DISSERTATION

submitted

to the

Combined Faculties for the Natural Sciences and for Mathematics

of the

Ruperto-Carola University of Heidelberg, Germany

for the degree of

Doctor of Natural Sciences

put forward by

Diplom-Statistician: Annette Möller

Born in: Hamburg

Oral examination:

Multivariate and spatial ensemble postprocessing methods

Advisor: Prof. Dr. Tilmann Gneiting

Zusammenfassung

In der jüngsten Vergangenheit wurde die Entwicklung von Wettervorhersagen über Ensemble Vorhersagesysteme zum Standard in der Meteorologie. Vorhersage Ensembles werden generiert aus mehreren Durchläufen dynamischer Wettervorhersagemodelle, mit unterschiedlichen Anfangs- sowie Randbedingungen oder Parametrisierungen. Jedoch sind auch die Ensemble Vorhersagen nicht in der Lage, die Vorhersageunsicherheit der numerischen Wettermodelle vollständig zu erfassen. Deshalb weisen Ensembles oft Verzerrungen und Dispersionsfehler auf, sind also mangelhaft kalibriert. Um dieses Problem zu beheben wurden bereits erfolgreich statistische Methoden zur Nachbearbeitung von Ensemblevorhersagen entwickelt. Dennoch sind viele dieser etablierten Verfahren ausgelegt auf die Anwendung auf eine einzelne Wettervariable, an einem festen Ort und für einen festen Vorhersagehorizont. Diese Arbeit präsentiert Erweiterungen der etablierten Nachbearbeitungsmethoden Bayesian model averaging (BMA) und Ensemble model output statistics (EMOS), mit dem Ziel Abhängigkeiten zwischen Wettervariablen sowie räumliche Abhängigkeiten zu erfassen, welche in den ursprünglichen Ensemblevorhersagen implizit vorhanden sind. Zu diesem Zweck wird eine Mehrschrittprozedur vorgeschlagen, welche sowohl für die Modellierung von Abhängigkeiten zwischen Wettervariablen als auch für die Modellierung räumlicher oder zeitlicher Abhängigkeitsstrukturen geeignet ist. Diese Prozedur kombiniert ein univariates Nachbearbeitungsmodell wie BMA oder EMOS mit einer multivariaten Abhängigkeitsstruktur, z.B. mit Hilfe einer Korrelationsmatrix oder der multivariaten Rangstruktur des Vorhersageensembles. Eine auf BMA basierende multivariate Nachbearbeitungsprozedur, welche die Abhängigkeit zwischen Wettervariablen modelliert, wird auf das 8-Mitglieder UWME Ensemble für den Nordwesten der USA angewendet. Das Ergebnis einer entsprechenden Fallstudie ist eine multivariate Vorhersageverteilung mit guter multivariater Kalibrierung und Schärfe. Eine räumliche Variante dieser multivariaten Nachbearbeitungsmethode basierend auf EMOS wird angewendet auf Temperatur Vorhersagen des 50 Mitglieder ECMWF Ensembles in Deutschland. Es wird eine räumlich adaptive Erweiterung von EMOS verwendet, welche von kürzlich entwickelten Methoden für eine schnelle und genaue Bayesianische Schätzung räumlicher Modelle profitieren kann. Die Prozedur führt zu guter univariater und multivariater Kalibrierung und Schärfe. Außerdem können die räumlichen Strukturen der beobachteten Wetterfelder durch die Methode angemessen wiedergegeben werden. Beide hier betrachteten Erweiterungen verbessern Kalibrierung und Schärfe im Vergleich zum ursprünglichen Vorhersageensemble und zu etablierten univariaten Nachbearbeitungsverfahren.

Abstract

In the recent past the state of the art in meteorology has been to produce weather forecasts from ensemble prediction systems. Forecast ensembles are generated from multiple runs of dynamical numerical weather prediction models, each with different initial and boundary conditions or parameterizations of the model. However, ensemble forecasts are not able to catch the full uncertainty of numerical weather predictions and therefore often display biases and dispersion errors and thus are uncalibrated. To account for this problem, statistical postprocessing methods have been developed successfully. However, many state of the art methods are designed for a single weather quantity at a fixed location and for a fixed forecast horizon. This work introduces extensions of two established univariate postprocessing methods, Bayesian model averaging (BMA) and Ensemble model output statistics (EMOS) to recover inter-variable and spatial dependencies from the original ensemble forecasts. For this purpose, a multi-stage procedure is proposed that can be applied for modeling dependence structures between different weather quantities as well as modeling spatial or temporal dependencies. This multi-stage procedure combines the postprocessing of the margins by the application of a univariate method as BMA or EMOS with a multivariate dependence structure, for example via a correlation matrix or via the multivariate rank structure of the original ensemble. The multivariate postprocessing procedure that models inter-variable dependence employs the UWME 8-member forecast ensemble over the North West region of the US and the standard BMA method, resulting in predictive distributions with good multivariate calibration and sharpness. The spatial postprocessing procedure is applied to temperature forecasts of the ECMWF 50-member ensemble over Germany. The procedure employs a spatially adaptive extension of EMOS, utilizing recently proposed methods for fast and accurate Bayesian estimation in a spatial setting. It yields excellent spatial univariate and multivariate calibration and sharpness. Further the method is able to capture the spatial structure of observed weather fields. Both extensions improve calibration and sharpness in comparison to the raw ensemble and to the respective standard univariate postprocessing methods.

Contents

List of Figures III

List of Tables IV

Acknowledgments **V**

1 Introduction **1**

2 Probabilistic Weather Forecasting **5**

2.1 Postprocessing methods for ensemble forecasts 7

2.1.1 Ensemble Model Output Statistics (EMOS) 8

2.1.2 Bayesian Model Averaging (BMA) 11

2.1.3 Geostatistical Output Perturbation Method (GOP) 15

2.1.4 Spatial BMA 17

2.1.5 Geostatistical Model Averaging (GMA) 19

2.1.6 Ensemble Copula Coupling (ECC) 21

2.2 Verification methods 24

2.2.1 Calibration 25

2.2.2 Sharpness 29

2.2.3 Proper scoring rules 30

3 Copulas **35**

3.1 Notation and preliminaries 35

3.2 Copulas and Sklar's Theorem 39

3.3 Copulas and random variables 41

3.4 Gaussian copulas 42
3.5 A Gaussian copula multi-stage procedure 46

4 Multivariate extension of BMA **49**
4.1 Combining BMA and Gaussian copulas 50
4.2 Data 55
4.3 Results 56
4.3.1 Results at Sea-Tac Airport 57
4.3.2 Aggregated results over the Northwest US 61
4.4 Assessing statistical significance 63

5 Spatially adaptive extension of EMOS **65**
5.1 Gaussian fields and Gaussian Markov random fields 66
5.2 Integrated Nested Laplace Approximation (INLA) 71
5.3 The SPDE approach 76
5.4 Markovian EMOS (MEMOS) 83
5.5 Multivariate dependence structures 90
5.6 Data 93
5.7 Mesh configuration 95
5.8 Training period 97
5.9 Results 99
5.9.1 Aggregated univariate results 99
5.9.2 Univariate results at single stations 104
5.9.3 Multivariate results 109
5.9.4 Univariate results for composite quantities 121
5.9.5 Samples from multivariate MEMOS versions 125

6 Discussion **129**

References **135**

List of Figures

2.1 Different shapes of PIT histograms 27

4.1 Estimated joint predictive distribution for January 1, 2008 at the observation station KSEA 59

4.2 Multivariate rank histograms for the station KSEA 60

4.3 Estimated correlation at 60 observation stations in the Northwest US 62

4.4 Multivariate rank histograms over all available observations at 60 stations in the Northwest US 62

4.5 Permutation distribution of energy score differences 64

5.1 Station locations for temperature in Germany along with an example of a mesh used for MEMOS 94

5.2 Mesh configurations with different parameters 96

5.3 CRPS, MAE and RMSE for local EMOS and MEMOS plotted against length of training period 99

5.4 Overall univariate calibration of MEMOS 103

5.5 Univariate rank histograms for MEMOS at the station Borkum 105

5.6 Univariate rank histograms for MEMOS at the station Frankfurt 107

5.7 Univariate rank histograms for MEMOS at the station Baden-Baden 108

5.8 Multivariate rank histograms over four North Sea islands 113

5.9 Multivariate rank histograms over eleven stations at the North Sea coastline 117

5.10 Multivariate rank histograms over three stations in East Germany 120

5.11 Univariate rank histograms for minimum temperature over three stations in East Germany 124

5.12 Samples from the multivariate MEMOS ECC distribution 126

List of Tables

2.1 Ensemble BMA kernel functions for different weather variables 14

4.1 Estimated correlation matrix at the station KSEA 57
4.2 Multivariate predictive performance at the station KSEA 60
4.3 Predictive performance averaged over 60 stations in the Northwest US . . . 63

5.1 Overall univariate predictive performance of MEMOS 102
5.2 Univariate predictive performance of MEMOS at the station Borkum 104
5.3 Univariate predictive performance of MEMOS at the station Frankfurt . . . 106
5.4 Univariate predictive performance of MEMOS at the station Baden-Baden . 107
5.5 Pairwise distances in km between four North Sea islands 110
5.6 Energy score over four North Sea islands 112
5.7 Determinant sharpness over four North Sea islands 112
5.8 Pairwise distances in km between eleven stations along the North Sea coastline 114
5.9 Energy score over eleven stations along the North Sea coastline 115
5.10 Determinant sharpness over eleven stations along the North Sea coastline . 116
5.11 Pairwise distances in km between three stations in East Germany 118
5.12 Energy score over three cities in East Germany 118
5.13 Determinant sharpness over three cities in East Germany 119
5.14 CRPS for minimum temperature over three stations in East Germany 122
5.15 MAE for minimum temperature over three stations in East Germany 122
5.16 RMSE for minimum temperature over three stations in East Germany . . . 123

Acknowledgements

I gratefully acknowledge the financial support by the German Research Foundation (DFG) within the programme "Spatio-/Temporal Graphical Models and Applications in Image Analysis" grant GRK 1653 and the scientific support by the Faculty of Mathematics and Computer Science of the University of Heidelberg.

Further, I would like to thank all the people that advised, supported and encouraged me during my PhD period. Specifically, I would like to acknowledge the assistance of my advisor Tilmann Gneiting, and of my mentors and coauthors Thordis Thorarinsdottir and Alex Lenkoski, providing very helpful discussions and ideas for developing the PhD project. Besides this I thank my colleagues Michael Scheuerer, Roman Schefzik and Kira Feldmann for helpful and inspiring discussions.

Apart from my scientific colleagues and advisors I would like to mention my friends Eva Didden and Anka George for uplifting me and cheering me up in times when I was suffering setbacks or having second thoughts. In particular I would like to thank my husband Jürgen for always backing me up, in scientific as well as as other concerns of real life.

Chapter 1

Introduction

Analysis and prediction of current and future weather conditions has always been of great interest. In this day and age, an accurate forecast of weather becomes increasingly important in many economic and social areas, in the management of safe air and ship traffic, in agriculture and forestry, in efficient generation of energy with sustainable technologies as wind power and solar cells, or in civil protection in case of natural disasters such as floods, droughts, storms, forest fire and others. It is even of interest in the financial sector, for example considering the weather derivative.

During the past decades a change in the practice of weather prediction has been observed. Up to the beginning of the 1990s, weather prediction was a purely deterministic venture. National and international weather research centers run different sophisticated numerical weather prediction (NWP) models. The output of these models are deterministic point forecasts for future weather conditions. However, deterministic point forecasts cannot address uncertainties such as imperfect model formulations or incorrect initial and boundary conditions used to run the model (Leutbecher and Palmer, 2008).

Numerical weather prediction models are still run today, but with a change towards the implementation of (dynamical) ensemble forecast methods. Ensembles of forecasts address prediction uncertainty and allow for probabilistic forecasting. A forecast ensemble is created by multiple runs of the numerical model, with a set of different initial conditions and lateral boundary conditions and/ or with different representations of the mathematical models used to describe the atmosphere. Dynamical ensembles can capture nonlinear uncertainties in forecasts, but they require considerable computational power.

However, ensemble forecasts still have their shortcomings. Often, ensembles exhibit systematic biases as well as dispersion errors. The ensembles therefore lack calibration (Hamill and Colucci, 1997) and are typically underdispersed. To address these shortcomings several statistical methods of postprocessing the ensemble forecasts have been developed (Wilks and Hamill, 2007). Such methods correct for biases and dispersion errors in light of the principle 'maximizing sharpness subject to calibration' proposed by Gneiting et al. (2007). An additional advantage of statistical postprocessing is the fact that the postprocessing models yield full predictive probability distributions for the weather quantities of interest. In many applications, the postprocessed ensemble outperforms the raw ensemble in terms of calibration and sharpness.

There are two main approaches of postprocessing, which are the basis for several extensions. The Ensemble Bayesian model averaging (BMA) proposed by Raftery et al. (2005) estimates the predictive density as a mixture of the individual densities associated with each ensemble member. The Ensemble model output statistics (EMOS) technique introduced in Gneiting et al. (2005) combines all ensemble forecasts in a multivariate regression approach. These methods, especially BMA, have been adapted in various ways to account for the characteristics of different weather quantities.

Many of the postprocessing techniques based on BMA and EMOS, however, are designed for a single weather quantity and a fixed location. They do not take into account dependencies between weather quantities or spatial dependencies between observation locations. For example for accurately forecasting composite quantities such as minima, maxima or averages, it is important to incorporate spatial structures and correlations among the location-specific forecast uncertainties.

Several approaches have already been developed on the basis of BMA or EMOS that account for spatial dependency structures in different ways. The main approaches are the Geostatistical model averaging (GMA) introduced by Kleiber et al. (2011a) and methods based on the Geostatistical output perturbation (GOP) method (Gel et al., 2004). While the first approach estimates spatially adaptive model parameters from all ensemble members, the second one aims at sampling spatially consistent and jointly calibrated samples of weather fields by employing only a single ensemble member or point forecast.

This work contributes to the postprocessing literature in two ways. On the one hand, it introduces a new way of multivariate postprocessing where dependencies between weather quantities are incorporated in the predictive distribution by combining existing variants of BMA for different types of weather quantities with a multivariate dependence structure induced by a Gaussian copula. This method yields a joint predictive distribution for several weather

quantities with good multivariate (as well as univariate) calibration properties. The presented approach was originally designed to employ a Gaussian copula for constructing the multivariate dependence structure. Nonetheless, a generalization to other types of dependence structures is straightforward and already utilized in the spatial approach discussed next. In the postprocessing literature multivariate postprocessing approaches are rarely found. Although Schoelzel and Friederichs (2008) consider the estimation of bivariate copulas to model the distribution of two weather quantities jointly (they apply this method to weather quantities as well as to other environmental variables), they are not employing these copula approaches in a context of ensemble postprocessing, only for observations of weather quantities. The work published by Möller et al. (2013) that is presented here, explicitly develops a multivariate ensemble postprocessing procedure to jointly model several weather quantities that closes a gap in the postprocessing literature.

On the other hand, this thesis develops a spatial postprocessing method by combining the advantages of GMA and GOP based methods. This postprocessing approach, called Markovian EMOS (MEMOS), extends the standard EMOS model to have spatially adaptive model parameters by assuming spatial Gaussian fields on them. The method additionally allows for producing samples from a spatial joint distribution by combining the basic MEMOS method with certain types of multivariate dependence structures in a multi-stage procedure. A further advantage is the possibility to utilize a recently developed methodology that allows to compute an explicit Gaussian Markov random field representation of a Gaussian field with Matérn covariance function (Lindgren et al., 2011). A direct link between the two types of random fields can be established via the solution of a stochastic partial differential equation (SPDE). This methodology yields a considerable computational benefit when estimating the parameters of the postprocessing model. The resulting postprocessing model was called Markovian EMOS to honor the specific markovian structure of the underlying GMRF representation.

The remainder of this work is organized as follows. Chapter 2 gives an overview on probabilistic weather forecasting in general as well as on several established univariate postprocessing methods. Chapters 4 and 5 present the multivariate and spatial extensions of the BMA and EMOS methods. For the spatial version the theory of Gaussian Fields, Gaussian Markov random fields and the link between both concepts via the solution of an SPDE is introduced. The respective chapters review the theory behind the developed methods and present the data employed for each of the approaches along with case studies analyzing their predictive performance.

In Chapter 3 some general theory about copulas is introduced, as Gaussian copula methods are employed in Chapter 4 to set up a multivariate dependence structure.

Chapter 6 briefly summarizes the key results of this work and discusses alternative approaches as well as an outlook on future research plans extending the obtained methods in various directions.

Chapter 2

Probabilistic Weather Forecasting

Current practise in predicting future weather is the use of numerical weather prediction (NWP) models. These are deterministic numerical simulation models representing the physics of the atmosphere. The NWP models are based on several dynamical partial differential equations in different variables describing the evolution of the states of atmosphere. To obtain deterministic forecasts for future atmospheric states the involved equations are discretized on a grid with a certain resolution (differing with the type of NWP model) and run forward in time. A set of initial and boundary conditions describing the current state of the atmosphere at time step t is assimilated from observations at time step t. The information is then used to run forward the NWP model in time and obtains predictions of future atmospheric states at time steps $t + l$, $l \geq 1$.

There are different types of NWP models that are run operationally, among them global models and mesoscale (limited-area) models. Different global models are run in real time all over the world, such as the European Centre for Medium-Range Weather Forecasts (ECMWF) model in Europe, the National Centers for Environmental Prediction (NCEP) model in the United States or the Meteorological Service of Canada (MSC) in Canada. Global models typically have a horizontal resolution of 36 - 100 km. The mesoscale models have a much higher resolution, typically about 2 - 36 km horizontally. They usually employ initial and boundary conditions provided by one or several global models.

An NWP model issues a single deterministic forecast and is therefore not capable of assessing the forecast uncertainty. However, there are major sources of uncertainty present in NWP models, such as an incomplete network of observations, measurement errors or incorrect model formulations due to incomplete knowledge of all physical processes in the

atmosphere.

The atmosphere is a chaotic system, very small perturbations of the current state can already lead to an increasingly large change in the evolution of the atmosphere so that after some time the development of the original state and the perturbed state are completely different (Lorenz, 1963). The chaotic nature of the atmosphere leads to significant forecast errors as minute errors in the initial conditions can grow exponentially during the integration process of the NWP model. This makes a single deterministic forecast useless when it comes to the assessment of the forecast uncertainty.

Ensemble methods are one manner of accounting for this problem. They lead to considerable improvements in the forecast skill. This was explored elaborately for medium-range forecasts (2-10 days) (Toth and Kalnay, 1993; Molteni et al., 1996) and there is also research concerning the application of ensembles to short-range forecasts (0-48h) (Eckel and Mass, 2005; Hamill and Colucci, 1997).

Ensemble forecasting adopts the idea of predicting probabilities for future weather events (Leutbecher and Palmer, 2008). Pushing forward the ensemble forecasting in daily weather prediction lead to a change from purely deterministic weather forecasts to a more probabilistic approach. The idea of probabilistic forecasting through ensemble forecasts is carried out by using a set of different initial and boundary conditions, all consistent with observations and observation errors, different model physics or model perturbations or even different numerical models. An ensemble of forecasts is then generated by running an NWP model with a varied set of initial conditions and/or perturbations or by employing a different unique numerical model for each run. There are several types of ensembles, according to the way their members are generated, see e.g. Eckel and Mass (2005):

Multi-analysis ensembles are produced by employing a single numerical model and run it multiple times, each time with a different set of initial conditions. Multi-model ensembles are obtained by employing a separate numerical model for each ensemble member, but run these multiple models with a single set of initial conditions. Multi-model multi-analysis ensembles are a combination of both types described above. Multiple numerical models are run, each with multiple sets of initial conditions, so each ensemble member is a specific combination of a certain numerical model with a certain set of initial conditions.

An example for a multi-analysis mesoscale ensemble system is the University of Washington Mesoscale Ensemble (UWME), operating at lead times up to three days (Grimit and Mass, 2002; Eckel and Mass, 2005). It comprises of eight members that are obtained by using eight analyses from different operational forecast centers as initial conditions in the

fifth-generation Pennsylvania State University - National Center for Atmospheric Research Mesoscale Model (PSU-NCAR MM5). The PSU-NCAR MM5 model is succeeded by the WRF (Weather Research and Forecasting) model presently. In Section 4 some more information on the UWME ensemble is given, as the WRF UWME ensemble is employed in the case study for the multivariate postprocessing extension.

A forecast ensemble allows to combine the individual members for example by using the ensemble mean as a single deterministic forecast, as well as to estimate the forecast uncertainty by considering the ensemble variance or the root mean square error. On average, the ensemble mean outperforms each of the individual ensemble members (Grimit and Mass, 2002).

For a unimodal predictive distribution, variations in the width of the distribution from forecast to forecast can be directly linked to the predictive skill of the mean forecast. The spread of the ensemble measures the width of the predictive distribution and therefore provides a tool to assess the predictive skill of the ensemble mean. Specifically, Whitaker and Loughe (1998) found a positive correlation between the ensemble spread, which is known a priori, and the forecast errors, only known a posteriori. This phenomenon is called spread-error correlation or spread-skill relationship.

2.1 Postprocessing methods for ensemble forecasts

Ensembles allow for probabilistic forecasts of continuous weather variables, such as sea level pressure or surface temperature. Ideally, a probabilistic forecast takes the form of a predictive cumulative distribution function (CDF) or a predictive probability density function (PDF). However, ensemble forecast systems are finite, they typically employ 5 up to 50 members. Hence, the raw ensemble output does not provide full predictive PDFs for continuous quantities, and some form of postprocessing is required.

Besides this, there are several additional challenges in the development of appropriate statistical postprocessing methods. The current NWP systems typically show systematic biases (forecast errors) and the ensembles lack calibration, for details see Section 2.2.1, they are very often underdispersed (Hamill and Colucci, 1997), as the ensembles only capture some of the uncertainties of numerical weather forecasting (Raftery et al., 2005).

Many different postprocessing methods are available these days, most of them can be modified to be suited for various weather quantities. There are two main approaches in the

existing postprocessing literature. The first type of methods is based on the ensemble BMA approach, where the predictive distribution is constructed as a mixture of individual kernel densities assigned to each forecast member. Each kernel density depends only on a specific forecast member. The weights of the mixture components are obtained from the predictive skill of the respective ensemble member. The second type of methods is based on the EMOS approach, where the predictive distribution is obtained via a multiple linear regression of the observations on all ensemble members.

These two basic postprocessing approaches have been extended and modified in various ways to account for the need of the applications to other than normally distributed weather quantities as well as for other requirements as taking into account dependency structures.

The following sections give an overview on the most important methods, that are frequently used, without any claim to be complete.

2.1.1 Ensemble Model Output Statistics (EMOS)

The ensemble model output statistics (EMOS) methodology was introduced by Gneiting et al. (2005). It addresses bias and dispersion errors of the raw forecast ensemble. The EMOS technique is motivated by the multiple linear regression model called model output statistics (MOS) in the context of applying linear regression equations to output of numerical weather prediction models (Glahn and Lowry, 1972; Klein and Glahn, 1974; Wilks, 2006, 2009). It has enjoyed popularity in the application to deterministic-style and probability of precipitation forecasts. The general idea of the MOS method is to combine the output of the numerical weather prediction models (numerical deterministic-style forecasts for different variables) with real observations made at observation stations to get improved forecasts at the considered stations. By taking into account the observations as well, the numerical forecasts can be verified or adapted to the situation at a specific station, after having interpolated the numerical forecast from the model grid to the respective station location. The regression model uses those parameters of the numerical model output that have an effect on a weather quantity of interest, like e.g. surface temperature or sea level pressure.

The EMOS technique is typically designed for the application to a univariate weather quantity, a fixed location and for a fixed forecast horizon. The basic method is not taking into account dependences with other weather quantities nor is it able to capture spatial or temporal dependencies. Specifically, let $x_1, \ldots, x_m$ denote a forecast ensemble of size m for a univariate quantity Y, where all individual forecast members are distinguishable. Using

the ensemble members as predictors in the multiple linear regression model this leads to the regression equation

$$Y = a + b_1 x_1 + \ldots + b_m x_m + \varepsilon, \tag{2.1}$$

where $a, b_1, \ldots, b_m$ are regression coefficients and ε is an error term with variance σ^2 and mean 0.

Equation (2.1) is defined for each considered observation case $i = 1, \ldots, N$ and the error terms ε_i are assumed to be independent of $\varepsilon_{i'}$ for all $i \neq i'$. For convenience, the index i for the observations is omitted.

The model (2.1) can easily be adapted for indistinguishable forecast ensembles, as e.g. the ECMWF ensemble utilized in Chapter 5, by assuming all multiplicative bias-correction parameters to be equal, that is, $b_k = b$, $k = 1, \ldots, m$.

Gneiting et al. (2005) propose this approach to improve the predictive performance of the ensemble with the additional benefit that the postprocessing method yields full predictive PDFs from a forecast ensemble for a continuous weather quantity. When assuming the error term to be normally distributed, that is $\varepsilon \sim N(0, \sigma^2)$, this leads to the following distribution of the quantity Y given the m ensemble members:

$$Y|x_1, \ldots, x_m \sim \mathrm{N}\Big(a + \sum_{k=1}^{m} b_k x_k, \sigma^2\Big). \tag{2.2}$$

This conditional distribution allows to obtain estimated PDFs and CDFs from the regression Equation (2.1) in a straightforward way. The EMOS approach corrects for model biases and accounts for dispersion errors by performing a simple bias correction through the linear combination of the ensemble members with the b_k as multiplicative coefficients and a as additive coefficient. However, with the assumption of a constant model variance the spread of the original ensemble is not taken into account in the model, and so it ignores the spread-skill relationship (Whitaker and Loughe, 1998; Barker, 1991). Gneiting et al. (2005) propose to model the variance of the error term as a linear function of the ensemble spread:
$\mathrm{Var}(\varepsilon) = c + dS^2$, where S^2 is the ensemble variance and c and d are nonnegative coefficients. In this case the predictive distribution of the quantity Y given the ensemble members is given as

$$Y|x_1, \ldots, x_m \sim \mathrm{N}\Big(a + \sum_{k=1}^{m} b_k x_k, c + dS^2\Big) \tag{2.3}$$

In the above model formulation the regression parameters b_k can be positive and negative. In applications positive as well as negative regression coefficients may occur. Negative co-

efficients are typically caused by collinearities of the ensemble member forecasts, but they are hard to interpret. To avoid this problem Gneiting et al. (2005) propose an alternative implementation of the EMOS technique, which constrains the coefficients $b_1, \ldots, b_m$ to be nonnegative. This variant of the EMOS technique is called EMOS$^+$. For details on the implementation see Gneiting et al. (2005).

The predictive mean of the EMOS model is a bias-corrected weighted average of the ensemble members and provides a deterministic forecast. The coefficients $b_1, \ldots, b_m$ reflect the overall performance of the individual ensemble members over the training period as well as correlations between ensemble members. The coefficients c and d reflect the ensemble spread and the performance of the ensemble mean forecast. All else being equal, larger values of d suggest a more pronounced spread-error relationship, while d will be estimated very small if spread and error are independent.

Estimation of the EMOS model parameters can be conducted with maximum likelihood methods (ML) from a set of training data. However, Gneiting et al. (2005) found that predictive PDFs estimated with ML methods tend to be overdispersive and to have wide prediction intervals with higher than nominal coverage. Therefore, they suggest an alternative approach. As the log-likelihood is essentially the negative of the ignorance score, they propose to choose a scoring rule of interest, express this score as a function of the coefficients for the training data and optimize this function with respect to the coefficient values. They argue that the continuous ranked probability score (CRPS, for definition and details see Section 2.2) is the most appropriate score. Due to the choice of the CRPS as reference score they call their estimation technique *minimum CRPS estimation*. This strategy obtains the values of $a, b_1, \ldots, b_m, c, d$ in the EMOS model that minimize the mean CRPS for a given set of training data.

In case of the EMOS model (2.3) with normally distributed error terms ε_i for each observation case i, it is possible to express the CRPS as an analytic function of the coefficients in closed form, as a sum over all observation cases in the training data:

$$\Gamma(a, b_1, \ldots, b_m, c, d) = \frac{1}{n} \sum_{i=1}^{n} (c + dS_i^1)^{1/2} \left[Z_i \left(2\Phi(Z_i) - 1\right) + 2\phi(Z_i) - \frac{1}{\sqrt{\pi}} \right], \tag{2.4}$$

where

$$Z_i = \frac{Y_i - (a + b_1 x_{i1} + \cdots + b_m x_{im})}{(c + dS_i^2)^{1/2}}$$

is a standardized forecast error, ϕ and Φ denote the PDF and the CDF of a $\mathrm{N}(0, 1)$ distribution, and the sum is taken over the training data. The coefficients that minimize (2.4) have to

be determined numerically.

To fit the nonnegative regression coefficients in the EMOS$^+$ approach the authors describe a stepwise procedure. They first estimate the full unconstrained EMOS model, then set all those coefficients to zero that have negative estimates and finally estimate the minimum CRPS estimators in this new model, where the ensemble variance has been recomputed as well. This procedure is iterated until all coefficients have nonnegative estimates.

Gneiting et al. (2005) applied the EMOS technique to surface temperature and sea level pressure. For both quantities it is reasonable to assume a normal distribution. It is possible to adapt the EMOS technique to continuous quantities with other distributions as well. However, the original EMOS model does not apply directly to non-negative weather quantities such as wind speed. For this case Thorarinsdottir and Gneiting (2010) propose an adaptation of the EMOS$^+$ model, employing truncated normal distributions with a cut-off at zero. Thorarinsdottir and Gneiting (2010) employ a heteroscedastic censored (tobit) regression model to estimate the model parameters. The truncated normal distribution is conditional on the forecast ensemble, while the unconditional (marginal) distribution of wind speed is often modeled with Weibull densities.

There are two different variants of estimating the EMOS parameters. The first possibility is to estimate only one set of parameters for all observation locations simultaneously for each given day. This is called global or regional EMOS. The second possibility is to estimate a separate set of parameters for each observation location on a given day. This procedure is called local EMOS.

While local EMOS leads to an improved local calibration as the forecast errors are estimated locally at each station, the resulting predictive weather fields are not necessarily spatially consistent. Research has been conducted that aimed at modifying the basic EMOS approach to account for spatial dependence structures. Similar to the spatial BMA approach described in Section 2.1.4 that combines the standard BMA method (Section 2.1.2) with the GOP approach (Section 2.1.3), a spatially adaptive extension of EMOS called spatial EMOS was developed in Feldmann (2012).

2.1.2 Bayesian Model Averaging (BMA)

In the EMOS approach the m ensemble forecast members are combined in a linear regression model $Y = a + b_1 x_1 + \cdots + b_m x_m + \varepsilon$, resulting in a single model formulation, where it

remains to specify the coefficients. Alternative models are not taken into account.

However, conditioning on a single selected model ignores model uncertainty and leads to underestimation of uncertainty when making inference. Raftery et al. (1997) and Hoeting et al. (1999) among others discussed this problem. A standard Bayesian approach to this problem is averaging over all possible models. Such a technique was originally developed by Leamer (1978) as a method to combine predictions and inferences from multiple statistical models. Let $\mathcal{M} = \{M_1, \ldots, M_k\}$ denote the set of all models taken into consideration and Δ the quantity of interest. Then the posterior distribution of Δ given the data D reads

$$P(\Delta|D) = \sum_{i=1}^{k} P(\Delta|M_i, D)\ P(M_i|D). \tag{2.5}$$

The sum in Equation (2.5) is an average of the posterior distributions of Δ under each model M_i, weighted by their posterior model probabilities. Raftery et al. (1997) call this procedure Bayesian Model Averaging (BMA). The BMA approach enjoys high popularity in the social and health sciences, where it is frequently applied.

Averaging over all possible models in the above described fashion improves predictive ability. The implementation of BMA, however, can be difficult in specific applications as the number of terms in (2.5) can be large. To get around this problem, Raftery et al. (1997) propose to reduce the set of possible models by some selection algorithm and only average over the reduced set. In case studies Raftery et al. (1997) showed that their reduced model averaging approach provides better predictive performance than any single model that might have been reasonably selected.

Several years later Raftery et al. (2005) extended the original BMA technique to statistical postprocessing of forecast ensembles. In this context the models M_k are not referring to different statistical models, but they are implicitly identified with the different runs of the dynamical numerical model(s) producing the ensemble members. The ensemble BMA method can be seen as a kernel dressing approach where each ensemble member x_k is associated with a kernel density $g_k(y|x_k)$. EMOS on the contrary assigns each member its own bias-correction parameter, but not its own density function.

Like EMOS, BMA is designed for a univariate weather quantity, for fixed locations and a fixed forecast horizon. In its basic form it cannot capture spatial or temporal dependencies. The ensemble BMA predictive density, as proposed for example in Raftery et al. (2005), is given by a mixture of individual kernel densities,

$$f(y|x_1, \ldots, x_m) = \sum_{k=1}^{m} \omega_k\, g_k(y|x_k), \tag{2.6}$$

where the weights ω_k are assumed to be non-negative with $\sum_{k=1}^{m} \omega_k = 1$. The weights can be interpreted as the posterior probability of forecast x_k being the best one, based on the performance of this member in the training period.

In the BMA model, bias-correction is achieved by a linear transformation $b_{0k} + b_{1k}x_k$ of each ensemble member. The linear transformation is linked to the mean of the kernel density g_k. The parameters b_{0k} are additive and the b_{1k} multiplicative bias correction terms, estimated member specifically. If all kernel densities g_k are chosen to be Gaussian, the predictive distribution of Y given a single ensemble member has the form

$$Y|x_k \sim \mathrm{N}(b_{0k} + b_{1k}x_k, \sigma^2).$$

The type of distribution assumed for the kernel g_k is not a feature estimated within the BMA procedure. It needs to be specified by the user and the choice depends on the weather variable of interest. Raftery et al. (2005) propose the application of BMA to surface temperature and sea level pressure. They apply Gaussian kernels in both cases. Sloughter et al. (2010) consider the postprocessing of wind speed forecasts and apply gamma kernels.

Sloughter et al. (2007) develop a BMA variant for precipitation amount, utilizing a discrete-continuous mixture distribution to account for the large number of zero observations, while precipitation observations in general are non-negative. The kernel density g_k is defined in two parts. One part models the probability of zero precipitation with a (discrete) point mass at zero, the other part describes the distribution of precipitation amount given that it is non-zero. As in the latter case the distribution can be assumed to be skewed, Sloughter et al. (2007) employ a gamma kernel. The gamma distribution $\Gamma(\alpha, \beta)$ is parameterized in terms of a shape parameter α and a scale parameter β. Its mean is obtained as $\alpha\beta$ and its variance as $\alpha\beta^2$. Sloughter et al. (2007) found that employing the cube root of the precipitation amount yields a more appropriate model fit than a model in terms of the original amounts. They introduce the kernels g_k in the overall BMA model (2.6) for the weather quantity y describing the cube root of precipitation amount as

$$g_k(y|x_k) = P(Y = 0|x_k)\ \mathbb{I}_{\{y=0\}} + P(Y > 0|x_k)\ h_k(y|x_k)\ \mathbb{I}_{\{y>0\}}.$$

In this equation, h_k denotes a gamma density in terms of the cube root of precipitation amount and $\mathbb{I}_{\{y\in A\}}$ the indicator function with $\mathbb{I}_{\{y\in A\}} = 1$ if $y \in A$ for a desired set A and $\mathbb{I}_{\{y\in A\}} = 0$ for $y \notin A$.

The probability of zero precipitation is modeled with a logistic regression approach where the predictor variable is defined as cube root of the original forecasts

$$P(Y = 0|x_k) = \frac{\exp(a_{0k} + a_{1k}x_k^{1/3} + a_{2k}\delta_k)}{1 + \exp(a_{0k} + a_{1k}x_k^{1/3} + a_{2k}\delta_k)}$$

Variable	Range	Kernel	Mean	Variance
Temperature	$y \in \mathbb{R}$	$\mathrm{N}(\mu_k, \sigma_k^2)$	$b_{0k} + b_{1k}x_k$	σ^2
Pressure	$y \in \mathbb{R}$	$\mathrm{N}(\mu_k, \sigma_k^2)$	$b_{0k} + b_{1k}x_k$	σ^2
Wind speed	$y \in \mathbb{R}_+$	$\Gamma(\alpha_k, \beta_k)$	$b_{0k} + b_{1k}x_k$	$c_0 + c_1 x_k$
Precipitation amount	$y^{1/3} \in \mathbb{R}_+$	$\Gamma(\alpha_k, \beta_k)$	$b_{0k} + b_{1k}x_k^{1/3}$	$c_0 + c_1 x_k$

Table 2.1: The ensemble BMA kernel functions for different weather variables and the associated link functions for mean value and variance

with $\delta_k = 1$ if $x_k = 0$ and $\delta_k = 0$ otherwise. Although the BMA model of Sloughter et al. (2007) is defined for the cube root of the precipitation amount, the resulting probabilistic forecast can easily be expressed in terms of the original amounts.

Table 2.1 is taken from Möller et al. (2013), where the BMA methodology is reviewed. It gives an overview on the four most popular BMA models for the quantities temperature, pressure, wind speed, precipitation amount and their model properties. The estimation procedure for the BMA models summarized in Table 2.1 are implemented in the `R` package `ensembleBMA` (R Development Core Team, 2011; Fraley et al., 2011).

The (member specific) bias correction parameters b_{0k} and b_{1k} specifying the mean of the kernel g_k are estimated with linear regression of the observations on the forecasts from the training data. In case of precipitation amount, the parameters for the probability of zero precipitation part, a_{0k}, a_{1k} and a_{2k}, are estimated by logistic regression. The weights w_k, the variance σ^2, and in case of wind speed and precipitation, the variance parameters c_0 and c_1, are estimated by maximum likelihood from the training data. As the likelihood cannot be analyzed analytically, an EM-algorithm is employed by introducing the unobserved quantities $z_{kst} = 1$ if ensemble member k is the best for observation site s at time point t and $z_{kst} = 0$ otherwise. For each (s, t) only one of the $\{z_{1st}, \ldots, z_{mst}\}$ is equal to 1, all others are zero. The estimate for σ^2 obtained by the EM algorithm can then be refined so that the CRPS is minimized for the training data. For this, a numerical search over a range of values of σ^2 is performed while all other parameters remain fixed. For details on how the estimation is conducted for a specific BMA model, see Raftery et al. (2005), Sloughter et al. (2007) and Sloughter et al. (2010).

Other variants of the ensemble BMA method are available that will not be discussed here, see for example the work by Roquelaure and Bergot (2008), Bao et al. (2010), and Chmielecki and Raftery (2010).

As for the EMOS model, there exists a global and a local BMA version, where on a given day either only one set of parameters is estimated over all locations or a separate set of parameters is provided for each station individually.

2.1.3 Geostatistical Output Perturbation Method (GOP)

Gel et al. (2004) introduced a quite different postprocessing approach than the others described above. Methods like EMOS and BMA are employed to perform postprocessing of an ensemble of forecasts. The forecast ensemble is available from large weather centers running the numerical models and the postprocessing is applied to improve the predictive quality of the ensemble.

In contrast to this, the Geostatistical output perturbation method (GOP) aims at producing probabilistic weather forecasts from a single numerical model, instead of postprocessing an already existing ensemble of forecasts. This method applies to whole weather fields simultaneously rather than just to weather events at individual locations. As BMA and EMOS, GOP is designed for univariate weather quantities and fixed forecast horizons. However, in contrast to these two methods it incorporates spatial dependencies between forecast errors at different locations.

The basic idea behind GOP is to perturb output from the numerical model, instead of going down the traditional road of perturbing model inputs to obtain an ensemble. This basic idea is not necessarily including spatial correlation. As actual error fields show substantial spatial correlation, the idea of GOP is to account for this by modeling errors using a geostatistical model preserving the spatial correlation structure of the weather field. Spatially consistent ensembles of forecasts fields can easily be generated by simulating a desired number of realizations from the resulting spatial random field model. The idea for developing this method originally came from the fact that generating a mesoscale forecast ensemble was typically not feasible for smaller weather organizations that perform mesoscale weather prediction locally without the computing resources of national weather centers.

Gel et al. (2004) consider a weather quantity $Y(s)$ at a spatial point $s \in \mathbb{R}^2$ verifying at a time point t, for a given forecast horizon and $X(s)$ denotes the forecast for this quantity. The focus lies on forecasting the spatial field $\mathbf{Y} = \{Y(s), s \in S\}$ for all points s on a large but finite set of locations S simultaneously using the spatial field of forecasts $\mathbf{X} = \{X(s), s \in S\}$, while the forecast horizon and the time point t remain fixed. The goal is to get jointly calibrated probabilistic forecasts of the whole weather field, rather than just forecasts that are calibrated

at each of the individual locations. In the GOP model of Gel et al. (2004), the finite set S can be distinct model grid points as well as observation locations.

To forecast $Y(s)$, a finite set of variables $Z(s)$ that are related to forecast bias might be included in the model. They can for example be functions of time (e.g. time of year or time of day) or functions of space, like latitude, longitude, altitude, land use type and others. The general form of the model for predicting $Y(s)$ proposed in Gel et al. (2004) is then given as

$$Y(s) = \mathbf{b}_0' Z(s) + (\mathbf{b}_1' Z(s)) X(s) + \varepsilon(s). \tag{2.7}$$

Here $\mathbf{b}_0$ and $\mathbf{b}_1$ are parameter vectors, while $\varepsilon(s)$ is a Gaussian stationary space-time process with mean zero and covariance matrix $\mathbf{\Sigma}$. Furthermore, $\mathbf{b}_0' Z(s)$ is an additive bias-correction term and $\mathbf{b}_1' Z(s)$ a multiplicative one.

For the spatial error field $\varepsilon(s)$, Gel et al. (2004) assume that

$$\frac{1}{2} \operatorname{Var}(\varepsilon(s_i) - \varepsilon(s_j)) = \rho^2 + \tau^2 \Big(1 - \exp\Big(- \frac{||s_i - s_j||}{r} \Big) \Big), \tag{2.8}$$

where $|| \cdot ||$ is the Euclidean norm and $s_i \neq s_j$ are locations in the set S. The parameter ρ^2 is called the nugget effect, describing the measurement error variance of the observations, $\rho^2 + \tau^2$ is the marginal variance of $\varepsilon(s)$ and r is a range parameter, indicating the rate at which the spatial correlations of the errors decay. In terms of a geostatistical model, Equation (2.8) implies that the error field has an exponential variogram (Gelfand et al., 2010; Diggle and Ribeiro Jr., 2010)

$$\gamma(d) = \rho^2 + \tau^2 (1 - \mathrm{e}^{-d/r}), \tag{2.9}$$

where $d = ||s_i - s_j||$ denotes the Euclidean distance between locations s_i and s_j, $i \neq j$, as in Equation (2.8).

Gel et al. (2004) propose a three-stage estimation method that approximates a full maximum likelihood approach to estimate the parameters in (2.7) and (2.8).

The GOP method allows for generating a spatial forecast ensemble of any desired size by simply simulating realizations from the process defined through (2.7) and (2.8). It is implemented in the `R` package `ProbForecastGOP`. For the simulation, the current forecast $X(s)$ and the parameters estimated from past data are needed.

The original GOP method is not belonging to the class of methods with spatially varying parameters. Spatial correlations are introduced by dressing the outputs of numerical weather models with spatially correlated error fields produced by a geostatistical model, where the

parameters are estimated globally. However, the possibility to generate spatially calibrated field forecast ensembles from the output of numerical models is a useful tool.

A disadvantage of the method proposed by Gel et al. (2004) is the fact that the technique does not honor the flow-dependent information of a forecast ensemble, as it only employs a single numerical forecast for the postprocessing (e.g. a single ensemble member). BMA and EMOS on the contrary account for the information of the full ensemble. However, the basic methods are not designed to include spatial correlations between forecast errors in the model. Therefore a natural extension is to combine both types of methods. The next section presents such an extension.

2.1.4 Spatial BMA

Berrocal et al. (2007) introduced the spatial BMA technique, which combines the standard BMA and GOP approaches, thus taking advantage of the benefits of both methods. Spatial BMA resembles the original BMA technique, the predictive PDF is a weighted average of the individual forecast densities with weights reflecting the skill of the forecast members.

However, in the spatial BMA model the forecast PDFs are multivariate densities with covariance structures that reflect the spatial dependency of the real observations. The technique has the ability to generate statistical ensembles of whole weather fields simultaneously, of any desired size, similar to GOP. At the same time spatial BMA reduces to the original BMA at any individual location, while it reduces to the original GOP method if only an ensemble of size $m = 1$ is considered, that is, when a single deterministic forecast is employed.

Berrocal et al. (2007) obtain the spatial BMA forecast members by dressing the weather field forecasts of the ensemble members with simulated spatially correlated error fields, in proportions corresponding to the BMA weights of each individual member.

Let $\mathbf{Y} = \{Y(s), s \in S\}$ denote an observational weather field for a large but finite set of locations S and consider the m ensemble forecast fields $\mathbf{X}_1 = \{x_1(s), s \in S\}, \ldots, \mathbf{X}_m = \{x_m(s), s \in S\}$. Denote by n the dimension of the finite set of locations S. In this setting, $x_1(s), \ldots, x_m(s), s \in S$ are the m (univariate) ensemble forecasts at location s and $Y(s)$ is the random variable describing the observational field at location s, with corresponding realization y_s.

The predictive PDF of the field $\mathbf{Y}$ is modeled as

$$f(\mathbf{Y}|\mathbf{X}_1, \ldots, \mathbf{X}_m) = \sum_{k=1}^{m} w_k \, g_k(\mathbf{Y}|\mathbf{X}_k). \tag{2.10}$$

The parameters are the same as in the original BMA technique with the only difference being that the conditional PDFs are multivariate densities, while the densities are univariate in the original BMA technique. Similar to the univariate BMA version a multivariate normal distribution can be assumed for the kernel densities g_k in case temperature or pressure is considered. The densities g_k are centered at the bias-corrected forecast fields, $b_{0k}\mathbf{1} + b_{1k}\,\mathbf{X}_k$, where $\mathbf{1}$ denotes the vector of ones with dimension equal to the dimension of the field $\mathbf{Y}$.

In case of kernel densities having a multivariate normal distribution, Equation (2.10) indicates that the predictive multivariate distribution of the weather field $\mathbf{Y}$ given one of the m ensemble field forecasts is multivariate normal with the following appearance:

$$\mathbf{Y}|\mathbf{X}_k \sim \mathrm{N}_n(b_{0k}\mathbf{1} + b_{1k}\,\mathbf{X}_k, \boldsymbol{\Sigma}_k^*), \tag{2.11}$$

where $\boldsymbol{\Sigma}_k^*$ is a spatially structured covariance matrix and $\mathrm{N}_n(b_{0k}\mathbf{1} + b_{1k}\,\mathbf{X}_k, \boldsymbol{\Sigma}_k^*)$ denotes the n-variate normal distribution of its arguments. Berrocal et al. (2007) showed that the matrix $\boldsymbol{\Sigma}_k^*$ for member k can be derived from the original BMA model variance σ^2, the GOP variance parameters ρ_k^2, τ_k^2 and the GOP covariance matrix $\boldsymbol{\Sigma}_k$ when applying GOP to member k. The resulting representation of the covariance matrix is the following:

$$\boldsymbol{\Sigma}_k^* = \frac{\sigma^2}{\rho_k^2 + \tau_k^2}\,\boldsymbol{\Sigma}_k. \tag{2.12}$$

The factor $\alpha_k = \frac{\sigma^2}{\rho_k^2+\tau_k^2}$ describes the ratio of the BMA variance to the GOP variance of member k for the errors. It is therefore called deflation factor for member k. The deflation factor stems from the assumption of forecast member k being the best in the ensemble. In the original GOP approach no such assumption is necessary, as only a single forecast is considered.

For the estimation of the full set of parameters of spatial BMA it is necessary to first fit the original BMA model. Given the estimates for the BMA parameters, the GOP model is fitted for each ensemble member separately. By combining the estimates from both procedures, the full set of spatial BMA parameters is obtained.

After parameter estimation it is possible to obtain a spatial BMA ensemble of any desired size by repeatedly choosing a random number $k \in \{1, \ldots, m\}$ and then drawing a sample from (2.11) for the chosen k. This sample is generated by dressing the bias-corrected weather

field forecasts $b_{0k}\mathbf{1} + b_{1k}\,\mathbf{X}_k$ with simulations from the spatially structured conditional error fields defined through the geostatistical model. For details on this sampling procedure see Berrocal et al. (2007).

The method described in Berrocal et al. (2007) was designed for weather quantities with a multivariate normal distribution. In Berrocal et al. (2008) a similar spatial postprocessing procedure was developed for precipitation amount, where the authors adapted the univariate BMA version for precipitation amount proposed by Sloughter et al. (2007).

Recently a similar extension was developed for the basic EMOS model. The method is called spatial EMOS in analogy to spatial BMA. It combines the standard EMOS model with GOP and accounts for extending the specific EMOS variance to a spatially structured covariance matrix. For details on this method see Feldmann (2012).

2.1.5 Geostatistical Model Averaging (GMA)

Kleiber et al. (2011a) modify the global BMA method to obtain a locally adaptive technique, called Geostatistical model averaging (GMA). This method estimates biases and variances at the observation stations and interpolates them to arbitrary points on the model grid with a geostatistical model. Kleiber et al. (2011a) apply the method to temperature forecasts, where the predictive distribution can be assumed to be normal. The authors devoted further research to the GMA model and modified it for application to quantitative precipitation (Kleiber et al., 2011b).

Let $s = s_1, \ldots, s_n \in \mathbb{R}^3$ denote a set of n training locations, $x_{1st}, \ldots, x_{mst}$ an m member forecast ensemble at location s, valid at time point t, and y_{st} the corresponding verifying observation.

Generalizing Raftery et al. (2005), the bias-correction coefficients and the variance are defined to vary not only by member and time point, but also by location. Therefore, GMA is an approach that uses spatially varying parameters. Concerning the bias correction parameters, Kleiber et al. (2011a) use a simplified version of the basic BMA model with only an additive bias-correction term:

$$y_{st} = x_{kst} - a_{ks} + \varepsilon_{st}, \tag{2.13}$$

where a_{ks} denotes an additive bias-correction parameter for member k at site s, x_{kst} the forecast of member k for (s, t) and $\varepsilon_{st} \sim \mathrm{N}(0, \sigma_s^2)$. The error term ε_{st} is assumed to have a spatially varying variance not depending on the time index t.

The GMA predictive density is given by

$$f(y_{st}|x_{1st},\ldots,x_{mst}) = \sum_{k=1}^{m} w_k\, g_k(y_{st}|x_{kst}), \tag{2.14}$$

and the predictive distribution of the quantity y given the kth ensemble member at location s and time point t reads

$$y_{st}|x_{kst} \sim \mathrm{N}(x_{kst} - a_{ks}, c\sigma_s^2), \tag{2.15}$$

where $\sigma_s^2 = \exp(\nu_s)$. The variance deflation factor c is supposed to improve the calibration of the forecast densities. Following Raftery et al. (2005), a common variance σ_s^2 among all members is used, and a_{ks} and σ_s^2 are estimated from training data as

$$\hat{a}_{ks} = \frac{1}{T}\sum_{t=1}^{T}(x_{kst} - y_{st}) \tag{2.16}$$

$$\hat{\sigma}_s^2 = \frac{1}{mT}\sum_{t=1}^{T}\sum_{k=1}^{m}(x_{kst} - y_{st} - \bar{e}_s)^2. \tag{2.17}$$

In these equations, T denotes the length of the training period and $\bar{e}_s$ is the average of the $m \times T$ errors $e_{st} = x_{kst} - y_{st}$.

The empirical estimates $\{\hat{a}_{ks_i}\}_{i=1}^n$ are viewed as a sample drawn from a stationary Gaussian random field with mean μ_{ak} and covariance function

$$C_{ak}(s_i, s_j) = C(a_{k,s_i}, a_{k,s_j}) = \rho_{ak}^2 + \tau_{ak}^2 \exp\Big(-\frac{||s_i - s_j||}{r_{a_{1k}}} - \frac{|h(s_i) - h(s_j)|}{r_{a_{2k}}}\Big), \tag{2.18}$$

where $||\cdot||$ is the Euclidean norm, ρ_{ak}^2 is the nugget effect describing the measurement error, τ_{ak}^2 is a variance parameter, $r_{a_{1k}}$ is the range corresponding to horizontal distance, $r_{a_{2k}}$ the range corresponding to vertical distance and $h(s)$ is the elevation at location s.

The estimates for $\nu_s = \log(\sigma_s^2)$ are $\hat{\nu}_s = \log(\hat{\sigma}_s^2)$. In a next step the $\{\hat{\nu}_{s_i}\}_{i=1}^n$ are viewed as a partial realization from a stationary Gaussian random field with mean $\mu_{\nu k}$ and covariance matrix

$$C_\nu(s_i, s_j) = C(\nu_{s_i}, \nu_{s_j}) = \rho_\nu^2 + \tau_\nu^2 \exp\Big(-\frac{||s_i - s_j||}{r_{\nu_1}} - \frac{|h(s_i) - h(s_j)|}{r_{\nu_2}}\Big). \tag{2.19}$$

The parameters of the random fields are estimated by maximum likelihood, using $\{\hat{a}_{ks_i}\}_{i=1}^n$ and $\{\hat{\nu}_{s_i}\}_{i=1}^n$ (estimated from the training data at the n locations) as data. Estimation has to be performed numerically. The BMA weights $w_1,\ldots,w_m$ and the deflation parameter c are

estimated with the EM algorithm, similar to the global BMA model. The predictive distribution at a location s for time point t is then given by (2.14), with the estimated parameters plugged in. However, this only holds for locations s that are training sites. For other locations s_0 there are no direct estimates of a_{ks_0} and $\sigma^2_{s_0}$. To obtain the predictive distribution for y_{s_0t} at a location s_0 not present in the training data, the geostatistical interpolation method kriging (Cressie, 1993; Stein, 1999) is employed to get estimates for a_{ks_0} and ν_{s_0} at the unknown site s_0. Plugging in these estimates in (2.14) for the location s_0 yields the predictive distribution for y_{s_0t}.

In contrast to GOP and spatial BMA, GMA first estimates forecast error characteristics in a spatially adaptive way at the available observation stations and then interpolates the estimates to unknown locations using kriging. With this technique, it is possible to compute forecast errors and the postprocessed predictive density at any arbitrary location, not only at the in-sample observation locations. However, the basic GMA model is not designed to sample ensembles of spatially correlated weather fields, as GOP or spatial BMA.

2.1.6 Ensemble Copula Coupling (ECC)

Many of the presented postprocessing methods are designed to be applied to a single weather quantity at a fixed location and for a fixed forecast horizon. They are not able to account for dependence structures between weather quantities or spatial and temporal dependencies. The ensemble forecasts on the other hand may capture many of these multivariate dependence structures present in the weather and climate processes, as they are solutions of the discetized differential equations describing the physics of the atmosphere. When individually postprocessing the univariate margins, one may fail to retain the dependence structure present in the ensemble forecasts. This implies a need for multivariate postprocessing and research has already been conducted in this direction. Methods such as spatial BMA and spatial EMOS extend the basic BMA and EMOS methods to account for spatial structures in forecast errors and forecast variances. Other parametric models, for example to postprocess bivariate wind vectors (Schuhen et al., 2012; Pinson, 2012; Sloughter et al., 2013) account for a special case of inter-variable dependency.

The ECC approach discussed in this section is more general with respect to the type of dependency that is considered. It can be applied to incorporate inter-variable as well as spatial or temporal dependencies. The general formulation allows for considering all dependence structures simultaneously.

The approach was developed in Schefzik (2011) and investigated further in Schefzik et al. (2013). It utilizes the information present in the rank structure of the original raw ensemble. As the raw ensemble represents some of the true multivariate dependence structures of the atmosphere, incorporating its information in the postprocessing procedure typically leads to an improvement in predictive performance.

ECC is a multi-stage procedure resulting in a postprocessed ensemble of the same size m as the raw ensemble. This procedure can be seen as a multivariate postprocessing method, being able to preserve dependence structures between different weather quantities as well as spatial and temporal dependence structures. This is achieved by individually postprocessing the margins of the considered multivariate forecasts and then reorder them according to the rank structure of the raw ensemble. The approach is flexible in that the postprocessing method for the margins can be chosen as desired. There is no restriction to a certain method.

Let $x_1^l, \ldots, x_m^l$, $l = 1, \ldots, L$, denote the l-th univariate margins of the raw ensemble. Here, l may indicate a multi-index, $l = (i, s, t)$, for weather quantity $i = \{1, \ldots, I\}$, location $s = \{1, \ldots, N\}$ and forecast horizon $t = \{1, \ldots, T\}$, in case of considering several multivariate variables simultaneously. However, ECC allows to fix one or more of these indices and consider only specific types of all possible dependencies. Then the ECC postprocessed ensemble can be generated as follows:

- Univariate postprocessing: For each margin l apply a univariate postprocessing technique to the raw ensemble

  $$x_1^l, \ldots, x_m^l \tag{2.20}$$

  and obtain a postprocessed predictive distribution F^l. The univariate postprocessing method utilized can be selected according to the user's needs.

- Quantization: Draw a discrete sample of size m,

  $$\tilde{x}_1^l, \ldots, \tilde{x}_m^l, \tag{2.21}$$

  from each univariate postprocessed predictive distribution F_l.

- Reordering: Determine the permutation σ_l of the integers $\{1, \ldots, m\}$ induced by the order statistics of the raw ensemble, $x_{(1)}^l, \ldots, x_{(m)}^l$, for each margin l. The permutation is defined by $\sigma_l(k) = \text{rank}(x_k^l)$ for $k = 1, \ldots, m$. In case of ties among the ensemble members the ranks are allocated randomly. Then the l-th margin of the ECC postprocessed ensemble is given by

  $$\hat{x}_1^l = \tilde{x}_{(\sigma_l(1))}^l, \ldots, \hat{x}_m^l = \tilde{x}_{(\sigma_l(m))}^l. \tag{2.22}$$

The discrete sample drawn in the quantization step can be generated in different ways. Schefzik et al. (2013) propose three different generation methods. The first method takes equidistant quantiles from the distribution F_l and is therefore called ECC-Q:

$$\tilde{x}_1^l = F_l^{-1}\Big(\frac{1}{m+1}\Big), \tilde{x}_2^l = F_l^{-1}\Big(\frac{2}{m+1}\Big), \ldots, \tilde{x}_m^l = F_l^{-1}\Big(\frac{m}{m+1}\Big). \tag{2.23}$$

The second method, called ECC-R, is just drawing a simple random sample

$$\tilde{x}_1^l = F_l^{-1}(u_1), \ldots, \tilde{x}_m^l = F_l^{-1}(u_m), \tag{2.24}$$

where $u_1, \ldots, u_m \overset{iid}{\sim} \text{Unif}[0,1]$.

The third method, the ECC-T, is a quantile transformation approach. It fits a parametric continuous cumulative distribution function S_l to the raw ensemble margin. Then the quantiles from F_l that correspond to the percentiles of the raw ensemble values in S_l are extracted, so that

$$\tilde{x}_1^l = F_l^{-1}(S_l(x_1^l)), \ldots, \tilde{x}_m^l = F_l^{-1}(S_l(x_m^l)). \tag{2.25}$$

For details on the sampling procedures see Schefzik et al. (2013). These authors also state that the ECC-Q approach is the most favorable, due to theoretical aspects as well as due to the experiences in their case study comparing the predictive performance of the three sampling methods.

A huge advantage of the ECC approach is its low computational cost. The modeling of the multivariate dependence structure of the considered quantities can be performed by only calculating marginal ranks.

Schefzik et al. (2013) show that the ECC method can be interpreted as a nonparametric empirical copula approach. How they define an empirical copula in the context of ECC is briefly summarized below.

Let $\{(x_k^1, \ldots, x_k^L) | k = 1, \ldots, m\}$ denote a data set of size m taking values in $\mathbb{R}^L$. Assuming that there are no ties, denote by $\text{rank}(x_k^l)$ the rank of x_k^l within $x_1^l, \ldots, x_m^l$. Then the corresponding empirical copula E_m is defined as

$$E_m\Big(\frac{i_1}{m}, \ldots, \frac{i_L}{m}\Big) = \frac{1}{m} \sum_{k=1}^{m} \mathbb{I}_{\{\text{rank}(x_k^1) \leq i_1, \ldots, \text{rank}(x_k^L) \leq i_L\}} \tag{2.26}$$

for integers $0 \leq i_1, \ldots, i_L \leq m$ and the indicator function $\mathbb{I}_A$ with $\mathbb{I}_A = 1$ on the chosen set A and $\mathbb{I}_A = 0$ on the complement of A.

Let then $R_1, \ldots, R_L$ denote the corresponding marginal empirical cumulative distribution functions of the raw ensemble, which take values in the set $I_m = \{0, \frac{1}{m}, \ldots, \frac{m-1}{m}, 1\}$. The multivariate empirical cumulative distribution function R of the raw ensemble is defined as the mapping $R : \mathbb{R}^L \to I_m$. Then there exists a uniquely defined empirical copula $E_m : I_m^L \to I_m$ such that

$$R(y_1, \ldots, y_L) = E_m(R_1(y_1), \ldots, R_L(y_L)), \tag{2.27}$$

for all $y_1, \ldots, y_L \in \mathbb{R}$. Additionally, let $\tilde{F}$ and $\hat{F}$ denote the multivariate cumulative distribution functions of the quantized independently postprocessed ensemble (2.21) and the ECC postprocessed ensemble (2.22), respectively. The marginal empirical cumulative distribution functions of the quantized independently postprocessed ensemble are denoted by $\tilde{F}_1, \ldots, \tilde{F}_L$, the corresponding copula denoted by $\tilde{E}_m$. Then, Schefzik et al. (2013) show that

$$\tilde{F}(y_1, \ldots, y_L) = \tilde{E}_m(\tilde{F}_1(y_1), \ldots, \tilde{F}_L(y_L)) \tag{2.28}$$

$$\hat{F}(y_1, \ldots, y_L) = E_m(\tilde{F}_1(y_1), \ldots, \tilde{F}_L(y_L)), \tag{2.29}$$

for all $y_1, \ldots, y_L \in \mathbb{R}$. Therefore, the quantized independently postprocessed ensemble and the ECC ensemble have the same margins, while the raw ensemble and the ECC ensemble have the same copula. The relationships to empirical copulas and the corresponding theoretical background are investigated further in Schefzik (2013).

Chapter 3 presents some basic and fundamental facts on parametric copulas along with examples of statistical approaches employing copulas. Chapters 4 and 5 introduce two different ways of combining a parametric Gaussian copula with a univariate postprocessing method.

2.2 Verification methods

This section presents methods to verify or assess the predictive quality of probabilistic forecasts, e.g. obtained from a statistical postprocessing method.

An important aim of probabilistic forecasting was stated for example in Gneiting et al. (2003), Gneiting et al. (2005) and Gneiting et al. (2007), and picked up by many others. Gneiting et al. (2005) characterize the main issue of probabilistic forecasting as follows:

> "The goal of probabilistic forecasting is to maximize the sharpness of the forecast PDFs subject to calibration (Gneiting et al., 2003). Calibration refers to the

> statistical consistency between the forecast PDFs and the verifications and is a joint property of the predictions and the verifications.[...] Sharpness refers to the spread of the forecast PDFs and is a property of the predictions only.[...] The more concentrated the forecast PDF, the sharper the forecast, and the sharper the better, subject to calibration."

Proper scoring rules play an important role in assessing forecasts as they address sharpness and calibration simultaneously. One example for a proper scoring rule is the CRPS mentioned in Section 2.1.1, which will be explained in more detail here.

As this work considers univariate as well as multivariate forecasts, verification tools for both cases are introduced.

Probabilistic forecasts can take the form of a discrete distribution associated with the forecast ensemble or of a full (continuous) predictive distribution, e.g. derived from the ensemble by employing a postprocessing method. When having a full predictive (postprocessed) distribution, it is nonetheless possible to sample a discrete ensemble of any desired size from this distribution. Therefore verification methods designed for a forecast ensemble can indirectly be applied to postprocessed distributions as well, by generating an ensemble from the respective distribution.

2.2.1 Calibration

Calibration describes the consistency between the ensemble forecasts or other probabilistic forecasts and the verifying observations. It is also referred to as reliability. A simple example can explain the basic idea of reliability: If the occurrence of precipitation is predicted to have probability p in certain circumstances and it really occurs with a frequency of (nearly) p in these circumstances, then the forecast is said to be reliable (Candille and Talagrand, 2005; Hamill, 2001). Calibration is a joint property of the forecasts and the observations.

Univariate calibration

To assess the calibration of a univariate probability distribution, the (relative) frequencies of the Probability Integral Transform (PIT) values can be plotted in equidistant bins. This representation of the frequencies of the PIT values is called the PIT histogram. The univariate ensemble forecasts are assessed with a discrete version of the PIT histogram, the Verification

rank histogram (VRH). The VRH can be employed to assess an ensemble sampled from a full predictive distribution as well.

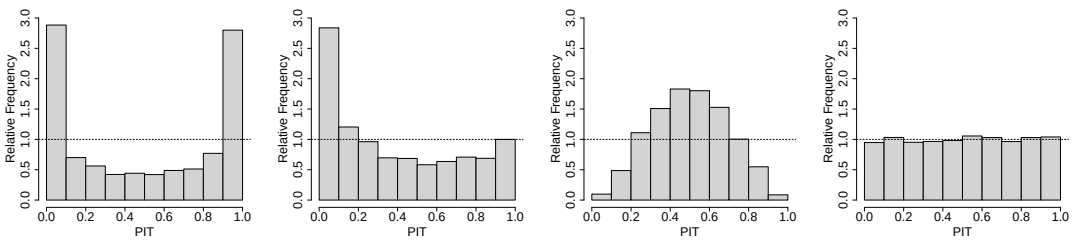
The basic idea of the PIT histogram is the assumption of a true (univariate) probability distribution F for the weather quantity considered (Dawid, 1984; Gneiting et al., 2007). An observation y can then be interpreted as a random sample from that true distribution. If the predicted distribution F' is identical to the true distribution F, the predictive CDF F' evaluated at an observation y, $p = F'(y)$, is a realization of a uniform distribution on [0,1]. In this case the plot obtained by computing the values p for all available observations and sort them into bins on the interval [0,1] produces a uniform histogram. Departures from uniformity can indicate that the predictive distribution is uncalibrated in some way. The number of bins on [0,1] used to obtain the PIT histogram can be chosen as desired, as long as they are equidistant.

There are some rules of thumb for the interpretation of different kinds of departure from uniformity. A U-shaped histogram can be seen as indicator for a predictive distribution with too small spread compared to the spread of the observations (underdispersion), while a hump-shaped histogram indicates the opposite, a distribution with too large spread (overdispersion). A skewed histogram indicates a biased predictive distribution.

Although the theory indicates that a well calibrated predictive distribution displays a uniform PIT histogram (or VHR histogram, see below), Hamill (2001) showed that an uncritical use of the histograms can lead to incorrect conclusions on the predictive distribution or the ensemble. Applying the principle of maximizing sharpness subject to calibration (Gneiting et al., 2007), further measures of predictive quality need to be employed to identify the predictive distribution/ensemble with the best qualities. In either case it is important to interpret the PIT (and VRH) histograms with care.

To illustrate possible appearances of the histograms, Figure 2.1 shows different shapes of PIT histograms. The U-shaped histogram in panel (a) indicates underdispersion, the skewed U-shape in panel (b) often appears when the predictive distribution exhibits a strong bias, the hump-shaped histogram in panel (c) is an indicator for overdispersion, while a nearly uniform PIT histogram like the one in panel (d) indicates a well calibrated predictive distribution.

To assess the calibration of a univariate forecast ensemble $x_1, \ldots, x_m$, $x_i \in \mathbb{R}$, for a real-valued univariate weather quantity, the verification rank histogram (VRH) or Talagrand diagram (Anderson, 1996; Hamill and Colucci, 1997; Talagrand et al., 1997; Hamill, 2001; Candille and Talagrand, 2005) can be employed.



(a) Underdispersion (b) Forecast Bias (c) Overdispersion (d) Calibration

Figure 2.1: Different shapes of PIT histograms

As it is possible to draw a sample from a predictive distribution and interpret this sample as an ensemble, the VRH can be employed for such a sampled ensemble as well. A sampled ensemble of size m can for example be generated by drawing the equidistant $\frac{i}{m+1}$-quantiles, $i = 1, \ldots, m$ from the respective predictive distribution.

The VRH can be obtained as follows (see for example Candille and Talagrand (2005); Talagrand et al. (1997)):

1. Compute the order statistics $x_{(1)}, \ldots, x_{(m)}$ of the ensemble. The $x_{(i)}$ partition the real line into $m + 1$ bins.

2. Obtain the empirical rank of the observation y by determining its position within the bins induced by $x_{(1)}, \ldots, x_{(m)}$.

3. The verification rank histogram plots the empirical frequency of the observation ranks.

If the ensemble members $x_1, \ldots, x_m$ and the observation y are statistically indistinguishable (exchangeable), the rank of the observation with respect to the ensemble members has a discrete uniform distribution on $\{1, \ldots, m + 1\}$.

In case of an ensemble of larger size, e.g. the ECMWF ensemble with $m = 50$ members employed in Section 5, the resolution of the rank histogram might be to high when considering all possible $m + 1$ bins to classify the ranks. For those larger ensembles it is often reasonable to reduce the resolution by classifying the ranks into only r bins, where r is an integer such that $\frac{m+1}{r}$ is an integer, too. The general interpretation of the appearance of the rank histogram is the same as for the PIT histogram.

The deviation from uniformity can be quantified by the reliability index Δ (Delle Monache

et al., 2006; Candille and Talagrand, 2005),

$$\Delta = \sum_{j=1}^{m+1} \left| f_j - \frac{1}{m+1} \right|,$$

where f_j is the observed relative frequency of rank j. The reliability index measures the deviation of the observed relative frequency f_j from its 'ideal' value $\frac{1}{m+1}$ occurring in case of perfect calibration.

Multivariate calibration

To assess the quality of multivariate forecasts, the methods described in Gneiting et al. (2008) can be applied. The authors develop multivariate extensions of the univariate assessment methods and apply them to bivariate wind vectors.

A direct generalization of the univariate assessment of calibration with the VRH is the multivariate rank histogram (MRH). It assesses calibration of an ensemble sampled from a multivariate predictive distribution or of a raw multivariate forecast ensemble. The only challenge lies in defining a multivariate rank order, as no natural ordering exists for multivariate vectors. Gneiting et al. (2008) propose a method for computing a multivariate ordering. This procedure is described below.

Given an ensemble of vector forecasts $x_1, \ldots, x_m$, with $x_k \in \mathbb{R}^p$ for $k = 1, \ldots, m$, for the respective verifying observation vector $x_0 \in \mathbb{R}^p$, the following steps yield the multivariate rank.

1. For $k = 0, \ldots, m$ assign pre-ranks ρ_k to each x_k within the pooled set of the observation and the ensemble members by computing

$$\rho_k = \sum_{j=0}^{m} \mathbb{I}_{\{x_j \preceq x_k\}}.$$

   The pre-ranks are integers between 1 and $m + 1$. Here $u \preceq v$ means that $u_l \leq v_l$ $\forall\, l = 1, \ldots, p$ and $u, v \in \mathbb{R}^p$.

2. Find the multivariate rank r, which is the rank of the observation's pre-rank (ties resolved randomly). For this, consider

$$s^{<} = \sum_{k=0}^{m} \mathbb{I}_{\{\rho_k < \rho_0\}} \quad \text{and} \quad s^{=} = \sum_{k=0}^{m} \mathbb{I}_{\{\rho_k = \rho_0\}}.$$

In step 1 and 2, $\mathbb{I}_A$ denotes the indicator function with $\mathbb{I}_A = 1$ on the chosen set A and $\mathbb{I}_A = 0$ otherwise. Then the multivariate rank r is chosen from a discrete uniform distribution on $\{s^< + 1, \ldots, s^< + s^=\}$ and is an integer between 1 and $m + 1$.

3. The multivariate rank histogram is a plot of the empirical frequency of the multivariate ranks.

Before computing the pre-ranks in step 1 of the procedure it is possible to standardize the observation and the forecast ensemble by a principal component transformation.

As in the univariate case, the reliability index can be computed to quantify the deviation from uniformity. The generalization to the multivariate case is straightforward. The definition is the same as before, namely

$$\Delta = \sum_{j=1}^{m+1} \left| f_j - \frac{1}{m+1} \right|,$$

where f_j is the observed relative frequency of the multivariate rank j.

An alternative to the multivariate rank histogram to assess the calibration of a multivariate forecast ensemble is the minimum spanning tree (MST) histogram (Smith, 2001). The interpretation is slightly different from the one of the (univariate or multivariate) rank histogram and the PIT histogram. Underdispersed or biased ensembles lead to a high number of low MST ranks, while overdispersed ensembles result in too many high MST ranks. Only in the case of calibration, the appearance of the rank or PIT histograms and the MST histograms have the same interpretation: a uniform histogram indicates good calibration.

2.2.2 Sharpness

The assessment of sharpness is also important. Sharpness is a property of the forecasts (or the predictive distributions) only, as it refers to the concentration of the distribution or a forecast ensemble. The sharper a (calibrated!) distribution, the less is the amount of prediction uncertainty, therefore a calibrated sharp predictive distribution provides more accurate forecasts.

Univariate case

The sharpness of an ensemble of forecasts can be assessed by its empirical standard deviation

or range. For a univariate predictive distribution, the standard deviation can be employed as well. Alternatively, the width of prediction intervals can be determined. To have a direct comparability between forecasts from a predictive distribution and a raw ensemble forecast with m members, the nominal $\frac{m-1}{m+1} \cdot 100\%$ prediction interval is employed, as its nominal coverage corresponds to that of the ensemble range.

Multivariate case

In the multivariate case the determinant sharpness (DS) is a useful multivariate measure for sharpness. It is a direct generalization of the standard deviation and is defined as

$$\text{DS} = (\det \Sigma)^{1/(2p)}, \tag{2.30}$$

where Σ is the covariance matrix of an ensemble or a multivariate predictive distribution for a p-dimensional quantity. For ensemble forecasts the matrix is generated by using the empirical variances and correlation of the ensemble. In case of a predictive distribution a requirement for the application of the DS is the existence of the second moment of the respective distribution.

However, the application of the DS has a major disadvantage in case of a nearly singular covariance matrix. This problem was discussed by Jolliffe (2008). For highly correlated variables the DS value can become very small, although the individual variances may be large. Therefore, the interpretation of the DS value in such situations can be misleading. A possible way to come around this problem is to reduce the dimension of the covariance matrix to nearly orthogonal components.

2.2.3 Proper scoring rules

Scoring rules for the verification of deterministic or probabilistic forecasts are well known and have been widely used in forecast assessment, as they provide summary measures for the quality of probabilistic forecasts. Very popular scores are for example the absolute error and the continuous ranked probability score (Matheson and Winkler, 1976; Hersbach, 2000; Gneiting and Raftery, 2007; Gneiting, 2011)

By assigning a numerical score based on the predictive distribution and the respective verifying observation, a scoring rule can be utilized to assess the quality of probabilistic forecasts. Besides this straightforward application, scoring rules can be employed as loss functions in estimation problems, as they are flexible to be tailored to the estimation problem at hand. A

detailed discussion of scoring rules and their mathematical background in measure theory is given in Gneiting and Raftery (2007).

A scoring rule is a function $S(P, y)$, where P denotes the predictive distribution used for forecasting and y denotes the verifying observation. Scoring rules are negatively orientated, a smaller value indicates a more appropriate forecast. The function $S(P, \cdot)$ takes values in $\mathbb{R}$ (or $\overline{\mathbb{R}}$). For y drawn from a distribution Q, the expected value of $S(P, \cdot)$ under Q is $S(P, Q)$.

As the intention is to minimize the value of the scoring rule a reasonable property is that the minimum value is attained if the true distribution is predicted. If

$$S(Q, Q) \leq S(P, Q)$$

holds for all P and Q, the scoring rule is called proper. It is called strictly proper in case equality holds if and only if $P = Q$. Strictly proper scoring rules provide attractive measures to assess the quality of probabilistic forecasts.

In meteorological applications, the scores are typically averaged over all locations and dates in a prescribed test set employed for prediction:

$$S_n = \frac{1}{n} \sum_{i=1}^{n} S(P_i, y_i),$$

where the sum is computed over all data points $i = (t, s)$ in the test set.

Univariate scoring rules

The continuous ranked probability score (CRPS) is a very attractive scoring rule as it is measured in the same units as the observations. Predictive PDFs and deterministic forecasts can both be assessed directly by the CRPS. The CRPS generalizes the absolute error AE and it reduces to it in case a deterministic forecast is used instead of a probability distribution F.

The CRPS for a predictive distribution F and a verifying observation y is defined as

$$\text{CRPS}(F, y) = \int_{-\infty}^{\infty} (F(x) - \mathbb{I}_{\{x \geq y\}})^2 \mathrm{d}x, \tag{2.31}$$

where $\mathbb{I}_{\{x \geq y\}}$ denotes the indicator function, taking the value 1 if $x \geq y$ and 0 otherwise. For some distributions, as for example the normal distribution, the CRPS can be evaluated analytically. However, in many cases it has to be approximated by simulation.

Gneiting and Raftery (2007) show that if F has a finite first moment then the CRPS can alternatively be written as

$$\text{CRPS}(F, y) = E_F|X - y| - \frac{1}{2} E_F|X - X'|, \tag{2.32}$$

where X and X' are independent random variables with distribution F. If F is the cumulative distribution function associated with a forecast ensemble of size m, the CRPS can be computed as

$$\text{CRPS}(F, y) = \frac{1}{m} \sum_{j=1}^{m} |x_j - y| - \frac{1}{2m^2} \sum_{i=1}^{m} \sum_{j=1}^{m} |x_i - x_j|. \tag{2.33}$$

Generally, the CRPS can be approximated by

$$\text{CRPS}(F, y) \approx \frac{1}{n} \sum_{j=1}^{n} |x_j - y| - \frac{1}{2n} \sum_{j=1}^{n} |x_j - x'_j|, \tag{2.34}$$

where $\{x_j\}_{j=1}^n$ and $\{x'_j\}_{j=1}^n$ are two independent samples from F. In case F is the normal distribution, a formula for the CRPS can be derived in closed form. See for example Equation (2.4), where the likelihood of the EMOS parameters is expressed in terms of the CRPS.

If μ_i denotes a deterministic forecast and y_i is the corresponding verifying observation, the absolute error is the absolute difference of the forecast and the observation,

$$\text{AE} = e_i = |\mu_i - y_i|. \tag{2.35}$$

When averaging the AE values over all data points i in the test set, this is called the mean absolute error (MAE):

$$\text{MAE}(\mu_i, y_i) = \frac{1}{n} \sum_{i=1}^{n} e_i. \tag{2.36}$$

In case of a predictive distribution F or if F is the cumulative distribution function associated with a forecast ensemble $x_1, \ldots, x_m$, μ ought to be taken as the median of F (Gneiting, 2011).

Alternatively to the absolute error, the squared error (SE) can be considered for a deterministic style forecast (Gneiting and Raftery, 2007; Gneiting, 2011). It is defined as

$$\text{SE} = (\mu_i - y_i)^2. \tag{2.37}$$

When averaging the SE values over all data points in the test set this results in the mean-square error (MSE).

In case of an ensemble or a predictive distribution F, μ ought to be taken as the mean of the respective distribution (Gneiting, 2011). The MSE is then defined as

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^{n} (y_i - \mu_i)^2. \tag{2.38}$$

The RMSE is the square root of the MSE,

$$\text{RMSE} = \sqrt{\text{MSE}} \tag{2.39}$$

and has the advantage of being recorded in the same unit as the verifying observations.

Multivariate scoring rules

Gneiting et al. (2008) developed generalizations of several univariate verification methods, such as scores and rank histograms. Among them are the Energy score and the multivariate rank histogram described above.

Similar to the univariate case the multivariate forecasts and predictive distributions can be assessed with multivariate generalizations of the univariate scoring rules. The Euclidean error (EE) is a multivariate generalization of the absolute error and is defined as

$$\text{EE}(F, \mathbf{y}) = \|\boldsymbol{\mu} - \mathbf{y}\|, \tag{2.40}$$

where $\boldsymbol{\mu}$ is taken to be the median of F and $||\cdot||$ the Euclidean norm. For an ensemble or a sample from a predictive distribution, the median $\boldsymbol{\mu}$ is defined as the vector that minimizes the sum of the Euclidean distance to the individual forecast vectors,

$$\mu = \min_{\boldsymbol{\lambda}} \Big\{ \sum_{i=1}^{m} \|\boldsymbol{\lambda} - \mathbf{x}_i\| \Big\}.$$

It can be determined numerically using the algorithm described in Vardi and Zhang (2000) as implemented in the `R` package `ICSNP`.

Gneiting and Raftery (2007) introduced the energy score (ES) as a multivariate generalization of the continuous ranked probability score defined in Equations (2.31) and (2.32). The ES is defined as

$$\text{ES}(F, \mathbf{y}) = E_F||\mathbf{X} - \mathbf{y}|| - \frac{1}{2} E_F||\mathbf{X} - \mathbf{X}'||, \tag{2.41}$$

where $||\cdot||$ denotes the Euclidean norm and $\mathbf{X}$ and $\mathbf{X}'$ are independent random vectors with distribution F and finite first moments. In case F is a forecast ensemble of size m, the energy

score can be computed as

$$\text{ES}(F, \mathbf{y}) = \frac{1}{m} \sum_{j=1}^{m} ||\mathbf{x}_j - \mathbf{y}|| - \frac{1}{2m^2} \sum_{i=1}^{m} \sum_{j=1}^{m} ||\mathbf{x}_i - \mathbf{x}_j||. \tag{2.42}$$

A general Monte-Carlo approximation for the energy score is given by

$$\text{ES}(F, \mathbf{y}) \approx \frac{1}{n} \sum_{j=1}^{n} ||\mathbf{x}_j - \mathbf{y}|| - \frac{1}{2n} \sum_{j=1}^{n} ||\mathbf{x}_j - \mathbf{x}'_j||. \tag{2.43}$$

In this approximation, $\{\mathbf{x}_j\}_{j=1}^n$ and $\{\mathbf{x}'_j\}_{j=1}^n$ are two independent samples from F. For further details, see Gneiting et al. (2008), where an overview on the ES is given along with an application to wind vectors.

Chapter 3

Copulas

This chapter provides a short review of the general theory of copulas. For more details on the topic see for example Nelsen (2006), Joe (1997) or Sempi (2011). Furthermore, a Gaussian copula multi-stage procedure is proposed, which is designed to set up multivariate ensemble postprocessing methods by utilizing univariate state of the art postprocessing models.

3.1 Notation and preliminaries

A copula is a function $C : [0,1]^n \to [0,1]$ connecting multivariate distribution functions with their one-dimensional marginal distribution functions. The special feature about the marginal distribution functions of a copula is that they are uniform on the interval $[0,1]$. In case of a continuous multivariate distribution function, the copula is simply the original multivariate distribution function with transformed univariate margins. However, in other cases the construction of a copula is not straightforward and it is necessary to define what is meant by a multivariate distribution function.

For that reason, this section gives an overview on some slightly more abstract concepts of copula theory. The presented notation, theorems and properties basically follow the line of Nelsen (2006). The following short introduction to copulas starts with general multivariate distributions and then moves on to the special subset of copulas and their properties. For a more details on copulas see Nelsen (2006), which is the basis for this short overview, or Sklar (1996) for a general overview on the topic.

Throughout the following, let $\overline{\mathbb{R}}$ denote the extended real line $\overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$, $\text{Dom}(H)$

the *domain of a function* H and $\text{Ran}(H)$ the *range* of H.
Furthermore, a one-dimensional function $f : \mathbb{R} \to \mathbb{R}$ is called *increasing (nondecreasing)* if $x \leq y$ implies $f(x) \leq f(y)$ and *strictly increasing* if $x < y$ implies $f(x) < f(y)$.
For points in $\mathbb{R}^n$, a vector notation is used: For $\mathbf{a} = (a_1, \ldots, a_n)$ and $\mathbf{b} = (b_1, \ldots, b_n)$, $\mathbf{a} \leq \mathbf{b}$ means $a_k \leq b_k$ for all $k \in \{1, \ldots, n\}$.

For a positive integer n let $\overline{\mathbb{R}}^n$ denote the Cartesian product $\overline{\mathbb{R}} \times \overline{\mathbb{R}} \times \ldots \times \overline{\mathbb{R}}$. For $\mathbf{a}, \mathbf{b} \in \overline{\mathbb{R}}^n$ with $\mathbf{a} \leq \mathbf{b}$, $[\mathbf{a}, \mathbf{b}]$ denotes the *n-dimensional box* $B = [a_1, b_1] \times \ldots \times [a_n, b_n]$, which is the product of n closed intervals. The *vertices* of such a box B are of the form $\mathbf{c} = (c_1, \ldots, c_n)$, where each c_k is equal to either a_k or b_k.
The *n-dimensional unit cube* is denoted by $\mathbf{I}^n = \underbrace{\mathbf{I} \times \ldots \times \mathbf{I}}_{n \text{ times}}$, where $\mathbf{I} = [0, 1]$ denotes the unit interval.
Further, an *n-place real function* H is a function whose domain $\text{Dom}(H)$ is a subset of $\overline{\mathbb{R}}^n$ and whose range $\text{Ran}(H)$ is a subset of $\mathbb{R}$.

To introduce copulas in a general setting, it is necessary to define multivariate distribution functions. Before introducing multivariate distribution functions, the one-dimensional version is presented along with the inverse and pseudo-inverse of a one-dimensional distribution function. These definitions are not only important in the framework of copula theory. They are fundamental concepts in general statistical theory and statistical modeling.

DEFINITION 3.1

A function $F : \overline{\mathbb{R}} \to \mathbb{R}$ is called *(one-dimensional) distribution function* if F is nondecreasing, $F(-\infty) = 0$ and $F(\infty) = 1$.

A continuous distribution function F which is strictly increasing has a unique inverse F^{-1}. However, in the case F is only nondecreasing or not necessarily continuous, a more general definition of an inverse is needed.

DEFINITION 3.2

Let F be a one-dimensional distribution function. A function $F^{(-1)}$ with domain $\mathbf{I} = [0, 1]$ such that

(a) $t \in \text{Ran}(F)$ implies that $F^{(-1)}(t)$ is any number $x \in \overline{\mathbb{R}}$ such that $F(x) = t$, that is $F(F^{(-1)}(t)) = t$ holds for all $t \in \text{Ran}(F)$, and

(b) $t \notin \text{Ran}(F)$ implies that $F^{(-1)}(t) = \inf\{x | F(x) \geq t\} = \sup\{x | F(x) \leq t\}$

is called *pseudo-inverse* of F.

If F is strictly increasing and continuous, the pseudo-inverse is equal to the ordinary inverse F^{-1}.

The definitions above were based on the concept of a (strictly) increasing one-dimensional function. The property of an increasing function can be generalized to n-place real functions defined on a subset of $\overline{\mathbb{R}}^n$. To introduce the generalized concept of an n-increasing function, the H-volume of a box B with respect to the n-place real function H needs to be considered. These definitions are the basis for defining a multivariate n-dimensional distribution function with properties analogue to the properties in the one-dimensional case. Having a definition of a multivariate distribution function allows to move on to the definition of a copula.

First, the H-volume of a box B with respect to the n-place real function H is introduced:

DEFINITION 3.3

Let $S_1, \ldots, S_n \subseteq \overline{\mathbb{R}}$, $S_i \neq \emptyset$, $i = 1, \ldots, n$. Further, let H be an n-place real function with domain $S_1 \times \ldots \times S_n$ and let $B = [\mathbf{a}, \mathbf{b}]$ be a box all of whose vertices lie in the domain of H. Then the H-volume of B is defined as

$$V_H(B) = \sum \text{sgn}(\mathbf{c})\, H(\mathbf{c}).$$

The sum is taken over all vertices $\mathbf{c}$ of B and the sgn function evaluated at a vertex $\mathbf{c}$ is given as

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

for $k \in \{1, \ldots, n\}$.

REMARK 3.4

The H-volume of B, $V_H(B)$ can equivalently be represented as the nth order difference of H on B:

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

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

After defining the H-volume of an n-dimensional box B, the concept of an n-increasing function can be introduced, along with the property of a function being grounded.

DEFINITION 3.5

(a) An n-place real function is said to be *n-increasing* if $V_H(B) \geq 0$ for all B whose vertices lie in the domain $\mathrm{Dom}(H)$ of H.

(b) Suppose the domain of an n-place real function H is given by $\mathrm{Dom}(H) = S_1 \times \cdots \times S_n$, where each S_k has a smallest element a_k. Then H is called *grounded* if $H(\mathbf{t}) = 0$ for all $\mathbf{t} = (t_1, \ldots, t_n)\mathrm{Dom}(H)$, such that $t_k = a_k$ for at least one k.
H is said to have *margins* if each S_k is nonempty and has a greatest element b_k.
The *one-dimensional margins* of H are given by the functions H_k with $\mathrm{Dom}(H_k) = S_k$, where $H_k(x) = H(b_1, \ldots, b_{k-1}, x, b_{k+1}, \ldots, b_n)$ for $x \in S_k$. Higher dimensional margins are defined by fixing fewer places in H.

The following definition extends the concept of an n-increasing function to an n-dimensional function that is increasing in each argument.

DEFINITION 3.6

Let $S_1, \cdots, S_n$ be nonempty subsets of $\overline{\mathbb{R}}$. A grounded and n-increasing function H with domain $S_1 \times \ldots \times S_n$ is called *increasing in each argument* if $(t_1, \ldots, t_{k-1}, x, t_{k+1}, \ldots, t_n)$, $(t_1, \ldots, t_{k-1}, y, t_{k+1}, \ldots, t_n) \in \mathrm{Dom}(H)$ and $x \leq y$ implies that $H(t_1, \ldots, t_{k-1}, x, t_{k+1}, \ldots, t_n)$ $\leq H(t_1, \ldots, t_{k-1}, y, t_{k+1}, \ldots, t_n)$.

LEMMA 3.7

From Definition 3.3, Remark 3.4 and Definition 3.5 (a) it follows directly, that for $S_1, \ldots, S_n$ nonempty subsets of $\overline{\mathbb{R}}$ and a grounded n-increasing function H with domain $\mathrm{Dom}(H) = S_1 \times \ldots \times S_n$, H is increasing in each argument.

With Definition 3.5 (a) and (b) it is now possible to formally define the n-dimensional distribution function. This definition is needed later on to state Sklar's theorem, which is fundamental for copula theory.

DEFINITION 3.8

A function $H : \overline{\mathbb{R}}^n \to \mathbb{R}$ is called an *n-dimensional distribution function* if

(a) H is n-increasing and

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

REMARK 3.9

From Lemma 3.7 it follows that the one-dimensional margins of the n-dimensional distribution function H are (one-dimensional) distribution functions. They are denoted by $F_1, \ldots, F_n$.

3.2 Copulas and Sklar's Theorem

After introducing preliminary notation and basic concepts such as distribution functions, n-dimensional distribution functions, grounded and n-increasing functions, a formal mathematical definition of a copula can be given. This definition leads to the important theorem of Sklar, that directly relates a multivariate distribution function to its one-dimensional margins.

One way to formulate a rigorous mathematical definition of a copula is to introduce subcopulas.

DEFINITION 3.10

An *n-dimensional subcopula* C'_n is a real-valued function with the following properties.

(a) The domain of C'_n is the set $S_1 \times \cdots \times S_n$ with $S_k \subseteq \mathbf{I}$, $k = 1, \ldots, n$, where 0 and 1 are contained in each S_k

(b) C'_n is grounded and n-increasing

(c) C'_n has (one-dimensional) margins C'_k, $k = 1, \ldots, n$, satisfying $C'_k(u) = u$ for all $u \in S_k$.

With Definition 3.10, an n-dimensional copula can be introduced as a subcopula with domain $\mathbf{I}^n$.

DEFINITION 3.11

An *n-dimensional copula* C_n is an n-dimensional subcopula whose domain is $\mathbf{I}^n$.

The following remark gives an alternative characterization of an n-dimensional copula.

REMARK 3.12

Equivalently, an n-dimensional copula C_n is a function $C_n : \mathbf{I}^n \to \mathbf{I}$ with the following properties:

(a) For every $\mathbf{u} \in \mathbf{I}^n$, $C_n(\mathbf{u}) = 0$ if at least one coordinate of $\mathbf{u}$ is 0. If all coordinates of $\mathbf{u}$ are 1 except u_k, then $C_n(\mathbf{u}) = u_k$.

(b) For each $\mathbf{a}, \mathbf{b} \in \mathbf{I}^n$ with $\mathbf{a} \leq \mathbf{b}$, it holds for the C_n-volume $V_{C_n}([\mathbf{a}, \mathbf{b}]) \geq 0$.

Definition 3.11 and Remark 3.12 state that a copula C_n is defined on the unit cube $\mathbf{I}^n$ and has uniform margins on $\mathbf{I}$. However, there are many more interesting and useful properties of copulas. The most important one is given in the theorem of Sklar, connecting copulas with multivariate distribution functions and their margins.

THEOREM 3.13 (SKLAR)

For any n-dimensional distribution function H with margins $F_1, \ldots, F_n$ there exists an n-dimensional copula C_n such that for all $\mathbf{x} \in \overline{\mathbb{R}}^n$

$$H(x_1, \ldots, x_n) = C_n(F_1(x_1), \ldots, F_n(x_n)).$$

If $F_1, \ldots, F_n$ are all continuous, then the copula C_n is unique, otherwise C_n is uniquely determined on $\text{Ran}(F_1) \times \cdots \times \text{Ran}(F_n)$. Conversely, for an n-dimensional copula C_n and distribution functions $F_1, \ldots, F_n$, the function H defined above is an n-dimensional distribution function with margins $F_1, \ldots, F_n$.

A proof of Sklar's theorem can be found in Nelsen (2006).

With Sklar's theorem it is possible to express a joint distribution function in terms of an n-dimensional copula and n univariate marginal distribution functions. A reverse characterization of the copula C_n in terms of a multivariate distribution function and pseudo-inverses of its margins follows directly from Theorem 3.13.

COROLLARY 3.14

Let H, C_n, $F_1, \ldots, F_n$ be as in Theorem 3.13. Further, let $F_1^{(-1)}, \ldots, F_n^{(-1)}$ be pseudo-inverses of $F_1, \ldots, F_n$. Then for any $\mathbf{u} \in \mathbf{I}^n$ it holds that

$$C_n(u_1, \ldots, u_n) = H\Big(F_1^{(-1)}(u_1), \ldots, F_n^{(-1)}(u_n)\Big).$$

In case the distribution functions $F_1, \ldots, F_n$ are all continuous, Corollary 3.14 provides a straightforward method for constructing copulas from n-dimensional distribution functions.

3.3 Copulas and random variables

Up to now copulas were introduced on a general mathematical background. No probabilistic concepts like random variables are mentioned in the previous section and the definitions of one-dimensional and multivariate distribution functions are non-probabilistic. However, distribution functions, their (pseudo-) inverses and the central theorem of Sklar can be formulated on a probabilistic background. Chapters 4 and 5 are concerned with an application of copulas in a statistical framework, where postprocessing distributions are estimated from ensemble forecasts and verifying observations. The postprocessing models are based on the implicit assumption of an underlying random variable describing the outcome of a weather quantity of interest.

In a probabilistic framework, a random variable X is said to have a (one-dimensional) distribution function F if $F(x) = P(X \leq x)$ for $x \in \mathbb{R}$ and F is assumed to be continuous from the right. A multivariate (joint) distribution of a vector of n random variables $\mathbf{X} = (X_1, \ldots, X_n)$ is defined in analogy as $F_{X_1,\ldots,X_n}(x_1, \ldots, x_n) = P(X_1 \leq x_1, \ldots, X_n \leq x_n)$ for $\mathbf{x} = (x_1, \ldots, x_n)' \in \mathbb{R}^n$.

The results described in the previous section hold for the probabilistic definition of a (multivariate) distribution function as well. In particular, Sklar's Theorem 3.13 can be formulated in terms of random variables.

THEOREM 3.15

Let $X_1, \ldots, X_n$ be random variables with distribution functions $F_1, \ldots, F_n$ and joint distribution function H. Then there exists an n-dimensional copula C_n such that

$$H(x_1, \ldots, x_n) = P(X_1 \leq x_1, \ldots, X_n \leq x_n) = C_n(F_1(x_1), \ldots, F_n(x_n))$$

for all $(x_1, \ldots, x_n)' \in \mathbb{R}^n$. If the distribution functions $F_1, \ldots, F_n$ are all continuous, C_n is uniquely defined. Otherwise, it is uniquely determined on $\text{Ran}(F_1) \times \cdots \times \text{Ran}(F_n)$.

Sklar's theorem for random variables allows for a connection to an important and well known relationship in statistics, the probability integral transform.

THEOREM 3.16

Let X be a random variable with values in $\mathbb{R}$ and a continuous distribution function F. Then $F(X) \sim \text{Unif}[0, 1]$.

The multivariate distribution function fully describes the dependence structure among the random variables $X_1, \ldots, X_n$. According to Theorem 3.15, the copula C_n separates the multivariate dependence structure from the marginal distribution functions $F_1, \ldots, F_n$ by applying the probability integral transform to the margins.

A random variable version of Corollary 3.14 can be formulated as well.

COROLLARY 3.17

Let H, C_n, $F_1, \ldots, F_n$ be as in Theorem 3.15. Further, let $F_1^{(-1)}, \ldots, F_n^{(-1)}$ be pseudo-inverses of the one-dimensional distribution functions $F_1, \ldots, F_n$. Then it holds that

$$C_n(u_1, \ldots, u_n) = H\Big(F_1^{(-1)}(u_1), \ldots, F_n^{(-1)}(u_n)\Big)$$

for all $(u_1, \ldots, u_n)' \in \mathbf{I}^n$.

Corollary 3.14 and 3.17 provide an explicit way to construct a copula from a given multivariate distribution H with margins $F_1, \ldots, F_n$. Besides this method known as *inversion method*, other ways of explicitly constructing a copula are available, see Nelsen (2006) for examples.

Another result that connects a well known fact from probability theory with copulas in the context of random variables is the following theorem that can be obtained from Sklar's theorem for random variables. It connects the so called product copula or independence copula $\Pi^n(\mathbf{u}) = u_1 \cdots u_n$ (which is an n-dimensional copula for all $n \geq 2$) with independence of random variables. The n random variables $X_1, \ldots, X_n$ are independent if and only if $H(x_1, \ldots, x_n) = F_1(x_1) \cdot \ldots \cdot F_n(x_n)$ for all $x_1, \ldots, x_n \in \mathbb{R}$ and joint probability distribution function H.

THEOREM 3.18

Let $X_1, \ldots, X_n$, $n \geq 2$, be continuous random variables. Then $X_1, \ldots, X_n$ are independent if and only if the n-dimensional copula C_n of $X_1, \ldots, X_n$ is the product copula, that is $C_n = \Pi^n$.

3.4 Gaussian copulas

A very common and useful parametric model is the Gaussian copula, which is an especially tractable type of copula. It is fully defined by the margins and the dependence parameter represented by a correlation matrix. Nelsen (2006) and Hoff (2007) for example present an

overview on Gaussian copulas. The Gaussian copula is of particular interest in this work, as it will be employed in Chapters 4 and 5 to recover the multivariate dependence structure that is not preserved when performing univariate postprocessing for each considered weather quantity and at each location individually.

The Gaussian copula belongs to the family of elliptical copulas defined for elliptical distributions like for example the normal distribution or the t-distribution. According to Embrechts et al. (2003), an elliptical distribution can be defined as follows:

DEFINITION 3.19

Let $\mathbf{X}$ be an n-dimensional random vector, $\boldsymbol{\mu} \in \mathbb{R}^n$, and $\boldsymbol{\Sigma}$ a nonnegative definite and symmetric $n \times n$ matrix. Further, let the characteristic function of $\mathbf{X} - \boldsymbol{\mu}$, $\varphi_{\mathbf{X}-\boldsymbol{\mu}}(\mathbf{t})$, be a function of the quadratic form $\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t}$, that is $\varphi_{\mathbf{X}-\boldsymbol{\mu}}(\mathbf{t}) = \phi(\mathbf{t}'\boldsymbol{\Sigma}\mathbf{t})$. Then, $\mathbf{X}$ has an *elliptical distribution* with parameters $\boldsymbol{\mu}, \boldsymbol{\Sigma}$ and ϕ. This is denoted by $\mathbf{X} \sim E_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \phi)$.

For a random vector $\mathbf{Y} = (Y_1, \ldots, Y_n)$ with marginal distribution functions $F_1, \ldots, F_n$, each Y_j can be transformed to a Gaussian random variable. By first employing the probability integral transform (3.16), a random variable with uniform distribution is obtained:

$$U_j = F_j(Y_j) \sim \text{Unif}[0, 1]$$

for $j \in \{1, \ldots, n\}$. Transforming the U_j with the inverse of the standard normal distribution Φ^{-1} yields

$$Z_j := \Phi^{-1}(U_j) = \Phi^{-1}(F_j(Y_j)) \sim N(0, 1).$$

Assume further that $\mathbf{Z} = (Z_1, \ldots, Z_n) \sim \text{N}_n(\mathbf{0}, \boldsymbol{\Sigma})$ with corresponding CDF Φ_n, where $\text{N}_n(\mathbf{0}, \boldsymbol{\Sigma})$ denotes the n-dimensional normal distribution with mean vector $\mathbf{0}$ and covariance matrix $\boldsymbol{\Sigma}$. With Sklar's Theorem for random variables the multivariate distribution function $H = \Phi_n$ of the vector $\mathbf{Z}$ can be expressed in terms of a copula $\mathcal{C}_n$ as follows

$$\begin{aligned} H(z_1, \ldots, z_n) &= \Phi_n(z_1, \ldots, z_n) && (3.1) \\ &= \mathcal{C}_n(F_{Z_1}(z_1), \ldots, F_{Z_n}(z_n)) && (3.2) \\ &= \mathcal{C}_n(\Phi(z_1), \ldots, \Phi(z_n)), && (3.3) \end{aligned}$$

where $F_{Z_j} = \Phi$ denotes the marginal distribution function of Z_j, $j = \{1, \ldots, n\}$, which is identical to the distribution function of the standard normal distribution for all j.

According to Corollary 3.17, the copula $\mathcal{C}_n$ of the n-dimensional normal distribution Φ_n for a random vector $\mathbf{Y} = (Y_1, \ldots, Y_n)$ is given as

$$\mathcal{C}_n(u_1, \ldots, u_n | \mathbf{C}) = \Phi_n(\Phi^{-1}(u_1), \ldots, \Phi^{-1}(u_n) | \mathbf{C}),$$

where $\mathbf{u} = (u_1, \ldots, u_n) \in \mathbf{I}^n$ and $\Phi_n(\cdot|\mathbf{C})$ is the CDF of the n-dimensional normal distribution having mean vector zero and $n \times n$ correlation matrix and $\mathbf{C}$, Φ^{-1} is the inverse of the CDF of the univariate standard normal distribution and $F_1, \ldots, F_n$ are the marginal distributions of $Y_1, \ldots, Y_n$. Due to the relationships obtained above, this can be equivalently represented as

$$\mathcal{C}_n(u_1, \ldots, u_n|\mathbf{C}) = \Phi_n(\Phi^{-1}(F_1(y_1)), \ldots, \Phi^{-1}(F_n(y_n))|\mathbf{C}), \tag{3.4}$$

Equation (3.4) shows that the Gaussian copula only requires the marginal distribution functions $F_1, \ldots, F_n$ of the original random variables $Y_1, \ldots, Y_n$ and the dependence parameter represented by the correlation matrix $\mathbf{C}$ to be specified. This makes the Gaussian copula model a particularly tractable copula model.

In case of the Gaussian copula (3.4), the parameter describing the dependence between the random variables $Y_1, \ldots, Y_n$ is therefore chosen to be the correlation matrix $\mathbf{C}$ obtained from the covariance matrix $\mathbf{\Sigma}$ of the n-variate normal distribution by standardizing. The standardization results in a positive definite matrix with ones on the diagonal and the pairwise correlations on the off-diagonals. Since the main interest of the Gaussian copula lies in modelling the dependence structure of the random variables $Y_1, \ldots, Y_n$ and not in the specific scaling of this dependence, the correlation matrix $\mathbf{C}$ instead of the covariance matrix $\mathbf{\Sigma}$ is employed.

Several statistical applications utilize Gaussian copula models to account for multivariate dependence structures.

Song (2000) for example discusses the Gaussian copula as a way of formulating multivariate models with given margins. He proposes a maximum likelihood method for estimating the parameters of the marginal models in case they are continuous.

Hoff (2007) introduces a semiparametric estimation method for a Gaussian copula model via an extended rank likelihood of the copula parameters, not depending on any parameters of the marginal distributions. This approach generalizes the marginal likelihood estimation, depending only on the parameters of interest but not on the nuisance parameters. The proposed extended rank likelihood does not need any (parametric) assumptions about the marginal distributions, it is a function of the association parameters of the Gaussian copula model only. Therefore the approach is suitable for any kind of multivariate data, even for mixed continuous and discrete marginal distributions. Parameter estimation is performed in the context of Bayesian inference using a Gibbs sampling scheme. The method is implemented in the `R` package `sbgcop`.

Pitt et al. (2006) propose an estimation procedure for Gaussian copula models that handles the problem of forming a multivariate distribution when some of the marginal distributions are non-gaussian. In their approach the marginal distributions are assumed to belong to specified parametric families, in contrast to the method of Hoff (2007), where assumptions about the distributions of the margins are not necessary and not needed for the estimation procedure. They present a general Bayesian approach for estimating the Gaussian copula model, applicable for arbitrary combinations of discrete and continuous margins and introduce a prior for the correlation matrix that relates to covariance selection models.

Masarotto and Varin (2012) discuss the use of Gaussian copulas to introduce multivariate dependence structures in marginal regression models and call their method Gaussian copula marginal regression (GCMR). This method is suitable for continuous as well as non-continuous margins and provides an extension of standard regression models with normal correlated errors. Masarotto and Varin (2012) suggest a likelihood inference and provide numerical approximations for the non-continuous case, while the likelihood is available in closed form for the continuous case. The procedure is available in the `R` package |gcmr|.

To conclude this short overview, Mikosch (2006) outlines drawbacks of Gaussian copula approaches and discusses some of the criticism that this method occasionally receives. Some of the arguments presented there are taken up and integrated in the general discussion at the end of this work.

Gaussian copulas also enjoy high popularity in applications related to the ensemble postprocessing of weather forecasts considered in this work. Among other application fields, they are popular in problems of hydrology. For some examples see the work of Genest and Favre (2007), Schoelzel and Friederichs (2008) and Kao and Govindaraju (2010).

Gaussian copulas provide convenient approaches for modeling multivariate dependence in statistical applications. However, other parametric and semiparametric copula models are available, as for example the family of elliptical copulas (Demarta and McNeil, 2005), Archimedian copulas (McNeil and Nešlehová, 2009), extremal copulas (Davison et al., 2012) or pair copulas (Aas et al., 2009). Especially in financial statistics and econometrics, these alternative copula families are employed frequently.

3.5 A Gaussian copula multi-stage procedure

To introduce a flexible method capable of extending the standard univariate postprocessing methods discussed in 2.1 to account for inter-variable as well as spatial and/or temporal dependencies, this work proposes a Gaussian copula multi-stage procedure. Developed for extending the univariate postprocessing methods to account for multivariate dependences, this procedure combines the individual univariate postprocessing of the margins with modeling a dependence structure via a correlation matrix $\mathbf{C}$.

The disadvantage of the Gaussian copula marginal regression model proposed by Masarotto and Varin (2012) is the fact, that their parametric multivariate likelihood estimation requires identical margins or at least margins that are all of continuous or all of non-continuous type. When developing a multivariate postprocessing procedure to recover inter-variable dependence a method is required that is capable of mixed types of marginal models. In contrast, the semi-parametric method of Hoff (2007) is not restricted to cases of identical types of margins as the margins are only nuisance parameters in the model. The main interest of Hoff (2007) lies in estimating the dependence parameter $\mathbf{C}$ of the Gaussian copula model. He suggests to combine the posterior distribution of the dependence parameter with the empirical univariate marginal distributions obtained from the data to sample from the multivariate distribution. However, when performing postprocessing of ensemble weather forecasts, the focus additionally lies in specifying the marginal postprocessing models to account for biases and dispersion errors of the raw ensemble.

This section proposes a multi-stage procedure employing a Gaussian copula as well as univariate postprocessing models. The approach allows for the marginal models to take any desired form, continuous as well as non-continuous, and is additionally capable of handling a multivariate vector with a mixture of continuous and non-continuous margins, which is highly useful in the context of multivariate ensemble postprocessing. Although the approach was originally developed for the application to ensemble postprocessing, it is not restricted to this context and may well be modified to be suitable for other applications, e.g. for a more general framework as the Gaussian copula marginal regression.

As the method needs to be suitable for various types of multivariate postprocessing, including to fit a dependence model for a vector with quite different types of margins, the multivariate distribution function defined through the Gaussian copula is not available in closed form in many considered situations. Thus a sampling scheme like the one described for example in Hoff (2007) is employed that yields an arbitrary number of Monte Carlo samples from the multivariate predictive distribution.

The proposed multi-stage Gaussian copula postprocessing procedure works as follows:

Let $\mathbf{Y} = (Y_1, \ldots, Y_k)$ denote the vector of weather variables of interest. The index $j = 1, \ldots, k$ can refer to any multivariate setting tailored to weather forecasting the user might desire. It can be the index of a set of (different) weather quantities, as e.g. temperature, pressure, precipitation or others, at a fixed location and for a fixed date. Alternatively it can denote a set of spatial locations considered for a single weather quantity for a fixed date or a set of dates considered for a single weather quantity at a fixed location.

1. The univariate marginal models F_j for each Y_j are specified by applying an ensemble postprocessing method for each $j = 1, \ldots, k$ individually. Additionally, a $k \times k$ correlation matrix $\mathbf{C}$ between the considered k margins is obtained within this procedure. Here, an appropriate way of estimating this correlation matrix may be chosen by the user.

2. In the second step the joint predictive distribution $\mathbf{F}$ of $\mathbf{Y}$ is set up via a Gaussian copula model as described in Section 3.4:

$$\mathbf{F}(y_1, \ldots, y_k|\mathbf{C}) = \Phi_k(\Phi^{-1}(F_1(y_1)), \ldots, \Phi^{-1}(F_k(y_k))|\mathbf{C}), \tag{3.5}$$

   where Φ is the CDF of the standard normal distribution and Φ_k the CDF of the k-variate standard normal distribution with mean vector 0 and correlation matrix $\mathbf{C}$.

3. To generate samples from the multivariate predictive distribution $\mathbf{F}$, the following sampling model is employed, that relates the margins Y_j to latent Gaussian values $\mathbf{Z}_i = (Z_{i1}, \ldots, Z_{ik})'$:

$$\mathbf{Z}_1, \ldots, \mathbf{Z}_n|\mathbf{C} \overset{iid}{\sim} \mathrm{N}_k(\mathbf{0}, \mathbf{C}), \tag{3.6}$$

$$Y_{ij} = F_j^{-1}(\Phi(Z_{ij})), \tag{3.7}$$

   where $i = 1, \ldots, n$ and $j = 1, \ldots, k$.

Via this procedure, an arbitrary number N of samples can be generated from the multivariate predictive distribution.

The method allows to construct a multivariate postprocessing procedure applicable in many situations concerning ensemble forecasts. Any desired univariate postprocessing model can be chosen to set up the margins F_j. Concerning the estimation strategies of the dependence parameter $\mathbf{C}$ of the Gaussian copula, several methods are possible. Chapter 4 gives an example for the application of this procedure to recover inter-variable dependence structure and presents a simple estimation strategy for $\mathbf{C}$. Some other alternatives are discussed at the end of this work.

Chapter 4

Multivariate extension of BMA

The state of the art postprocessing techniques discussed in Section 2.1 are designed to be applied to a univariate weather quantity. Many of these methods are flexible to be adapted to weather quantities of various types. However, none of these methods take into account the fact that different weather quantities may exhibit different types of interaction. A joint predictive distribution of several quantities, that reflects the dependency structure properly can improve predictive performance. There is for example a well known inverse relationship between temperature and pressure, and precipitation and temperature might be correlated to some extent as well. When applying a postprocessing technique designed only for a univariate quantity to several variables independently, the postprocessed forecasts may violate the multivariate correlation structure of the original ensemble forecasts and the observations.

In low-dimensional multivariate settings it is possible to model the correlation structure directly with a parametric model. This has already been explored in applications to bivariate wind vectors, see for example the work of Pinson (2012), Schuhen et al. (2012) and Sloughter et al. (2013).

In higher dimensions, joint parametric modeling becomes much more of a problem, especially in the case of fitting a joint distribution of weather quantities with marginal distributions that are assumed to be of different types as e.g. in the case of temperature, precipitation and wind speed. In such situations copula methods can be a useful tool. Gaussian copula methods related to the approach presented in the following have already been discussed in the previous chapter.

The next section proposes an application of the Gaussian copula multi-stage procedure presented in Section 3.5 in the context of recovering inter-variable dependence structure from

ensemble forecasts. The univariate BMA method was chosen for individually postprocessing the margins of the considered weather quantities.

In Möller et al. (2013), the multivariate postprocessing procedure involving a Gaussian copula and the BMA methodology is described in detail along with a case study where the method is applied to 48-h forecasts of the 8-member University of Washington Mesoscale Ensemble (UWME, see Chapter 2) for five weather quantities. The following sections give an account of the theory of the multivariate procedure and present the results of the case study in Möller et al. (2013). The review of the theory part and especially the results of the case study are taken almost literally from Möller et al. (2013).

4.1 Combining BMA and Gaussian copulas

The standard ensemble BMA methods described in Section 2.1.2 have proven to work well for post-processing ensemble forecasts of univariate quantities. As already mentioned, treating each weather quantity of interest individually ignores multivariate dependence structures that may be present in the raw ensemble.

To account for this specific shortcoming of the standard BMA method, Möller et al. (2013) propose to model the joint distribution of several weather quantities. They propose to set up a Gaussian copula model, that has already been presented in Section 3.4 in a general framework. Möller et al. (2013) combine the individual postprocessing of the marginal distributions $F_1, \ldots, F_p$ of the p considered weather quantities with a Gaussian copula model as defined in Equation (3.4) in Section 3.4. The procedure of combining univariate postprocessing models and a Gaussian copula is performed according to the general multi-stage procedure described in Section 3.5. This provides the possibility to estimate the joint dependence structure represented by the Gaussian copula parameter $\mathbf{C}$ in Equation (3.4) and the univariate postprocessed marginal distributions separately. In the following, the approach of Möller et al. (2013) is outlined in detail.

Let $F_1, \ldots, F_p$ be the marginal distribution functions (CDFs) of the p weather quantities of interest, represented by random variables $Y_1, \ldots, Y_p$ with $Y_j \sim F_j$ for $j = 1, \ldots, p$. It is assumed that for each considered observation day t, the random variable $Y_j^{(t)}$ has its own distribution model $F_j^{(t)}$. However, for a general introduction of the postprocessing model the time index is often omitted for convenience. The F_j need not to come from the same

distribution family. They are generally defined by

$$F_j(y) = \int_{-\infty}^{y} f_j(u|x_{1j}, \ldots, x_{mj})\,\mathrm{d}u, \tag{4.1}$$

where $f_j(u|x_{1j}, \ldots, x_{mj})$ represents the ensemble BMA density introduced in Equation 2.6 for quantity j, evaluated at u and depending on the ensemble members x_{1j},. . .,x_{mj}.

Further, let $\mathbf{C}$ be a $p \times p$ correlation matrix, i.e. a positive definite matrix with unit diagonal as in Equation (3.4). As outlined in Section 3.4, the dependence parameter $\mathbf{C}$ of the Gaussian copula is chosen to be the correlation matrix obtained from the covariance matrix of the n-dimensional normal distribution by standardization. This standardized dependence structure ensures that the characteristics of the marginal distributions $F_1, \ldots, F_p$ of each univariate quantity remain unchanged, while the dependence structure of the p random variables is fully defined. In the case of forming a multivariate distribution of several (different) weather quantities, $\mathbf{C}$ represents the dependence structure between these weather quantities. As outlined in Section 3.4 a Gaussian copula model combines the marginal distributions F_j in a p-dimensional normal distribution to form the joint distribution function $\mathbf{F}$.

As explained in Section 3.4, it follows from Corollary 3.14 that the joint distribution $\mathbf{F}$ of the p weather quantities takes the form

$$\mathbf{F}(y_1, \ldots, y_p|\mathbf{C}) = \Phi_p(\Phi^{-1}(F_1(y_1)), \ldots, \Phi^{-1}(F_p(y_p))|\mathbf{C}), \tag{4.2}$$

where $\Phi^{-1}(\cdot)$ is the inverse CDF of a standard Gaussian distribution, and $\Phi_p(\cdot|\mathbf{C})$ is the CDF of a p-dimensional Gaussian distribution with mean vector $\mathbf{0}$ and correlation matrix $\mathbf{C}$. See Section 3.4 for a detailed account of Gaussian copulas.

In terms of the general multi-stage procedure described in Section 3.5, the index $j = 1, \ldots, k$ corresponds to the different weather quantities, the total number of considered quantities is $k = p$. The index $t = 1, \ldots, n$ corresponds to the days utilized for the estimation of the latent Gaussian random variables Z_j explained below (not to be confused with the training days employed for each of the BMA models), where the total number of days is $n = T$. Step 1 of the procedure outlined in Section 3.5 involves performing postprocessing with standard BMA for each univariate weather quantity, for each day and at each considered station individually. The model for the joint distribution (4.2) equals the model (3.5) in step 2, with $k = p$.

The Gaussian copula model only requires the marginal distributions $F_1, \ldots, F_p$ and the correlation matrix $\mathbf{C}$ to be fully defined. The marginal distributions of the weather quantities

F_j are directly estimated by the univariate BMA procedure described in Section 2.1.2. It remains to consider a suitable way of estimating the dependence parameter $\mathbf{C}$.

The Gaussian copula sampling model described in Equations (3.6) and (3.7) in Section 3.5 can be set up for the specific situation at hand as follows (Möller et al., 2013).

Let

$$\mathbf{Z}^{(1)}, \ldots, \mathbf{Z}^{(T)} | \mathbf{C} \overset{iid}{\sim} \mathrm{N}_p(\mathbf{0}, \mathbf{C}) \tag{4.3}$$

and for $t = 1, \ldots, T$ and $j = 1, \ldots, p$ set

$$Y_j^{(t)} = F_j^{-1}(\Phi(Z_j^{(t)})), \tag{4.4}$$

where

$$F_j^{-1}(u) = \sup\{y : F_j(y) \leq u\}$$

denotes the pseudo-inverse of the marginal F_j, as introduced in Definition 3.2. Furthermore, $\mathbf{Z}^{(t)} = (Z_1^{(t)}, \ldots, Z_p^t)$ is the vector of p latent Gaussian random variables associated with each of the quantities $1, \ldots, p$ at day $t = 1, \ldots, T$. The corresponding realizations are denoted by $\mathbf{z}^{(t)} = (z_1^{(t)}, \ldots, z_p^t)$. The set of considered weather quantities at a specific day t is represented by the vector of random variables $\mathbf{Y}^{(t)} = (Y_1^{(t)}, \ldots, Y_p^{(t)})$ with corresponding realizations $\mathbf{y}^{(t)} = (y_1^{(t)}, \ldots, y_p^{(t)})$, as noted at the beginning of this section.

With the representation (4.2) it follows from (4.3) and (4.4) that $\mathbf{Y}^{(t)} = (Y_1^{(t)}, \ldots, Y_p^{(t)}) \sim \mathbf{F}$, for all $t = 1, \ldots, T$. The construction also highlights that each $Y_j^{(t)}$ is marginally distributed according to F_j and creates a link between a vector $\mathbf{Y}$ distributed according to $\mathbf{F}$ and a latent Gaussian variable $\mathbf{Z}$.

In the following, the index t is omitted for convenience. In case of a continuous marginal distribution F_j it is immediately obvious that

$$Z_j = \Phi^{-1}(F_j(Y_j)). \tag{4.5}$$

Therefore, is is possible for the majority of weather quantities to directly infer a latent realization z_j, given F_j (or at least an estimate of F_j) and an observed y_j.

However, when building a model that forms a joint distribution of arbitrary weather quantities, there might occur quantities not having a continuous distribution. The general Gaussian copula model (4.2) is capable of any type of marginal distribution F_j. Nevertheless, for applying the sampling model described in (3.6) and (3.7), an appropriate way of defining the inverse of a discrete marginal distribution F_j is necessary.

In case of considering precipitation, the standard approach in ensemble postprocessing is a mixed discrete-continuous distribution with a point mass at zero (Sloughter et al., 2007). The following construction can be employed in case weather quantity j describes precipitation amount: Suppose $Y_j \in [0, +\infty)$, where $F_j(0) = \alpha$ with $0 < \alpha \leq 1$ and F_j is otherwise continuously increasing on $(0, +\infty)$. Then

$$-\infty < Z_j \leq \Phi^{-1}(\alpha) \tag{4.6}$$

holds for $Y_j = 0$ and

$$Z_j = \Phi^{-1}(F_j(Y_j)) \tag{4.7}$$

for $Y_j > 0$.

With a collection of observations $\mathbf{y}^{(1)}$,...,$\mathbf{y}^{(T)}$ is is therefore possible to infer the latent Gaussian values $\mathbf{z}^{(1)}$,...,$\mathbf{z}^{(T)}$, where $\mathbf{z}^{(t)} = (z_1^{(t)}, \ldots, z_p^{(t)})$ and $\mathbf{y}^{(t)} = (y_1^{(t)}, \ldots, y_p^{(t)})$, for $t = 1, \ldots, T$. The latent Gaussian values are then employed to estimate the correlation matrix $\mathbf{C}$. Due to (4.6) and (4.7) this procedure can also be performed for the case of a mixed discrete-continuous distribution assumed for precipitation amount. In case F_j is purely discrete, a similar construction as described above can be used.

Now, let $\mathbf{y}^{(1)}, \ldots, \mathbf{y}^{(T)}$ denote a collected set of multivariate observations of the p weather quantities over T days, which are assumed to be realizations of $\mathbf{Y}^{(1)}$,...,$\mathbf{Y}^{(T)}$. As noted above, $\mathbf{Y}^{(t)} = (Y_1^{(t)}, \ldots, Y_p^{(t)})$ and $\mathbf{y}^{(t)} = (y_1^{(t)}, \ldots, y_p^{(t)})$, for $t = 1, \ldots, T$. According to Möller et al. (2013), each $\mathbf{Y}^{(t)}$ is associated with its own Gaussian copula $\mathbf{F}^{(t)}$:

$$\mathbf{Y}^{(t)} \sim \mathbf{F}^{(t)}(\cdot|\mathbf{C}), \tag{4.8}$$

meaning that the model (4.2) is set up for each day t separately as

$$\mathbf{F}^{(t)}\left(y_1^{(t)}, \ldots, y_p^{(t)}|\mathbf{C}\right) = \Phi_p\left(\Phi^{-1}\left(F_1^{(t)}(y_1^{(t)})\right), \ldots, \Phi^{-1}\left(F_p^{(t)}(y_p^{(t)})\right)|\mathbf{C}\right). \tag{4.9}$$

Here, $F_j^{(t)}$ denotes the ensemble BMA marginal distribution for weather quantity j at day t using the (multivariate) m ensemble members $\mathbf{x}_1^{(t)}$,... $\mathbf{x}_m^{(t)}$, where $\mathbf{x}(t)_k = (x_{1k}^{(t)}, \ldots, x_{pk}^{(t)})$ is the vector containing the k-th ensemble forecast for each variable $1 \ldots, p$.

Although a separate multivariate model $\mathbf{F}^{(t)}$ is assumed on each random vector $\mathbf{Y}^{(t)}$ for a day t, the distribution of the latent factor $\mathbf{Z}^{(t)}$ remains independent of t, despite the time index t in the superscript. This fact was already indicated by the model specification in (4.3), where the parameter $\mathbf{C}$ of the Normal distribution is independent of t. Equation (4.3) therefore means that $\mathbf{Z}^{(t)} \sim N_p(0, \mathbf{C})$ for all t. This allows for utilizing the latent Gaussian observations of

the days $t = 1, \ldots, T$ to estimate the dependence parameter $\mathbf{C}$. Möller et al. (2013) employ the sample correlation matrix of the latent Gaussian values $\mathbf{z}^{(1)}, \ldots, \mathbf{z}^{(T)}$ as an estimator for $\mathbf{C}$.

As pointed out above, Möller et al. (2013) assume that the dependence parameter $\mathbf{C}$ of the Gaussian copula is constant over time. In the general Gaussian copula multi-stage procedure for multivariate postprocessing proposed in Section 3.5 no specific estimation strategy is predetermined. The Gaussian copula model itself is not requiring any specific properties on $\mathbf{C}$ except that it is a correlation matrix (see Section 3.4 for details). A constant correlation as well as a time varying version fits the context of ensemble postprocessing. What estimating strategy is the most appropriate depends on the specific problem hat hand. If the method aims at taking into account seasonal effects present in the data, a time varying correlation matrix might be preferred. Instead of choosing a parametric estimation method like the sample estimator for the correlation, a Bayesian approach as discussed in Hoff (2007) is possible as well. Möller et al. (2013) refer to this topic in a short discussion, outlining other alternatives and possible benefits.

To form the *predictive* distribution for a day s coming some time after T, the (multivariate) m ensemble members $\mathbf{x}_1^{(s)}, \ldots, \mathbf{x}_m^{(s)}$ at time point s and the estimate $\hat{\mathbf{C}}$ obtained from the latent Gaussian vectors $\mathbf{z}^{(1)}, \ldots, \mathbf{z}^{(T)}$ are needed.

The first step in the multivariate postprocessing procedure conducted in Möller et al. (2013) is performed according to step 1 of the multi-stage procedure in Section 3.5, namely to estimate the ensemble BMA predictive marginal distributions $F_1^{(s)}, \ldots, F_p^{(s)}$ from a training period consisting of days prior to s. Note that the training period for forming the predictive ensemble BMA margins on day s is not necessarily identical with the days $t = 1, \ldots, T$ denoting the indices of the training days used to estimate the dependence parameter $\mathbf{C}$ of the Gaussian copula.

Step 2 of the multi-stage procedure involves defining the Gaussian copula model (3.5) for each considered s and plug in the estimate for the parameter $\mathbf{C}$:

$$\mathbf{Y}^{(s)} \sim \mathbf{F}^{(s)}(\cdot|\hat{\mathbf{C}}). \tag{4.10}$$

This joint predictive distribution does not have an easy analytic structure and a parametric estimation is not at all trivial, see e.g. Song (2000).

Step 3 of the multi-stage procedure comprises of producing a desired number of Monte Carlo samples $\hat{\mathbf{Y}}$ from the joint predictive distribution $\mathbf{F}^{(s)}$ by employing the sampling procedure

(3.6) and (3.7): This implies first sampling

$$\hat{\mathbf{Z}} \sim \mathrm{N}_p(\mathbf{0}, \hat{\mathbf{C}}) \tag{4.11}$$

and then setting

$$\hat{Y}_j = (\hat{F}_j^{(s)})^{-1}(\Phi(\hat{Z}_j)). \tag{4.12}$$

Here $\hat{\mathbf{Z}} = (\hat{Z}_1, \ldots, \hat{Z}_p)$ and $\hat{F}_j^{(s)}$, $j = 1, \ldots, p$ denotes the estimated ensemble BMA marginal distribution for day s. A large number of these $\hat{\mathbf{Y}}$ samples effectively describes the entire joint predictive distribution. As noted before, the marginal distributions of $\mathbf{F}^{(s)}$ for each individual quantity remain the ensemble BMA margins $F_1^{(s)}, \ldots, F_p^{(s)}$ due to the properties of a Gaussian copula.

To close the description of the multivariate postprocessing procedure, a brief comment is given on the interpretation of the correlation matrix $\mathbf{C}$ representing the dependence parameter in the Gaussian copula. It describes the correlation between the quantiles of the predictive distribution after postprocessing the raw ensemble. It is a residual correlation matrix, not directly modelling the physical relationship between the considered weather quantities. The true relationships are largely accounted for directly by the ensemble itself. Instead $\mathbf{C}$ accounts for any subsequent correlation between the weather quantities after performing univariate postprocessing.

4.2 Data

For the case study conducted in Möller et al. (2013), daily 48-h forecasts based on the University of Washington Mesoscale Ensemble (UMWE, Eckel and Mass, 2005) with valid dates in the calendar year 2008 are employed. The UWME is an eight-member multi-analysis ensemble using initial and lateral boundary conditions from operational centers around the world. Currently, the UWME employs the WRF mesoscale model that is succeeding the Fifth-Generation Penn State/NCAR Mesoscale Model (MM5). Further information as well as real time forecasts and observations can be found on the web page `http://www.atmos.washington.edu/~ens/uwme.cgi`.

The forecasts are made on a 12 km grid over the Pacific Northwest region of Western North America. As the observation stations are scattered around the region and usually do not correspond to any of the grid points, the forecasts need to be bilinearly interpolated from the four surrounding grid points to the station locations of interest. The observation locations

considered for the case study are situated in the US states of Washington, Oregon, Idaho, California, and Nevada, see Figure 4.3 for an impression of the region and the distribution of the stations. The daily observations are provided by weather observation stations in the Automated Surface Observing Network (National Weather Service, 1998).

Möller et al. (2013) aim at forming a multivariate predictive distribution of five weather quantities: 2-m surface maximum and minimum temperature, sea level pressure, 10-m maximum wind speed, and 24-h precipitation accumulation. Both forecasts and observations are initialized at 00 UTC (Universal Time Coordinated) which is 5pm local time when daylight saving time operates and 4pm local time otherwise. To remove dates and locations with any missing forecasts or observations, quality control procedures as described by Baars (2005) were applied to the entire data set.

The complete data employed in the case study contains dates from the calender years 2006 up to 2008. From the data of the complete year 2008, 60 distinct observation locations are used for the analysis of predictive performance of the proposed method. The 60 stations offer between 95 and 271 days in which all ensemble forecasts and verifying observations were available. The additional data from 2006 and 2007 was utilized to provide an appropriate rolling training period for all days in 2008 and 2007 and to estimate the correlation matrix for the Gaussian copula. The estimate $\hat{\mathbf{C}}$ is obtained from the latent Gaussian observations $\mathbf{z}^{(t)}$ for all days $t = 1, \ldots, T$ available in the 2007 data. To compute the latent $\mathbf{z}^{(t)}$, the marginal BMA densities for the T days in 2007 were estimated. The 2006 data was only employed to provide appropriate training periods for the beginning dates of 2007. The procedure results in a single correlation matrix based on all available dates in 2007. The data of 2008 was employed for forming the predictive distributions that are assessed with multivariate verification tools.

4.3 Results

This section reproduces (in some parts literally) the results of the case study conducted in Möller et al. (2013). They applied the multi-stage Gaussian copula procedure to maximum and minimum temperature, sea level pressure, maximum wind speed and precipitation over the North American Pacific Northwest in 2008. The univariate postprocessing of each quantity with the BMA method is performed at each observation location separately (local BMA). Based on an exploratory analysis using a subset of the data set, a 40-day sliding training period was employed for the parameter estimation of the BMA model. This time period

	maxwsp	precip	mintemp	maxtemp	pressure
maxwsp	1	−0.016	0.032	0.139	−0.123
precip	−0.016	1	−0.001	−0.174	−0.015
mintemp	0.032	−0.001	1	0.239	−0.110
maxtemp	0.139	−0.174	0.239	1	−0.203
pressure	−0.123	−0.015	−0.110	−0.203	1

Table 4.1: Estimated correlation matrix at the station KSEA, Sea-Tac Airport, based on data from the calendar year 2007

consists of the 40 most recent days prior to the forecast date for which the ensemble forecast and the verifying observation were available. Typically, this period corresponds to more than 40 calendar days (see e.g. Wilson et al., 2007, for similar settings).

As the values of the five weather variables considered are given on scales that vary by several orders of magnitude, the components are normalized (using observed mean values and standard deviations) before the scores are calculated over the test set.

4.3.1 Results at Sea-Tac Airport

The first part of the case study conducted in Möller et al. (2013) focused on the predictive performance of the proposed multivariate postprocessing method at a single station. The authors consider the KSEA observation station as an example for the single station performance. The station is located at Sea-Tac Airport, a major transportation hub in the area.

The first step within the multivariate postprocessing procedure is the estimation of the correlation matrix from the latent Gaussian observations. Using all available data from 2007, the ensemble BMA methodology is run for each of the five variables as described in Section 2.1.2. The observations for these data and the estimated marginal distributions are employed to infer the latent Gaussian observations $\mathbf{z}$ as described in Section 4.1. This is performed separately for each day in 2007 and the resulting latent data are then used to estimate a single correlation matrix. Table 4.1 shows the entries of this correlation matrix for the five quantities at the station KSEA.

The displayed correlation matrix shows strong negative correlations between the forecast errors of pressure and both maximum and minimum temperatures. This is in line with the known inverse relationship between temperature and pressure systems. A positive correlation

between the forecast errors of the minimum and maximum temperatures is visible as well, which is intuitively right. These patterns within the correlation matrix imply that a joint distribution of the five weather quantities captures additional information that is ignored in a univariate postprocessing approach.

After estimating the correlation matrix from the data of 2007, the univariate ensemble BMA method is run individually for each of the five quantities to obtain the marginal predictive distributions for all dates in the test data set of 2008. The estimated correlation matrix and these marginal distributions are employed to obtain 20,000 samples from the joint predictive distribution for each day, as described in Section 4.1.

Figure 4.1, taken from Möller et al. (2013), shows a pairwise plot of the estimated joint predictive distribution for January 1, 2008 at the station KSEA. The plot presents a heat map for each pair of quantities, where lighter regions correspond to higher values of probability mass. The 8 ensemble members are represented by red circles and the verifying observation is displayed as a blue square. The diagonal elements of the plot show the marginal ensemble BMA predictive distributions for each of the five quantities. Wind speed is given in meters per second, precipitation in millimeters, temperature in degrees C, and pressure in millibar.

The correlation structure for KSEA presented in Table 4.1 has clearly been carried over to the predictive distribution. The positive correlation between maximum and minimum temperatures is evident as well as the negative correlation of each of these quantities with pressure. The predictive distribution in Figure 4.1 exhibits the effect of postprocessing the marginal distributions with BMA as well: On the one hand, the predictive distributions are often centered away from the ensemble members, which is an effect of the bias-correction embedded in the postprocessing procedure. On the other hand, the spread of the distributions increased in comparison to the raw ensemble. By construction, the margins of the joint distribution displayed on the diagonals remain the unchanged univariate BMA distributions.

To assess the predictive quality of the multivariate postprocessing distribution, the multivariate verification tools described in Section 2.2 are employed.

Table 4.2 presents the values of multivariate verification metrics, namely the energy score (ES), the Euclidean error (EE), the determinant sharpness (DS) and finally the reliability index (Δ). Table 4.2 compares the predictive performance of three different methods, where the scores are averaged over all dates in 2008 at the station KSEA.

The first method is the Gaussian copula approach proposed by Möller et al. (2013), where the correlation structure between the weather quantities is estimated from past data. The

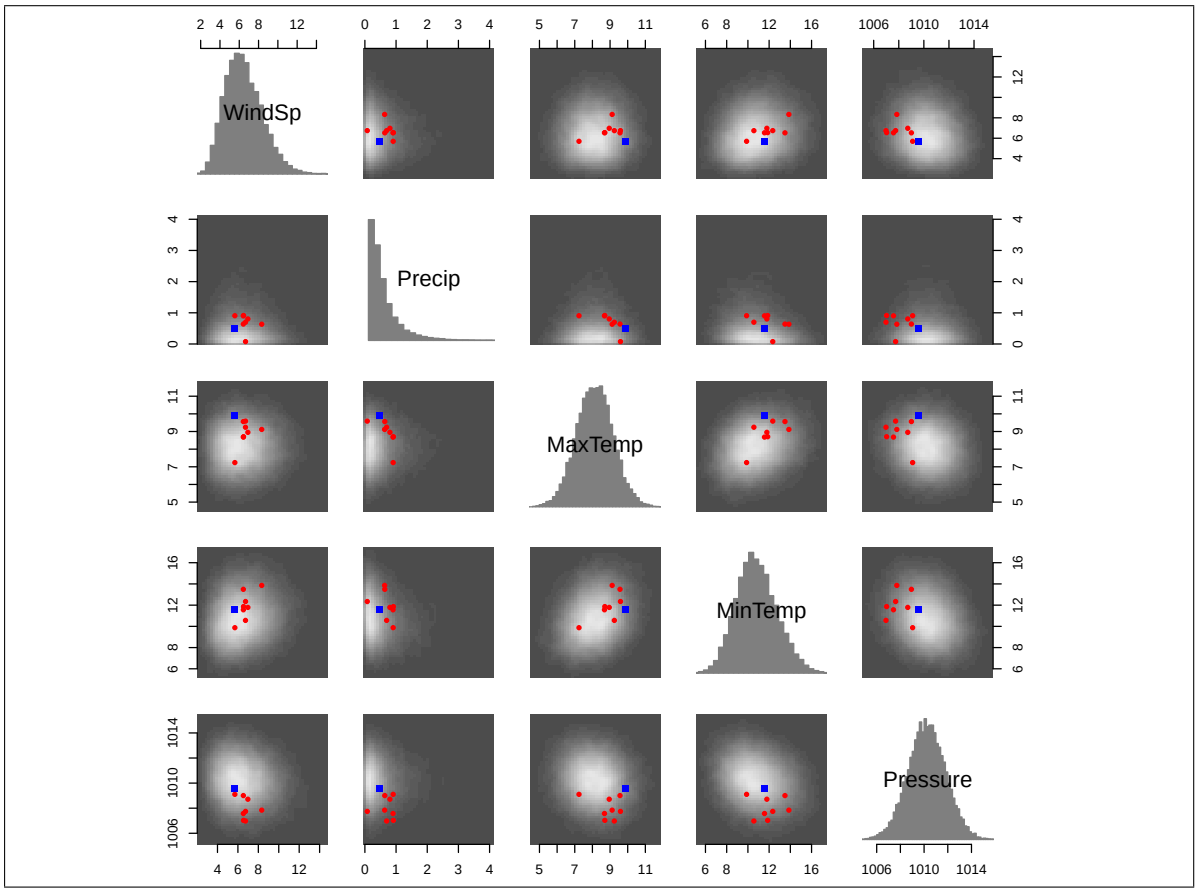


Figure 4.1: Estimated joint predictive distribution for January 1, 2008 at the KSEA observation station along with ensemble predictions (circles) and verifying observation (square)

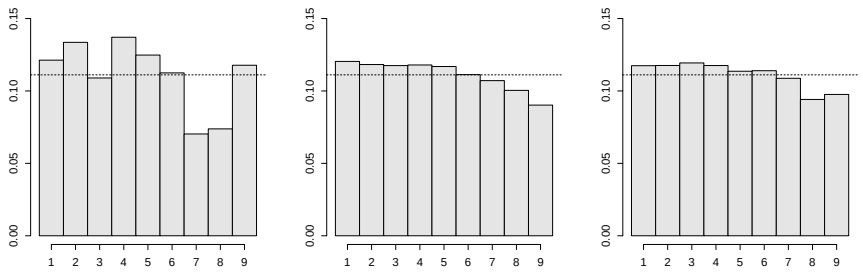
second method called Independence is a Gaussian copula approach with correlation matrix C equal to the identity matrix and therefore assuming that the margins of the five quantities are independent. In principle this approach is simply performing univariate postprocessing of the margins with BMA and combines them in a multivariate distribution not taking into account any dependence structures estimated from the observation data. Note that the Copula and Independence approach have the same marginal distributions and thus differ only in the manner in which the joint distribution is constructed. These two postprocessing approaches are then compared to the raw UWME ensemble.

The Copula and the Independence approach both improve considerably on the raw ensemble in all metrics except the determinant sharpness (DS). However, as the 8 ensemble members naturally impart greater sharpness in comparison to an estimated multivariate predictive distribution, this is not surprising. The high level of sharpness of the raw ensemble is not of great use as the ensemble lacks calibration.

It can be noticed further that the Copula approach improves on the Independence approach

	ES	EE	Δ	DS
UWME	0.938	1.081	0.185	0.566
Independence	0.637	0.982	0.047	7.516
Copula	0.636	0.982	0.019	6.971

Table 4.2: Predictive performance at the KSEA observation station, averaged over the 271 available days in 2008



(a) UWME (b) Independence (c) Copula

Figure 4.2: Multivariate rank histograms for the Copula and Independence approaches as well as the raw UWME ensemble for the KSEA observation station over the 295 available days in 2008

under all metrics. According to the values of the reliability index and the DS, the predictive distribution of the copula approach is both better calibrated and somewhat sharper than the distribution of the Independence approach. The Euclidean scores are essentially the same for the two approaches, since the different correlation matrices cannot be expected to have much influence on the median values of the underlying distributions. According to Equation (2.40), the Euclidean errors of both approaches are based on the median values of their respective distributions. Therefore similar median values lead to similar results for the Euclidean errors. The fact that both distributions have similar median values while the Copula approach displays an improved sharpness leads to a lower energy score value for the Copula approach in comparison to the Independence approach.

To investigate the multivariate calibration of the three approaches, Möller et al. (2013) consider multivariate rank histograms, displayed in Figure 4.2. Both postprocessing methods improve calibration considerably over the raw ensemble. However, in the Independence approach the final bins are somewhat less filled than in the multivariate rank histogram for the Copula approach, though neither returns a completely uniform rank histogram.

4.3.2 Aggregated results over the Northwest US

Möller et al. (2013) conducted a similar analysis for 60 distinct observation stations in the Northwest US. The univariate ensemble BMA model was run individually for each station, day and weather quantity during 2007. Then a correlation matrix was estimated from the verifying observations separately for each observation station. Figure 4.3 shows that there is considerable agreement in the estimated correlations between individual stations, although they were estimated locally.

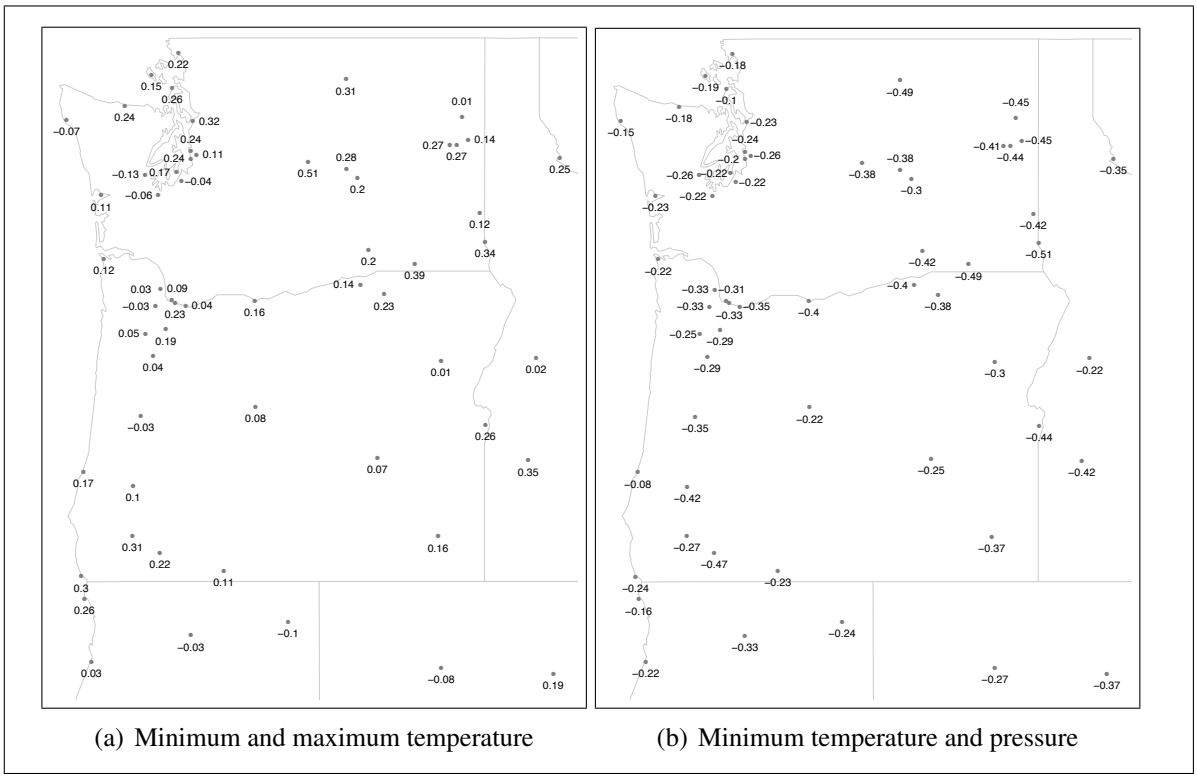
Figure 4.3(a) shows the pairwise estimated correlation between minimum and maximum temperature at each observation station considered for the analysis. As is to be expected, the majority of estimates are positive, up to values of 0.51.

Figure 4.3(b) shows the correlation between minimum temperature and pressure plotted for each observation station. While the previous figure showed limited similarity in the estimated correlations, this figure exhibits considerable agreement between observation stations. All values are negative and the majority lies between -0.2 and -0.4. Apart from this, the plot shows some spatial structure in the correlation estimates. The values near the northern end of the Puget Sound are all roughly between -0.1 and -0.2, while those closer to the Seattle/Tacoma area are grouped between -0.22 and -0.26. A tight group of observation stations in the Columbia River Valley, on the border of Washington and Oregon, all have correlations between -0.31 and -0.35 and finally those in Eastern Washington and Eastern Oregon exhibit stronger correlations typically below -0.4.

The correlations among the weather quantities as well their spatial structure visible in panels (a) and (b) of Figure 4.3 suggest that incorporating this structure by using the Gaussian copula methodology captures important features of the true multivariate distribution that are ignored in the Independence approach.

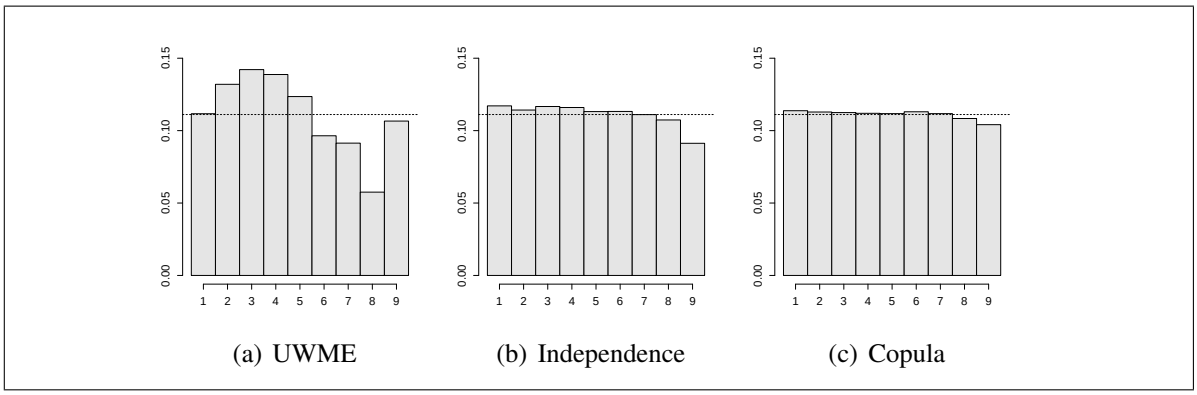
Figure 4.4 compares the multivariate rank histograms (averaged over all observation stations) for the Copula approach, the Independence approach and the raw UWME ensemble. The overall results are similar to those obtained for the station KSEA. Both multivariate postprocessing methods improve the calibration considerably in comparison to the raw ensemble. The effect already described for the station KSEA is visible as well: namely that the last bin is slightly less filled in the Independence approach than in the Copula approach.

The verification metrics results averaged over the 60 stations and all available days in 2008 are presented in Table 4.3. The scores exhibit essentially the same structure as for the station KSEA. The determinant sharpness value improves for the Copula versus the Independence



(a) Minimum and maximum temperature (b) Minimum temperature and pressure

Figure 4.3: Estimated correlation at 60 observation stations in the Northwest US using 2007 data and the Gaussian copula approach



(a) UWME (b) Independence (c) Copula

Figure 4.4: Multivariate Rank histograms for the Copula, the Independence approach and the raw UWME ensemble over all available observations at 60 observation stations in 2008

	ES	EE	Δ	DS
UWME	0.881	1.061	0.161	0.811
Independence	0.586	0.914	0.071	1.945
Copula	0.585	0.914	0.066	1.905

Table 4.3: Predictive performance of the Copula and Independence approaches and the raw UWME ensemble, averaged over 60 observation stations in the Northwest US and all available days in 2008

approach, while the Euclidean score is essentially the same. The energy score improves for the Copula approach, as well as the level of calibration measured by the reliability index.

4.4 Assessing statistical significance

The results presented in Table 4.3 show a small improvement in the energy score when moving from the Independence to the Copula approach, which is considerably less than the improvement when moving from the raw UWME ensemble to the Independence approach.

To assess the significance of the presented results, Möller et al. (2013) conduct a permutation test to show that the small magnitude of the difference is nevertheless not purely the result of sampling variability.

The permutation test is treating the two populations of energy scores as interchangeable under a null hypothesis that the scores come from the same distribution (see Good, 1995, for a detailed discussion of the permutation test and its properties).

Then a large number of synthetic datasets is constructed under this assumption of exchangeability. Thus, for every day and station combination in which the two energy scores were calculated, one score is randomly assigned to the Copula group and the other to the Independence group. This pairwise reassignment is performed since the magnitude of the single energy score values can vary dramatically throughout the year, while the differences vary to a much smaller degree.

Once a synthetic dataset is constructed in this manner, the difference in the means of the two groups is computed and retained. The entire process is repeated a large number of times. Möller et al. (2013) construct 10,000 synthetic datasets to conduct the permutation test.

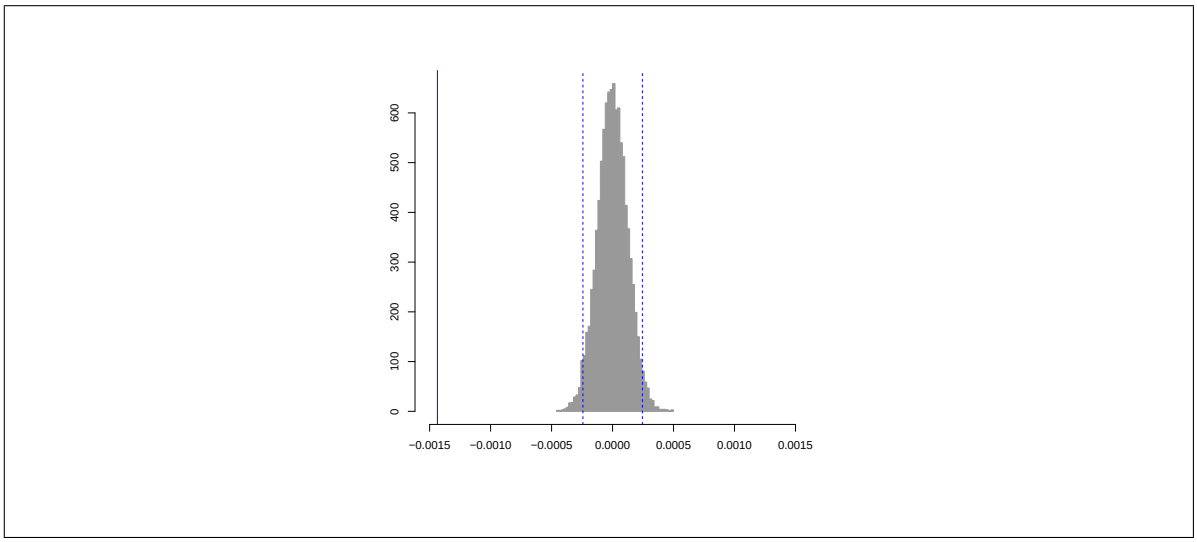


Figure 4.5: Permutation distribution of difference between energy scores for the Copula and Independence approaches along with true value (solid vertical line) as well as 0.025 and 0.975 quantiles (dotted vertical lines)

Figure 4.5 shows the distribution of these permutation scores along with the true difference of -0.00145. As would be expected, the permutation distribution is centered about 0, with 0.025 and 0.975 quantiles of -0.000243 and 0.000245, respectively. None of the 10,000 sampled values from the permutation distribution approaches the true value.

This shows the statistical significance of the results presented in Table 4.3. While the magnitude of the difference is not large, especially in comparison to the initial improvement over the raw ensemble, the result could not have come about purely from sampling variability.

Möller et al. (2013) argue that intuitively, it is reasonable that the improvement when employing the Copula approach instead of the Independence approach would be orders of magnitude smaller than the initial improvement over the raw ensemble, especially for the dataset under consideration. The raw ensemble consists of only 8 forecasts, and thus the UWME results are computed over a 5-dimensional predictive distribution with only 8 points to fill the entire space. By returning a full distribution, as opposed to such a sparse, discrete distribution, ensemble BMA will undoubtedly yield a substantial improvement in the performance of the predictive distribution.

By contrast, the improvement in moving from the Independence approach to the Copula approach is understandably less dramatic, as both procedures return full predictive distributions and also have the same marginal distributions. Nevertheless, the permutation test indicates that the small improvement in Table 4.3 is not due to sampling variability.

Chapter 5

Spatially adaptive extension of EMOS

While the previous section was concerned with extending a univariate postprocessing method to incorporate inter-variable dependencies, this section aims at spatially adaptive estimation of a standard postprocessing method for a single weather quantity. To introduce a spatial dependence structure, the ECC procedure in Section 2.1.6 can be applied to the individually postprocessed margins. This method recovers multivariate dependencies from the rank structure of the original ensemble.

Currently, the method developed in this section is designed for a normally distributed weather quantity such as temperature, but adaptations to weather quantities with other distributions are possible. The basic univariate method employed for the extension is the EMOS model, see Section 2.1.1. However, similar procedures may be developed for other univariate postprocessing models as well.

In a case study analyzing the predictive performance of the EMOS extension, not only the univariate margins from different competing postprocessing methods are compared. An additional case study assesses the multivariate predictive performance of the predictive distributions obtained by providing each univariate method with the ECC induced dependence structure or an Independence structure.

In methods such as GMA (Section 2.1.5) a geostatistical model is used to interpolate estimates to any arbitrary location where no observations are available.

The method developed in this chapter has similar goals.

The standard EMOS method is extended by assuming Gaussian fields (GF) on the bias-

correction parameters. These Gaussian fields are approximated with Bayesian methods, where a recently developed Gaussian Markov random field (GMRF) representation of a GF is utilized to benefit from the computational speed-up.

The following sections outline some of the theory employed in the approach. The theory of GFs and GMRFs is briefly reviewed, along with the INLA methodology to estimate Bayesian models fast and accurately and the SPDE approach of finding a GMRF representation of a GF.

5.1 Gaussian fields and Gaussian Markov random fields

This section gives an overview on the theory of Gaussian fields (GFs) and Gaussian Markov random fields (GMRFs). The presented theory is the background for the SPDE approach introduced in Section 5.3 and employed for the spatially adaptive extension of EMOS in Section 5.4.

Gaussian Fields

A Gaussian field is a specific stochastic process on a domain $D \subseteq \mathbb{R}^d$ of interest. In the context of this work, spatial stochastic processes are of particular interest. For an overview on (spatial) stochastic processes and especially Gaussian fields see for example the monograph of Gelfand et al. (2010) and the references therein. Diggle and Ribeiro Jr. (2010) also provide a short introduction to Gaussian fields in the context of geostatistical models. The main facts presented here were taken from these two monographs. For some basic mathematical properties of stochastic processes in general see e.g. Liptser and Shiryaev (2010).

Let $D \subseteq \mathbb{R}^d$ and $\mathbf{s} \in D$. In spatial applications typically $d = 2$ or $d = 3$. Then a spatial stochastic process $\{X(\mathbf{s}), \mathbf{s} \in D \subseteq \mathbb{R}^d\}$ is defined as the collection of random variables

$$X(\mathbf{s}) = X(\mathbf{s}, \omega), \mathbf{s} \in D, \omega \in \Omega$$

with a well-defined joint distribution. For any fixed $\mathbf{s} \in D$, $X(\mathbf{s})$ is a random variable $X(\mathbf{s}, \omega)$, with Ω being some abstract sample space and $\omega \in \Omega$. For each fixed finite set of spatial locations $(\mathbf{s}_1, \ldots, \mathbf{s}_n)' \subset D$, the vector

$$(X(\mathbf{s}_1), \ldots, X(\mathbf{s}_n))'$$

is a random vector with a specific multivariate distribution. The distribution of the process $\{X(\mathbf{s}), \mathbf{s} \in D \subseteq \mathbb{R}^d\}$ is given by the finite-dimensional distribution function F of the finite

collections $\mathbf{s}_1, \ldots, \mathbf{s}_n$:

$$F(x_1, \ldots, x_n; \mathbf{s}_1, \ldots, \mathbf{s}_n) = P(X(\mathbf{s}_1) \leq x_1, \ldots, X(\mathbf{s}_n) \leq x_n)$$

of the random vector $(X(\mathbf{s}_1), \ldots, X(\mathbf{s}_n))'$ for any arbitrary n and any collection $\mathbf{s}_1, \ldots, \mathbf{s}_n \in D$. For a fixed $\omega \in \Omega$,

$$(x_1, \ldots, x_n)' = (X(\mathbf{s}_1, \omega), \ldots, X(\mathbf{s}_n, \omega))'$$

are realizations of the spatial process $\{X(\mathbf{s}), \mathbf{s} \in D \subseteq \mathbb{R}^d\}$.

In the special case of a spatial Gaussian process, $\{X(\mathbf{s}), \mathbf{s} \in D \subseteq \mathbb{R}^d\}$ is called a Gaussian Field (GF) if and only if

$$(X(\mathbf{s}_1), \ldots, X(\mathbf{s}_n)) \sim \mathrm{N}_n(\boldsymbol{\mu}_n, \boldsymbol{\Sigma}_n)$$

for all finite collections of points $\mathbf{s}_1, \ldots, \mathbf{s}_n \in D$.

A GF is fully determined by its mean function $\mu(\mathbf{s}) = \mathrm{E}(X(\mathbf{s}))$ and its covariance function $C(\mathbf{s}, \mathbf{s}') = \mathrm{Cov}(X(\mathbf{s}), X(\mathbf{s}'))$, $\mathbf{s}, \mathbf{s}' \in D$. For any arbitrary finite set of locations $\mathbf{s}_1, \ldots, \mathbf{s}_n$, the resulting n-dimensional vector $(X(\mathbf{s}_1), \ldots, X(\mathbf{s}_n))$ has an n-variate normal distribution with n-dimensional mean vector $\mu_n = (\mu(\mathbf{s}_1), \ldots, \mu(\mathbf{s}_n))'$ and an $n \times n$ covariance matrix Σ_n with elements $\Sigma_{ij} = C(\mathbf{s}_i, \mathbf{s}_j)$, $i, j = 1, \ldots, n$.

A spatial stochastic process defined on $\mathbb{R}^d$ is called weakly stationary, if and only if for any $\mathbf{s}, \mathbf{h} \in \mathbb{R}^d$

$$\mathrm{E}(X(\mathbf{s})) = \mathrm{E}(X(\mathbf{s}+\mathbf{h})) = \mu \tag{5.1}$$

$$C(\mathbf{s}, \mathbf{s}+\mathbf{h}) = \mathrm{Cov}(X(\mathbf{s}), X(\mathbf{s}+\mathbf{h})) = \mathrm{Cov}(X(0), X(\mathbf{h})) = C(\mathbf{h}). \tag{5.2}$$

In the special case of a GF, weak stationarity is equivalent to strict stationarity. A general spatial stochastic process is called strictly stationary, if for all finite dimensional collections $\mathbf{s}_1, \ldots, \mathbf{s}_n \in \mathbb{R}^d$, all $\mathbf{h} \in \mathbb{R}^d$ and all $x_1, \ldots, x_n \in \mathbb{R}$ the distribution function F fulfills

$$F(x_1, \ldots, x_n; \mathbf{s}_1 + \mathbf{h}, \ldots, \mathbf{s}_n + \mathbf{h}) = F(x_1, \ldots, x_n; \mathbf{s}_1, \ldots, \mathbf{s}_n).$$

The covariance function of a weakly stationary stochastic process needs to be positive definite to be valid. A function C defined on $\mathbb{R}^d$ is called positive definite if the respective covariance matrix of a finite dimensional distribution is nonnegative definite for any collection $\mathbf{s}_1, \ldots, \mathbf{s}_n$. Furthermore, the covariance matrix of a finite dimensional vector $\mathbf{X}(\mathbf{s}) = (X(\mathbf{s}_1), \ldots, X(\mathbf{s}_n))$ needs to be nonnegative definite in general to be valid: For any vector $\mathbf{a} = (a_1, \ldots, a_n)$, the marginal variance σ_L^2 of any linear combination $L = \mathbf{a}'\mathbf{X}(\mathbf{s})$ is

required to be nonnegative. As $\mathbf{a}'\Sigma_n\mathbf{a} = \sigma_L^2 \overset{!}{\geq} 0$, it follows directly that this constraint is only fulfilled if Σ_n is nonnegative definite.

A case of special interest in many applications is when for a weakly stationary process the covariance function depends on $\mathbf{h}$ only in terms of the Euclidian distance $||\mathbf{h}||$. A covariance function with this property is called isotropic. Setting (without loss of generality) $C(0) = 1$, an isotropic covariance function can be written as

$$C(\mathbf{h}) = g(||\mathbf{h}||), \quad \mathbf{h} \in \mathbb{R}^d, \tag{5.3}$$

for a continuous function $g : [0, \infty) \to \mathbb{R}$ with $g(0) = 1$. When defining G_d as the class of all continuous functions g generating a valid isotropic covariance function on $\mathbb{R}^d$ via (5.3), the following properties hold:

$$G_1 \supseteq G_2 \supseteq \ldots \qquad \text{and} \qquad G_d \downarrow G_\infty = \bigcap_{d \geq 1} G_d,$$

as it is possible to restrict an isotropic process in $\mathbb{R}^d$ to any lower-dimensional subspace.

The class of valid isotropic covariance functions contains many different parametric families of covariance functions. The most popular one is the Matérn family (Matérn, 1986; Guttorp and Gneiting, 2008). Other parametric families of interest are for example the powered exponential family or the Cauchy family. The Matérn, powered exponential and Cauchy family belong to the class G_∞, meaning they generate valid isotropic covariance functions in all dimensions $d \geq 1$. For more details on some families of covariance functions and their properties see for example Gelfand et al. (2010) and the references therein. The family of Matérn covariance functions will be of particular interest in Section 5.3, where their properties are briefly discussed.

Gaussian Markov random fields

While GFs are specific continuous (spatial) random processes, a Markov random field (MRF) is a finite-dimensional vector $(X_1, \ldots, X_n)'$ having a specific parametric distribution and satisfying certain conditional independence properties inducing a Markovian structure. A Gaussian Markov random field (GMRF) denotes the special case where the finite-dimensional vector has a normal distribution. For details on GMRFs see the monograph of Rue and Held (2005), which is essentially the basis for the overview presented here. Further, in Gelfand et al. (2010) some facts on general MRFs as well as on GMRFs are presented. For convenience, the GMRF is introduced in the general notation of an arbitrary random vector $\mathbf{X} = (X_1, \ldots, X_n)'$ in the following. However, all the definitions and properties directly

translate to the special case of a spatial GMRF $\mathbf{X} = (X(\mathbf{s}_1), \ldots, X(\mathbf{s}_n))'$ for a finite collection of spatial locations $\mathbf{s}_1, \ldots, \mathbf{s}_n \in \mathbb{R}^d$.

To understand the underlying structure of a GMRF (or a MRF in general), a short review of the definition of (marginal) independence and conditional independence is necessary. Two continuous random variables X and Y are independent if and only if and only if $p_{X,Y}(x, y) = p_X(x) \cdot p_Y(y)$ for the respective marginal PDFs $p_X(\cdot)$, $p_Y(\cdot)$ and the joint PDF $p_{X,Y}(\cdot, \cdot)$. This property is denoted by $X \perp Y$. The random variables X and Y are called conditionally independent given a third random variable Z, if and only if $p_{X,Y|Z}(x, y|z) = p_{X|Z}(x|z) \cdot p_{Y|Z}(y|z)$, where $p_{X,Y|Z}(\cdot, \cdot|\cdot)$ is the joint PDF of X and Y given Z, $p_{X|Z}(\cdot|\cdot)$ the conditional PDF of X given Z and $p_{Y|Z}(\cdot|\cdot)$ the conditional PDF of Y given Z. The conditional independence is denoted by $X \perp Y|Z$. In the case of conditional independence of X and Y, the two random variables can still be marginally dependent. Conditional independence can be defined for random vectors as well: $\mathbf{X}$ and $\mathbf{Y}$ are conditionally independent given $\mathbf{Z}$ if and only if $p_{\mathbf{X},\mathbf{Y}|\mathbf{Z}}(\mathbf{x}, \mathbf{y}|\mathbf{z}) = p_{\mathbf{X}|\mathbf{Z}}(\mathbf{x}|\mathbf{z}) \cdot p_{\mathbf{Y}|\mathbf{Z}}(\mathbf{y}|\mathbf{z})$ for the respective conditional densities.

A knowledge of some basic notions about graph theory is useful in the context of random variables, as the conditional independence structure of a GMRF can be visualized with an undirected graph $G = (V, E)$, where V denotes the set of vertices (also called nodes) in the graph and E denotes the set of edges. In the special case of $V = \{1, \ldots, n\}$, V defines a labelled graph. The set of edges contains all $\{i, j\}$ with $i, j \in V$ and $i \neq j$. If $\{i, j\} \in E$, there is an edge from node i to node j. In case $\{i, j\} \in E$ for all $i, j \in V$ with $i \neq j$, G is called fully connected. See for example Lauritzen (1996) and Whittaker (1990) for an introduction to graphical models, their basic notions and applications in (multivariate) statistics.

Before giving a formal definition of a GMRF, some preliminary notation has to be introduced. An $(n \times n)$ matrix $\mathbf{A}$ is called positive definite if and only if

$$\mathbf{x}' \mathbf{A} \mathbf{x} > 0 \quad \forall \mathbf{x} \neq 0.$$

In case $\mathbf{A}$ is also symmetric, it is called symmetric positive definite (SPD). In the following only SPD matrices are considered.

Regarding a GMRF, there is a connection between the graph G and the parameters of the normal distribution (that need not to be valid in a general MRF), which is described by the following theorem (Rue and Held, 2005):

THEOREM 5.1

Let the random vector $\mathbf{X} = (X_1, \ldots, X_n)'$ be normally distributed with mean $\boldsymbol{\mu}$ and an SPD precision matrix $\mathbf{Q}$, where $\mathbf{Q} = \boldsymbol{\Sigma}^{-1}$ is the inverse of the SPD covariance matrix $\boldsymbol{\Sigma}$. Then for $i \neq j$:

$$X_i \perp X_j | \mathbf{X}_{-ij} \iff Q_{ij} = 0.$$

The theorem states that the non-zero pattern of $\mathbf{Q}$ directly relates to the structure of the graph G. If an element Q_{ij} of the precision matrix is zero, X_i and X_j are conditionally independent given $\mathbf{X}_{-ij}$.

This leads directly to the definition of a GMRF.

DEFINITION 5.2

Let $\mathbf{X} = (X_1, \ldots, X_n)'$ be a random vector with realizations in $\mathbb{R}^n$. Then $\mathbf{X}$ is called a GMRF with respect to a labelled graph $G = (V, E)$ with mean $\boldsymbol{\mu}$ and SPD precision matrix $\mathbf{Q}$, if and only if its density has the form

$$p(\mathbf{x}) = (2\pi)^{-n/2} \, |\mathbf{Q}|^{1/2} \, \exp\Big(-\frac{1}{2} \, (\mathbf{x} - \boldsymbol{\mu})' \mathbf{Q} (\mathbf{x} - \boldsymbol{\mu}) \Big) \tag{5.4}$$

and for all $i \neq j$

$$Q_{ij} \neq 0 \iff \{i, j\} \in E.$$

Besides the well known parametrization via the mean μ and the precision $\mathbf{Q}$ matrix used in the definition of a GMRF above, there is a canonical parametrization of a GMRF. It is given in the following definition.

DEFINITION 5.3

A GMRF $\mathbf{X}$ with respect to a graph G with canonical parameters $\mathbf{b}$ and $\mathbf{Q}$, where $\mathbf{Q}$ is an SPD matrix, has a density $p(\mathbf{x})$ with the property

$$p(\mathbf{x}) \propto \exp\Big(-\frac{1}{2} \, \mathbf{x}' \mathbf{Q} \mathbf{x} + \mathbf{b}' \mathbf{x} \Big),$$

where $\mathbf{Q} = \Sigma^{-1}$ is the precision matrix and the mean of the distribution can be obtained via the relationship $\boldsymbol{\mu} = \mathbf{Q}^{-1}\mathbf{b}$. The canonical representation is denoted by $\mathbf{X} \sim N_C(\mathbf{b}, \mathbf{Q})$.

There is a direct relationship between the canonical and the standard parametrization of the normal distribution, given by $\mathrm{N}(\boldsymbol{\mu}, \mathbf{Q}^{-1}) = \mathrm{N}_C(\mathbf{Q}\boldsymbol{\mu}, \mathbf{Q})$.

As mentioned above, a GMRF is characterized in terms of the precision matrix $\mathbf{Q}$ instead of the covariance matrix $\mathbf{\Sigma}$. The elements of $\mathbf{Q}$ have a direct conditional interpretation. However, the precision matrix $\mathbf{Q}$ is hard to interpret in terms of marginal dependence. The diagonal elements of $\mathbf{Q}$ are the precisions of X_i given $\mathbf{X}_{-i}$, while the off-diagonal elements describe the conditional correlation (with proper scaling) between the random variables X_i and X_j given $\mathbf{X}_{-ij}$. Conversely, the covariance matrix $\mathbf{\Sigma}$ provides information on the marginal variance of X_i and the marginal correlation between X_i and X_j, but do not provide any information on the conditional dependence structure.

GMRFs are a useful concept in hierarchical models, where they can describe stochastic dependence between a set of unknown parameters. A typical application defines a GMRF $\mathbf{X}$ on a vector of unknown hyperparameters $\boldsymbol{\theta}$ in a three stage procedure:

$$\boldsymbol{\theta} \sim p(\boldsymbol{\theta}) \tag{5.5}$$

$$\mathbf{X} \sim p(\mathbf{x}|\boldsymbol{\theta}) \tag{5.6}$$

$$Y_i \overset{iid}{\sim} p(y_i|x_i), \quad i = 1, \ldots, n. \tag{5.7}$$

Such hierarchical model connects the latent field $\mathbf{X}$ to the observational variables $\mathbf{Y}$, which are assumed to be conditionally independent given $\mathbf{X}$. Then the joint posterior of the latent field and the hyperparameters reads

$$p(\mathbf{x}, \boldsymbol{\theta}|\mathbf{y}) \propto p(\boldsymbol{\theta})p(\mathbf{x}|\boldsymbol{\theta}) \, \Pi_{i=1}^{n} \, p(y_i|x_i). \tag{5.8}$$

This way of employing a GMRF in a hierarchical model is also utilized in the recently proposed INLA approach, developed by Rue et al. (2009) and described in the next section.

5.2 Integrated Nested Laplace Approximation (INLA)

Bayesian methods have become more and more popular during the last three decades, as they provide several advantages. The possibility to specify prior distributions allows for including information obtained in advance, e.g. from previous studies or experts in the considered field. Bayesian models readily allow for defining a hierarchical structure on the data or the parameters of the model, making prediction or imputation of missing values easy.

A main challenge when employing Bayesian methods is computational. Markov chain Monte Carlo (MCMC) methods provide an extremely flexible approach to estimate Bayesian models, they can essentially deal with any type of Bayesian model and any type of data.

Nonetheless, MCMC simulations to obtain posterior distributions can become computationally intensive as the model gets complex. Therefore, an application of MCMC based inference can be restricted to relatively simple models, while complex models and high-dimensional data still pose challenges.

The recently developed Integrated Nested Laplace Approximation (INLA), proposed by Rue et al. (2009), provides a computationally efficient alternative to MCMC methods. The INLA methodology has a wide range of possible applications, as it is designed for the class of latent Gaussian models. This class contains for example (generalized) linear mixed models, spatial and spatio-temporal models, and many more standard models widely applied in statistics. As INLA is available as an `R` package, `R-INLA`, the method can easily be accessed by any researcher. The web page `www.r-inla.org` provides information on the use of the `R-INLA` package, papers on background theory and applications as well as discussion groups and further help.

In the following, the basic ideas of the INLA procedure described in Rue et al. (2009) are explained.

Latent Gaussian models are a subclass of the so called structured additive regression models. In this class of models the response variables $Y_i, i \in \mathcal{I}$ with corresponding realizations y_i (the observations) are assumed to belong to an exponential family. Here, $\mathcal{I}$ is a finite index set that can be defined according to the needs of the application. The aim of a model belonging to this class is to estimate the effect of a set of covariates on some functional, e.g. the mean μ_i, of the conditional predictive distributions.

To derive a general model, the mean μ_i is linked to a structured additive linear predictor η_i through a link function $g(\cdot)$ by $g(\mu_i) = \eta_i$. The structured additive linear predictor includes the effects of various covariates. It is generally defined as

$$\eta_i = \gamma + \sum_{k=1}^{K} \beta_k \, z_{ki} + \sum_{l=1}^{L} f_l(u_{li}) + \xi_i, \tag{5.9}$$

where γ denotes a scalar overall intercept, the vector $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_K)$ contains the coefficients for the linear effects of the covariates $\mathbf{z} = (z_1, \ldots, z_K)$ on the response variable, $\mathbf{f} = \{f_1(\cdot), \ldots, f_L(\cdot)\}$ are unknown functions of the covariates $\mathbf{u} = (u_1, \ldots, u_L)$, and the components of $\boldsymbol{\xi} = (\xi_i, i \in \mathcal{I})$, are unstructured terms. The vector of all linear predictor components is given as $\boldsymbol{\eta} = (\eta_i, i \in \mathcal{I})$ As the functions $f_l(\cdot)$ can take many different forms, this model has a wide range of applications.

Rue et al. (2009) denote by $\mathbf{x} := (\gamma, \boldsymbol{\beta}, \mathbf{f}, \boldsymbol{\eta})$ the vector of all parameters of interest in the

model (5.9).

The class of latent Gaussian models is a subclass of additive Bayesian models, where a Gaussian prior is assumed on the components of $\mathbf{x}$. For the INLA procedure, Rue et al. (2009) go one step further and assume a GMRF prior with mean zero and precision matrix $\mathbf{Q}(\boldsymbol{\theta_1})$ on $\mathbf{x}$, depending on a vector of hyperparameters $\boldsymbol{\theta}_1$. This is denoted by

$$\mathbf{x}|\boldsymbol{\theta}_1 \sim \mathrm{N}\big(\mathbf{0}, \mathbf{Q}^{-1}(\boldsymbol{\theta_1})\big), \tag{5.10}$$

where $\mathrm{N}\big(\mathbf{0}, \mathbf{Q}^{-1}(\boldsymbol{\theta_1})\big)$ denotes the multivariate normal distribution with mean vector $\mathbf{0}$ and covariance matrix $\mathbf{Q}^{-1}(\boldsymbol{\theta_1})$. This precision matrix has a sparse structure and therefore allows for sparse matrix algorithms (Rue and Held, 2005). This assumption is reasonable as many latent Gaussian models exhibit the properties underlying this assumption. The specific Markovian structure of such a model is useful to facilitate the approximations performed by INLA.
Rue et al. (2009) further assume that the response variables $\mathbf{Y} = (Y_i, i \in \mathcal{I})$ are conditionally independent given $\mathbf{x}$ and a second vector of hyperparameters $\boldsymbol{\theta}_2$. The specific distribution of $\mathbf{Y}|\mathbf{x}, \boldsymbol{\theta}_2$ depends on the distributional family that was assumed for the response variables $\mathbf{Y} = (Y_i, i \in \mathcal{I})$ themselves. Let then

$$\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) \tag{5.11}$$

denote the vector of all hyperparameters of the specified model, where $\boldsymbol{\theta}_1$ is the sub-vector of hyperparameters associated with the parameter vector $\mathbf{x}$ and $\boldsymbol{\theta}_2$ the sub-vector of hyperparameters associated with the response variables $\mathbf{Y} = (Y_i, i \in \mathcal{I})$. Rue et al. (2009) state that the dimension m of $\boldsymbol{\theta}$ needs to be much smaller than the dimension of $\mathbf{x}$ ($m \leq 6$) for the INLA procedure to work properly. The vector of hyperparameters is not necessarily Gaussian.

In the following, let $p(\cdot|\cdot)$ denote a conditional density. The main objective of the INLA procedure is to obtain accurate approximations for the posterior margins of the x_i, the components of the latent Gaussian vector $\mathbf{x}$, given the observations $\mathbf{y} = (y_i, i \in \mathcal{I})$, namely

$$p(x_i|\mathbf{y}) = \int p(\boldsymbol{\theta}|\mathbf{y})\, p(x_i|\boldsymbol{\theta}, \mathbf{y})\, \mathrm{d}\boldsymbol{\theta}, \tag{5.12}$$

as well as for the posterior margins of the vector of hyperparameters $\boldsymbol{\theta}$ and its components θ_j given $\mathbf{y} = (y_i, i \in \mathcal{I})$, namely

$$p(\theta_j|\mathbf{y}) = \int p(\boldsymbol{\theta}|\mathbf{y})\, \mathrm{d}\boldsymbol{\theta}_{-j}. \tag{5.13}$$

To determine the integrals (5.12) and (5.13), $p(\boldsymbol{\theta}|\mathbf{y})$ and $p(x_i|\boldsymbol{\theta},\mathbf{y})$ need to be computed or approximated.

The basic idea of the INLA approach is to construct nested approximations of the form

$$\tilde{p}(x_i|\mathbf{y}) = \int \tilde{p}(\boldsymbol{\theta}|\mathbf{y})\ \tilde{p}(x_i|\boldsymbol{\theta},\mathbf{y})\ \mathrm{d}\boldsymbol{\theta} \tag{5.14}$$

and

$$\tilde{p}(\theta_j|\mathbf{y}) = \int \tilde{p}(\boldsymbol{\theta}|\mathbf{y})\ \mathrm{d}\boldsymbol{\theta}_{-j}, \tag{5.15}$$

where $\tilde{p}(\cdot|\cdot)$ is an approximation to the respective conditional density. After approximating the conditional densities $p(\boldsymbol{\theta}|\mathbf{y})$ and $p(x_i|\boldsymbol{\theta},\mathbf{y})$ by $\tilde{p}(\boldsymbol{\theta}|\mathbf{y})$ and $\tilde{p}(x_i|\boldsymbol{\theta},\mathbf{y})$ respectively, numerical integration is performed to integrate out $\boldsymbol{\theta}$ in Equations (5.12) and (5.13). This is possible because of the small dimension of $\boldsymbol{\theta}$.

Approximating the posterior margins of interest takes advantage of the Gaussian model assumptions. These approximations are further based on the Laplace approximation (Tierney and Kadane, 1986). Due to the nested approach, the Laplace approximation works especially well when applied to latent Gaussian models. Because of this feature, Rue et al. (2009) called their approach Integrated Nested Laplace Approximations (INLAs). The INLA appraoch for approximating the posterior margins $p(x_i|\mathbf{y})$ of the components of the latent field $\mathbf{x}$ is performed in three steps. The first step is concerned with the approximation of the posterior margin $p(|\mathbf{y})$ of the vector of hyperparameters $\boldsymbol{\theta}$ as outlined below in Equations (5.18) and (5.19). The second step computes the Laplace approximation or the simplified Laplace approximation of $p(x_i|\boldsymbol{\theta},\mathbf{y})$ for selected values of $\boldsymbol{\theta}$, to improve the Gaussian approximation $\tilde{p}_G(\mathbf{x}|\boldsymbol{\theta},\mathbf{y})$. The Gaussian approximation itself is shortly explained below. These two steps are combined in a third step by using a numerical integration of the form (5.25).

The approximation of the posterior density $p(\boldsymbol{\theta}|\mathbf{y})$ utilizes a Gaussian approximation in the way described below.

Since $p(\mathbf{x},\boldsymbol{\theta}|\mathbf{y}) = p(\mathbf{x}|\boldsymbol{\theta},\mathbf{y}) \cdot p(\boldsymbol{\theta}|\mathbf{y})$ it follows that

$$\begin{aligned} p(\boldsymbol{\theta}|\mathbf{y}) = \frac{p(\mathbf{x},\boldsymbol{\theta}|\mathbf{y})}{p(\mathbf{x}|\boldsymbol{\theta},\mathbf{y})} &= \frac{p(\mathbf{x},\boldsymbol{\theta},\mathbf{y})/p(\mathbf{y})}{p(\mathbf{x}|\boldsymbol{\theta},\mathbf{y})} && (5.16)\\ &\propto \frac{p(\mathbf{x},\boldsymbol{\theta},\mathbf{y})}{p(\mathbf{x}|\boldsymbol{\theta},\mathbf{y})} && (5.17)\\ &\approx \frac{p(\mathbf{x},\boldsymbol{\theta},\mathbf{y})}{\tilde{p}_G(\mathbf{x}|\boldsymbol{\theta},\mathbf{y})}\Big|_{\mathbf{x}=\mathbf{x}^*(\boldsymbol{\theta})} && (5.18)\\ &=: \tilde{p}(\boldsymbol{\theta}|\mathbf{y}), && (5.19) \end{aligned}$$

where $\tilde{p}_G(\mathbf{x}|\boldsymbol{\theta}, \mathbf{y})$ denotes the Gaussian approximation of $p(\mathbf{x}|\boldsymbol{\theta}, \mathbf{y})$ and $\mathbf{x}^*(\boldsymbol{\theta})$ its mode. Expression (5.18) is equivalent to the Laplace approximation of a marginal posterior distribution presented in Tierney and Kadane (1986). The INLA approach of Rue et al. (2009) is based on Gaussian approximations to densities that are of the form

$$p(\mathbf{x}) \propto \exp\Big(-\frac{1}{2}\,\mathbf{x}'\,\mathbf{Q}\,\mathbf{x} + \sum_{i\in\mathcal{I}} g_i(x_i)\Big), \tag{5.20}$$

where $g_i(x_i) = \log(p(y_i|x_i, \boldsymbol{\theta}))$ and $\mathbf{Q}$ is the precision matrix of $\mathbf{x}$. The Gaussian approximation $\tilde{p}(\mathbf{x})$ is obtained by matching the curvature of the density and the values of $\mathbf{x}$ at the mode. The computation of the mode is performed iteratively with a Newton-Raphson algorithm known as scoring algorithm or its variant, the Fisher scoring algorithm. For details on this procedure see Rue et al. (2009).

For computing $p(x_i|\boldsymbol{\theta}, \mathbf{y})$, Rue et al. (2009) discuss three different alternatives. A straightforward strategy would be to use the Gaussian approximation $\tilde{p}_G(x_i|\boldsymbol{\theta}, \mathbf{y})$, which is a computationally cheap way. As $\tilde{p}_G(\mathbf{x}|\boldsymbol{\theta}, \mathbf{y})$ was already computed within the exploration of $\tilde{p}(\boldsymbol{\theta}|\mathbf{y})$, the only remaining task would be the additional computation of the marginal variances. The drawback of the Gaussian approximation is that there can be errors in the location of the posterior density and/or errors produced by a lack of skewness (Rue and Martino, 2007).

As a way to improve the Gaussian approximation, Rue et al. (2009) propose a Laplace approximation. As in most cases $\mathbf{x}$ has more elements than the hyperparameter vector $\boldsymbol{\theta}$, the computation is more expensive. When rewriting $\mathbf{x}$ as $\mathbf{x} = (x_i, \mathbf{x}_{-i})$ and with the identity $p(\mathbf{x}_{-i}|x_i, \boldsymbol{\theta}, \mathbf{y}) \cdot p(x_i|\boldsymbol{\theta}, \mathbf{y}) = p((x_i, \mathbf{x}_{-i})|\boldsymbol{\theta}, \mathbf{y})$ Laplace approximation yields

$$\begin{aligned}
p(x_i|\boldsymbol{\theta}, \mathbf{y}) = \frac{p((x_i, \mathbf{x}_{-i})|\boldsymbol{\theta}, \mathbf{y})}{p(\mathbf{x}_{-i}|x_i, \boldsymbol{\theta}, \mathbf{y})} &= \frac{p(x_i, \mathbf{x}_{-i}, \boldsymbol{\theta}, \mathbf{y})/p(\boldsymbol{\theta}, \mathbf{y})}{p(\mathbf{x}_{-i}|x_i, \boldsymbol{\theta}, \mathbf{y})} && (5.21)\\
&\propto \frac{p(x_i, \mathbf{x}_{-i}, \boldsymbol{\theta}, \mathbf{y})}{p(\mathbf{x}_{-i}|x_i, \boldsymbol{\theta}, \mathbf{y})} && (5.22)\\
&\approx \frac{p(x_i, \mathbf{x}_{-i}, \boldsymbol{\theta}, \mathbf{y})}{\tilde{p}_{GG}(\mathbf{x}_{-i}|x_i, \boldsymbol{\theta}, \mathbf{y})}\Big|_{\mathbf{x}_{-i}=\mathbf{x}^*_{-i}(x_i,\boldsymbol{\theta})} && (5.23)\\
&=: \tilde{p}_{LA}(x_i|\boldsymbol{\theta}, \mathbf{y}), && (5.24)
\end{aligned}$$

where $\tilde{p}_{GG}(\mathbf{x}_{-i}|x_i, \boldsymbol{\theta}, \mathbf{y})$ is the Gaussian approximation to $\mathbf{x}_{-i}|x_i, \boldsymbol{\theta}, \mathbf{y}$ and $\mathbf{x}^*_{-i}(x_i, \boldsymbol{\theta})$ denote the values of $\mathbf{x}_{-i}$ at its mode. The Gaussian approximation $\tilde{p}_{GG}$ is different from the conditional density corresponding to $\tilde{p}_G(\mathbf{x}|\boldsymbol{\theta}, \mathbf{y})$. The Gaussian approximation $\tilde{p}_{GG}$ in Equation (5.23) needs to be recomputed for each x_i and $\boldsymbol{\theta}$, as the precision matrix of $\mathbf{x}_{-i}|x_i, \boldsymbol{\theta}, \mathbf{y}$ depends on x_i and $\boldsymbol{\theta}$. This re-computation makes the above described approximation $\tilde{p}_{LA}(x_i|\boldsymbol{\theta}, \mathbf{y})$ very expensive. Therefore, modifications are necessary that make the approximation computationally more feasible. For details on this see Rue et al. (2009).

The third alternative discussed in Rue et al. (2009) is a simplified Laplace approximation $\tilde{p}_{SLA}(x_i|\boldsymbol{\theta}, \mathbf{y})$. It can be obtained via a Taylor series expansion of $\tilde{p}_{LA}(x_i|\boldsymbol{\theta}, \mathbf{y})$ around $x_i = \mu_i(\boldsymbol{\theta})$. This procedure corrects the Gaussian approximation $\tilde{p}_G(x_i|\boldsymbol{\theta}, \mathbf{y})$ for location and skewness resulting in sufficiently accurate approximation for many common observational models. This procedure yields a high computational benefit. See Rue et al. (2009) for details on the simplified Laplace approximation.

INLA starts exploring the marginal joint posterior for the hyperparameters, $\tilde{p}(\boldsymbol{\theta}|\mathbf{y})$, to locate its mode. Within a grid search, a set of J 'relevant' points $\{\boldsymbol{\theta}_k\}$ along with a corresponding set of weights $\{\Delta_k\}$, $k = 1, \ldots, J$, is determined to approximate $\tilde{p}(\boldsymbol{\theta}|\mathbf{y})$. The marginal posteriors $\tilde{p}(\theta_j|\mathbf{y})$ can be computed by using interpolation based on the computed 'relevant' values. Then for each $\boldsymbol{\theta}_k$ the conditional posteriors $\tilde{p}(x_i|\boldsymbol{\theta}_k, \mathbf{y})$ of the latent Gaussian field components are evaluated on a grid of selected x_i values. By employing numerical integration, the marginal posteriors of the latent Gaussian field can be determined as

$$\tilde{p}(x_i|\mathbf{y}) \approx \sum_{k=1}^{J} \tilde{p}(x_i|\boldsymbol{\theta}_k, \mathbf{y})\, \tilde{p}(\boldsymbol{\theta}_k|\mathbf{y})\, \Delta_k. \tag{5.25}$$

For further details on the INLA procedure and its applications see Rue et al. (2009). They also discuss the error rates of INLA and its performance in some simulated and real examples. A short introduction to the INLA as well as to the SPDE method (see Section 5.3) within the `R-INLA` package together with examples of applications can be found in Blangiardo et al. (2013).

5.3 The SPDE approach

In Section 5.1 the concepts of continuous random fields (such as Gaussian fields) and of finite-dimensional random fields admitting conditional independence properties (such as Gaussian Markov random fields) were introduced.

In many applications a continuously indexed field is an appropriate model with an intuitive interpretation of the resulting field. When the aim is, for example, to predict surface temperature over Germany, it is of interest to issue predictions at any arbitrary location, including locations where no observations are available.

While GFs are very popular in geostatistics (Cressie, 1993; Stein, 1999), have good analytic properties and a natural interpretation, as they are specified in terms of the mean and covariance function, they are hampered with the 'big n problem' for high dimensions n:

The general cost of factorizing dense $n \times n$ (covariance) matrices is $\mathcal{O}(n^3)$. As GFs are parameterized in terms of (dense) covariance matrices, this results in high computational costs. Especially in Bayesian hierarchical models, where often repeated computations are necessary to perform simulation-based model fitting, this can be very slow.

In contrast, GMRFs have good computational properties, which is of high importance in Bayesian inference. As they are parameterized in terms of a (sparse) precision matrix, they allow for fast numerical algorithms as, for example, sparse matrix routines (Rue and Held, 2005). These advantages of GMRFs have been enhanced with the recent development of INLA (Rue et al., 2009) (see the previous Section 5.2), as INLA allows for fast and accurate Bayesian inference in latent Gaussian models.

Despite these good computational properties, current statistical models based on GMRFs are relatively simple. The reason is that it is not straightforward to parameterize the precision matrix of a GMRF to obtain the desired behavior in terms of marginal correlations and variances. This difficulty is due to the need of constructing a positive definite precision matrix to obtain a positive definite covariance matrix as its inverse. It is not in general evident how this global positive definiteness constraint influences the parametrization of the full conditionals.

Much research has been conducted in the direction of finding GMRF approximations for common covariance functions in geostatistics and using these approximations instead of the respective GFs. However, past approaches were restricted to relatively simple domains like lattices or toruses and included time consuming computational steps themselves.

Nonetheless, the research so far leads to a general strategy tackling the computational problems of a GF:
Employ a GF on a set of locations $\{\mathbf{s}_i\}$ to set up the statistical model and construct a discretized version of this GF with covariance matrix Σ. Then find an appropriate GMRF representation of this GF (with local neighbourhood and precision matrix $\mathbf{Q}$) and perform the computations for model fitting with the GMRF representation. This strategy is only reasonable in case it is possible to compute this GMRF representation at any desired location and in case the computation of the GRMF representation is fast enough that an overall computational-speed up is achieved in comparison to treating the GF directly.

This leads to the question, when does such a GMRF representation exist and how can it be obtained?

Lindgren et al. (2011) showed that for certain GFs with a Matérn covariance function on $\mathbb{R}^d$ as defined in Equation (5.26), a direct GMRF representation is available. An explicit

representation can be obtained via a certain stochastic partial differential equation (SPDE), having GFs with Matérn covariance function as a solution when driven by Gaussian white noise. The result can be formulated as a basis function representation with piecewise linear basis functions and Gaussian weights defined on a triangulation of the original domain, inducing the specific Markovian dependence structure. A clear advantage of this approach is that the covariance function is implicitly defined through the SPDE equation and needs not be computed directly. Further, the GMRF representation can be constructed from the basic SPDE equation in a straightforward way, without the need of going into the theory of stochastic differential equations in general.

Surprisingly, the basic (stationary) SPDE approach can easily be extended to cover more involved modelling problems. Lindgren et al. (2011) and Bolin and Lindgren (2011) discuss extensions to Matérn fields on manifolds, to non-stationary fields, to fields with oscillating or anisotropic covariance functions, and to non-separable space-time models. Even for these complex models an explicit GMRF representation may still be available.

Because of its usefulness in model fitting and its direct connection to the INLA methodology, the SPDE method was implemented within the `R-INLA` package, see `www.r-inla.org` for references and tutorials on INLA as well as SPDE-INLA. Although some of the implementation is still in progress, many SPDE features are already directly usable within the `R-INLA` package and can be combined with the standard INLA features. Examples for applications of the SPDE-INLA methodology to environmental data that are related to the approach presented in Chapter 5 can be found in Cameletti et al. (2013) and Blangiardo et al. (2013).

In the following, basic facts about the SPDE approach presented in Lindgren et al. (2011) are summarized.

Stationary SPDEs

Before introducing the SPDE equation connecting the parameters of a GF with Matérn covariance function to the parameters of the GMRF representation, the general definition of the Matérn covariance function is recapitulated.

The Matérn covariance function is defined as (Matérn, 1986; Gelfand et al., 2010; Guttorp and Gneiting, 2008)

$$C(\mathbf{s}, \mathbf{s}') = \frac{\sigma_C^2}{2^{\nu-1}\,\Gamma(\nu)}\,(\kappa||\mathbf{s}-\mathbf{s}'||)^{\nu}\,K_{\nu}(\kappa||\mathbf{s}-\mathbf{s}'||), \tag{5.26}$$

where $\mathbf{s}, \mathbf{s}' \in \mathbb{R}^d$, $||\cdot||$ denotes the Euclidean norm, K_ν is the modified Bessel function of

the 2nd kind and order $\nu > 0$, $\kappa > 0$ is a scaling parameter, and $\sigma_C^2 = \frac{\Gamma(\nu)}{\Gamma(\nu+d/2)(4\pi)^{d/2}\kappa^{2\nu}}$ is the marginal variance.

The parameter ν describes the smoothness of the underlying process in terms of mean-square differentiability, in that the associated Gaussian sample paths are m-times differentiable if and only if $m < \nu$. In most applications ν is set to a fixed value, as its identification is difficult. The scaling parameter κ is related to the range parameter r describing the distance at which $X(\mathbf{s})$ and $X(\mathbf{s}')$, the values of the underlying GF at $\mathbf{s}$ and $\mathbf{s}'$, become nearly independent. Lindgren et al. (2011) use the empirically derived relationship $r = \sqrt{8\nu}/\kappa$ throughout their paper, which corresponds to correlations near 0.1 at distance r, for all ν.

As discussed in Section 5.1, the Matérn covariance family belongs to the class of isotropic covariance functions, so according to (5.3), (5.26) can be written as

$$C(\mathbf{h}) = \frac{\sigma_C^2}{2^{\nu-1}\,\Gamma(\nu)}\ (\kappa||\mathbf{h}||)^\nu\ K_\nu(\kappa||\mathbf{h}||) = g(||\mathbf{h}||), \tag{5.27}$$

where $g(t) = \frac{\sigma_C^2}{2^{\nu-1}\,\Gamma(\nu)}\ (\kappa t)^\nu\ K_\nu(\kappa t)$ with $g \in G_\infty$, as noted in Section 5.1, as Matérn covariance functions are valid for all dimensions $d \geq 1$.

Lindgren et al. (2011) noted that a GF $X(\mathbf{s}) = \{X(\mathbf{s}), \mathbf{s} \in D \subseteq \mathbb{R}^d\}$ with Matérn covariance function (called Matérn field in the following) is a solution to the linear fractional SPDE

$$(\kappa^2 - \triangle)^{\alpha/2}\, X(\mathbf{s}) = W(\mathbf{s}), \tag{5.28}$$

where $\mathbf{s} \in \mathbb{R}^d$, $\alpha = \nu + d/2$, $\kappa > 0$, $\nu > 0$, $\triangle = \sum_{i=1}^d \frac{\partial^2}{\partial \mathbf{s}_i^2}$ defines the Laplace operator, and W is spatial white noise with unit variance. For all $\kappa > 0$ and $\nu > 0$, the solution $X(\mathbf{s})$ of the SPDE is a Matérn field. Familiar special cases are for example $\kappa = 1$, $\nu = \frac{1}{2}$, where the resulting field has exponential covariance function, or the limiting solution $\nu \to \infty$, $\kappa = \frac{1}{2\sqrt{\nu}}$, leading to a squared exponential covariance function.

For the special limiting cases $\kappa \to 0$ or $\nu \to 0$, the SPDE still has well-defined solutions. However, they do not have covariance functions belonging to the Matérn family.

The parameter α in (5.28) is the power of the fractional differential operator. Its value influences the structure of the precision matrix Q of the GMRF approximation. The differential operator can be defined through its spectral properties by using the Fourier transform definition of the fractional Laplace operator in $\mathbb{R}^d$. Whittle (1954) and Whittle (1963) show that the wave number spectrum of a stationary solution is given as

$$S(\mathbf{k}) = (2\pi)^{-d}\,(\kappa^2 + ||\mathbf{k}||^2)^{-\alpha}, \tag{5.29}$$

where $\mathbf{k} \in \mathbb{R}^d$. A Matérn field has the continuous Markov property (Rue and Held, 2005), if and only if and only if the spectrum of $X(\mathbf{s})$ is proportional to the inverse of a non-negative symmetric polynomial $P(k)$: $S_X(\mathbf{k}) \propto P(\mathbf{k})^{-1}$ (Rozanov, 1977). In the case of the SPDE (5.28), the Matérn field $X(\mathbf{s})$ is Markov if $\alpha = \nu + d/2$ is an integer.

To obtain a discrete approximation to the SPDE solution, the spatial domain is discretized into some kind of (regular or irregular) grid or lattice. Lindgren et al. (2011) investigate a regular two-dimensional lattice as well as an irregular triangulation of the domain of interest. For the regular lattice, the approximate solution to the SPDE can be obtained in a straightforward way, see Lindgren et al. (2011) for the details. However, a triangulation where the size and the number of the triangles can be adapted to the situation at hand, is more flexible and more appropriate in practical applications. Lindgren et al. (2011) derive explicit results for triangulated domains on $\mathbb{R}^2$. However, their results are valid on $\mathbb{R}^d$ as well as on general manifolds. They refer to these generalizations in the appendix.
The following overview is restricted to the results on $\mathbb{R}^2$, as this case is needed for the application in this chapter. The triangulation is obtained by dividing $\mathbb{R}^2$ into a set of non-intersecting triangles, where two triangles have at most one common edge or corner. The three corners of a triangle are called vertices. A set of initial vertices (the locations of the observations) is used and further vertices are added to satisfy overall constraints. Those points where triangle vertices meet are called nodes. The set of all nodes of the triangulation is denoted by $\{1, \ldots, n\}$. In the final triangulation, each node is surrounded by a set of triangles having this node as a vertex. Lindgren et al. (2011) propose the use of the so called Delaunay triangulation, which maximizes the minimum interior angle of each triangle. This method makes transitions between large and small triangles smoother. For further details and references see Lindgren et al. (2011).
After a Delaunay triangulation of the domain of interest is obtained, the stochastic weak solution to the SPDE (5.28) is approximated in a discretized space spanned by a set of chosen basis functions ψ_k. The stochastic weak solution can be obtained by requiring

$$\{\langle \phi_j, (\kappa^2 - \triangle)^{\alpha/2} X \rangle, j = 1, \ldots, m\} \stackrel{D}{=} \{\langle \phi_j, W \rangle, j = 1, \ldots, m\} \tag{5.30}$$

for any set of test functions $\{\phi_j\}$. Here, X is the Matérn field and W spatial white noise as in Equation (5.28), $\langle f, g \rangle = \int f(\mathbf{u}) g(\mathbf{u}) \, \mathrm{d}\mathbf{u}$, $\mathbf{u} \in \mathbb{R}^2$, denotes the inner product of f and g, and D refers to equality in distribution.

The approximation of the stochastic weak solution to the SPDE is constructed via a finite element representation of the Matérn field. It is a weighted sum

$$X_n(\mathbf{s}) := \sum_{k=1}^{n} \psi_k(\mathbf{s}) \, w_k. \tag{5.31}$$

of basis functions $\psi_k(\mathbf{s})$, $k = 1, \ldots, n$, $\mathbf{s} \in \mathbb{R}^2$, where the w_k are Gaussian distributed weights and n is the total number of nodes in the triangulation. Lindgren et al. (2011) choose the basis functions ψ_k to be continuous piecewise linear with support on the triangles that are attached to the node k. That is, ψ_k is a linear function from $\mathbb{R}^2$ to $[0, 1]$ such that $\psi_k(r) = 1$ for $r = k$ and $\psi_k(r) = 0$ for $r \neq k$, $r, k = 1, \ldots, n$.

With this choice of the functions ψ_k, the weights w_k determine the values of the field $X_n(\mathbf{s})$ at locations $\mathbf{s} \in \mathbb{R}^2$ that correspond to the nodes $k = 1, \ldots, n$ in the triangulation, while values in the interior of the triangles can be determined by linear interpolation. Therefore, the distribution of the continuous field $X(\mathbf{s})$, $\mathbf{s} \in \mathbb{R}^2$, is already fully determined by the joint distribution of the weights $\mathbf{w} = (w_1, \ldots, w_n)'$.

To obtain the approximate solution X_n it is required to determine the distribution of the weights w_k such that it fulfills the stochastic weak formulation (5.30) of the SPDE for a specific set of test functions with $m = n$. The specific choice of the test functions in relation to the basis functions influences the approximation properties of the resulting representation. Lindgren et al. (2011) choose $\phi_k = (\kappa^2 - \triangle)^{1/2}\, \psi_k$ for $\alpha = 1$ and $\phi_k = \psi_k$ for $\alpha = 2$. The solutions for $\alpha \geq 3$ are obtained recursively: For $l = 3, 4, \ldots, \alpha$, set $\alpha = l$ on the left-hand side of Equation 5.28, replace the right-hand side with a field generated from the case $l - 2$ and choose $\phi_k = \psi_k$. Details on this recursive approach can be found in the appendix of Lindgren et al. (2011).

Lindgren et al. (2011) show that the vector of weights $\mathbf{w} = (w_1, \ldots, w_n)'$ is a GMRF with mean zero and precision matrix $\mathbf{Q}_\mathbf{w}$, which is a function of the Matérn covariance parameter κ^2. This establishes an explicit link between the parameters of the covariance function of the Matérn field and the elements of the precision matrix $\mathbf{Q} = \mathbf{Q}_\mathbf{w}$ of the GMRF representation.

The approximate solution X_n obtained in the above described way converges weakly to the true solution X. Weak convergence of a bounded sequence $X_n \in L^2$ is defined as:

$$\begin{aligned} \mathrm{E}(\langle f, X_n \rangle) &\rightarrow \mathrm{E}(\langle f, X \rangle), \\ \mathrm{Cov}(\langle f, X_n \rangle, \langle g, X_n \rangle) &\rightarrow \mathrm{Cov}(\langle f, X \rangle, \langle g, X \rangle), \end{aligned}$$

for any $f, g \in L^2$ (say: $X_n \overset{d(L^2)}{\rightarrow} X$).

With this definition it can be shown that if X is the stochastic weak solution to the SPDE $(\kappa^2 - \triangle)^{\alpha/2} X = W$ and $X_n = \sum_{k=1}^{n} \psi_k\, w_k$ the finite weak approximation (5.31) to X,

then:

$$X_n \overset{d(L^2)}{\to} X,$$
$$(\kappa^2 - \triangle)^{\alpha/2} X_n \overset{d(L^2)}{\to} (\kappa^2 - \triangle)^{\alpha/2} X.$$

See the appendix in Lindgren et al. (2011) for proofs.

Extension to nonstationary SPDEs

The above described approach is suitable for stationary models. However, the SPDE approach can easily be extended to handle non-stationary models with the advantage that a GMRF approximation is still available in explicit form.

To introduce non-stationarity, the parameters in the SPDE (5.28) are allowed to depend on the location $\mathbf{s} \in \mathbb{R}^d$. Instead of introducing a non-constant innovation variance in the spatial white noise $W(\mathbf{s})$, an additional scaling parameter τ is introduced and the non-stationary SPDE can be written as

$$(\kappa(\mathbf{s})^2 - \triangle)^{\alpha/2} (\tau(\mathbf{s}) X(\mathbf{s})) = W(\mathbf{s}). \tag{5.32}$$

As in the stationary version, $\mathbf{s} \in \mathbb{R}^d$, $\alpha = \nu + d/2$, $\kappa > 0$, $\nu > 0$, $\triangle = \sum_{i=1}^{d} \frac{\partial^2}{\partial \mathbf{s}_i^2}$ and W is spatial white noise. By letting one or both of the SPDE parameters vary by location, non-stationarity is achieved in the resulting model.

A special case that is very useful is when the SPDE parameters vary slowly with $\mathbf{s}$, as in a low-dimensional log-linear representation like

$$\log(\kappa^2(\mathbf{s})) = \sum_i \beta_i^{(\kappa^2)} B_i^{(\kappa^2)}(\mathbf{s}), \tag{5.33}$$
$$\log(\tau(\mathbf{s})) = \sum_i \beta_i^{(\tau)} B_i^{(\tau)}(\mathbf{s}), \tag{5.34}$$

with basis functions $\{B_i^{(\cdot)}(\cdot)\}$ that are smooth over the domain of interest. This approach preserves the local interpretation of the SPDE (5.32) as a Matérn field, while the actual form of the non-stationarity covariance function might be unknown.

Further extensions of the SPDE model

Apart from the extension to non-stationarity, Lindgren et al. (2011) introduce five further extensions of the basic SPDE method. In the following, these approaches are mentioned

briefly. Bolin and Lindgren (2011) additionally introduce a more general way of constructing the SPDE to obtain many of the extensions described in Lindgren et al. (2011). The development of these more complex SPDE variants is currently a very active line of research.

The first extension is to consider solutions to the SPDE on manifolds. Here, the main interest are Matérn fields on spheres, as this is particularly useful for the analysis of global spatial and spatio-temporal data. A second extension is employing a complex-valued version of the basic SPDE (5.28), resulting in oscillating covariance functions. In the complex version two independent white noise processes and an oscillation parameter ϑ are introduced. With increasing ϑ, oscillations are increasing. In the special case of $\vartheta = 0$, the regular Matérn covariance function is recovered. A third extension is achieved by allowing a non-isotropic Laplace operator and adding a directional derivative term. This results in an anisotropic SPDE model. It is a generalization of the non-stationary SPDE model described above, which locally behaves isotropic, while globally behaving non-stationarily. The last extension presented by Lindgren et al. (2011) is a non-separable space-time SPDE model. While the construction of non-separable covariance functions is difficult, a non-separable SPDE model can be obtained easily using locally specified parameters. A simple but useful example is the transport-diffusion equation. Allowing the parameters of this equation to vary with space and/or time yields a large class of interesting models.

5.4 Markovian EMOS (MEMOS)

This section introduces a spatially adaptive extension of the standard univariate EMOS post-processing method described in 2.1.1. The aim of this extension is to improve the predictive performance of EMOS by incorporating spatial dependence structures. The proposed extension is called Markovian EMOS due to the Markovian dependence structure present in the model.

The current version of MEMOS is developed for temperature, which can be assumed to have a normal distribution. In its current form however, it can be employed for the postprocessing of other weather quantities with a normal distribution as well. As the MEMOS model utilizes the SPDE and the INLA methodology and both methods are not restricted to normally distributed responses, it is in general possible to extend the idea of MEMOS to weather quantities with other than a normal distribution. However, as the basic EMOS method itself is designed for a normally distributed weather quantity and only a few extensions exist, it might be necessary either to develop an EMOS version suitable for the desired distribution

or to choose another univariate postprocessing method and develop an extension similar to the one described below.

In the following, the MEMOS method designed for the weather quantity surface temperature is described. Possible modifications are mentioned in the discussion.

Let $\{Y(\mathbf{s}), \mathbf{s} \in D \subseteq \mathbb{R}^2\}$ be a spatial process representing surface temperature in a domain $D \subseteq \mathbb{R}^2$ for a fixed date t, where $t = 1, \ldots, T$. As the current model is not considering temporal dependence structures, the time index t is omitted in the notation below. In this analysis, the domain D corresponds to the spatial region of Germany, defined in terms of the respective longitude and latitude values covering that region. In the following, a finite set of spatial indices $\mathbf{s} = s_1, \ldots, s_N$ is employed for model estimation, representing the locations of the observation stations in Germany. For convenience, the random variables $Y(\mathbf{s})$ at the locations $\mathbf{s} = s_1, \ldots, s_N$ are denoted by $Y_1, \ldots, Y_N$ and the respective realizations by $y_1, \ldots, y_N$. In analogy, the random variables $\varepsilon(\mathbf{s})$ at $\mathbf{s} = s_1, \ldots, s_N$ introduced in (5.35) are denoted by $\varepsilon_1, \ldots, \varepsilon_N$.

The general form of the proposed extension has a similar appearance as the basic regression equation of the standard EMOS model (2.1):

$$Y_s = \gamma + a_s + b_{1,s}\, x_{1,s} + \ldots + b_{m,s}\, x_{m,s} + \varepsilon_s, \tag{5.35}$$

$$\varepsilon_s \sim \mathrm{N}(0, \sigma^2), \tag{5.36}$$

with $s = 1, \ldots, N$. Here, $x_{1,s}, \ldots, x_{m,s}$ denote the m raw ensemble members at location s, Y_s is the random variable representing surface temperature at location s as introduced above, γ an overall fixed effect intercept parameter, and ε_s a normal variable representing measurement error at location s. Furthermore, ε_s is assumed to have variance σ^2, and ε_s is independent of $\varepsilon_{s'}$ for $s \neq s'$. Moreover, a_s and $b_{1,s}, \ldots, b_{m,s}$ are the additive and multiplicative bias-correction parameters of the postprocessing model.

The special feature about this proposed extension is that a_s and $b_{1,s}, \ldots, b_{m,s}$ are no fixed-effects parameters as in (2.1), but assumed to be realizations of latent GFs $\mathbf{a}(s)$, $\mathbf{b}_1(s)$, ..., $\mathbf{b}_m(s)$ with Matérn covariance function as defined in (5.26). The proposed method aims at utilizing the SPDE approach described in Section 5.3, meaning that GMRF representations $\mathbf{a}_n(s)$, $\mathbf{b}_{n,1}(s), \ldots, \mathbf{b}_{n,m}(s)$ will be obtained via the approximation of the stationary solution to the SPDE (5.28). Due to the Markovian structure induced by the GMRF representations of the GFs, the EMOS extension is called Markovian EMOS, in short MEMOS. For the basic MEMOS method developed and applied in this work, the simplest stationary form of the SPDE (5.28) is employed. However, research to modify MEMOS by utilizing the extended

SPDE versions discussed in Lindgren et al. (2011) is planned and will be referred to in the discussion.

The model (5.35) is designed for a non-exchangeable forecast ensemble (as e.g. the UWME ensemble discussed in Section 4). In the case study conducted in the following sections, however, the exchangeable ECMWF ensemble is employed. As already mentioned in Section 2.1.1, in case of an exchangeable ensemble, the multiplicative bias-correction parameters can be chosen as $b_{1,s} = \ldots = b_{m,s} = b_s$ for all s and the model (5.35) reduces to

$$Y_s = \gamma + a_s + b_s \bar{x}_s + \varepsilon_s, \tag{5.37}$$

$$\varepsilon_s \sim \mathrm{N}(0, \sigma^2), \tag{5.38}$$

where all parameters have the same meaning as in (5.35), while $\bar{x}_s = \frac{1}{m} \sum_{k=1}^{m} x_{ks}$ represents the mean over all m ensemble members x_{ks}, $k = 1, \ldots, m$ at location s, and again a_s and b_s are realizations of latent GFs $\mathbf{a}(s)$ and $\mathbf{b}(s)$, for which GMRF representations are constructed via the stationary SPDE equation (5.28).

In the current MEMOS version, the Gaussian error process $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_N)'$ is spatially unstructured, so that

$$\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_N)' \sim \mathrm{N}_N(\mathbf{0}, \sigma^2 \mathbf{I}_N),$$

with a spatially constant variance term σ^2, introduced in Equation (5.35), on the diagonal. Here, $\mathrm{N}_N(\mathbf{0}, \sigma^2 \mathbf{I}_N)$ denotes the N-dimensional normal distribution with mean vector zero and covariance matrix equal to $\sigma^2 \mathbf{I}_N$, where $\mathbf{I}_N$ is the identity matrix of dimension N. In contrast to this, the standard EMOS model (2.3) includes the (spatially varying) ensemble spread in the variance term for the error process. Employing the non-stationary SPDE model (5.32) results in spatially varying SPDE parameters that correspond to a non-stationary precision matrix of the GMRF representation of the fields $\mathbf{a}(s)$ and $\mathbf{b}(s)$ assumed for the bias-correction parameters in the MEMOS model. However, this model still lacks a spatially varying variance in the error process $\boldsymbol{\varepsilon}$ in a similar way to the EMOS model. To achieve a more complex version of MEMOS with a variance term similar to the EMOS model (2.3), some of the other SPDE extensions need to be considered in future research.

As the SPDE approach as well as the INLA procedure are utilized to fit the MEMOS model, a slight reformulation of the model (5.37) is presented that directly corresponds to the model specification within INLA. The model (5.37) can equivalently be written as

$$Y_s = \eta_s + \varepsilon_s, \tag{5.39}$$

where η_s denotes the linear predictor utilized to formulate a latent Gaussian model within

the INLA framework as described in (5.9). For model (5.37), η_s is defined as

$$\eta_s = \gamma + a_s + b_s \, \bar{x}_s. \tag{5.40}$$

Then it follows directly from (5.39), (5.38), and from the INLA theory outlined in Section 5.2 that Y_s given η_s, σ^2, and the (deterministic) m ensemble forecasts at location s has the normal distribution

$$Y_s | \eta_s, \sigma^2, x_{1,s}, \ldots, x_{m,s} \sim \mathrm{N}(\eta_s, \sigma^2), \tag{5.41}$$

where η_s is part of the vector $\mathbf{x}$ of all model parameters, and σ^2 is the variance of the error term ε_s introduced in (5.37), which is part of the hyperparameter vector $\boldsymbol{\theta}_2$ associated with the response variables Y_s. The hyperparameter vectors $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ will be introduced below.

According to the theory presented in Rue et al. (2009) and briefly summarized in Section 5.2, the vector $\mathbf{x}$ contains all model parameters present in the linear predictor (5.9), including the linear predictor components themselves. The vector of hyperparameters $\boldsymbol{\theta}_1$ associated with $\mathbf{x}$ contains additional parameters that are required to define the prior distribution of $\mathbf{x}$. For the linear predictor (5.40) of the MEMOS model the vector $\mathbf{x}$ is defined as

$$\mathbf{x} = (\gamma, \mathbf{f} = (f_1(s), f_2(s)), \boldsymbol{\eta}) = (\gamma, \mathbf{a}, \mathbf{b}, \boldsymbol{\eta}). \tag{5.42}$$

Here, the components γ, $\boldsymbol{\eta}$, $\mathbf{a}$ and $\mathbf{b}$ of $\mathbf{x}$ are defined in analogy to Section 5.2. For the linear predictor (5.40) of the MEMOS model defined in Equations (5.37), the sub-components of $\mathbf{x}$ are $\boldsymbol{\eta} = (\eta_s, s = s_1, \ldots, s_N)$, $\mathbf{a} = (a_s, s = s_1, \ldots, s_N)$, and $\mathbf{b} = (b_s, s = s_1, \ldots, s_N)$. The parameter γ is a scalar component not depending on the locations $s_1, \ldots, s_N$. The associated vector of hyperparameters is

$$\boldsymbol{\theta}_1 = (\kappa_a, \kappa_b, \sigma^2_{C,a}, \sigma^2_{C,b}). \tag{5.43}$$

This vector contains the parameters that are required to define the prior distributions of the sub-components $\mathbf{a}$ and $\mathbf{b}$ of $\mathbf{x}$, the other components of $\mathbf{x}$ are not contributing additional parameters. More precisely, κ_a and $\sigma^2_{C,a}$ are the parameters of the Matérn covariance function (5.26) of the GF $\mathbf{a}(s)$, and κ_b and $\sigma^2_{C,b}$ are the Matérn covariance function (5.26) parameters of the GF $\mathbf{a}(s)$. In Equations (5.47) and (5.48) introduced later, the hyperparameters of the GFs $\mathbf{a}(s)$ and $\mathbf{b}(s)$ are collected within sub-vectors $\boldsymbol{\theta}_a = (\kappa_a, \sigma^2_{C,a})$ and $\boldsymbol{\theta}_b = (\kappa_b, \sigma^2_{C,b})$, respectively.

Besides the hyperparameter vector $\boldsymbol{\theta}_1$ associated with $\mathbf{x}$, Rue et al. (2009) define a second hyperparameter vector $\boldsymbol{\theta}_2$ containing all additional parameters required to define the conditional distribution of the response variables Y_s that are not already part of $\boldsymbol{\theta}_1$. From the

definition of the MEMOS model in Equations (5.37) and (5.39) it follows that the only additional parameter required to specify the conditional distribution of the Y_s that is not part of $\boldsymbol{\theta}_1$ is the variance σ^2 of the error term ε_s introduced in Equation (5.37) and (5.38). Therefore, it follows that $\boldsymbol{\theta}_2 = \sigma^2$.

According to Equation (5.11) and Rue et al. (2009) the full vector of hyperparameters $\boldsymbol{\theta}$ for the MEMOS model is then given as

$$\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2) = (\kappa_a, \kappa_b, \sigma^2_{C,a}, \sigma^2_{C,b}, \sigma^2). \tag{5.44}$$

In the general INLA framework, the mean $\boldsymbol{\mu}$ of the distribution of $\mathbf{Y}|\mathbf{x}, \boldsymbol{\theta}_2$ is linked to the linear predictor through a link function $g(\cdot)$ via $g(\mu_i) = \eta_i$, $i \in \mathcal{I}$. As the conditional distribution $Y_s|\eta_s, \sigma^2$ is Gaussian in the MEMOS model, the link function g is the identity mapping, hence $\mu_s = \eta_s$. Further, in the Gaussian case the INLA calculations are exact (the only approximation is performed when computing $\tilde{p}(\boldsymbol{\theta}|\mathbf{y})$ as explained in Equation (5.18)).

The η_s are components of the parameter vector $\mathbf{x}$ defined in (5.42) and σ^2 is part of the hyperparameter vector $\boldsymbol{\theta}$ defined in (5.44). Both, $\mathbf{x}$ and $\boldsymbol{\theta}$ are assumed to be random vectors that are assigned a prior distribution in the INLA framework. INLA then approximates the posterior distribution of each component of $\boldsymbol{\theta}$ and utilizes the results to approximate the posterior distribution of each component of $\mathbf{x}$, see Section 5.2 for details on the INLA procedure. In practise it is therefore necessary to first obtain samples from the marginal posteriors of η_s and σ^2 to be able to generate a sample from the conditional distribution (5.41). When having estimated the MEMOS model for day t from previous training data, the SPDE-INLA output provides approximations of the posterior predictive margins of η_s, σ^2 and the other components of $\mathbf{x}$ and $\boldsymbol{\theta}$ at all locations s that are present on prediction day t. By default, the output further provides summary statistics (posterior mean, posterior standard deviation, and the posterior 0.25-, 0.5- and 0.75-quantiles) for each approximated predictive posterior margin.

To generate samples from an unconditional version of (5.41), the first step is to obtain the desired number of samples $\eta_{s,i}$, $i = 1 \ldots,$ N.sample, from the posterior predictive marginal of η_s for every location s present on the currently considered prediction day t. In a second step, the posterior mean $\hat{\mu}_{\sigma^2}$ of the posterior predictive marginal for σ^2 for the current prediction day t is extracted from the INLA output. As σ^2 was assumed to be constant over s, its posterior predictive marginals (and therefore its posterior mean values) are identical at all locations s. This posterior mean value is then used to obtain a sample of size N.sample from a $\text{N}(0, \hat{\mu}_{\sigma^2})$ distribution. This sample from this distribution is denoted by Z_i, $i = 1, \ldots,$ N.sample. The last step then sets $Y_{s,i} = \eta_{s,i} + Z_i$. For the univariate case

study conducted in Sections 5.9.1 and 5.9.2, N.sample $= 5,000$ samples were obtained separately for each location s present on the current prediction day t. Algorithm 5.1 presents pseudo-code that sums up the procedure of sampling from an approximation to the MEMOS predictive distribution.

The above defined MEMOS model lacks the parameter vector $\boldsymbol{\beta}$ present in the general definition of $\boldsymbol{\eta}$ in (5.9), meaning covariates with fixed-effects parameters are not included, except for the overall intercept γ. Besides the intercept, two spatially structured effects defined through functions f_l are included, $a_s = f_1(s)$ and $b_s = f_2(s)$. As already mentioned, the structured effects a_s and b_s are assumed to be realizations of GFs $\mathbf{a}(s)$ and $\mathbf{b}(s)$, for which GMRF representations $\mathbf{a}_n(s)$ and $\mathbf{b}_n(s)$ are obtained via the SPDE equation (5.28) by setting up a basis function representation (5.31) defined on a triangulation with n nodes:

$$\mathbf{a}_n(s) = \sum_{k=1}^{n} \psi_k(s)\,\tilde{a}_k, \tag{5.45}$$

$$\mathbf{b}_n(s) = \sum_{k=1}^{n} \psi_k(s)\,\tilde{b}_k. \tag{5.46}$$

Here, the ψ_k are piecewise linear basis functions, while $\tilde{a}_k$ and $\tilde{b}_k$ both are zero mean Gaussian distributed weights, as described in Section 5.3. According to the theory of Lindgren et al. (2011) presented in Section 5.3, the weights $\tilde{\mathbf{a}} = (\tilde{a}_1, \ldots, \tilde{a}_n)'$ and $\tilde{\mathbf{b}} = (\tilde{b}_1, \ldots, \tilde{b}_n)'$ are Gaussian with mean zero:

$$\tilde{\mathbf{a}} = (\tilde{a}_1, \ldots, \tilde{a}_n)' \sim \mathrm{N}(0, \mathbf{Q}^{-1}(\boldsymbol{\theta}_\mathbf{a})) \tag{5.47}$$

$$\tilde{\mathbf{b}} = (\tilde{b}_1, \ldots, \tilde{b}_n)' \sim \mathrm{N}(0, \mathbf{Q}^{-1}(\boldsymbol{\theta}_\mathbf{b})), \tag{5.48}$$

where n is the total number of nodes in the triangulation that was used to approximate the solution to the SPDE, and $\mathbf{Q}(\boldsymbol{\theta}_\mathbf{a})$ and $\mathbf{Q}(\boldsymbol{\theta}_\mathbf{b})$ are the precision matrices of the Gaussian distributions of $\tilde{\mathbf{a}}$ and $\tilde{\mathbf{b}}$. Each precision matrix depends on a set of hyperparameters $\boldsymbol{\theta}_a = (\kappa_a, \sigma^2_{C,a})$ and $\boldsymbol{\theta}_b = (\kappa_b, \sigma^2_{C,b})$, linking the GMRF representations (5.45) and (5.46) to the GFs $\mathbf{a}$ and $\mathbf{b}$ via the elements of $\boldsymbol{\theta}_a$ and $\boldsymbol{\theta}_b$, which are the parameters of the respective Matérn covariance functions (5.26). The two vectors $\boldsymbol{\theta}_a$ and $\boldsymbol{\theta}_b$ are also part of the overall hyperparameter vector $\boldsymbol{\theta}$ of the MEMOS model defined in (5.44).

Despite utilizing finite element versions $\mathbf{a}_n(s)$ in (5.45) and $\mathbf{b}_n(s)$ in (5.46), the SPDE approach provides a full representation of the processes $\mathbf{a}(s)$ and $\mathbf{b}(s)$ varying continuously over the considered domain D, at any desired location s, not only at those locations corresponding to the nodes $s = 1, \ldots, n$ used to approximate the SPDE solution. The precision matrices $\mathbf{Q}_a$ and $\mathbf{Q}_b$ of the Gaussian distributed weights $\tilde{\mathbf{a}}$ and $\tilde{\mathbf{b}}$ are obtained from the approximate solution to the SPDE (5.28) at the nodes $k = 1, \ldots, n$ of the triangulation that

is employed. With these precision matrices the distributions of the two Gaussian weights vectors are fully determined. Due to the properties of the basis function representations (5.45) and (5.46), the precision matrices of the Gaussian weights vectors are identical to the precision matrices of the GMRF representations $\mathbf{a}_n(s)$ and $\mathbf{b}_n(s)$.

To utilize the SPDE model, the parameter α in the SPDE equation (5.28) needs to be chosen to run the `SPDE-INLA` procedure. As mentioned in Section 5.3, α needs to be an integer to obtain a GMRF representation of a GF with Matérn covariance (5.26). In the current version of the `R-INLA` package, the values $\alpha = 1$ and $\alpha = 2$ are available. Preliminary case studies for different types of meshes revealed that a MEMOS model with $\alpha = 1$ yields better predictive performance than $\alpha = 2$. As $\alpha = 2$ corresponds to a smoother spatial field than $\alpha = 1$, it seems that a field with a too smooth spatial structure is not appropriate for the temperature data in Germany analyzed here. Therefore all subsequent analyzes are performed with $\alpha = 1$.

The variance parameter σ^2 that is a component of the hyperparameter vector (5.44) is representing the variance of the conditional distribution of the response variables, $Y_s|\eta_s, \sigma^2$, in Equation (5.41). Within the `R-INLA` framework this parameter is internally parameterized as $\sigma^2 = \frac{1}{q\tau}$, where $q > 0$ is a fixed scaling factor, that can be specified by the user and τ is the precision parameter of the conditional distribution of $Y_s|\eta_s, \sigma^2$ in (5.41). `R-INLA` internally approximates the posterior marginal for the hyperparameter τ instead of the one for σ^2. However, `R-INLA` provides a way to transform the approximated posterior marginal for τ into a marginal for σ^2 with the call `inla.tmarginal()`. The default value for the scaling is $q = 1$, which implies a nonspecific scaling constant over all data points. It is possible to choose any other value of q that is constant over all data points as well as a non-constant scaling by specifying q as a vector of values for each observation case handed over to the `INLA` procedure.

The MEMOS model (5.41) employed here was chosen to have a constant scaling $q = 1$, resulting in a spatially constant variance σ^2. Choosing q to be a vector equal to the ensemble variance S^2 or the ensemble standard deviation S at each date-station pair present in the observations would result in a similar variance approach like in the EMOS model (2.3), only without an additive variance parameter c as in (2.3). The absence of an additive parameter in the variance term can lead to a very large value of $\frac{1}{q\tau}$ in case q is small, while $\frac{1}{q\tau}$ gets very small in case q is large. Thus, when choosing $q = S^2$ or $q = S$ the variance σ^2 can get very extreme when the ensemble spread or variance takes too large or too small values. This might result in a predictive distribution that lacks calibration and sharpness, as the spread of the predictive distribution increases too much. A small empirical case study showed that

$q = 1$ yields better predictive performance than $q = S$ or $q = S^2$ in the MEMOS model (5.41). To obtain a MEMOS variant with a variance approach as in (2.3), other features of the SPDE-INLA approach need to be investigated. This could be the subject of further research.

For a final overview, the pseudo code in Algorithm 5.1 sums up the procedure of estimating the MEMOS model and generating samples from an approximation to the MEMOS predictive distribution (5.41).

Algorithm 5.1 Procedure to obtain samples from the predictive MEMOS distribution

for t in 1:T **do**

Set up the training data consisting of the n.train days prior to prediction day t

Construct mesh from merged training-prediction data and set up SPDE object

for s in $(s_1, \ldots, s_N)$ **do**

Approximate posterior predictive marginals of the hyperparameter-vector,
namely $\sigma^2, \kappa_a, \kappa_b, \sigma^2_{C,a}, \sigma^2_{C,b}$
and of the components of the latent field $\mathbf{x}$, namely γ, a_s, b_s, η_s
for location s and prediction day t with R-SPDE-INLA
by utilizing current training data prior to day t

Extract posterior predictive mean $\hat{\mu}_{\sigma^2}$ of posterior marginal for hyperparameter component σ^2 for prediction day t, where $\hat{\mu}_{\sigma^2}$ is constant over all locations s

for i in 1:N.sample **do**

sample $\eta_{s,i}$ from the posterior predictive marginal of η_s
for location s and prediction day t
sample $Z_i \sim \mathrm{N}(0, \hat{\mu}_{\sigma^2})$
put $Y_{s,i} \leftarrow \eta_{s,i} + Z_i$

end for

end for

end for

return $(Y_{t,s,1}, \ldots, Y_{t,s,\text{N.sample}})$, $t = 1, \ldots, T, s = s_1, \ldots, s_N$

5.5 Multivariate dependence structures

The INLA-procedure itself only provides approximations for the posterior distributions of all marginal components of the parameter vector $\mathbf{x}$. An approximation of the joint posterior dis-

tribution of the complete $\mathbf{x}$ vector or of a sub-vector like for example the full linear predictor vector $\boldsymbol{\eta}$ would be computationally expensive and is therefore not performed directly.

In case of the MEMOS postprocessing model, the marginal posterior distributions of the linear predictor components η_s are the posterior predictive MEMOS postprocessed distributions for temperature at location s (on a fixed day t) whose predictive performances are to be assessed with verification methods. Therefore the $\mathbf{x}$ components η_s are of main interest here. Approximations of the posterior predictive distributions along with their posterior mean, standard deviation and a set of quantiles are automatically provided within the INLA-procedure for each component η_s (as well as for all other marginal components of $\mathbf{x}$), namely at the observation locations $s = s_1, \ldots, s_N$. These posterior approximations can directly be assessed with univariate verification methods.

As the MEMOS model indirectly incorporates spatial dependencies through the SPDE models placed on the bias-correction parameters and therefore is not a univariate postprocessing model in the classical sense, it is of interest to assess the predictive quality of a multivariate predictive MEMOS distribution with multivariate verification methods.

The following paragraphs describe different ways of supplying the basic MEMOS method with a multivariate dependence structure in a multi-stage procedure. To compare the predictive performance of the combined MEMOS versions with the performance of other postprocessing methods, the other postprocessing methods in question are supplied with the dependence structures as well.

For MEMOS as well as for other univariate postprocessing methods, the multi-stage procedure works as follows:

1. Perform univariate postprocessing of the raw ensemble with the desired postprocessing model separately for each location $s = s_1, \ldots, s_N$.
2. Combine the N spatial margins within the chosen dependence structure and sample from the resulting multivariate distribution.

The samples obtained from these different multivariate predictive distributions are assessed with multivariate verification methods.

Independence

The first and most basic type of dependence structure is the Independence structure. Pro-

viding a univariate postprocessing method with the Independence structure results in a multivariate predictive distribution with independent margins. For each combined multivariate postprocessing distribution, the margins are equal to the marginal predictive distributions obtained with the respective univariate postprocessing method.

For the raw ensemble, a sample from the multivariate raw ensemble Independence distribution is obtained by randomly permuting the univariate m ensemble members at each considered observation location s and each fixed day t separately.

For providing one of the univariate postprocessing methods as e.g. EMOS or MEMOS with the Independence structure, one simply obtains a large number of random samples independently from each marginal postprocessed distribution.

Ensemble Copula Coupling (ECC)

Here, the univariate postprocessing methods are combined within the ensemble copula coupling (ECC) procedure, which recovers the multivariate dependence structure present in the original raw ensemble forecast. The ECC method is a multi-stage procedure as well, where in a first step the margins are postprocessed individually with a method of choice. Then a sample of the size m of the original ensemble is drawn from each predictive distribution. This sample is reordered according to the rank order structure of the raw ensemble. See Section 2.1.6 for details.

For the raw unpostprocessed ensemble, raw ensemble ECC is simply the raw ensemble itself, as it already possesses the rank structure employed for the reordering.

In case of applying the ECC procedure to the margins of a postprocessing method such as EMOS or MEMOS, the regular ECC steps can be performed:

1. Postprocess the margins individually with the method of choice.
2. Obtain a sample of size m, the size of the original ensemble, from each postprocessed margin by applying the ECC-Q sampling strategy described in Equation (2.23) in Section 2.1.6.
3. Reorder the samples from each margin according to the rank structure of the corresponding raw ensemble forecasts.
4. The result is a multivariate ECC ensemble of the size m of the original ensemble.

5.6 Data

The data used for the analysis contains observations and forecasts for surface (2-m) temperature in Germany. The time period considered ranges from February 2010 to April 2011. The observation data was provided by the German Weather Service (DWD). For the case study, observations that were initialized at 00 UTC are employed, which corresponds to 1am local time in Germany, and to 2am local time during the daylight saving period. For temperature observations there are 518 stations available during the considered time period.

The forecast ensemble is provided by the European Centre for Medium-Range Weather Forecasts (ECMWF, Molteni et al. (1996); Palmer et al. (1997); Buizza (2006)). The ensemble comprises of one control forecast (not used here) and 50 (exchangeable) members. The forecasts are initialized at 00 UTC for different forecast horizons in 3 hour steps up to 144 hours. In the conducted case study the 24-h ahead ECMWF forecasts are utilized. The original ECMWF forecasts have been issued on a grid over Germany, at 31 km resolution. To use them in combination with the observations, bilinear interpolation of the forecasts from the four surrounding grid points to the station locations of interest is performed.

Panel (a) of Figure 5.1 shows the observation stations in Germany available for temperature, where stations that are considered as examples in the case study are marked in red. Details on the different case studies with the selected stations can be found in Sections 5.9.2, 5.9.3 and 5.9.4. The considered stations are the four North Sea Island stations Borkum, Norderney, Sylt, and Helgoland, Frankfurt and Baden-Baden in the South of Germany, eleven stations along the North Sea coastline, and finally three stations in the very east of Germany, Bertsdorf, Görlitz, and Bad Muskau. Panel (b) shows an example of a triangulation (also called mesh), namely the mesh that was constructed to estimate the MEMOS model from training data prior to October 3, 2010, which was then used to issue predictions on October 3, 2010. This specific day is considered later on in the case study as example for samples from different versions of the MEMOS method. Panel (c) shows the same mesh with additional orange points at the locations of the observation stations present on October 3, 2010. The mesh can be constructed automatically with the functions

```
inla.mesh.create(loc=, ...)
```

or

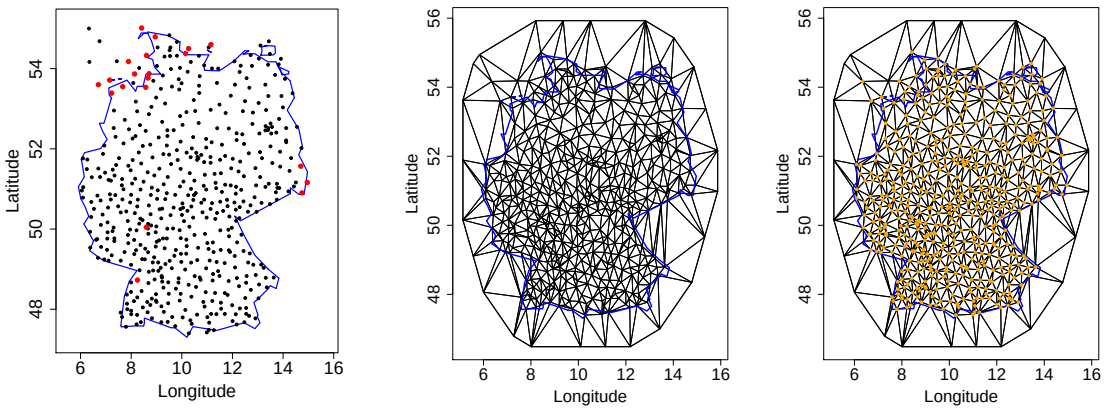
```
inla.mesh.create.helper(points=, ...)
```

within the `R-SPDE-INLA` package. To obtain a mesh with these functions, a matrix with the coordinates of the observation locations needs to be specified along with additional parameters controlling the appearance of the mesh, like for example the length of the triangle

edges or the size of the interior angles, which are controlled by the argument `refine`. The mesh in Figure 5.1 (b) was created by the call

```
inla.mesh.create(loc=locations, cutoff=0,
 refine=list(max.edge=5, min.angle=0.1)),
```

where the locations specified for `loc` are all observation locations present in the combined data set containing the training data and the data for the prediction day October 3, 2010. In Figure 5.1 (b), the training data that was used consists of the recent 25 days prior to October 3, 2010, namely all dates from September 8, 2010 to October 2, 2010. This means the combined training-prediction data set employed to construct the mesh consists of all dates from September 8, 2010 to October 3, 2010 and the observation locations present in this data set are the initial vertices in the triangulation algorithm performed by `inla.mesh.create()`. The `cutoff` value chosen to create the mesh implies that no locations that are close together will be merged to a single location. The `refine` argument gives additional constraints on the maximum edge length and the minimum interior angle in each triangle. For further details on the mesh parameters and their effect on the appearance of the mesh see Section 5.7.



(a) Observation stations

(b) MEMOS mesh on October 3, 2010 with 536 vertices and 1046 triangles

(c) MEMOS mesh on October 3, 2010 along with the station locations on that day

Figure 5.1: Available station locations for temperature, where the stations considered in the case study are displayed in red (a), along with the mesh used for MEMOS on October 3, 2010 constructed by R-SPDE-INLA (b), and the same mesh together with the locations of the observation stations on that day (c)

The density of the stations varies over Germany. In regions with more cities and high industrial development such as, for example, the Rhine-Neckar region, the Rhine-Ruhr region and

big cities like Berlin or Munich, there are more observation stations than in the rural areas of north Germany. In addition to this there are a few stations outside of Germany and several offshore stations on the islands in the German North Sea and Baltic Sea. The flexibility of the triangulation approach is useful, as it is possible to construct triangles of different size, adapting locally to the density of observation locations. The mesh in Figure 5.1 (b) constitutes a middle course between large and small triangles, where the total number of vertices (536 vertices) is close to the number of observation stations (508 stations). Panel (c) shows that most of the vertices in the mesh correspond to real station locations. Only a few locations are not lying on a triangle vertex. However, in addition to those vertices overlapping with actual observation locations, there are several vertices with no observation locations close by. These vertices are mainly outside of Germany. So a triangulation with larger triangles outside of Germany and smaller triangles in the interior is reasonable for constructing the approximate solution of the SPDE (5.28), which rules the dependence structure of the GMRF representation.

5.7 Mesh configuration

To get smoother transitions between regions with a dense network of station locations and regions with only a few observation stations, control parameters in the mesh functions can be set. Depending on the specific choice of these mesh parameters, the desired mesh configuration can be created.

There are two main parameters influencing the mesh configuration. The `cutoff` value specifies the minimum distance between two points. Points that are closer to each other than the value chosen for `cutoff` are merged to a single location in the mesh construction. The `refine` parameter specifies whether the mesh construction should be refined by setting it either to `TRUE` or `FALSE`. Alternatively, the `refine` argument allows to specify constraints for the refinement, like the maximum length of the edges and the minimum interior angle of the triangles. By choosing different values for these two parameters different mesh types can be obtained: Meshes with very small or very large triangles, meshes with smooth transitions between the triangle sizes, or meshes with an abrupt change of the triangle sizes. A small empirical case study showed that the mesh configuration has a large influence on the predictive quality of the estimated MEMOS model. Figure 5.2 shows examples of four different mesh configurations obtained by choosing different values for `cutoff` and `refine` for the station locations in Germany. All four meshes in Figure 5.2 were obtained by specifying the argument `loc` in `inla.mesh.create()` as the 518 distinct observation locations of the

full data set containing the dates from February 2010 to April 2011.

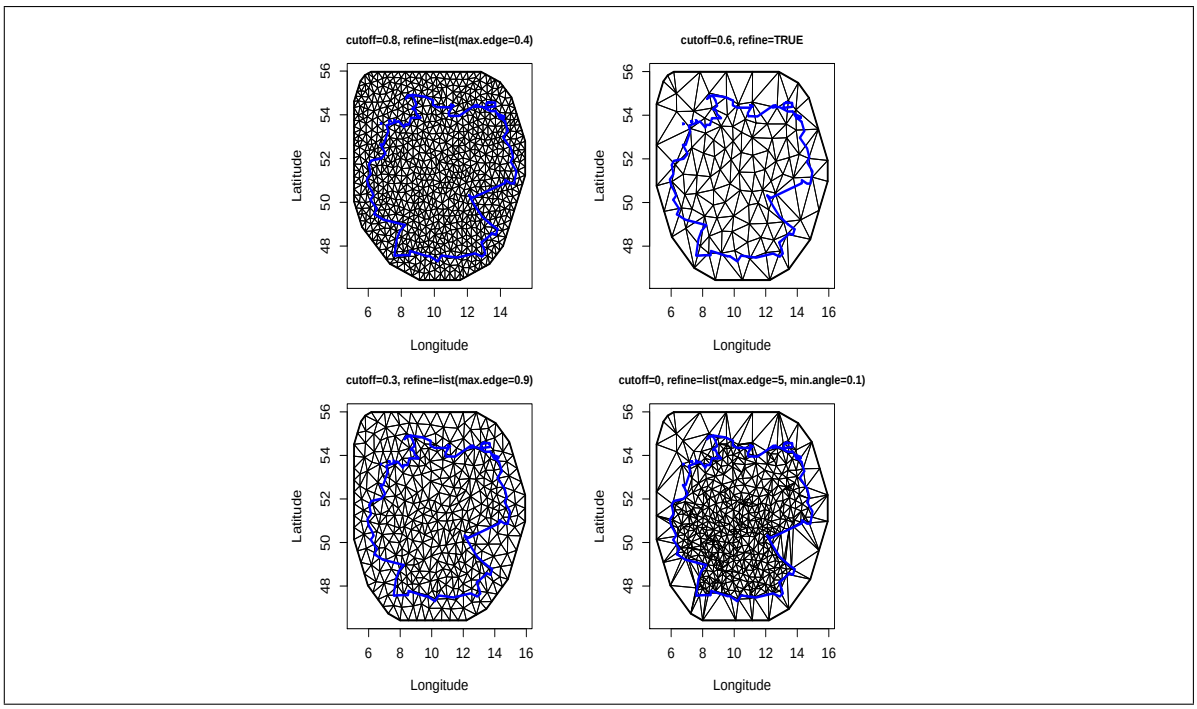


Figure 5.2: Mesh configurations with different parameters

The mesh in the top left panel of Figure 5.2 has a high `cutoff` value and allows for short edges only, resulting in a high number of small triangles. The call to obtain this mesh is

```
inla.mesh.create(loc=locations, cutoff=0.8,
        refine=list(max.edge=0.4)).
```

The mesh in the top right panel has a high `cutoff` value as well, but no constraints on the edge length. Therefore the mesh has less and larger triangles. While the transition between the triangle sizes is very smooth in the top left mesh, it is less smooth in the top right mesh. Here the triangles covering the borders of Germany are larger than in the interior. However, the transition is still smoother than in the bottom right mesh. This mesh is obtained via the call

```
inla.mesh.create(loc=locations, cutoff=0.5, refine=TRUE).
```

The bottom left mesh has a small `cutoff` value, yet the constraint on the edge length is not as tough as in the top left mesh. This results in somewhat larger triangles than in the top left mesh, while the transition between the triangle sizes in the top left mesh is still quite smooth. This mesh is produced via the call

```
inla.mesh.create(loc=locations, cutoff=0.3,
        refine=list(max.edge=0.9)).
```

The bottom right mesh has a `cutoff` value of zero and very soft constraints on the edge length. This choice of parameters results in a crude triangulation with a very rough transition between large and smaller triangles. The call for this mesh is

```
inla.mesh.create(loc=locations, cutoff=0,
 refine=list(max.edge=5, min.angle=0.1)).
```

It should be noted, that the arguments in this call are the same as for the mesh in Figure 5.1. The only difference is that the mesh in Figure 5.1 is constructed from the locations present in the combined data set containing the 25 training days September 8, 2010 to October 2, 2010 as well as the data for the prediction day October 3, 2010, as already described in Section 5.6. In contrast, the meshes in Figure 5.2 are constructed from the observation locations present in the complete data set containing dates from February 2010 to April 2011.

From the empirical inspection of different mesh types, a rule of thumb for the application within the MEMOS model was derived.
A mesh where the triangles and vertices reflect the structure of the observation stations yields better predictive performance. Where the density of stations is higher, small triangles should be employed, and in regions with very low density only a few and large triangles should be placed. Using an overall smooth triangulation with small triangles everywhere, regardless of the real density of stations, yields a deterioration of the predictive performance. A further experience is that the model yields better predictive performance when the number of vertices resembles the number of observation stations. The predictive performance deteriorates when the number of vertices becomes much larger than the number of observations available, such as in the top left panel of Figure 5.2. A mesh reflecting the structure of the observation stations, such as the one in the bottom right panel of Figure 5.2, yields the best results concerning predictive quality. A possible explanation is that the vertices of the mesh are the discrete points where the SPDE solution is approximated. If there are more locations where an approximation is computed than observations, this might have a negative effect on the quality of results. Especially in the case of a too large number of vertices not having any observations in their vicinity, the accuracy of the approximated SPDE solution at those vertices might deteriorate.

5.8 Training period

When choosing an appropriate training period for estimating the model parameters there is a trade-off. A short training period can adapt quickly to seasonally varying model biases,

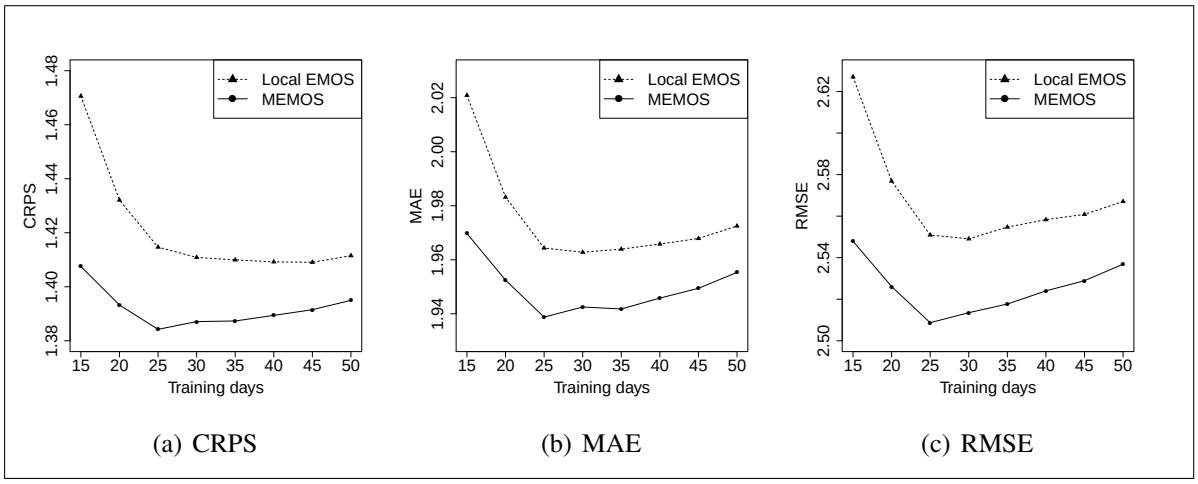
changes in the performance of the ensemble member models, and changes in environmental conditions. On the other hand, longer training periods provide more data for the estimation procedure and reduce the statistical variability in the estimation of the parameters.

In the postprocessing literature, a common practise is to use rolling training periods between 20 and 40 days. A rolling training period of length T consists of the most recent T days for which data is available. Therefore, it can consist of more than T calender days. To investigate the best length of a rolling training period for the MEMOS method, the predictive performance of MEMOS for 15, 20, 25, 30, 35, 40, 45, and 50 training days is investigated in terms of the CRPS, the MAE, and the RMSE, see Section 2.2. These training lengthes consist of the most recent 15, 20, 25, 30, 35, 40, or 50 days prior to the day on which a prediction is issued with the respective estimated postprocessing model. As the CRPS assesses sharpness and calibration simultaneously and is measured in the same unit as the observations, it is a reasonable choice to judge the performance of MEMOS for different training periods. However, the MAE and RMSE are additionally considered as a function of the length of the training period.

The scores for each training length were computed for the same prediction days, for a fair comparison. Therefore, the prediction days available for the longest period of 50 days are considered as prediction days in all training length settings. This approach ensures that the scores are based on the same dates for each training length. The set of test days therefore contains March 24, 2010 to April 30, 2011.

As the optimal length of training period for local EMOS, which is a natural benchmark for the MEMOS method, is not necessarily the same as for MEMOS, MEMOS might be favored when judging on the basis of the best training length for MEMOS. To have a fair comparison, the optimal training length for local EMOS is additionally investigated, for the same training lengths of 15, 20, 25, 30, 35, 40, 45, and 50 days as described above.

Figure 5.3 shows the CRPS, MAE, and RMSE values of local EMOS and MEMOS for all tested training periods.



(a) CRPS (b) MAE (c) RMSE

Figure 5.3: CRPS, MAE and RMSE for local EMOS and MEMOS plotted against length of training period

This analysis implies that the best training length for MEMOS is 25 days, as all scores have their minimum in that value. On the other hand, the different scores do not suggest a unique optimal value for local EMOS. A length of 30 days is suggested when considering MAE and RMSE, while the minimum CRPS value is attained for 45 days. Despite slightly different results concerning the optimal training length of both methods, Figure 5.3 reveals that the values of the three considered scores for MEMOS are always smaller than the values of local EMOS, for each of the considered training lengths.

In conclusion, a training length of 25 days is employed for all postprocessing methods analyzed in the case study.

5.9 Results

5.9.1 Aggregated univariate results

The first objective of the case study is to assess the predictive quality of the basic MEMOS method described in Section 5.4 in comparison to standard univariate postprocessing methods. This section presents results of the univariate methods aggregated over all available stations in Germany. As the considered postprocessing methods (including MEMOS) issue predictions for days after the last day of the rolling training period, the presented results are out-of-sample regarding time. However, as the predictive distributions include the prediction at all stations that were present in the training data as well, the results are in-sample

concerning the spatial component.

The complete data employed for the case study consists of February 2, 2010 to April 30, 2011, which are 453 days in total. As already mentioned in Section 5.8, where training period lengthes from 15 up to 50 days were assessed, the prediction days finally considered for verification are only those available for the longest training period of 50 days. These days comprise of all dates from March 24, 2010 on. Therefore only the days from March 24, 2010 to April 30, 2011 were considered for prediction in the subsequent analysis. This is a total of 403 days. Section 5.8 suggested 25 days as the optimal length of training period. The total number of distinct stations present in this data set is 518.

The univariate methods compared in the case study are the raw ensemble, global EMOS, local EMOS, and the proposed extension MEMOS. Their predictive performance is measured in terms of the CRPS, the MAE, and the RMSE, calibration is assessed in terms of VRH histograms (in case of the raw ensemble) and PIT histograms (in case of the predictive postprocessed distributions). For the raw ensemble, the ensemble versions of the scores are employed, for global and local EMOS, the predictive distribution is a normal distribution with estimated parameters. Therefore exact formulas for normal distributions can be employed. The posterior predictive MEMOS distribution is not exactly normal, so approximative versions of the scores are utilized.

The global and local EMOS predictive distributions are estimated in the standard way as described in Section 2.1.1 with a rolling training period. The estimation of the MEMOS model is performed automatically with the SPDE-INLA function according to the theory in Section 5.4. However, the generation of the the training and prediction data used for constructing the mesh and computing the approximate posterior predictive distributions for desired components of the parameter vector $\mathbf{x}$ defined in (5.42) is slightly more involved as in the case of EMOS. The procedure of data extraction is described in the following:

The data corresponding to the current 25 days rolling training period and the data corresponding to the associated prediction day (the day after the last training day, as 24-h forecasts are employed) are extracted.

At the beginning of each estimation-prediction step, both data parts are merged and all distinct observation stations contained in this training-prediction data are used to construct the mesh for this step. The mesh is reconstructed in each estimation-prediction step from the current training-prediction data, to take into account the specific spatial structure present on that day. The call to generate the mesh in each step is the same as for the mesh in Figure 5.1, namely

```
inla.mesh.create(loc=locations, cutoff=0,
 refine=list(max.edge=5, min.angle=0.1)),
```

where the locations in the `loc` argument in each estimation-prediction step are specified as all locations present in the combined data set containing the data of the current training period and the data for the current prediction day. In Figure 5.1 this combined data consists of the 25 training days September 8, 2010 to October 2, 2010 along with the data for October 3, 2010, as described in Section 5.6.

The data part corresponding to the current training period is used for estimating the MEMOS model parameters, the components x_i of the full parameter vector $\mathbf{x}$ in (5.42). INLA approximates those components x_i not defined through an SPDE model directly for all observation cases, here corresponding to the observation locations $s = s_1, \ldots, s_N$ handed over to INLA with the observation vector. The only parameter in (5.42) not connected to an SPDE model is the fixed effect intercept γ. As the linear predictor $\boldsymbol{\eta}$ depends on the two bias-correction parameters defined through an SPDE model, it is implicitly derived from the posterior distributions of those two parameters. The components of $\mathbf{a}$ and $\mathbf{b}$ are approximated via an SPDE model at the nodes $k = 1, \ldots, n$ of the current mesh. The SPDE-INLA procedure automatically interpolates them internally to the observation locations $s = s_1, \ldots, s_N$ to obtain approximations of the linear predictor components η_s at $s = s_1, \ldots, s_N$. By specifying appropriate linear combinations $\mathbf{B}\boldsymbol{\eta}$ within the `inla()` call as described in the third paragraph of Section 5.5, the approximations for the part of $\boldsymbol{\eta}$ corresponding to the current prediction day t and all locations s present on that day are extracted and displayed separately in the SPDE-INLA output. This part of the output can directly employed for computing verification measures for the current prediction day t.

Table 5.1 presents the verification results for the raw ensemble, global and local EMOS, and MEMOS in terms of the univariate scores CRPS, MAE, and RMSE. The scores are averaged over all prediction days and all available stations. Note that the local EMOS method was estimated by using a 25 days training period for each station, even if this corresponds to more than 25 calendar days.

Table 5.1 shows that the MEMOS method performs best in terms of all considered scores. The largest improvement can be seen for the CRPS. For all scores, global EMOS improves significantly on the raw ensemble, as it corrects for biases and dispersion errors in. However, global EMOS estimates only a single set of parameters for all stations and is thus not accounting for local differences. In contrast to this, local EMOS estimates a separate set of parameters at each considered station, resulting in better local calibration. As stations in quite different regions with a high variability in temperature are considered, the information

	CRPS	MAE	RMSE
Raw ECMWF	2.50	2.81	3.76
Global EMOS	1.79	2.49	3.24
Local EMOS	1.42	1.96	2.55
MEMOS	1.38	1.94	2.51

Table 5.1: Univariate predictive performance of the MEMOS methodology with $\alpha = 1$ in comparison to the raw ensemble, global and local EMOS, over all stations and dates in March 24, 2010 to April 30, 2011

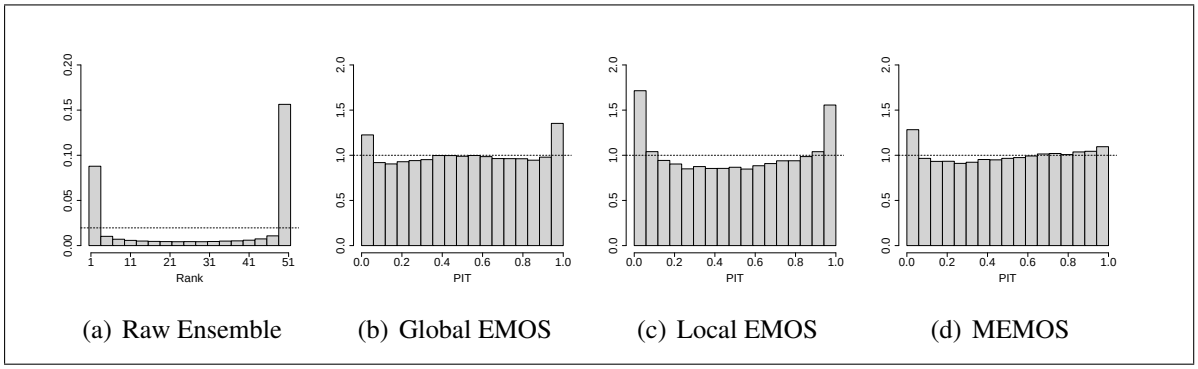
incorporated in the local EMOS parameters leads to improved scores. However, the improvement from global to local EMOS is not as large as the improvement from the raw ensemble to global EMOS. Going from local EMOS to MEMOS yields another small improvement. Performing any type of postprocessing has the largest effect on the aggregated results and yields a significant improvement in the scores, while employing more complex postprocessing methods in comparison to a simple model often yields only small improvements. As the raw ensemble provides only 50 discrete forecasts, the ECMWF results are computed over a distribution with a spatial dimension of approximately 500 with only 50 points to fill this entire space. A full predictive distribution as returned by global EMOS therefore results in a substantial improvement in comparison to the sparse and discrete ECMWF distribution. Local EMOS additionally accounts for local differences. The weather over the North Sea region is for example similar for all stations there, but different in comparison to the weather predominant in the slightly continental East Germany or the warm Rhine-Neckar Region. A postprocessing method like MEMOS that is able to borrow information from neighbouring stations through the Markovian dependence structure of the GMRF approximations can improve predictive performance further. MEMOS yields the best results in all scores, so introducing a Markovian dependence structure among the stations provides useful information that local EMOS cannot capture.

Figure 5.4 presents the verification rank histogram for the raw ECMWF ensemble (a) along with the PIT histograms for global (b) and local EMOS (c) and MEMOS (d). The ranks and PIT values are aggregated over all available stations and dates in the period of March 24, 2010 up to April 30, 2011. As the resolution of the verification rank histogram is quite high when using all 51 bins to classify the ranks, a slightly lower (1/3 of the original) resolution was applied by classifying the 51 possible outcomes for the ranks in 17 bins. The frequencies

of the ranks falling in each of the 17 bins are then plotted as usual. The PIT values in panels (b)-(d) are classified into 17 bins as well, to have better comparability to the rank histogram.

For the raw ensemble the verification rank histogram is obtained by computing the rank of the observation within the 50 member ensemble for each date-station pair. This partitions the real line into 51 bins. The occurring ranks are then classified in 17 bins as described above.

For the predictive distributions, PIT values are obtained by evaluating the estimated predictive cumulative distribution function for each date-station pair at the respective observation. In case of local and global EMOS, the PIT values are computed by evaluating the normal cumulative distribution function with parameters estimated from the respective EMOS method at the observations. For MEMOS, the PIT values are obtained by evaluating the empirical CDF at the observations, as the posterior predictive MEMOS distribution is not exactly normal.



(a) Raw Ensemble (b) Global EMOS (c) Local EMOS (d) MEMOS

Figure 5.4: Univariate verification rank histogram and PIT histograms over all stations and dates in March 24, 2010 to April 30, 2011

The VRH in panel (a) indicates a heavy underdispersion of the raw ECMWF ensemble and an additional bias in the forecasts, as the last bin is more occupied than the first one. This suggests a strong need for postprocessing. The PIT histogram for global EMOS in panel (b) already shows a quite good calibration. This is in line with the large improvement in the scores in 5.1. While the scores improve for local EMOS, the PIT in panel (c) exhibits a slight underdispersion, that is clearly more pronounced than in panel (b). However, a more or less pronounced underdispersion is typical for local EMOS, as it only uses the data of a single station to estimate the model parameters for this station. This can result in a too small variance of the predictive distribution, as the observations in the vicinity of the station are not taken into account. The PIT histogram for MEMOS in panel (d) exhibits the best calibration. The histogram is closer to uniformity than the histograms for global and local EMOS. The first and last bin are less pronounced for MEMOS than they are for global and local EMOS.

These calibration results line up with the results for the scores in Table 5.1.

5.9.2 Univariate results at single stations

This section presents univariate verification results for the raw ensemble and the three competitive postprocessing methods at selected single stations situated in different regions of Germany. The results at each considered station are aggregated over all available prediction days, containing the same dates from March 24, 2010 to April 30, 2011 that have been considered in Section 5.9.1. As in the previous section, the ranks and PIT values of the single station examples are classified into 17 equidistant bins to reduce the resolution of the histograms.

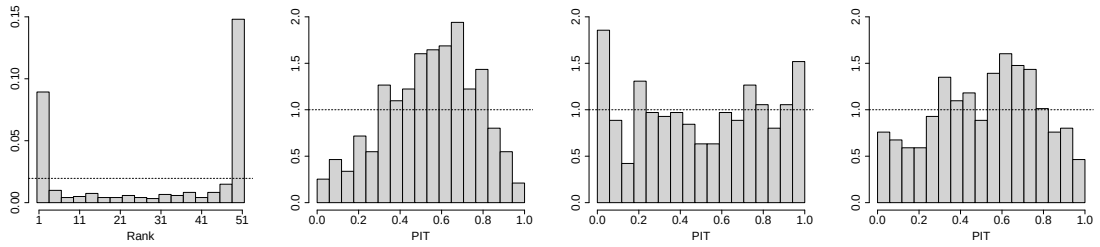
The first station considered is a station on the north sea island Borkum belonging to the East Frisia region and having a healthy offshore climate. Borkum is the westernmost island of the seven East Frisian islands, with a size of $31\,\text{km}^2$. The only town on the island is named Borkum as well. The elevation of the Borkum observation station is 3 m above sea level, and the site situated near the small airport of the island. Table 5.2 shows the scores for Borkum, Figure 5.5 shows the verification rank histograms.

	CRPS	MAE	RMSE
Raw ECMWF	1.59	1.89	2.35
Global EMOS	1.14	1.52	1.92
Local EMOS	1.08	1.49	1.93
MEMOS	1.06	1.47	1.86

Table 5.2: Univariate predictive performance at the station Borkum over all dates in March 24, 2010 to April 30, 2011

The results in Table 5.2 exhibit a similar structure to the overall results in Table 5.1. The largest improvement in all three scores is achieved when going from the raw ensemble to global EMOS. The local EMOS method estimating the bias-correction parameters locally at the station Borkum improves the CRPS and MAE further. The calibration of the predictive distribution for Borkum improves by estimating local parameters, resulting in improved CRPS and MAE. However, the RMSE deteriorates slightly for local EMOS. In combination with the shape of the local EMOS PIT histogram, this indicates that the local EMOS

predictive distribution still has a small bias paired with a too low spread in the distribution. For MEMOS however, all scores improve further. The ability of MEMOS to incorporate information from stations in the vicinity through a Markovian dependence structure makes it possible to account for the information of other stations in offshore regions and along the coastline of North Germany and by this allowing to learn more about the local setting. This reduces the bias of the predictive distribution for Borkum and simultaneously increases the spread in comparison to the too small spread of the local EMOS predictive distributions.



(a) Raw ensemble (b) Global EMOS (c) Local EMOS (d) MEMOS

Figure 5.5: Univariate verification rank histogram and PIT histograms for Borkum, over all dates in March 24, 2010 to April 30, 2011

Panel (a) of Figure 5.5 shows a heavy underdispersion of the raw ensemble at the station Borkum and an additional forecast bias. The PIT histogram in panel (b) indicates an overdispersion of the global EMOS predictive distribution. As global EMOS estimates only a single set of parameters for all stations, it is not calibrated for a specific station like for example Borkum, leading to a lack of calibration in the sense that the spread of the predictive distribution is too large in comparison to observed values. The global EMOS methods seems to overestimate the local variability present at an offshore island. However, in comparison to the raw ensemble, there is a huge improvement, which is visible in the scores in Table 5.2 as well. In contrast, the PIT histogram for local EMOS in panel (c) indicates a slight underdispersion of the predictive distribution, although not at all that severe as for the raw ensemble. This goes in line with the scores for the station Borkum. Compared to the histograms of global and local EMOS the histogram of MEMOS is much closer to calibration. It still exhibits a slight tendency to overdispersion. However, this is much less pronounced as for global EMOS.

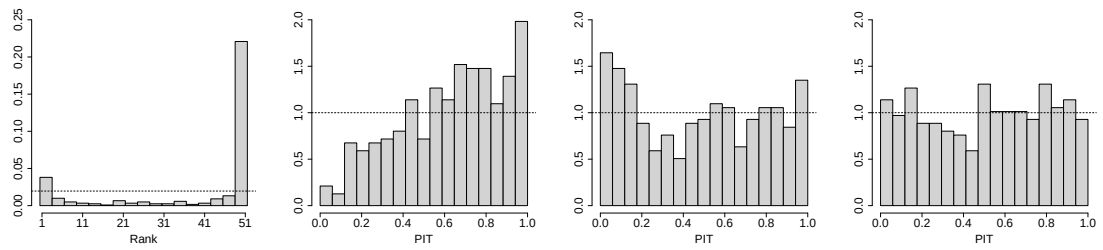
The next station considered is at Frankfurt airport, at an elevation of 113 m above sea level. Frankfurt is a city in the south-west of Germany, situated at the river Main. The area of Frankfurt is an area of very high traffic density, but on the other hand it is close to the Taunus, a popular local recreation area. It is the largest city of Hessen, famous for being a

financial and industrial center in Germany. Due to its central position within Germany, the airport of Frankfurt is one of the largest in the world. The climate in the Rhine-Main region is very mild and belongs to the warmest within Germany. Table 5.3 shows the scores for this station and Figure 5.6 the verification rank and PIT histograms.

	CRPS	MAE	RMSE
Raw ECMWF	2.10	2.41	2.90
Global EMOS	1.60	2.24	2.84
Local EMOS	1.32	1.86	2.31
MEMOS	1.28	1.83	2.26

Table 5.3: Univariate predictive performance at the station Frankfurt over all dates in March 24, 2010 to April 30, 2011

The results for Frankfurt are similar to the results at Borkum. The largest and most significant improvement in all three scores is when going from the raw ensemble to global EMOS. Panel (a) of Figure 5.6 shows the verification rank histogram for the raw ensemble, exhibiting heavy underdispersion and significant bias. The last bin is significantly higher than the first one. Panel (b) shows the PIT histogram for global EMOS. It displays the same skewed structure as panel (a) to some extent, however, much less pronounced. Moving from the raw ensemble to global EMOS corrects for the strong bias present in the raw ensemble forecasts and improves calibration. This effect is consistent with the highly improved scores. The PIT histogram of local EMOS in panel (c) displays a further improvement in calibration, although underdispersion is still present. However, the skewed form of the histogram is attenuated. The PIT histogram for MEMOS in panel (d) exhibits the best calibration. The U-shape is not visible any more and the histogram is close to uniformity. In comparison to the station Borkum the improvement achieved with MEMOS is slightly more distinct, the same holds for the improvements when moving vom the raw ensemble to global EMOS and from global to local EMOS. It should be noted that the weather on a North sea island as Borkum is quite special in comparison to other regions of Germany, like e.g. the Rhine-Main region around Frankfurt, where the weather is more stable than around the coast region of North Germany. Postprocessing can generally improve the shortcomings of the raw ensemble, however the improvement obtained by performing postprocessing seems to be larger in regions where the temperature or the weather in general is more stable.



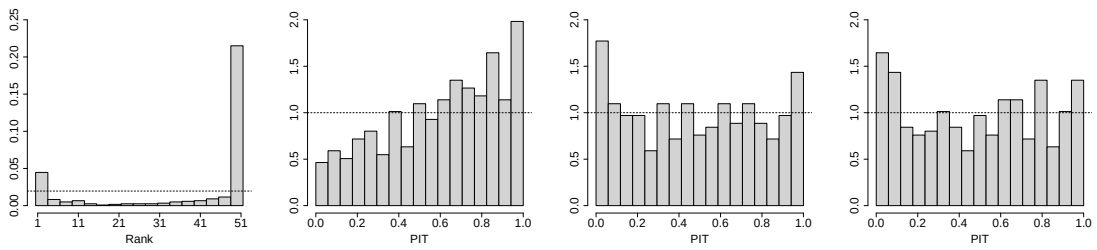
(a) Raw ensemble (b) Global EMOS (c) Local EMOS (d) MEMOS

Figure 5.6: Univariate verification rank histogram and PIT histograms for Frankfurt, over all dates in March 24, 2010 to April 30, 2011

The next station considered is even more southward than Frankfurt. The city of Baden-Baden is in the western part of the northern Black Forest, one of the largest forested areas within Germany and a very popular recreation area. The station is at 240 m above sea level. Baden-Baden is a world famous health resort with genuine mineral springs and with a not less famous casino. Table 5.4 shows the scores for this station and Figure 5.7 the verification rank and PIT histograms.

	CRPS	MAE	RMSE
Raw ECMWF	2.53	2.83	3.54
Global EMOS	1.80	2.50	3.27
Local EMOS	1.64	2.24	2.95
MEMOS	1.56	2.15	2.79

Table 5.4: Results for univariate predictive performance at the station Baden-Baden over all dates in March 24, 2010 to April 30, 2011



(a) Raw ensemble (b) Global EMOS (c) Local EMOS (d) MEMOS

Figure 5.7: Univariate verification rank histogram and PIT histograms for Baden-Baden, over all dates in March 24, 2010 up to April 30, 2011

The scores are similar to those at Frankfurt. All three scores improve significantly when going from the raw ensemble to global EMOS. The largest improvement can be seen for the CRPS. The verification rank and PIT histograms are shown in Figure 5.7. The raw ensemble is heavily underdispersed, quite similar to the raw ensemble for Frankfurt. As for Frankfurt, the PIT histogram for global EMOS at Baden-Baden displays a bias and a lack of calibration. However, compared to the raw ensemble, the calibration improves. In case of local EMOS, the PIT histogram for Baden-Baden exhibits underdispersion. All scores for MEMOS improve on local EMOS for Baden-Baden. The calibration is improved slightly in comparison to local EMOS, nonetheless, the PIT histogram of MEMOS still exhibits a trace of underdispersion. The additional information provided by the Markovian structure utilized by MEMOS constitutes an improvement in comparison to only utilizing the information of the station alone as local EMOS does.

The three examples show that the univariate MEMOS method outperforms the standard EMOS postprocessing methods with respect to the three considered aggregated verification scores CRPS, MAE, and RMSE.

The stations considered in the analysis had a very moderate elevation. A preliminary case study for a larger set of single stations revealed that the results deteriorate when considering stations with extreme elevation (e.g. stations on the mountains Zugspitze, Kahler Asten, Feldberg). In those cases MEMOS can perform worse than local EMOS, as MEMOS is not considering the station elevation as an explicit covariate in the model. Although local EMOS is not doing so either, its advantage is the separate parameter estimation for each station. Only data from the station in question is employed for parameter estimation, which improves the predictive performance at that specific station. However, when it comes to coherently predict the temperature for a (large) region, a postprocessing method accounting for the spatial dependence structure has the edge over a purely univariate method.

An extended postprocessing model considering other covariates explaining the response variable, such as elevation, land use type, or other weather quantities, might lead to further improvement in predictive performance. Extensions in this direction will be mentioned in the discussion.

5.9.3 Multivariate results

In this section the predictive performance of the univariate postprocessing methods provided with different types of multivariate dependence structure is compared.

The univariate methods to be combined with the dependence structures are the same as in the fully univariate analysis. Two types of dependence structure are considered for the analysis, the Independence structure and the ECC structure, which are described in Section 5.5. The predictive performance of these multivariate methods are analyzed with the multivariate verification methods described in Section 2.2. To inspect multivariate calibration, the multivariate rank histogram is used, which is described in Section 2.2.1. As in the univariate analysis in Sections 5.9.1 and 5.9.2, the possible 51 outcomes for the multivariate ranks are classified into 17 bins to reduce the resolution of the rank histogram. Further, the determinant sharpness (2.30) and the energy score (2.41) are employed.

For each of the multivariate raw ensemble versions, samples of the same size as the original ensemble, $m = 50$, are obtained. These samples are analyzed with the ensemble versions of the scores. The covariance matrix used for the determinant sharpness (2.30) is the empirical covariance matrix of the raw ensemble. The ensemble version of the energy score is given in Equation (2.42). When combining the postprocessing methods global EMOS, local EMOS and MEMOS with the Independence structure, 5,000 samples are obtained for each method. The multivariate ranks, the determinant sharpness, and the energy score are computed from the respective samples. For the determinant sharpness (2.30), the empirical covariance matrix of the respective postprocessing method is employed, which is a diagonal matrix with the estimated marginal variances on the diagonal. To compute the energy score from the 5,000 multivariate samples, the approximative version (2.43) is used. In case of combining global EMOS, local EMOS and MEMOS with the ECC structure, samples of the same size as the original ensemble, $m = 50$, are obtained, as for the multivariate raw ensemble versions. The samples from the different ECC variants are analyzed with the ensemble versions of the scores as well, as described above. Note that for all forecast methods and the corresponding observations only those days where used for verification, on which all of the selected stations are available. Those days on which at least one of the selected stations is missing were

	Borkum	Norderney	Helgoland	List, Sylt
Borkum	*3*	31	100	190
Norderney		*11*	70	165
Helgoland			*4*	98
List, Sylt				*26*

Table 5.5: Pairwise distances in km (upper triangle) between the stations on the North Sea islands Borkum, Norderney, Helgoland, and Sylt and their elevation in m (diagonal)

removed.

For analyzing multivariate predictive performance, several sets of stations in different regions of Germany are considered. The multivariate predictive performance over these chosen stations is analyzed with the multivariate verification methods described in Section 2.2.

The first example consists of stations situated on four North Sea islands: Borkum, Norderney, Sylt, and Helgoland. Borkum, Norderney, and Helgoland belong to the East Frisian islands, while Sylt is a North Frisian islands near the border to Denmark. Table 5.5 shows the pairwise distances between the four stations along with the elevation of each station. All four stations have an elevation close to sea level. While Borkum and Norderney are part of the East Frisian chain of islands and very close to each other, the position of Helgoland is much more offshore. The position of the stations within Germany is presented in panel (a) of Figure 5.1.

The distance between Borkum and Norderney is about the same value as the horizontal resolution of the ECMWF forecast grid. So the two islands may well be within the same grid box. The distances between Norderney and Helgoland as well as between Borkum and Helgoland and the distances of all other islands to Sylt are larger. Panel (b) and (c) of Figure 5.1 indicate further that in the presented mesh used for the MEMOS model, Borkum and Norderney (the two red dots side by side in the East Frisian Islands chain) are indeed located in two different triangles though sharing one edge, while Helgoland lies in a triangle at least sharing a vertex with the Norderney triangle.

Table 5.6 presents the energy score (ES) for the raw ensemble, global and local EMOS, and MEMOS, each provided with the twi types of dependence structures described above. Table 5.7 shows the respective values for the determinant sharpness (DS), while Figure 5.8

presents the multivariate rank histograms for all methods. The results for the ES show a general improvement when going down each column: When going from the raw ensemble to MEMOS, the results improve for each of the dependence structures. This is in line with the univariate results of Section 5.9.1. The improvement in multivariate predictive performance when going from the raw ensemble to global EMOS, to local EMOS and then to MEMOS holds both for the Independence and the ECC approach.

The results for the ES lead to the conclusion that for a fixed method of constructing the multivariate predictive distribution from the univariate margins of the respective postprocessing model, the best results are obtained when choosing MEMOS as the model for the margins. This is in line with the univariate results presented in Section 5.9.1.

The next question is that of the influence of the way the multivariate dependence structure is constructed. For the raw ensemble, the best results are obtained when providing it with the Independence structure. However, when applying any of the competitive postprocessing methods to the margins, the best results are obtained for the ECC structure.

These ES values can be explained as follows. The raw ensemble exhibits strong underdispersion and therefore lacks calibration in the univariate case. Passing from the ECC structure to the Independence structure decreases the sharpness of the predictive distribution, see Table 5.7. In this case the increased spread in the predictive distribution has a positive effect due to the lack of calibration in the raw ensemble, resulting in better ES values. However, when combining the EMOS or MEMOS margins with dependence structures, the effect is the other way around. The margins obtained from univariate global and local EMOS and MEMOS correct the raw ensemble for dispersion errors and display quite good calibration. In this case the decreased sharpness of the Independence approach in comparison to the ECC approach (see again Table 5.7) has a negative effect on the ES values, as the distributions already have a good univariate calibration.

The DS values in Table 5.7 complement and support the discussion of the ES values in Table 5.6. As already mentioned, the highest level of sharpness is obtained for the ECC dependence structure. The Independence structure exhibits the highest level of spread, therefore the lowest level of sharpness. As the margins are assumed to be independent this results in more variability in the joint distribution.

The columns of Table 5.7 show a specific pattern as well. For each of the dependence types, the raw ensemble is sharpest. This is reasonable, as the 50 ensemble members impart greater sharpness than a full predictive distribution estimated in the case of EMOS and MEMOS. Global EMOS reduces the sharpness significantly, du to the global parameter estimation

	Independence	ECC
Raw ECMWF	3.80	3.81
Global EMOS	2.76	2.66
Local EMOS	2.37	2.35
MEMOS	2.37	2.29

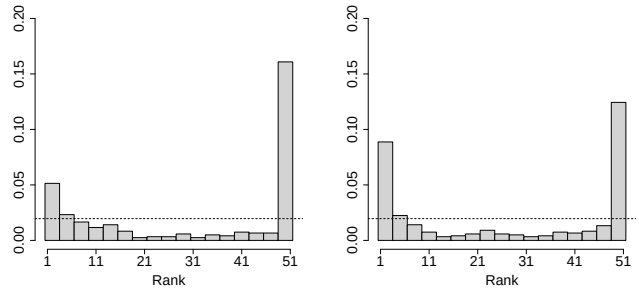
Table 5.6: Multivariate predictive performance measured with the energy score (ES) for the North Sea islands Borkum, Norderney, Helgoland, and Sylt, aggregated over all days in March 24, 2010 to April 30, 2011

	Independence	ECC
Raw ECMWF	0.63	0.39
Global EMOS	2.91	1.72
Local EMOS	1.43	0.84
MEMOS	2.30	1.36

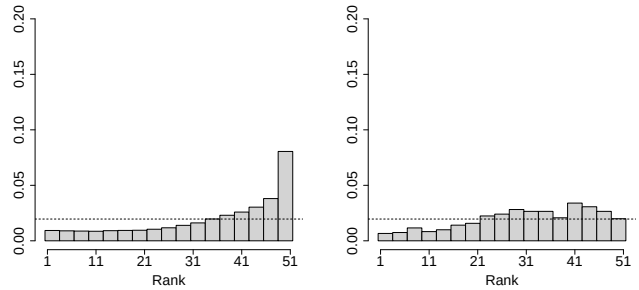
Table 5.7: Multivariate predictive performance measured with the determinant sharpness (DS) for the North Sea islands Borkum, Norderney, Helgoland, and Sylt, aggregated over all days in March 24, 2010 to April 30, 2011

in the model that admits more variability. For all local EMOS variants, the sharpness is increased. For each station a separate set of parameters is estimated, based only on data for that specific station. This univariate procedure results in margins with a high sharpness level, that is adopted by the multivariate distribution. For MEMOS, the sharpness is again reduced due to the fact that the method is taking into account information from several neighbouring stations. However, in comparison to global EMOS, MEMOS exhibits higher sharpness for all the multivariate versions.

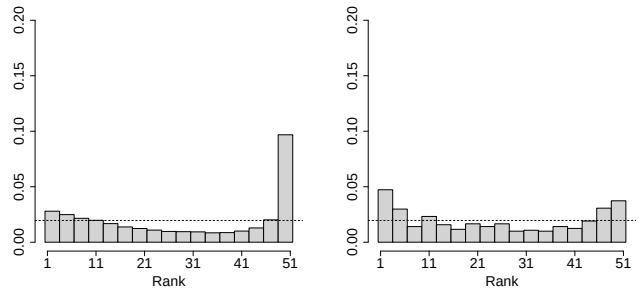
The multivariate rank histograms for the different methods are presented in Figure 5.8. Although the ECC distributions are the sharpest ones, the inheritance of the rank structure of the original ensemble improves the calibration as it utilizes the information from the NWP models underlying the raw ensemble forecasts.



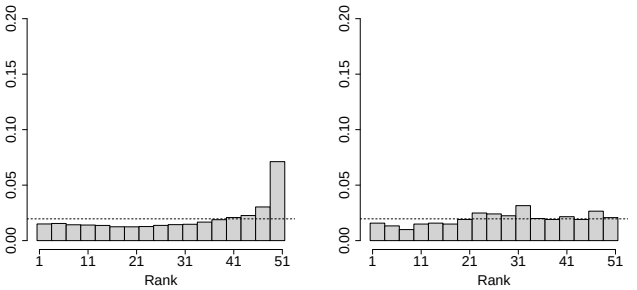
(a) Raw ECMWF Ind (b) Raw ECMWF ECC



(c) Global EMOS Ind (d) Global EMOS ECC



(e) Local EMOS Ind (f) Local EMOS ECC



(g) MEMOS Ind (h) MEMOS ECC

Figure 5.8: Multivariate verification rank histograms for the North Sea islands, over all dates in March 24, 2010 to April 30, 2011

	L	PO	FB	LK	KH	LW	WH	BH	CH	NH	E
Leck (L)	*7*	57	143	91	90	115	161	142	103	115	193
St. Peter-Ording (PO)		*5*	168	110	100	58	105	87	49	60	137
Fehmarnbelt (FB)			*4*	58	70	212	255	206	178	186	290
LH Kiel (LK)				*5*	17	155	200	154	123	132	235
Kiel-Holtenau (KH)					*27*	142	186	138	109	117	220
LH Alte Weser (LW)						*32*	47	48	38	37	81
Wittmundhafen (WH)							*8*	60	77	70	35
Bremerhaven (BH)								*7*	39	27	91
Cuxhaven (CH)									*5*	12	112
Nordholz (NH)										*25*	104
Emden (E)											*0*

Table 5.8: Pairwise distances in km (upper triangle) between the 11 stations along the North Sea coastline and their elevation in m (diagonal)

The second multivariate example consists of eleven stations along the North Sea coastline. The chosen stations cover a large region, ranging from the East Frisian region to the North Frisian region and even over to the coast of the Baltic Sea. Most of the stations are directly situated at the onshore coast. However, some of the stations are on small offshore islands or lighthouses very close to the onshore coast. Together with the stations from the other multivariate and univariate examples, the eleven stations are marked in red in Figure 5.1 (a) as well.

Table 5.8 shows the pairwise distances between the eleven stations along with the elevation of each station. All stations have an elevation close to sea level, the city of Emden is actually at sea level.

Table 5.9 presents the energy score (ES) values for the raw ensemble, global and local EMOS, and MEMOS, each provided with the types of dependence structure described above. Table 5.10 shows the respective results for the determinant sharpness (DS), while Figure 5.9 presents the corresponding multivariate rank histograms. The patterns are quite similar to those in the North Sea islands example. Even the multivariate rank histograms have nearly the same appearance.

For each dependence structure, the predictive performance increases when going from the raw ensemble to global EMOS, local EMOS, and finally MEMOS. MEMOS is the best

	Independence	ECC
Raw ECMWF	6.33	6.37
Global EMOS	5.41	5.24
Local EMOS	4.80	4.74
MEMOS	4.74	4.61

Table 5.9: Multivariate predictive performance measured with the energy score (ES) for eleven locations along the North Sea coastline, aggregated over all days in March 24, 2010 to April 30, 2011

choice for the marginal model, regardless which dependence structure is added. MEMOS even improves on its natural benchmark, local EMOS. This result is in line with the univariate results presented in Section 5.9.1 and with the results from the North Sea islands example.

As in the previous example it is not only of interest which of the marginal models yields the best predictive performance, but the aim is to additionally investigate the question to what extent the multivariate dependence structure influences the predictive quality. The best results are obtained for the ECC variant. Only for the raw ensemble the pattern is different.

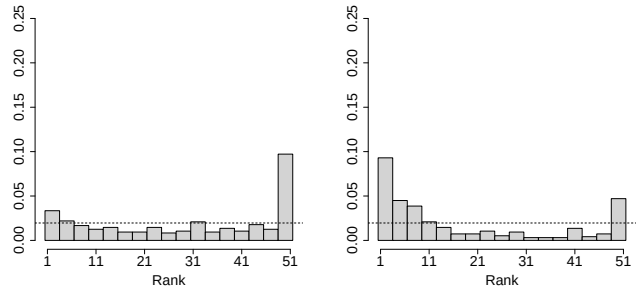
This structure within the ES values can be explained as already mentioned in the North Sea islands example. The univariate raw ensemble is heavily underdispersed. The DS values in Table 5.10 show that the sharpness of the multivariate predictive distributions is highest for the ECC approach. In case of the univariate raw ensemble not yet corrected for dispersion errors, a larger spread in the multivariate distribution has a positive effect. Therefore raw ensemble Independence yields the best ES value, while the ES value for the ECC approach deteriorates due to a too small spread in the multivariate distribution.

Going down the columns of Table 5.10 the sharpness decreases when moving from the raw ensemble to global EMOS, it increases from global to local EMOS, but decreases again when going from local EMOS to MEMOS. This is reasonable, as a predictive global EMOS distribution exhibits more spread than the 50 raw ensemble members. As local EMOS estimates the predictive distributions from the data of each station individually, this imparts greater sharpness than global EMOS. MEMOS utilizes a Markovian structure, taking into account the information from neighbouring stations. This increases the spread in the predictive distribution in comparison to local EMOS. However, as MEMOS is a spatially adaptive method, it produces predictive distributions with higher sharpness than global EMOS.

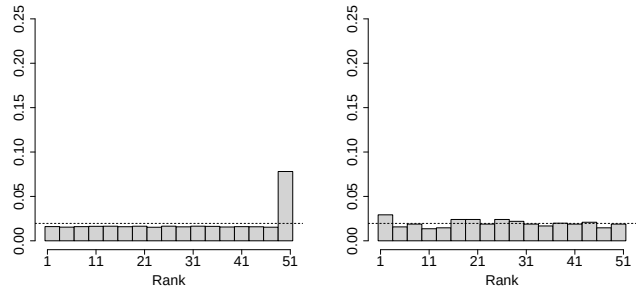
	Independence	ECC
Raw ECMWF	0.62	0.19
Global EMOS	2.96	1.01
Local EMOS	1.75	0.60
MEMOS	2.36	0.81

Table 5.10: Multivariate predictive performance measured with the determinant sharpness (DS) for eleven locations along the North Sea coastline, aggregated over all days in March 24, 2010 to April 30, 2011

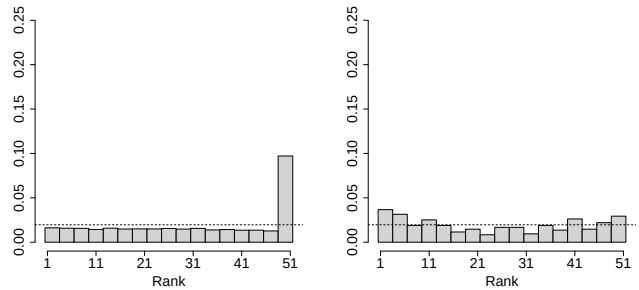
Figure 5.9 presents the multivariate rank histograms for all considered methods. The histograms for the raw ensemble versions are in the top row. Raw ensemble Independence displays a bias, the bin for the highest ranks is more occupied than the bin for the lowest ranks. Raw ensemble ECC exhibits underdispersion, but here the low ranks are more pronounced than for raw ensemble Independence. The second row presents the multivariate rank histograms for the global EMOS versions. For global EMOS ECC the improvement is significantly, the multivariate rank histogram is very close to uniformity. The underdispersion of raw ensemble ECC is not visible any more. These observations for the global EMOS histograms are in line with the ES values in Table 5.9. The multivariate rank histograms for local EMOS in the second row and for MEMOS in the bottom row are similar to the global EMOS histograms. However, local EMOS ECC exhibits a weak tendency to underdispersion in contrast to global EMOS ECC. MEMOS ECC is very close to uniformity.



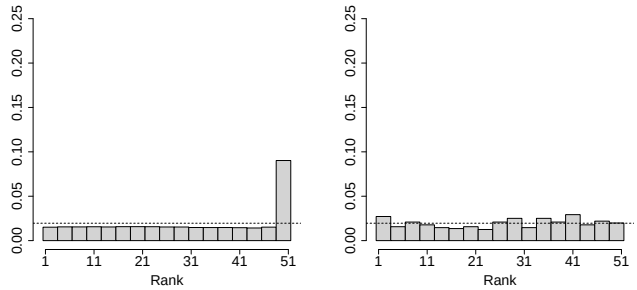
(a) Raw ECMWF Ind (b) Raw ECMWF ECC



(c) Global EMOS Ind (d) Global EMOS ECC



(e) Local EMOS Ind (f) Local EMOS ECC



(g) MEMOS Ind (h) MEMOS ECC

Figure 5.9: Multivariate verification rank histograms for the North Sea coastline, over all dates in March 24, 2010 to April 30, 2011

	Bad Muskau	Bertsdorf	Görlitz
Bad Muskau	*125*	74	50
Bertsdorf		*270*	30
Görlitz			*238*

Table 5.11: Pairwise distances in km (upper triangle) between the stations Bad Muskau, Görlitz and Bertsdorf and their elevation in m (diagonal)

	Independence	ECC
Raw ECMWF	4.02	4.04
Global EMOS	3.38	3.32
Local EMOS	3.08	3.03
MEMOS	3.06	3.01

Table 5.12: Multivariate predictive performance measured with the energy score (ES) for the cities Bad Muskau, Görlitz, and Bertsdorf near the Polish border, aggregated over all days in March 24, 2010 to April 30, 2011

The third multivariate example concerns three stations in the eastern part of Germany, at the Polish border, namely Bad Muskau, Bertsdorf, and Görlitz, the easternmost city of Germany. Table 5.11 shows the pairwise distances between the three stations and their elevations. Figure 5.1 (a) shows the three stations marked in red in the very eastern part of Germany. When looking additionally at panel (b) and (c) it is clear that two of the cities are vertices of the same triangle, while the third one is at the vertex of an immediate neighbour triangle.

Tables 5.12 and 5.13 present the ES and DS values, the respective multivariate rank histograms can be found in Figure 5.10. The pattern within the ES and DS values directly corresponds to the results in the two North Sea examples. Going down each column of the ES values in Table 5.12 yields an improvement. The most significant one is visible from the raw ensemble to global EMOS, as in the other examples. The improvements from the simple to the more refined marginal postprocessing models yields smaller improvements.

The pattern of the ES values within each row is similar to the North Sea examples as well. For the raw ensemble, the best results are obtained with the Independence approach. For

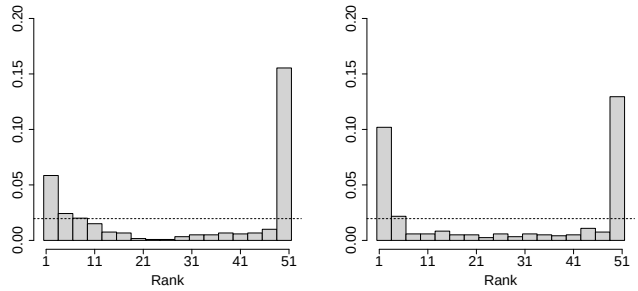
	Independence	ECC
Raw ECMWF	0.67	0.41
Global EMOS	2.90	1.74
Local EMOS	2.36	1.41
MEMOS	2.30	1.38

Table 5.13: Multivariate predictive performance measured with the determinant sharpness (DS) for the cities Bad Muskau, Görlitz, and Bertsdorf near the Polish border, aggregated over all days in March 24, 2010 to April 30, 2011

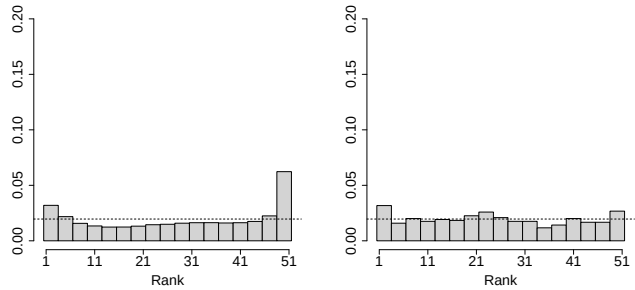
ECC, the ES value deteriorates and for the Copula approach it deteriorates slightly more. This is explained by the heavy underdispersion present in the univariate raw ensemble, as already discussed in the previous examples. A lower level of sharpness as for Independence (see Table 5.13) yields better results in case of underdispersion, while an underdispersed and sharp distribution as in the case of ECC deteriorates the results. Increased sharpness only has a positive effect on the scores subject to good calibration.

For the three postprocessing models, the effect is the other way around, as in the previous examples. The best ES values are obtained when combining the respective univariate post-processing model with ECC.

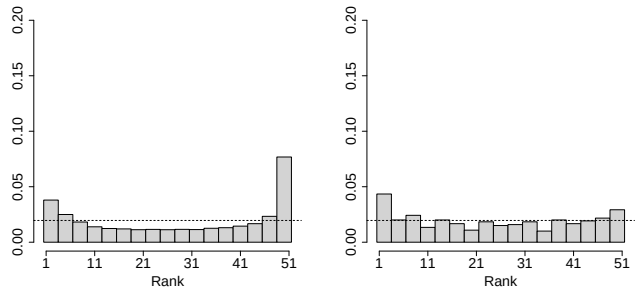
The respective multivariate rank histograms can be found in Figure 5.10. Their appearance is similar to the histograms in the other two examples. It is clearly visible that ECC obtains the best calibration for all univariate postprocessing models and the raw ensemble. In the top row the histograms for the raw ensemble versions are presented. Raw ensemble Independence exhibits a strong underdispersion and bias, the last bin is over-occupied. Raw ensemble ECC has a reduced bias, but still displays heavy underdispersion. The second row shows the multivariate rank histograms for the global EMOS variants. Calibration is improved significantly in comparison to the raw ensemble. For global EMOS Independence, a small bias and dispersion error is still visible, while the calibration of global EMOS ECC is pretty good. The histograms for local EMOS are similar to the ones for global EMOS. However, the local EMOS variants exhibit slightly stronger underdispersion than the global EMOS variants. Nonetheless, local EMOS ECC yields the best calibration, very close to uniformity. The bottom row presents the histograms for the MEMOS variants. They are nearly identical to the histograms of local EMOS. Although the scores improve for MEMOS in comparison to local EMOS, the multivariate rank histograms stay essentially the same.



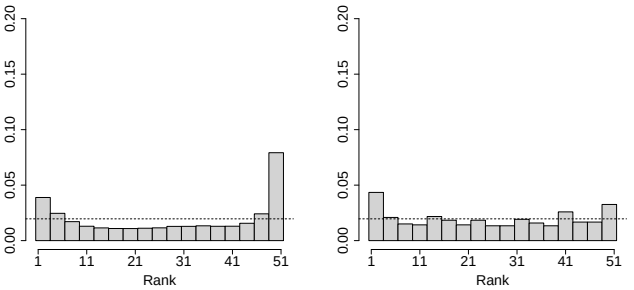
(a) Raw ECMWF Ind (b) Raw ECMWF ECC



(c) Global EMOS Ind (d) Global EMOS ECC



(e) Local EMOS Ind (f) Local EMOS ECC



(g) MEMOS Ind (h) MEMOS ECC

Figure 5.10: Multivariate verification rank histograms for the cities Bad Muskau, Görlitz, and Bertsdorf over all dates in March 24, 2010 to April 30, 2011

To sum up the results of the considered multivariate examples, MEMOS improves the predictive performance in comparison to the multivariate versions of standard EMOS. The MEMOS versions yielded the best ES values in the different station settings. It can be noted further that for each univariate postprocessing model the best results are obtained when combining it with ECC. This holds for global and local EMOS, as well as for MEMOS. Adding a multivariate dependence structure to the univariate postprocessing models yields an improvement in predictive performance. Especially the combination of MEMOS with the ECC structure constitutes a multivariate postprocessing procedure that results in high univariate and multivariate predictive quality.

5.9.4 Univariate results for composite quantities

The previous section assessed the multivariate predictive performance of the univariate methods combined with two types of dependence structure. Now, the univariate predictive performance of a composite quantity like minimum, maximum, or average temperature over a set of stations is assessed with univariate verification methods. Such a composite quantity displays the information about spatial structures in a compressed form.

The composite quantities are directly computed from the multivariate samples that were obtained from the combined methods as described in Section 5.9.3. The samples already produced and employed in the previous section can be used for this analysis, the only extra work is to compute a desired composite quantity from the original samples. The composite quantity of choice is computed over all selected stations, for each prediction day separately. The same procedure is applied to the observations. On every day, the composite quantity over the observations at the chosen stations is computed. Note that for all considered methods and the corresponding observations only those days where used in the verification process, on which all of the selected stations are available. Days on which at least one of the selected stations is missing were removed, as in Section 5.9.3.

For the case study that analyzes the univariate predictive performance of a composite quantity, the Polish border example already considered in Section 5.9.3 was chosen, to have a direct comparison of multivariate and univariate performance. The analyzed composite quantity is the minimum temperature over the three cities.

Tables 5.14, 5.15, 5.16 show the CRPS, the MAE, and the RMSE values for the three multivariate variants of each univariate method. The respective univariate rank histograms can be found in Figure 5.11. The pattern is similar to the one within the multivariate results.

The only difference is the raw ensemble, where now ECC exhibits the best predictive performance among the raw ensemble versions. In the multivariate examples as well as in this composite example, the ECC dependence structure yields the best scores and calibration for all considered univariate postprocessing models.

	Independence	ECC
Raw ECMWF	2.12	2.05
Global EMOS	2.01	1.80
Local EMOS	1.83	1.70
MEMOS	1.82	1.68

Table 5.14: CRPS for minimum temperature over the cities Bad Muskau, Görlitz, and Bertsdorf near the Polish border, aggregated over all days in March 24, 2010 to April 30, 2011

	Independence	ECC
Raw ECMWF	2.38	2.36
Global EMOS	2.70	2.45
Local EMOS	2.45	2.34
MEMOS	2.42	2.31

Table 5.15: MAE for minimum temperature over the cities Bad Muskau, Görlitz, and Bertsdorf near the Polish border, aggregated over all days in March 24, 2010 to April 30, 2011

The univariate rank histograms presented in Figure 5.11 have a similar appearance as the multivariate rank histograms in Figure 5.10. However, the univariate rank histograms for the raw ensemble versions displayed in the top row exhibit a stronger underdispersion than the respective multivariate rank histograms. The improvement in calibration when moving from the raw ensemble to one of the postprocessing models is clearly visible.

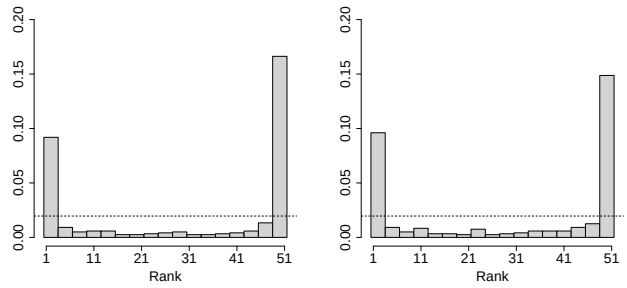
The current and the previous section were presenting two different ways of analyzing samples from a multivariate predictive distribution, constructed by providing univariate methods with certain types of multivariate dependence structures. The first possibility is to directly assess samples from the multivariate distribution with multivariate assessment tools. The second possibility is to reduce the dimension of the multivariate distribution to a univariate

	Independence	ECC
Raw ECMWF	3.00	2.96
Global EMOS	3.46	3.23
Local EMOS	3.12	2.98
MEMOS	3.05	2.92

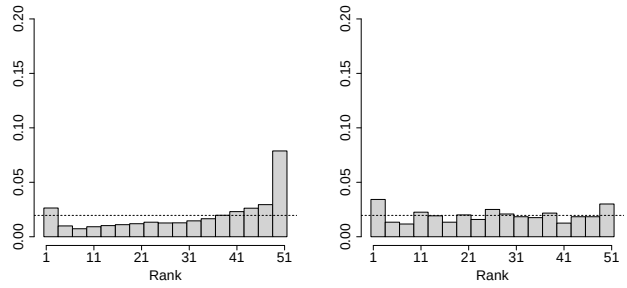
Table 5.16: RMSE for minimum temperature over the cities Bad Muskau, Görlitz, and Bertsdorf near the Polish border, aggregated over all days in March 24, 2010 to April 30, 2011

one by considering a composite quantity over the spatial dimensions. This composite values can be assessed with the standard univariate verification tools.

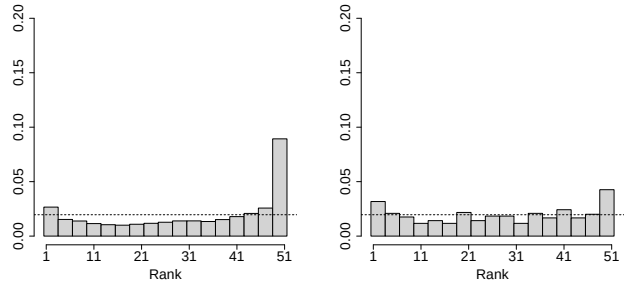
Both procedures showed that the multivariate versions of MEMOS outperform the respective multivariate versions of global and in most cases even of local EMOS, although the improvement when moving from local EMOS to MEMOS is often only a small one. However, the choice of the type of dependence structure employed to construct the multivariate distribution from the original univariate methods has a significant influence. The ECC procedure yields strong improvements in predictive performance, regardless which univariate model is employed.



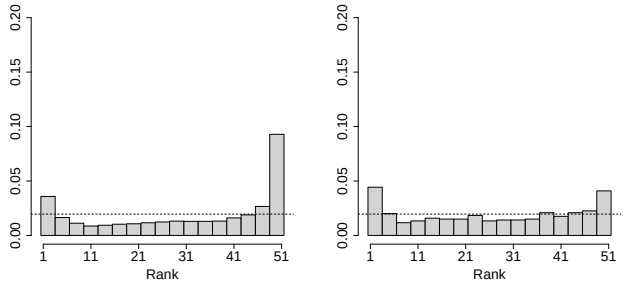
(a) Raw ECMWF Ind (b) Raw ECMWF ECC



(c) Global EMOS Ind (d) Global EMOS ECC



(e) Local EMOS Ind (f) Local EMOS ECC



(g) MEMOS Ind (h) MEMOS ECC

Figure 5.11: Univariate verification rank histograms for minimum temperature over the cities Bad Muskau, Görlitz, and Bertsdorf, over all dates in March 24, 2010 to April 30, 2011

5.9.5 Samples from multivariate MEMOS versions

Sections 5.9.3 and 5.9.4 showed results for predictive performance of the multivariate postprocessing methods, obtained by providing each univariate model with different dependence structures as described in Section 5.5. To conclude the previous analysis, this section exemplarily presents samples at all observation stations present at a chosen prediction day from the multivariate predictive MEMOS distributions obtained by combining the basic MEMOS model with the ECC structure. The samples are compared with the true observations and with randomly selected raw ensemble members.

The prediction day October 3, 2010 was chosen for displaying the observations and samples from the multivariate MEMOS methods, as the spatial temperature distribution over Germany was particularly interesting during the night from October 2 to October 3, 2010. The mesh employed for estimation of the basic MEMOS model that is used for prediction on October 3, 2010, along with the 508 available observation stations on that day, is displayed in Figure 5.1 in Section 5.6. The original raw ECMWF 24-h ahead forecasts and observations employed in the case study were initialized at 00 UTC. For this example 00 UTC corresponds to 2am local time (daylight saving time operates until the end of October). The raw ECMWF forecasts and the samples obtained from the predictive postprocessing distributions displayed in Figure 5.12 are valid at 2am on October 3, 2010. In that specific night, quite interesting and for autumn unusual change concerning the temperature took place.

In the top row of Figure 5.12 three randomly selected raw ensemble members are shown. The second row displays three of the 50 MEMOS ECC members, where the rank dependence structure is adopted from the raw ensemble member in the respective place in the row above. The true observations on October 3, 2010 are shown in the left and right panel of the bottom row.

According to information from achieves about past weather concerning the night in question[1] there was a low pressure area over the Northern Atlantic Ocean ranging to Western Europe and deep into southern regions. The tail of the low pressure area was affecting middle Europe from a westward direction. The result was cloud cover and precipitation in the western and middle parts of Germany. However, the eastern part of Germany benefited from the influence of a high pressure area being centered on the western part of Russia. This lead to sunny weather in the easternmost regions of Germany. Between these two pressure areas masses

[1] `http://www.wetter24.de/wetter-news/archiv.html`, `http://www.wetter.de/wetterarchiv/wetterbericht/2010-10-02`, and `http://www.wetterzentrale.de/topkarten/fscfsreaeur.html`

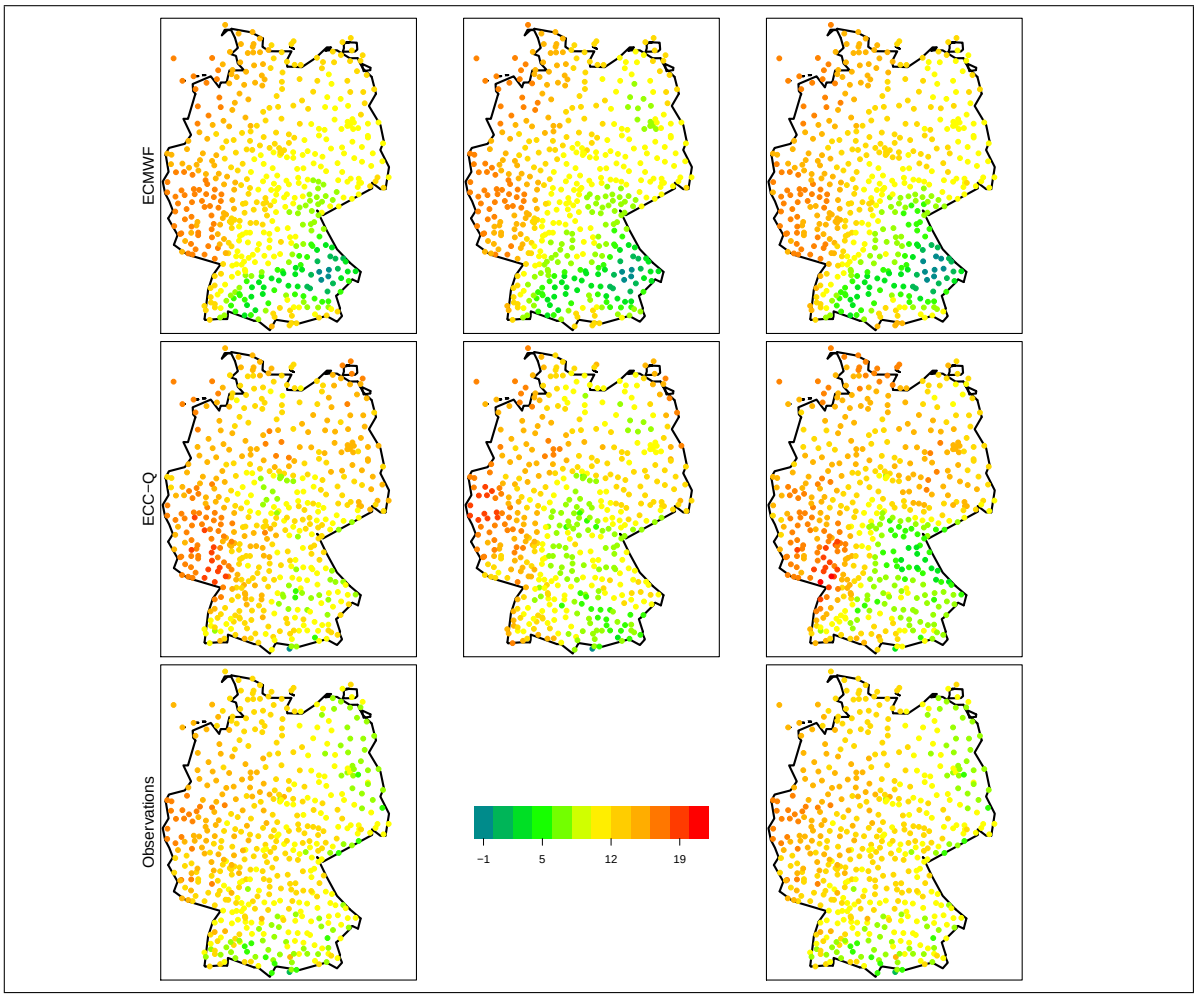


Figure 5.12: Samples from MEMOS ECC predictive distribution for surface temperature over Germany in degrees Celsius, along with raw ensemble forecasts and the observations at all 508 available stations. All forecasts and the observations are valid 2am on October 3, 2010.

of warm air coming from the Mediterranean Sea streamed over to central Europe, yielding very warm late summer temperatures in Germany, slightly untypical for the beginning of October. However, the masses of warm air in combination with the high cloud cover in the western part and a nearly cloudless night in the eastern part induced an east-west partitioning in the temperature level. While the temperatures in the western regions of Germany stayed very mild throughout the night, they dropped heavily in the eastern regions of Germany. In some geographically sheltered regions in Northern East Germany and in Eastern Bavaria the temperature could have even dropped to freezing levels.

This east-west temperature partition with especially low temperatures in the easternmost parts and some south-eastern parts of Germany is clearly visible in the observations (bottom row) for the night of October 3, 2010. The warmest temperatures can be found in the mid-

west of Germany, where the influence of the low-pressure area is strongest. The temperatures decrease in eastward direction where the influence of the high-pressure area increases. The described east-west pattern of the temperature is visible in the raw ensemble members as well. However, all three considered members show a strong tendency to issue more extreme temperature forecasts than the observations. In the mid-west region of Germany with quite warm temperatures for an October night the considered ensemble members issue even higher temperature forecasts than the observations. For the regions of Bavaria with very low temperatures, the raw ensemble forecasts provide even cooler temperature forecasts than present in the observations. The east-west partitioning in the temperature level is much more pronounced in the raw ensemble forecasts. The raw ensemble overestimates the temperature at most of the stations with an already high observed temperature and underestimates it at stations with an already low observed temperature. However, the general spatial structure of the temperature observations is clearly present in the raw ensemble forecasts.

The samples from MEMOS ECC correct at least for the extreme low temperature forecasts, resulting in a more balanced spatial structure. Nonetheless, at some of the stations with an already high observed temperature value, the method still issues too high temperatures in comparison to the observations. In the MEMOS ECC samples, the general spatial structure present in the original raw ensemble is clearly visible, only the levels of the temperature forecasts are adapted by employing univariate MEMOS to the margins. MEMOS ECC predicts too warm temperatures in East Germany. The ECC member in the left column provides quite suitable forecasts for the south-east region. The other two members, especially the ECC member in the right column, issue too cold forecasts for that region.

Obviously, the structure of the east-west partition present in Germany during the night in question is captured by the raw ensemble. Nonetheless, the raw ensemble cannot forecast the temperature levels themselves adequately. The postprocessing methods are able to correct the temperature levels in the right direction. However, the unusual east-west structure is not captured to its full extent. As this night presents not a common everyday weather situation, it is harder for a postprocessing method to catch up with such a situation when having only past data not foretelling such a change. At days with a more standard weather situation that slowly develops over the training period, the MEMOS postprocessing versions may well be capturing the spatial structure even more adequately.

Chapter 6

Discussion

The standard practice to forecast future weather variables such as surface temperature, wind speed, precipitation amount or other quantities of interest is to employ deterministic numerical weather prediction (NWP) models on the basis of of differential equations describing the physics of the atmosphere. However, the outputs of these models provide single deterministic point forecasts, not allowing to assess the forecast uncertainty of the NWP models properly. To account for this, ensembles of NWP forecasts have been proposed. The ensemble members are generated by running the NPW models with different initial conditions or model formulations. Forecast ensembles allow for probabilistic weather forecasting, and it is possible to obtain information about predictive uncertainty via the ensemble spread. Nonetheless, ensembles tend to be biased and often exhibit dispersion errors. In particular, many ensembles are underdispersed. To correct for these shortcomings, statistical postprocessing of the NWP output has become routine. However, many of the standard postprocessing methods are designed for a single weather quantity, a fixed forecast horizon and for fixed locations. As the raw ensemble forecasts consist of NWP model output describing a variety of aspects of the atmospheric processes, it can be assumed that multivariate structures are present in the ensemble forecasts. By applying univariate postprocessing methods to the forecast ensemble, these multivariate structures are ignored.

These shortcomings require the development of postprocessing methods taking into account inter-variable, spatial, or temporal dependence structures. Several techniques of that type have already been proposed. This work introduces two extensions of already existing standard postprocessing methods, namely BMA and EMOS, to account for shortcomings of the basic versions. Similar procedures to the ones presented here can be developed on the basis of any desired univariate postprocessing method.

The first extension proposed by Möller et al. (2013) and discussed in Chapter 4 aims at recovering inter-variable dependence structures that are ignored by standard univariate postprocessing methods. By employing a univariate postprocessing model it is possible to correct for dispersion errors and biases present in the raw ensemble. However, the dependence structures of the raw ensemble are not inherited. The proposed extension compensates for this by providing the univariate margins with a multivariate dependence structure induced by a Gaussian copula model. The dependence structure is represented by a correlation matrix. The margins for each weather quantity are postprocessed with the basic univariate method at hand, here the BMA model. Apart from the postprocessing distributions of the margins the only parameter that needs to be estimated is the dependence parameter of the Gaussian copula model. The Gaussian copula model separates the estimation of the margins from the estimation of the dependence structure. The introduced postprocessing procedure is simple and not restricted to BMA. It should be noticed, however, that the Gaussian copula procedure models the multivariate correlation structure only, leaving the margins unchanged. A useful feature about this approach is the fact that the procedure models the multivariate distribution *after* postprocessing. By first applying a univariate postprocessing model to the margins and then learning about the dependence parameter and the latent Gaussian factors using a verifying observation, the joint residual structure implied by the univariate postprocessing method is implicitly modeled. This strategy reflects the modeling process in the context of ensemble postprocessing in an appropriate way. The case study for five weather quantities conducted in Chapter 4 utilizes the UWME ensemble and revealed that the Gaussian copula multi-stage procedure yields a multivariate predictive distribution with good multivariate calibration and sharpness properties. A further advantage of the multi-stage procedure is its flexibility in the application: It is not restricted to the weather quantities considered there. In general, the method can be applied to arbitrary weather quantities, as long as a suitable univariate postprocessing model is at hand.

While the approach presented in Chapter 4 employs a constant correlation matrix for all observations, more involved estimation strategies might be considered, e.g. a time-varying correlation matrix. Further investigations performed by Möller et al. (2013) revealed no additional benefit in the conducted case study. However, for future research and different data sets the type of the employed correlation matrix should be reconsidered, especially in high-dimensional situations.

Some alternatives to the Gaussian copula approach use discrete copulas. Examples are the above mentioned ECC method (Schefzik et al., 2013) or the Schaake shuffle (Clark et al., 2004). ECC learns the multivariate rank structure from the original raw ensemble, while the Schaake shuffle investigates historical observations. The advantage of the Gaussian copula

procedure discussed in this work is that it is not requiring subsequent samples from the predictive distribution to strictly obey any observed rank structure. While the Schaake shuffle additionally requires a large number of observations, ECC relies on a sample of the same size as the original ensemble. Thus, a disadvantage of the current ECC version is the fact that it only generates an ECC sample of the size of the original ensemble, while the Gaussian copula approach allows to obtain an arbitrary number of samples. Further research should be undertaken to investigate the quality of these and other alternative methods to construct multivariate distributions. The statistics literature has started to discuss these questions (Hoff et al., 2011) and subsequent work to compare these methods in the ensemble postprocessing context will be necessary and fruitful.

The Gaussian copula model has received occasional criticism (Mikosch, 2006) as it belongs to the family of elliptic distributions and is thus not capturing the dependence in the tails of a multivariate distribution adequately. Other copula models exist that focus on modelling extreme events and heavy tailed distributions. However, in the context of the developed multistage procedures for ensemble postprocessing this concern is not of main interest. The goal of the procedures developed in this work is to construct a multivariate predictive distribution after postprocessing a raw forecast ensemble with univariate methods. In case of extreme weather events, this information is likely to be incorporated to some extent in the ensemble itself. The Gaussian copula would simply indicate the variability about these extreme values. The modeling task of main interest in ensemble postprocessing focuses on day-to-day prediction of weather quantities. In the context of predicting and analyzing extreme events, other copula models may be preferable. Further research is planned in the direction of employing other copula models, especially vine copulas (Aas et al., 2009), for ensemble postprocessing and compare the predictive quality to the Gaussian copula approach. An additional research project that aims at joint modelling of two weather quantities, surface temperature and wind speed, was started recently. The 2-dimensional Gaussian copula approach will be compared to a bivariate version of BMA for the ALADIN-HUNEPS ensemble provided by the Hungarian Meteorological Service (HMS). For this setting, an alternative BMA version for wind speed, proposed by Baran (2013) and Baran et al. (2013) will be employed.

The procedure discussed in Chapter 5 introduces a spatially adaptive extension of EMOS, while considering a single weather quantity. The procedure utilizes two recently developed methods, the INLA-technique (Rue et al., 2009) and the SPDE approach (Lindgren et al., 2011), which allows for computing a GMRF representation of a GF. Both techniques are conveniently implemented in the `R-INLA` package and can directly be used for model estimation. The key feature is the spatially adaptive extension of the univariate EMOS model, where the bias-correction parameters are assumed to be realizations of latent GFs. This is

combined with the SPDE approach allowing to obtain a GMRF representation of these GFs, resulting in additional computational benefits. The method was developed for temperature, which can be assumed to have a normal distribution. However, extensions for weather quantities with other marginal distributions are possible, as INLA is not restricted to normally distributed responses. A future area of research may for example be the development of spatially adaptive models for precipitation or wind speed. In the case of wind speed an modified version of univariate EMOS exists (see Section 2.1.1 and Thorarinsdottir and Gneiting, 2010) that can directly be extended in a similar way as the EMOS version for normally distributed variables. For precipitation, a postprocessing model needs to be employed, that is designed for the specific situation of precipitation, as for example the EMOS variant proposed by Scheuerer (2013). However, the discussed extension may well be expanded to other univariate postprocessing models, as the INLA procedure captures a wide range of models and is very flexible. A generalization of this approach may be a conceivable and interesting area of research in the context of ensemble postprocessing.

The case study over Germany conducted in this work revealed that MEMOS outperforms global and local EMOS in its univariate version. The multivariate versions with different dependence structures improve the multivariate predictive performance as well. When inspecting the samples from the multivariate MEMOS distributions visually, it can be seen that these methods are able to capture the spatial structure of temperatures in Germany.

The current MEMOS method relies on the simplest stationary version of the SPDE model with Matérn covariance function. More involved versions of the SPDE methodology are possible and the theoretical development is in general straightforward. Of special interest in climate and weather prediction are non-stationary and anisotropic fields as well as fields capturing spatio-temporal dependencies. To forecast extreme weather events like a long period of heavy rain, heavy storms or sudden hail showers, a method taking into account temporal structures may provide additional benefit. The derivation of more general versions of the SPDE model as well as their implementation in the `R-SPDE-INLA` framework is in progress, for details see Bolin and Lindgren (2011) and Lindgren et al. (2011). Therefore, a refinement and extension of the basic MEMOS method within the SPDE framework might be available in near future and further research can be conducted in this direction.

The approximative procedure to sample from the MEMOS predictive distribution (5.41) described in Algorithm 5.1 is not a genuinely Bayesian approach, as it only involves sampling from the posterior marginal distribution of η_s. Instead of additionally sampling from the posterior marginal of σ^2, the posterior mean of σ^2 is plugged in and samples $Z_i \sim \mathrm{N}(0, \hat{\mu}_{\sigma^2})$ are obtained. These Z_i are used to generate a sample $Y_{s,i} = \eta_{s,i} + Z_i$. A more refined and

purely Bayesian version of this sampling procedure will be developed in future research. The refined approach involves to also obtain samples σ_i^2 from the posterior marginal of σ^2. The sample from the MEMOS predictive distribution (5.41) can then be obtained by sampling $Z_i^* \sim \mathrm{N}(0,1)$ and setting $Y_{s,i} = \eta_{s,i} + \sigma_i \cdot Z_i^*$. This approach is likely to increase the spread in the samples $Y_{s,1}, \ldots, Y_{\text{s,N.sample}}, s = s_1, \ldots, s_N$ and therefore may yield an additional improvement in the scores and the calibration of the MEMOS predictive distribution. Due to the Bayesian nature of MEMOS, a further extension of the multivariate MEMOS ECC procedure is possible that yields a predictive sample of any desired size. That is, for a single set of parameter values drawn from the posterior distributions of η_s and σ^2, a multivariate MEMOS ECC sample is constructed conditional on the current parameter values. This procedure may then be repeated arbitrarily often until the predictive sample has reached the required size. Such an ECC variant would provide a more accurate visualization of the multivariate predictive distribution and might yield an additional improvement in predictive performance.

As already discussed in Chapter 5, the current MEMOS version employs a constant variance in space. A possibility to obtain a model with spatially structured variance may be via one of the more complex SPDE extensions. Another alternative is to consider a different model that directly incorporates a spatially varying model variance, as for example stochastic volatility models. Huang et al. (2011) investigate the application of stochastic volatility models in an environmental context. They extend existing models to a heteroscedastic version with a non-stationary covariance structure in space and time. However, they do not employ INLA or the SPDE method to estimate their model. Further research in this direction may include a modification of the volatility models implemented in INLA to fit the context of ensemble postprocessing or alternatively move away from INLA and extend the model of Huang et al. (2011) to be suitable for the desired purpose of ensemble postprocessing. A comparison between INLA-based models, standard postprocessing models including a spatially structured variance (e.g. GMA, see below) and stochastic volatility models similar to the one proposed by Huang et al. (2011) might be of interest in the ensemble postprocessing context.

Several postprocessing methods utilizing spatially adaptive model parameters or aiming at producing coherent and calibrated weather fields are already available. Among them are, for example, methods like spatial BMA (Berrocal et al., 2007, 2008), spatial EMOS (Feldmann, 2012) and Geostatistical Model Averaging (GMA, Kleiber et al., 2011a,b). A comparison of the predictive performance of the proposed MEMOS method to these state of the art methods is of interest for future research. A very recently developed spatially adaptive extension of EMOS utilizing an intrinsic and stationary Gaussian random field model to perform spatial interpolation of the model parameters was proposed by Scheuerer and Büer-

mann (2013). This approach represents the predictive mean as a sum of short-term averages of local temperature observations and functions of the ensemble forecasts. It constitutes a useful alternative to the current MEMOS version for temperature. Especially concerning spatial out-of-sample predictive performance, a comparison between MEMOS, the method proposed by Scheuerer and Büermann (2013), spatial BMA and GMA is of high interest. As in the current work only spatial in-sample performance was considered, a special interest for future research will be the assessment of spatial out-of-sample performance of the univariate as well as the multivariate MEMOS variants in comparison to other spatially adaptive postprocessing methods.

To improve the predictive performance of MEMOS further one may think about incorporating other covariates in the model that explain the temperature level, such as the elevation at the stations. Preliminary case studies revealed that the predictive performance of the MEMOS model deteriorates at stations with a very high elevation. For example, the mountains Zugspitze and Wendelstein in Bavaria, the mountain Feldberg in the Black Forest, or the mountain Brocken in the Harz region differ significantly in their elevation from stations in the very north of Germany, which are only several meters above sea level. This suggests an inclusion of an additional covariate describing the elevation of the observation stations, or to perform other correction methods within the estimation procedure. Future research may investigate wether this leads to an improvement in the predictive performance of MEMOS.

To conclude, the two proposed extensions presented in this work lead to an improvement in predictive performance and yield multivariate predictive distributions with improved multivariate calibration properties. The methods proposed utilize already existing univariate postprocessing methods, which makes them easy to handle. The conjunction of marginal and joint calibration is a very promising concept that can be followed up in further research. The fact that the statistical postprocessing literature is currently investigating numerous multivariate postprocessing procedures indicates the relevance of this topic. Both extensions presented in this work aim at recovering multivariate dependence structures. The first method models several weather quantities jointly and thus accounts for dependencies among these quantities while ignoring spatial structures. The second method incorporates spatial dependencies but only for a single weather quantity. A natural extension of high relevance is the combination of both ideas to model weather quantities jointly in a spatio-temporal domain. Such a procedure can utilize advanced features of the SPDE approach, where spatio-temporal modelling is possible. With the ongoing development and implementation of more complex features in the `R-SPDE-INLA` framework, the tools to analyse such models are provided. Several years from now, joint univariate and multivariate postprocessing in a combined intervariable and spatio-temporal setting may well become feasible.

Bibliography

K. Aas, C. Czado, A. Frigessi, and H. Bakken. Pair-copula constructions of multiple dependence. *Insurance: Mathematics & Economics*, 44:182–198, 2009.

J. L. Anderson. A method for producing and evaluating probabilistic forecasts from ensemble model integrations. *Journal of Climate*, 9:1518–1530, 1996.

J. Baars. Observations QC documentation. Available at `http://www.atmos.washington.edu/mm5rt/qc_obs/qc_doc.html`, 2005. ISSN 0935-8943.

L. Bao, T. Gneiting, E. P. Grimit, P. Guttorp, and A. E. Raftery. Bias correction and Bayesian model averaging for ensemble forecasts of surface wind direction. *Monthly Weather Review*, 138:1811–1821, 2010.

S. Baran. Probabilistic wind speed forecasting using Bayesian model averaging with truncated normal components. *Computational Statistics and Data Analysis*, 2013. To appear.

S. Baran, A. Horányi, and D. Nemoda. Statistical post-processing of probabilistic wind speed forecasting in Hungary. *Meteorologische Zeitschrift*, 22(3):273–282, 2013.

T. W. Barker. The relationship between spread and forecast error in extended-range forecasts. *Journal of Climate*, 4(7):733–742, 1991.

V. J. Berrocal, A. E. Raftery, and T. Gneiting. Combining spatial statistical and ensemble information in probabilistic weather forecasts. *Monthly Weather Review*, 135:1386–1402, 2007.

V. J. Berrocal, A. E. Raftery, and T. Gneiting. Probabilistic quantitative precipitation field forecasting using a two-stage spatial model. *Annals of Applied Statistics*, 2(4):1170–1193, 2008.

M. Blangiardo, M. Cameletti, G. Baio, and H. Rue. Spatial and spatio-temporal models with R-INLA. *Spatial and Spatio-Temporal Epidemiology*, 4:33 – 49, 2013.

D. Bolin and F. Lindgren. Spatial models generated by nested stochastic partial differential equations, with an application to global ozone mapping. *Annals of Applied Statistics*, 5 (1):523 – 550, 2011.

R. Buizza. The ECMWF ensemble prediction system. In T. N. Palmer and R. Hagedorn, editors, *Predictability of Weather and Climate*, pages 459–489. Cambridge University Press, 2006.

M. Cameletti, F. Lindgren, D. Simpson, and H. Rue. Spatio-temporal modeling of particulate matter concentration through the SPDE approach. *AStA Advances in Statistical Analysis*, 97(2):109–131, 2013.

G. Candille and O. Talagrand. Evaluation of probabilistic prediction systems for a scalar variable. *Quarterly Journal of the Royal Meteorological Society*, 131(609):2131–2150, 2005.

R. M. Chmielecki and A. E. Raftery. Probabilistic visibility forecasting using Bayesian model averaging. *Monthly Weather Review*, 139:1626–1636, 2010.

M. P. Clark, S. Gangopadhyay, L. E. Hay, B. Rajagopalan, and R. L. Wilby. The Schaake shuffle: a method for reconstructing space-time variability in forecasted precipitation and temperature fields. *Journal of Hydrometeorology*, 5:243–262, 2004.

N. A. C. Cressie. *Statistics for Spatial Data*. Wiley, 1993.

A. C. Davison, S. Padoan, and M. Ribatet. Statistical modelling of spatial extremes. *Statistical Science*, 27:161–186, 2012.

A. P. Dawid. Statistical theory: The prequential approach (with discussion and rejoinder). *Journal of the Royal Statistical Society Ser. A*, 147:278–292, 1984.

L. Delle Monache, J. P. Hacker, Y. Zhou, X. Deng, and R. B. Stull. Probabilistic aspects of meteorological and ozone regional ensemble forecasts. *Journal of Geophysical Research*, 111:D24307, 2006. doi: 10.1029/2005JD006917.

S. Demarta and A. J. McNeil. The t copula and related copulas. *International Statistical Review*, 73:111–129, 2005.

P. J. Diggle and P. J. Ribeiro Jr. *Model-based Geostatistics*. Springer, 2010.

F. A. Eckel and C. F. Mass. Aspects of effective mesoscale, short-range ensemble forecasting. *Weather and Forecasting*, 20:328–350, 2005.

P. Embrechts, F. Lindskog, and A. McNeil. Modelling dependence with copulas and applications to risk management. In S. T. Rachev, editor, *Handbook of Heavy Tailed Distributions in Finance*, pages 329–384. Elsevier, 2003.

K. Feldmann. *Statistical Postprocessing of Ensemble Forecasts for Temperature: The Importance of Spatial Modeling*. Diploma thesis, Faculty of Mathematics and Informatics, University of Heidelberg, 2012.

C. Fraley, A. E. Raftery, T. Gneiting, J. M. Sloughter, and V. J. Berrocal. Probabilistic weather forecasting in R. *R Journal*, 3:55–63, 2011.

Y. Gel, A. E. Raftery, and T. Gneiting. Calibrated probabilistic mesoscale weather field forecasting: The geostatistical output perturbation (GOP) method (with discussion and rejoinder). *Journal of the American Statistical Association*, 99:575–590, 2004.

A. Gelfand, P. Diggle, M. Fuentes, and P. Guttorp, editors. *Handbook of Spatial Statistics*. Chapman and Hall, 2010.

C. Genest and A.-C. Favre. Everything you always wanted to know about copula modeling but where afraid to ask. *Journal of Hydrologic Engineering*, 12(4):347–368, 2007.

H. R. Glahn and D. A. Lowry. The use of model output statistics (MOS) in objective weather forecasting. *Journal of Applied Meteorology*, 11:1203–1211, 1972.

T. Gneiting. Making and evaluating point forecasts. *Journal of the American Statistical Association*, 106:746–762, 2011.

T. Gneiting and A. E. Raftery. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102:359–378, 2007.

T. Gneiting, A. E. Raftery, F. Balabdaoui, and A. Westveld. Verifying probabilistic forecasts: Calibration and sharpness. In *Proceedings of the Workshop on Ensemble Forecasting, Val-Morin, Québec*, 2003.

T. Gneiting, A. Raftery, A. Westveld, and T. Goldman. Calibrated probabilistic forecasting using ensemble model output statistics and minimum CRPS estimation. *Monthly Weather Review*, 33:1098–1118, 2005.

T. Gneiting, F. Balabdaoui, and A. E. Raftery. Probabilistic forecasts, calibration and sharpness. *Journal of the Royal Statistical Society Ser. B*, 69:243–268, 2007.

T. Gneiting, L. I. Stanberry, E. P. Grimit, L. Held, and N. A. Johnson. Assessing probabilistic forecasts of multivariate quantities, with applications to ensemble predictions of surface winds (with discussion and rejoinder). *Test*, 17:211–264, 2008.

P. I. Good. *Permutation Tests*. Springer, 1995.

E. P. Grimit and C. F. Mass. Initial results of a mesoscale short-range ensemble forecasting system over the Pacific Northwest. *Weather and Forecasting*, 17:192–205, 2002.

P. Guttorp and T. Gneiting. Studies in the history of probability and statistics XLIX: On the Matérn correlation family. *Biometrika*, 93:989–995, 2008.

T. M. Hamill. Interpretation of rank histograms for verifying ensemble forecasts. *Monthly Weather Review*, 129:550–560, 2001.

T. M. Hamill and S. J. Colucci. Verification of Eta-RSM short-range ensemble forecasts. *Monthly Weather Review*, 125:1312–1327, 1997.

H. Hersbach. Decomposition of the continuous ranked probability score for ensemble prediction systems. *Weather and Forecasting*, 15:559–570, 2000.

J. A. Hoeting, D. Madigan, A. E. Raftery, and C. T. Volinsky. Bayesian model averaging: A tutorial. *Statistical Science*, 14(4):382–417, 1999.

P. D. Hoff, X. Niu, and J. A. Wellner. Information bounds for Gaussian copulas. Technical Report 586, Department of Statistics, University of Washington, 2011. `http://arxiv.org/abs/1110.3572`.

Peter D. Hoff. Extending the rank likelihood for semiparametric copula estimation. *Annals of Applied Statistics*, 1(1):265–283, 2007.

W. Huang, K. Wang, F. J. Breidt, and R. A. Davis. A class of stochastic volatility models for environmental applications. *Journal of Time Series Analysis*, 32:364–377, 2011.

H. Joe. *Multivariate Models and Dependence Concepts*. Chapman and Hall, London, 1997.

I. T. Jolliffe. Comments on: Assessing probabilistic forecasts of multivariate quantities with an application to ensemble predictions of surface winds. *Test*, 17:249–250, 2008.

S.-C. Kao and R. S Govindaraju. A copula-based joint deficit index for droughts. *Journal of Hydrology*, 380(1-2):121–134, 2010.

W. Kleiber, A. E. Raftery, J. Baars, T. Gneiting, C. Mass, and E. P. Grimit. Locally calibrated probabilistic temperature foreasting using geostatistical model averaging and local Bayesian model averaging. *Monthly Weather Review*, 139(8):2630–2649, 2011a.

W. Kleiber, A. E. Raftery, and T. Gneiting. Geostatistical model averaging for locally calibrated probabilistic quantitative precipitation forecasting. *Journal of the American Statistical Association*, 106(496):1291–1303, 2011b.

W. H. Klein and H. R. Glahn. Forecasting local weather by means of model output statistics. *Bulleting of the American Meteorological Society*, 55(10):1217–1227, 1974.

S. L. Lauritzen. *Graphical Models*. Oxford University Press, 1996.

E. E. Leamer. *Specification Searches*. Wiley, 1978.

M. Leutbecher and T. N. Palmer. Ensemble forecasting. *Journal of Computational Physics*, 227:3515–3539, 2008.

F. Lindgren, H. Rue, and J. Lindström. An explicit link between Gaussian fields and Gaussian Markov random fields: The stochastic partial differential equation approach (with discussion). *Journal of the Royal Statistical Society, Series B*, 73(4):423–498, 2011.

R. S. Liptser and A. N. Shiryaev. *Statistics of Random Processes*. Springer, 2nd edition, 2010.

E. N. Lorenz. Deterministic nonperiodic flow. *Journal of the Atmospheric Sciences*, 20: 130–141, 1963.

G. Masarotto and C. Varin. Gaussian copula marginal regression. *Electronic Journal of Statistics*, 6:1517–1549, 2012.

B. Matérn. *Spatial Variation: Stochastic Models and their Application to some Problems in Forest Surveys and other Sampling Investigations*, volume 32 of *Lecture Notes in Statistics*. Springer, 2nd edition, 1986.

J. E. Matheson and R. L. Winkler. Scoring rules for continuous probability distributions. *Management Science*, 22:1087–1096, 1976.

A. McNeil and J. Nešlehová. Multivariate Archimedean copulas, d-monotone functions and ℓ_1-norm symmetric distributions. *Annals of Statistics*, 37:3059–3097, 2009.

T. Mikosch. Copulas: Tales and facts. *Extremes*, 9:3–20, 2006.

A. Möller, A. Lenkoski, and T. L. Thorarinsdottir. Multivariate probabilistic forecasting using Bayesian model averaging and copulas. *Quarterly Journal of the Royal Meteorological Society*, 139(673):982–991, 2013.

F. Molteni, R. Buizza, T. N. Palmer, and T. Petroliagis. The new ECMWF ensemble prediction system: Methodology and validation. *Quarterly Journal of the Royal Meteorological Society*, 122:73–119, 1996.

National Weather Service. Automated Surface Observing System (ASOS) User's Guide. Available at `http://www.weather.gov/asos/aum-toc.pdf`, 1998.

R. B. Nelsen. *An Introduction to Copulas*. Springer, New York, 2nd edition, 2006.

T. N. Palmer, J. Barkmeijer, R. Buizza, and T. Petroliagis. The ECMWF ensemble prediciton system. *Meteorological Applications*, 4:301–304, 1997.

P. Pinson. Adaptive calibration of (u, v)-wind ensemble forecasts. *Quarterly Journal of the Royal Meteorological Society*, 138(666):1273–1284, 2012.

M. Pitt, D. Chan, and R. Kohn. Efficient Bayesian inference for Gaussian copula regression models. *Biometrika*, 93:537–554, 2006.

R Development Core Team. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria, 2011. URL `http://www.R-project.org`.

A. E. Raftery, D. Madigan, and J. A. Hoeting. Bayesian model averaging for linear regression models. *Journal of the American Statistical Association*, 92(437):179–191, 1997.

A. E. Raftery, T. Gneiting, F. Balabdaoui, and M. Polakowski. Using Bayesian model averaging to calibrate forecast ensembles. *Monthly Weather Review*, 133:1155–1174, 2005.

S. Roquelaure and T. Bergot. A local ensemble prediction system for fog and low clouds: Construction, Bayesian model averaging calibration, and validation. *Journal of Applied Meteorology and Climatology*, 47:3072–3088, 2008.

Yu. A. Rozanov. Markov random fields and stochastic partial differential equations. *Math. USSR Sbornik*, 32(4):515–534, 1977.

H. Rue and L. Held. *Gaussian Markov Random Fields. Theory and Applications.* Chapman and Hall/CRC, 2005. 263 pp.

H. Rue and S. Martino. Approximate Bayesian inference for hierarchical Gaussian Markov random fields models. *Journal of Statistical Planning and Inference*, 137:3177 – 3192, 2007.

H. Rue, S. Martino, and N. Chopin. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximation (with discussion). *Journal of the Royal Statistical Society, Series B*, 71(2):319 – 392, 2009.

R. Schefzik. *Ensemble Copula Coupling*. Diploma thesis, Faculty of Mathematics and Informatics, University of Heidelberg, 2011.

R. Schefzik. Ensemble copula coupling as a multivariate discrete copula approach. *arXiv:1305.3445v1*, 2013.

R. Schefzik, T. L. Thorarinsdottir, and T. Gneiting. Uncertainty quantification in complex simulation models using ensemble copula coupling. *Statistical Science*, 28:616–640, 2013.

M. Scheuerer. Probabilistic quantitative precipitation forecasting using ensemble model output statistics. *Quarterly Journal of the Royal Meteorological Society*, 2013. doi: 10.1002/qj.2183.

M. Scheuerer and L. Büermann. Spatially adaptive post-processing of ensemble forecasts for temperature. *Journal of the Royal Statistical Society, Series C*, 2013. doi: 10.1111/rssc.12040.

C. Schoelzel and P. Friederichs. Multivariate non-normally distributed random variables in climate research - introduction to the copula approach. *Nonlinear Processes in Geophysics*, 15:761–772, 2008.

N. Schuhen, T. L. Thorarinsdottir, and T. Gneiting. Ensemble model output statistics for wind vectors. *Monthly Weather Review*, 140(10):3204–3219, 2012.

C. Sempi. Copulae: Some mathematical aspects. *Applied Stochastic Models in Business and Industry*, 27:37–50, 2011.

A. Sklar. Random variables, distribution functions, and copulas — a personal look backward and forward. In L. Rüschendorf, B. Schweizer, and M. D. Taylor, editors, *Distributions with Fixed Marginals and Related Topics*, pages 1–14. Institute of Mathematical Statistics, Hayward, 1996.

J. M. Sloughter, A. E. Raftery, T. Gneiting, and C. Fraley. Probabilistic quantitative precipitation forecasting using Bayesian model averaging. *Monthly Weather Review*, 135: 3209–3220, 2007.

J. M. Sloughter, T. Gneiting, and A. E. Raftery. Probabilistic wind speed forecasting using ensembles and Bayesian model averaging. *Journal of the American Statistical Association*, 105:25–35, 2010.

J. M. Sloughter, T. Gneiting, and A. E. Raftery. Probabilistic wind vector forecasting using ensembles and Bayesian model averaging. *Monthly Weather Review*, 141:2107–2119, 2013.

L. A. Smith. Disentangling uncertainty and error: On the predictability of nonlinear systems. In A. I. Mees, editor, *Nonlinear Dynamics and Statistics*, pages 31–64. Birkhäuser, Boston, 2001.

P. Xu-Kung Song. Multivariate dispersion models generated from Gaussian copula. *Scandinavian Journal of Statistics*, 27(2):305–320, 2000.

M. L. Stein. *Interpolation of Spatial Data: Some Theory for Kriging*. Springer, 1999.

O. Talagrand, R. Vautard, and B. Strauss. Evaluation of probabilistic prediction systems. In *Proc. Workshop on Predictability*, pages 1–25. Reading, UK, European Centre for Medium-Range Weather Forecasts, 1997.

T. L. Thorarinsdottir and T. Gneiting. Probabilistic forecasts of wind speed: Ensemble model output statistics using heteroskedastic censored regression. *Journal of the Royal Statistical Society Ser. A*, 173:371–388, 2010.

L. Tierney and J. B. Kadane. Accurate approximations for posterior moments and marginal densities. *Journal of the American Statistical Association*, 81:82 –86, 1986.

Z. Toth and E. Kalnay. Ensemble forecasting at NMC: The generation of perturbations. *Bulletin of the American Meteorological Society*, 74:2317–2330, 1993.

Yehuda Vardi and Cun-Hui Zhang. The multivariate L_1-median and associated data depth. *Proceedings of the National Academy of Sciences of the United States of America*, 97: 1423–1426, 2000.

J. S. Whitaker and A. F. Loughe. The relationship between ensemble spread and ensemble mean skill. *Monthly Weather Review*, 126:3292–3302, 1998.

J. Whittaker. *Graphical Models in Applied Multivariate Statistics*. Wiley, 1990.

P. Whittle. On stationary processes in the plane. *Biometrika*, 41:434 – 449, 1954.

P. Whittle. Stochastic processes in several dimensions. *Bulletin of the International Statistical Institute*, 40:974 – 994, 1963.

D. S. Wilks. *Statistical Methods in the Atmospheric Sciences*. Elsevier Academic Press, 2nd edition, 2006.

D. S. Wilks. Extending logistic regression to provide full-probability-distribution MOS forecasts. *Meteorological Applications*, 16:361–368, 2009.

D. S. Wilks and T. M. Hamill. Comparison of ensemble-MOS methods using GFS reforecasts. *Monthly Weather Review*, 135:2379–2390, 2007.

L. J. Wilson, S. Beauregard, A. E. Raftery, and R. Verret. Calibrated surface temperature forecasts from the Canadian ensemble prediction system using Bayesian model averaging. *Monthly Weather Review*, 135:1364–1385, 2007.