be found by tallying the length of the MST, which links the M ensemble members within the combined set of the $M+1$ lengths of the ensemble-only MST to the M MSTs achieved by the substitution of the observations for each of the ensemble members.

In the case of exchangeability of the ensemble members and the observation, the lengths are exchangeable and the MST rank is uniform.
The empirical frequency of the MST ranks is plotted in a rank histogram, which can be used to check calibration. For an underdispersed or biased ensemble, there is an overpopulation of the lowest MST ranks and for an overdispersed ensemble, the highest ranks materialise too often.
The MSTRH cannot be constructed for dimension $L = 1$.

The two methods for the multivariate case presented above are closely related to tests for exchangeability. Below, we assume the invariance of the function
$F : \mathbb{R}^L \times \underbrace{\mathbb{R}^L \times \cdots \times \mathbb{R}^L}_{M \text{ times}} \to \mathbb{R}$ under permutations of its final M arguments.
For a given ensemble forecast $\mathbf{x}_1, .., \mathbf{x}_M \in \mathbb{R}^L$ and the realizing observation $\mathbf{x}_0 \in \mathbb{R}^L$, we define z_m via $z_m = F(\mathbf{x}_m; \mathbf{x}_{-m})$ for $m = 0, ..., M$, where $\mathbf{x}_{-m}$ is the set $\{\mathbf{x}_0, \mathbf{x}_1, ..., \mathbf{x}_M\} \setminus \{\mathbf{x}_m\}$. If there exists exchangeability between the observation and the ensemble members, the rank of z_0, which might be randomised and is calculated with the observation in the first argument, among $\{z_0, z_1, ..., z_M\}$ is uniform on $\{1, ..., M + 1\}$.
By determining a rank for each individual forecast, collecting them and checking the corresponding rank histogram for uniformity, we can judge exchangeability.
Special choices of the function F yield the techniques previously introduced as follows.

- If F is a coordinative projection for the first argument, our procedure leads to the univariate VRH.
- By using standardised values and setting $F(\mathbf{x}_m; \mathbf{x}_{-m}) = \sum_{n=0}^{M} \mathbb{1}_{\{\mathbf{x}_n \leq \mathbf{x}_m\}}$, we obtain the pre-rank as in the construction of the MRH.
- If the value of F is the length of the MST for the second set-valued argument, we get the MSTRH.

We now focus on calibration checks for density forecasts.
The MRH and the MSTRH cannot only be used to assess calibration for ensemble forecasts, but also for density forecasts if we generate an ensemble forecast by sampling from the predictive density.
Nevertheless, other techniques depending on the predictive density itself or a function thereof are available, and we point out some below.
For the univariate case, we can employ the probability integral transform (PIT), which is the value the predictive CDF attains at the observation. If the predictive distribution is continuous, and the observation is drawn from it, the PIT is uniformly distributed on $\mathbb{I} = [0, 1]$. By plotting the histogram of the PIT values, we can check for uniformity and consequently for calibration. The interpretation is the same as for the VRH.
For the multivariate case, we might look at PIT histograms for projections and scan for non-uniform directions, but this is not a genuinely multivariate approach. Amongst other methods, the Box density ordinate transform (BOT) is able to handle calibration

checks for multivariate density forecasts. Let p denote the predictive density for a future quantity and $\mathbf{x}_0$ be the materialisation. Moreover, let $\mathbf{X}$ be a random vector with density p. Then, a version of the BOT is given by $u = 1 - \mathbb{P}[p(\mathbf{X}) \leq p(\mathbf{x}_0)]$.
In this connection, we consider the following example. If $\mathbf{X} \sim \mathcal{N}_L(\boldsymbol{\mu}, \Sigma)$, then $u = 1 - \chi^2_L((\mathbf{x}_0 - \boldsymbol{\mu})^t \Sigma^{-1}(\mathbf{x}_0 - \boldsymbol{\mu}))$ equals one minus the CDF of a χ^2-distribution with L degrees of freedom, when evaluated at the standardised observation.
If $\mathbf{x}_0$ has density p and $p(\mathbf{X})$ is continuously distributed, then u is standard uniform. In addition to this, a randomised version of the BOT stays uniform if $p(\mathbf{X})$ has discrete components. However, its power is diminished.

Under the assumption of exchangeability between the ensemble members and the verifying observation, we expect the empirical distribution of the ensemble values and that of the realisations to be statistically consistent if compositing over forecast cases in a steady state. These two empirical distributions can be compared by plotting a so-called marginal calibration diagram. In so doing, dispersion errors, as well as forecast biases, can be unmasked.

4.2 Assessing sharpness

Sharpness refers to the concentration of the predictive distribution. The sharper the probabilistic forecast, the less uncertainty, and the sharper, the better, subject to calibration.
For an ensemble forecast for a univariate quantity, this fact can be expressed as follows: The smaller the ensemble spread, the sharper, and the sharper, the better, subject to proper coverage. Ensemble spread is mostly quantified by the ensemble range or the ensemble standard deviation.
In the multivariate case, a generalisation of the univariate standard deviation as a measure of spread is given by the determinant sharpness (DS), namely

$$DS = (\det(\Sigma))^{1/(2L)},$$

where $\Sigma \in \mathbb{R}^{L \times L}$ is the covariance matrix of an ensemble or density forecast for a variable in $\mathbb{R}^L$. The DS measure can be applied to both ensembles of size $M > L$ and predictive densities with finite second moments.
However, according to a comment of Jolliffe in [16], the usage of the DS for predictive distributions with nearly singular covariance matrix Σ may pose problems. A possible resolution is to reduce the dimension to nearly orthogonal components.
Besides the DS, there exist other tools to quantify sharpness, like for example the volume of the convex hull or bounding box of an ensemble forecast, see [58] or [22]. Furthermore, Angulo and Ruiz-Medina suggested to quantify sharpness via entropy criteria, see the comment in [16].

4.3 Proper scoring rules as performance measures

A scoring rule is a summary measure for the assessment of probabilistic forecasts in the sense that a numerical score based on the predictive distribution and on the occurring event or value is assigned.
As a convention, we consider scoring rules to be negatively orientated penalties $s(P, \mathbf{x})$ which should be minimised by the forecaster, where P is the predictive distribution suggested by the forecaster and $\mathbf{x} \in \mathbb{R}^L$ the occurring value.
Let $s(P, Q)$ denote the expected value of $s(P, \mathbf{X})$ for the case of $\mathbf{X} \sim Q$.
An exceedingly important property of scoring rules is propriety, which is defined as follows. Let the predictive distribution Q be the forecaster's best judgement. Then, a scoring rule is called proper if

$$s(Q, Q) \leq s(P, Q) \tag{4.2}$$

for all P and Q. A scoring rule is called strictly proper if (4.2) holds with equality if and only if $P = Q$.
According to the definition of propriety, a proper scoring rule causes that the forecaster has no incentive to predict any $P \neq Q$ and encourages him to quote his true belief $P = Q$. Propriety ensures that a scoring rule takes account of calibration and sharpness simultaneously.
A well-known and widely used proper scoring rule for probabilistic forecasts of a univariate scalar quantity is the continuous ranked probability score (CRPS) defined as

$$\mathrm{crps}(P, x) = \underbrace{\int_{-\infty}^{\infty} (F(y) - \mathbb{1}_{\{y \geq x\}})^2 dy}_{\text{standard form}} = \underbrace{\mathbb{E}_P\left[|X - x|\right] - \frac{1}{2}\mathbb{E}_P\left[|X - X'|\right]}_{\text{kernel score representation}},$$

where F denotes the CDF associated with the predictive distribution P, and X and X' are independent random variables that have distribution P and finite first moment.
A direct generalisation of the CRPS to evaluate probabilistic forecasts of a multivariate quantity is given by the energy score (ES), which is a proper scoring rule defined by

$$\mathrm{es}(P, \mathbf{x}) = \mathbb{E}_P\left[||\mathbf{X} - \mathbf{x}||\right] - \frac{1}{2}\mathbb{E}_P\left[||\mathbf{X} - \mathbf{X}'||\right], \tag{4.3}$$

where $||.||$ denotes the Euclidean norm, and $\mathbf{X}$ and $\mathbf{X}'$ are independent random vectors with distribution P and $\mathbb{E}_P[||\mathbf{X}||] < \infty$.
If $P = P_{\mathrm{ens}}$ is an ensemble forecast comprising M members, the ES can be explicitly determined as follows. Since the predictive distribution P_{ens} places point mass $\frac{1}{M}$ on the ensemble members $\mathbf{x}_1, ..., \mathbf{x}_M \in \mathbb{R}^L$, inserting in (4.3) yields

$$\mathrm{es}(P_{\mathrm{ens}}, \mathbf{x}) = \frac{1}{M}\sum_{m=1}^{M} ||\mathbf{x}_m - \mathbf{x}|| - \frac{1}{2M^2}\sum_{n=1}^{M}\sum_{m=1}^{M} ||\mathbf{x}_n - \mathbf{x}_m||. \tag{4.4}$$

For a deterministic forecast, namely if $P = \delta_{\boldsymbol{\mu}}$ is the point measure in $\boldsymbol{\mu} \in \mathbb{R}^L$, the ES reduces to $\mathrm{es}(\delta_{\boldsymbol{\mu}}, \mathbf{x}) = ||\boldsymbol{\mu} - \mathbf{x}||$.

Thus, the ES can be applied to both ensemble forecasts and density forecasts and is a direct way of comparing them by using a single metric.
If the expectations in (4.3) cannot be derived in closed forms, we replace (4.3) by a computationally efficient Monte Carlo approximation.
Moreover, if the components are incomparable in magnitude, it might be useful or even necessary to employ a standardisation of both observations and ensemble forecasts, depending on the application we consider. This is due to the fact that the Euclidean variant of the ES does not make any distinction between the components of the forecast vector.

Besides the ES, which also holds for ensemble forecasts, there are several other proper scoring rules for assessing the forecast performance if the predictive distribution P of a multivariate quantity has a density function p. The most important ones are

- the logarithmic score $\text{logs}(P, \mathbf{x}) = -\log p(\mathbf{x})$,
- the quadratic score $\text{qs}(P, \mathbf{x}) = -2p(\mathbf{x}) + ||p||^2$ and
- the spherical score $\text{sphs}(P, \mathbf{x}) = -\frac{p(\mathbf{x})}{||p||}$,

see for example [33]. In this context, $||p||^2 = \int p(\mathbf{y})^2 \, \mathrm{d}\mathbf{y}$.
Those three scoring rules can be applied to density forecasts exclusively.
Of course, they are applicable to univariate quantities, too.

Assessment methods for probabilistic forecasting

Characteristic	Univariate Quantity	Multivariate Quantity
calibration (statistical consistency between the probabilistic forecasts and the observations)	ensemble forecasts: • Talagrand histogram / verification rank histogram density forecasts: • probability integral transform	ensemble forecasts: • multivariate rank histogram • minimum spanning tree rank histogram density forecasts: • Box density ordinate transform
sharpness (measure of spread)	• ensemble standard deviation • ensemble range	• determinant sharpness
proper scoring rules (ranking and comparison of competing forecast methods)	ensemble forecasts: • continuous ranked probability score density forecasts: • continuous ranked probability score • logarithmic score • quadratic score • spherical score	ensemble forecasts: • energy score density forecasts: • energy score • logarithmic score • quadratic score • spherical score

Table 4.2: Assessment methods

4.4 Summary

The tools presented above possess some appealing properties like simplicity, generality and interpretability. They can be applied in both parametric and non-parametric settings and do not require models to be nested.
Table 4.2 shows a summary of the most important assessment techniques for probabilistic forecasts named in this chapter.

Chapter 5

Practical applications and assessment of the ECC technique

After having studied of the methodological background and the theoretical development of the ECC approach in the precedent chapter, we now want to test and illustrate our novel technique in some practical applications with real data. Moreover, its predictive performance shall be evaluated and compared to that of other methods.

To be precise, we apply ECC to the 50-member ensemble of the European Centre for Medium-Range Weather Forecasts (ECMWF). The ECMWF is an international and intergovernmental organisation located in Reading, England, and established in 1975, which is supported by 32 states and provides operational medium- and extended-range weather forecasts and a state-of-the-art super-computing facility for scientific research. Furthermore, the ECMWF develops several numerical methods for medium-range weather forecasting, does scientific and technical research in order to improve them and collects and stores appropriate meteorological data. In addition, the organisation collaborates with satellite agencies and the European Commission. These and further facts about the ECMWF can be found on the website `http://www.ecmwf.int`.

We start with a description of the ECMWF data set and the preliminaries like data cleaning or data transformation, for instance, that have to be done before accomplishing the case study, in the first section of this chapter.
The procedural method and some results of the applications of the ECC technique are the subjects being dealt with in the second section.
This chapter is completed by an evaluation of ECC, as far as calibration and sharpness of the prediction are concerned, by employing some of the assessment tools presented in Chapter 4 to our ECC forecasts based on the ECMWF data set and comparing the predictive performance with that of the unprocessed raw ensemble and other techniques in the third and final section.

	Berlin (Tegel)	Hamburg (Fuhlsbüttel)	Frankfurt (Airport)
identification number	10382	10147	10637
longitude in degrees'minutes	13'19 east	10'00 east	8'35 east
latitude in degrees'minutes	52'34 north	53'37 north	50'02 north
elevation in metres	37	16	113

Table 5.1: Geographical facts about the observation sites Berlin, Hamburg and Frankfurt

5.1 Data set and preliminaries

5.1.1 Observation sites and data set

We start with a description of the ECMWF data set we work with in what follows.

The data consist of observations and ensemble forecasts for the three locations Berlin, Hamburg and Frankfurt. Before pointing out the details of the composition of our data set, we comment briefly on the three sites we focus on and collect some interesting geographical and climatological facts about them. Those are mainly taken from `http://en.wikipedia.org` or are given in the data set.

Table 5.1 provides geographical information about our three observation sites. Berlin, Hamburg and Frankfurt are German cities having 3,400,000, 1,800,000 and 672,000 inhabitants, respectively, and being located in the north temperate zone and the middle latitudes, respectively, having an oceanic and semi-humid climate, compare the climatic map on pages 220 and 221 in [59], which means that the climate is strongly influenced by the sea, there are relatively small variations in temperature, comparably high wind speeds, moderate temperatures and between six and nine humid months at an average, where precipitation is greater than evaporation.

In contrast to Hamburg and Frankfurt, Berlin is situated near the transition zone from the oceanic to the continental climate, which is characterised by great variations of temperature in form of hot summers and cold winters, little precipitation and little influence of the seas. In Berlin, there are warm summers and cold winters, whereas spring and autumn are chilly to mild, in general. It is remarkable that in Berlin's densely built-up area, temperatures can be much higher than in the surrounding areas because the heat is stored by the city's buildings. Moreover, Berlin has a moderate rainfall throughout the year and only light snowfall, mainly from December through March. In addition, it can be observed that wind in Berlin mostly comes from northwest and southwest, often with rather high speed and transporting maritime air. However, wind in Berlin often comes from southeast and east as well, which is characteristic for high-pressure fronts of continental air masses.

Berlin

Month	Mean Temperature (°C)		Mean Total Rainfall (mm)	Mean Number of Rain Days
	Daily Minimum	Daily Maximum		
Jan	-1.9	2.9	42.3	10.0
Feb	-1.5	4.2	33.3	8.0
Mar	1.3	8.5	40.5	9.1
Apr	4.2	13.2	37.1	7.8
May	9.0	18.9	53.8	8.9
Jun	12.3	21.6	68.7	9.8
Jul	14.3	23.7	55.5	8.4
Aug	14.1	23.6	58.2	7.9
Sep	10.6	18.8	45.1	7.8
Oct	6.4	13.4	37.3	7.6
Nov	2.2	7.1	43.6	9.6
Dec	-0.4	4.4	55.3	11.4

Table 5.2: Climate data for Berlin

Hamburg in Northern Germany has a mild climate, which is due to the proximity of the city to coastal areas arousing oceanic influences. Furthermore, Hamburg's climate is humid throughout the year, and snowfall is rare. Especially in the winter, wind speeds can be very high.

Frankfurt is one of the warmest cities in Germany with a very mild climate including cool winters and warm summers. Wind in Frankfurt mostly comes from west.

Tables 5.2, 5.3 and 5.4 summarise some facts about the climate in Berlin, Hamburg and Frankfurt, respectively. They show climatological information based on monthly averages for the 30-year period from 1971 to 2000 and can be found on the World Weather Information Service (WWIS) website `http://www.worldweather.org`, which is sponsored by the World Meteorological Organization (WMO). We note that the mean number of rain days corresponds to the mean number of days with at least one millimetre of rain.

We now go into the structure of our raw data set provided by the ECMWF in detail. Note that we are given both observations and forecasts for the three locations Berlin, Hamburg and Frankfurt. However, they partly differ in their representation and form, which is specified in the following.

Observations are available for Berlin, Hamburg and Frankfurt, respectively,

- from 1 January 1980 to 25 October 2010, every three hours, namely at 0000, 0300, 0600, 0900, 1200, 1500, 1800 and 2100 Coordinated Universal Time (UTC) and

Hamburg

Month	Mean Temperature (°C)		Mean Total Rainfall (mm)	Mean Number of Rain Days
	Daily Minimum	Daily Maximum		
Jan	-1.4	3.5	64.4	12.1
Feb	-1.2	4.4	42.4	9.2
Mar	1.1	8.0	62.9	11.3
Apr	3.3	12.3	45.6	8.9
May	7.4	17.5	53.7	9.6
Jun	10.5	19.9	76.9	11.3
Jul	12.7	22.1	74.7	11.4
Aug	12.5	22.2	73.0	10.2
Sep	9.6	17.9	68.4	10.8
Oct	6.0	13.0	63.6	10.5
Nov	2.4	7.5	69.4	11.7
Dec	0.0	4.6	77.7	12.4

Table 5.3: Climate data for Hamburg

Frankfurt

Month	Mean Temperature (°C)		Mean Total Rainfall (mm)	Mean Number of Rain Days
	Daily Minimum	Daily Maximum		
Jan	-1.3	4.0	42.5	9.8
Feb	-1.2	5.6	37.1	7.1
Mar	1.9	10.4	47.6	9.3
Apr	4.1	14.5	42.8	8.5
May	8.4	19.5	60.2	9.8
Jun	11.7	22.3	60.6	10.3
Jul	13.7	24.8	64.9	9.3
Aug	13.4	24.8	52.9	7.8
Sep	10.1	20.1	50.0	8.1
Oct	6.0	14.0	54.6	9.3
Nov	2.1	7.7	51.8	9.7
Dec	0.0	5.0	55.7	9.9

Table 5.4: Climate data for Frankfurt

- for the following weather variables:
  - wind direction in degrees,
  - wind speed in metres per second,
  - temperature T_K in Kelvin, where $T_K = T_C + 273.16$ if T_C denotes the temperature in degrees Celsius,
  - precipitation in millimetres over the last six hours and
  - sea level pressure in Pascal.

As far as the forecasts at Berlin, Hamburg and Frankfurt, respectively, are concerned, we have to remark that ECMWF ensemble forecast models run on a regular grid consisting of longitude and latitude grid points, which, in general, do not correspond exactly to the observation sites. Hence, bilinear interpolation is employed to obtain the ensemble forecasts at our three observation sites. This is a method that can be used in order to determine intermediate values within a two-dimensional regular grid and which is common in the meteorological community and also in image processing.

Ensemble forecasts are available for Berlin, Hamburg and Frankfurt, respectively,

- in the period from 1 February 2010 to 21 October 2010, that is, for 263 initialisation days, using initialisations at 0000 UTC on each day,
- made by 50 ensemble members,
- for 49 prediction horizons in steps of three hours and
- for the following weather variables:
  - wind vectors, namely the u- and v-components, in metres per second,
  - wind speed in metres per second,
  - temperature in Kelvin,
  - precipitation in metres, cumulative over all prediction horizons, and
  - sea level pressure in Pascal.

5.1.2 Data cleaning

Plots of the temporal development of the observations suggest that it might be useful and convenient to clean the raw observation data in the following way:

- In the original data, missing values are given the value -99. We change all -99-values to "Not available (NA)"- entries.
- We assign "NA" to all observation values that are greater or less, respectively, than the maximum or minimum weather records, respectively, of the corresponding

Weather Quantity	Record(s)
wind speed	maximum of 103.3 m/s
temperature	maximum of 321.15 K
	minimum of 215.05 K
precipitation (over 12 hours, tropical cyclone)	maximum of 1144 mm
sea level pressure	maximum of 108560 Pa
	minimum of 85000 Pa

Table 5.5: Weather records

weather quantity. In this connection, we employ the weather records presented in Table 5.5 being based upon information from `http://en.wikipedia.org`.

We note that, if the wind speed is bad, we do not trust the wind direction either. Therefore, we assign "NA" to the observation value of both wind speed and wind direction in this case.

5.1.3 Data transformation

As mentioned before, the data form of the observations and the forecasts does not always coincide. To be precise, transformations, as far as wind data and precipitation are concerned, are obviously necessary in order to obtain a common data frame. In what follows, we describe the procedure of transformation we employ in the cases of the weather variables named above.

Wind

We recall that the observations are available for wind direction and wind speed, while our forecasts are given in the form of wind vectors consisting of u- and v-components. The relationship between the different variables is explained in the following.

The wind direction θ in degrees, where $\theta \in [0, 360)$, describes the direction **from** which the wind is blowing. It increases clockwise from north, when viewed from above, and hence, a wind direction of 0, 90, 180 and 270 degrees, respectively, describes a northerly, easterly, southerly and westerly wind, respectively.

Wind vectors consist of two orthogonal velocity components called u and v, where u is the zonal velocity, that is, the component of the horizontal wind **towards** east, and v the meridional velocity, that is, the component of the horizontal wind **towards** north. Hence, positive (negative) values of u imply a horizontal wind component **towards** east (west), and positive (negative) values of v imply a horizontal wind component **towards** north (south) and the other way round.

Wind data can be expressed either by wind vectors with u- and v-components, as in

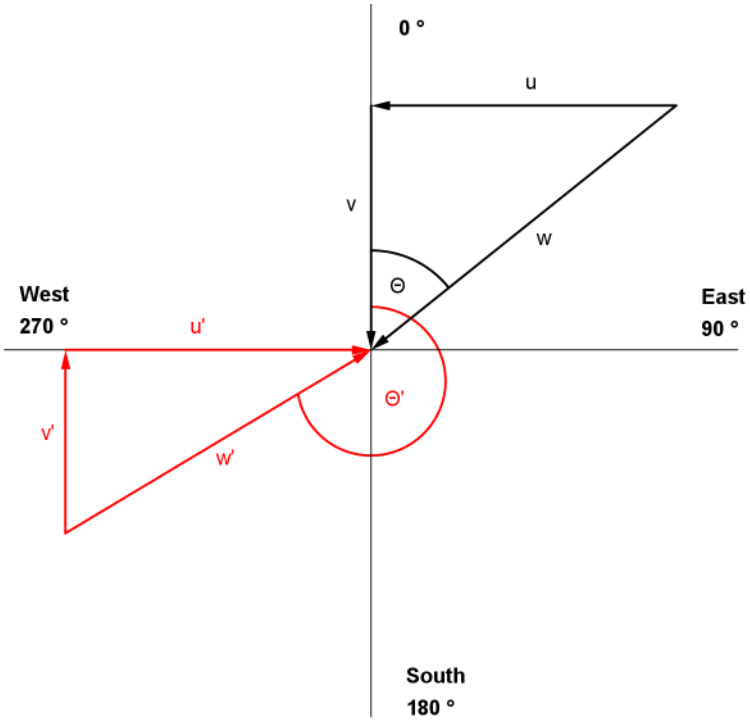

Figure 5.1: Wind: relationship between the variables. Exemplary illustrations for $\theta \in (0, 90)$ (black) and $\theta' \in (180, 270)$ (red)

the case of our ensemble forecasts, or by wind direction θ and wind speed w being a non-negative weather quantity, as in the case of our observations.
Figure 5.1 illustrates the relationship between the different variables. Elementary geometrical considerations including vector addition, addition theorems for sine and cosine and a case-by-case analysis among other things lead to the fact that the u- and v-component can be expressed by

$$u = -w \cdot \sin\left(\frac{\pi}{180} \cdot \theta\right) \text{ and } v = -w \cdot \cos\left(\frac{\pi}{180} \cdot \theta\right), \tag{5.1}$$

respectively, where $\theta \in [0, 360)$ denotes the wind direction in degrees and w the non-negative wind speed. Further information about wind vectors can be found online at `http://mst.nerc.ac.uk/wind_vect_convs.html`.
Hence, we use the formulas (5.1) to transform the observation data for wind given by wind direction and wind speed into observation wind vectors consisting of u- and v-components, while adopting the forecasts which are already given by u- and v-components. We note that both the u- and the v-component can be assumed to be normally distributed — a fact which is employed in the applications later on.
Finally, we mention that the wind speed w can be expressed in terms of u- and v-components via $w = \sqrt{u^2 + v^2}$. According to Figure 5.1 and Pythagoras' theorem, this

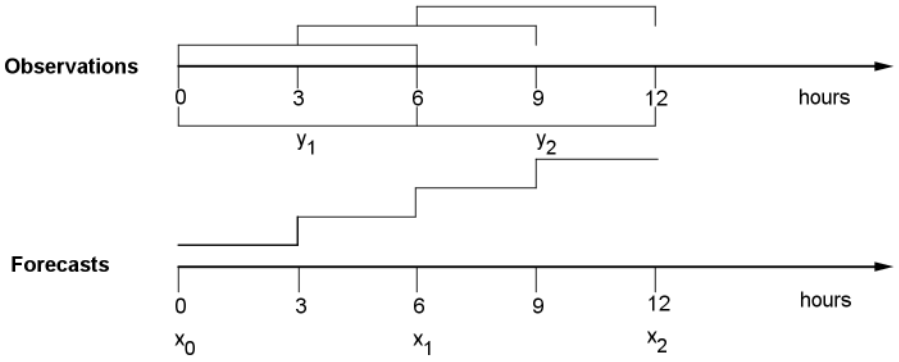

Figure 5.2: Precipitation transformation scheme

is obvious.

Precipitation

In the case of precipitation, we proceed as described in the following.
Since the precipitation observations are given over the last six hours, and the forecasts are cumulative over the prediction horizons, the following transformation is self-evident and makes sense.
If the notation follows Figure 5.2, we have the correspondences

$$\begin{aligned} x_1 - x_0 &\leftrightarrow y_1, \\ x_2 - x_1 &\leftrightarrow y_2, \end{aligned}$$

and so on. In this sense, we implement our precipitation transformations such that we always consider the precipitation over the last six hours for both forecasts and observations.
Moreover, we choose millimetres as the unit to be employed and thus have to transform the forecasts into this order of magnitude as well.
Very occasionally, we obtain transformed precipitation values that are slightly less than zero, which does not make sense, of course, and is likely due to rounding errors within the data set. In those cases, the negative values are set to zero.

5.1.4 Overview of dependence structures

After having arranged the data cleaning and transformation as proposed before, we are ready to apply the ECC method to our data set and see how it performs. Before presenting the results in the next two sections of this chapter, we occupy ourselves with the correlation structure within our data set first.
To get both an impression and an overview of the dependence structure between the

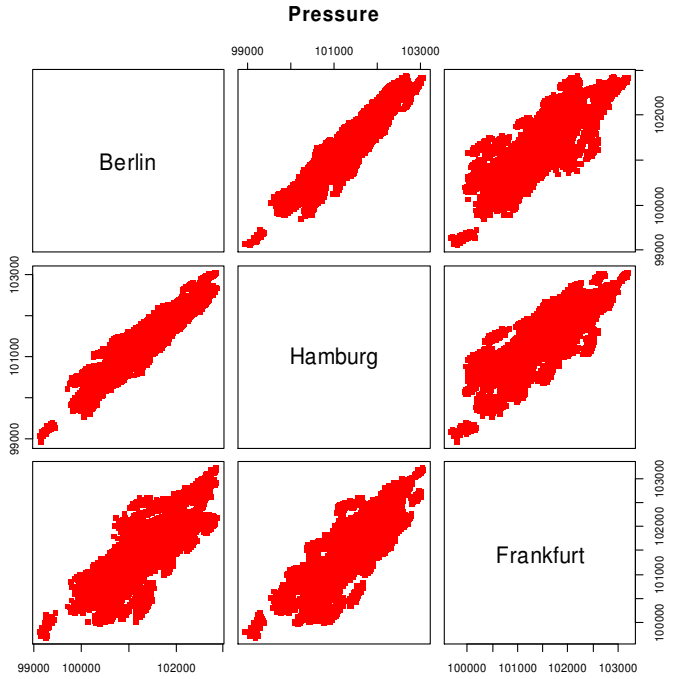

Figure 5.3: Local correlations of the ensemble forecasts of pressure

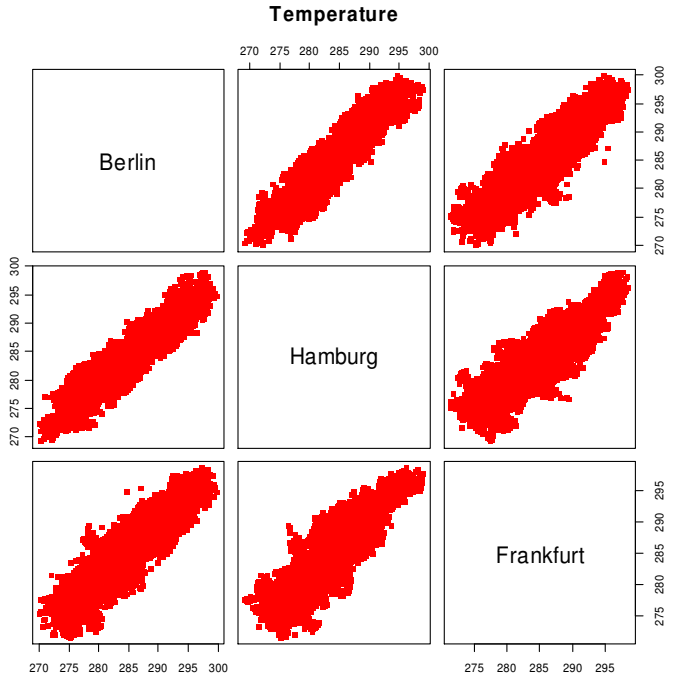

Figure 5.4: Local correlations of the ensemble forecasts of temperature

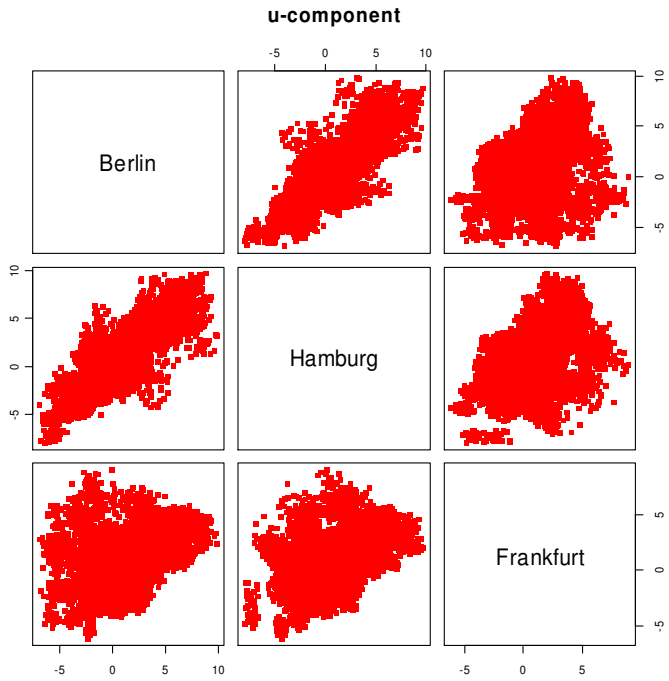

Figure 5.5: Local correlations of the ensemble forecasts of the u-component

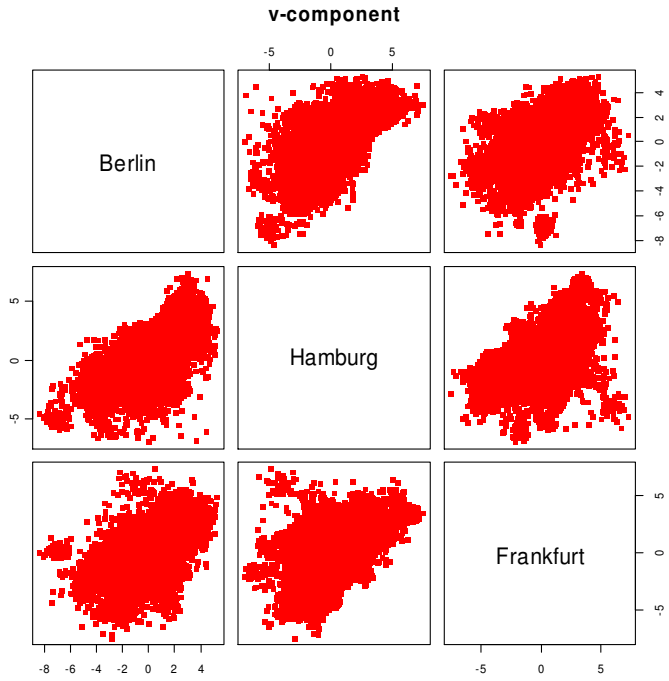

Figure 5.6: Local correlations of the ensemble forecasts of the v-component

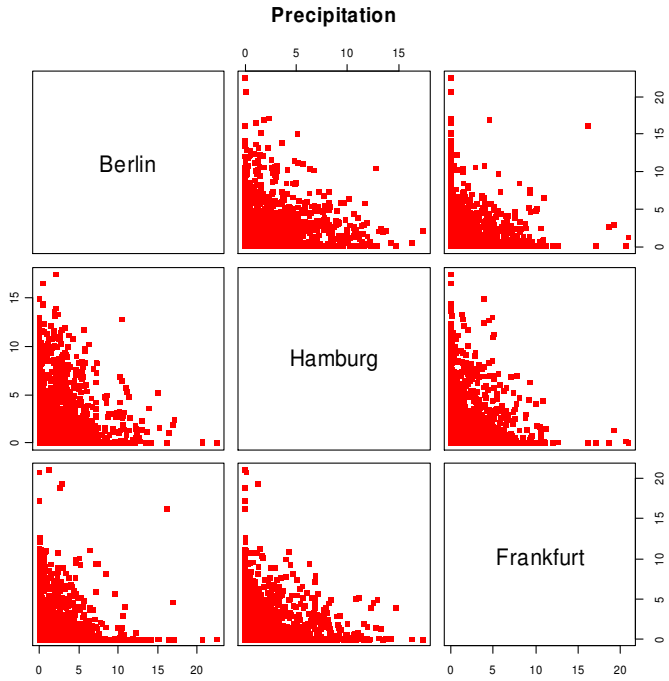

Figure 5.7: Local correlations of the ensemble forecasts of precipitation

weather quantities and the locations in our case study, we consider the scatterplot matrices in Figures 5.3 to 5.10. Those are based upon 24 hours-ahead forecasts made by the 50 ensemble members valid for the period from 14 March 2010 to 22 October 2010 at 0000 UTC on each of the 223 days. Figures 5.3 to 5.7 show the local correlations of individual weather variables at our three observation sites Berlin, Hamburg and Frankfurt.
In addition, Figures 5.8, 5.9 and 5.10 illustrate the cross-variate dependence structures at each location individually.
In a nutshell, sea level pressure and temperature show the strongest local correlation, which is positive in these cases, among the individual weather quantities. For wind vectors and precipitation, there are also certain dependence patterns, but those are less pronounced than in the cases of pressure and temperature. As far as cross-variate correlations at a fixed observation site are concerned, we can observe a negative correlation between temperature and sea level pressure, for instance.
As a rule, the dependence structures in the case of forecasts with a prediction horizon of 48 hours look very similar to those for 24 hours-ahead forecasts.

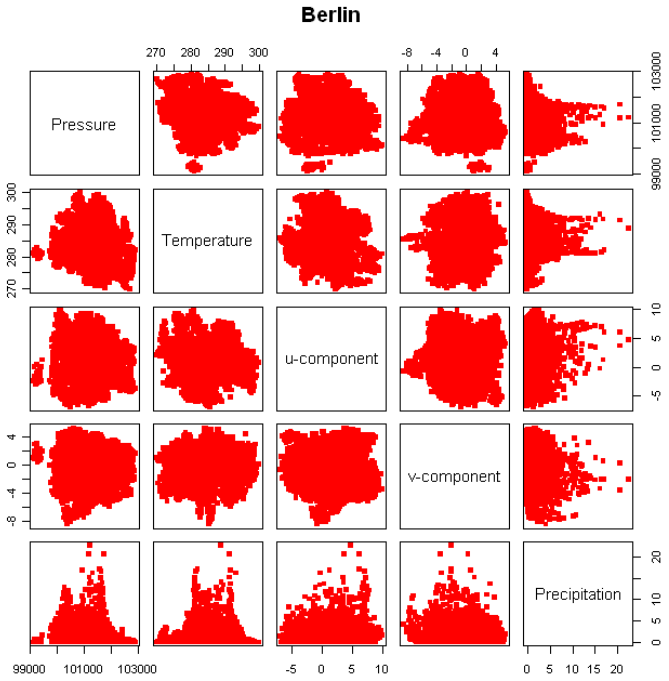

Figure 5.8: Cross-variate dependence structures of the ensemble forecasts at Berlin

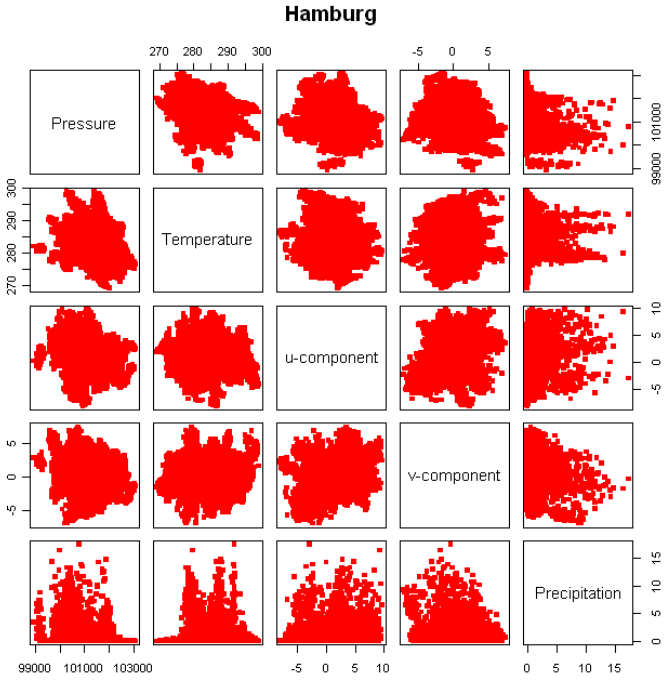

Figure 5.9: Cross-variate dependence structures of the ensemble forecasts at Hamburg

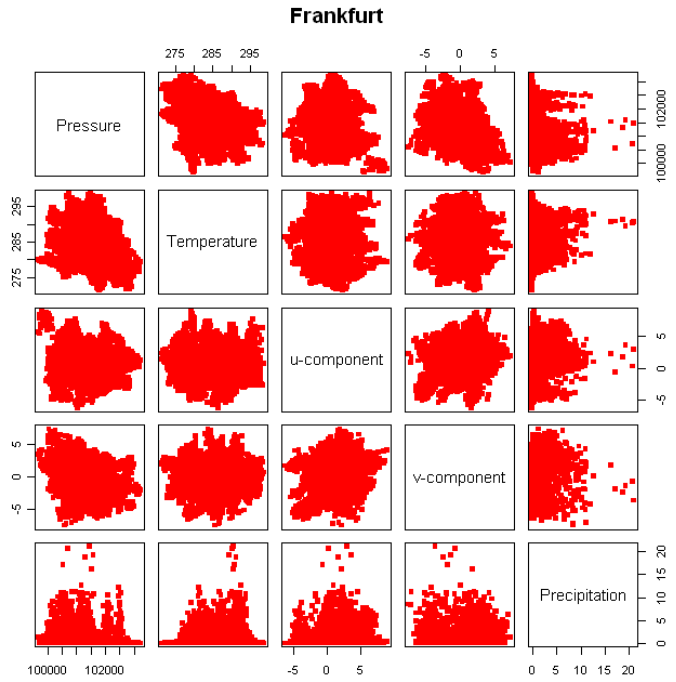

Figure 5.10: Cross-variate dependence structures of the ensemble forecasts at Frankfurt

5.2 Procedural method and some results

We now apply the ECC approach to the cleaned and transformed data set presented in the preceding section.
Before showing some results and commenting on them, we recall the procedure of ECC, compare Section 3.3, and explain how it is realised and implemented in our case study.

We are given the unprocessed raw ensemble forecasts with their corresponding order statistics by the ECMWF ensemble consisting of 50 members for temperature, sea level pressure, precipitation and wind vectors in form of u- and v-components at Berlin, Hamburg and Frankfurt for 49 prediction horizons. In our examples in this thesis, we confine ourselves to look-ahead times of 24 and 48 hours, respectively.
First of all, we employ the BMA technique, compare Subsection 2.2.1, to obtain statistically postprocessed univariate ensemble forecasts for each weather variable at each location individually, in each case for prediction horizons of 24 and 48 hours, respectively. In this connection, we employ the R package `ensembleBMA` to do the computation. Our test period comprises the 223 days from 14 March 2010 to 22 October 2010 in the case of 24 hours-ahead forecasts and the 222 days from 16 March 2010 to 23 October 2010 in the case of 48 hours-ahead forecasts, respectively. We employ a sliding training period of 40 days in the BMA postprocessing procedure for both prediction horizons. Note that a 40-day sliding training period consists of the 40 most recent days prior to the forecast

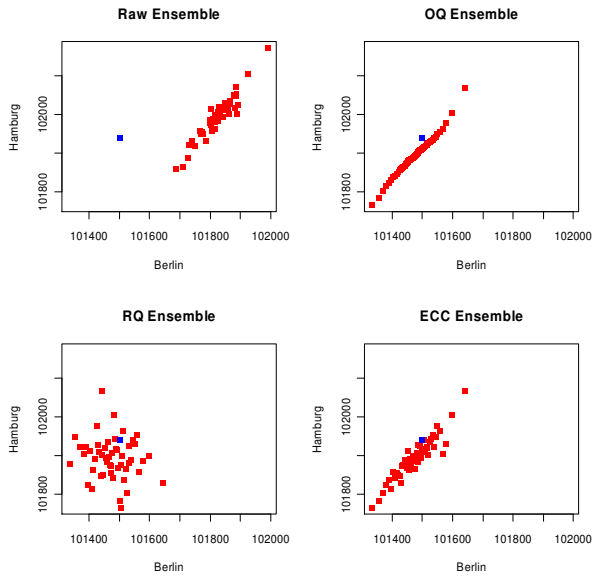

Figure 5.11: Raw, OQ, RQ and ECC ensemble for 24 hours-ahead forecasts for pressure in Pascal at Berlin and Hamburg valid at 0000 UTC on 15 June 2010. The verifying observation is indicated by a blue dot.

for which ensemble output and verifying observations are available. Hence, in terms of calendar days, this period typically corresponds to more than 40 days. Moreover, we assume that the ensemble members are exchangeable, which means that the BMA weights are equal for all ensemble members. We recall that the BMA predictive distribution is modelled by a mixture of normal distributions in the case of temperature, sea level pressure and u- and v-components and a mixture of gamma distributions with point mass at zero in the case of precipitation, respectively. Both versions are implemented in the `ensembleBMA` package.

In the next step, we have to generate a sample of size 50 from each BMA predictive distribution. As suggested in the ECC description in Section 3.3, this is achieved by taking the equally spaced $\frac{m}{50.5}$-quantiles, where m runs through all natural numbers from 1 to 50, as samples in each case. Those can be calculated easily by using the corresponding command in the `ensembleBMA` package.

Before doing the final ECC postprocessing, we first consider two other postprocessing approaches. Under the naive assumption of independence, we simply take the quantiles above of each BMA predictive distribution as the postprocessed multivariate ensemble. In each case, we can either employ the increasing $\frac{1}{50.5}, \frac{2}{50.5}, ..., \frac{50}{50.5}$-quantiles in **this** order or take the 50 different $\frac{m}{50.5}$-quantiles, where $m \in \{1, ..., 50\}$, in an order being randomly chosen. In what follows, the corresponding ensembles obtained by those two techniques are called ordered quantiles (OQ) ensemble and random quantiles (RQ) en-

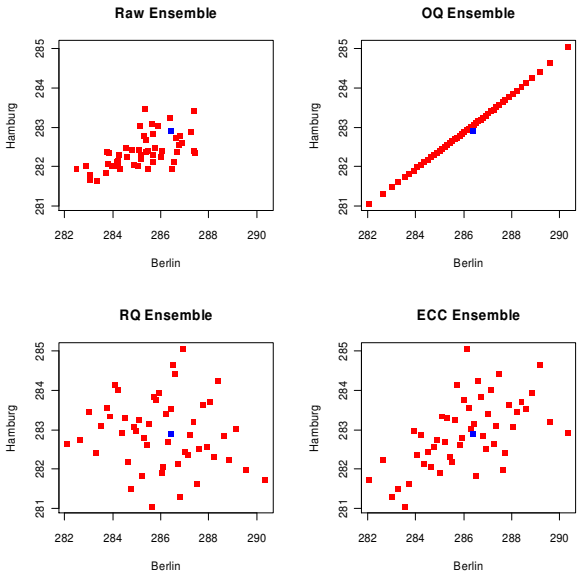

Figure 5.12: Raw, OQ, RQ and ECC ensemble for 24 hours-ahead forecasts for temperature in Kelvin at Berlin and Hamburg valid at 0000 UTC on 15 June 2010. The verifying observation is indicated by a blue dot.

semble, respectively. However, the methods introduced above ignore spatial, temporal and cross-variate dependencies in the sense that the multivariate rank order of the raw ensemble is lost, as we will illustrate in our plots later on.

Finally, the last ECC step corrects those shortcomings by rearranging the 50 quantiles with respect to the raw ensemble ranks, which leads to an inheritance of the multivariate dependence structure given by the unprocessed original ensemble and a preservation of the dynamical information from the physical model.

After those preliminary notes, we continue with the presentation of some selected results now.
Beforehand, we remark that the focus in this section is on the predictive performance of the ECC technique for certain single days only, whereas that over the whole test period is assessed in the next section.
The figures presented in what follows consist of four scatterplots in each case. More precisely, in each figure, we show the plots of the raw ensemble and the postprocessed OQ, RQ and ECC ensembles. Thereby, we confine ourselves to forecasts for a prediction horizon of 24 hours.

In Figure 5.11, the different ensemble forecasts for pressure at Berlin and Hamburg valid at 0000 UTC on 15 June 2010 are illustrated, where the verifying observation is

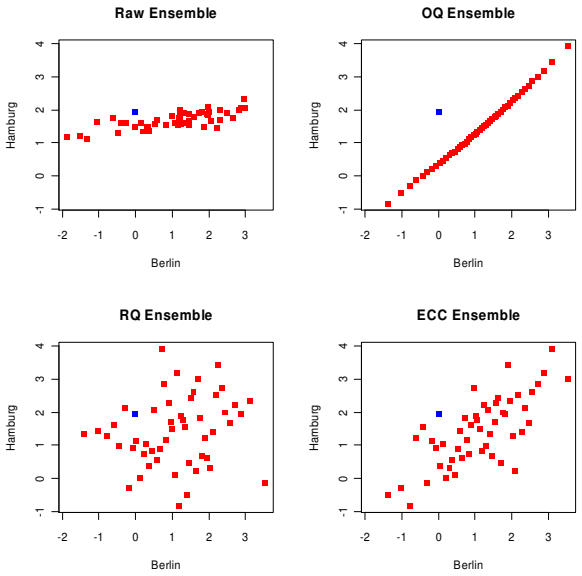

Figure 5.13: Raw, OQ, RQ and ECC ensemble for 24 hours-ahead forecasts for the wind vector u-component in metres per second at Berlin and Hamburg valid at 0000 UTC on 15 June 2010. The verifying observation is indicated by a blue dot.

indicated by a blue dot. As we can see, the ECC ensemble performs very well in this case by correcting biases and dispersion errors within the raw ensemble and retaining the dynamic properties of the original ensemble, whereas the OQ and the RQ ensembles obviously lose the bivariate rank order information from the raw ensemble. Analogously, we obtain similar results when considering all three locations Berlin, Hamburg and Frankfurt together for the situation described before.

Of course, scatterplots in the same manner as before can be made for the other weather variables, too. Exemplarily, we show the four ensemble forecast plots for temperature and the wind vector u-component, respectively, at Berlin and Hamburg valid at 0000 UTC on 15 June 2010 in Figures 5.12 and 5.13, respectively.

Finally, we consider two examples for cross-variate dependence structures. In Figure 5.14, the four ensembles for forecasts of temperature and pressure at Hamburg valid at 0000 UTC on 15 June 2010 are presented. Figure 5.15 shows the corresponding ensemble forecasts for the u- and v-wind vector components at Hamburg valid at 0000 UTC on 20 June 2010.

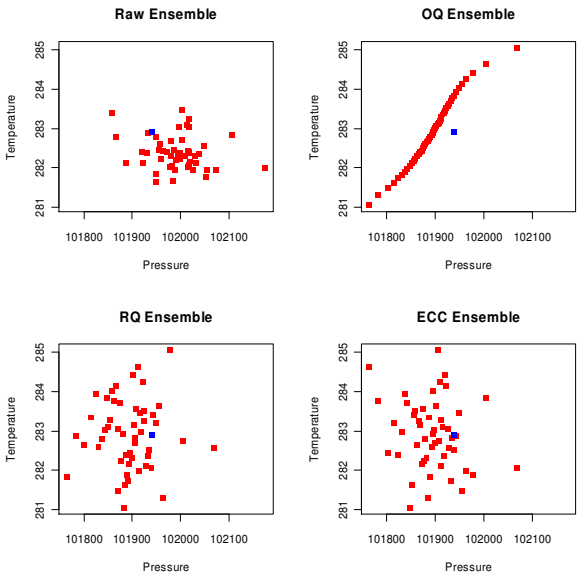

Figure 5.14: Raw, OQ, RQ and ECC ensemble for 24 hours-ahead forecasts for pressure in Pascal and temperature in Kelvin at Hamburg valid at 0000 UTC on 15 June 2010. The verifying observation is indicated by a blue dot.

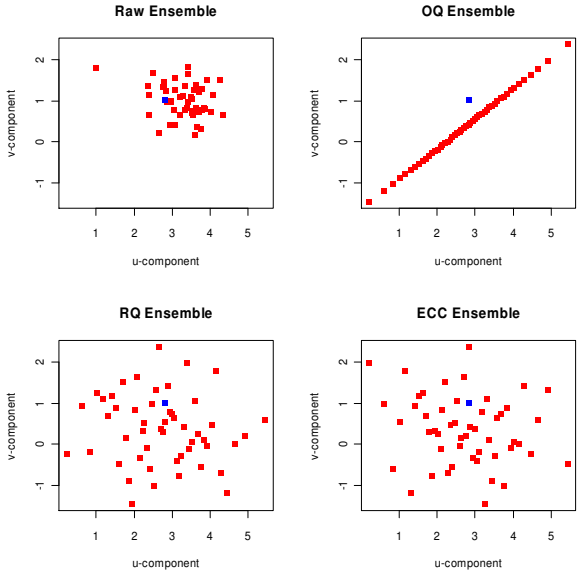

Figure 5.15: Raw, OQ, RQ and ECC ensemble for 24 hours-ahead forecasts for the wind vector u-and v-component in metres per second at Hamburg valid at 0000 UTC on 20 June 2010. The verifying observation is indicated by a blue dot.

5.3 Assessment of the ECC approach

In this section, we want to evaluate the predictive performance of our novel ECC technique over a test period by employing some of the assessment methods for multivariate probabilistic forecasting proposed by Gneiting et al. in [16] and introduced in Chapter 4. Moreover, the ECC ensemble performance is compared to that of both the raw ensemble and the two other postprocessed ensemble versions presented in the preceding section, namely the OQ and the RQ ensemble.

Before presenting our results, we briefly recall the tools from Chapter 4 that are used in the following considerations.
Note that the general aim of probabilistic forecasting is to achieve sharpness of the forecasts subject to calibration, see [14].
To check calibration, we employ the multivariate rank histogram (MRH) and compute the MRH discrepancy Δ, which quantifies deviation from uniformity, according to equation (4.1).
As an overall performance measure for multivariate forecasting techniques, we use the negatively orientated energy score (ES). Since we deal with ensemble forecasts, we employ the ensemble version of the ES given by equation (4.4) in our case study. In this context, note that we have to standardise both the observations and the ensemble forecasts whenever quantities which are incomparable in magnitude are involved in the computation of the ES. This is done as follows: For a fixed prediction horizon — as is the case in our examples later, where we consider 24 hours- and 48 hours-ahead forecasts —, we calculate the empirical mean

$$\mu^{ij} = \frac{1}{D}\sum_{d=1}^{D} y_d^{ij}$$

and the empirical standard deviation

$$\sigma^{ij} = \sqrt{\frac{1}{D-1}\sum_{d=1}^{D}(y_d^{ij} - \mu^{ij})^2}$$

for each weather quantity $i \in \{1, ..., I\}$ at each location $j \in \{1, ..., J\}$ and transform each observation y^{ij} and each forecast x^{ijm} of ensemble member m into the standardised versions $\tilde{y}^{ij}$ and $\tilde{x}^{ijm}$ by setting $\tilde{y}^{ij} := \frac{y^{ij}-\mu^{ij}}{\sigma^{ij}}$ and $\tilde{x}^{ijm} := \frac{x^{ijm}-\mu^{ij}}{\sigma^{ij}}$, respectively, on each day $d \in \{1, ..., D\}$. As a consequence, the ES has no unit when employing standardised values. However, if we determine the ES for a single weather variable at different locations, for instance, we need not standardise, and the unit of the ES is the same as that of the corresponding quantity in those cases.

Now we are ready to start with the discussion of our assessment results. Recall that in our case study, we focus on prediction horizons of 24 and 48 hours, respectively. The corresponding test periods comprise the 223 days from 14 March 2010 to 22 October

2010 in the case of 24 hours-ahead forecasts and the 222 days from 16 March 2010 to 23 October 2010 in the case of 48 hours-ahead forecasts, respectively. Remember that the sliding training period employed in the univariate ensemble postprocessing via BMA consists of 40 days for both prediction horizons.

Of course, there is a variety of possible scenarios for our case study by examining different combinations of weather quantities and/or locations. In this thesis, we confine ourselves to several exemplary situations, which might be particularly interesting. A summarising discussion of the results is given for each scenario.
Thereby, note that for the RQ ensemble, the ES values are an average over 100 runs in what follows.

Scenario 1

We assess the predictive performance of 24 hours-ahead and 48 hours-ahead forecasts for each weather variable separately for different combinations of our locations Berlin, Hamburg and Frankfurt. Hence, we are faced with two- and three-dimensional forecasts, respectively, here.
Selected MRHs for several cases and the corresponding discrepancies are presented in Figures 5.16 to 5.25 and Tables 5.6 and 5.7, respectively.
Tables 5.8 and 5.9 summarise the values for the ES in these situations.

For prediction horizons of both 24 and 48 hours, we observe that the ECC ensemble is much better calibrated than the unprocessed raw ensemble and the OQ ensemble in each of the cases we consider in this scenario. As a rule, as far as calibration is concerned, ECC also slightly outperforms the RQ ensemble. Especially in case of pressure, the ECC calibration performance turns out to be very nice compared to that of the other methods.
However, we have to be conscious of the fact that due to the random component in the development of the MRH, it is rather difficult to extract minor differences in the calibration of our different approaches. Therefore, for a deeper analysis, it might be better to take an average of for example 100 MRH simulations in each case as final MRH in order to diminish the random component. Furthermore, it would be helpful if we were able to apply our approaches to an extended data set consisting of longer test periods of for example a whole year in order to improve both the expressiveness of our results and the ECC calibration performance once again. Thereby, note that at the moment, we have 51 possible multivariate ranks, but "only" 222 and 223 days, respectively., to make the assessment. Nevertheless, the ECC calibration results presented in this scenario as a first step are convincing, and we look forward to future and additional case studies.

As far as the average ES as an overall performance measure over our test periods is concerned, ECC always outperforms both the raw and the OQ ensemble when dealing with pressure, temperature and wind vectors.

Ensemble	Pressure	Temperature	u-component	v-component	Precipitation
Raw	1.79	0.75	0.93	0.85	0.66
OQ	0.55	0.80	0.77	0.75	0.52
RQ	0.56	0.41	0.39	0.38	0.44
ECC	**0.30**	**0.40**	**0.36**	**0.36**	**0.43**

Table 5.6: Discrepancies Δ for prediction horizons of 24 hours for each weather variable separately for the three locations Berlin, Hamburg and Frankfurt together. The values correspond to the MRHs in Figures 5.16 to 5.20.

Ensemble	Pressure	Temperature	u-component	v-component	Precipitation
Raw	1.77	0.63	0.96	0.83	0.63
OQ	0.56	0.72	0.78	0.71	0.54
RQ	0.49	0.46	0.39	0.41	**0.35**
ECC	**0.36**	**0.44**	**0.36**	**0.33**	0.38

Table 5.7: Discrepancies Δ for prediction horizons of 48 hours for each weather variable separately for the three locations Berlin, Hamburg and Frankfurt together. The values correspond to the MRHs in Figures 5.21 to 5.25.

However, in the scenarios involving precipitation, ECC unfortunately fails to have a better ES than the raw ensemble, although it predominantly performs better than the OQ and RQ ensemble. Hence, it is obvious that this shortcoming arises from the univariate ensemble postprocessing made by the BMA method. Indeed, a detailed analysis of this circumstance shows that already the univariate BMA technique performs worse than the raw ensemble for the locations of Hamburg and Frankfurt in case of 24 hours-ahead forecasts and for all three observation sites in case of 48 hours-ahead forecasts, as far as the average continuous ranked probability score (CRPS), which is just the univariate variant of the ES, is concerned. In particular, it seems that BMA leads to bad results for days on which the verifying precipitation amount is rather high, while in the surrounding period of those days no or rather little precipitation occurs, among other possible error sources. Thus, the employment of an alternative univariate ensemble postprocessing method might yield results that are more satisfactory for the case of precipitation.
As a rule, ECC has a lower and thus a better ES than the RQ ensemble when considering pressure and temperature, while in case of wind vectors, that is, for u- and v-components, there are situations in which the ES of the RQ ensemble is lower than that of the ECC ensemble. As we have seen in Subsection 5.1.4, the spatial correlation structure for pressure and temperature at our three observation sites is rather strong, while that for both the u- and the v-component is less pronounced. Hence, we might conclude that the stronger the dependence pattern is the better ECC performs, and the full potential of our novel ECC technique is realised best in cases involving high correlation structures.

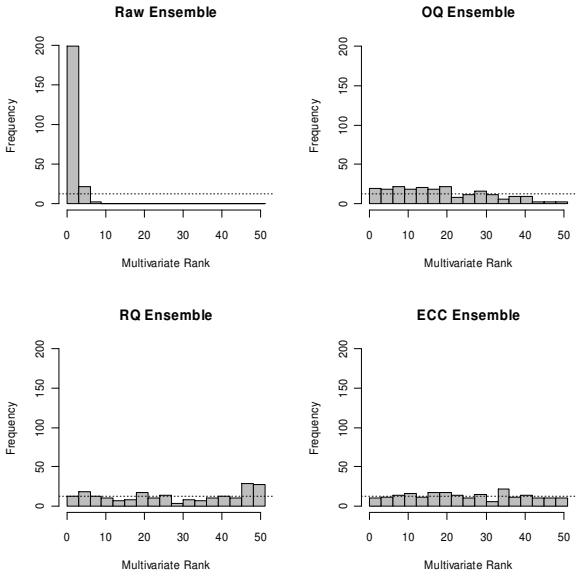

Figure 5.16: MRHs for 24 hours-ahead forecasts for pressure at Berlin, Hamburg and Frankfurt valid at 0000 UTC during the period from 14 March 2010 to 22 October 2010 comprising 223 days

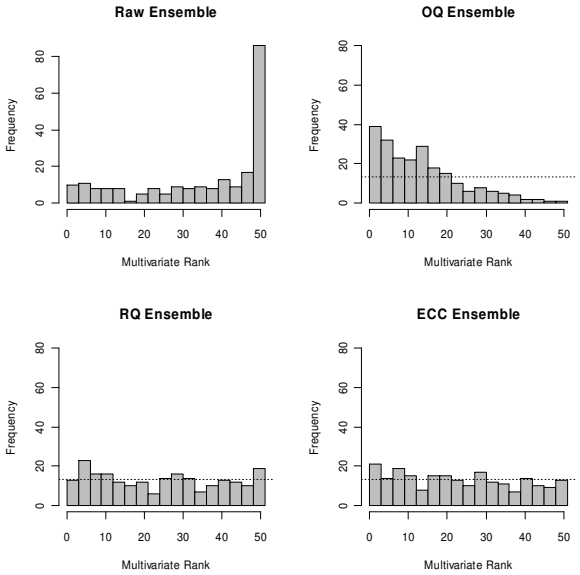

Figure 5.17: MRHs for 24 hours-ahead forecasts for temperature at Berlin, Hamburg and Frankfurt valid at 0000 UTC during the period from 14 March 2010 to 22 October 2010 comprising 223 days

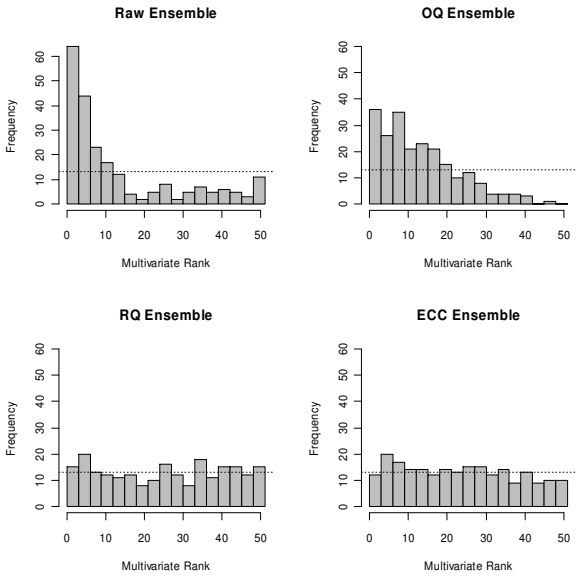

Figure 5.18: MRHs for 24 hours-ahead forecasts for the wind vector u-component at Berlin, Hamburg and Frankfurt valid at 0000 UTC during the period from 14 March 2010 to 22 October 2010 comprising 223 days

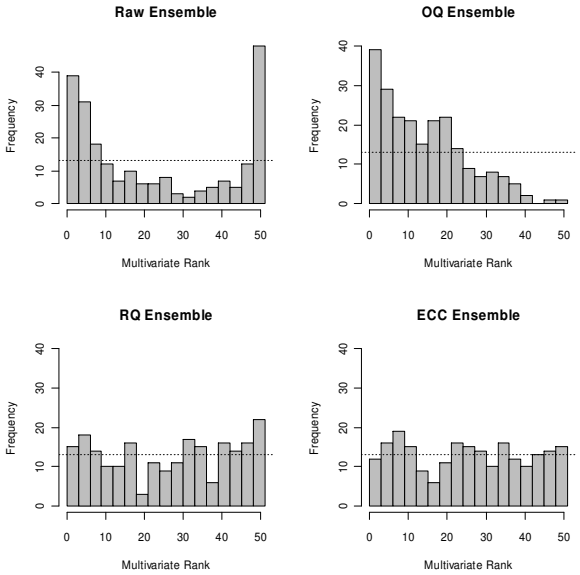

Figure 5.19: MRHs for 24 hours-ahead forecasts for the wind vector v-component at Berlin, Hamburg and Frankfurt valid at 0000 UTC during the period from 14 March 2010 to 22 October 2010 comprising 223 days

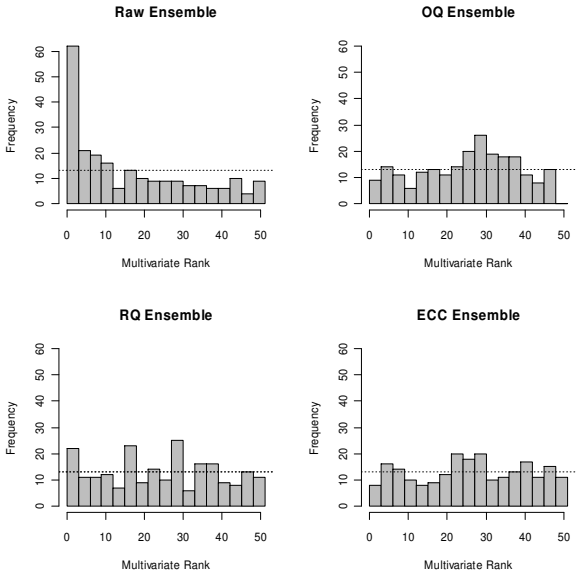

Figure 5.20: MRHs for 24 hours-ahead forecasts for precipitation at Berlin, Hamburg and Frankfurt valid at 0000 UTC during the period from 14 March 2010 to 22 October 2010 comprising 223 days

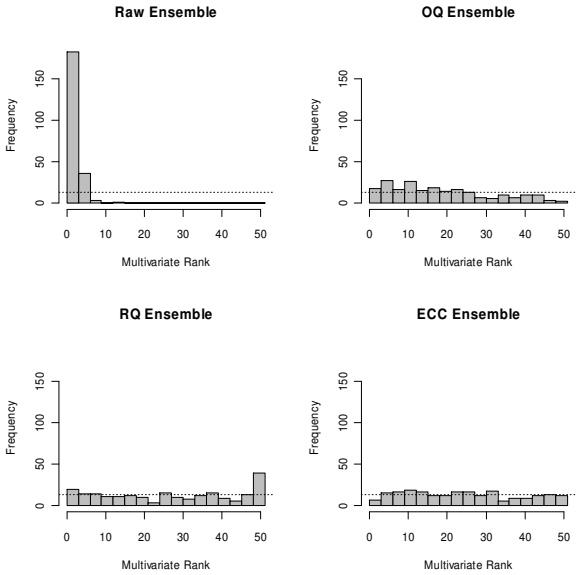

Figure 5.21: MRHs for 48 hours-ahead forecasts for pressure at Berlin, Hamburg and Frankfurt valid at 0000 UTC during the period from 16 March 2010 to 23 October 2010 comprising 222 days

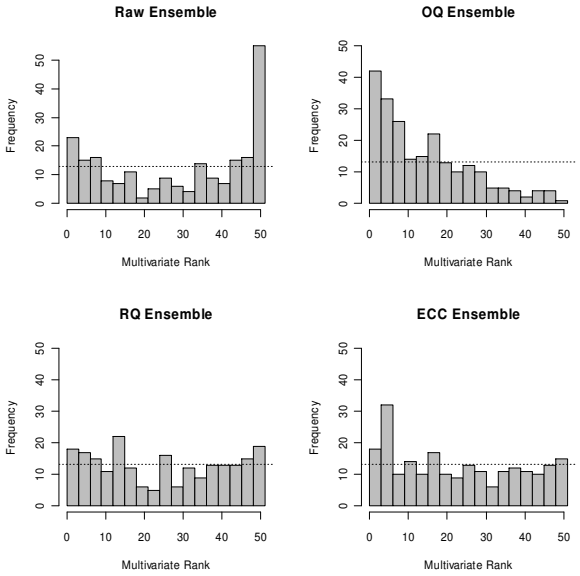

Figure 5.22: MRHs for 48 hours-ahead forecasts for temperature at Berlin, Hamburg and Frankfurt valid at 0000 UTC during the period from 16 March 2010 to 23 October 2010 comprising 222 days

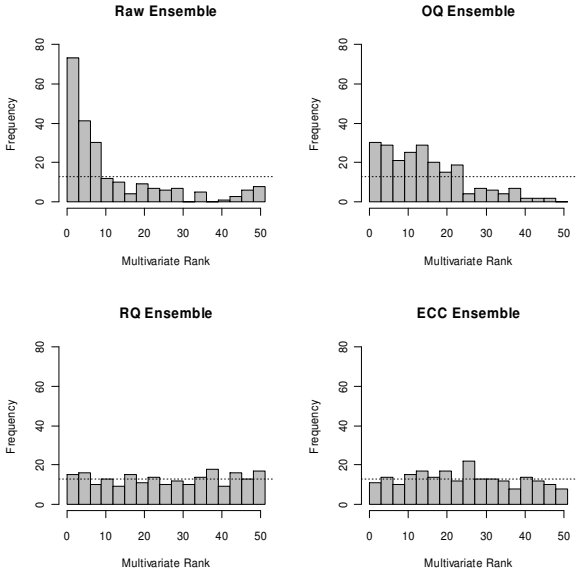

Figure 5.23: MRHs for 48 hours-ahead forecasts for the wind vector u-component at Berlin, Hamburg and Frankfurt valid at 0000 UTC during the period from 16 March 2010 to 23 October 2010 comprising 222 days

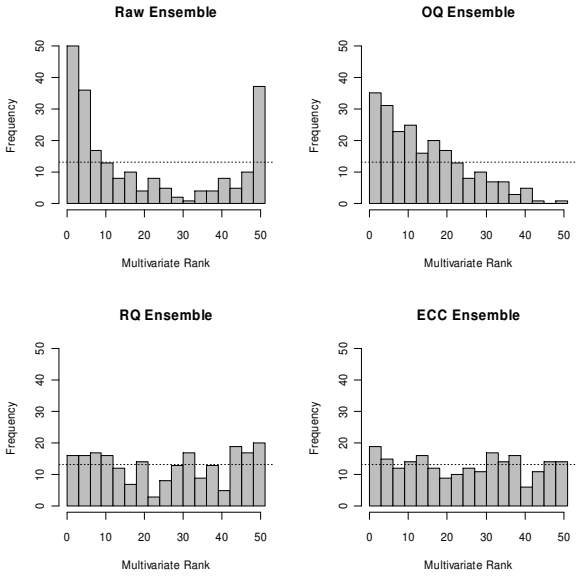

Figure 5.24: MRHs for 48 hours-ahead forecasts for the wind vector v-component at Berlin, Hamburg and Frankfurt valid at 0000 UTC during the period from 16 March 2010 to 23 October 2010 comprising 222 days

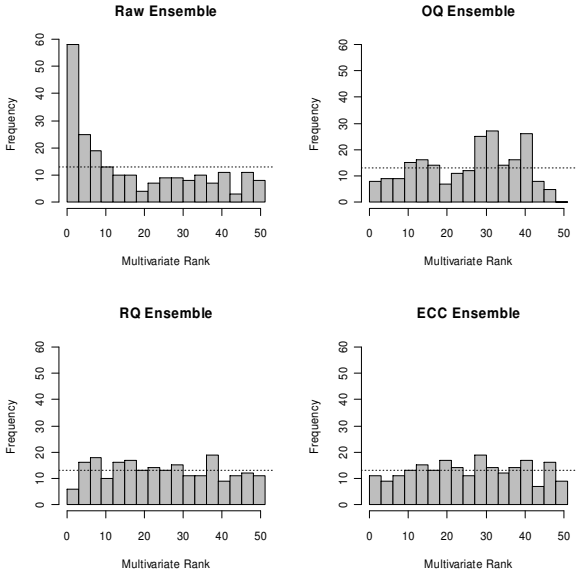

Figure 5.25: MRHs for 48 hours-ahead forecasts for precipitation at Berlin, Hamburg and Frankfurt valid at 0000 UTC during the period from 16 March 2010 to 23 October 2010 comprising 222 days

Berlin, Hamburg and Frankfurt

Ensemble	Pressure (Pa)	Temperature (K)	u-component (m/s)	v-component (m/s)	Precipitation (mm)
Raw	1271.6	2.183	1.710	1.813	**1.139**
OQ	76.5	1.735	1.371	1.566	1.241
RQ	76.7	1.638	**1.288**	**1.486**	1.205
ECC	**74.6**	**1.636**	1.292	1.489	1.206

Hamburg and Frankfurt

Ensemble	Pressure (Pa)	Temperature (K)	u-component (m/s)	v-component (m/s)	Precipitation (mm)
Raw	1232.3	1.630	1.407	1.535	**0.952**
OQ	60.8	1.260	1.039	1.344	1.043
RQ	60.6	1.212	**0.980**	1.299	1.028
ECC	**59.6**	**1.210**	**0.980**	**1.297**	1.028

Hamburg and Berlin

Ensemble	Pressure (Pa)	Temperature (K)	u-component (m/s)	v-component (m/s)	Precipitation (mm)
Raw	334.4	1.702	1.233	1.183	**0.715**
OQ	57.6	1.316	1.081	0.974	0.777
RQ	58.6	1.264	**1.046**	**0.923**	0.773
ECC	**56.8**	**1.262**	1.047	0.928	0.764

Berlin and Frankfurt

Ensemble	Pressure (Pa)	Temperature (K)	u-component (m/s)	v-component (m/s)	Precipitation (mm)
Raw	1278.9	1.845	1.414	1.556	**0.769**
OQ	63.0	1.438	1.056	1.292	0.788
RQ	62.4	**1.361**	**0.997**	**1.245**	0.777
ECC	**61.5**	1.362	0.999	1.247	0.776

Table 5.8: Average energy scores for prediction horizons of 24 hours for each weather variable separately for different combinations of locations

Berlin, Hamburg and Frankfurt

Ensemble	Pressure (Pa)	Temperature (K)	u-component (m/s)	v-component (m/s)	Precipitation (mm)
Raw	1229.3	2.216	1.644	1.753	**1.204**
OQ	127.4	1.939	1.467	1.678	1.297
RQ	129.8	1.850	**1.392**	**1.586**	1.266
ECC	**125.4**	**1.846**	**1.392**	1.588	1.267

Hamburg and Frankfurt

Ensemble	Pressure (Pa)	Temperature (K)	u-component (m/s)	v-component (m/s)	Precipitation (mm)
Raw	1194.2	1.661	1.356	1.493	**1.031**
OQ	100.7	1.426	1.112	1.432	1.085
RQ	101.6	1.380	**1.054**	**1.377**	1.074
ECC	**99.5**	**1.375**	1.056	**1.377**	1.073

Berlin and Hamburg

Ensemble	Pressure (Pa)	Temperature (K)	u-component (m/s)	v-component (m/s)	Precipitation (mm)
Raw	323.7	1.768	1.242	1.150	**0.743**
OQ	97.7	1.517	1.200	1.065	0.806
RQ	100.8	1.470	1.170	**1.013**	0.801
ECC	**97.1**	**1.466**	**1.165**	1.016	0.796

Berlin and Frankfurt

Ensemble	Pressure (Pa)	Temperature (K)	u-component (m/s)	v-component (m/s)	Precipitation (mm)
Raw	1240.8	1.833	1.308	1.494	**0.796**
OQ	104.3	1.570	1.087	1.371	0.839
RQ	104.5	1.4	**1.036**	**1.316**	0.826
ECC	**102.4**	1.499	1.040	1.318	0.825

Table 5.9: Average ES values for prediction horizons of 48 hours for each weather variable separately for different combinations of locations

Ensemble	Ber	Ham	Fra	Ber+Ham	Ber+Fra	Ham+Fra	Ber+Ham +Fra
Raw	0.5739	0.2466	2.2337	0.6263	2.2134	2.1439	2.2106
OQ	0.1874	0.1685	0.2078	0.2767	0.3089	0.2935	0.3723
RQ	0.1772	**0.1614**	0.1972	0.2590	0.2863	0.2751	0.3453
ECC	**0.1766**	0.1617	**0.1970**	**0.2581**	**0.2858**	**0.2746**	**0.3447**

Table 5.10: Average (standardised) ES values for prediction horizons of 24 hours when considering pressure and temperature for different combinations of the locations Berlin (Ber), Hamburg (Ham) and Frankfurt (Fra)

Ensemble	Ber	Ham	Fra	Ber+Ham	Ber+Fra	Ham+Fra	Ber+Ham +Fra
Raw	0.5523	0.2739	2.1116	0.6222	2.1559	2.0872	2.1483
OQ	0.2399	0.2286	0.2539	0.3551	0.3768	0.3674	0.4613
RQ	0.2248	0.2146	0.2372	0.3320	0.3493	0.3425	0.4285
ECC	**0.2246**	**0.2140**	**0.2365**	**0.3296**	**0.3477**	**0.3406**	**0.4258**

Table 5.11: Average (standardised) ES values for prediction horizons of 48 hours when considering pressure and temperature for different combinations of the locations Berlin (Ber), Hamburg (Ham) and Frankfurt (Fra)

Scenario 2

Now we consider examples, in which both cross-variate and inter-locational dependence structures are involved. We examine the predictive performances in case of pressure and temperature on the one hand and u- and v-wind vector components on the other hand for several combinations of our three locations, which leads to two-, four-, or six-dimensional forecasts, respectively.

Scenario 2a — Pressure and temperature

On the basis of Tables 5.10 and 5.11, we recognise that whenever considering pressure and temperature simultaneously for different combinations of locations, ECC has the best ES values in those situations nearly throughout, for prediction horizons of both 24 and 48 hours. In particular, it is apparent that the ECC ensemble outperforms the raw ensemble to a great extent, and thus, the improvement of the forecasts obtained by the ECC approach with respect to the ES is remarkable. Note that in those examples, we quasi deal with a kind of "maximal" correlation structure within our data set, that is, we consider those weather quantities showing the strongest cross-variate and inter-locational dependence patterns.

Ensemble	24 hours-ahead	48 hours-ahead
Raw	0.88	0.84
OQ	1.13	1.09
RQ	0.45	**0.37**
ECC	**0.35**	0.39

Table 5.12: Discrepancies Δ for prediction horizons of 24 and 48 hours, respectively, when considering pressure and temperature at all the three locations Berlin, Hamburg and Frankfurt simultaneously. The values belong to the MRHs in Figures 5.26 and 5.27.

Ensemble	24 hours-ahead	48 hours-ahead
Raw	1.17	1.01
OQ	0.96	1.03
RQ	0.36	0.47
ECC	**0.34**	**0.39**

Table 5.13: Discrepancies Δ for prediction horizons of 24 and 48 hours, respectively, when considering pressure and temperature at the locations Berlin and Hamburg simultaneously. The values belong to the MRHs in Figures 5.28 and 5.29.

Ensemble	24 hours-ahead	48 hours-ahead
Raw	1.18	1.17
OQ	1.06	1.02
RQ	0.43	0.40
ECC	**0.37**	**0.34**

Table 5.14: Discrepancies Δ for prediction horizons of 24 and 48 hours, respectively, when considering pressure and temperature at the locations Berlin and Frankfurt simultaneously. The values belong to the MRHs in Figures 5.30 and 5.31.

To illustrate the calibration performance of our different ensembles, we take a look at the MRHs in Figures 5.26 to 5.31, which show exemplary results for selected situations. The corresponding discrepancies can be found in Tables 5.12 to 5.14. It can be seen that the convincing performance of the ECC method is corroborated by the MRH plots such that, in a nutshell, the results in this scenario are especially nice.

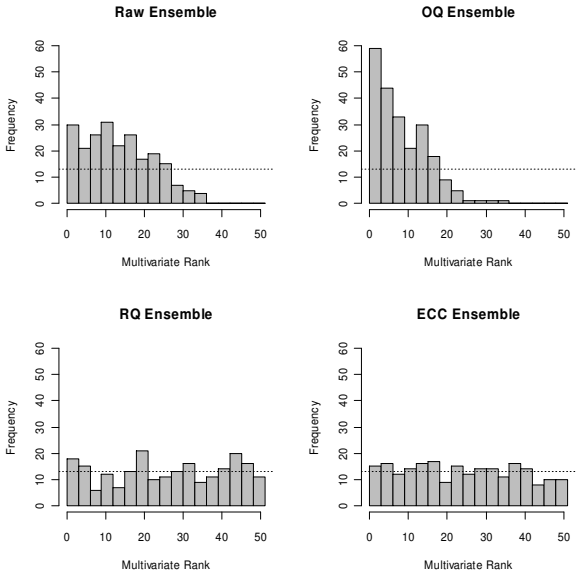

Figure 5.26: MRHs for 24 hours-ahead forecasts for pressure and temperature at Berlin, Hamburg and Frankfurt valid at 0000 UTC during the period from 14 March 2010 to 22 October 2010 comprising 223 days

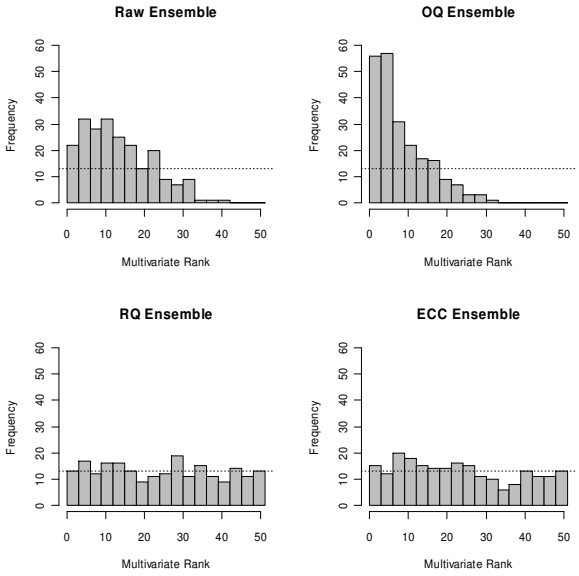

Figure 5.27: MRHs for 48 hours-ahead forecasts for pressure and temperature at Berlin, Hamburg and Frankfurt valid at 0000 UTC during the period from 16 March 2010 to 23 October 2010 comprising 222 days

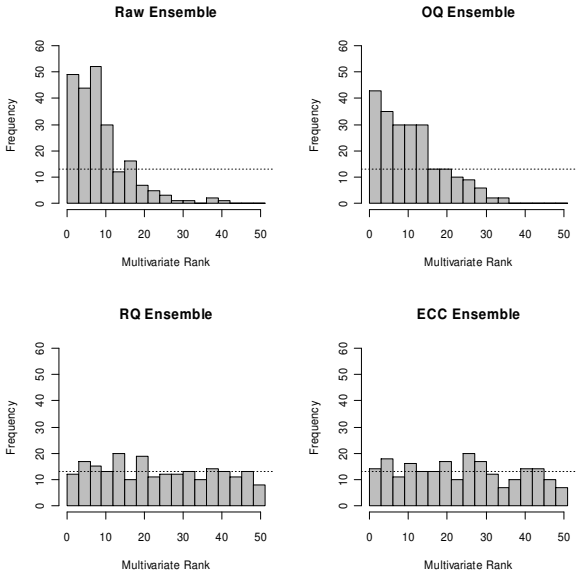

Figure 5.28: MRHs for 24 hours-ahead forecasts for pressure and temperature at Berlin and Hamburg valid at 0000 UTC during the period from 14 March 2010 to 22 October 2010 comprising 223 days

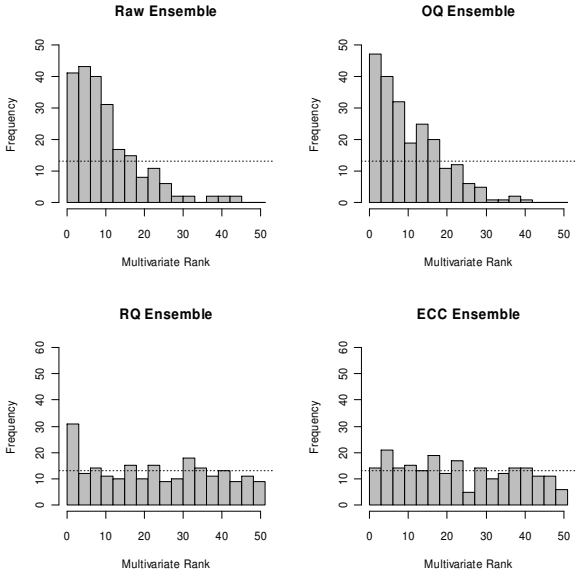

Figure 5.29: MRHs for 48 hours-ahead forecasts for pressure and temperature at Berlin and Hamburg valid at 0000 UTC during the period from 16 March 2010 to 23 October 2010 comprising 222 days

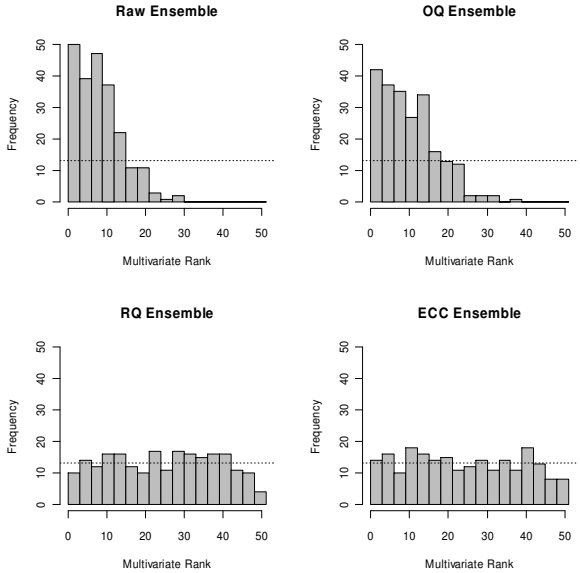

Figure 5.30: MRHs for 24 hours-ahead forecasts for pressure and temperature at Berlin and Frankfurt valid at 0000 UTC during the period from 14 March 2010 to 22 October 2010 comprising 223 days

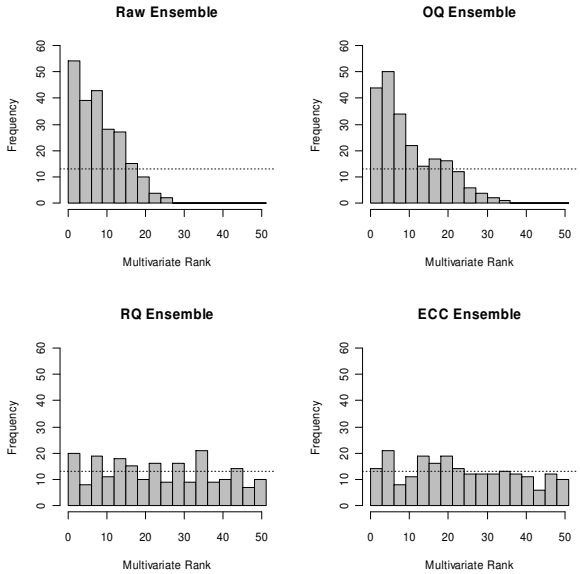

Figure 5.31: MRHs for 48 hours-ahead forecasts for pressure and temperature at Berlin and Frankfurt valid at 0000 UTC during the period from 16 March 2010 to 23 October 2010 comprising 222 days

Ensemble	Ham	Ham+Fra	Ber+Ham+Fra
Raw	0.538	1.088	1.289
OQ	0.472	0.882	1.063
RQ	**0.447**	**0.826**	**0.988**
ECC	0.450	0.830	0.992

Table 5.15: Average (standardised) ES values for prediction horizons of 24 hours when considering u- and v-wind vector components for different combinations of the locations Berlin (Ber), Hamburg (Ham) and Frankfurt (Fra)

Ensemble	Ham	Ham+Fra	Ber+Ham+Fra
Raw	0.548	1.068	1.257
OQ	0.533	0.947	1.146
RQ	**0.502**	**0.886**	**1.065**
ECC	0.505	0.890	1.070

Table 5.16: Average (standardised) ES values for prediction horizons of 48 hours when considering u- and v-wind vector components for different combinations of the locations Berlin (Ber), Hamburg (Ham) and Frankfurt (Fra)

Scenario 2b — u- and v-wind vector components

In contrast to the preceding Scenario 2a in which we dealt with pressure and temperature involving rather strong dependence structures, we now consider wind vectors, that is, u- and v-components, which show correlation patterns being not that pronounced according to the plots in Subsection 5.1.4.
Thereby, we concentrate on three examples, namely wind vectors at Hamburg only, Hamburg and Frankfurt together and Hamburg, Berlin and Frankfurt together, for prediction horizons of both 24 and 48 hours.

The ES values of this scenario, which are given in Tables 5.15 and 5.16, turn out to be as expected: The ECC ensemble outperforms both the raw and the OQ ensemble, but has a slightly higher ES than the RQ ensemble in all the examples.

In addition, we also obtain reasonable results as far as calibration is concerned, where we only focus on wind vectors at all the three observation sites together here, compare Figures 5.32 and 5.33 and the corresponding discrepancies in Table 5.17.

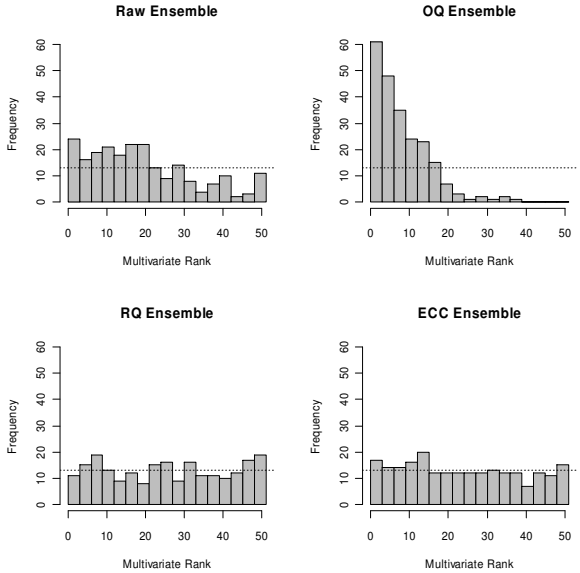

Figure 5.32: MRHs for 24 hours-ahead forecasts for u- and v-wind vector components at Berlin, Hamburg and Frankfurt valid at 0000 UTC during the period from 14 March 2010 to 22 October 2010 comprising 223 days

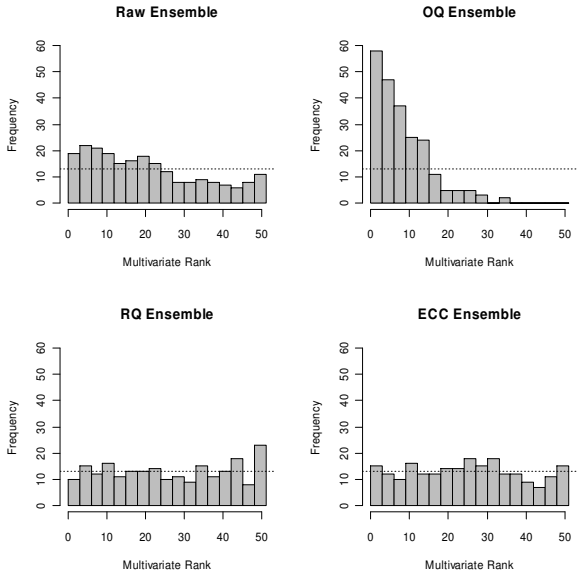

Figure 5.33: MRHs for 48 hours-ahead forecasts for u- and v-wind vector components at Berlin, Hamburg and Frankfurt valid at 0000 UTC during the period from 16 March 2010 to 23 October 2010 comprising 222 days

Ensemble	24 hours-ahead	48 hours-ahead
Raw	0.57	0.46
OQ	1.16	1.15
RQ	0.43	0.37
ECC	**0.35**	**0.35**

Table 5.17: Discrepancies Δ for prediction horizons of 24 and 48 hours, respectively, when considering u- and v-wind vector components at all the three locations Berlin, Hamburg and Frankfurt simultaneously. The values belong to the MRHs in Figures 5.32 and 5.33.

Ensemble	24 hours-ahead	48 hours-ahead
Raw	2.73	2.63
OQ	1.48	1.58
RQ	**1.38**	**1.47**
ECC	1.39	1.48

Table 5.18: Average (standardised) ES values for prediction horizons of 24 and 48 hours, respectively, when considering all weather variables and all the three locations Berlin, Hamburg and Frankfurt simultaneously.

Ensemble	24 hours-ahead	48 hours-ahead
Raw	**0.32**	0.45
OQ	1.41	1.43
RQ	0.34	**0.33**
ECC	0.40	0.40

Table 5.19: Discrepancies Δ for prediction horizons of 24 and 48 hours, respectively, when considering all weather variables and all the three locations Berlin, Hamburg and Frankfurt simultaneously. The values belong to the MRHs in Figures 5.34 and 5.35.

Scenario 3

We close our evaluation examples with the examination of the different ensemble predictive performances when considering all three locations and all five weather variables together, which means that we deal with 15-dimensional forecasts in this case.
The corresponding ES values for this scenario are shown in Table 5.18, while the MRHs and the corresponding discrepancies are presented in Figures 5.34 and 5.35 and in Table 5.19, respectively.

With regard to the average ES over our test periods for prediction horizons of 24 and 48 hours, respectively, we obtain reasonable results. The ECC ensemble performs substantially better than both the unprocessed raw ensemble and the OQ ensemble. Due to the

fact that the cross-variate dependence structures are rather weak when considering all weather quantities simultaneously, compare Subsection 5.1.4, the ES of the RQ ensemble is slightly lower than that of the ECC ensemble.

In contrast to the expected results with respect to the ES, the calibration performances illustrated by the MRHs are not really that nice in this scenario. The OQ ensemble turns out to be extremely uncalibrated, while the calibration of the unprocessed raw ensemble, the RQ and the ECC ensemble is essentially very similar. However, this seems unreasonable because in all the situations we considered in the lower-dimensional scenarios before, the raw ensemble was calibrated worst, and ECC generally performed best. Hence, it is not obvious that the raw ensemble performs that well in our scenario here. Generally, we observe that the higher the considered dimension of the forecasts is the worse the OQ ensemble is calibrated and the more the distinction between the calibration of the raw, RQ and ECC ensembles, respectively, vanishes. Perhaps, the MRH is inadequate to assess calibration when sufficiently high dimensions are involved and should only be employed in lower-dimensional situations. Thus, when applying the MRH concept, it is advisable to focus on preferably few dimensions which are especially of interest depending on the situation we consider. For the stated reasons, the development of alternative multivariate calibration evaluation tools is required to a great extent.

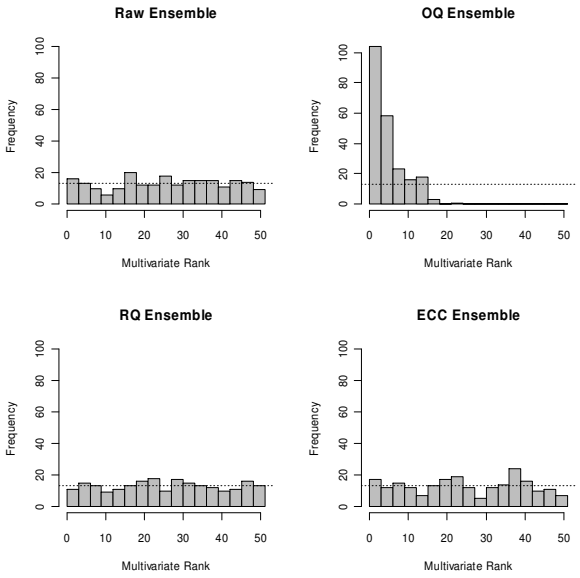

Figure 5.34: MRHs for 24 hours-ahead forecasts for all weather variables at Berlin, Hamburg and Frankfurt valid at 0000 UTC during the period from 14 March 2010 to 22 October 2010 comprising 223 days

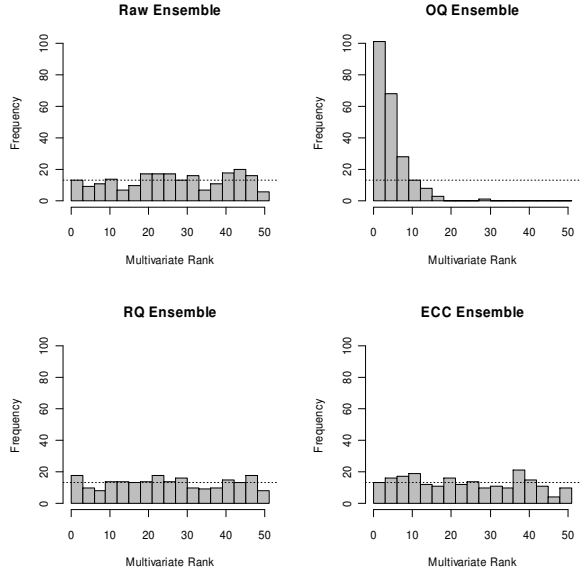

Figure 5.35: MRHs for 48 hours-ahead forecasts for all weather variables at Berlin, Hamburg and Frankfurt valid at 0000 UTC during the period from 16 March 2010 to 23 October 2010 comprising 222 days

Chapter 6

Summary, discussion and annotations

In this final chapter, we want to summarise the main aspects in the context of the novel ECC approach introduced in the thesis on hand, discuss our results and provide both an overview of topics and an outlook on possible future work connected to ECC. In particular, we state several proposals for further contributions to the ECC method, that might lead to an improvement and advancement by solving problems ECC is not yet able to handle.

The motivation to develop multivariate ensemble postprocessing methods for probabilistic weather forecasting is due to the fact that state-of-the-art ensemble postprocessing techniques like BMA or EMOS, compare Section 2.2, can be applied to a single weather quantity, at a single location and for a single prediction horizon only, which means that spatial, temporal and cross-variate dependencies cannot be handled by those methods. However, dependence structures are exceedingly important in a wide range of applications, and thus, there is a critical need of techniques addressing them.

There exist several techniques being able to take account of spatial dependencies.
In [10], Gel et al. propose the geostatistical output perturbation (GOP) method to obtain calibrated probabilistic mesoscale weather field forecasts. In a nutshell, GOP dresses a single deterministic weather field prediction with Gaussian error fields which are produced by geostatistical methods in order to get statistical ensembles of weather fields respecting spatial correlations. However, this method does not apply to an ensemble of weather field forecasts.
In [2], Berrocal et al. introduce the so-called spatial BMA approach, which combines the GOP technique with the BMA approach described in Subsection 2.2.1 in this thesis. Since spatial BMA adopts the advantages of both the GOP and the BMA methods, it applies to a dynamical ensemble of weather field forecasts and honours the flow-dependent information contained therein, while taking account of spatial correlations.
In the case of quantitative precipitation, the two-stage spatial model, which is proposed

by Berrocal et al. in [3], provides a useful tool producing correlated probabilistic forecasts of precipitation accumulation at multiple sites simultaneously.
However, all those approaches fail to take temporal or cross-variate dependencies into account, which is very unfortunate.

In this thesis, the challenge of physically coherent spatio-temporal probabilistic forecasting has been solved by introducing a novel tool called ECC, which essentially consists of two steps. In the ECC approach, we begin with the employment of well-known univariate ensemble postprocessing techniques to obtain calibrated and sharp predictive distributions for each weather quantity, at each location and for for each look-ahead time separately. Then, in a second step, we generate as many samples from each of the predictive distributions in the first step as there are ensemble members, which can be done by taking the $\frac{m}{M+0.5}$-quantiles, where M denotes the ensemble size and m passes all natural numbers from 1 to M in each case, and rearrange them in the order of the original raw ensemble forecasts. Hence, the multivariate rank structure of the raw ensemble and thus the dynamic information from the physical model are kept by the postprocessed ECC ensemble.

Moreover, we have shown that ECC — as suggested by the name — can indeed be considered as a copula approach by pointing out its relationships to discrete copulas. Consequently, ECC joins the long list of methods based on the concept of copulas, which is employed in a wealth of applications, not only in meteorology and climatology, but also in hydrology, extreme value theory, engineering and mainly in economics and mathematical finance. One reason for the popularity of copulas is that they capture the dependence structure among the quantities involved, which are described by random variables, as is the case in our ECC method. Hence, the adoption of a copula often replaces the assumption of independence, which is unjustified in the majority of cases. Since the field of copulas has rapidly developed over the last years and is of immense interest in current research, affecting both theoretical and practical aspects, this notion will be likely to yield further interesting and useful stuff in the future, and there might be some new application areas of copulas to discover as well.
We recommend [52] for a brief overview of purely mathematical aspects of copulas with predominantly recent research results.

In this thesis, the ECC approach has been applied for 24 and 48 hours-ahead forecasts in a case study to the ECMWF ensemble consisting of 50 members over test periods in the year 2010, and we have illustrated the results in several scatterplots clarifying the general idea of ECC. Furthermore, the predictive performance of our novel tool has been assessed by employing some of the multivariate verification methods proposed in [16] and compared to that of the raw ensemble and other postprocessed ensembles, namely the ordered quantiles (OQ) and random quantiles (RQ) ensemble as introduced in Chapter 5.
The ECC ensemble turned out to be much better calibrated than both the raw and the

OQ ensemble in each case if we consider scenarios in rather low dimensions, for which the multivariate rank histogram (MRH) is an adequate assessment tool to check deviation from uniformity. In general, the calibration of our ECC approach was also slightly better than that of the RQ ensemble in the situations we considered in this thesis. Especially in case of pressure, the ECC calibration performance was very nice. However, note that due to the random component involved in the MRH, it is difficult to extract minor differences in calibration between the techniques. The random component might be diminished by taking the average of many histogram simulations as final MRH.
Moreover, as far as the average energy score (ES) over our test periods are concerned, ECC remarkably outperformed both the raw and the OQ ensemble in each scenario dealing with pressure, temperature and/or wind vectors. However, in case of precipitation, ECC unfortunately failed to have a better ES than the raw ensemble, but note that in this situation, already the univariate postprocessed ensemble obtained by the BMA technique partly performed worse than the raw ensemble if we consider the average continuous ranked probability score (CRPS), which is just the univariate variant of the ES. The employment of a different univariate ensemble postprocessing method might correct this shortcoming in our case study. In addition, as a rule, ECC performed better than the RQ ensemble when considering combinations with pressure and temperature, the weather variables which show the strongest dependence structure, while in case of wind vectors, for which the dependence patterns are not that pronounced, there were scenarios in which the ES of the RQ ensemble was lower than or equal to that of the ECC ensemble. In general, we observed that the stronger the correlation structure among the ensemble members is the better ECC performed in our case study.
In a nutshell, ECC showed a convincing performance in our case study and provided an improvements of the forecasts.
However, note that we employed a sliding training period of 40 days for our computations without exceptions, which need not be the optimal choice in each scenario. Varying the training period with respect to the considered weather quantity, for example, could lead to an improvement of the BMA and ECC predictive performance, respectively. Moreover, in order to intensify the expressiveness of our results, it is convenient to extend the data set to the period of a whole year covering all seasons, for instance, which cannot be done in this thesis, but is a task for the future.

Against the backdrop of the fact that there are already R packages like `ensembleBMA` or `ensembleMOS`, for instance, doing the computation of the univariate predictive distributions needed in the ECC technique, another merit of our novel method, besides the nice properties mentioned above, is that it can be implemented easily. Due to the simplicity and clarity of the ECC notion, the R code of the multivariate copula step itself is straightforward and comparably short. Note that there is no general computational limit for the ECC procedure as far as high forecast dimensions and/or ensemble sizes are concerned.

The design of ECC allows using this tool for problems, in which the physical consistency

of spatio-temporal predictive distributions plays a decisive role. Typical applications may include air traffic management or ship routeing, as well as flood management in the case of quantitative precipitation forecasts. In addition, it is conceivable to employ the ECC approach in the context of event planning, agricultural issues and energy recovery and supply, respectively. Perhaps, the basic idea of ECC might be useful in other fields apart from weather forecasting, too.

Obviously, ECC offers a simple, but very useful technique for multivariate probabilistic forecasting and multivariate ensemble postprocessing, respectively.
Since parameters are only estimated for the univariate distributions in the first ECC step, but not in the crucial multivariate copula step itself, our novel tool can be considered as a non-parametric multivariate approach.
While ECC only takes account of dependencies within the ensemble forecasts by focusing on the rank structure and not of past observation dependence patterns, for instance, methods with other priorities to address the challenge of multivariate probabilistic forecasting are developing, too. In this context, the construction of multivariate predictive distributions for weather variables incorporating dependencies in a parametric technique, which is surely more complex than the clear ECC notion, is a current research topic. A first approach to estimate the joint multivariate CDF of several weather quantities might be to employ a Gaussian copula model.
Moreover, it is very promising to use the concept of so-called graphical models in order to model spatial, temporal and cross-variate dependencies in weather patterns. Hence, this is also of great interest in topical research. For an introduction to the interesting and useful field of graphical models, we refer to [7] or [32].
All in all, there are various multivariate probabilistic forecasting approaches differing in basic concepts and complexity, as we have pointed out.

Despite of the numerous positive and nice properties of the ECC approach named before, there are still some open questions in that context.
At the end of this chapter, we discuss several additional issues and suggestions for possible improvement, as far as the ECC technique and its methodological frame are concerned, in what follows.
From a theoretical point of view, it is a very interesting task to evolve the theory of discrete copulas, which is strongly connected to the ECC notion as we have seen, by studying the possibility to generalise at least some of the results presented in Subsection 3.4.5 for the bivariate case to similar ones in the multivariate situation. A first step in this direction might be to employ the frame of n-stochastic matrices or stochastic arrays, compare Remark 3.26, in order to extend the theory to the multivariate case. However, the details would still have to be worked out carefully. A rigorous mathematical treatment of this open question would contribute to an advanced theoretical frame for the novel ECC technique and is therefore of great interest.
Another issue, which is of both theoretical and practical interest, is the development of further assessment methods for probabilistic forecasts of multivariate quantities which go

beyond those presented in [16]. Unfortunately, the latter are not appropriate to handle problems of sufficiently high dimensions, as may be the case in many applications ECC is confronted with, and hence, there is a need of further research in this field. In our case study, for example, the MRH is likely to be inadequate to assess the calibration of the ECC technique and does not yield plausible and satisfactory results if we consider all five weather variables at all three locations for a fixed look-ahead time simultaneously, that is, if we operate with fifteen dimensions.
Furthermore, the ECC tool in its present version introduced in this thesis is applicable to exchangeable ensemble members only. Hence, the development of a modified ECC technique for the case of non-exchangeability provides material for further work.
In addition, our state-of-the-art ECC approach might not be appropriate if we are faced with very high forecast dimensions, that is, with a great amount of weather quantities, locations and prediction horizons, while the number M of ensemble members is rather small. In those high-dimensional cases, it might be unsatisfactory to sample only M times from the corresponding univariate predictive distributions, and hence, we would have a problem if the desired ensemble size after the ensemble postprocessing exceeded that of the raw ensemble, which requires the development of an alternative procedure being able to handle this shortcoming.

All in all, ECC represents a well-performing multivariate ensemble postprocessing method, which mainly impresses with its simplicity and clarity. However, there are still various proposals for contributions in order to improve, modify and supplement, respectively, the ECC technique as well as related aspects and topics, both from a theoretical and a practical point of view.

Bibliography

[1] Bao, L., Gneiting, T., Grimit, E. P., Guttorp, P. and Raftery, A. E. (2010). Bias correction and Bayesian model averaging for ensemble forecasts of surface wind direction. Monthly Weather Review, 138, 1811-1821.

[2] Berrocal, V. J., Raftery, A. E. and Gneiting, T. (2007). Combining spatial statistical and ensemble information for probabilistic weather forecasting. Monthly Weather Review, 135, 1386-1402.

[3] Berrocal, V. J., Raftery, A. E. and Gneiting, T. (2008). Probabilistic quantitative precipitation field forecasting using a two-stage spatial model. Annals of Applied Statistics, 2, 1170-1193.

[4] Bremnes, J. B. (2004). Probabilistic forecasts of precipitation in terms of quantiles using NWP model output. Monthly Weather Review, 132, 338-347.

[5] Bremnes, J. B. (2007). Improved calibration of precipitation forecasts using ensemble techniques. Part 2: Statistical calibration methods. Norwegian Meteorological Institute, Technical Report no. 04/2007, available online at `http://met.no/Forskning/Publikasjoner/Publikasjoner_2007/filestore/report04_2007.pdf`.

[6] Darsow, W. F., Nguyen, B. and Olsen, E. T. (1992). Copulas and Markov processes. Illinois Journal of Mathematics, 36, 600-642.

[7] Edwards, D. (2000). Introduction to Graphical Modelling, 2nd edition. Springer, New York.

[8] Embrechts, P., McNeil, A. and Straumann, D. (2002). Correlation and dependence in risk management: properties and pitfalls. In Risk Management: Value at Risk and Beyond, ed. M. A. H. Dempster. Cambridge University Press, Cambridge, 176-223.

[9] Embrechts, P., Lindskog, F. and McNeil, A. (2003). Modelling dependence with copulas and applications to risk management. In Handbook of Heavy Tailed Distributions in Finance, ed. S. T. Rachev. Elsevier, Amsterdam, 329-384.

[10] Gel, Y., Raftery, A. E. and Gneiting, T. (2004). Calibrated probabilistic mesoscale weather field forecasting: The geostatistical output perturbation (GOP) method (with discussion). Journal of the American Statistical Association, 99, 575-587.

[11] Genest, C., Quesada Molina, J. J. and Rodríguez Lallena, J. A. (1995). De l'impossibilité de construire des lois à marges multidimensionelles données à partir de copules. Comptes rendus de l'Académie des sciences, Paris, série I, mathématique, 320, 723-726.

[12] Gneiting, T. and Raftery, A. E. (2005). Weather forecasting with ensemble methods. Science, 310, 248-249.

[13] Gneiting, T., Raftery, A. E., Westveld, A. H. and Goldman, T. (2005). Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation. Monthly Weather Review, 133, 1098-1118.

[14] Gneiting, T., Balabdaoui, F. and Raftery, A. E. (2007). Probabilistic forecasts, calibration and sharpness. Journal of the Royal Statistical Society Series B: Statistical Methodology, 69, 243-268.

[15] Gneiting, T. and Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation. Journal of the American Statistical Association, 102, 359-378.

[16] Gneiting, T., Stanberry, L. I., Grimit, E. P., Held, L. and Johnson, N. A. (2008). Assessing probabilistic forecasts of multivariate quantities, with applications to ensemble predictions of surface winds (with discussion and rejoinder). Test, 17, 211-264.

[17] Gneiting, T. (2008). Editorial: Probabilistic forecasting. Journal of the Royal Statistical Society Series A: Statistics in Society, 171, 319-321.

[18] Hamill, T. M., Whitaker, J. S. and Wei, X. (2004). Ensemble reforecasting: Improving medium-range forecast skill using retrospective forecasts. Monthly Weather Review, 132, 1434-1447.

[19] Hamill, T. M., Hagedorn, R. and Whitaker, J. S. (2008). Probabilistic forecast calibration using ECMWF and GFS ensemble reforecasts. Part II: Precipitation. Monthly Weather Review, 136, 2620-2632.

[20] Joe, H. (1990). Multivariate concordance. Journal of Multivariate Analysis, 35, 12-30.

[21] Johnson, R. A. and Wichern, D. W. (1992). Applied Multivariate Statistical Analysis, 3rd edition. Prentice Hall, Englewood Cliffs, New Jersey.

[22] Judd, K., Smith, L. A. and Weisheimer, A. (2007). How good is an ensemble at capturing truth? Using bounding boxes for forecast evaluation. Quarterly Journal of the Royal Meteorological Society, 133, 1309-1325.

[23] Kimberling, C. H. (1974). A probabilistic interpretation of complete monotonicity. Aequationes Mathematicae, 10, 152-164.

[24] Kolesárová, A., Mesiar, R., Mordelová, J. and Sempi, C. (2006). Discrete copulas. IEEE Transactions on Fuzzy Systems, 14, 698-705.

[25] Krzysztofowicz, R. and Maranzano, C. J. (2006). Bayesian processor of output: Probability of precipitation occurrence. Research paper RK-0601, University of Virginia, available online at http://test.faculty.virginia.edu/rk/BPO%20For%20Probability%20of%20Precipitation%20Occurrence.pdf.

[26] Krzysztofowicz, R. and Maranzano, C. J. (2006). Bayesian processor of output: Probabilistic quantitative precipitation forecast. Research paper RK-0602, University of Virginia, available online at http://test.faculty.virginia.edu/rk/BPO%20For%20Probabilistic%20Quantitative%20Precipitation.pdf.

[27] Krzysztofowicz, R. (2007). New statistical techniques for probabilistic weather forecasting — stage II. Project summary, available online at http://faculty.virginia.edu/rk/Project%20Summary%202.pdf.

[28] Krzysztofowicz, R. and Toth, Z. (2007). Bayesian processor of ensemble for precipitation forecasting: A development plan. Abstract, American Geophysical Union, fall meeting 2007, available online at http://adsabs.harvard.edu/abs/2006AGUSM.H31B..02T.

[29] Krzysztofowicz, R. (2008). Bayesian processor of ensemble (BPE): Concept and development. Conference slides 19th Conference on Probability and Statistics in the Atmosphere Sciences, New Orleans, Louisiana, available online at http://faculty.virginia.edu/rk/Bayesian%20Processor%20of%20Ensemble%20(BPE)%20Concept%20and%20Development.pdf.

[30] Krzysztofowicz, R. and Toth, Z. (2008). Bayesian processor of ensemble (BPE): Concept and implementation. Workshop slides 4th NCEP/NWS Ensemble User Workshop, Laurel, Maryland, available online at http://www.emc.ncep.noaa.gov/gmb/ens/ens2008/Krzysztofowicz_Presentation_Web.pdf.

[31] Krzysztofowicz, R. (2010). Decision criteria, data fusion and prediction calibration: A Bayesian approach. Research paper RK-0901, October 2009, revised March 2010, available online at http://faculty.virginia.edu/rk/DecisionCriteria.pdf.

[32] Lauritzen, S. L. (1996). Graphical Models. Clarendon Press, Oxford.

[33] Matheson, J. E. and Winkler, R. L. (1976). Scoring rules for continuous probability distributions. Management Science, 22, 1087-1096.

[34] Mayor, G., Suñer, J. and Torrens, J. (2005). Copula-like operations on finite settings. IEEE Transactions on Fuzzy Systems, 13, 468-477.

[35] Mayor, G., Suñer, J. and Torrens, J. (2007). Sklar's theorem in finite settings. IEEE Transactions on Fuzzy Systems, 15, 410-416.

[36] McNeil, A. and Nešlehová, J. (2009). Multivariate Archimedean copulas, d-monotone functions and ℓ_1-norm symmetric distributions. The Annals of Statistics, 37, 3059-3097.

[37] Mesiar, R. (2005). Discrete copulas — what they are. In Joint EUSFLAT-LFA 2005, Conference Proceedings, eds. E. Montseny and P. Sobrevilla. Universitat Politècnica de Catalunya, Barcelona, 927-930.

[38] Mesiar, R. and Komorníková, M. (2009). Copulas: An approach how to model the dependence structure of random vectors. Acta Polytechnica Hungarica, 6, 5-19.

[39] Nelsen, R. B. (2002). Concordance and copulas: A survey. In Distributions with Given Marginals and Statistical Modelling, eds. C. M. Cuadras, J Fortiana and J. A. Rodríguez Lallena. Kluwer Academic Publishers, Dordrecht, 169-178.

[40] Nelsen, R. B. (2006). An Introduction to Copulas, 2nd edition. Springer, New York.

[41] Nešlehová, J. (2007). On rank correlation measures for non-continuous random variables. Journal of Multivariate Analysis, 98, 544-567.

[42] Palmer, T. N. (2002). The economic value of ensemble forecasts as a tool for risk assessment: From days to decades. Quarterly Journal of the Royal Meteorological Society, 128, 747-774.

[43] Quesada Molina, J. J. and Rodríguez Lallena, J. A. (1994). Some advances in the study of the compatibility of three bivariate copulas. Journal of the Italian Statistical Society, 3, 397-417.

[44] Quesada Molina, J. J., Rodríguez Lallena, J. A. and Úbeda Flores, M. (2003). What are copulas? Monografías del Seminario Matematico García de Galdeano, 27, 499-506.

[45] R Development Core Team (2011). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL `http://www.R-project.org/`.

[46] Raftery, A. E., Gneiting, T., Balabdaoui, F. and Polakowski, M. (2005). Using Bayesian model averaging to calibrate forecast ensembles. Monthly Weather Review, 133, 1155-1174.

[47] Scarsini, M. (1984). On measures of concordance. Stochastica, 8, 201-218.

[48] Schmid, F. and Schmidt, R. (2007). Multivariate extensions of Spearman's rho and related statistics. Statistics & Probability Letters, 77, 407-416.

[49] Schmitz, V. (2003). Copulas and Stochastic Processes. Dissertation, available online at http://sylvester.bth.rwth-aachen.de/dissertationen/2004/198/04_198.pdf.

[50] Schoelzel, C. and Friederichs, P. (2008). Multivariate non-normally distributed random variables in climate research-introduction to the copula approach. Nonlinear Processes in Geophysics, 15, 761-772.

[51] Schweizer, B. and Sklar, A. (1983). Probabilistic Metric Spaces. Elsevier, New York.

[52] Sempi, C. (2011). Copulae: Some mathematical aspects. Applied Stochastic Models in Business and Industry, 27, 37-50.

[53] Sklar, A. (1996). Random variables, distribution functions, and copulas — a personal look backward and forward. In Distributions with Fixed Marginals and Related Topics, eds. L. Rüschendorf, B. Schweizer and M. D. Taylor. IMS Lecture Notes — Monograph Series Number 28, Hayward, California, 1-14.

[54] Sloughter, J. M., Raftery, A. E., Gneiting, T. and Fraley, C. (2007). Probabilistic quantitative precipitation forecasting using Bayesian model averaging. Monthly Weather Review, 135, 3209-3220.

[55] Sloughter, J. M., Gneiting, T. and Raftery, A. E. (2010). Probabilistic forecasts of wind speed using ensembles and Bayesian model averaging. Journal of the American Statistical Association, 105, 25-35.

[56] Smith, L. A. (2001). Disentangling uncertainty and error: On the predictability of nonlinear systems. In Nonlinear Dynamics and Statistics, ed. A. I. Mees. Birkhäuser, Boston, 31-64.

[57] Thorarinsdottir, T. L. and Gneiting, T. (2010). Probabilistic forecasts of wind speed: Ensemble model output statistics using heteroskedastic censored regression. Journal of the Royal Statistical Society Series A: Statistics in Society, 173, 371-388.

[58] Weisheimer, A., Smith, L. A. and Judd, K. (2005). A new view of seasonal forecast skill: Bounding boxes from the DEMETER ensemble forecasts. Tellus A, 57, 265-279.

[59] Westermann Verlag (ed.) (2002). Diercke Weltatlas, 5., aktualisierte Auflage. Westermann, Braunschweig.

[60] Wilks, D. S. and Hamill, T. M. (2007). Comparison of ensemble-MOS methods using GFS reforecasts. Monthly Weather Review, 135, 2379-2390.

[61] Wilks, D. S. (2009). Extending logistic regression to provide full-probability-distribution MOS forecasts. Meteorological Applications, 16, 361-368.

Acknowledgements

I thank both Prof. Dr. Tilmann Gneiting and Dr. Thordis Thorarinsdottir for providing the opportunity to work on the very interesting topic of this thesis, for introducing me to the exciting field of probabilistic forecasting in theory and applications and for the excellent supervision of my work. Moreover, I give hearty thanks to the whole statistics research group for fruitful discussions with valuable suggestions, ideas and hints during our routine group meetings.

Erklärung

Hiermit versichere ich, dass ich meine Arbeit selbständig unter Anleitung verfasst habe, dass ich keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe, und dass ich alle Stellen, die dem Wortlaut oder dem Sinne nach anderen Werken entlehnt sind, durch die Angabe der Quellen als Entlehnungen kenntlich gemacht habe.

Datum Unterschrift