Technische Universität München

Fakultät für Mathematik

Predicting temperature time series using spatial vine copulae

Master's Thesis

von

Tobias M. Erhardt

Themenstellerin:	Prof. Claudia Czado, Ph.D.
Betreuer:	Ulf Schepsmeier
Abgabetermin:	20. August 2013

Hiermit erkläre ich, dass ich die Master's Thesis selbständig und nur mit den angegebenen Hilfsmitteln angefertigt habe.

München, den 20. August 2013

Acknowledgments

Contents

Chapter 1

Introduction

Weather prediction and prediction of temperatures has a long history. Due to the fact that the lives of humans have always been affected by weather in lots of different fields like agriculture, economy in general or even in everyday life, there was always great interest in predicting the weather. However it is not possible to fix a specific date of birth of weather prediction.

In this thesis we are going to present new, vine copula based approaches for weather prediction. Utilization of available spatial information for the purpose of model building will lead to a distinct reduction in the number of parameters needed to parametrize these high dimensional vine copula models.

Copulae are multivariate distribution functions which can be understood as a tie between a distribution function and its marginals and which capture all dependency information. Vine copulae, which were introduced by Bedford and Cooke (2001, 2002) and trace back to ideas of Joe (1996), are constructions of d-dimensional copulae built on bivariate copulae only. Bivariate copulae are well understood and easy to compute. Due to the paper of Aas, Czado, Frigessi, and Bakken (2009) on pair-copula constructions, regular vines (a special class of vine copulae) got into the focus of a broad field of research.

Throughout the years lots of different methods and ideas of weather prediction and the (joint) modeling of weather characteristic variables were developed. The research was intensified in several different directions. The bandwidth of methods range from simple heuristic predictions to complex mathematical and physical models which for example rely on dynamical systems. Nowadays also statistical methods play an important role in weather prediction. In practice approaches from different disciplines are used in a combined fashion. All modern weather prediction methods have in common, that they are based on data measurements from a dense network of observation stations all across the world and data collected by satellites, amongst others.

Also the area of geostatistics or more general spatial statistics aims to propose suitable models and methods to handle such kind of data. Georges Matheron (1962) was the one who established geostatistics. He was amongst others inspired by work of the South African mining engineer Danie G. Krige, after whom Matheron named the popular method of spatial prediction called kriging. The book of Cressie (1991) is still considered as an extensive standard reference on spatial statistics, covering a wide range of classical spatial statistical methods. A continuation of this book to the treatment of spatio-temporal data

is provided by Cressie and Wikle (2011).

On February 21th, 2013 a Spatial Copula Day organized by Prof. Claudia Czado and Ulf Schepsmeier was held at the Chair of Mathematical Statistics at the Technische Universität München. The aim of the seminar was to present and discuss different recent approaches to model spatial data with the help of copulae, which allow a more flexible modeling compared to the classical geostatistical approaches. Amongst others, first ideas of the modeling approaches presented in this master's thesis which are based on vine copulae were discussed.

In the following we are going to present two new approaches intended to model the spatial dependencies of spatio-temporal data. Furthermore, prediction methods based on the presented models are introduced. The exploratory investigations initiating the development of the spatial models are based on daily mean temperature data collected over the period 01/01/2010-12/31/2012 by the German Meteorological Service (Deutscher Wetterdienst) at 54 selected observation stations across Germany. The following investigations on the goodness of the models fitted to the training data set are conducted on a validation data set consisting of mean temperature time series over the same period as before for 24 additional observation stations in Germany.

Subsequent to this introduction we prepare the model building process by clarification of some fundamental concepts applied throughout the thesis. Besides some theory on copulae, the dependence measure Kendall's τ is introduced and the basic statistical concepts of least squares estimation and linear regression are repeated. Moreover, Ljung-Box tests are introduced and a summary on elementary graph theory needed to organize the vine construction in trees is provided. Furthermore, a tool called Fisher z-transform is presented and two different types of scoring rules which will be used for the purpose of model validation are given.

The actual model building process is already initiated in Chapter 3. Bit by bit we develop a joint model for the marginal densities corresponding to the variables defined by the 54 observation stations of the training data. In the end of this chapter the calculation of so called copula data by means of the joint marginal model is explained.

Chapter 4 summarizes the basic findings on regular vines (R-vines) of the last years, which are required to understand the models presented in the subsequent chapters. The theoretical part on regular vine models is followed by an analysis of the spatial dependencies of the copula data obtained in Chapter 3, which includes the investigation of the structure of an R-vine model fitted to the data.

The investigations in Chapter 4 initiate the model building process of the first spatial dependency model presented in Chapter 5. The new model which we name spatial R-vine model relies on a reparametrization of an R-vine model, which exploits the relationship between the model parameters and the available spatial information. Different reparametrizations based on distances and elevation differences are compared and the most promising reparametrization is chosen. An explanation of maximum likelihood estimation of the model parameters and its performance are followed by the illustration of the prediction at a new location based on the specified spatial R-vine model.

The model presented in Chapter 6 is built on the theory of composite likelihood methods. First the necessary theory on composite likelihood models is introduced, then the new model structure is set up. The presented model is a composition of low dimensional

canonical vine models which model dependencies locally. Canonical vines (C-vines) are a special class of vine copulae, having a star like tree structure with a unique root node. Similarly to the spatial R-vine models a joint reparametrization of the C-vine copula parameters is selected. This time the parameters are estimated by maximum composite likelihood estimation. Prediction at new locations is performed by simulation from conditional distributions which are constructed based on the newly developed model.

Finally an evaluation and comparison of the new models is conducted in Chapter 8. For this purpose a further spatial model is introduced in Chapter 7, which is set up by means of classical geostatistical methods.

Chapter 2

Preliminaries

Throughout this thesis we are going to repeatedly apply different kinds of basic tools. We briefly present the most important of these recurring concepts in the following.

First of all copulae, the basis for all of our new developed models, are introduced in Section 2.1. The second section treats dependence measures like Kendall's τ, which plays a central role in the modeling process of the spatial models presented in this thesis. Section 2.3 summarizes two important classes of bivariate copulae, elliptical and Archimedean copulae. The basic statistical concepts of least squares and linear regression are repeated in Section 2.4. Furthermore, the so called Ljung-Box test, a test for serial autocorrelation is presented in Section 2.5. A short summary of some elementary graph theory concepts is provided with Section 2.6. Moreover Section 2.7 gives a definition of the Fisher z-transform and its inverse. Finally two different types of scoring rules are presented in Section 2.8.

2.1   Copulae

Copulae are an important tool to analyze the multivariate dependence structure of random variables. The fundamental building blocks of our models are bivariate copulae, so we will give a short overview over copulae in the following, including a definition, some of their main properties and a concept called pair-copula construction, which plays an important role in the model building process of the subsequent chapters. Different types and examples of bivariate copulae are given in Section 2.3.

For a penetrative introduction to copulae we refer to Nelsen (2006). Our short introduction here is based on Joe (2001, Section 1.6) and Fischer (2011).

2.1.1   Definition and properties

Generally speaking, copulae are multivariate cumulative distribution functions on the unit hypercube $[0, 1]^d$ with uniform margins. For our purpose and for the sake of convenience it is sufficient to give a definition for the bivariate case.

Definition 2.1 (Copula, bivariate case)**.** Let $U, V \sim \mathcal{U}(0, 1)$. Then the bivariate cumulative distribution function $C : [0, 1]^2 \to [0, 1]$ with $(U, V) \sim C$ is called *(bivariate) copula.*   ⌟

The fundamental theorem of Sklar (1959) associates a copula C to any arbitrary (bivariate) distribution function $\boldsymbol{F}$.

Theorem 2.2 (Sklar, bivariate case)**.**

(i) *Let $(X, Y) \sim \boldsymbol{F}$, where $\boldsymbol{F}$ is a bivariate cumulative distribution function and F_X, F_Y are the continuous marginal distribution functions of X and Y, respectively. Then there exists a unique copula $C : [0, 1]^2 \to [0, 1]$ such that for all $(x, y)^\top \in \mathbb{R}^2$ holds, that*

$$\boldsymbol{F}(x, y) = C\left(F_X(x), F_Y(y)\right). \tag{2.1.1}$$

(ii) *Let F_X, F_Y be arbitrary univariate distribution functions and C be a bivariate copula. Then $\boldsymbol{F} : \mathbb{R}^2 \to [0, 1]$ defined through Equation (2.1.1) is a bivariate distribution function with marginals F_X and F_Y.*

Remark 2.3.

(i) Equation (2.1.1) explains where the name copula stems from. The copula C "couples" the marginal distributions F_X and F_Y to the joint distribution $\boldsymbol{F}$.

(ii) If $\boldsymbol{F}$ in Equation (2.1.1) is continuous, and F_X^{-1}, F_Y^{-1} are the quantile functions of F_X and F_Y, then the uniquely determined copula C is given by

$$C(u, v) = \boldsymbol{F}\left(F_X^{-1}(u), F_Y^{-1}(v)\right). \tag{2.1.2}$$

$\lrcorner$

An important concept which is frequently used in the context of copulae is the *probability integral transform*. It says that on the one hand it holds for a univariate random variable $Z \sim F$ that $U = F(Z) \sim \mathcal{U}(0, 1)$ and on the other hand $Z = F^{-1}(U) \sim F$ for $U \sim \mathcal{U}(0, 1)$ and F a continuous distribution function. If we consider now the setting given in Remark 2.3 (ii), this yields, that $(F_X(X), F_Y(Y)) \sim C$, if $(X, Y) \sim \boldsymbol{F}$. Moreover if $(U, V) \sim C$, then $(F_X^{-1}(U), F_Y^{-1}(V)) \sim \boldsymbol{F}$.

Furthermore we give some basic properties of (bivariate) copulae in the following.

Proposition 2.4. *Let $C : [0, 1]^2 \to [0, 1]$ be a bivariate copula and $u, v \in [0, 1]$. Then*

$$\begin{aligned}
C(u, 0) &= 0, \\
C(0, v) &= 0, \\
C(u, 1) &= u \quad \text{and} \\
C(1, v) &= v.
\end{aligned}$$

These properties can be easily checked in the case of Remark 2.3 (ii) or for the following basic examples of bivariate copulae.

Example 2.5 (Independence copula and Fréchet-Hoeffding bounds)**.** Let $u, v \in [0, 1]$.

(i) *Independence copula*: In the case of independent random variables X, Y, Equation (2.1.1) can be written as $F_X(x) \cdot F_Y(y) = C\left(F_X(x), F_Y(y)\right)$, which leads to the independence copula

$$C^\perp(u, v) = u \cdot v.$$

(ii) *Fréchet-Hoeffding bounds*: The upper (bivariate) Fréchet-Hoeffding bound is given through

$$C^U(u, v) := \min\left\{u, v\right\},$$

whereas the lower (bivariate) Fréchet-Hoeffding bound is defined as

$$C^L(u, v) := \max\left\{u + v - 1, 0\right\}.$$

For every arbitrary copula the inequality

$$C^L(u, v) \leq C(u, v) \leq C^U(u, v)$$

holds.
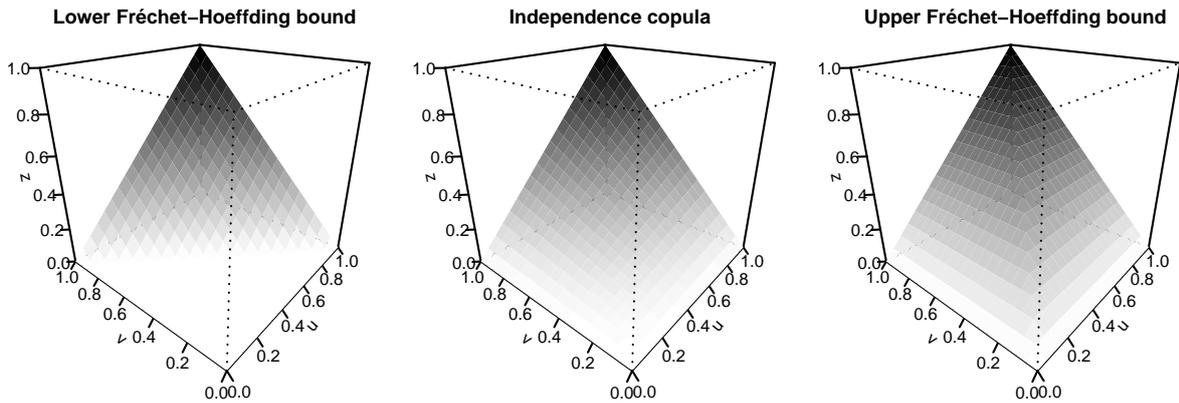
Figure 2.1.1: Visualization of the three copulae of Example 2.5. Left: Lower Fréchet-Hoeffding bound $(z = C^L(u, v))$. Middle: Independence copula $(z = C^\perp(u, v))$. Right: Upper Fréchet-Hoeffding bound $(z = C^U(u, v))$.

2.1.2   Pair-copula construction

Following Aas, Czado, Frigessi, and Bakken (2009) we briefly introduce the concept of *pair-copula construction (PCC)*. For some theoretical background on pair-copula constructions we refer the reader to Joe (1996), Bedford and Cooke (2001, 2002) and Kurowicka and Cooke (2006).

Aas et al. (2009) summarize the pair-copula construction method in a concise fashion: The "radically new way of constructing complex multivariate highly dependent models [...] is based on a decomposition of a multivariate density into a cascade of pair copulae, applied on original variables and on their conditional and unconditional distribution functions." Moreover they (Aas et al., 2009) call it "a simple and powerful tool for model building."

For a formal explanation of pair-copula constructions we consider a random vector $\boldsymbol{X} = (X_1, \ldots, X_d)^\top$, with absolute continuous distribution function $\boldsymbol{F}$ and continuous, strictly increasing marginal densities, denoted as $F_1, \ldots, F_d$. Furthermore we make the general assumption, that pair copulae do not depend on the conditioning variables. Usually this assumption is called *simplifying assumption*.

First, the joint probability density function $f(x_1, \ldots, x_d)$ of $\boldsymbol{X}$ can be decomposed uniquely (aside from relabeling) into a product of a marginal density and $d-1$ conditional densities, i.e.

$$f(x_1, \ldots, x_d) = f_1(x_1) \cdot f_{2|1}(x_2 \,|\, x_1) \cdot f_{3|12}(x_3 \,|\, x_1, x_2) \cdots f_{d|1\cdots d-1}(x_d \,|\, x_1, \ldots, x_{d-1}). \quad (2.1.3)$$

The next step is to depict the $d-1$ conditional densities in (2.1.3) in terms of pair-copulae respectively bivariate copulae and marginal densities. From (2.1.1) we obtain

$$f_{12}(x_1, x_2) = \frac{\partial}{\partial x_1} \frac{\partial}{\partial x_2} \boldsymbol{F}_{12}(x_1, x_2) = c_{12}\left(F_1(x_1), F_2(x_2)\right) \cdot f_1(x_1) \cdot f_2(x_2) \quad (2.1.4)$$

by partial differentiation, where $c_{12}\left(\cdot\right)$ denotes the density of the bivariate copula corresponding to $\boldsymbol{F}_{12}$ respectively f_{12}. From this follows

$$f_{2|1}(x_2 \,|\, x_1) = c_{12}\left(F_1(x_1), F_2(x_2)\right) \cdot f_2(x_2) \quad (2.1.5)$$

for the second term in (2.1.3). The calculations to get the third factor of (2.1.3) are a little bit more involved. We advance as follows: First we calculate

$$f_{3|12}(x_3 \,|\, x_1, x_2) = \frac{f_{123}(x_1, x_2, x_3)}{f_{12}(x_1, x_2)} = \frac{f_{23|1}(x_2, x_3 \,|\, x_1) \cdot f_1(x_1)}{f_{12}(x_1, x_2)} = \frac{f_{23|1}(x_2, x_3 \,|\, x_1)}{f_{2|1}(x_2 \,|\, x_1)}.$$

Together with

$$
\begin{aligned}
f_{23|1}(x_2, x_3 \,|\, x_1) &= \frac{\partial}{\partial x_2} \frac{\partial}{\partial x_3} \boldsymbol{F}_{23|1}(x_2, x_3 \,|\, x_1) \\
&= \frac{\partial}{\partial x_2} \frac{\partial}{\partial x_3} C_{23;1}\left(\left(F_{2|1}(x_2 \,|\, x_1), F_{3|1}(x_3 \,|\, x_1)\right)\right) \\
&= c_{23;1}\left(F_{2|1}(x_2 \,|\, x_1), F_{3|1}(x_3 \,|\, x_1)\right) \cdot f_{2|1}(x_2 \,|\, x_1) \cdot f_{3|1}(x_3 \,|\, x_1)
\end{aligned}
$$

and a relabeled version of (2.1.5) we obtain

$$f_{3|12}(x_3 \,|\, x_1, x_2) = c_{23;1}\left(F_{2|1}(x_2 \,|\, x_1), F_{3|1}(x_3 \,|\, x_1)\right) \cdot c_{13}\left(F_1(x_1), F_3(x_3)\right) \cdot f_3(x_3). \quad (2.1.6)$$

We easily recognize that this decomposition is not unique. If we would have decomposed $f_{123}(x_1, x_2, x_3)$ in the first step as

$$f_{123}(x_1, x_2, x_3) = f_{13|2}(x_1, x_3 \,|\, x_2) \cdot f_2(x_2)$$

an analogous derivation would have led to the decomposition

$$f_{3|12}(x_3 \,|\, x_1, x_2) = c_{13;2}\left(F_{1|2}(x_1 \,|\, x_2), F_{3|2}(x_3 \,|\, x_2)\right) \cdot c_{23}\left(F_2(x_2), F_3(x_3)\right) \cdot f_3(x_3). \quad (2.1.7)$$

We conclude that in general there is no unique pair-copula construction.

Before we finish the general d-dimensional case of the decomposition, we have a look at the following three-dimensional example.

Example 2.6 (Pair-copula construction, $d = 3$)**.** According to (2.1.3) we obtain

$$f(x_1, x_2, x_3) = f_1(x_1) \cdot f_{2|1}(x_2 \,|\, x_1) \cdot f_{3|12}(x_3 \,|\, x_1, x_2)$$

for $d = 3$. With (2.1.5), (2.1.6) and (2.1.7) we get the two alternative pair-copula constructions

$$\begin{aligned} f(x_1, x_2, x_3) = {}& f_1(x_1) \cdot f_2(x_2) \cdot f_3(x_3) \\ & \cdot c_{12}\left(F_1(x_1), F_2(x_2)\right) \cdot c_{13}\left(F_1(x_1), F_3(x_3)\right) \\ & \cdot c_{23;1}\left(F_{2|1}(x_2 \,|\, x_1), F_{3|1}(x_3 \,|\, x_1)\right) \end{aligned}$$

and

$$\begin{aligned} f(x_1, x_2, x_3) = {}& f_1(x_1) \cdot f_2(x_2) \cdot f_3(x_3) \\ & \cdot c_{12}\left(F_1(x_1), F_2(x_2)\right) \cdot c_{23}\left(F_2(x_2), F_3(x_3)\right) \\ & \cdot c_{13;2}\left(F_{1|2}(x_1 \,|\, x_2), F_{3|2}(x_3 \,|\, x_2)\right). \end{aligned}$$

Note that there is (only) one further way to decompose the density $f(x_1, x_2, x_3)$, i.e. there are exactly three different pair-copula constructions in the three-dimensional case. ⌐

To finalize the general case of the pair-copula construction we need a generalized formula for the conditional densities

$$f_{i|1\cdots i-1}(x_i \,|\, x_1, \ldots, x_{i-1}), \quad i = 2, \ldots, d,$$

in (2.1.3). To shorten the notation, we write

$$\boldsymbol{x}_{\mathcal{I}} = \{x_l : l \in \mathcal{I}\},$$

for arbitrary index sets $\mathcal{I}$. Then we obtain

$$f_{i|1\cdots i-1}(x_i \,|\, x_1, \ldots, x_{i-1}) = c_{ij;\mathcal{I}_{-j}}\left(F_{i|\mathcal{I}_{-j}}(x_i \,|\, \boldsymbol{x}_{\mathcal{I}_{-j}}), F_{j|\mathcal{I}_{-j}}(x_j \,|\, \boldsymbol{x}_{\mathcal{I}_{-j}})\right) f_{i|\mathcal{I}_{-j}}(x_i \,|\, \boldsymbol{x}_{\mathcal{I}_{-j}}),$$

analogous to the derivation in the three-dimensional special case above, where $i = 2, \ldots, d$, j is an arbitrary index out of $\{1, \ldots, i - 1\}$,

$$\mathcal{I}_{-j} := \{1, \ldots, i-1\} \backslash \{j\} = \{1, \ldots, j-1, j+1, \ldots, i-1\}$$

is the index set up to $i - 1$, where the j-th index is removed, and $c_{ij;\mathcal{I}_{-j}}(\cdot, \cdot)$ is a bivariate copula density. This formula allows it to rewrite (2.1.3) in the desired product form, where all building blocks are bivariate copulae or marginal densities.

It remains to give an expression for the calculation of the conditional distribution functions which occur in the pair-copula construction. Therefore we consider the two indices $k, l \in \{1, \ldots, d\}$, $k \neq l$, and an arbitrary index set $\mathcal{J}$ which satisfies $\{l\} \subset \mathcal{J} \subset \{1, \ldots, d\} \backslash \{k\}$. Moreover we define $\mathcal{J}_{-l} := \mathcal{J} \backslash \{l\}$. According to Joe (1996) we have

$$F_{k|\mathcal{J}}(x_k \,|\, \boldsymbol{x}_{\mathcal{J}}) = \frac{\partial \, C_{kl;\mathcal{J}_{-l}}\left(F_{k|\mathcal{J}_{-l}}(x_k \,|\, \boldsymbol{x}_{\mathcal{J}_{-l}}), F_{l|\mathcal{J}_{-l}}(x_l \,|\, \boldsymbol{x}_{\mathcal{J}_{-l}})\right)}{\partial F_{l|\mathcal{J}_{-l}}(x_l \,|\, \boldsymbol{x}_{\mathcal{J}_{-l}})}, \tag{2.1.8}$$

where $C_{kl;\mathcal{J}_{-l}}\left(\cdot,\cdot\right)$ is a bivariate copula. In the univariate case, where $\mathcal{J} = \{l\}$ holds, (2.1.8) can be written as

$$F_{k|l}(x_k\,|\,x_l) = \frac{\partial\,C_{kl}\left(F_k(x_k), F_l(x_l)\right)}{\partial F_l(x_l)}. \tag{2.1.9}$$

The special case of Equation (2.1.9), where the univariate margins are uniform, i.e. $u_k \coloneqq F_k(x_k) = x_k$, $u_l \coloneqq F_l(x_l) = x_l$, and a dependence of the copula $C_{kl;\mathcal{J}_{-l}}\left(\cdot,\cdot\right)$ on some parameters $\boldsymbol{\theta}$ is indicated, plays an important role in some simulation and maximum likelihood estimation algorithms. That's why it got its own name in the literature. The so called *h-functions* are defined as

$$h(u_k, u_l, \boldsymbol{\theta}) \coloneqq \frac{\partial\,C_{kl}\left(u_k, u_l; \boldsymbol{\theta}\right)}{\partial u_l}.$$

This concludes the section about pair-copula constructions. In Chapter 4 the developed formulas and the concept of pair-copula construction will be picked up again and applied to a special case.

2.2   Dependence measures

The data investigated in this thesis stems from spatially spread observation stations. The reason why we investigate such kind of data is the spatial dependence of the observations at different locations. Hence we need to introduce some measures of dependence, to be able to model and to capture these dependencies. We will focus on two very basic dependence measures, Kendall's τ and tail dependence. For further dependence concepts and details we refer the reader to Joe (2001, Chapter 2).

2.2.1   Kendall's τ

Kendall's τ plays a central role in the new spatial models that are introduced within this work, as the parametrization and the dependence structure of these models can be characterized via Kendall's τ's. Another rank correlation measure is the so called Spearman's ρ. For an extensive introduction to rank correlation methods see for example Kendall (1970). Besides we refer to Joe (2001, Section 2.1) and Nelsen (2006, Chapter 5).

Definition 2.7 (Kendall's τ)**.** Let (X, Y) and (X', Y') be independent pairs of continuous random variables, each with bivariate, cumulative distribution function $\boldsymbol{F}$. Then the rank correlation coefficient *Kendall's τ* is defined as

$$\begin{aligned} \tau_{X,Y} &\coloneqq \mathrm{P}((X - X')(Y - Y') > 0) - \mathrm{P}((X - X')(Y - Y') < 0) \\ &= 2\,\mathrm{P}((X - X')(Y - Y') > 0) - 1. \end{aligned} \tag{2.2.1}$$

The first term in (2.2.1) is the probability that the two pairs (X, Y) and (X', Y') are *concordant*, i.e. that both components of one of the two pairs are greater than the corresponding components of the other pair. In formulas this means, that either $X > X'$ and $Y > Y'$ or $X < X'$ and $Y < Y'$. Accordingly the second term in (2.2.1) represents the probability that one component of a pair is larger than the corresponding one of the other pair and the other component is smaller, i.e. $X > X'$ and $Y < Y'$ or $X < X'$ and $Y > Y'$. In these two cases the two pairs are called *discordant*.

Now we want to give a definition for an empirical version of Kendall's τ. Therefore we have to note, that in general concordance and disconcordance are not the only cases that can occur. It is also possible that so called *ties* occur. This is the case when either $X = X'$, $Y = Y'$ or $X = X'$ and $Y = Y'$. But if we consider only continuous random variables, ties occur with a probability of zero.
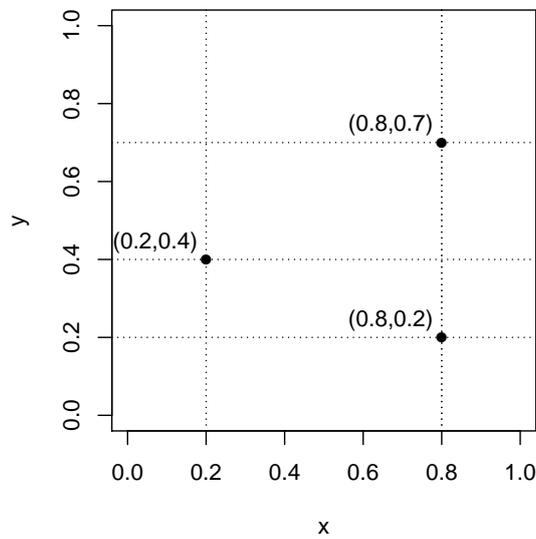
Figure 2.2.1: Visualization of concordance, disconcordance and ties. The points $(0.2, 0.4)$ and $(0.8, 0.7)$ are concordant, $(0.2, 0.4)$ and $(0.8, 0.2)$ are disconcordant and $(0.8, 0.2)$ and $(0.8, 0.7)$ are tied.

Definition 2.8 (Empirical Kendall's τ in the case of no ties). Let (x_i, y_i), $i = 1, \ldots, N$, be N independently sampled observations of a pair of continuous random variables $(X, Y) \sim \boldsymbol{F}$. Then the *empirical Kendall's* τ is defined as

$$\tau = \frac{c}{\binom{N}{2}} - \frac{d}{\binom{N}{2}} = \frac{2c}{\binom{N}{2}} - 1, \tag{2.2.2}$$

where

$$c = \# \text{ concordant pairs} = \# \left\{ i < j : x_i < x_j, y_i < y_j \text{ or } x_i > x_j, y_i > y_j \right\}$$

and

$$d = \# \text{ disconcordant pairs} = \# \left\{ i < j : x_i < x_j, y_i > y_j \text{ or } x_i > x_j, y_i < y_j \right\}.$$

Note that $\binom{N}{2} = N(N-1)/2 = c + d$ is the total number of distinct pairs (x_i, y_i), (x_j, y_j), $i < j$. Comparison of (2.2.1) and (2.2.2) yields that the empirical Kendall's τ is defined analogously to the theoretical version of Definition 2.7.

From Definition 2.7 some useful properties of Kendall's τ can be derived.

Proposition 2.9 (Properties of Kendall's τ). *Let* $(X, Y) \sim \boldsymbol{F}$ *be a pair of continuous random variables with copula C. Then:*

(i) $-1 \leq \tau_{X,Y} \leq 1$;

(ii) *if* X, Y *are independent, it holds* $\tau_{X,Y} = 0$;

(iii) *if* $g : \mathbb{R} \to \mathbb{R}$ *is a strictly increasing function and* $Y = g(X)$, *then* $\tau_{X,Y} = 1$;

(iv) *if* $g : \mathbb{R} \to \mathbb{R}$ *is a strictly decreasing function and* $Y = g(X)$, *then* $\tau_{X,Y} = -1$;

(v) *if* $g_1 : \mathbb{R} \to \mathbb{R}$, $g_2 : \mathbb{R} \to \mathbb{R}$ *are strictly increasing functions, then* $\tau_{g_1(X), g_2(Y)} = \tau_{X,Y}$;

(vi) $\tau_{X,Y} = 4 \int C \mathrm{d}C - 1$.

While properties (i)-(iv) of Proposition 2.9 are the typical properties which we expect to be fulfilled by correlation measures, the properties (v) and (vi) depict the great advantage of the rank correlation measure Kendall's τ. Property (vi) yields, that $\tau_{X,Y}$ only depends on the copula and not on the hole bivariate distribution function $\boldsymbol{F}$. This means, that the dependence given through the copula is captured in τ.

2.2.2 Tail dependence

Additionally to Kendall's τ, which is a measure for overall correlation, we consider the dependence concept called tail dependence, measuring correlation in the tails of multivariate distributions. For this purpose we follow the approach of Joe (2011). For properties of tail dependence in the context of vine copulae (which will be introduced later on), we advert to Joe, Li, and Nikoloulopoulos (2010).

In contrast to rank correlation measures, tail dependence is a measure for the dependence of extreme events. Informally spoken there exists tail dependence if there is a positive probability that an extreme value close to 0 or 1 within $[0, 1]$ occurs for one variable of a copula, given that an other variable already takes an extreme value close to one bound of $[0, 1]$.

Definition 2.10 (Positive tail dependence in the case of bivariate copulae). Let $(U, V) \sim C$. If the limit

$$\lambda_{LL} := \lim_{u \searrow 0} \mathrm{P}(U \leq u \,|\, V \leq u) = \lim_{u \searrow 0} \mathrm{P}(V \leq u \,|\, U \leq u) = \lim_{u \searrow 0} \frac{C(u, u)}{u}$$

exists and $\lambda_{LL} \in (0, 1]$, then C has *lower tail dependence*. If $\lambda_{LL} = 0$, C has no lower tail dependence. Accordingly, if the limit

$$\lambda_{UU} := \lim_{u \nearrow 1} \mathrm{P}(U > u \,|\, V > u) = \lim_{u \nearrow 1} \mathrm{P}(V > u \,|\, U > u) = \lim_{u \nearrow 1} \frac{1 - 2u + C(u, u)}{1 - u}$$

exists and $\lambda_{UU} \in (0,1]$, then C has *upper tail dependence*. If $\lambda_{UU} = 0$, C has no upper tail dependence. ⌋

Definition 2.11 (Negative tail dependence in the case of bivariate copulae)**.** Let $(U, V) \sim C$. If the limit

$$\lambda_{LU} := \lim_{u \nearrow 1} \mathrm{P}(1 - U > u \,|\, V > u) = \lim_{u \searrow 0} \mathrm{P}(1 - V \leq u \,|\, U \leq u) = \lim_{u \searrow 0} \frac{u - C(u, 1 - u)}{u}$$

exists and $\lambda_{LU} \in (0,1]$, then C has *(negative) lower-upper tail dependence*. If $\lambda_{LU} = 0$, C has no (negative) lower-upper tail dependence. Accordingly, if the limit

$$\lambda_{UL} := \lim_{u \nearrow 1} \mathrm{P}(U > u \,|\, 1 - V > u) = \lim_{u \searrow 0} \mathrm{P}(V \leq u \,|\, 1 - U \leq u) = \lim_{u \searrow 0} \frac{u - C(1 - u, u)}{1 - u}$$

exists and $\lambda_{UL} \in (0,1]$, then C has *(negative) upper-lower tail dependence*. If $\lambda_{UL} = 0$, C has no (negative) upper-lower tail dependence. ⌋

From the definitions follows, that $\lambda_{LL}, \lambda_{UU}, \lambda_{LU}, \lambda_{UL} \in [0,1]$. Figure 2.2.2 illustrates the different types of tail dependence. The corresponding Kendall's τ's are indicated in brackets.

2.3 Important classes of bivariate copulae

In this section we want to present some important bivariate copulae, which are utilized in the model building processes of the subsequent sections. The copulae, which we take into consideration, stem from two different classes of copulae, the elliptical and the Archimedean copulae. For a broad study of different kinds of copulae, we refer the reader to Joe (2001).

2.3.1 Elliptical copulae

According to Fischer (2011), *elliptical copulae* are copulae that are linked to elliptical distributions. They are applied in different areas such as statistics and econometrics. To construct a bivariate elliptical copula, (2.1.2) is employed. In the following we highlight the construction and some properties of two prominent (bivariate) elliptical copulae, the Gaussian and the Student-t copula in two dimensions.

Example 2.12 (Gaussian copula, bivariate case)**.** Let $\mathbf{\Phi}_\rho$ be the distribution function of the bivariate normal distribution $\mathcal{N}_2(\mathbf{0}, \Sigma)$ with zero mean, unit variances and correlation $\rho \in (-1, 1)$, i.e. covariance matrix

$$\Sigma = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}.$$

Furthermore let Φ be the distribution function of the standard normal distribution $\mathcal{N}(0, 1)$ and Φ^{-1} the corresponding quantile function. Application of (2.1.2) yields the (bivariate) Gaussian copula

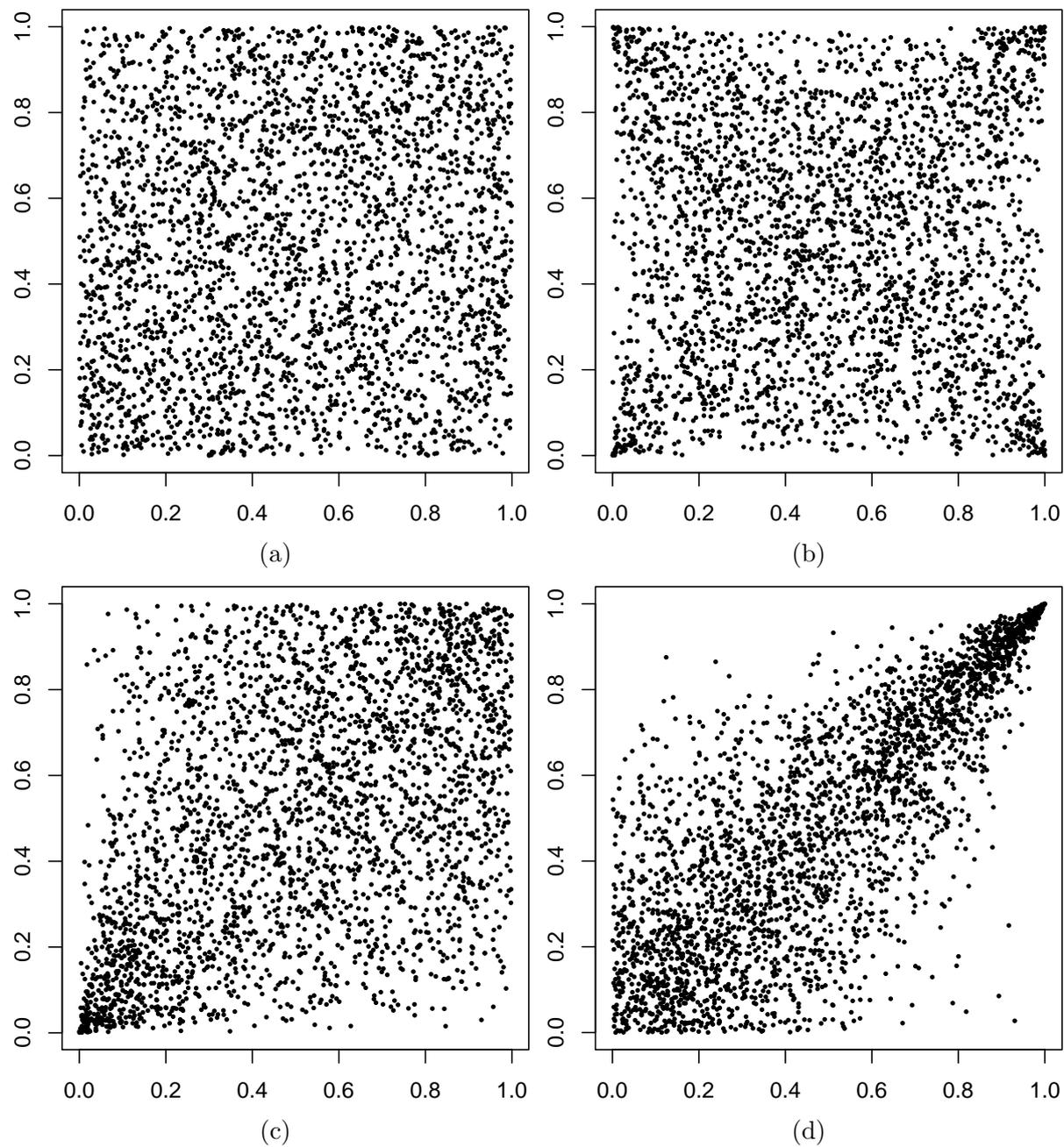$$C(u, v; \rho) = \mathbf{\Phi}_\rho \left(\Phi^{-1}(u), \Phi^{-1}(v) \right).$$

Figure 2.2.2: Sample scatter plots of bivariate copulae with (a) no tail dependence ($\tau \approx 0.06$), (b) all four different types of tail dependence ($\tau = 0$), (c) lower tail dependence ($\tau \approx 0.33$) and (d) upper tail dependence ($\tau \approx 0.61$).

Its density is given as

$$c\left(u, v; \rho\right) = \frac{1}{\sqrt{1-\rho^2}} \exp\left\{-\frac{x^2+y^2-2\rho xy}{2(1-\rho^2)}\right\} \exp\left\{\frac{1}{2}\left(x^2+y^2\right)\right\},$$

where $x = \Phi^{-1}(u)$ and $y = \Phi^{-1}(v)$ (cp. Joe, 2001). One can easily see that the density degenerates to 1 if $\rho = 0$, i.e. for $\rho = 0$ the Gaussian copula coincides with the independence copula $C^{\perp}$.

Example 2.13 (Student-t copula, bivariate case)**.** Compare Demarta and McNeil (2005) for the following derivation of the Student-t copula and the stated properties. Let $\boldsymbol{T}_{\nu,R}$ be the cumulative distribution function of the bivariate Student-t distribution with $\nu > 2$ degrees of freedom and density corresponding to

$$t_\nu(x,y) = \frac{\Gamma\left(\frac{\nu+2}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)\sqrt{(\pi\nu)^2|R|}}\left(1+\frac{(x,y)R^{-1}(x,y)^\top}{\nu}\right)^{-(\nu+2)/2},$$

where

$$R = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$$

and $\rho \in (-1,1)$. Furthermore let T_ν be the cumulative distribution function of the univariate Student-t distribution with $\nu > 2$ degrees of freedom and T_ν^{-1} the corresponding quantile function. Application of (2.1.2) yields the (bivariate) Student-t copula

$$C\left(u,v;\rho\right) = \boldsymbol{T}_{\nu,R}\left(T_\nu^{-1}(u), T_\nu^{-1}(v)\right).$$

Its density is provided by

$$c\left(u,v;\rho\right) = \frac{1}{\sqrt{1-\rho^2}}\frac{\Gamma\left(\frac{\nu+2}{2}\right)\Gamma\left(\frac{\nu}{2}\right)}{\left[\Gamma\left(\frac{\nu+1}{2}\right)\right]^2}\cdot\frac{\left[\left(1+\frac{x^2}{\nu}\right)\left(1+\frac{y^2}{\nu}\right)\right]^{(\nu+1)/2}}{\left(1+\frac{x^2+y^2-2\rho xy}{(1-\rho^2)\nu}\right)^{(\nu+2)/2}},$$

where $x = T_\nu^{-1}(u)$ and $y = T_\nu^{-1}(v)$. For the positive tail dependence holds, that

$$\lambda_{LL} = \lambda_{UU} = 2\, T_{\nu+1}\left(-\sqrt{\nu+1}\frac{\sqrt{1-\rho}}{\sqrt{1+\rho}}\right).$$

Contrary we have

$$\lambda_{LU} = \lambda_{UL} = 2\, T_{\nu+1}\left(-\sqrt{\nu+1}\frac{\sqrt{1+\rho}}{\sqrt{1-\rho}}\right).$$

for the negative tail dependence (see Joe, 2011).

The Gaussian copula of Example 2.12 is the limiting case of the Student-t copula for $\nu \to \infty$. This yields zero tail dependence for the Gaussian copula. Figure 2.3.1 illustrates the contour plots of Student-t respectively Gaussian copulae for different combinations of the parameters ρ and ν, given normal margins.

Fang, Fang, and Kotz (2002) deliver a simple relationship between Kendall's τ and ρ which is valid for elliptical copulae.
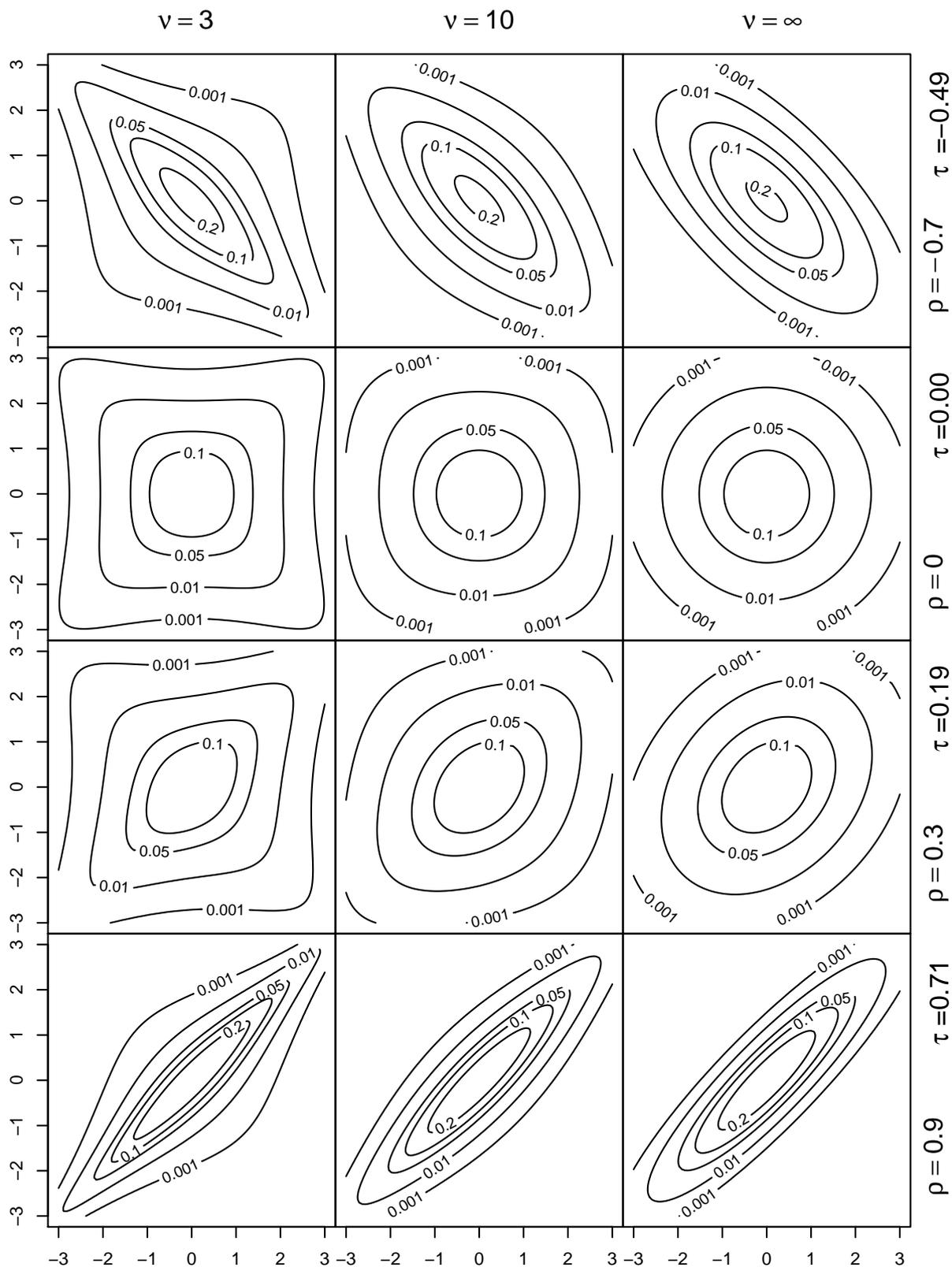
Figure 2.3.1: Contour plots with normal margins of Student-t copulae, for different combinations of the parameters ν and ρ.

Proposition 2.14 (Kendall's τ for Gaussian and Student-t copula)**.** *For $(X, Y) \sim C$, where C is either the Gaussian or the Student-t copula, we have*

$$\tau_{X,Y} = \frac{2}{\pi} \arcsin(\rho).$$

Finally we want to address the symmetry properties of the investigated copulae. For a copula C we talk of reflection symmetry if $c(u, v) = c(1 - u, 1 - v)$. This is true for both, the Gaussian copula of Example 2.12 and the Student-t copula of Example 2.13. Moreover both copula types are symmetric in their arguments. That means that $C(u, v) = C(v, u)$.

2.3.2 Archimedean copulae

The second class of copulae which we want to introduce are (bivariate) Archimedean copulae. Due to their properties and ability to model different kinds of (tail) dependence, they became quite popular in the area of pair-copula constructions. The following definitions and propositions are inspired by the compact introduction on Archimedean copulae in Kurowicka and Cooke (2006).

Definition 2.15 (Generator of an (bivariate) Archimedean copula and its inverse)**.** A strictly convex, strictly decreasing function $\varphi : [0, 1] \rightarrow [0, \infty]$ with existing second derivative $\varphi''(x) > 0$ for all $x \in (0, 1]$ and $\varphi(1) = 0$ is called *generator of an Archimedean copula*. Its *(quasi) inverse* is defined as

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

⌟

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

Definition 2.16 (Archimedean copula, bivariate case)**.** Let φ be a generator with inverse $\varphi^{[-1]}$ and $u, v \in [0, 1]$. Then C defined by

$$C(u, v) := \varphi^{[-1]}(\varphi(u) + \varphi(v))$$

is called *(bivariate) Archimedean copula*. ⌟

Proposition 2.17 (Density of an Archimedean copula, bivariate case)**.** *Let C be an Archimedean copula with generator φ. Then the copula density is given as*

$$c(u, v) = \begin{cases} -\frac{\varphi''(C(u,v))\varphi'(u)\varphi'(v)}{[\varphi'(C(u,v))]^3} & \text{for } \varphi(u) + \varphi(v) \leq \varphi(0) \\ 0 & \text{for } \varphi(u) + \varphi(v) > \varphi(0). \end{cases}$$

Eventually we state some properties that are valid for Archimedean copulae in general.

Proposition 2.18 (Properties of Archimedean copulae)**.** *Let C be an Archimedean copula with generator φ and $u, v \in [0, 1]$. Then the following properties hold:*

(i) C is a copula.

(ii) C is symmetric in its arguments, i.e. $C(u,v) = C(v,u)$.

(iii) If $\varphi(x) = -a\ln(x)$ with $a > 0$, then $C \equiv C^{\perp}$, i.e. C equals the independence copula.

One further advantage of Archimedean copulae is that there exists an easy formula for the calculation of Kendall's τ in dependence of the generator function and its first derivative (cp. Genest and MacKay, 1986).

Proposition 2.19 (Kendall's τ for Archimedean copulae)**.** *For $(X,Y) \sim C$, where C is an Archimedean copula, we have*

$$\tau_{X,Y} = 4 \int_0^1 \frac{\varphi(t)}{\varphi'(t)} \mathrm{d}t + 1.$$

In the following we are going to state some examples. Since most of our later calculations base on the `R` (R Development Core Team, 2012) packages `CDVine` (Brechmann and Schepsmeier, 2013) and `VineCopula` (Schepsmeier, Stöber, and Brechmann, 2013), we will stick to their naming conventions and parameter restrictions. Besides, formulas for generators, copulae, their respective densities, expressions for Kendall's τ's and for tail dependence as well as further information is taken from Joe (2001), Nelsen (2006) and Embrechts, Lindskog, and McNeil (2003), if not stated differently.

Example 2.20 (Clayton copula)**.** This bivariate copula family was first discussed by Clayton (1978). It depends on one parameter $\theta \in (0, \infty)$ and is generated by

$$\varphi(x) = \frac{1}{\theta}\left(\frac{1}{x^\theta} - 1\right).$$

We can write the copula as

$$C(u,v;\theta) = \left(\frac{1}{u^\theta} + \frac{1}{v^\theta} - 1\right)^{-1/\theta}$$

and get

$$c(u,v;\theta) = \frac{1+\theta}{(uv)^{1+\theta}}\left(\frac{1}{u^\theta} + \frac{1}{v^\theta} - 1\right)^{-(2+1/\theta)}$$

for its density. Kendall's τ can be expressed as

$$\tau_{X,Y} = \frac{\theta}{\theta + 2}.$$

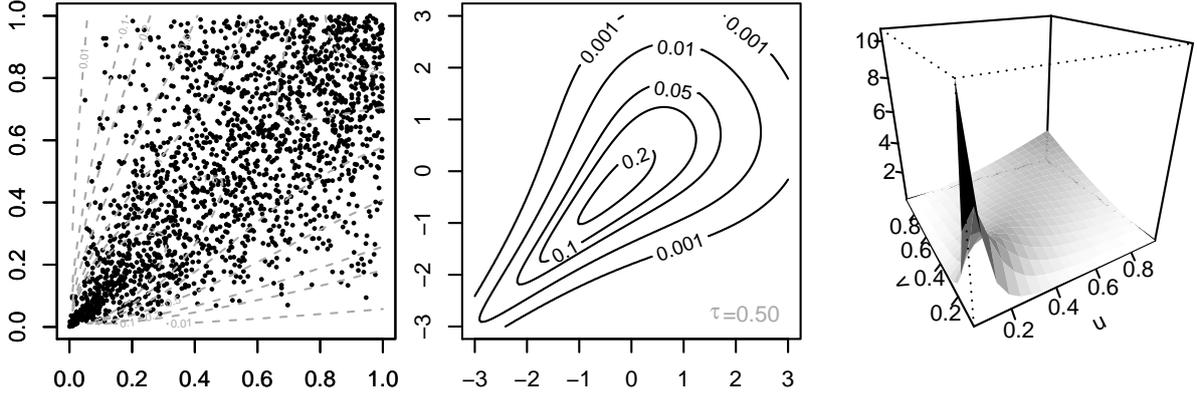For the (positive) tail dependence we obtain $\lambda_{LL} = 2^{-1/\theta}$ and $\lambda_{UU} = 0$. ⌟

Figure 2.3.2: Visualization of a Clayton copula with $\theta = 2$. Left: Random sample ($N = 1000$). Middle: Contour plot with normal margins. Right: Density on $[0.05, 0.95]^2$.

Example 2.21 (Gumbel copula)**.** Gumbel (1960) was the first to discuss this bivariate copula family. It is generated by the one parametric generator

$$\varphi(x) = (-\ln x)^{\theta},$$

where it holds $\theta \in [1, \infty)$ for the parameter. Following we can write

$$C(u, v; \theta) = \exp\left(-\left((-\ln u)^{\theta} + (-\ln v)^{\theta}\right)^{1/\theta}\right)$$

for the copula and

$$c(u, v; \theta) = \frac{C(u, v; \theta)(\ln u \ln v)^{\theta-1}}{(uv)\left((-\ln u)^{\theta} + (-\ln v)^{\theta}\right)^{2-1/\theta}}\left(\left((-\ln u)^{\theta} + (-\ln v)^{\theta}\right)^{1/\theta} + \theta - 1\right)$$

for the corresponding density. We obtain

$$\tau_{X,Y} = \frac{\theta - 1}{\theta}$$

for Kendall's τ. There is no lower (positive) tail dependence and it holds $\lambda_{UU} = 2 - 2^{1/\theta}$ for the upper (positive) tail dependence. ⌟

Example 2.22 (Frank copula)**.** The bivariate copula family, which is nowadays known as the Frank copula family, was first treated by Frank (1979) in a non-statistical context. Dependent on one parameter $\theta \in \mathbb{R} \backslash \{0\}$, the generator is given by

$$\varphi(x) = \ln \frac{1 - e^{-\theta}}{1 - e^{-\theta x}}.$$

This yields

$$C(u, v; \theta) = -\frac{1}{\theta} \ln\left(1 - \frac{(1 - e^{-\theta u})(1 - e^{-\theta v})}{1 - e^{-\theta}}\right)$$
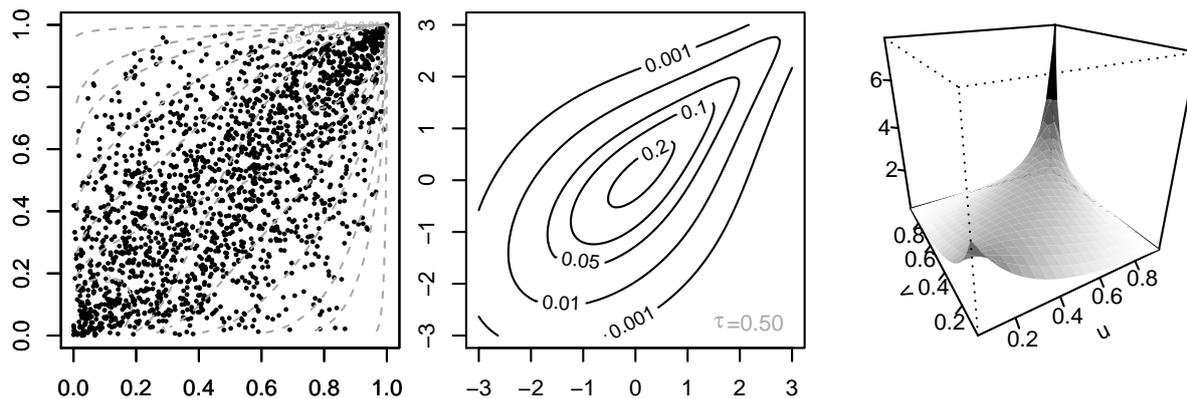
Figure 2.3.3: Visualization of a Gumbel copula with $\theta = 2$. Left: Random sample ($N = 1000$). Middle: Contour plot with normal margins. Right: Density on $[0.05, 0.95]^2$.

for the copula and we get

$$c\left(u, v; \theta\right) = \frac{\theta\left(1 - e^{-\theta}\right) e^{-\theta(u+v)}}{\left(1 - e^{-\theta} - (1 - e^{-\theta u})(1 - e^{-\theta v})\right)^2}$$

for its density. The respective Kendall's τ is

$$\tau_{X,Y} = 1 - \frac{4}{\theta}\left(1 - D_1(\theta)\right),$$

where

$$D_k(x) = \frac{k}{x^k} \int_0^x \frac{y^k}{e^y - 1} \mathrm{d}y, \quad k \in \mathbb{N},$$

is the Debye function. The Frank copula exhibits no (positive) tail dependence. Moreover it is reflection symmetric, i.e. it holds $c\left(1 - u, 1 - v; \theta\right) = c\left(u, v; \theta\right)$ for all $u, v \in [0, 1]$.   ⌟

Figure 2.3.4: Visualization of a Frank copula with $\theta = 8$. Left: Random sample ($N = 1000$). Middle: Contour plot with normal margins. Right: Density on $[0.05, 0.95]^2$.

Example 2.23 (Joe copula)**.** This last example of a bivariate copula family was first treated intensely by Joe (1993). It depends on one parameter $\theta > 1$ and is generated by

$$\varphi(x) = \ln\left(1 - (1-x)^\theta\right)^{-1}.$$

The copula is given by

$$C\left(u, v; \theta\right) = 1 - \left((1-u)^\theta + (1-v)^\theta - (1-u)^\theta(1-v)^\theta\right)^{1/\theta}$$

and its density is

$$c\left(u, v; \theta\right) = \frac{(1-u)^{\theta-1}(1-v)^{\theta-1}\left(\theta - 1 + (1-u)^\theta + (1-v)^\theta - (1-u)^\theta(1-v)^\theta\right)}{\left((1-u)^\theta + (1-v)^\theta - (1-u)^\theta(1-v)^\theta\right)^{2-1/\theta}}.$$

In terms of θ, Kendall's τ can be expressed as

$$\tau_{X,Y} = 1 + \frac{4}{\theta^2}\int_0^1 y\ln(y)(1-y)^{2(1-\theta)/\theta}\mathrm{d}y.$$

Once again there is no lower (positive) tail dependence, but it holds $\lambda_{UU} = 2 - 2^{1/\theta}$ for the upper (positive) tail dependence. Here the results for Kendall's τ and the tail dependence are taken from Schepsmeier (2010).
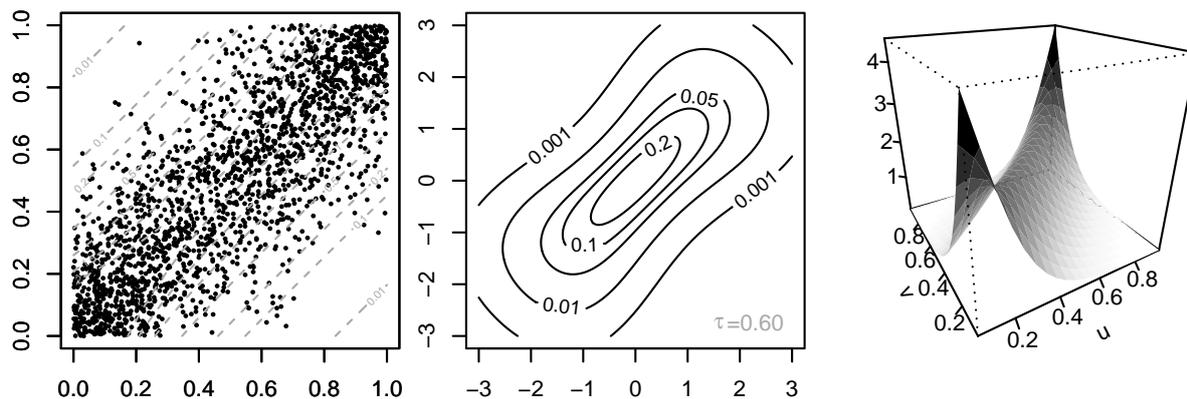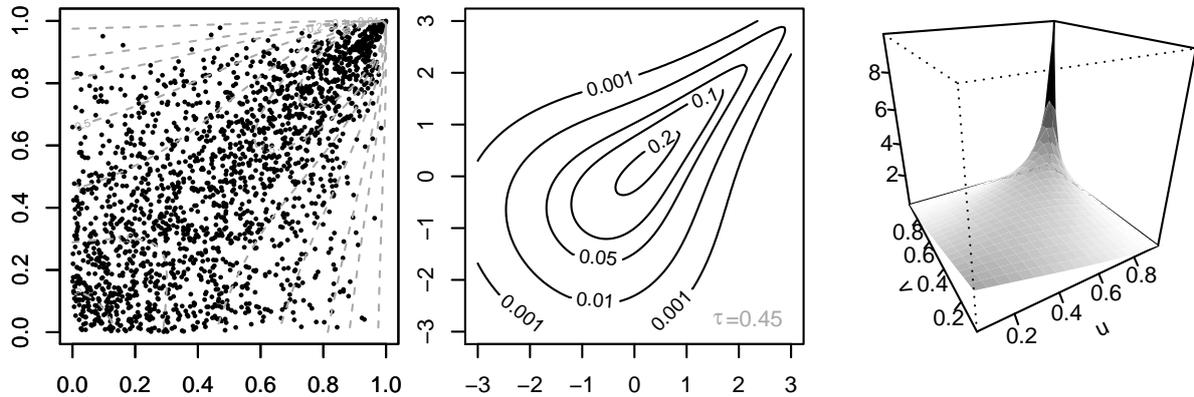

Figure 2.3.5: Visualization of a Joe copula with $\theta = 2.5$. Left: Random sample ($N = 1000$). Middle: Contour plot with normal margins. Right: Density on $[0.05, 0.95]^2$.

2.3.3 Survival copulae and copulae with negative dependence structures

In the previous subsections we have seen, that all of the considered copula families exhibit symmetry in their both arguments, but not all of them are reflection symmetric, e.g. the Clayton and the Gumbel copula. These copulae are moreover asymmetric in terms of tail dependence. Sometimes there is only upper tail dependence and sometimes only lower tail dependence. This leads to the idea to extend the previously considered copulae by

reflecting their arguments, i.e. we consider the reflection of the first, the second and both arguments in an arbitrary copula $C(u, v)$ at $1/2$.

Let's first consider reflection in the first argument, i.e. let $(1-U, V) \sim C$. The resulting copula is given as

$$C^1(u, v) := v - C(1 - u, v) \quad \text{with density} \quad c^1(u, v) := c(1 - u, v).$$

Now we investigate reflection in the second argument, i.e. let $(U, 1 - V) \sim C$. We obtain

$$C^2(u, v) := u - C(u, 1 - v) \quad \text{with density} \quad c^2(u, v) := c(u, 1 - v).$$

Finally we are also interested in the case where both arguments are reflected. We have to consider $(1 - U, 1 - V) \sim C$. This results in the copula

$$C^{12}(u, v) := u + v - 1 + C(1 - u, 1 - v) \quad \text{with density} \quad c^{12}(u, v) := c(1 - u, 1 - v),$$

which is called *survival copula*.

Our idea for these extensions stems from Joe (1993). Often these copulae are also called *rotated copulae*. In the order given above, they are referred to as 90°, 270° and 180°-rotated copulae, respectively. For this naming convention we want to give Czado and Stöber (2012) as a reference.

2.4   Least squares and linear regression

As least squares estimation and linear regression are essential tools in lots of our modeling processes, we give a short overview of the most important concepts and results on these topics. The following is based on Czado and Schmidt (2011).

In the context of regression, we consider a setting as follows. The goal is to model a response vector $\boldsymbol{Y} = (Y_1, \ldots, Y_n)^\top$ with realizations $\boldsymbol{y} = (y_1, \ldots, y_n)^\top$. The models are build on explanatory variables, i.e. we have k covariates $\boldsymbol{x}_j = (x_{1j}, \ldots, x_{nj})^\top$, $j = 1, \ldots, k$.

2.4.1   General regression models

Now let us first define a general regression model.

Definition 2.24 (General regression model)**.** Let $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_r)^\top \in \Theta$ be an (unknown) parameter vector out of the r-dimensional parameter space $\Theta \subseteq \mathbb{R}^r$. Moreover let $g_1, \ldots, g_n : \mathbb{R}^k \times \Theta \to \mathbb{R}$ be parametric functions. Then

$$Y_i = g_i(x_{i1}, \ldots, x_{ik}; \boldsymbol{\theta}) + \varepsilon_i, \qquad i = 1, \ldots, n, \tag{GRM}$$

is called *general regression model*, if $\varepsilon_1, \ldots, \varepsilon_n$ fulfill the white noise conditions (R).

(R) It holds for the random variables $\varepsilon_1, \ldots, \varepsilon_n$, that

(i) $\mathrm{E}(\varepsilon_i) = 0$, $i = 1, \ldots, n$,

(ii) $\mathrm{Var}(\varepsilon_i) = \sigma^2$, $i = 1, \ldots, n$, for $\sigma > 0$ unknown and

(iii) $\mathrm{Cov}(\varepsilon_i, \varepsilon_j) = 0$, $1 \leq i \neq j \leq n$. ⌟

Consequently the errors of these regression models exhibit a constant (homogeneous) variation around zero and they are not correlated among each other. However it may occur in practice, that condition (ii) of (R) is to strong. To weaken it, we consider (known) *weights $w_i > 0$, $i = 1, \ldots, n$,* and replace (ii) by

(ii') $\mathrm{Var}(\varepsilon_i) = \sigma^2 w_i$, $w_i > 0$, $i = 1, \ldots, n$, known, $\sigma > 0$ unknown,

which allows for *heteroscedasticity.* We denote the conditions (R), where (ii) is replaced by (ii'), as (R'). If we now set

$$Z_i := \frac{Y_i}{\sqrt{w_i}}, \qquad i = 1, \ldots, n,$$

this leads to a *weighted regression model*

$$Z_i = h_i(x_{i1}, \ldots, x_{ik}; \boldsymbol{\theta}) + \varepsilon_i^*, \qquad i = 1, \ldots, n, \tag{2.4.1}$$

with $h_1, \ldots, h_n : \mathbb{R}^k \times \Theta \to \mathbb{R}$ and $\varepsilon_i^* = \varepsilon_i w_i^{-1/2}$, $i = 1, \ldots, n$. For $\varepsilon_1, \ldots, \varepsilon_n$ which satisfy the conditions (R'), we can follow directly, that $\varepsilon_1^*, \ldots, \varepsilon_n^*$ satisfy (R), since

$$\mathrm{Var}(\varepsilon_i^*) = \frac{\mathrm{Var}(\varepsilon_i)}{w_i} = \sigma, \qquad i = 1, \ldots, n.$$

Following we can treat the model given in (2.4.1) in the same way as a general regression model (GRM), i.e. we found a way to solve a heteroscedastic regression by means of homoscedastic regression.

2.4.2 Least squares estimation

We proceed with a definition of the residual sum of squares in the framework of general regression models and eventually present the concept of least squares estimation.

Definition 2.25 (Residual sum of squares)**.** For a general regression model (GRM) the *residual sum of squares $Q : \Theta \times \mathbb{R}^n \to [0, \infty)$* is defined by

$$Q(\boldsymbol{\theta}, \boldsymbol{y}) := \sum_{i=1}^{n} \left(y_i - g_i(x_{i1}, \ldots, x_{ik}; \boldsymbol{\theta}) \right)^2. \tag{2.4.2}$$

⌟

Definition 2.26 (Least squares estimator)**.** If there exists a measurable function $\widehat{\boldsymbol{\theta}} : \mathbb{R}^n \to \Theta$, such that

$$Q\left(\widehat{\boldsymbol{\theta}}(\boldsymbol{y}), \boldsymbol{y}\right) \leq Q\left(\widetilde{\boldsymbol{\theta}}, \boldsymbol{y}\right) \qquad \text{for all } \widetilde{\boldsymbol{\theta}} \in \Theta, \boldsymbol{y} \in \mathbb{R}^n,$$

then we call $\widehat{\boldsymbol{\theta}}(\boldsymbol{Y})$ *least squares estimator (LSE)* of a general regression model (GRM). ⌟

If the parameter space Θ is open, the least squares estimator $\widehat{\boldsymbol{\theta}}$ has to fulfill the so called *normal equations*

$$\frac{\partial}{\partial \theta_j} Q(\boldsymbol{\theta}, \boldsymbol{y})\big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}(\boldsymbol{y})} = 0, \quad j = 1, \ldots, r, \tag{2.4.3}$$

$$\overset{(2.4.2)}{\Leftrightarrow} \quad \sum_{i=1}^{n} \left(\big(y_i - g_i(x_{i1}, \ldots, x_{ik}; \boldsymbol{\theta})\big) \cdot \frac{\partial}{\partial \theta_j} g_i(x_{i1}, \ldots, x_{ik}; \boldsymbol{\theta})\Big|_{\boldsymbol{\theta}=\widehat{\boldsymbol{\theta}}(\boldsymbol{y})} \right) = 0, \quad j = 1, \ldots, r.$$

Sometimes these equations may be solved explicitly, e.g. if the g_i, $i = 1, \ldots, n$, are linear functions, but often numerical methods are needed.

2.4.3   Linear models

Next, an important special case of regression models, the linear models, and some properties of these are summarized.

Definition 2.27 (General linear model)**.** A general regression model (GRM) is called *general linear model*, if the functions $g_1, \ldots, g_n : \mathbb{R}^k \times \Theta \to \mathbb{R}$, are given by

$$g_i(x_{i1}, \ldots, x_{ik}; \boldsymbol{\beta}) = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_k x_{ik}, \qquad i = 1, \ldots, n, \tag{LM}$$

where $\boldsymbol{\beta} = (\beta_0, \ldots, \beta_k)^\top \in \mathbb{R}^r$ with $r = k + 1$ and if it holds for the errors that $\varepsilon_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2)$, $i = 1, \ldots, n$, with $\sigma > 0$. ⌟

If one merges the covariates into a *design matrix*

$$\boldsymbol{X} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1k} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{nk} \end{pmatrix},$$

and consolidates the errors in a vector $\boldsymbol{\varepsilon} = (\varepsilon_1, \ldots, \varepsilon_n)^\top$, the general linear model (LM) can be rewritten in matrix-vector notation as

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \qquad \boldsymbol{\varepsilon} \sim \mathcal{N}_n\left(\boldsymbol{0}_n, \sigma^2 I_n\right), \tag{2.4.4}$$

where $\boldsymbol{0}_n$ is a vector consisting of n zeros and I_n an $n \times n$ identity matrix. Note that the least squares estimator of (2.4.4) can be calculated explicitly from the normal equations (2.4.3) as

$$\widehat{\boldsymbol{\beta}}(\boldsymbol{Y}) = (\boldsymbol{X}^\top \boldsymbol{X})^{-1} \boldsymbol{X}^\top \boldsymbol{Y},$$

if $\boldsymbol{X}$ has full rank. Moreover it holds that

$$\widehat{\boldsymbol{\beta}}(\boldsymbol{Y}) \sim \mathcal{N}\left(\boldsymbol{\beta}, \sigma^2 (\boldsymbol{X}^\top \boldsymbol{X})^{-1}\right).$$

Based on this property it is possible to perform hypothesis tests on the significance of the parameter estimates $\widehat{\boldsymbol{\beta}}(\boldsymbol{y})$.

Finally we will present the so called coefficient of determination and an adjusted version thereof.

Definition 2.28 (Coefficient of determination). For a general linear model (LM) we define the *total sum of squares* by

$$\mathrm{SS}_{\mathrm{tot}} := \left\| \boldsymbol{y} - \boldsymbol{1}_n \overline{y} \right\|^2,$$

where $\boldsymbol{1}_n$ is a vector of n ones and $\overline{y} = \frac{1}{n} \sum_{i=1}^n y_i$. The *residual sum of squares* is given as

$$\mathrm{SS}_{\mathrm{res}} := \left\| \boldsymbol{y} - \boldsymbol{X} \widehat{\boldsymbol{\beta}}(\boldsymbol{y}) \right\|^2.$$

Then the *coefficient of determination* and its *adjusted* version are defined as

$$\mathrm{R}^2 := 1 - \frac{\mathrm{SS}_{\mathrm{res}}}{\mathrm{SS}_{\mathrm{tot}}} \qquad \text{and} \qquad \mathrm{R}^2_{\mathrm{adj}} := 1 - \frac{(n-1)\mathrm{SS}_{\mathrm{res}}}{(n-r)\mathrm{SS}_{\mathrm{tot}}}.$$

⌟

The coefficient of determination measures the percentage of the total variability in the data explained by a linear model. Whereas R^2 increases when an additional covariate is added to the model, $\mathrm{R}^2_{\mathrm{adj}}$ is adjusted for this effect and designed for the comparison of models with different numbers of covariates. We are going to use the above defined measures for the purpose of model building, in order to compare models and model components by means of explanatory power.

2.5 The Ljung-Box test

The Ljung-Box test (cp. Ljung and Box, 1978) is a statistical test intended to check if the error terms specified by a model for a discrete time series $\{X_t, t = 1, \ldots, N\}$ exhibit no autocorrelation up to a certain lag m.

Usually the time series under consideration are modeled in terms of a so called *autoregressive-moving average (ARMA(p,q)) model*, i.e. as

$$X_t = c + \sum_{i=1}^p \phi_i X_{t-i} + \varepsilon_t + \sum_{i=1}^q \vartheta_i \varepsilon_{t-i},$$

where the components of the process $\{\varepsilon_t, t = 1, \ldots, N\}$ are i.i.d. with zero mean and finite variance σ^2. Then the null hypothesis tested by the *Ljung-Box test* is given as

$$\mathrm{H}_0 : \rho_1(\varepsilon_t) = \ldots = \rho_m(\varepsilon_t) = 0,$$

where $m > p + q$ and

$$\rho_k(\varepsilon_t) = \frac{\sum_{t=k+1}^N \varepsilon_{t-k} \varepsilon_t}{\sum_{t=1}^N \varepsilon_t^2}$$

is the autocorrelation with lag k of the process $\{\varepsilon_t, t \in \mathbb{N}\}$. The test statistic on which the test is based is given by

$$\widetilde{Q}(\rho) := N(N+2) \sum_{k=1}^m \frac{\rho_k^2(\varepsilon_t)}{N-k}.$$

For large N it is approximately χ^2_{m-p-q} distributed.

2.6   Graph theory

Furthermore, we want to take a brief look at graph theory, since the so called vine copula models, which build the base for our models in the subsequent chapters, are systematized via graphs. We refer to the comprehensive introduction of graph theory in Diestel (2010, Chapter 1) and pick out the concepts and definitions which are important for our purposes.

Definition 2.29 (Graph). Let $\mathcal{V}$ be an arbitrary set and $\mathcal{E} \subseteq \{\{v, w\} : v, w \in \mathcal{V}\}$. Then the pair $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ of the sets $\mathcal{V}$ and $\mathcal{E}$ is called *graph*. An element v of $\mathcal{V}$ is called *vertex* or *node* of the graph $\mathcal{G}$. Each element $e \in \mathcal{E}$ is named *edge*.   ⌟

Usually graphs are visualized in a way, where the vertices $v \in \mathcal{V}$ depicted as

are arranged (arbitrarily) over a plane. Moreover two vertices $v, w \in \mathcal{V}$ are connected by lines, if $\{v, w\} \in \mathcal{E}$.

Example 2.30. Figure 2.6.1 is an example of a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with vertex set $\mathcal{V} = \{1, \ldots, 8\}$ and edges in $\mathcal{E} = \{\{1, 2\}, \{2, 3\}, \{3, 4\}, \{3, 7\}, \{4, 5\}, \{5, 6\}, \{6, 7\}\}$.   ⌟
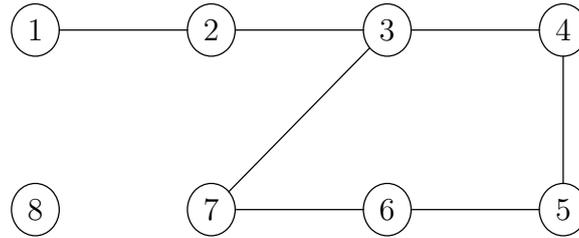

Figure 2.6.1: Example of a graph.

As we are interested in a special class of graphs we need some further definitions.

Definition 2.31 (Degree). Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a graph. Then the set $\mathcal{E}(v) \subseteq \mathcal{E}$ of all edges at a vertex $v \in \mathcal{V}$ is defined as

$$\mathcal{E}(v) := \{e \in \mathcal{E} : v \in e\}.$$

This allows to define the *degree* of a vertex $v \in \mathcal{V}$ as

$$\deg(v) = \#\mathcal{E}(v),$$

i.e. $\deg(v)$ equals the number of edges at the vertex v.   ⌟

Definition 2.32 (Path). Let $\mathcal{P} = (\mathcal{V}', \mathcal{E}')$ be a sub-graph of a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with vertex set of the form

$$\mathcal{V}' = \{v_1, \ldots, v_n\} \subseteq \mathcal{V}$$

and edge set

$$\mathcal{E}' = \{\{v_1, v_2\}, \ldots, \{v_{n-1}, v_n\}\} \subseteq \mathcal{E},$$

where $n \in \mathbb{N}$, and let $v_i \neq v_j$ for all $i = 1, \ldots, n$, $j = 1, \ldots, n$, $i \neq j$. Then we call $\mathcal{P}$ *path* in $\mathcal{G}$. In this case the vertices v_1 and v_n are called *ends*.   ⌟

Definition 2.33 (Connected). Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a non-empty graph. We call $\mathcal{G}$ *connected* if for every pair v, w of vertices in $\mathcal{V}$ there is a path $\mathcal{P} \subseteq \mathcal{G}$ with ends v and w. ⌟

Let us reconsider Example 2.30 and Figure 2.6.1. We easily see that $\deg(1) = 1$, $\deg(3) = 3$, $\deg(2) = \deg(4) = \deg(5) = \deg(6) = \deg(7) = 2$ and $\deg(8) = 0$. For example, the unique path between the vertices 1 and 3 is given as $\mathcal{P} = (\{1, 2, 3\}, \{\{1, 2\}, \{2, 3\}\})$. Notice that $\mathcal{P}' = (\mathcal{V}', \mathcal{E})$ with $\mathcal{V}' = \mathcal{V} \setminus \{8\}$ is no path, since the maximum allowed degree of a vertex of a path is two, but it holds $\deg(3) = 3$. Moreover there is no edge that links the vertex 8 to any other vertex. It follows, that $\mathcal{G}$ is not connected.

Definition 2.34 (Cycle). Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be a graph and $\mathcal{P} = (\mathcal{V}', \mathcal{E}') \subset \mathcal{G}$ be an arbitrary path in $\mathcal{G}$ with $\mathcal{V}' = \{v_1, \dots, v_n\}$ and $\mathcal{E}' = \{\{v_1, v_2\}, \dots, \{v_{n-1}, v_n\}\}$, where $n \in \mathbb{N}$. Let $e = \{v_1, v_n\} \in \mathcal{E}$. Then $\mathcal{C} := (\mathcal{V}', \mathcal{E}' \cup e) \subseteq \mathcal{G}$ is called *cycle* in $\mathcal{G}$. ⌟

Example 2.30 yields an example for a cycle. Here a cycle is given through the sub-graph $\mathcal{C} = (\mathcal{V}', \mathcal{E}')$ of $\mathcal{G}$ with $\mathcal{V}' = \{3, \dots, 7\}$ and $\mathcal{E}' = \{\{3, 4\}, \{4, 5\}, \{5, 6\}, \{6, 7\}, \{3, 7\}\}$.

Finally we are ready to define the special class of graphs, which we are interested in.

Definition 2.35 (Tree). A connected graph $\mathcal{T} = (\mathcal{V}, \mathcal{E})$ without cycles is called *tree*. ⌟

To learn about the properties of trees, we state the following

Proposition 2.36 (Properties of trees). *For a graph $\mathcal{T} = (\mathcal{V}, \mathcal{E})$ the following properties are equivalent:*

(i) $\mathcal{T}$ is a tree.

(ii) There is a unique path between two arbitrary nodes in $\mathcal{V}$.

(iii) $\mathcal{T}$ is minimally connected, i.e. $\mathcal{T}$ is connected but for $e \in \mathcal{E}$ and $\mathcal{E}' = \mathcal{E} \setminus \{e\}$, $\mathcal{T}' = (\mathcal{V}, \mathcal{E}')$ is not connected.

(iv) $\mathcal{T}$ is maximally acyclic, i.e. $\mathcal{T}$ has no cycle but for $v, w \in \mathcal{V}$ such that $e = \{v, w\} \notin \mathcal{E}$ and $\mathcal{E}' = \mathcal{E} \cup \{e\}$, $\mathcal{T}' = (\mathcal{V}, \mathcal{E}')$ has got a cycle.

With the help of Proposition 2.36 we verify, that Figure 2.6.2 depicts an example of a tree. Property (ii) can be checked easily and adding or removing any edge to respectively of the graph would lead to a violation of the properties (iii) respectively (iv) of a tree.
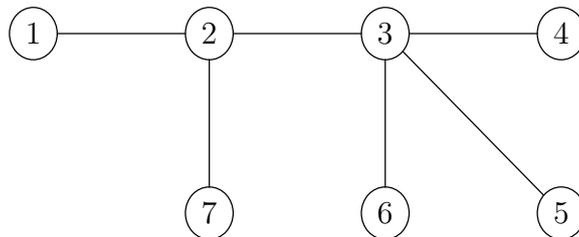

Figure 2.6.2: Example of a tree.

2.7   The Fisher z-transform

We are also going to apply a tool called *Fisher z-transform*. It is a transformation between the interval $(-1, 1)$ and the real numbers $\mathbb{R}$ and is often applied to transform correlation coefficients. The Fisher z-transform $F_z(\cdot)$, first introduced by Fisher (1915), is given by

$$F_z(r) = \frac{1}{2} \ln\left(\frac{1+r}{1-r}\right), \ r \in (-1, 1), \tag{2.7.1}$$

while its inverse

$$F_z^{-1}(z) = \frac{e^{2z} - 1}{e^{2z} + 1}, \ z \in \mathbb{R}, \tag{2.7.2}$$

transforms back to the interval $(-1, 1)$.

Whereas the two functions (2.7.1) and (2.7.2) are illustrated in Figure 2.7.1, some values of the Fisher z-transform and its inverse are given in Table 2.1 and Table 2.2, respectively. Both, the figures and the tables, show that it holds (approximately)

$$F_z(r) \approx r,$$

for r close to 0, and that correlations r with a high absolute value, let's say $|r| > 0.9$ are widely spread over $(-\infty, F_z(-0.9)) \cup (F_z(0.9), \infty)$.

r:	0	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	1.0
$F_z(r)$:	0	0.100	0.203	0.310	0.424	0.549	0.693	0.867	1.099	1.472	∞

Table 2.1: The Fisher z-transform for a range of r's.

z:	0	0.25	0.50	0.75	1.00	1.25	1.50	2.00	2.50	3.00
$F_z^{-1}(z)$:	0	0.245	0.462	0.635	0.762	0.848	0.905	0.964	0.987	0.995

Table 2.2: The inverse of the Fisher z-transform for a range of z's.
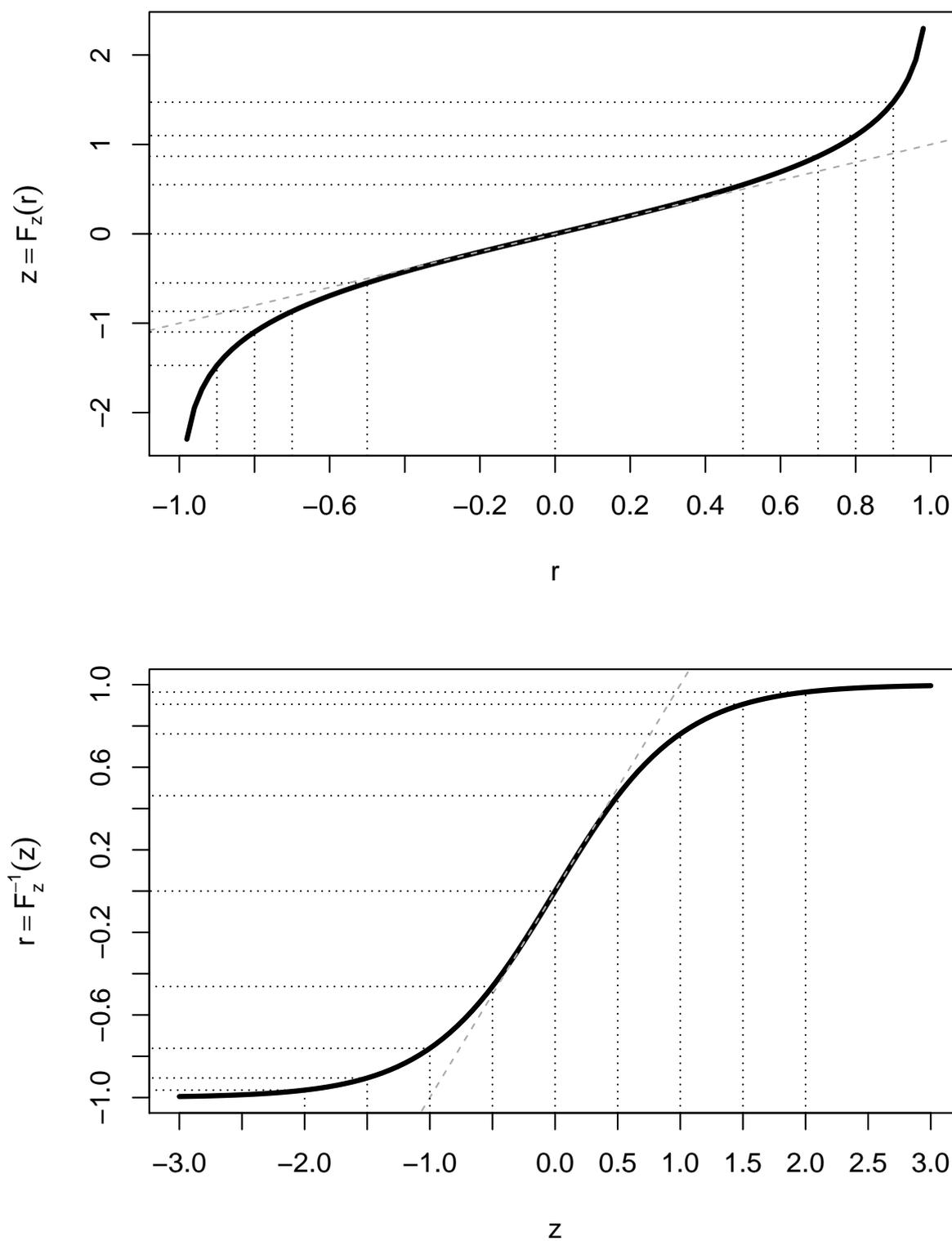
Figure 2.7.1: The Fisher z-transform and its inverse.

2.8   Scoring

Since we are going to predict based on the models developed in this thesis, we are also interested in evaluating these predictions. One possibility to do such evaluations are scoring rules. Following the extensive summary of different scoring rules in Gneiting and Raftery (2007) we briefly introduce to different kinds of scoring rules presented there.

In general scoring rules can be considered as quality measures of probabilistic predictions. Thus they can be used to rank different prediction procedures.

The first scoring rule is a popular scoring rule for continuous variables. It is defined in terms of the predictive cumulative distribution function F under consideration. For an arbitrary element y of the sample space $\mathbb{R}$ the so called *continuous ranked probability score* is defined as

$$\text{CRPS}(F, y) := - \int_{-\infty}^{\infty} \big(F(x) - \mathbb{1}_{\{x \geq y\}}\big)^2 \mathrm{d}x. \tag{2.8.1}$$

Here $\mathbb{1}$ is the indicator function. Since the continuous ranked probability score takes only negative values, bigger scores, i.e. scores close to zero, are preferred.

A case in which the continuous ranked probability score can be stated explicitly is the case when the predictive distribution is a Gaussian distribution. We obtain with ϕ the probability density function and Φ the cumulative distribution function corresponding to the standard normal distribution, that

$$\text{CRPS}(\mathcal{N}\big(\mu, \sigma^2\big), y) = \sigma \left[\frac{1}{\sqrt{\pi}} - 2\phi\left(\frac{y - \mu}{\sigma}\right) - \frac{y - \mu}{\sigma}\left(2\Phi\left(\frac{y - \mu}{\sigma}\right) - 1\right)\right].$$

Hersbach (2000) explains the calculation of (2.8.1) when only a sample from the predictive distribution F is available.

Sometimes it may be that one cannot access the full predictive distribution. However if one is able to calculate $(1 - \alpha) \cdot 100\%$ central prediction intervals, the interval score might be considered. Thus let l denote the $\alpha/2$ quantile of F and u the respective $1 - \alpha/2$ quantile. Then for an arbitrary realization $y \in \mathbb{R}$ the *interval score* is defined as

$$\text{IS}_\alpha(l, u; y) := (l - u) - \frac{2}{\alpha}(l - y)\mathbb{1}_{\{y < l\}} - \frac{2}{\alpha}(y - u)\mathbb{1}_{\{y > u\}}.$$

Obviously the interval score favors prediction procedures whose prediction intervals are small and are likely to cover the y which materializes.

We conclude the preliminaries with this introduction to scoring rules and advance to a preparatory analysis and modeling of the data which we are going to investigate throughout the thesis.

Chapter 3

Modeling the margins

Now we are ready to initiate the model building process for an exemplary spatial data set. The goal of this chapter is to find a joint model for the marginals given in this data set. Whereas the multivariate data set is presented in the first section of this chapter, the subsequent sections discuss the components of models intended to model each margin separately. Based on the parameter estimates obtained from these separate models, a joint model for all margins is developed in Section 3.6. In the last section the transformation of the data to so called copula data is presented.

3.1   Mean temperature data

The data which we are going to investigate in the following, is daily mean temperature data in °C collected over the period 01/01/2010-12/31/2012 by the German Meteorological Service (Deutscher Wetterdienst) at 54 selected observation stations across Germany. As an example one of these 54 time series, i.e. the 1096 data points of the observation station "München" (Munich), is depicted in Figure 3.1.1.
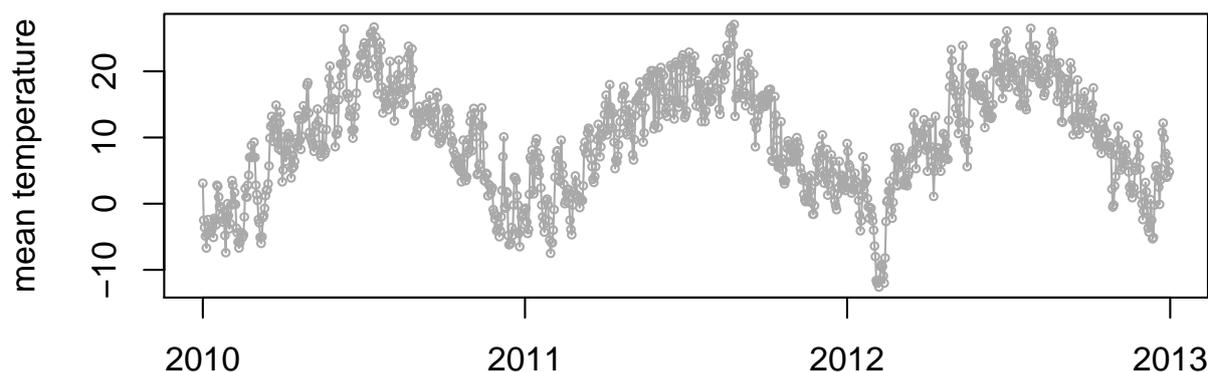

Figure 3.1.1: Mean temperature in °C in München over the years 2010 to 2012.

A list of all 54 observation stations is given by Table 3.1. Besides the full station name a short name consisting of four letters and an ID number are assigned to each station. Moreover longitude ($x_{\mathrm{long},s}$), latitude ($x_{\mathrm{lat},s}$) and elevation ($x_{\mathrm{elev},s}$) are given for each observation station.

31

s/ ID	short name	full name	$x_{\mathrm{long},s}$ (longitude)	$x_{\mathrm{lat},s}$ (latitude)	$x_{\mathrm{elev},s}$ (elevation)
1	ange	Angermünde	13.99	53.03	54.00
2	auee	Aue	12.72	50.59	387.00
3	brle	Berleburg, Bad-Stünzel	8.37	50.98	610.00
4	brln	Berlin-Buch	13.50	52.63	60.00
5	bonn	Bonn-Roleber	7.19	50.74	159.00
6	brau	Braunschweig	10.45	52.29	81.20
7	brem	Bremen	8.80	53.05	4.00
8	cott	Cottbus	14.32	51.78	69.00
9	cuxh	Cuxhaven	8.71	53.87	5.00
10	dres	Dresden-Hosterwitz	13.85	51.02	114.00
11	emde	Emden	7.23	53.39	0.00
12	esse	Essen-Bredeney	6.97	51.41	150.00
13	fran	Frankfurt/Main	8.60	50.05	112.00
14	frei	Freiburg	7.84	48.02	236.30
15	fuer	Fürstenzell	13.35	48.55	476.40
16	gard	Gardelegen	11.40	52.51	47.00
17	gies	Gießen/Wettenberg	8.64	50.60	202.70
18	goer	Görlitz	14.95	51.16	238.00
19	goet	Göttingen	9.95	51.50	167.00
20	hall	Halle-Kröllwitz	11.95	51.51	93.00
21	hamb	Hamburg-Fuhlsbüttel	9.99	53.63	11.00
22	hann	Hannover	9.68	52.47	55.00
23	holz	Holzdorf (Flugplatz)	13.17	51.77	81.00
24	jena	Jena (Sternwarte)	11.58	50.93	155.00
25	kais	Kaiserslautern	7.74	49.42	285.00
26	kass	Kassel	9.44	51.30	231.00
27	kauf	Kaufbeuren	10.60	47.87	716.00

s/ ID	short name	full name	$x_{\text{long},s}$ (longitude)	$x_{\text{lat},s}$ (latitude)	$x_{\text{elev},s}$ (elevation)
28	kons	Konstanz	9.19	47.68	442.50
29	kron	Kronach	11.32	50.25	312.00
30	ling	Lingen	7.31	52.52	22.00
31	lipp	Lippstadt-Bökenförde	8.39	51.63	92.00
32	lueb	Lübeck-Blankensee	10.70	53.80	15.50
33	magd	Magdeburg	11.58	52.10	76.00
34	mann	Mannheim	8.55	49.51	96.10
35	mein	Meiningen	10.38	50.56	450.00
36	mont	Montabaur	7.81	50.44	265.00
37	mnch	München-Stadt	11.54	48.16	515.20
38	mnst	Münster/Osnabrück	7.70	52.14	47.80
39	neub	Neuburg/Donau (Flugplatz)	11.21	48.71	380.00
40	neur	Neuruppin	12.81	52.91	38.00
41	nuer	Nürnberg	11.06	49.50	314.00
42	pidi	Piding	12.91	47.77	458.00
43	rege	Regensburg	12.10	49.04	365.40
44	rost	Rostock-Warnemünde	12.08	54.18	4.00
45	saar	Saarbrücken-Ensheim	7.11	49.21	320.00
46	schl	Schleswig	9.55	54.53	43.00
47	stut	Stuttgart (Neckartal)	9.22	48.79	224.00
48	trie	Trier-Zewen	6.61	49.73	131.50
49	ueck	Ueckermünde	14.07	53.74	1.20
50	uelz	Uelzen	10.53	52.94	50.00
51	ulmm	Ulm	9.95	48.38	566.80
52	ware	Waren	12.67	53.52	70.00
53	weid	Weiden	12.19	49.67	439.60
54	wuer	Würzburg	9.96	49.77	268.00

Table 3.1: Observation stations: ID, short name, full name, longitude, latitude and elevation in meters.

The location of the observation stations in Germany is illustrated in Figure 3.1.2, where the particular stations are marked with their ID number and their respective short name. The elevation is depicted based on an 80×120 elevation grid, which is obtained from the NGA Raster Roam of the National Geospatial-Intelligence Agency. Dark colored cells indicate low elevation, whereas light cells represent high elevations.

The now following modeling process was inspired by Pachali (2012), who developed different marginal models for daily mean temperature data.

Figure 3.1.2: The 54 observation stations across Germany.

3.2 A model for annual seasonality

As we all know temperature is subject to daily and yearly fluctuations. Since we are interested in daily means of temperature we do not have to take the daily fluctuations into consideration. Contrariwise we need to capture the yearly fluctuations in our model, which are caused amongst others by the inclination of the Earth's axis. These annual fluctuations in temperature can be clearly observed for our data. Figure 3.2.1 illustrates, that a sine curve of the form

$$\lambda \sin(\omega t + \delta) \tag{3.2.1}$$

parametrized by

$$\lambda : \text{amplitude,}$$
$$\omega : \text{angular frequency,}$$
$$\delta : \text{phase shift,}$$

is able to capture the seasonal trend of our data.
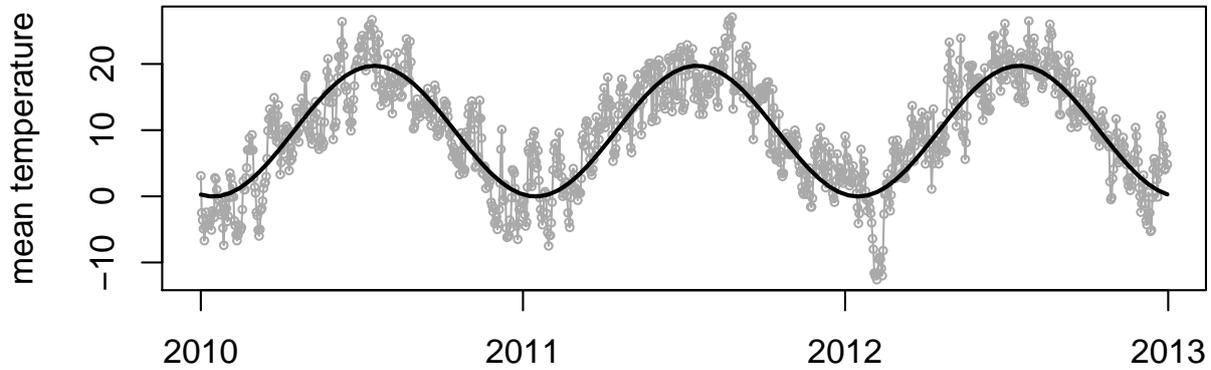
Figure 3.2.1: Mean temperature in °C in München over the years 2010 to 2012 (gray) and sine curve fitted to the data (black).

Inspired by Simmons (1990) we substitute the parameters in (3.2.1) according to

$$
\begin{aligned}
\cos(\delta) &= \frac{\beta_{\sin}}{\lambda}, \\
\sin(\delta) &= \frac{\beta_{\cos}}{\lambda}, \\
\lambda &= \sqrt{\beta_{\sin}^2 + \beta_{\cos}^2},
\end{aligned}
\tag{3.2.2}
$$

and use the trigonometric identity

$$\sin(x + y) = \sin x \cos y + \cos x \sin y$$

to transform (3.2.1) to

$$\beta_{\sin} \sin(\omega t) + \beta_{\cos} \cos(\omega t). \tag{3.2.3}$$

Due to the fact that we are going to use (3.2.3) in an annual context, we set $\omega = 2\pi/365.25$. Hereby we select 365.25 for the denominator to account for leap years. Thus (3.2.3) can be used as a linear component of a regression model.

3.3   Autoregression

Furthermore, the errors of our regression models shouldn't exhibit any dependence. But if we have a look at the deseasonalized time series, see for instance Figure 3.3.1, we clearly observe that the temperature on a particular day depends on the previous days.
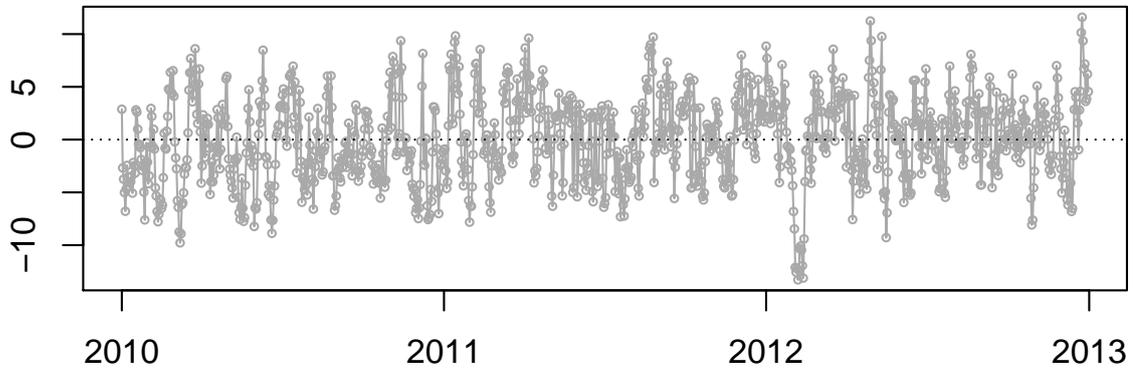
Figure 3.3.1: Deseasonalized mean temperature time series (München, 2010-2012).

To model this kind of dependence on previous points in time, we apply the concept of autoregression. This simply means that we add some lagged responses as covariates to our current models, i.e. we consider regression models of the form

$$Y_t = \beta_0 + \beta_{\sin} \sin\left(\frac{2\pi t}{365.25}\right) + \beta_{\cos} \cos\left(\frac{2\pi t}{365.25}\right) + \sum_{j=1}^{q} \gamma_j Y_{t-j} + \varepsilon_t, \quad t = 1, \ldots, N, \quad (3.3.1)$$

where the errors $\varepsilon_1, \ldots, \varepsilon_N$ are independent and identically distributed random variables with $\mathrm{E}(\varepsilon_t) = 0$ and $\mathrm{Var}(\varepsilon_t) = \sigma^2 > 0$, $t = 1, \ldots, N$.
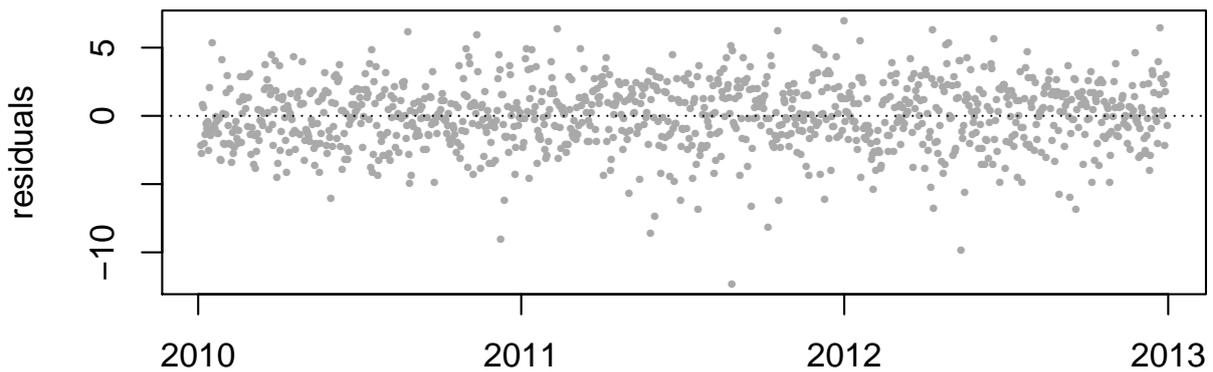
Figure 3.3.2: Residuals of the mean temperature time series for München according to model (3.3.1) with $q = 3$.

It remains to choose q in (3.3.1), such that the errors are no longer dependent. Some further investigations in Section 3.5 will show that $q = 3$ is an adequate choice for our data.

As an example we consider again the time series for München. Whereas the corresponding residuals of Model (3.3.1) with $q = 3$ are depicted in Figure 3.3.2, Figure 3.3.3 shows the autocorrelation function of these residuals. From the figures we see, that there is no significant autocorrelation left. Moreover, we cannot detect further functional trends in the residuals.
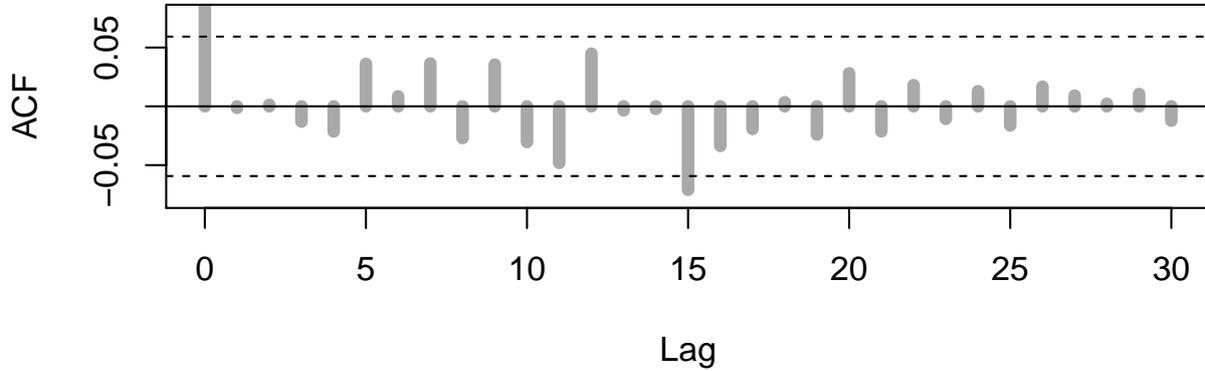
Figure 3.3.3: Plot of the autocorrelation function of the residuals depicted in Figure 3.3.2.

For a theoretical background on time series regression models, which cover linear regression models that include autoregression, we advert to Durbin (1960).

3.4 Skew-t distribution

Until now we have not specified a distribution for the errors $\varepsilon_1, \ldots, \varepsilon_N$ of our model given in (3.3.1). To find a suitable distribution family we fit different distributions (normal, skew normal and skew-t distribution) to the corresponding residuals and compare the fit of the respective densities to the histograms of the residuals. Moreover we investigate the corresponding Q-Q-plots.

The mentioned plots for München are depicted in Figure 3.4.1. From both, the histograms and the Q-Q-plots, we clearly see that the skew-t distribution yields the best fit in the tails.

Furthermore we observe that a skew-t distribution is most suitable in general, since it is able to capture skewness on the one hand and heavy tails on the other hand. For the construction of the skew-t distribution we follow Azzalini and Capitanio (2003), defining the density of a skew-t distributed random variable $X \sim$ skew-$t\,(\xi, \omega, \alpha, \nu)$ as

$$f(x) = \frac{2}{\omega} t_\nu(\widetilde{x}) T_{\nu+1}\left(\alpha\widetilde{x}\left(\frac{\nu+1}{\nu+\widetilde{x}^2}\right)^{1/2}\right),$$

where $\widetilde{x} = {(x-\xi)}/{\omega}$, t_ν is the density and $T_{\nu+1}$ the cumulative distribution function of a usual, univariate Student-t distribution with ν and $\nu+1$ degrees of freedom, respectively. Whereas the parameters ξ, ω and α can be interpreted as location, scale and shape parameter, respectively, ν denotes the degree of freedom parameter of the skew-t distribution.
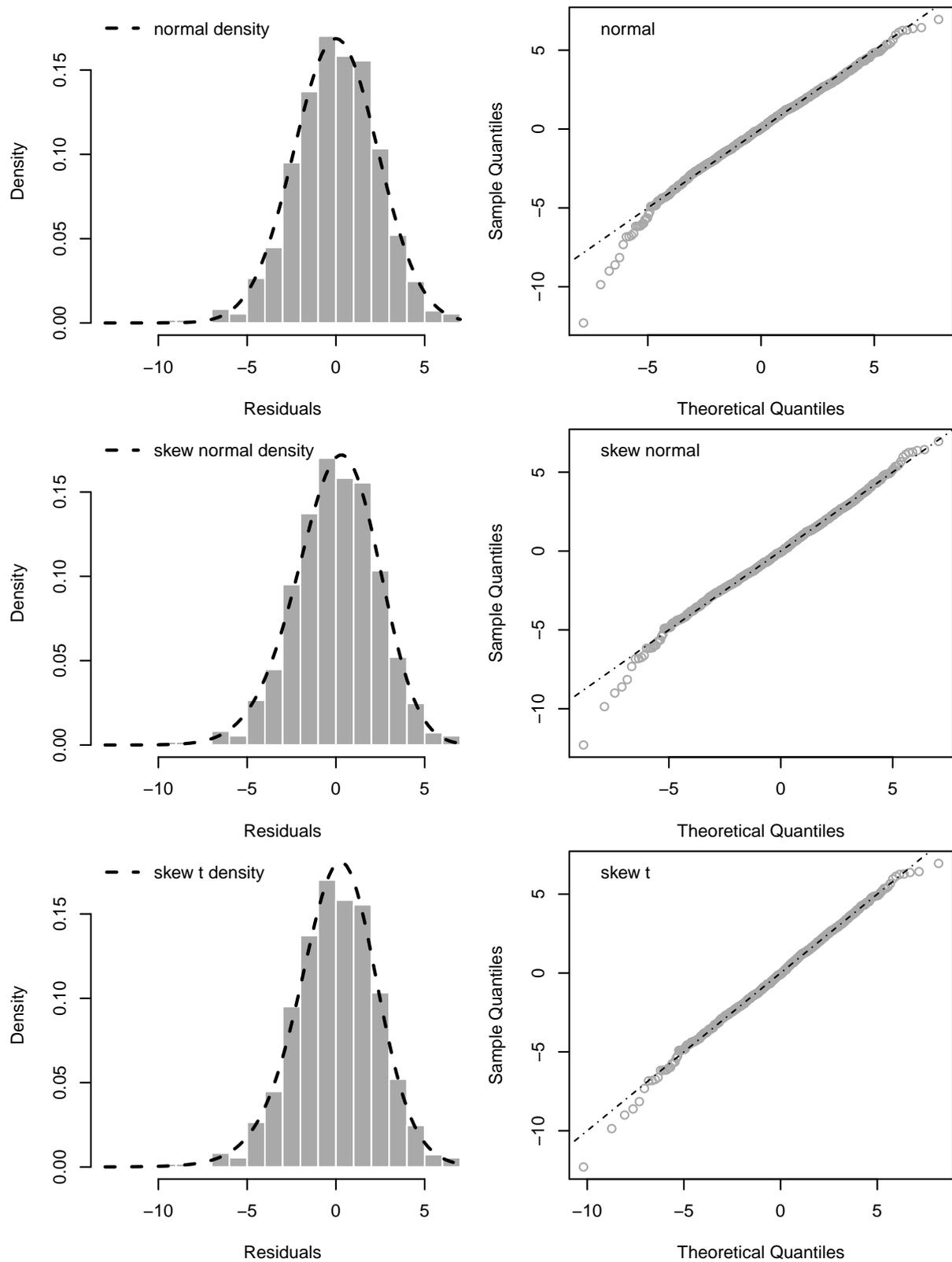
Figure 3.4.1: Histograms of the residuals of the mean temperature time series for München overlaid with densities and respective Q-Q-plots for normal, skew normal and skew-t distribution.

Defining

$$\delta := \frac{\alpha}{\sqrt{(1 + \alpha^2)}} \qquad \text{and} \qquad \mu := \delta \sqrt{\frac{\nu}{\pi}} \frac{\Gamma\left(\frac{\nu-1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)}, \qquad (3.4.1)$$

the expectation and the variance of a random variable $X \sim \text{skew-}t\,(\xi, \omega, \alpha, \nu)$ are given by

$$\text{E}(X) = \xi + \omega\mu \qquad \text{and} \qquad \text{Var}(X) = \omega^2 \left(\frac{\nu}{\nu - 2} - \mu^2\right), \qquad (3.4.2)$$

and exist for $\nu > 1$ and $\nu > 2$, respectively. The skewness is of the form

$$\text{skw}(X) = \mu \left(\frac{\nu(3 - \delta^2)}{\nu - 3} - \frac{3\nu}{\nu - 2} + 2\mu^2\right) \left(\frac{\nu}{\nu - 2} - \mu^2\right)^{-3/2} \qquad (3.4.3)$$

and exists for $\nu > 3$, whereas the kurtosis exists for $\nu > 4$ and is given as

$$\text{krt}(X) = \left(\frac{3\nu^2}{(\nu - 2)(\nu - 4)} - \frac{4\mu^2\nu(3 - \delta^2)}{\nu - 3} + \frac{6\mu^2\nu}{\nu - 2} - 3\mu^4\right) \left(\frac{\nu}{\nu - 2} - \mu^2\right)^{-2} - 3. \quad (3.4.4)$$

Eventually, as the most important theoretical issues on the skew-t distribution are clarified, we agree on

$$\varepsilon_1, \ldots, \varepsilon_N \overset{\text{i.i.d.}}{\sim} \text{skew-}t\,(\xi, \omega, \alpha, \nu) \qquad (3.4.5)$$

as distribution model for the errors. Note that the location parameter has to fulfill the condition $\xi = -\omega\mu$, in order that the zero mean condition for the errors holds.

3.5 Separate marginal models for daily mean temperatures at different locations

Now that we have discussed the single components of our model for daily mean temperature time series, we summarize our considerations by presenting the 54 final separate marginal models and how their parameters are estimated.

Let's recall, that our data consists of d times N ($d = 54$, $N = 1096$) observations y_t^s ($t = 1, \ldots, N$, $s = 1, \ldots, d$) of daily mean temperatures. We consider the observations $y_1^s, \ldots, y_N^s$ collected at the observation stations $s = 1, \ldots, d$, as realizations of time series $Y_1^s, \ldots, Y_N^s$, $s = 1, \ldots, d$.

Before we can apply Model (3.3.1) to each of the time series $Y_1^s, \ldots, Y_N^s$, $s = 1, \ldots, d$, we have to ensure that the homoscedasticity assumption of the errors $\varepsilon_1^s, \ldots, \varepsilon_N^s$, $s = 1, \ldots, d$, is fulfilled. In fact the errors are heteroscedastic, i.e. there exist weights $w_t^s > 0$, $t = 1, \ldots, N$, $s = 1, \ldots, d$, such that

$$\text{Var}(\varepsilon_t^s) = \sigma^2 w_t^s, \quad t = 1, \ldots, N, s = 1, \ldots, d,$$

for $\sigma > 0$. In order to achieve (approximate) homoscedasticity we have to estimate appropriate weights $\widehat{w}_t^s$, $t = 1, \ldots, N$, $s = 1, \ldots, d$, and consider the scaled time series

$$\frac{Y_1^s}{w_1^s}, \ldots, \frac{Y_N^s}{w_N^s}, \quad s = 1, \ldots, d,$$

instead of the original ones. In a first step we estimate common weights $\widetilde{w}_t$, $t = 1, \dots, N$, for all stations $s = 1, \dots, d$, by calculating the sample variances

$$\widetilde{w}_t := \frac{1}{d-1} \sum_{s=1}^{d} (y_t^s - \overline{y}_t)^2, \quad t = 1, \dots, N,$$

where

$$\overline{y}_t := \frac{1}{d} \sum_{s=1}^{d} y_t^s, \quad t = 1, \dots, N,$$

are the sample means. Since these weights, which we from now on will call *raw weights*, exhibit big jumps from one point in time to the next, we smooth our first weight estimates in a second step. For this purpose we apply least squares estimation, to fit a polynomial

$$q(t; \boldsymbol{\alpha}) := \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \dots + \alpha_9 t^9, \qquad \boldsymbol{\alpha} = (\alpha_0, \dots, \alpha_9)^\top,$$

of degree nine to the logarithmized raw weights $\ln(\widetilde{w}_t)$, $t = 1, \dots, N$. With

$$\widehat{\boldsymbol{\alpha}} = \operatorname*{argmin}_{\boldsymbol{\alpha}} \sum_{t=1}^{N} \big(q(t; \boldsymbol{\alpha}) - \ln(\widetilde{w}_t)\big)^2$$

we obtain the *smoothed weights*

$$\widehat{w}_t = \exp\left\{q(t; \widehat{\boldsymbol{\alpha}})\right\}, \quad t = 1, \dots, N.$$

Figure 3.5.1 depicts the raw as well as the smoothed weights. We observe higher variance during winter and relatively lower variances in the summers.
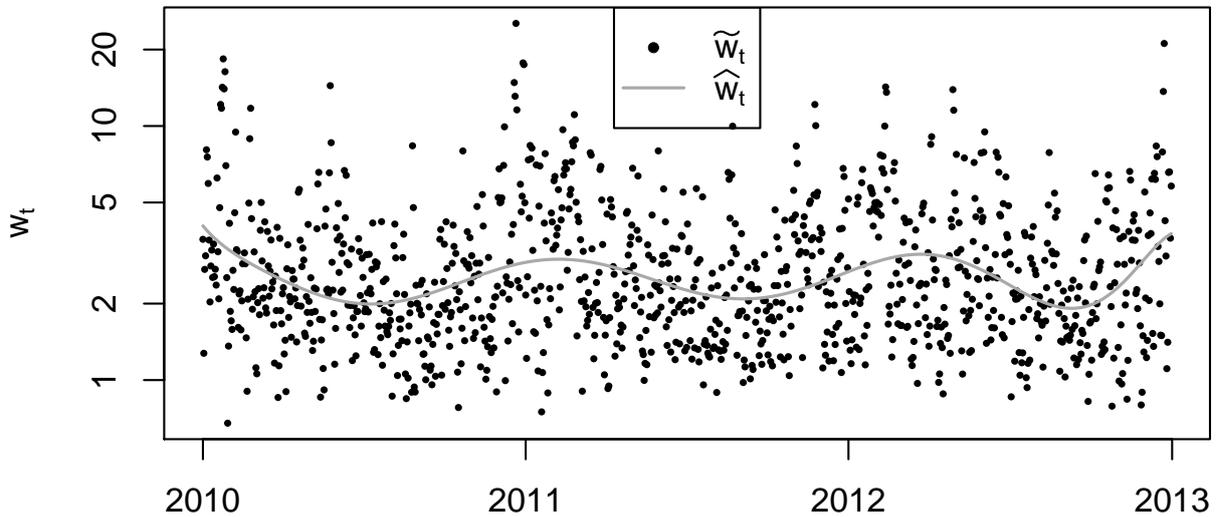
Figure 3.5.1: Plot of the estimated raw weights $\widetilde{w}_t$, $t = 1, \dots, N$, and the smoothed weights $\widehat{w}_t$, $t = 1, \dots, N$.

Now we are ready to apply (3.3.1) to the weighted responses

$$\widetilde{Y}_t^s = \frac{Y_t^s}{\sqrt{\widehat{w}_t}}, \quad t = 1, \dots, N, s = 1, \dots, d.$$

Therefore we define the 54 parameter vectors

$$\boldsymbol{\beta}^s_{\mathrm{ls}} := (\beta^s_0, \beta^s_{\sin}, \beta^s_{\cos}, \gamma^s_1, \gamma^s_2, \gamma^s_3)^\top$$

and the functions

$$g^s(t) := g^s\left(t, \widetilde{Y}^s_{t-1}, \widetilde{Y}^s_{t-2}, \widetilde{Y}^s_{t-3}; \boldsymbol{\beta}^s_{\mathrm{ls}}\right)$$

$$:= \beta^s_0 + \beta^s_{\sin} \sin\left(\frac{2\pi t}{365.25}\right) + \beta^s_{\cos} \cos\left(\frac{2\pi t}{365.25}\right) + \gamma^s_1 \widetilde{Y}^s_{t-1} + \gamma^s_2 \widetilde{Y}^s_{t-2} + \gamma^s_3 \widetilde{Y}^s_{t-3}$$

for all $s = 1, \ldots, d$. Together with (3.4.5) this yields the models

$$\widetilde{Y}^s_t = g^s\left(t, \widetilde{Y}^s_{t-1}, \widetilde{Y}^s_{t-2}, \widetilde{Y}^s_{t-3}; \boldsymbol{\beta}^s_{\mathrm{ls}}\right) + \varepsilon^s_t, \qquad \varepsilon^s_t \sim \text{skew-}t\left(\xi^s, \omega^s, \alpha^s, \nu^s\right), \qquad (3.5.1)$$

where $t = 1, \ldots, N$, $s = 1, \ldots, d$, and

$$\boldsymbol{\beta}^s_{\text{skew-}t} := (\xi^s, \omega^s, \alpha^s, \nu^s)^\top$$

are the parameters of the skew-t distribution corresponding to the stations $s = 1, \ldots, d$.

The parameters $\boldsymbol{\beta}^s_{\mathrm{ls}}$ of the models (3.5.1) are estimated via least-squares estimation. For this purpose we consider the weighted observations

$$\widetilde{y}^s_t = \frac{y^s_t}{\sqrt{\widehat{w}_t}}, \quad t = 1, \ldots, N, s = 1, \ldots, d.$$

Then we obtain the estimates of the parameter vectors $\boldsymbol{\beta}^s_{\mathrm{ls}}$, $s = 1, \ldots, d$, as

$$\widehat{\boldsymbol{\beta}}^s_{\mathrm{ls}} = \operatorname*{argmin}_{\boldsymbol{\beta}^s_{\mathrm{ls}}} \sum_{t=4}^{N} \left(\widetilde{y}^s_t - g^s\left(t, \widetilde{y}^s_{t-1}, \widetilde{y}^s_{t-2}, \widetilde{y}^s_{t-3}; \boldsymbol{\beta}^s_{\mathrm{ls}}\right)\right)^2, \quad s = 1, \ldots, d,$$

which yields the residuals

$$\widehat{\varepsilon}^s_t = \widetilde{y}^s_t - g^s\left(t, \widetilde{y}^s_{t-1}, \widetilde{y}^s_{t-2}, \widetilde{y}^s_{t-3}; \widehat{\boldsymbol{\beta}}^s_{\mathrm{ls}}\right), \quad t = 4, \ldots, N, s = 1, \ldots, d.$$

To conclude the parameter estimation we fit a skew-t distribution parametrized by $\boldsymbol{\beta}^s_{\text{skew-}t}$ to each of the residual series $\widehat{\varepsilon}^s_4, \ldots, \widehat{\varepsilon}^s_N$, $s = 1, \ldots, d$. This is done by maximum-likelihood estimation. Since there are no explicit expressions for the parameter estimates, the estimation has to be performed numerically. All in all this yields the vector

$$\widehat{\boldsymbol{\beta}}^s := \begin{pmatrix} \widehat{\boldsymbol{\beta}}^s_{\mathrm{ls}} \\ \widehat{\boldsymbol{\beta}}^s_{\text{skew-}t} \end{pmatrix} = \left(\widehat{\beta}^s_0, \widehat{\beta}^s_{\sin}, \widehat{\beta}^s_{\cos}, \widehat{\gamma}^s_1, \widehat{\gamma}^s_2, \widehat{\gamma}^s_3, \widehat{\xi}^s, \widehat{\omega}^s, \widehat{\alpha}^s, \widehat{\nu}^s\right)^\top \qquad (3.5.2)$$

of estimated model parameters for $s = 1, \ldots, d$.

To check if our models (3.5.1) eliminate any kind of temporal dependence, respectively to verify if it is sufficient to include autoregression components up to a lag $q = 3$, we investigate the residuals obtained from the fitting process described above. For this purpose we apply the Ljung-Box test presented in Section 2.5. More precisely we compute

lag=	10	15	20	30	lag=	10	15	20	30
ange	0.282	0.404	0.240	0.325	kons	0.410	0.589	0.799	0.984
auee	0.440	0.207	0.360	0.786	kron	0.919	0.607	0.703	0.784
brle	0.576	0.540	0.723	0.951	ling	0.375	0.383	0.144	0.153
brln	0.146	0.144	0.089	0.174	lipp	0.245	0.566	0.735	0.694
bonn	0.176	0.385	0.649	0.937	lueb	0.288	0.311	0.377	0.114
brau	0.183	0.150	0.310	0.308	magd	0.074	0.058	0.156	0.283
brem	0.475	0.316	0.081	0.056	mann	0.206	0.330	0.521	0.860
cott	0.300	0.273	0.389	0.779	mein	0.053	0.112	0.241	0.174
cuxh	0.064	0.160	0.364	*0.011*	mont	0.212	0.443	0.517	0.886
dres	0.241	0.196	0.298	0.757	mnch	0.497	0.123	0.233	0.691
emde	0.280	0.335	0.164	*0.038*	mnst	0.338	0.375	0.198	0.285
esse	0.232	0.342	0.299	0.529	neub	0.124	*0.011*	0.063	0.356
fran	0.369	0.522	0.678	0.928	neur	0.151	0.122	0.077	0.063
frei	0.089	0.179	0.302	0.231	nuer	0.601	0.245	0.447	0.806
fuer	0.883	0.137	0.166	0.301	pidi	0.132	0.161	0.253	0.431
gard	0.209	0.121	0.105	0.171	rege	0.392	0.055	0.152	0.332
gies	0.097	0.209	0.437	0.790	rost	0.121	0.234	0.312	0.057
goer	0.168	0.207	0.105	0.321	saar	0.535	0.800	0.834	0.906
goet	0.199	0.335	0.519	0.754	schl	0.179	0.321	0.501	0.405
hall	0.124	0.232	0.336	0.637	stut	0.492	0.455	0.720	0.960
hamb	0.426	0.423	0.339	0.080	trie	0.130	0.298	0.277	0.426
hann	0.426	0.144	0.095	0.093	ueck	0.083	0.072	0.077	0.134
holz	0.272	0.159	0.265	0.657	uelz	0.306	0.318	0.059	0.051
jena	0.124	0.186	0.422	0.557	ulmm	0.348	0.254	0.567	0.928
kais	0.625	0.848	0.835	0.922	ware	0.287	0.167	0.136	0.070
kass	0.230	0.285	0.606	0.849	weid	0.783	0.183	0.406	0.807
kauf	0.139	*0.044*	0.131	0.268	wuer	0.151	0.157	0.323	0.773

Table 3.2: p-values of Ljung-Box tests with lags 10, 15, 20 and 30.

the respective p-values for different lags (10, 15, 20 and 30) for all 54 residual time series $\widehat{\varepsilon}_4^s, \ldots, \widehat{\varepsilon}_T^s$, $s = 1, \ldots, d$. The results of these computations are summarized in Table 3.2.

Only the four p-values for Kaufbeuren (15), Neuburg/Donau (15), Cuxhaven (30) and Emden (30) in Table 3.2 are smaller than 0.05. These would indicate the rejection of the serial independence hypothesis, but notice that the respective p-values for lower or higher maximum lags given in the table do not support the rejection. The overall picture emerging from Table 3.2 leads to the conclusion that our models are appropriate with regard to autoregression.

Finally, we investigate estimates of the expectation, standard deviation, skewness and kurtosis of the error series $\varepsilon_1^s, \ldots, \varepsilon_T^s$, $s = 1, \ldots, d$. These are summarized in Table 3.3 and are calculated by means of the formulas (3.4.2)-(3.4.4), based on the parameter estimates $\widehat{\xi}^s$, $\widehat{\omega}^s$, $\widehat{\alpha}^s$ and $\widehat{\nu}^s$, $s = 1, \ldots, d$. From the table we see, that $\widehat{\mathrm{E}(\varepsilon_t^s)} \approx 0$ for all 54 stations, i.e. the zero mean assumption for the errors is fulfilled. Moreover we observe, that the

standard deviations for the 54 observation stations lie between 1.115 and 1.521 and that most of them scatter closely around 1.35. We conclude that our weighting approach yields reasonable homoscedasticity.

	$\widehat{\mathrm{E}(\varepsilon_t^s)}$	$\widehat{\mathrm{sd}(\varepsilon_t^s)}$	$\widehat{\mathrm{skw}(\varepsilon_t^s)}$	$\widehat{\mathrm{krt}(\varepsilon_t^s)}$		$\widehat{\mathrm{E}(\varepsilon_t^s)}$	$\widehat{\mathrm{sd}(\varepsilon_t^s)}$	$\widehat{\mathrm{skw}(\varepsilon_t^s)}$	$\widehat{\mathrm{krt}(\varepsilon_t^s)}$
ange	0.001	1.344	0.035	0.650	kons	0.001	1.283	-0.333	1.149
auee	0.000	1.521	-0.239	0.598	kron	0.000	1.370	-0.081	0.518
brle	0.000	1.335	-0.163	0.359	ling	0.001	1.300	0.201	0.640
brln	0.000	1.377	-0.013	0.351	lipp	0.001	1.408	0.119	0.483
bonn	0.000	1.329	-0.001	0.336	lueb	0.003	1.326	0.446	1.395
brau	0.001	1.365	0.116	0.773	magd	0.001	1.361	0.033	0.437
brem	0.001	1.308	0.358	1.122	mann	0.000	1.308	-0.074	0.385
cott	0.000	1.488	-0.081	0.255	mein	0.001	1.363	-0.115	0.470
cuxh	0.001	1.115	0.537	2.090	mont	-0.000	1.263	0.057	0.499
dres	0.000	1.416	-0.180	0.264	mnch	0.003	1.502	-0.315	0.799
emde	0.001	1.204	0.334	1.099	mnst	0.000	1.361	0.194	0.516
esse	0.000	1.368	-0.022	0.414	neub	0.000	1.363	-0.182	1.000
fran	0.000	1.287	0.053	0.281	neur	0.000	1.296	0.048	0.445
frei	-0.000	1.385	-0.075	0.826	nuer	0.001	1.395	-0.133	0.756
fuer	0.001	1.337	-0.396	0.946	pidi	-0.001	1.351	-0.324	0.940
gard	0.003	1.417	0.159	0.923	rege	0.001	1.308	-0.255	0.752
gies	0.000	1.289	0.110	0.332	rost	0.002	1.165	0.513	1.407
goer	-0.000	1.446	-0.275	0.317	saar	0.000	1.304	-0.048	0.505
goet	0.001	1.322	0.070	0.464	schl	0.000	1.178	0.357	1.940
hall	0.001	1.414	0.014	0.482	stut	0.000	1.385	-0.162	0.971
hamb	0.001	1.313	0.366	1.177	trie	-0.000	1.254	0.087	0.342
hann	0.001	1.369	0.200	0.796	ueck	0.001	1.340	0.114	0.653
holz	0.000	1.470	-0.018	0.194	uelz	0.003	1.347	0.266	1.309
jena	0.001	1.419	-0.040	0.498	ulmm	0.002	1.389	-0.294	0.852
kais	0.000	1.312	-0.000	0.469	ware	0.001	1.285	0.178	0.664
kass	0.000	1.297	0.041	0.266	weid	0.001	1.408	-0.292	0.860
kauf	0.001	1.453	-0.208	0.788	wuer	0.001	1.358	-0.057	0.666

Table 3.3: Estimates of expectation, standard deviation, skewness and kurtosis of the error series for all 54 observation stations under the assumption of a skew-t distribution, calculated based on the parameter estimates $\widehat{\xi^s}$, $\widehat{\omega^s}$, $\widehat{\alpha^s}$ and $\widehat{\nu^s}$, $s = 1, \ldots, d$.

In Figure 3.5.2 the estimates $\widehat{\mathrm{skw}(\varepsilon_t^s)}$ and $\widehat{\mathrm{krt}(\varepsilon_t^s)}$ given in Table 3.3 are plotted against the latitude of the respective observation stations. From the plots we see, that both skewness and kurtosis alter from the north to the south. We observe positive skewness in the north and negative skewness in the south. For some of the observation stations there would be no need to allow for skewness in the model. But since there are stations whose skewness deviates evidently from zero, the allowance of skewness in our model is justified. Contrariwise the kurtosis is always positive and exhibits lower values in middle Germany, whereas the kurtosis in north and south Germany is relatively larger. The

presence of positive kurtosis justifies the model assumption of a leptokurtic distribution, i.e. the application of a skew-t distribution is justified.
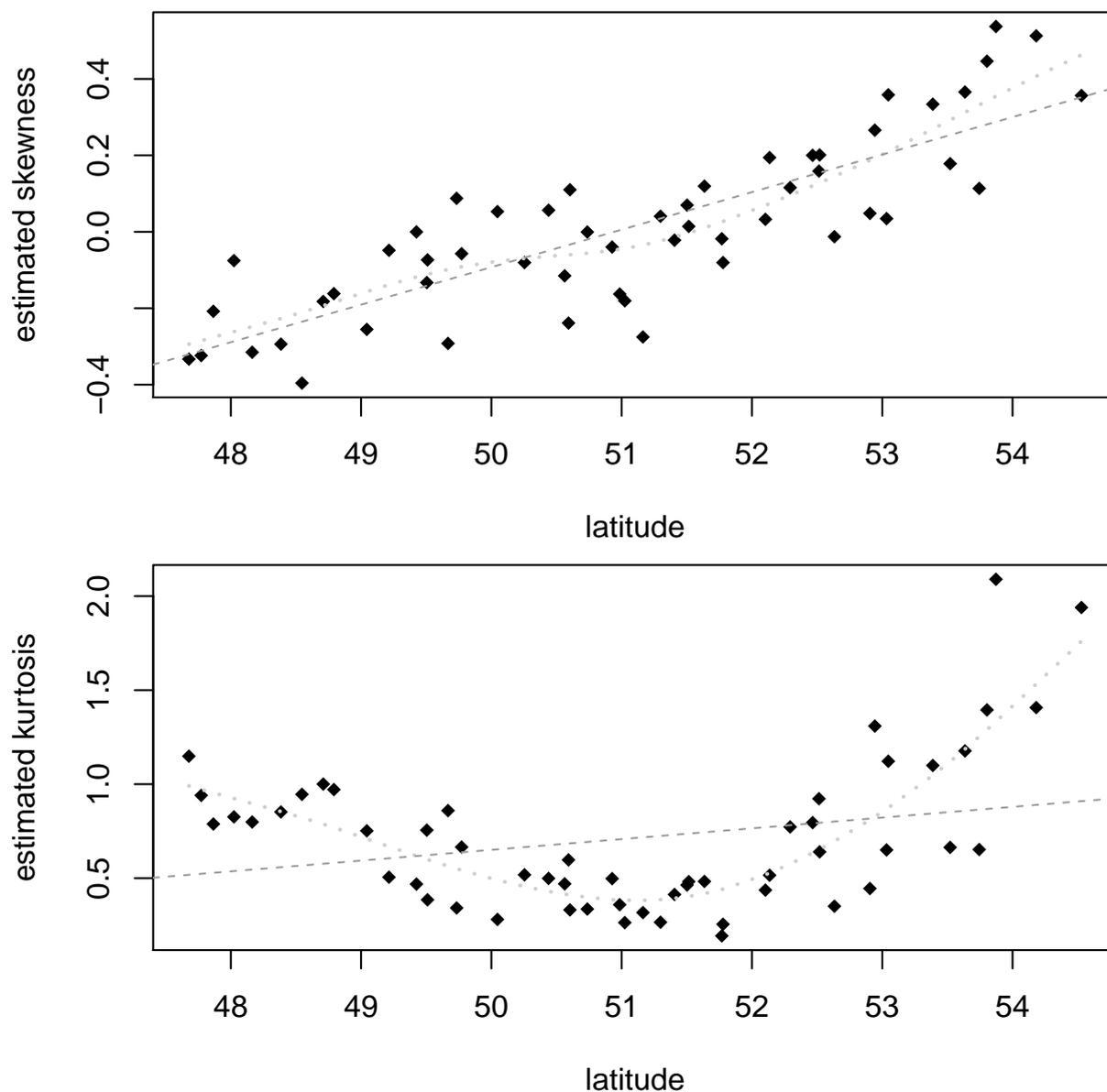
Figure 3.5.2: Plots of the skewness and kurtosis estimates given in Table 3.3 against latitude.

The trends observed in Figure 3.5.2 will be captured by the joint marginal model, which we are going to develop next.

3.6 Joint marginal model for daily mean temperature data

In this section we will investigate potential dependencies of the separate marginal model parameters on the covariates $x_{\mathrm{elev},s}$ (elevation), $x_{\mathrm{long},s}$ (longitude) and $x_{\mathrm{lat},s}$ (latitude) and develop a joint (space-time) marginal model, which models the mean temperatures Y_t^s over time $t = 1, \ldots, N$ and space $s = 1, \ldots, d$. In contrast to the separate models, the goal is a joint model for all observation stations where the common parameters do not depend on $s = 1, \ldots, d$.

From Figures 3.6.1-3.6.10 we clearly observe that the separate parameter estimates exhibit (polynomial-) trends with respect to the three covariates elevation, longitude and latitude. Therefore we investigate linear models for each of the ten components of the parameter estimates given in (3.5.2), where we regress each parameter estimate on polynomials of the covariates $x_{\mathrm{elev},s}$, $x_{\mathrm{long},s}$ and $x_{\mathrm{lat},s}$ up to a degree of nine, i.e. we search for models of the form

$$\widehat{\theta}^s \approx \theta_0 + \sum_{j=1}^{p} \theta_{1j} x_{\mathrm{elev},s}^j + \sum_{k=1}^{q} \theta_{2k} x_{\mathrm{long},s}^k + \sum_{l=1}^{r} \theta_{3l} x_{\mathrm{lat},s}^l \qquad (3.6.1)$$

where $\widehat{\theta}^s$ stands for an arbitrary parameter of the parameter vector $\widehat{\boldsymbol{\beta}}^s$ in (3.5.2) (or its logarithm, if the parameter is restricted to the positive real numbers) and the degrees $p, q, r \leq 9$ of the polynomials are the maximum permissible degrees such that the parameters θ_{1p}, θ_{2q}, and θ_{3r} corresponding to the covariates $x_{\mathrm{elev},s}^p$, $x_{\mathrm{long},s}^q$ and $x_{\mathrm{lat},s}^r$ are significant.

	p	q	r	# par	R^2_{adj}	R^2_{adj} (3)	R^2_{adj} (9)
β_0	1	0	1	3	0.8043	0.8028	0.7431
$\beta_{\sin}$	4	1	6	12	0.5550	0.4400	0.5078
$\beta_{\cos}$	6	2	1	10	0.9383	0.9245	0.9306
γ_1	1	2	6	10	0.7051	0.6478	0.6452
γ_2	1	2	6	10	0.6879	0.6101	0.6686
γ_3	0	4	7	12	0.5723	0.4331	0.5635
ξ	1	2	1	5	0.8345	0.8416	0.8568
ω	3	1	6	11	0.5370	0.5123	0.5164
α	4	2	1	8	0.8689	0.8618	0.8947
ν	2	2	4	9	0.7517	0.6990	0.7302
Sum	**23**	**18**	**39**	**90**			

Table 3.4: Analysis of the dependence of univariate marginal model parameters on the covariates elevation, longitude and latitude.

Table 3.4 summarizes the selected models and compares the R^2_{adj} to those of the corresponding models with p, q and r fixed to 3 or 9 respectively. Except for the parameters ξ and α, the R^2_{adj} of the selected models, which are quite high, are bigger than the ones of the comparison models. Due to the fact that our models for ξ and α need much less
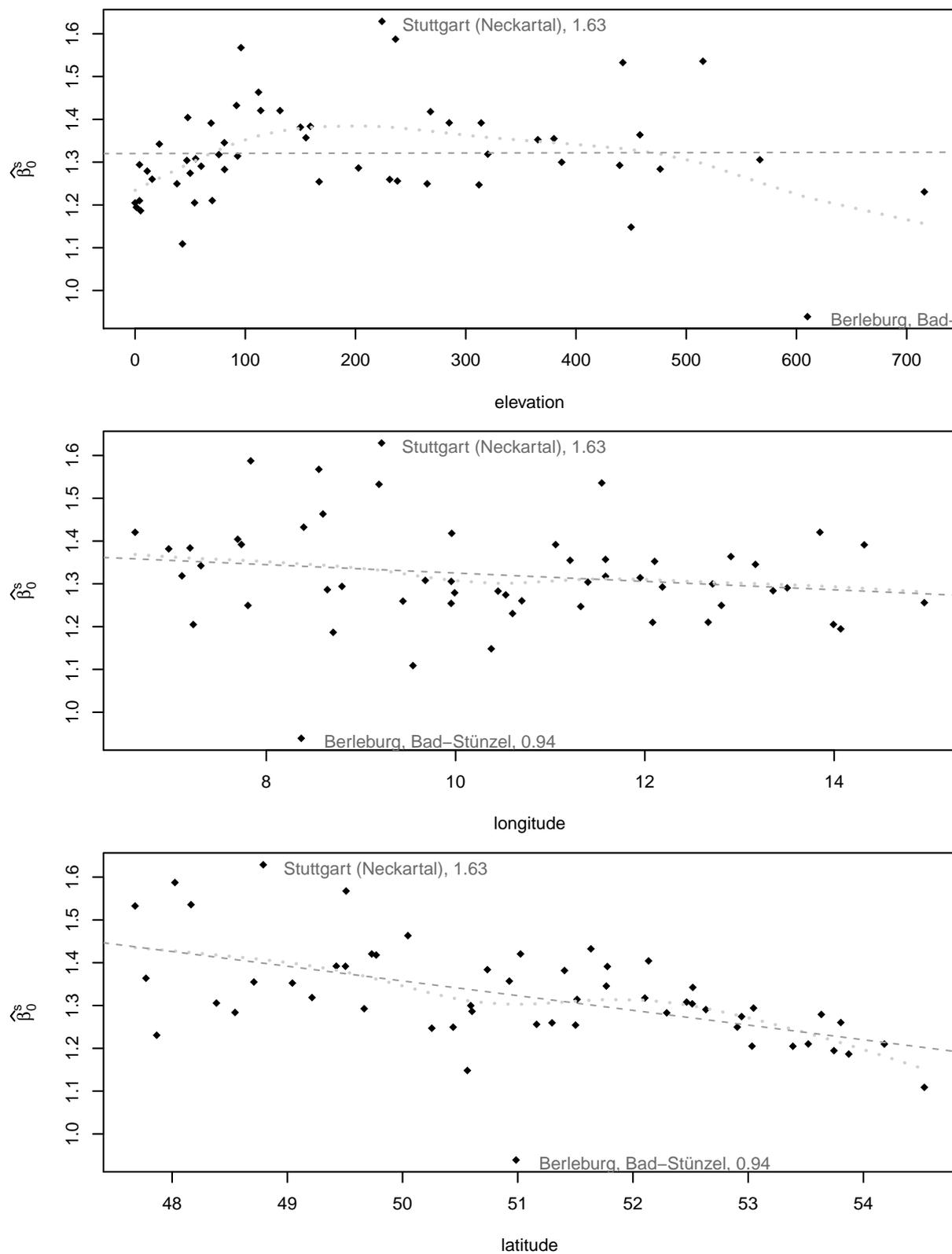
Figure 3.6.1: Plots of the parameter estimates $\widehat{\beta}_0^s$ against elevation, longitude and latitude, respectively.
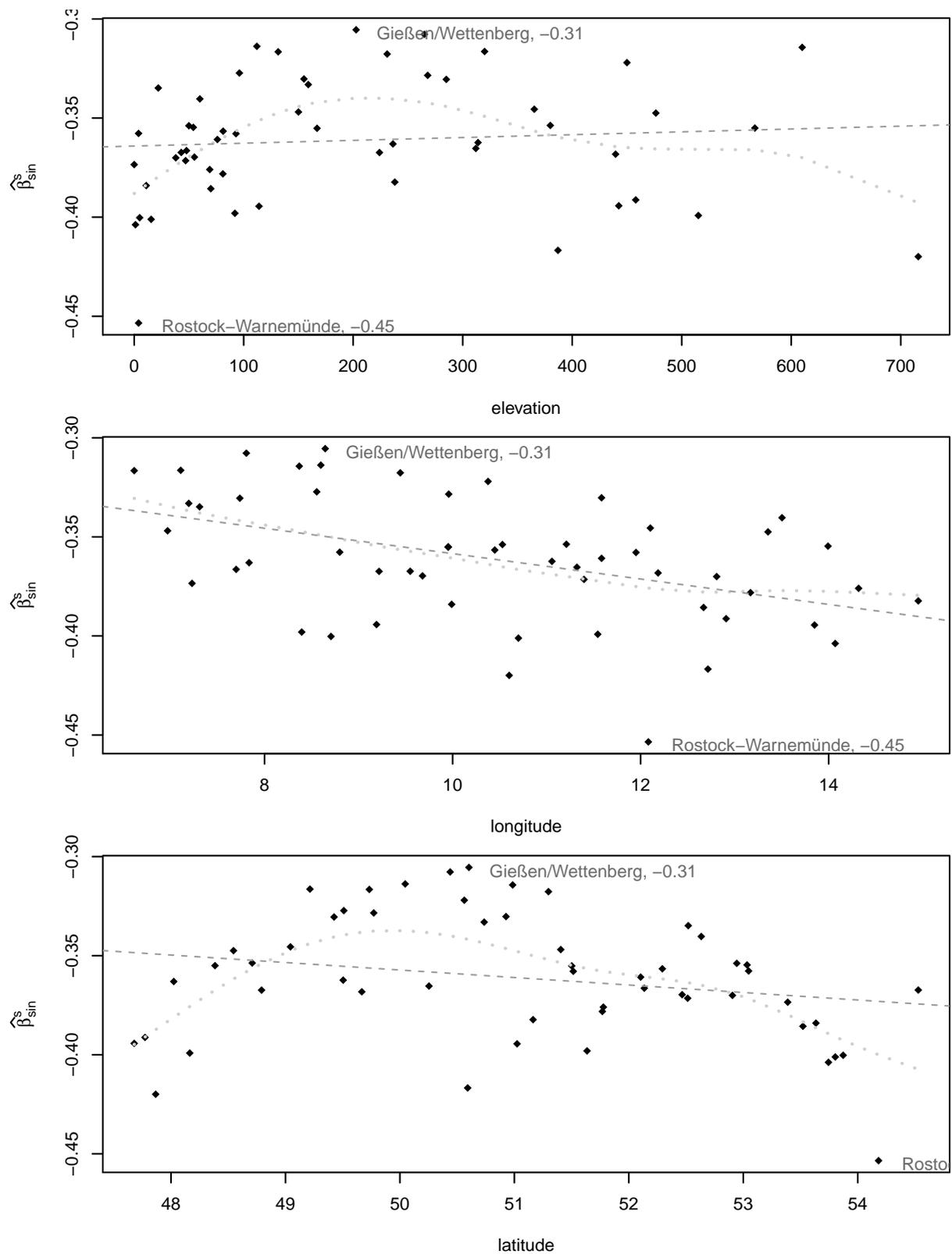
Figure 3.6.2: Plots of the parameter estimates $\widehat{\beta}_{\sin}^s$ against elevation, longitude and latitude, respectively.
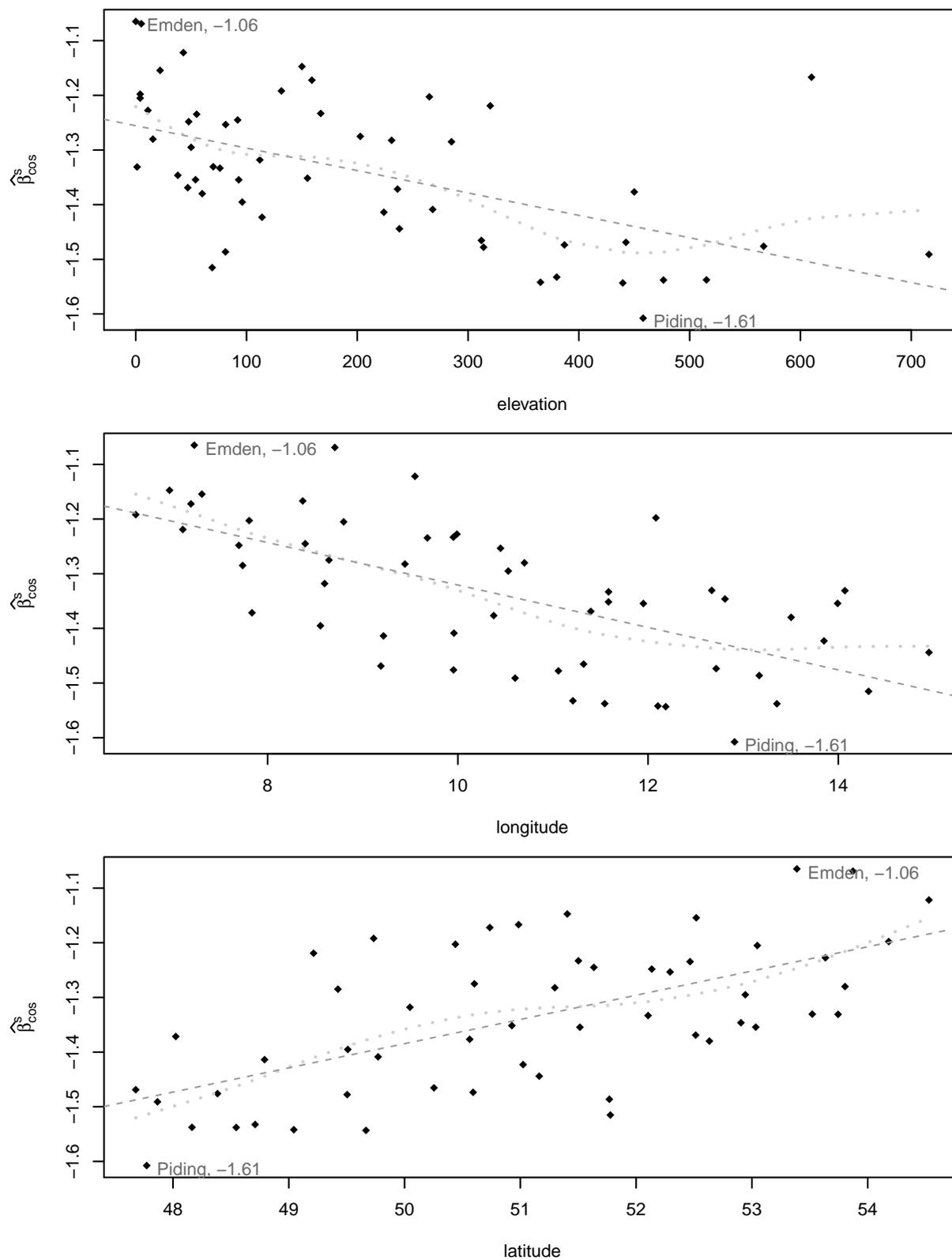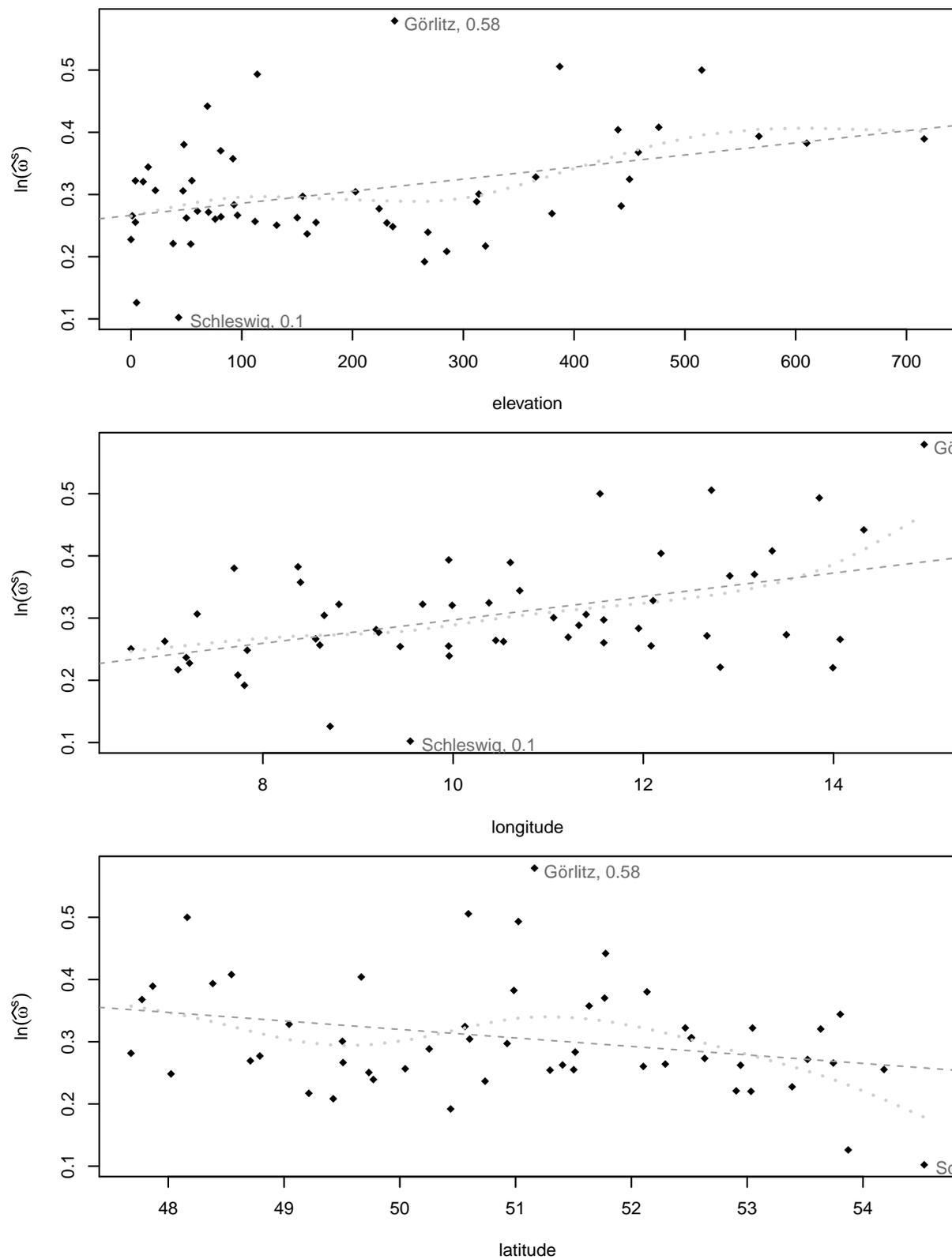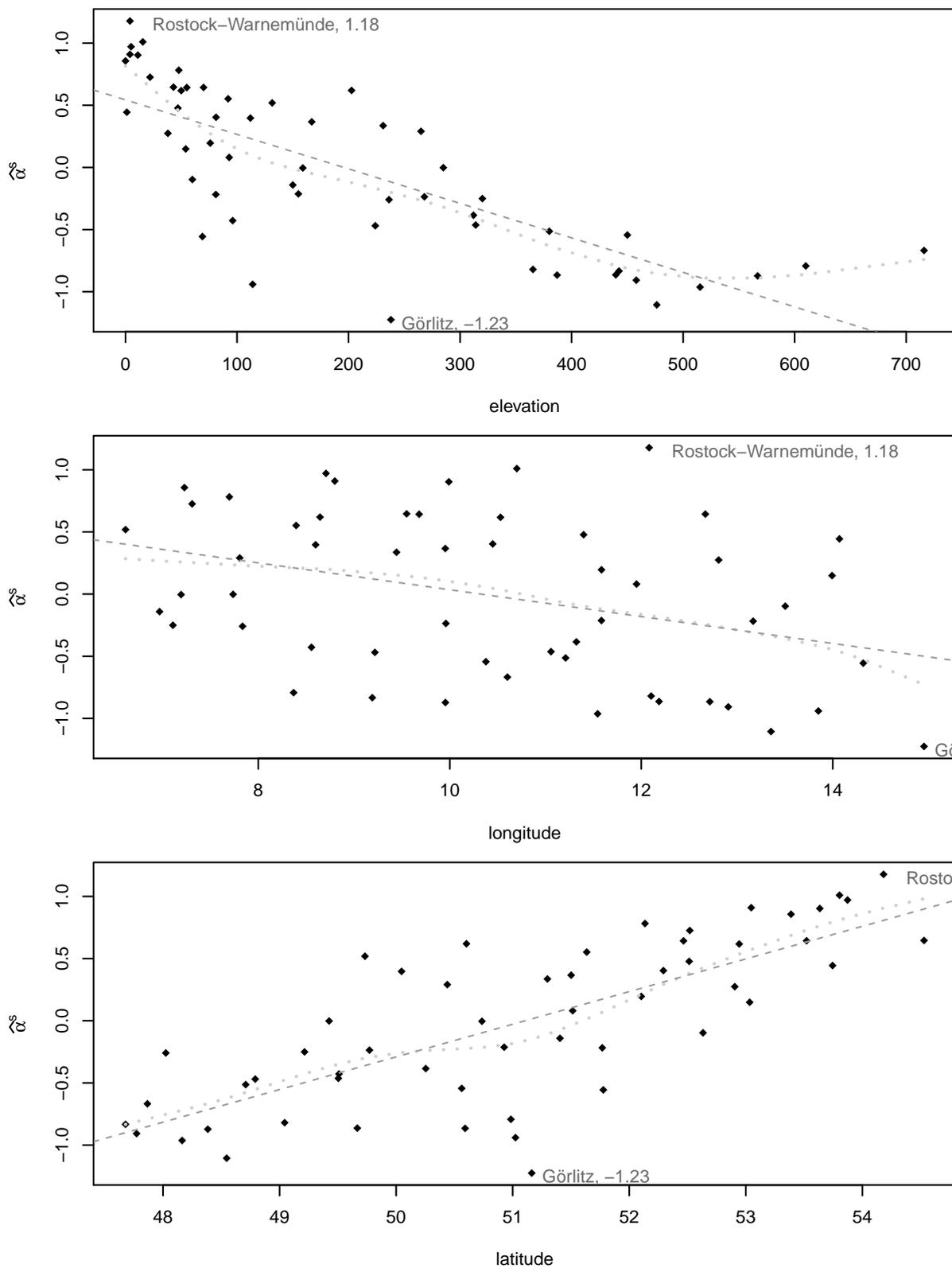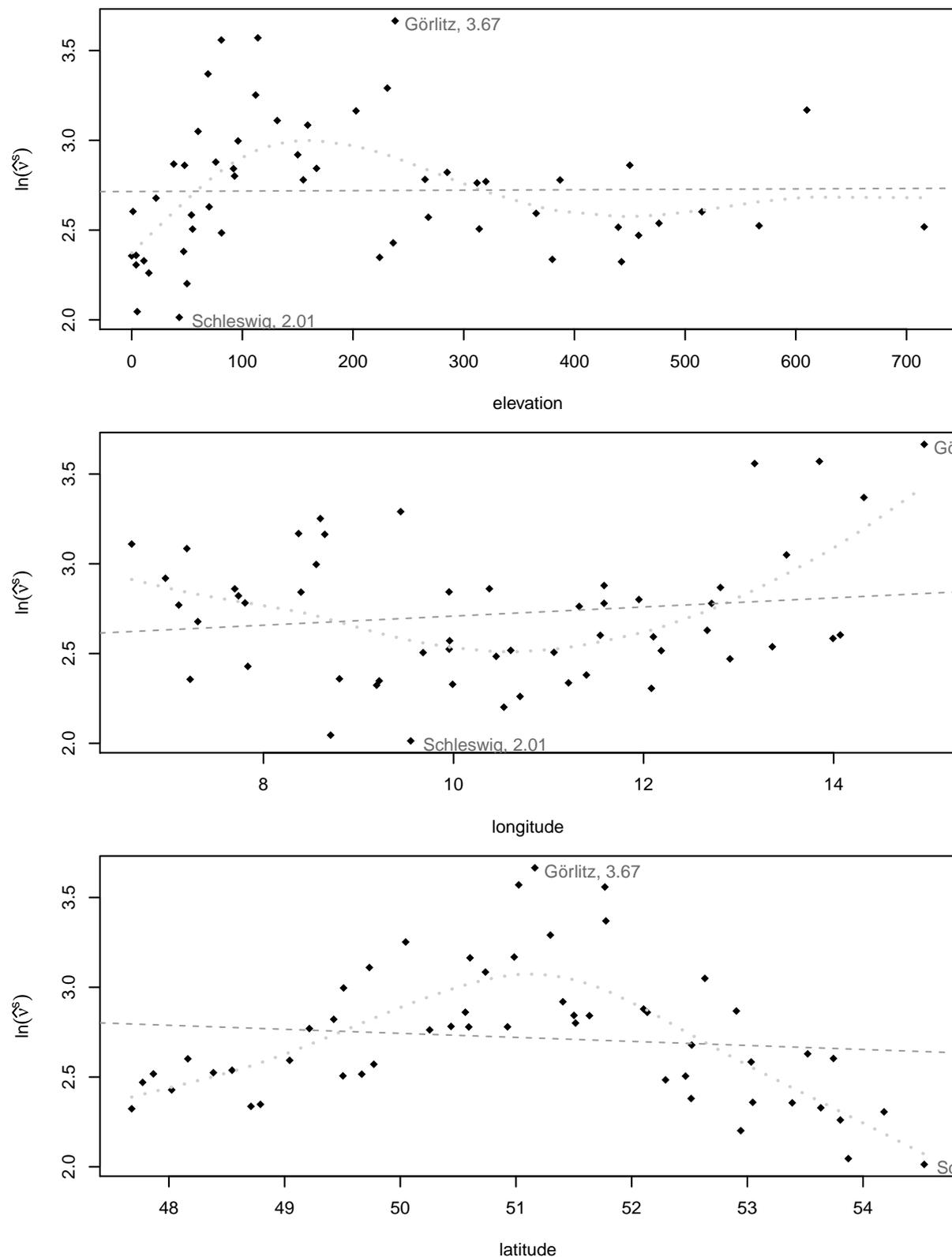
Figure 3.6.3: Plots of the parameter estimates $\widehat{\beta}_{\cos}^s$ against elevation, longitude and latitude, respectively.
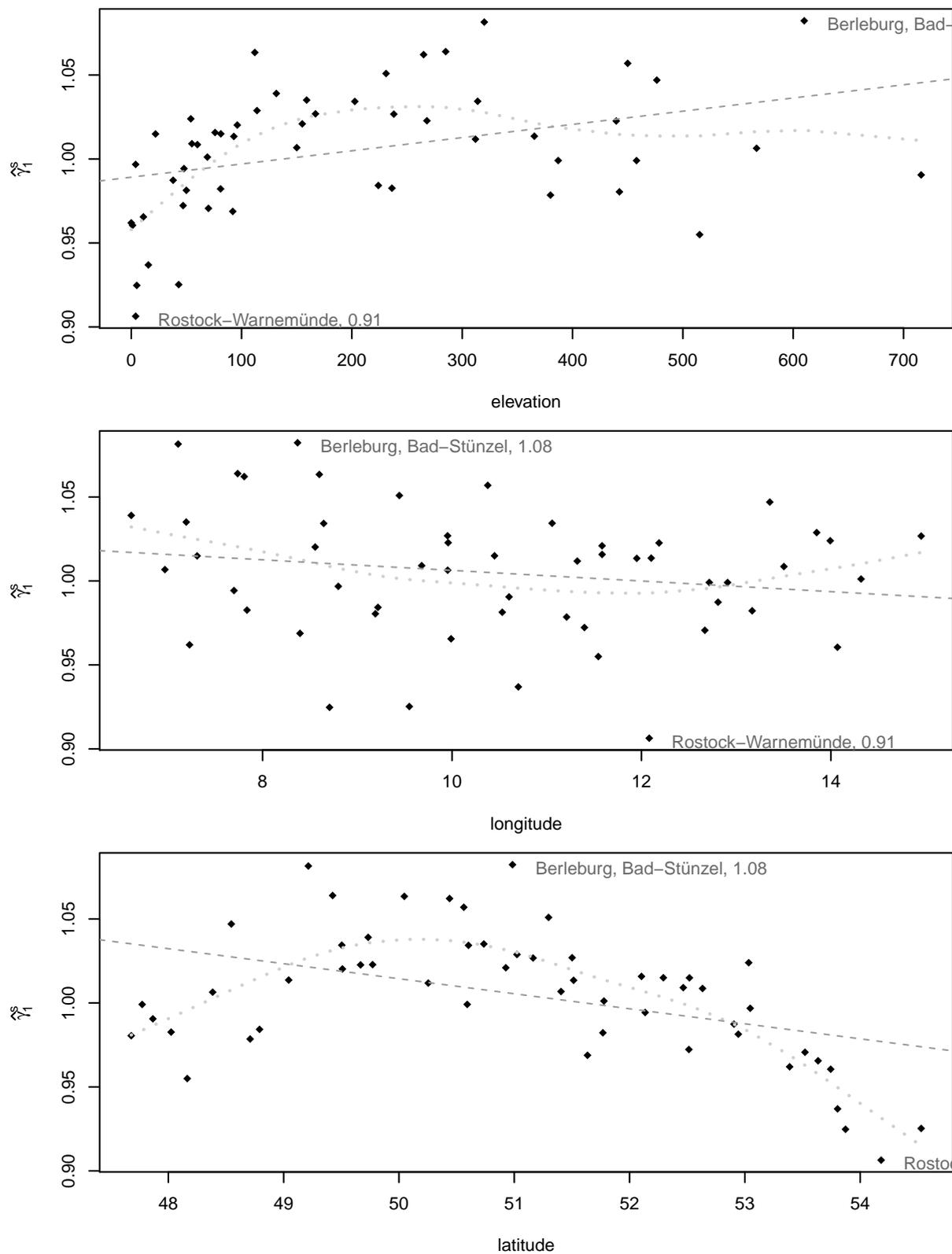
Figure 3.6.4: Plots of the parameter estimates $\widehat{\gamma}_1^s$ against elevation, longitude and latitude, respectively.
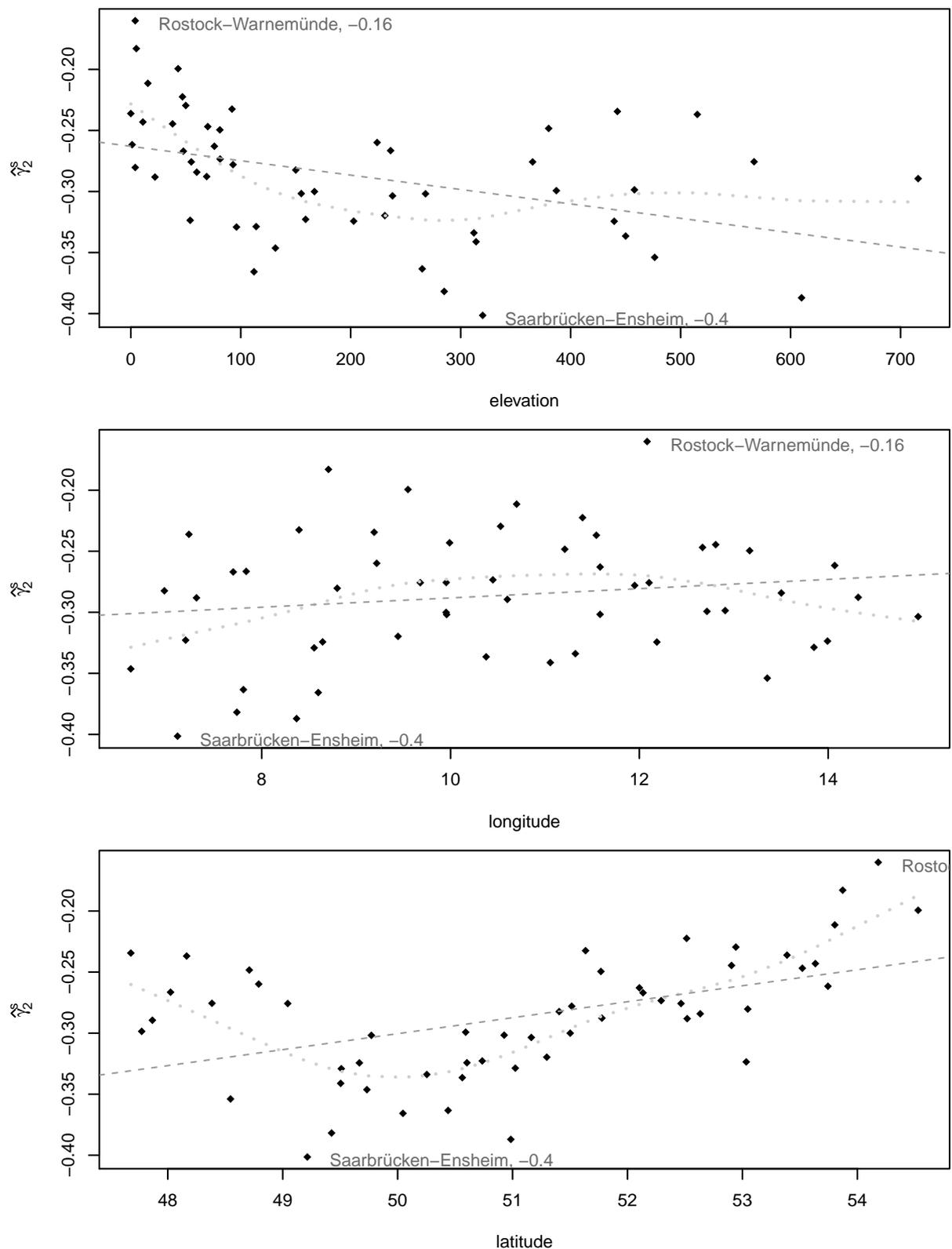
Figure 3.6.5: Plots of the parameter estimates $\widehat{\gamma}_2^s$ against elevation, longitude and latitude, respectively.
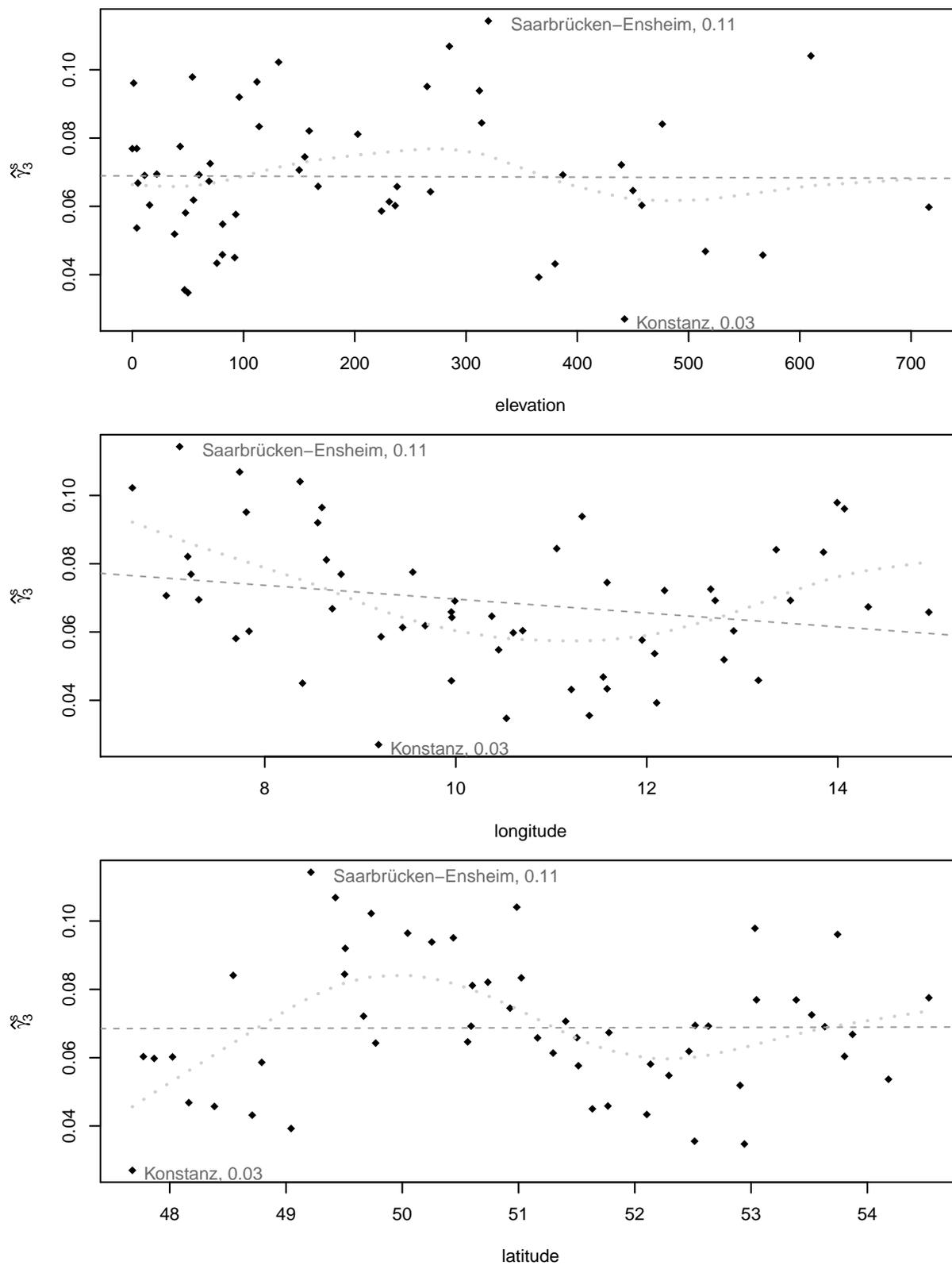
Figure 3.6.6: Plots of the parameter estimates $\widehat{\gamma}_3^s$ against elevation, longitude and latitude, respectively.
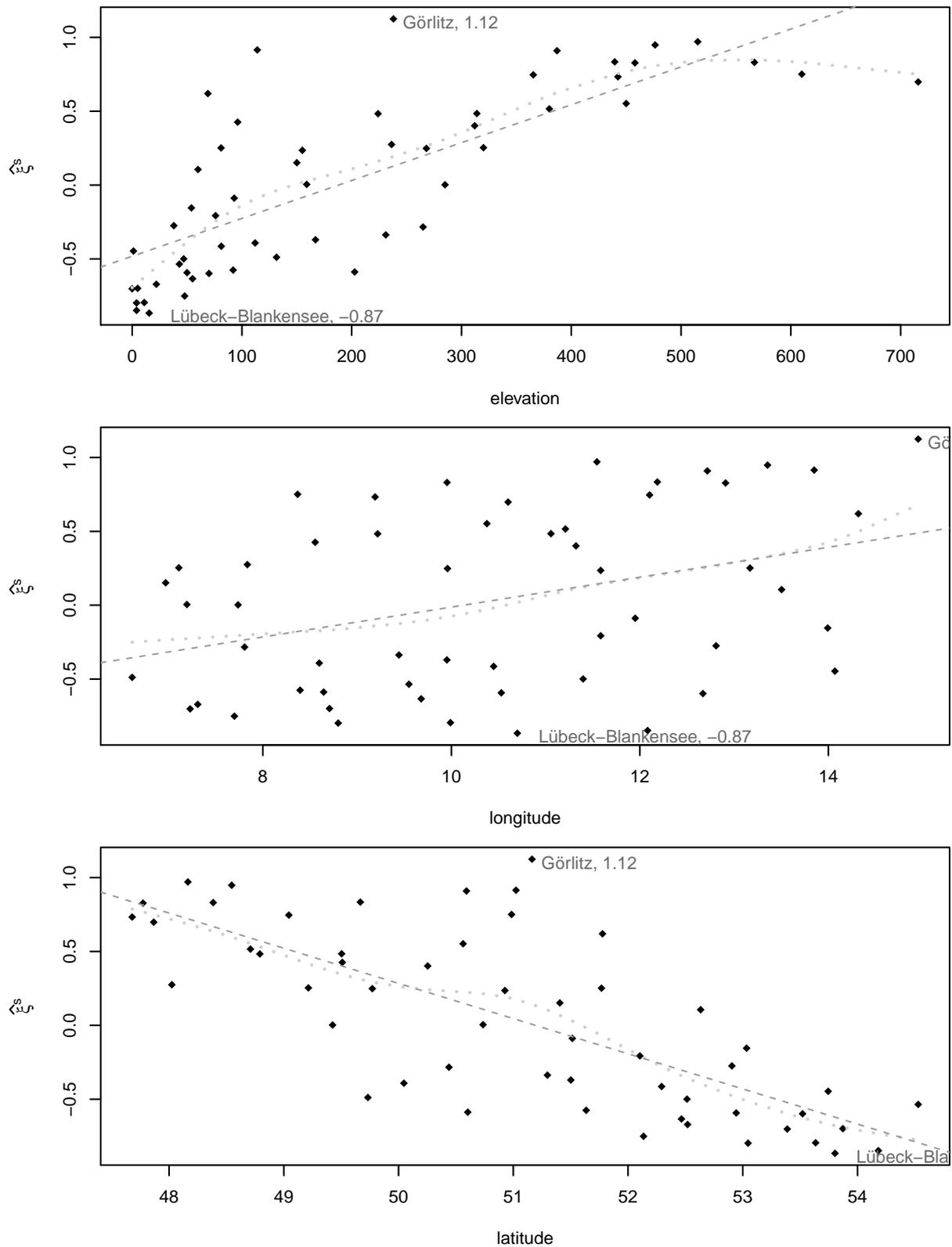
Figure 3.6.7: Plots of the parameter estimates $\widehat{\xi}^s$ against elevation, longitude and latitude, respectively.

Figure 3.6.8: Plots of the logarithmized parameter estimates $\ln(\widehat{\omega}^s)$ against elevation, longitude and latitude, respectively.

Figure 3.6.9: Plots of the parameter estimates $\widehat{\alpha}^s$ against elevation, longitude and latitude, respectively.

Figure 3.6.10: Plots of the logarithmized parameter estimates $\ln(\widehat{\nu}^s)$ against elevation, longitude and latitude, respectively.

parameters than the corresponding models with polynomial degrees fixed to nine and that the difference in the $\mathrm{R}^2_{\mathrm{adj}}$ of these models is not that big, we are satisfied with our selection. Thus we obtain the polynomials respectively exponentials of polynomials

$$\beta_0(s) := \beta_{00} + \sum_{j=1}^{1} \beta_{01j} x^j_{\mathrm{elev},s} + \sum_{l=1}^{1} \beta_{03l} x^l_{\mathrm{lat},s},$$

$$\beta_{\sin}(s) := \beta_{\sin 0} + \sum_{j=1}^{4} \beta_{\sin 1j} x^j_{\mathrm{elev},s} + \sum_{k=1}^{1} \beta_{\sin 2k} x^k_{\mathrm{long},s} + \sum_{l=1}^{6} \beta_{\sin 3l} x^l_{\mathrm{lat},s},$$

$$\beta_{\cos}(s) := \beta_{\cos 0} + \sum_{j=1}^{6} \beta_{\cos 1j} x^j_{\mathrm{elev},s} + \sum_{k=1}^{2} \beta_{\cos 2k} x^k_{\mathrm{long},s} + \sum_{l=1}^{1} \beta_{\cos 3l} x^l_{\mathrm{lat},s},$$

$$\gamma_1(s) := \gamma_{10} + \sum_{j=1}^{1} \gamma_{11j} x^j_{\mathrm{elev},s} + \sum_{k=1}^{2} \gamma_{12k} x^k_{\mathrm{long},s} + \sum_{l=1}^{6} \gamma_{13l} x^l_{\mathrm{lat},s},$$

$$\gamma_2(s) := \gamma_{20} + \sum_{j=1}^{1} \gamma_{21j} x^j_{\mathrm{elev},s} + \sum_{k=1}^{2} \gamma_{22k} x^k_{\mathrm{long},s} + \sum_{l=1}^{6} \gamma_{23l} x^l_{\mathrm{lat},s},$$

$$\gamma_3(s) := \gamma_{30} + \sum_{k=1}^{4} \gamma_{32k} x^k_{\mathrm{long},s} + \sum_{l=1}^{7} \gamma_{33l} x^l_{\mathrm{lat},s},$$

$$\xi(s) := \xi_0 + \sum_{j=1}^{1} \xi_{1j} x^j_{\mathrm{elev},s} + \sum_{k=1}^{2} \xi_{2k} x^k_{\mathrm{long},s} + \sum_{l=1}^{1} \xi_{3l} x^l_{\mathrm{lat},s},$$

$$\omega(s) := \exp\left\{ \omega_0 + \sum_{j=1}^{3} \omega_{1j} x^j_{\mathrm{elev},s} + \sum_{k=1}^{1} \omega_{2k} x^k_{\mathrm{long},s} + \sum_{l=1}^{6} \omega_{3l} x^l_{\mathrm{lat},s} \right\},$$

$$\alpha(s) := \alpha_0 + \sum_{j=1}^{4} \alpha_{1j} x^j_{\mathrm{elev},s} + \sum_{k=1}^{2} \alpha_{2k} x^k_{\mathrm{long},s} + \sum_{l=1}^{1} \alpha_{3l} x^l_{\mathrm{lat},s},$$

$$\nu(s) := \exp\left\{ \nu_0 + \sum_{j=1}^{2} \nu_{1j} x^j_{\mathrm{elev},s} + \sum_{k=1}^{2} \nu_{2k} x^k_{\mathrm{long},s} + \sum_{l=1}^{4} \nu_{3l} x^l_{\mathrm{lat},s} \right\}.$$

For reasons of convenience we summarize the parameters in the two vectors

$$\boldsymbol{\beta}_{\mathrm{ls}} := \left(\boldsymbol{\beta}_0^\top, \boldsymbol{\beta}_{\sin}^\top, \boldsymbol{\beta}_{\cos}^\top, \boldsymbol{\gamma}_1^\top, \boldsymbol{\gamma}_2^\top, \boldsymbol{\gamma}_3^\top \right)^\top$$

and

$$\boldsymbol{\beta}_{\mathrm{skew}\text{-}t} := \left(\boldsymbol{\xi}^\top, \boldsymbol{\omega}^\top, \boldsymbol{\alpha}^\top, \boldsymbol{\nu}^\top \right)^\top,$$

where

$$\boldsymbol{\beta}_0 := (\beta_{00}, \beta_{011}, \beta_{031})^\top,$$

$$\boldsymbol{\beta}_{\sin} := (\beta_{\sin 0}, \beta_{\sin 11}, \dots, \beta_{\sin 14}, \beta_{\sin 21}, \beta_{\sin 31}, \dots, \beta_{\sin 36})^\top,$$

$$\boldsymbol{\beta}_{\cos} := (\beta_{\cos 0}, \beta_{\cos 11}, \dots, \beta_{\cos 16}, \beta_{\cos 21}, \beta_{\cos 22}, \beta_{\cos 31})^\top,$$

$$
\begin{aligned}
\boldsymbol{\gamma}_1 &:= \left(\gamma_{10}, \gamma_{111}, \gamma_{121}, \gamma_{122}, \gamma_{131}, \ldots, \gamma_{136}\right)^\top, \\
\boldsymbol{\gamma}_2 &:= \left(\gamma_{20}, \gamma_{211}, \gamma_{221}, \gamma_{222}, \gamma_{231}, \ldots, \gamma_{236}\right)^\top, \\
\boldsymbol{\gamma}_3 &:= \left(\gamma_{30}, \gamma_{321}, \ldots, \gamma_{324}, \gamma_{331}, \ldots, \gamma_{337}\right)^\top, \\
\boldsymbol{\xi} &:= \left(\xi_0, \xi_{11}, \xi_{21}, \xi_{22}, \xi_{31}\right)^\top, \\
\boldsymbol{\omega} &:= \left(\omega_0, \omega_{11}, \ldots, \omega_{13}, \omega_{21}, \omega_{31}, \ldots, \omega_{36}\right)^\top, \\
\boldsymbol{\alpha} &:= \left(\alpha_0, \alpha_{11}, \ldots, \alpha_{14}, \alpha_{21}, \alpha_{22}, \alpha_{31}\right)^\top, \\
\boldsymbol{\nu} &:= \left(\nu_0, \nu_{11}, \nu_{12}, \nu_{21}, \nu_{22}, \nu_{31}, \ldots, \nu_{34}\right)^\top.
\end{aligned}
$$

The results above allow us to state our final joint marginal model. For this purpose we replace the parameters of the models (3.5.1) by the functions $\beta_0(s)$, $\beta_{\sin}(s)$, $\beta_{\cos}(s)$, $\gamma_1(s)$, $\gamma_2(s)$, $\gamma_3(s)$, $\xi(s)$, $\omega(s)$, $\alpha(s)$, $\nu(s)$ derived above. Eventually we consider the model

$$
\begin{aligned}
\widetilde{Y}_t^s &= g\left(t, \widetilde{Y}_{t-1}^s, \widetilde{Y}_{t-2}^s, \widetilde{Y}_{t-3}^s, x_{\mathrm{elev},s}, x_{\mathrm{long},s}, x_{\mathrm{lat},s}; \boldsymbol{\beta}_{\mathrm{ls}}\right) + \varepsilon_t^s, \\
\varepsilon_t^s &\sim \mathrm{skew}\text{-}t\left(\xi(s), \omega(s), \alpha(s), \nu(s)\right), \\
&\quad t = 1, \ldots, N, s = 1, \ldots, d.
\end{aligned}
\tag{3.6.2}
$$

where

$$
\begin{aligned}
g(t,s) &:= g\left(t, \widetilde{Y}_{t-1}^s, \widetilde{Y}_{t-2}^s, \widetilde{Y}_{t-3}^s, x_{\mathrm{elev},s}, x_{\mathrm{long},s}, x_{\mathrm{lat},s}; \boldsymbol{\beta}_{\mathrm{ls}}\right) \\
&:= \beta_0(s) + \beta_{\sin}(s)\sin\left(\frac{2\pi t}{365.25}\right) + \beta_{\cos}(s)\cos\left(\frac{2\pi t}{365.25}\right) \\
&\quad + \gamma_1(s)\widetilde{Y}_{t-1}^s + \gamma_2(s)\widetilde{Y}_{t-2}^s + \gamma_3(s)\widetilde{Y}_{t-3}^s.
\end{aligned}
$$

The parameter estimation for our joint model (3.6.2) is performed analogously to Section 3.5. Consequently the estimates $\widehat{\boldsymbol{\beta}}_{\mathrm{ls}}$ are obtained by least-squares estimation, i.e. as

$$
\widehat{\boldsymbol{\beta}}_{\mathrm{ls}} = \underset{\boldsymbol{\beta}_{\mathrm{ls}}}{\mathrm{argmin}} \sum_{s=1}^{d}\sum_{t=4}^{N}\left(\widetilde{y}_t^s - g\left(t, \widetilde{y}_{t-1}^s, \widetilde{y}_{t-2}^s, \widetilde{y}_{t-3}^s, x_{\mathrm{elev},s}, x_{\mathrm{long},s}, x_{\mathrm{lat},s}; \boldsymbol{\beta}_{\mathrm{ls}}\right)\right)^2.
$$

Thereafter a reparametrized skew-t distribution is fitted to the residuals resulting from the least-squares estimation. Again maximum-likelihood estimation is applied, but this time the parameters of the likelihood corresponding to an ordinary skew-t distribution are replaced by the polynomials $\xi(s)$, $\alpha(s)$, $\omega(s)$ and $\nu(s)$ from above. The whole procedure results in the vector

$$
\widehat{\boldsymbol{\beta}} := \begin{pmatrix} \widehat{\boldsymbol{\beta}}_{\mathrm{ls}} \\ \widehat{\boldsymbol{\beta}}_{\mathrm{skew}\text{-}t} \end{pmatrix} = \left(\widehat{\boldsymbol{\beta}}_0^\top, \widehat{\boldsymbol{\beta}}_{\sin}^\top, \widehat{\boldsymbol{\beta}}_{\cos}^\top, \widehat{\boldsymbol{\gamma}}_1^\top, \widehat{\boldsymbol{\gamma}}_2^\top, \widehat{\boldsymbol{\gamma}}_3^\top, \widehat{\boldsymbol{\xi}}^\top, \widehat{\boldsymbol{\omega}}^\top, \widehat{\boldsymbol{\alpha}}^\top, \widehat{\boldsymbol{\nu}}^\top\right)^\top
$$

of parameter estimates for Model (3.6.2).

Figures 3.6.11-3.6.13 illustrate the components of the estimated model (3.6.2) by means of level plots. The depicted components are evaluated on the 80×120 grid, which was already introduced in Figure 3.1.2. For all figures a scale within a reasonable range was

chosen. Thus all plots show some white regions for extreme locations, where the polynomials of our model yield inadequate values. These extreme locations are locations which lie distinctly outside of the longitude, latitude and elevation range of the data on which the model is based. A prediction of mean temperatures of such locations based on our marginal model will lead to values diverging considerably from the temperatures which occur in reality. To prevent this, some further adjustments of the model will be necessary.

From Figure 3.6.11 we see that the mean value $\widehat{\beta}_0(s)$ of the daily mean temperature yields higher values for upcountry parts of Germany which are low and comparatively low values for mountainous areas and areas close to the sea. Moreover we observe that the autoregression components with lag one and two ($\widehat{\gamma}_1(s)$ and $\widehat{\gamma}_2(s)$) act in an opponent fashion. For northern and southern parts of Germany the mean temperature tends to decrease compared to the previous day, whereas it tends to increase in the rest of Germany. The mean temperature depends negatively on the temperature observed two days ago. Furthermore we notice that the influence of three days ago ($\widehat{\gamma}_3(s)$) is relatively low.

The spatial visualization of the skew-t parameters is given in Figures 3.6.12. Whereas the location parameter $\widehat{\xi}(s)$ increases from north-west to south-east, the shape parameter $\widehat{\alpha}(s)$ behaves the other way round. The structure of the shape parameters yields that for areas close to the North Sea positive temperature extremes are more likely and for more continental areas negative temperature extremes occur more often. The scale parameter $\widehat{\omega}(s)$ shows its local maxima in the eastern respectively the south-eastern parts of Germany. Following, these regions exhibit relatively broader fluctuations in mean temperature. Moreover the degrees of freedom $\widehat{\nu}(s)$ are relatively low in the north and in the south on the one hand and they increase westward and eastward on the other hand, meaning that heavy distribution tails are more present in the north and in the south of Germany, yielding more temperature extremes in these areas.

Figure 3.6.13 represents the seasonality component of our model. Since a meaningful interpretation of $\widehat{\beta}_{\sin}(s)$ and $\widehat{\beta}_{\cos}(s)$ is not possible, we use (3.2.2) to transform our parameters back to the amplitude $\widehat{\lambda}(s)$ and the phase shift $\widehat{\delta}(s)$ of the sinusoidal model for the seasonality. This allows for an interpretation of the annual temperature oscillations. We clearly observe higher amplitude of the seasonal fluctuations if we move further to the south-east, which is due to the fact that the climate in the north-east of Germany is rather maritime, whereas the south-east exhibits more continental climate. The parameters $\widehat{\delta}(s)$ are governed by the seasons. The other way round we can deduce from these parameters, on which days the colder respectively the warmer half-year begins. For instance in the south-east of Germany where $\widehat{\delta}(s)$ exhibits values around -1.8, the warmer half-year lasts from the middle of April to the middle of October. For the north-east of Germany this period starts up to eight days later, if we rely on our model.
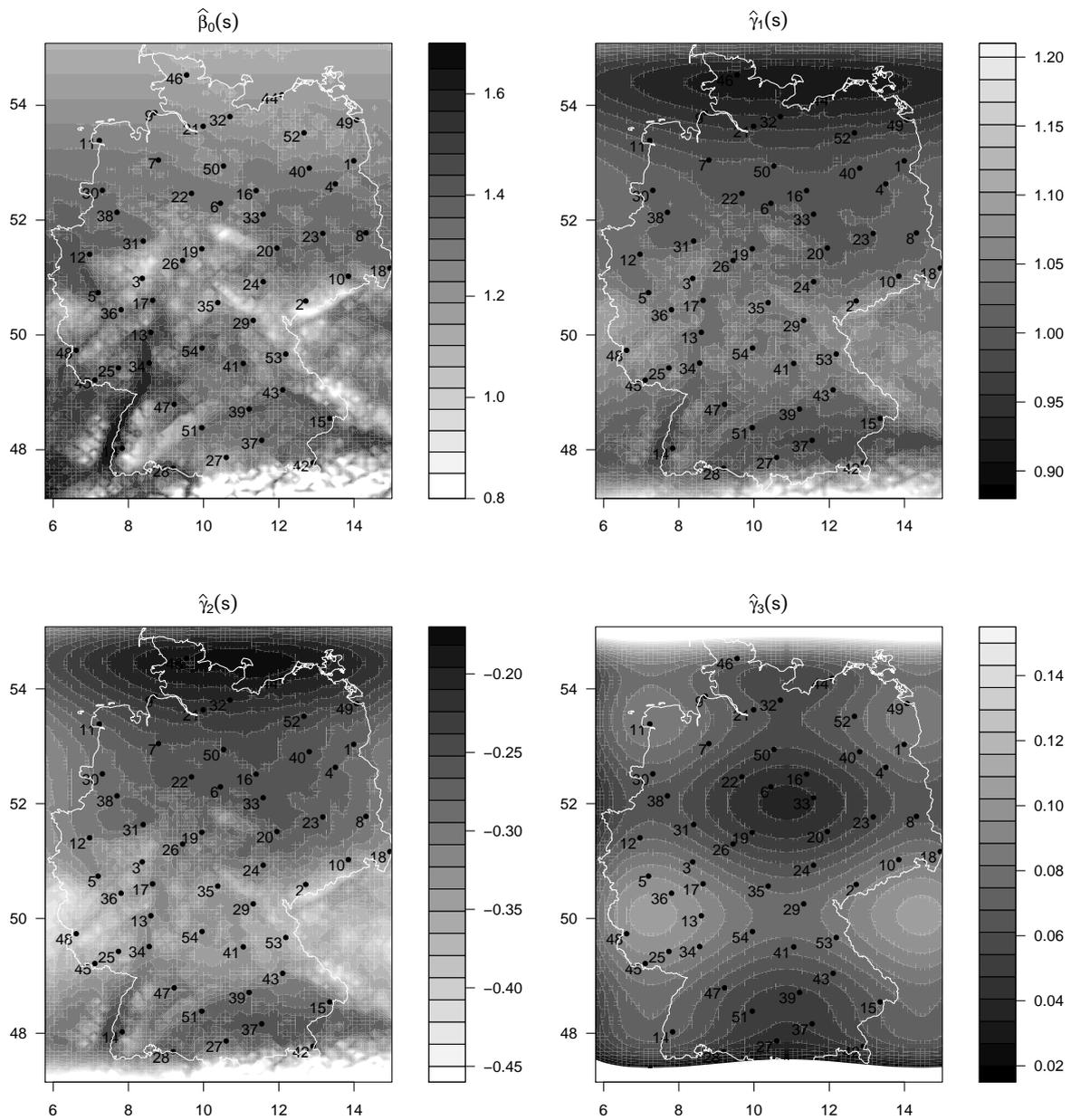
Figure 3.6.11: Level plots of the estimated polynomials $\widehat{\beta}_0(s)$, $\widehat{\gamma}_1(s)$, $\widehat{\gamma}_2(s)$ and $\widehat{\gamma}_3(s)$ with the locations of the observation stations indicated.
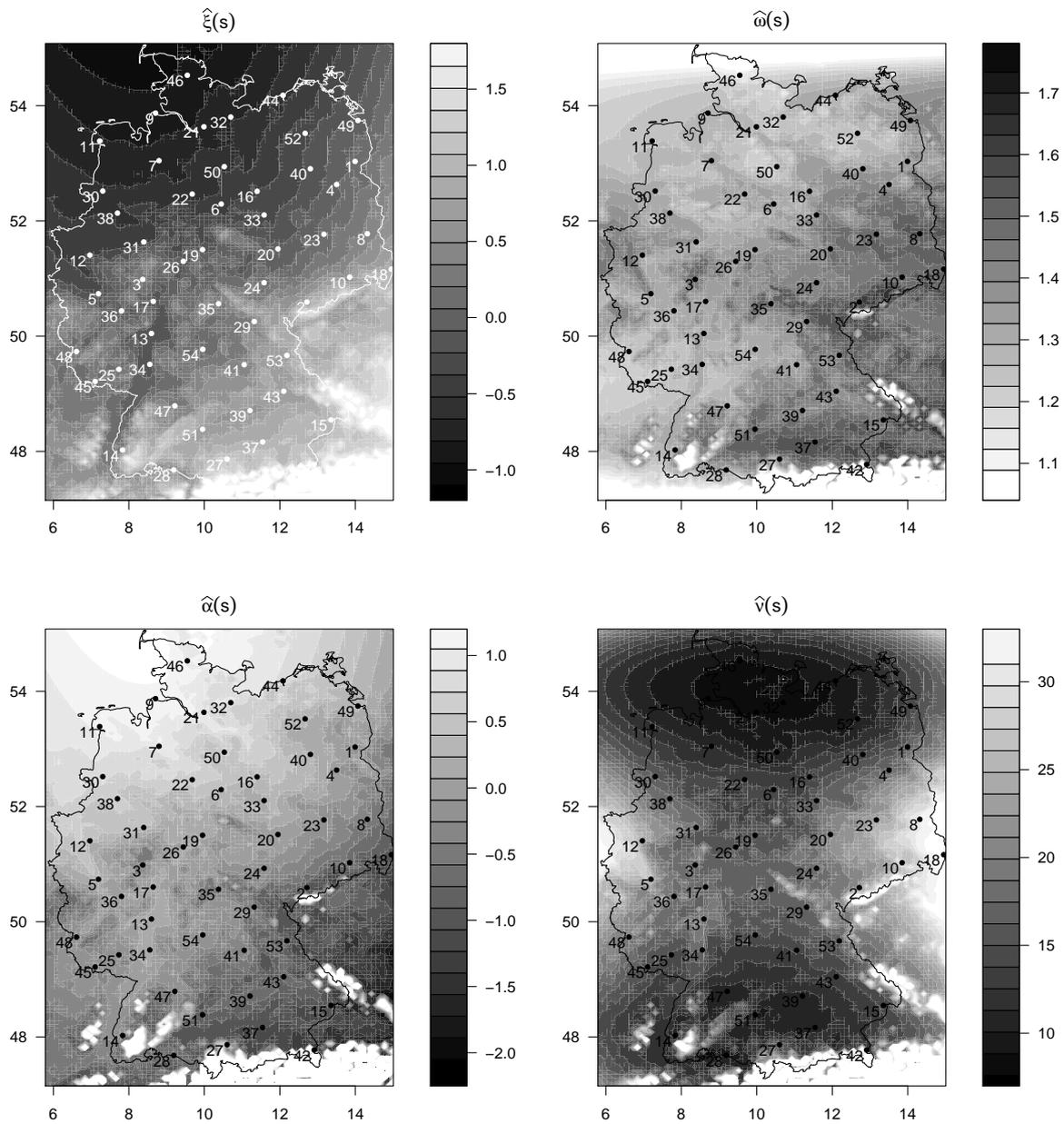
Figure 3.6.12: Level plots of the estimated polynomials $\widehat{\xi}(s)$, $\widehat{\omega}(s)$, $\widehat{\alpha}(s)$ and $\widehat{\nu}(s)$ with the locations of the observation stations indicated.
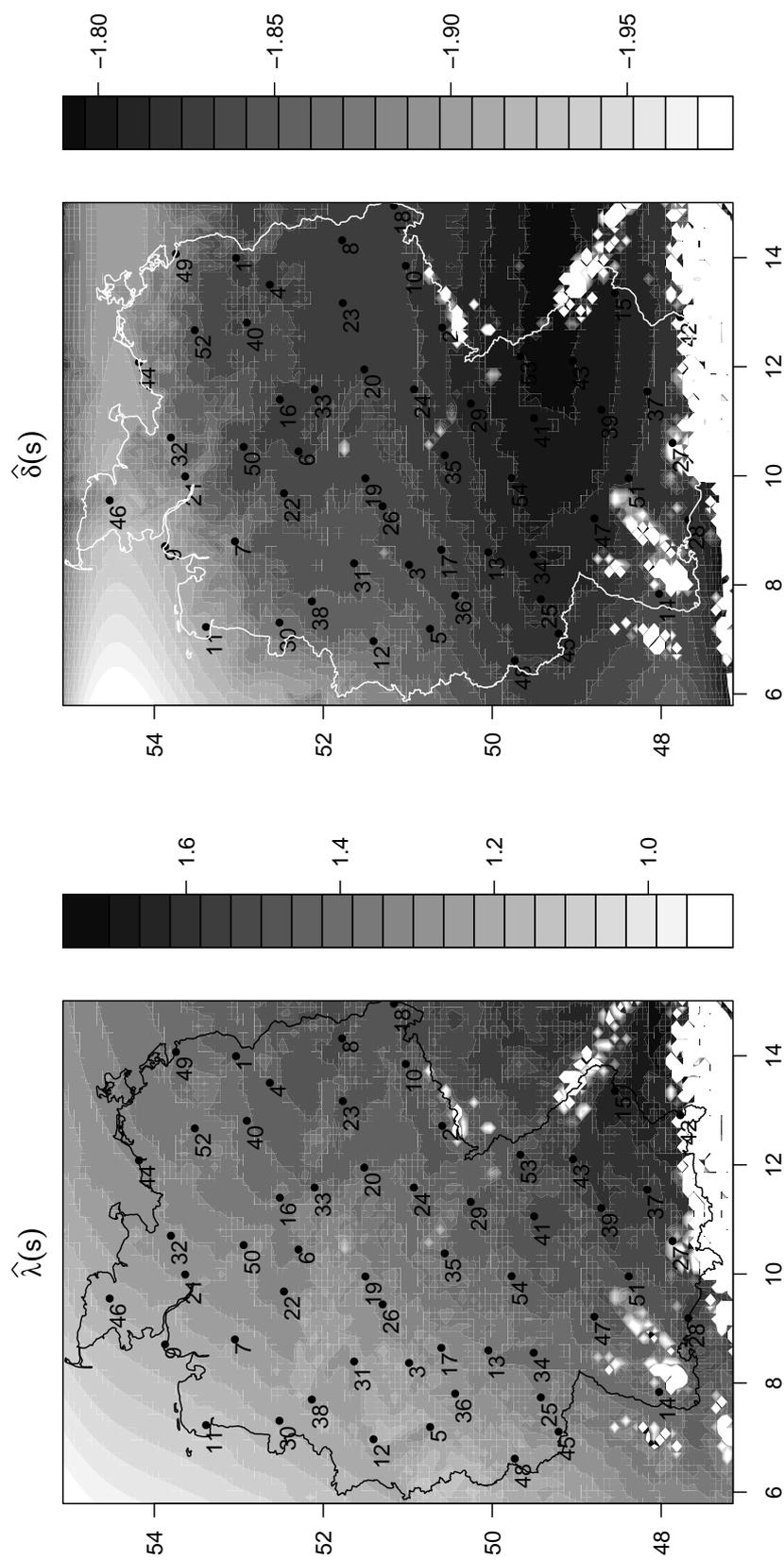
Figure 3.6.13: Level plots of the parameters $\widehat{\lambda}(s)$ and $\widehat{\delta}(s)$ calculated from the estimated polynomials $\widehat{\beta}_{\sin}(s)$, $\widehat{\beta}_{\cos}(s)$ with the locations of the observation stations indicated.

3.7   Transformation to copula data

This section aims to conclude our investigations on the marginals of our mean temperature data set. In the following chapters we are going to consider different kinds of models, which are based on the concept of pair-copula construction and intended to model the dependence structure among several variables. Since the building blocks of these models are copulae and the marginals are going to take a back seat, we transform our data to so called *copula data*, i.e. we transform our original time series $Y_1^s, \ldots, Y_N^s$, $s = 1, \ldots, d$, to $u_1^s, \ldots, u_N^s$, $s = 1, \ldots, d$, such that

$$u_1^s, \ldots, u_N^s \overset{\text{i.i.d.}}{\sim} \mathcal{U}(0,1) \quad \text{for all } s = 1, \ldots, d. \tag{3.7.1}$$

The transformation to copula data is carried out by means of the probability integral transform. All required tools were already presented previously. We saw that for all $s = 1, \ldots, d$ the errors $\varepsilon_1^s, \ldots, \varepsilon_N^s$ of our model are i.i.d. skew-t distributed with the parameters $\xi(s)$, $\omega(s)$, $\alpha(s)$ and $\nu(s)$, respectively. Hence the desired copula data is gained as

$$\widehat{u}_t^s = T_{\text{skew}} \left(\widehat{\varepsilon}_t^s \,|\, \widehat{\xi}(s), \widehat{\omega}(s), \widehat{\alpha}(s), \widehat{\nu}(s) \right), \quad t = 4, \ldots, N, s = 1, \ldots, d,$$

where $T_{\text{skew}} \left(\cdot \,|\, \xi, \omega, \alpha, \nu \right)$ is the cumulative distribution function of a skew-t distribution with location parameter ξ, scale parameter ω, shape parameter α and ν degrees of freedom.

To see if the transformed data series $\widehat{u}_4^s, \ldots, \widehat{u}_N^s$, $s = 1, \ldots, d$, fulfill the property given in (3.7.1), we consider the histograms given in Figure 3.7.1 and Figure 3.7.2. This consideration yields uniformity for most the 54 copula data series, only the transformed data of some of the observation stations like Aue, Cuxhaven and München yield obvious deviations from being uniformly distributed. This might be due to the fact that the overall number of parameters in the joint marginal model is reduced compared to the total number of parameters needed to model each margin separately, which may result in a loss of precision of the estimates $\widehat{\beta}_0(s)$, $\widehat{\beta}_{\sin}(s)$, $\widehat{\beta}_{\cos}(s)$, $\widehat{\gamma}_1(s)$, $\widehat{\gamma}_2(s)$, $\widehat{\gamma}_3(s)$, $\widehat{\xi}(s)$, $\widehat{\omega}(s)$, $\widehat{\alpha}(s)$, $\widehat{\nu}(s)$ compared to the respective estimates $\widehat{\beta}_0^s$, $\widehat{\beta}_{\sin}^s$, $\widehat{\beta}_{\cos}^s$, $\widehat{\gamma}_1^s$, $\widehat{\gamma}_2^s$, $\widehat{\gamma}_3^s$, $\widehat{\xi}^s$, $\widehat{\omega}^s$, $\widehat{\alpha}^s$, $\widehat{\nu}^s$ of the separate marginal models, for all $s = 1, \ldots, d$.
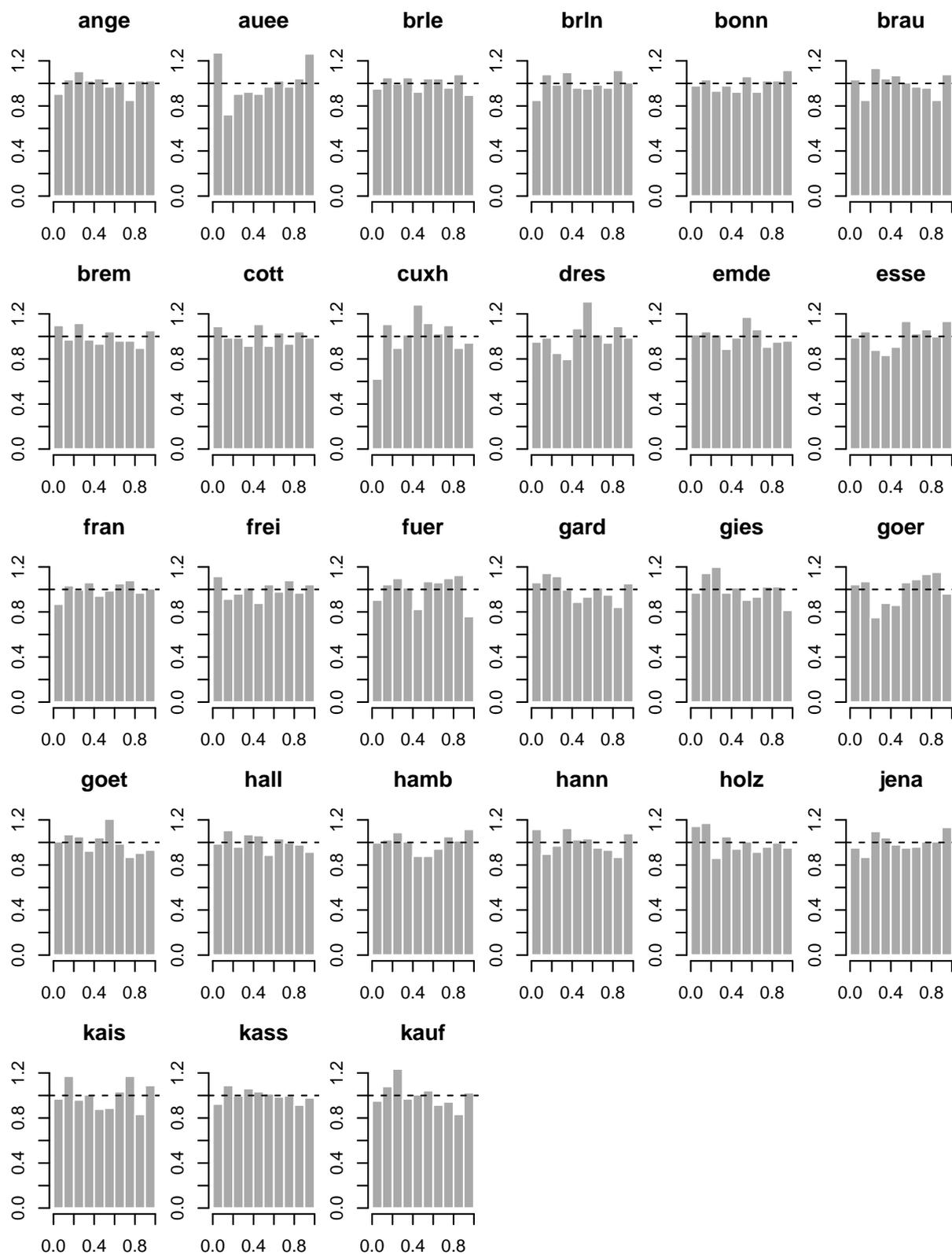
Figure 3.7.1: Histograms of the copula data obtained from the joint marginal model for the observation stations $s = 1, \ldots, 27$.
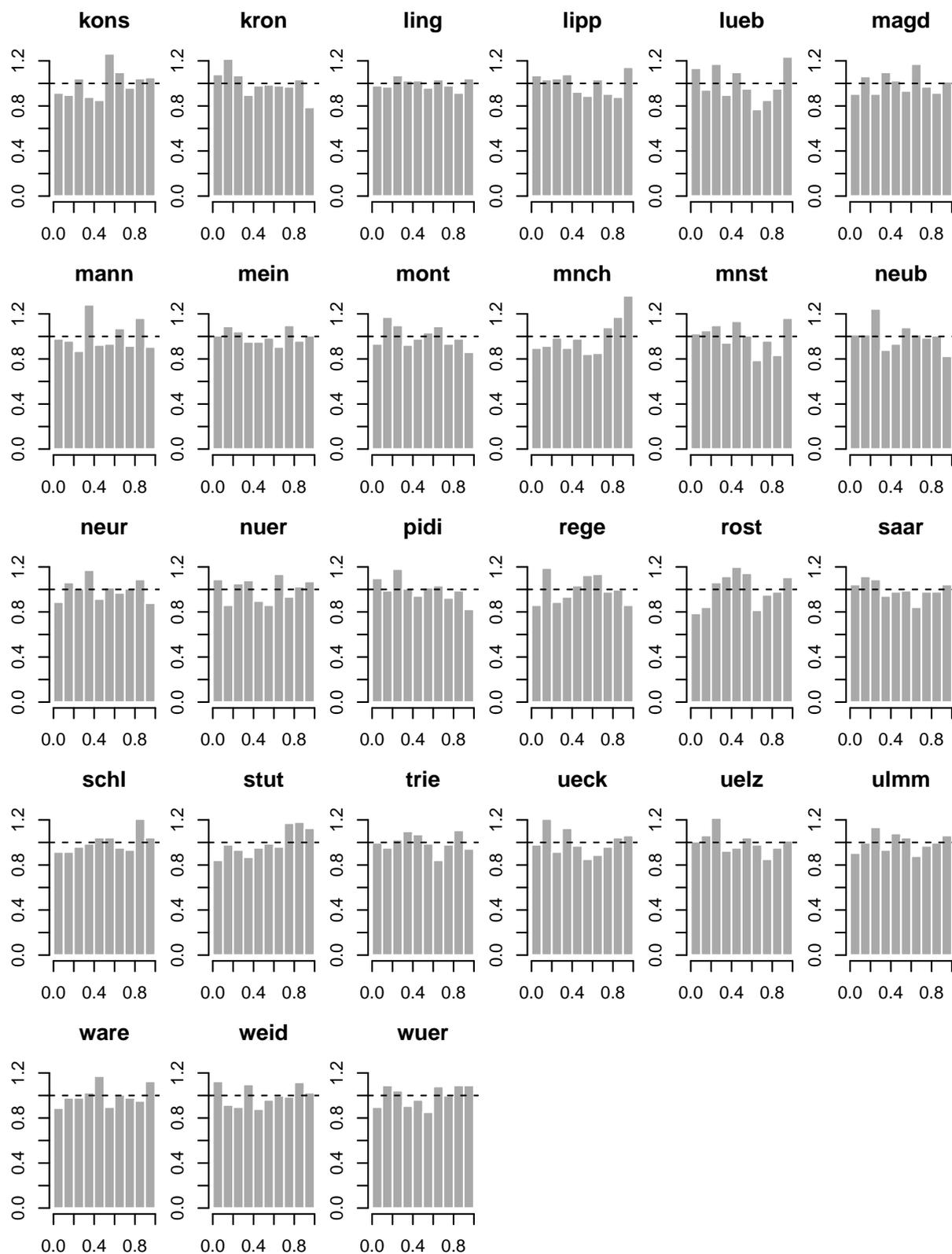
Figure 3.7.2: Histograms of the copula data obtained from the joint marginal model for observation stations $s = 28, \ldots, 54$.

Chapter 4

Regular vines

The aim of this chapter is to briefly introduce the so called regular vine (R-vine) copula models. This special case of a vine model, which is a graphical tool to model the dependence of random variables, is the base for our new models developed in the following chapters. Vine models in general were introduced by Bedford and Cooke (2001, 2002) and trace back to ideas of Joe (1996). Due to the paper of Aas et al. (2009) on pair-copula constructions, which we already introduced in Section 2.1.2, regular vines got into the focus of a broad field of research, since they systematize the composition of pair-copula constructions.

In the following sections we will define, amongst others, regular vines and two special cases hereof. Furthermore, we present a concept to organize R-vines in matrices for the purpose of computations and we discuss how to select R-vine distributions for a given set of multivariate data.

4.1   R-vines and their distributions

The subject matter of this section are R-vines and their corresponding distributions. As a reference for the following definitions in the setting of vines in general, we advert to Bedford and Cooke (2001, 2002). With regard to the fact that we focus on R-vines, we stick to Dißmann et al. (2013) in the following.

As already mentioned, regular vines are graphical models, with the aim to ease the depiction and the structuring of the dependence between random variables. They are graphical in the sense, that they are composed out of trees, which fulfill certain conditions. These conditions are summarized in the following formal definition of an R-vine.

Definition 4.1 (Regular vine (R-vine))**.** If

(i) $\mathcal{V} = (\mathcal{T}_1, \ldots, \mathcal{T}_{d-1})$,

(ii) $\mathcal{T}_1 = (\mathcal{V}_1, \mathcal{E}_1)$ is a tree with vertices $\mathcal{V}_1 = \{1, \ldots, d\}$ and edges $\mathcal{E}_1$,

(iii) $\mathcal{T}_i = (\mathcal{V}_i, \mathcal{E}_i)$ is a tree with vertices $\mathcal{V}_i = \mathcal{E}_{i-1}$ and edges $\mathcal{E}_i$, for all $i = 2, \ldots, d-1$ and

(iv) if it holds for every two vertices $a = \{a_1, a_2\}, b = \{b_1, b_2\} \in \mathcal{V}_i$ connected by an edge $e = \{a, b\} \in \mathcal{E}_i$, $i = 2, \ldots, d - 1$, that $\#(a \cap b) = 1$,

then we call $\mathcal{V}$ d-dimensional *regular vine* respectively *R-vine*.   ⌟

The definition of an R-vine by itself says nothing about any multivariate distribution of random variables, but its interpretation is as follows: The vertices in $\mathcal{V}_1$ of the first tree $\mathcal{T}_1$ represent d random variables $X_1, \ldots, X_d$, whose dependence is modeled by the R-vine. The edges $\mathcal{E}_i$, $i = 1, \ldots, d-1$, of the trees represent all bivariate (conditional) dependencies of these random variables, which are needed to completely specify the multivariate distribution of the random variables $X_1, \ldots, X_d$, according to the pair-copula construction principle. Moreover we stress that condition number (iv) of Definition 4.1, which is called *proximity condition*, is crucial in order to obtain a valid pair-copula construction. All this will become clearer in line with the example after the next definition.

Definition 4.2 (Complete union, conditioning set, conditioned set, constraint set)**.** Let $\mathcal{V}$ be an R-vine on d elements. For an arbitrary edge $e = \{a, b\} \in \mathcal{E}_i$, $i = 1, \ldots, d-1$, we define

(i) the *complete union* of the edge e

$$\mathcal{U}_e := \{v \in \mathcal{V}_1 : \exists e_j \in \mathcal{E}_j, j = 1, \ldots, i - 1, \text{ such that } v \in e_1 \in \ldots \in e_{i-1} \in e\},$$

(ii) the *conditioning set* associated with e

$$\mathcal{D}_e := \mathcal{U}_a \cap \mathcal{U}_b,$$

(iii) the *conditioned set(s)* associated with e

$$\mathcal{C}_e := \mathcal{C}_{e,a} \cup \mathcal{C}_{e,b}, \quad \text{where} \quad \mathcal{C}_{e,a} := \mathcal{U}_a \backslash \mathcal{D}_e \quad \text{and} \quad \mathcal{C}_{e,b} := \mathcal{U}_b \backslash \mathcal{D}_e,$$

and

(iv) the *constraint set* for the R-vine $\mathcal{V}$

$$\mathscr{C}(\mathcal{V}) := \left\{ \{\mathcal{C}_{e,a}, \mathcal{C}_{e,b} | \mathcal{D}_e\} : e = \{a, b\} \in \mathcal{E}_i, i = 1, \ldots, d - 1 \right\}.$$

⌟

In the following, whenever we depict R-vines as graphs, we will use a type of short notation which is based on Definition 4.2, for to label vertices and edges. The vertices $\mathcal{V}_1$ of tree one will be labeled with the numbers or names of the corresponding random variables, whereas the vertices $\mathcal{V}_i$ of the trees of higher order will inherit their label from their corresponding edge in $\mathcal{E}_{i-1}$. The edges $e = \{a, b\} \in \mathcal{E}_1$ of tree one are labeled according to the scheme $\mathcal{C}_{e,a}\mathcal{C}_{e,b}$ and the edges $e = \{a, b\} \in \mathcal{E}_i$, $i = 2, \ldots, d - 1$, are labeled $\mathcal{C}_{e,a}\mathcal{C}_{e,b} | \mathcal{D}_e$.
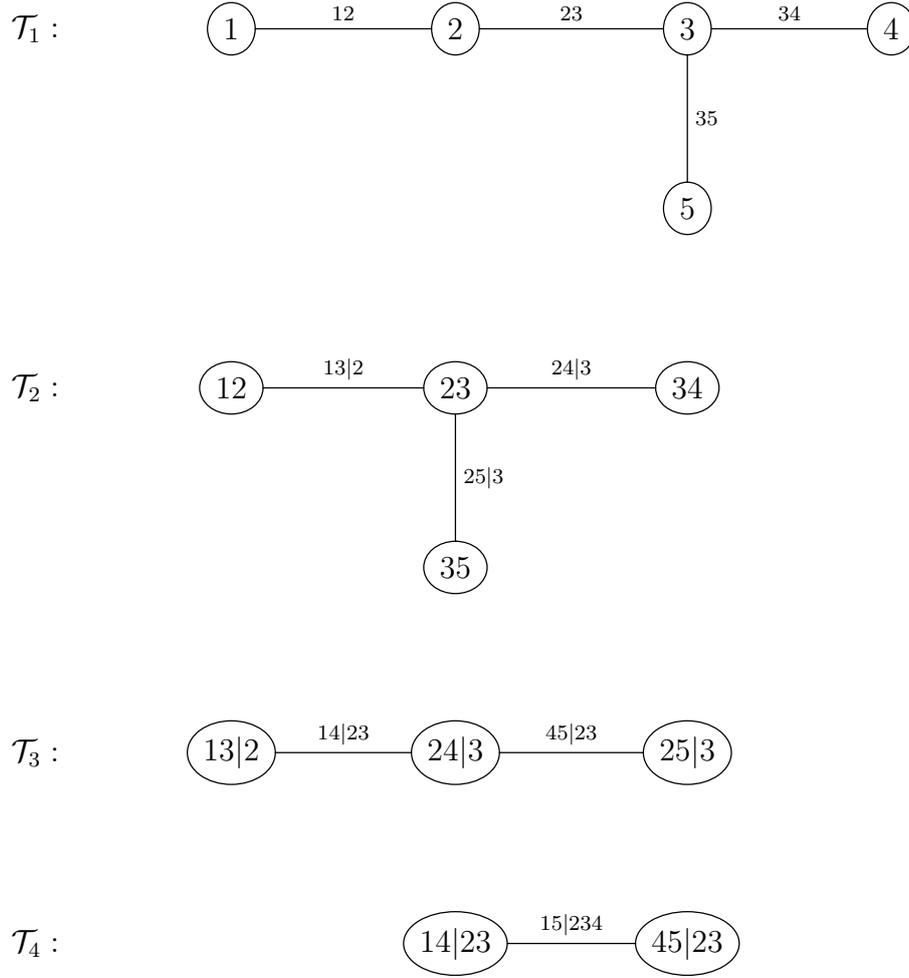
Figure 4.1.1: Example for an R-vine.

Example 4.3 (R-vine: $d = 5$)**.** An example for a five dimensional R-vine is given in Figure 4.1.1. The four trees of the vine $\mathcal{V} = (\mathcal{T}_1, \ldots, \mathcal{T}_4)$ are depicted, where the vertices and edges are named according to the scheme described above. From the figure we see, that

$$\mathcal{V}_1 = \{1, \ldots, 5\},$$

$$\mathcal{E}_1 = \mathcal{V}_2 = \big\{\{1,2\}, \{2,3\}, \{3,4\}, \{3,5\}\big\},$$

$$\mathcal{E}_2 = \mathcal{V}_3 = \Big\{\big\{\{1,2\}, \{2,3\}\big\}, \big\{\{2,3\}, \{3,4\}\big\}, \big\{\{2,3\}, \{3,5\}\big\}\Big\},$$

$$\mathcal{E}_3 = \mathcal{V}_4 = \bigg\{\Big\{\big\{\{1,2\}, \{2,3\}\big\}, \big\{\{2,3\}, \{3,4\}\big\}\Big\}, \Big\{\big\{\{2,3\}, \{3,4\}\big\}, \big\{\{2,3\}, \{3,5\}\big\}\Big\}\bigg\},$$

$$\mathcal{E}_4 = \Bigg\{\bigg\{\Big\{\big\{\{1,2\}, \{2,3\}\big\}, \big\{\{2,3\}, \{3,4\}\big\}\Big\}, \Big\{\big\{\{2,3\}, \{3,4\}\big\}, \big\{\{2,3\}, \{3,5\}\big\}\Big\}\bigg\}\Bigg\}.$$

It can be checked easily that the proximity condition is fulfilled. In order to illustrate the sets defined in Definition 4.2, let's consider $\mathcal{E}_3 = \{e', e^*\}$, where $e' \coloneqq \{a', b'\}$, $e^* \coloneqq \{a^*, b^*\}$,

$a' := \{\{1,2\},\{2,3\}\}$, $b' := a^* := \{\{2,3\},\{3,4\}\}$ and $b^* := \{\{2,3\},\{3,5\}\}$. The complete unions of the edges a', b', a^* and b^* are

$$\mathcal{U}_{a'} = \{1,2,3\}, \quad \mathcal{U}_{b'} = \mathcal{U}_{a^*} = \{2,3,4\} \quad \text{and} \quad \mathcal{U}_{b^*} = \{2,3,5\}.$$

The conditioning sets of the edges e' and e^* are

$$\mathcal{D}_{e'} = \mathcal{U}_{a'} \cap \mathcal{U}_{b'} = \{2,3\} = \mathcal{U}_{a^*} \cap \mathcal{U}_{b^*} = \mathcal{D}_{e^*}$$

and the conditioned sets are

$$\mathcal{C}_{e'} := \mathcal{C}_{e',a'} \cup \mathcal{C}_{e',b'} = \{1,4\} \quad \text{and} \quad \mathcal{C}_{e^*} := \mathcal{C}_{e^*,a^*} \cup \mathcal{C}_{e^*,b^*} = \{4,5\}.$$

⌐

By means of the next two definitions we introduce the already mentioned link of an R-vine to the multivariate distribution of some random variables $X_1, \ldots, X_d$.

Definition 4.4 (Regular vine copula specification)**.** Let

(i) $\vec{F} = (F_1, \ldots, F_d)$ be a vector of d univariate continuous invertible distribution functions,

(ii) $\mathscr{V}$ be a d-dimensional R-vine and

(iii) $\mathcal{B} := \{C_b : b \in \mathscr{C}(\mathscr{V})\} = \left\{ C_{\mathcal{C}_{e,a},\mathcal{C}_{e,b};\mathcal{D}_e} : e = \{a,b\} \in \mathcal{E}_i, i = 1, \ldots, d-1 \right\}$ be a set of bivariate copulae.

Then we call $(\vec{F}, \mathscr{V}, \mathcal{B})$ *regular vine copula specification.* ⌐

Definition 4.5 (Realization of a regular vine copula specification, R-vine distribution)**.** Let $\boldsymbol{F}$ be the joint distribution of the random vector $\boldsymbol{X} = (X_1, \ldots, X_d)^\top$. $\boldsymbol{F}$ is called *realization of a regular vine copula specification* $(\vec{F}, \mathscr{V}, \mathcal{B})$, if for all $e \in \mathcal{E}_i$, $i = 1, \ldots, d-1$, the bivariate copula $C_{\mathcal{C}_{e,a},\mathcal{C}_{e,b};\mathcal{D}_e}$ of $X_{\mathcal{C}_{e,a}}$ and $X_{\mathcal{C}_{e,b}}$ given $\boldsymbol{X}_{\mathcal{D}_e}$ lies in $\mathcal{B}$, and if F_i is the marginal distribution of X_i for all $i = 1, \ldots, d-1$. In this case $\boldsymbol{F}$ is also called *R-vine distribution.* ⌐

Now that a connection between R-vines and corresponding R-vine distributions is established we are interested in how to depict the pair-copula construction respectively the R-vine distribution density in terms of an R-vine copula specification. Moreover, the expression for the density given in the following theorem is unique.

Theorem 4.6 (Uniqueness of R-vine distribution)**.** *Let $\mathscr{V}$ be a d-dimensional R-vine and $(\vec{F}, \mathscr{V}, \mathcal{B})$ a corresponding R-vine copula specification. Then there exists a unique R-vine distribution $\boldsymbol{F}$, that realizes this R-vine copula specification. For the density f corresponding to $\boldsymbol{F}$ we get*

$$f(\boldsymbol{x}) = \prod_{k=1}^{d} f_k(x_k) \prod_{j=1}^{d-1} \prod_{e \in \mathcal{E}_j} c_{\mathcal{C}_{e,a},\mathcal{C}_{e,b};\mathcal{D}_e} \left(F_{\mathcal{C}_{e,a}|\mathcal{D}_e}(x_{\mathcal{C}_{e,a}} \mid \boldsymbol{x}_{\mathcal{D}_e}), F_{\mathcal{C}_{e,b}|\mathcal{D}_e}(x_{\mathcal{C}_{e,b}} \mid \boldsymbol{x}_{\mathcal{D}_e}) \right), \quad (4.1.1)$$

where $\boldsymbol{x} = (x_1, \ldots, x_d)^\top$, f_k, $k = 1, \ldots, d$, are the marginal densities corresponding to F_k, $k = 1, \ldots, d$, and $c_{\mathcal{C}_{e,a},\mathcal{C}_{e,b};\mathcal{D}_e}(\cdot,\cdot)$ are the densities corresponding to the bivariate copulae $C_{\mathcal{C}_{e,a},\mathcal{C}_{e,b};\mathcal{D}_e} \in \mathcal{B}$.

To evaluate such a density, we need to calculate the so called *transformed variables* $F_{\mathcal{C}_{e,a}|\mathcal{D}_e}(x_{\mathcal{C}_{e,a}} \mid \boldsymbol{x}_{\mathcal{D}_e})$, $F_{\mathcal{C}_{e,b}|\mathcal{D}_e}(x_{\mathcal{C}_{e,b}} \mid \boldsymbol{x}_{\mathcal{D}_e})$ of Equation (4.1.1). We do this in analogy to the formula stated in (2.1.8).

Lemma 4.7 (Recursive calculation of transformed variables). *Let $\mathcal{V}$ be a d-dimensional R-vine and $a = \{a_1, a_2\}, b = \{b_1, b_2\} \in \mathcal{V}_i$ be arbitrary vertices connected by an edge $e = \{a, b\} \in \mathcal{E}_i$, $i = 2, \ldots, d-1$. Then we obtain*

$$F_{\mathcal{C}_{e,a}|\mathcal{D}_e}(x_{\mathcal{C}_{e,a}} \mid \boldsymbol{x}_{\mathcal{D}_e}) = \frac{\partial\, C_{\mathcal{C}_a;\mathcal{D}_a}\left(F_{\mathcal{C}_{a,a_1}|\mathcal{D}_a}(x_{\mathcal{C}_{a,a_1}} \mid \boldsymbol{x}_{\mathcal{D}_a}), F_{\mathcal{C}_{a,a_2}|\mathcal{D}_a}(x_{\mathcal{C}_{a,a_2}} \mid \boldsymbol{x}_{\mathcal{D}_a})\right)}{\partial F_{\mathcal{C}_{a,a_2}|\mathcal{D}_a}(x_{\mathcal{C}_{a,a_2}} \mid \boldsymbol{x}_{\mathcal{D}_a})}$$

for $i = 3, \ldots, d-1$ and

$$F_{\mathcal{C}_{e,a}|\mathcal{D}_e}(x_{\mathcal{C}_{e,a}} \mid \boldsymbol{x}_{\mathcal{D}_e}) = \frac{\partial\, C_{\mathcal{C}_a}\left(F_{\mathcal{C}_{a,a_1}}(x_{\mathcal{C}_{a,a_1}}), F_{\mathcal{C}_{a,a_2}}(x_{\mathcal{C}_{a,a_2}})\right)}{\partial F_{\mathcal{C}_{a,a_2}}(x_{\mathcal{C}_{a,a_2}})}$$

for $i = 2$. The respective expressions for b are obtained analogously.

4.2 C- and D-vines

Now we are going to consider two special cases of R-vines, therefore we refer to Kurowicka and Cooke (2003) and Kurowicka and Cooke (2005) for the definitions and to Aas et al. (2009) for the density formulas.

Definition 4.8 (Canonical vine (C-vine)). A regular vine (R-vine) $\mathcal{V} = (\mathcal{T}_1, \ldots, \mathcal{T}_{d-1})$ is called d-dimensional *canonical vine (C-vine)* if each tree $\mathcal{T}_i$, $i = 1, \ldots, d-1$, has one vertex $v \in \mathcal{V}_i$ of maximal degree $\deg(v) = d - i$. These vertices of maximal degree are called *root vertices* or *root nodes*. ⌋

From the definition of the canonical vine it follows directly, that all vertices which are no root vertices have a degree of one, i.e. each tree has a star structure. The whole structure of a C-vine is determined uniquely, if one fixes an ordering of the root vertices. A simple, four dimensional example of a C-vine is depicted in Figure 4.2.1. Here the root vertices are 1, $\{1, 2\}$ and $\{\{1, 2\}, \{1, 3\}\}$, respectively.

Since a C-vine is a special case of an R-vine, we can use Formula (4.1.1) to get the density for a general d-dimensional C-vine. Due to the special structure of C-vines the density exhibits a special structure as well. It holds

$$f(\boldsymbol{x}) = \prod_{k=1}^{d} f_k(x_k) \prod_{j=1}^{d-1} \prod_{i=1}^{d-j} c_{j,j+i;\mathcal{I}_j}\left(F_{j|\mathcal{I}_j}(x_j \mid \boldsymbol{x}_{\mathcal{I}_j}), F_{j+i|\mathcal{I}_j}(x_{j+i} \mid \boldsymbol{x}_{\mathcal{I}_j})\right),$$

where $\boldsymbol{x} = (x_1, \ldots, x_d)^\top$, $\mathcal{I}_j = \{1, \ldots, j-1\}$ and f_k, $k = 1, \ldots, d$, are the marginal densities corresponding to F_k, $k = 1, \ldots, d$.

Now we advance to the second special case of an R-vine.

Definition 4.9 (Drawable vine (D-vine)). A regular vine (R-vine) $\mathcal{V} = (\mathcal{T}_1, \ldots, \mathcal{T}_{d-1})$ is called d-dimensional *drawable vine (D-vine)* if it holds for the vertices $v \in \mathcal{V}_1$ of the first tree $\mathcal{T}_1$ that their degree is only one or two $(\deg(v) \leq 2)$. ⌋
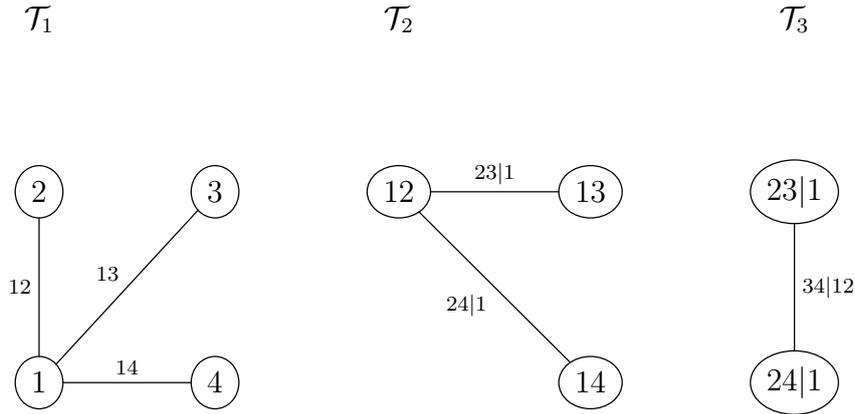
Figure 4.2.1: Example for a four dimensional C-vine.

From the definition of the D-vine follows, that all trees of a D-vine are paths. In the case of a D-vine the whole structure is determined uniquely, by fixing the structure of tree one. Again a four dimensional example, this time of a D-vine, is depicted in Figure 4.2.2. Here the structure is determined through the path $\mathcal{T}_1 = (\mathcal{V}_1, \mathcal{E}_1)$ with edge set $\mathcal{E}_1 = \{\{1,2\}, \{2,3\}, \{3,4\}\}$.
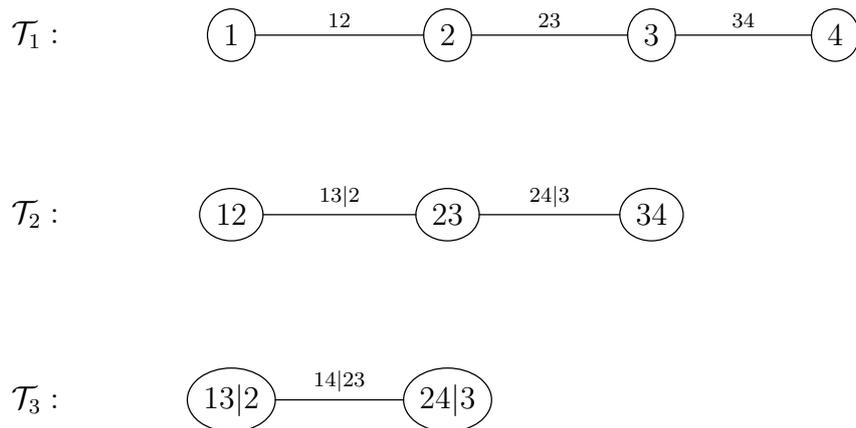
Figure 4.2.2: Example for a four dimensional D-vine.

Application of Formula (4.1.1) yields now the density for a general d-dimensional D-vine, which exhibits again a special structure. It holds

$$f(\boldsymbol{x}) = \prod_{k=1}^{d} f_k(x_k) \prod_{j=1}^{d-1} \prod_{i=1}^{d-j} c_{i,i+j;\mathcal{I}_{ij}} \left(F_{i|\mathcal{I}_{ij}}(x_i \mid \boldsymbol{x}_{\mathcal{I}_{ij}}), F_{i+j|\mathcal{I}_{ij}}(x_{i+j} \mid \boldsymbol{x}_{\mathcal{I}_{ij}}) \right),$$

where $\boldsymbol{x} = (x_1, \ldots, x_d)^{\top}$, $\mathcal{I}_{ij} = \{i+1, \ldots, i+j-1\}$ and f_k, $k = 1, \ldots, d$, are the marginal densities corresponding to F_k, $k = 1, \ldots, d$.

For a software package which is designed to work with C- and D-vines and an extensive guideline on it, we advert to Brechmann and Schepsmeier (2013).

4.3 Regular vine matrices

In order to enable efficient algorithms for statistical inference on R-vines, it is crucial to be able to access all necessary information on an R-vine in an efficient manner. Therefore we present a storage scheme for R-vines according to Dißmann (2010) and Dißmann et al. (2013). Here the main idea is to implicitly store the structure of R-vines in lower triangular matrices instead of storing the whole set of nested trees.

In the following we give the definition of an R-vine matrix, which captures the structure of R-vines and ensures that the proximity condition is fulfilled. Moreover we state two properties of these R-vine matrices.

Definition 4.10 (R-vine matrix). Let $M = (m_{i,j})_{i,j=1,\ldots,d} \in \mathbb{R}^{d \times d}$ be a lower triangular matrix. For $j = 1, \ldots, d-1$ let

$$\mathscr{A}(M_{.,j}) := \left\{ \{m_{j,j} | \widetilde{\mathcal{D}}\} : \widetilde{\mathcal{D}} = \{m_{k,j}, \ldots, m_{d,j}\}, k = j+1, \ldots, d \right\},$$

$$\mathscr{B}(M_{.,j}) := \left\{ \{m_{k,j} | \widetilde{\mathcal{D}}\} : \widetilde{\mathcal{D}} = \{m_{j,j}\} \cup \{m_{k+1,j}, \ldots, m_{d,j}\}, k = j+1, \ldots, d \right\}.$$

M is called *R-vine matrix* if for all $i = 1, \ldots, d-2$, $k = i+1, \ldots, d-1$, there is a $j \in \{i+1, \ldots, d-1\}$ such that

$$\mathscr{Z}_{k,i}(M) := \{m_{k,i} | \{m_{k+1,i}, \ldots, m_{d,i}\}\} \in \mathscr{A}(M_{.,j}) \quad \text{or} \quad \mathscr{Z}_{k,i}(M) \in \mathscr{B}(M_{.,j}). \tag{4.3.1}$$

⌟

Proposition 4.11 (Properties of R-vine matrices). *Let $M = (m_{i,j})_{i,j=1,\ldots,d} \in \mathbb{R}^{d \times d}$ be an R-vine matrix. Then*

(i) $\{m_{j,j}, \ldots, m_{d,j}\} \subset \{m_{i,i}, \ldots, m_{d,i}\}$ *for $1 \leq i < j \leq d$ and*

(ii) $m_{j,j} \notin \{m_{j+1,j+1}, \ldots, m_{d,j+1}\}$ *for $j = 1, \ldots, d-1$.*

We can follow directly, that all entries in a column of an R-vine matrix differ. Proposition 4.11 (i) says that all the entries of a column of an R-vine matrix are also included in each column with lower index, whereas (ii) yields, that each diagonal entry of an R-vine matrix is new compared to the entries of the columns with higher index.

At first glance the definition of an R-vine matrix is rather counterintuitive. To illustrate Definition 4.10 and the subsequent proposition, we have a look at the following five dimensional example.

Example 4.12 (R-vine matrices: Example 4.3 continued). We consider the two triangular matrices

$$M^* = \begin{pmatrix} 1 & & & & \\ 5 & 4 & & & \\ 4 & 5 & 2 & & \\ 3 & 2 & 5 & 3 & \\ 2 & 3 & 3 & 5 & 5 \end{pmatrix} \quad \text{and} \quad M' = \begin{pmatrix} 5 & & & & \\ 1 & 4 & & & \\ 4 & 1 & 3 & & \\ 2 & 2 & 1 & 2 & \\ 3 & 3 & 2 & 1 & 1 \end{pmatrix}.$$

In order to see that M^* is an R-vine matrix, we examine the sets $\mathscr{A}(M^*_{.,j})$, $\mathscr{B}(M^*_{.,j})$, $j = 1, \ldots, 4$ and $\mathscr{Z}_{k,i}(M^*)$, $i = 1, \ldots, 4$, $k = i+1, \ldots, 4$, given in Table 4.1, Table 4.2 and

Table 4.3, respectively. We easily check that $\mathscr{Z}_{2,1}(M^*) \in \mathscr{B}(M^*_{.,2})$, $\mathscr{Z}_{3,1}(M^*) \in \mathscr{A}(M^*_{.,2})$, $\mathscr{Z}_{4,1}(M^*) \in \mathscr{B}(M^*_{.,3})$, $\mathscr{Z}_{3,2}(M^*) \in \mathscr{B}(M^*_{.,3})$, $\mathscr{Z}_{4,2}(M^*) \in \mathscr{A}(M^*_{.,3})$ and $\mathscr{Z}_{4,3}(M^*) \in \mathscr{B}(M^*_{.,4})$. This yields that M^* is an R-vine matrix. The proof, that M' is an R-vine matrix as well, is left to the reader. Furthermore both matrices apparently fulfill the properties (i) and (ii) from Proposition 4.11.

	$\mathscr{A}(M^*_{.,1})$	$\mathscr{A}(M^*_{.,2})$	$\mathscr{A}(M^*_{.,3})$	$\mathscr{A}(M^*_{.,4})$
$k=2$	$\{1|\{2,3,4,5\}\}$			
$k=3$	$\{1|\{2,3,4\}\}$	$\{4|\{2,3,5\}\}$		
$k=4$	$\{1|\{2,3\}\}$	$\{4|\{2,3\}\}$	$\{2|\{3,5\}\}$	
$k=5$	$\{1|\{2\}\}$	$\{4|\{3\}\}$	$\{2|\{3\}\}$	$\{3|\{5\}\}$

Table 4.1: $\mathscr{A}(M^*_{.,j})$ for $j = 1,\ldots,4$.

	$\mathscr{B}(M^*_{.,1})$	$\mathscr{B}(M^*_{.,2})$	$\mathscr{B}(M^*_{.,3})$	$\mathscr{B}(M^*_{.,4})$
$k=2$	$\{5|\{1,2,3,4\}\}$			
$k=3$	$\{4|\{1,2,3\}\}$	$\{5|\{2,3,4\}\}$		
$k=4$	$\{3|\{1,2\}\}$	$\{2|\{3,4\}\}$	$\{5|\{2,3\}\}$	
$k=5$	$\{2|\{1\}\}$	$\{3|\{4\}\}$	$\{3|\{2\}\}$	$\{5|\{3\}\}$

Table 4.2: $\mathscr{B}(M^*_{.,j})$ for $j = 1,\ldots,4$.

$\mathscr{Z}_{k,i}(M^*)$	$i=1$	$i=2$	$i=3$
$k=2$	$\{5|\{2,3,4\}\}$		
$k=3$	$\{4|\{2,3\}\}$	$\{5|\{2,3\}\}$	
$k=4$	$\{3|\{2\}\}$	$\{2|\{3\}\}$	$\{5|\{3\}\}$

Table 4.3: $\mathscr{Z}_{k,i}(M^*)$ for $i = 1,\ldots,4$, $k = i+1,\ldots,4$.

Since the structure of an R-vine $\mathscr{V}$ is, amongst others, specified by a constraint set $\mathscr{C}(\mathscr{V})$, as we saw in Section 4.1, an analogous set for lower triangular matrices is considered.

Definition 4.13 (Matrix constraint set). Let $M = (m_{i,j})_{i,j=1,\ldots,d} \in \mathbb{R}^{d\times d}$ be a lower triangular matrix. We define

(i) the *j-th (matrix) constraint set* for M

$$\mathscr{C}(M_{.,j}) := \{\{m_{j,j}, m_{k,j}|\mathcal{D}\} : \mathcal{D} = \{m_{k+1,j},\ldots,m_{d,j}\}, k = j+1,\ldots,d\},$$

for $j = 1,\ldots,d-1$ and $\mathcal{D} := \emptyset$ for $k = d$, and

(ii) the *(matrix) constraint set* for M

$$\mathscr{C}(M) := \mathscr{C}(M_{.,1}) \cup \ldots \cup \mathscr{C}(M_{.,d-1}).$$

For the elements $\{m_{j,j}, m_{k,j} | \mathcal{D}\}$, $j = 1, \ldots, d-1$, $k = j+1, \ldots, d$ of the constraint set $\mathscr{C}(M)$ we call

(iii) $\{m_{j,j}, m_{k,j}\}$ *(matrix) conditioned set* and

(iv) $\mathcal{D}$ *(matrix) conditioning set.* ⌟

Dißmann (2010) presents two algorithms, that allow to transform between R-vines $\mathscr{V}$ and R-vine matrices $M = (m_{i,j})_{i,j=1,\ldots,d} \in \mathbb{R}^{d \times d}$. Moreover it is proven that for an R-vine $\mathscr{V}$ there exists an R-vine matrix M and vice versa, such that the constraint set $\mathscr{C}(\mathscr{V})$ of the R-vine equals the (matrix) constraint set $\mathscr{C}(M)$ of the R-vine matrix.

Until now, nothing was said about how to extract the information stored in an R-vine matrix. Now we know that the solution for this lies in Definition 4.13. For an R-vine matrix M, the elements of the matrix constraint set $\mathscr{C}(M)$ represent all the bivariate building blocks of the R-vine, respectively the pair-copula construction. To get the desired information, we just have to treat the matrix conditioned sets and the matrix conditioning sets in the same way as we use the corresponding sets of a constraint set $\mathscr{C}(\mathscr{V})$ for an R-vine $\mathscr{V}$. An Example follows.

Example 4.14 (R-vine: Example 4.12 continued)**.** Let's reconsider the two triangular matrices M^* and M' from Example 4.12. The components of $\mathscr{C}(M^*)$ are given by

$$\mathscr{C}(M^*_{.,1}) = \{\{1,5|2,3,4\}, \{1,4|2,3\}, \{1,3|2\}, \{1,2|\emptyset\}\},$$
$$\mathscr{C}(M^*_{.,2}) = \{\{4,5|2,3\}, \{2,4|3\}, \{3,4|\emptyset\}\},$$
$$\mathscr{C}(M^*_{.,3}) = \{\{2,5|3\}, \{2,3|\emptyset\}\},$$
$$\mathscr{C}(M^*_{.,4}) = \{\{3,5|\emptyset\}\}.$$

The unconditioned elements of $\mathscr{C}(M^*)$ yield the first tree of the R-vine, the elements with one conditioning variable the second, and so on. We obtain the R-vine of Figure 4.1.1 which already was investigated in Example 4.3. For M' we get

$$\mathscr{C}(M'_{.,1}) = \{\{1,5|2,3,4\}, \{4,5|2,3\}, \{2,5|3\}, \{3,5|\emptyset\}\},$$
$$\mathscr{C}(M'_{.,2}) = \{\{1,4|2,3\}, \{2,4|3\}, \{3,4|\emptyset\}\},$$
$$\mathscr{C}(M'_{.,3}) = \{\{1,3|2\}, \{2,3|\emptyset\}\},$$
$$\mathscr{C}(M'_{.,4}) = \{\{1,2|\emptyset\}\}.$$

This yields that $\mathscr{C}(M^*) = \mathscr{C}(M')$, i.e. M^* and M' represent the same R-vine. We conclude that in general there is no unique R-vine matrix representation of an R-vine. ⌟

For reasons of convenience information about copula families and the parameters corresponding to the bivariate copulae of an R-vine are also stored in $d \times d$ matrices. Whereas the parameters are real numbers, it is convenient to code the copula family types with natural numbers.

Definition 4.15 (R-vine family and parameter matrices)**.** Let $M = (m_{i,j})_{i,j=1,\ldots,d} \in \mathbb{R}^{d \times d}$ be an R-vine matrix. Then we store information about the copula family of the bivariate copulae $C_{m_{j,j},m_{i,j};m_{i+1,j},\ldots,m_{d,j}}$, $j = 1, \ldots, d-1$, $i = j+1, \ldots, d-1$ in the entry $t_{i,j}$ of the lower triangular matrix $T = (t_{i,j})_{i,j=1,\ldots,d}$, which we call *R-vine family matrix*. Analogously the parameters are stored in the entries $\theta_{i,j}^l$ of the matrices $P^l = (\theta_{i,j}^l)_{i,j=1,\ldots,d} \in \mathbb{R}^d$, where $1 \le l \le l_{\max}$ and $l_{\max}$ is the maximum number of parameters needed to parametrize the bivariate copulae which occur. The matrices P^l are called *R-vine parameter matrices*. ⌟

Example 4.16 (R-vine: Families and parameters)**.** Let's again reconsider the R-Vine matrix $M^* \in \mathbb{R}^{5 \times 5}$ of Example 4.12. For $l_{\max} = 2$ the R-vine family matrix and the R-vine parameter matrices could be given by

$$
T = \begin{pmatrix} \Pi & & & \\ G & J & & \\ C & \Phi & F & \\ t & t & \Phi & t \end{pmatrix}, \quad
P^1 = \begin{pmatrix} 0 & & & \\ 3.5 & 5.4 & & \\ 1.1 & -0.3 & -2.1 & \\ 0.7 & 0.6 & 0.8 & 0.7 \end{pmatrix}, \quad
P^2 = \begin{pmatrix} 0 & & & \\ 0 & 0 & & \\ 0 & 0 & 0 & \\ 3.5 & 5.3 & 0 & 9.1 \end{pmatrix},
$$

where the coding of the families used in T is given in Table 4.4.

no.	code	copula family name
0	Π	Independence
1	Φ	Gaussian
2	t	Student-t
3	C	Clayton
4	G	Gumbel
5	F	Frank
6	J	Joe

Table 4.4: Coding for copula families.

Let us consider the bivariate copula $C_{m_{j,j}^*,m_{i,j}^*;m_{i+1,j}^*,\ldots,m_{5,j}^*}\left(\cdot,\cdot;\theta_{i,j}^1,\theta_{i,j}^2,t_{i,j}\right)$ for $i = 3$ and $j = 1$. The matrix conditioned set equals $\{m_{1,1}^*, m_{3,1}^*\} = \{1,4\}$ and the matrix conditioning set is given by $\{m_{4,1}^*, m_{5,1}^*\} = \{2,3\}$, i.e. we consider $C_{1,4;2,3}\left(\cdot,\cdot;\theta_{3,1}^1,\theta_{3,1}^2,t_{3,1}\right)$. The copula family is given by $t_{3,1} = 4$, which means that the copula is a Gumbel copula. The Gumbel copula is one-parametric, that's the reason why $\theta_{3,1}^2$ is set to zero. For the first parameter we have $\theta_{3,1}^1 = 3.5$. ⌟

Corollary 4.17 (Density of an R-vine distribution based on R-vine matrix)**.** *Let $M = (m_{i,j})_{i,j=1,\ldots,d} \in \mathbb{R}^{d \times d}$ be an R-vine matrix corresponding to an R-vine $\mathcal{V}$. Let moreover $T = (t_{i,j})_{i,j=1,\ldots,d}$ be the R-vine family matrix and $P^l = (\theta_{i,j}^l)_{i,j=1,\ldots,d} \in \mathbb{R}^{d \times d}$, $1 \le l \le l_{\max}$, be the R-vine parameter matrices corresponding to M. Then it holds for the R-vine distribution density f, that*

$$
f(\boldsymbol{x}) = \prod_{k=1}^{d} f_k(x_k) \prod_{j=1}^{d-1} \prod_{i=j+1}^{d} c_{m_{j,j},m_{i,j};m_{i+1,j},\ldots,m_{d,j}}\left(\widetilde{u}_{j,j}, \widetilde{u}_{i,j}; \boldsymbol{\theta}_{i,j}, t_{i,j}\right), \tag{4.3.2}
$$

where $\boldsymbol{x} = (x_1, \ldots, x_d)^\top$, f_k, $k = 1, \ldots, d$, are the marginal densities of the R-vine distribution $\boldsymbol{F}$,

$$\widetilde{u}_{j,j} := F_{m_{j,j} \mid m_{i+1,j}, \ldots, m_{d,j}}(x_{m_{j,j}} \mid \boldsymbol{x}_{m_{i+1,j}, \ldots, m_{d,j}}),$$
$$\widetilde{u}_{i,j} := F_{m_{i,j} \mid m_{i+1,j}, \ldots, m_{d,j}}(x_{m_{i,j}} \mid \boldsymbol{x}_{m_{i+1,j}, \ldots, m_{d,j}}),$$
$$\boldsymbol{\theta}_{i,j} := (\theta_{i,j}^1, \theta_{i,j}^2, \ldots)^\top,$$

and $c_{m_{j,j}, m_{i,j}; m_{i+1,j}, \ldots, m_{d,j}}(\widetilde{u}_{j,j}, \widetilde{u}_{i,j}; \boldsymbol{\theta}_{i,j}, t_{i,j})$ are the densities of bivariate copulae with family $t_{i,j}$ and parameters $\boldsymbol{\theta}_{i,j}$.

Of course for $i = d$ the transformed variables simplify to $\widetilde{u}_{j,j} = F_{m_{j,j}}(x_{m_{j,j}})$ and $\widetilde{u}_{d,j} = F_{m_{d,j}}(x_{m_{d,j}})$. In general the transformed variables are again computed in analogy to Lemma 4.7.

For the purpose of inference, we want to calculate the log-likelihood of an R-vine $\mathscr{V}$ with given R-vine matrix $M = (m_{i,j})_{i,j=1,\ldots,d} \in \mathbb{R}^{d \times d}$. This is easily done, since we already have got an expression for the R-vine density based on the information stored in an R-vine matrix and the corresponding R-vine family and R-vine parameter matrices. With Equation (4.3.2) we get

$$\ell(\boldsymbol{\theta} \mid \boldsymbol{x}) = \sum_{k=1}^{d} \ln f_k(x_k) + \sum_{j=1}^{d-1} \sum_{i=j+1}^{d} \ln c_{m_{j,j}, m_{i,j}; m_{i+1,j}, \ldots, m_{d,j}}(\widetilde{u}_{j,j}, \widetilde{u}_{i,j}; \boldsymbol{\theta}_{i,j}, t_{i,j}) \qquad (4.3.3)$$

for the log-likelihood.

4.4 Selection of R-vine distributions

Now that we are provided with the necessary knowledge for parameter estimation in the context of R-vines, it remains to answer the question, how to select an appropriate R-vine copula specification, i.e. the parametric pair-copula construction, for a given data set. To this end we keep on following Dißmann et al. (2013).

Let us first of all emphasize that the R-vine model selection procedure consists of three main steps. The first step is the choice of the R-vine tree structure. This is equivalent to the determination of the corresponding R-vine matrix. The second step deals with the selection of the (parametric) copula families for the bivariate pairs, which were chosen in the previous step. The estimation of the model parameters is the final step needed to complete the selection procedure. Whereas all work for the third step is already done, we will subsequently focus on the first two steps.

Usually it is the goal of model selection to choose the model with the highest likelihood. But due to a high number of different models that have to be taken into consideration and high computational effort, this is intractable for R-vines in higher dimensions, as they, for instance, occur in our applications. Consequently Dißmann et al. (2013) developed a heuristic, *sequential method* for the purpose of R-vine structure selection. They called the method "sequential", since the R-vine structure is chosen tree-wise, as we will see next.

For more details on the infeasibility of a "full" model selection procedure and a substantial reasoning for the adequacy of the sequential approach see Dißmann et al. (2013).

Now we will present the sequential method, which interweaves the three modeling steps outlined above. As already mentioned the structure of the R-vine is chosen tree by tree. Starting with the first tree $\mathcal{T}_1$, the model is set up by building one tree onto the other, i.e. the now following steps have to be performed iteratively for each tree $\mathcal{T}_i$, $i = 1, \ldots, d-1$.

The (conditioned) variable pairs to be modeled by the tree under consideration ($\mathcal{T}_i$) are chosen such that the strongest pairwise dependencies that occur are modeled (first step). To access these pairwise dependencies different approaches could be followed, but normally Kendall's τ (see Section 2.2.1) is used as measure of dependence. Note that for $i = 2, \ldots, d-1$ these calculations are performed on the transformed variables from the previous iteration.

On the base of the quantities, which measure the dependence of all the variable pairs that are relevant according to the proximity condition, a graph $\mathcal{G}_i = (\mathcal{V}_i, \mathcal{E}_i')$ is set up. The edges of $\mathcal{G}_i$ represent the valid variable pairs and are weighted with the absolute values of the quantities described above. Starting from $\mathcal{G}_i$ the tree $\mathcal{T}_i = (\mathcal{V}_i, \mathcal{E}_i)$ that maximizes the sum of the weights of its edges (maximum spanning tree) is determined.

The next step is the choice of the bivariate copula families for the selected pairs. To do so, we choose the family separately for each pair. Moreover we choose among plenty of different kinds of families (see Section 2.3 for a selection of bivariate copula families), so it is appropriate to utilize Kendall's τ in the prior step, since rank correlation measures like that do not depend on the chosen distributions.

For the family selection, first of all the independence copula is considered. An independence test, which can be found in Genest and Favre (2007) and is based on Kendall's τ, is performed. If the null hypothesis of independence cannot be rejected, the independence copula is chosen. Otherwise all other copula families under consideration are compared by means of their Akaike Information Criterion (AIC). The family that yields the smallest AIC is selected.

Finally we have to estimate one or more parameters for each of the specified bivariate copulae. This is done by maximum-likelihood estimation. It remains to calculate the transformed variables according to Lemma 4.7, which are the base for the calculations in the next iteration. This concludes the last step for tree number i.

Algorithm 4.1 (cp. Dißmann et al. (2013)) summarizes the previous explanation of the sequential model selection procedure for R-vines. It yields an R-vine copula specification $(\vec{F}, \mathcal{V}, \mathcal{B})$. Together with an R-vine matrix calculated from $\mathcal{V}$, we have all the necessary information to be able to jointly estimate the R-vine parameters by maximizing the log-likelihood given in Equation (4.3.3).

We finally note, that the modifications on Algorithm 4.1 are small if one is only interested in selecting C- respectively D-vines. One simply has to select a spanning star respectively a spanning path, in Line 6 and Line 21 of the algorithm.

Input: Data $(x_{t,1}, \ldots, x_{t,d})^{\top}$, $t = 1, \ldots, T$, (realizations of i.i.d. random vectors).
Output: R-vine copula specification $(\vec{F}, \mathscr{V}, \mathcal{B})$.
 1: Determine adequate marginals $\vec{F} = (F_1, \ldots, F_d)$.
 2: Set up a complete graph $\mathcal{G}_1 = (\mathcal{V}_1 = \{1, \ldots, d\}, \mathcal{E}_1' = \{\{j, k\} \in \mathcal{V}_1^2 : 1 \leq j < k \leq d\})$.
 3: **for** $\{j, k\} \in \mathcal{E}_1'$ **do**
 4:     Calculate the measure of dependence $\widehat{m}_{j,k}$ (e.g. Kendall's τ).
 5: **end for**
 6: Select the spanning tree $\mathcal{T}_1 = (\mathcal{V}_1, \mathcal{E}_1)$ of $\mathcal{G}_1$ that maximizes $\sum_{\{j,k\} \in \mathcal{E}_1} |\widehat{m}_{j,k}|$.
 7: **for** $e = \{j, k\} \in \mathcal{E}_1$ **do**
 8:     **if** H_0 of the independence test for the pair $\{j, k\}$ cannot be rejected, **then**
 9:         choose the independence copula,
10:     **else**
11:         compare the other copula families under consideration by their AIC and choose the family with the smallest AIC ($C_{jk}(\cdot, \cdot; \boldsymbol{\theta})$).
12:     **end if**
13:     Estimate the parameter(s) $\boldsymbol{\theta}$ corresponding to the selected copula family.
14:     Calculate the transformed variables $\widehat{F}_{j|k}(x_{t,j} \,|\, x_{t,k})$ and $\widehat{F}_{k|j}(x_{t,k} \,|\, x_{t,j})$, $t = 1, \ldots, T$, by applying Lemma 4.7 with $C_{jk}(\cdot, \cdot; \widehat{\boldsymbol{\theta}})$.
15: **end for**
16: **for** $i = 2, \ldots, d-1$ **do** {Iteration over the trees $\mathcal{T}_2, \ldots, \mathcal{T}_{d-1}$.}
17:     Set up the graph $\mathcal{G}_i = (\mathcal{V}_i = \mathcal{E}_{i-1}, \mathcal{E}_i' = \{\{a, b\} \in \mathcal{V}_i^2 : \#(a \cap b) = 1\})$.
18:     **for** $\{j, k | \mathcal{D}\}$ corresponding to the edges in $\mathcal{E}_i'$ **do**
19:         Calculate the measures of dependence $\widehat{m}_{j,k|\mathcal{D}}$ (based on the transformed variables from the previous iteration).
20:     **end for**
21:     Select the spanning tree $\mathcal{T}_i = (\mathcal{V}_i, \mathcal{E}_i)$ of $\mathcal{G}_i$ that maximizes $\sum_{e \in \mathcal{E}_i} |\widehat{m}_{j,k|\mathcal{D}}|$.
22:     **for** $\{j, k | \mathcal{D}\}$ corresponding to the edges $\mathcal{E}_i$ **do**
23:         **if** H_0 of the independence test for the pair $\{j, k | \mathcal{D}\}$ cannot be rejected, **then**
24:             choose the independence copula,
25:         **else**
26:             compare the other copula families under consideration by their AIC and choose the family with the smallest AIC ($C_{jk;\mathcal{D}}(\cdot, \cdot; \boldsymbol{\theta})$).
27:         **end if**
28:         Estimate the parameter(s) $\boldsymbol{\theta}$ corresponding to the selected copula family.
29:         Calculate the transformed variables $\widehat{F}_{j|k \cup \mathcal{D}}(x_{t,j} \,|\, x_{t,k}, \boldsymbol{x}_{t,\mathcal{D}})$ and $\widehat{F}_{k|j \cup \mathcal{D}}(x_{t,k} \,|\, x_{t,j}, \boldsymbol{x}_{t,\mathcal{D}})$, $t = 1, \ldots, T$, by applying Lemma 4.7 with $C_{jk;\mathcal{D}}(\cdot, \cdot; \widehat{\boldsymbol{\theta}})$.
30:     **end for**
31: **end for**

Algorithm 4.1: Sequential R-vine model selection.

4.5   R-vines on spatial mean temperature data

To conclude our chapter about regular vines we investigate the spatial dependencies of our daily mean temperature data and fit an R-vine to the data or more precisely to the copula data calculated from the original mean temperature data set as presented in Chapter 3.

Initially we have a look on the copula data and the bivariate dependencies for five selected observation stations, namely Hamburg-Fuhlsbüttel (21) and Lübeck-Blankensee (32) which are located in northern Germany, as well as Nürnberg (41), Regensburg (43) and Weiden (53) in Bavaria, i.e. south-east Germany. We expect that the data corresponding to observation stations which are located close to each other exhibits strong correlation, whereas observations which are collected far apart should be rather weakly correlated. This expectation is confirmed by Figure 4.5.1, which depicts a pairs plot for the five observation stations addressed above. The three Bavarian stations yield Kendall's τ's bigger than 0.65. For the Kendall's τ corresponding to the pair Hamburg and Lübeck a value of 0.78 is calculated. The remaining results for the rank correlation measures indicate a low correlation between the data stemming from the north and the data from the south.

Besides Figure 4.5.1 depicts histograms of the copula data on the diagonal panels, scatter plots on the upper off diagonal panels and contour plots on the lower off diagonal panels. From the histograms we see that the copula data fulfill the requirement of being uniform on $[0, 1]$. Both the scatter plots and the contour plots indicate the magnitude of the correlation in accordance with the observations on the Kendall's τ's from above. The scatter and contour plots corresponding to the six north-south pairs suggest slightly asymmetric copulae in case these pairs should be modeled in the first tree of an R-vine.

Since the variable pairs of the current tree of an R-vine are selected in such a way that the highest dependencies are modeled and since we expect that high dependencies occur for station pairs with short distance, we consider the contour plots of the station pairs with the twelve shortest distances. They are illustrated in Figure 4.5.2, where the two numbers in the upper left corner of each plot are the distance in kilometers and the elevation difference in meters. The respective Kendall's τ's in the lower right corners are all larger than 0.7, with the exception of the pair Gießen (17) - Berleburg (3). It is conspicuous that this pair exhibits a large elevation difference, which seems to be the reason for the relatively low Kendall's τ. Compared to the eleven other contour plots, the contour plot for this outstanding pair seems to be the only one which shows slight asymmetry. The eleven remaining contour plots possess an elliptical shape, which is a hint that a big share of the bivariate copulae in tree one of the subsequently investigated R-vine will be either Student-t or even Gaussian copulae.

In order to learn more about the structure of an R-vine for our daily mean temperature data, we fitted an R-vine to the respective copula data. For this purpose we took Gaussian, Student-t, Clayton, Gumbel, Frank and Joe copulae and rotated versions thereof into consideration. Moreover an independence test was performed for each variable pair, to check if an independence copula would be more appropriate than any other copula type. Figure 4.5.3 shows the first tree of this R-vine embedded into a map of Germany. A comparison with Figure 4.5.2 yields, that eleven of the twelve pairs, excluding the pair Gießen (17) - Berleburg (3), are building blocks included in the first tree of our fitted
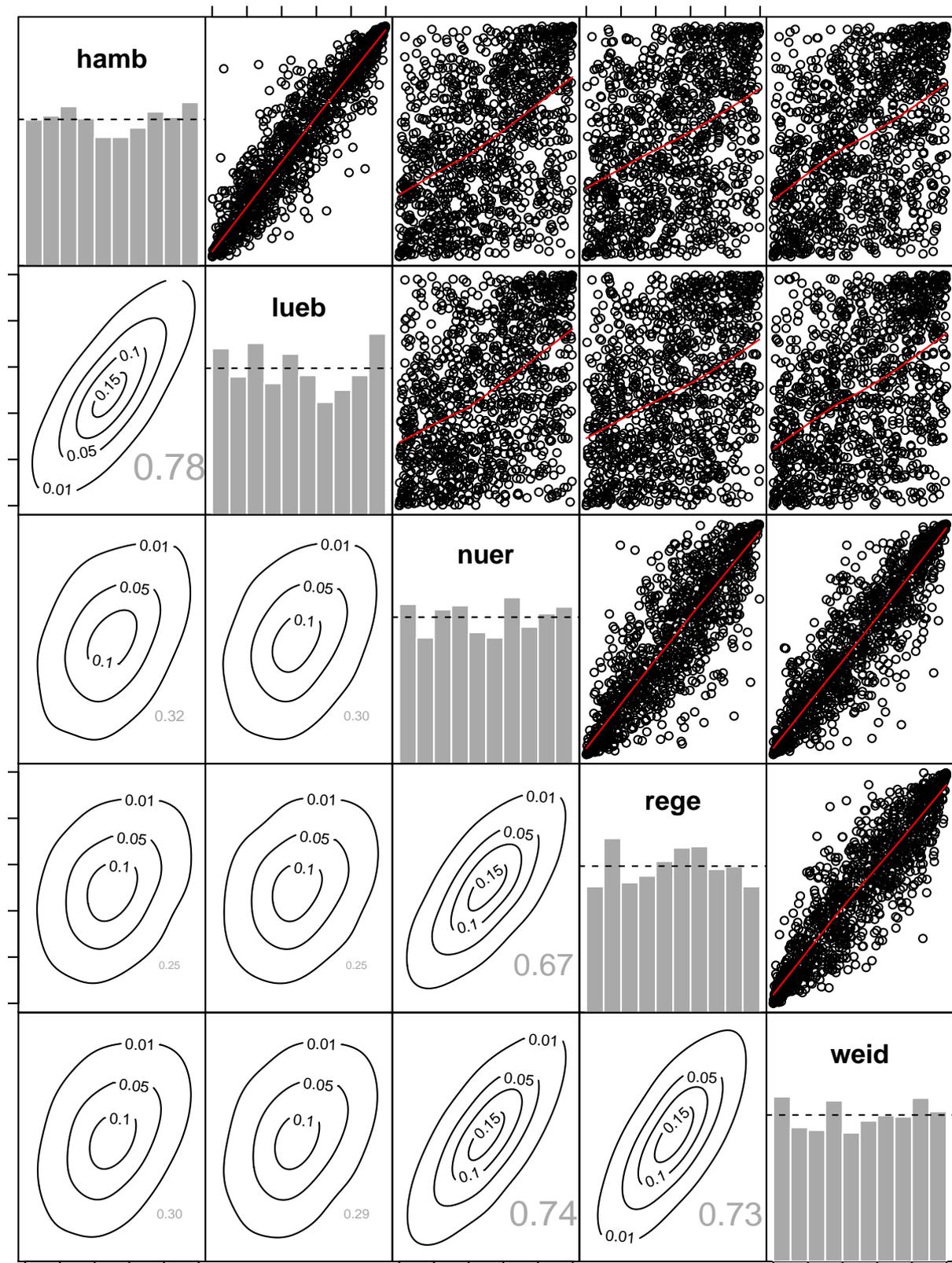
Figure 4.5.1: Pairs plot of the copula data of Hamburg-Fuhlsbüttel (21), Lübeck-Blankensee (32), Nürnberg (41), Regensburg (43) and Weiden (53), with contour plots on the lower off diagonal panels, histograms on the diagonal panels and scatter plots on the upper off diagonal panels.
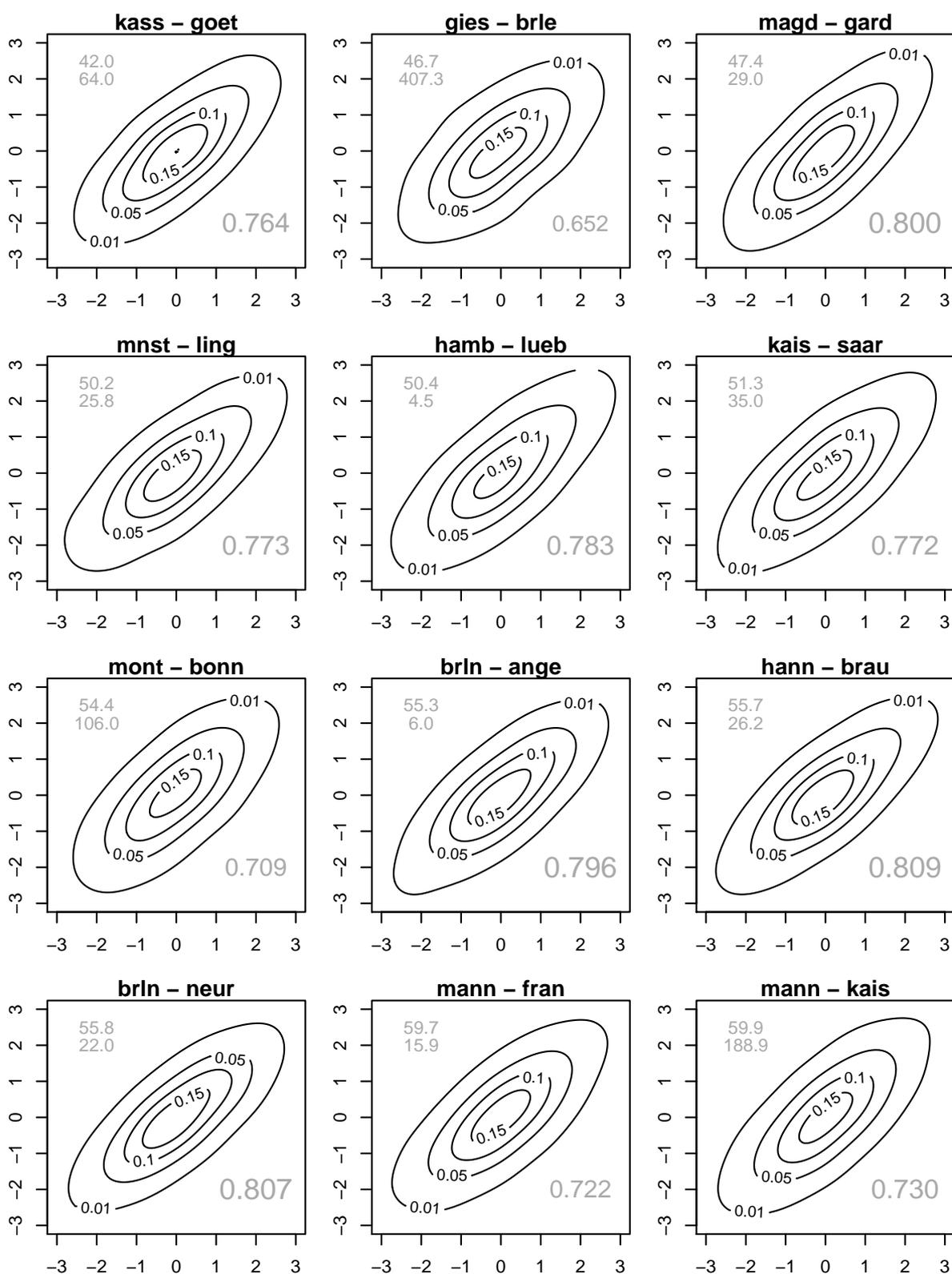
Figure 4.5.2: Contour plots for the station pairs with the twelve shortest distances.
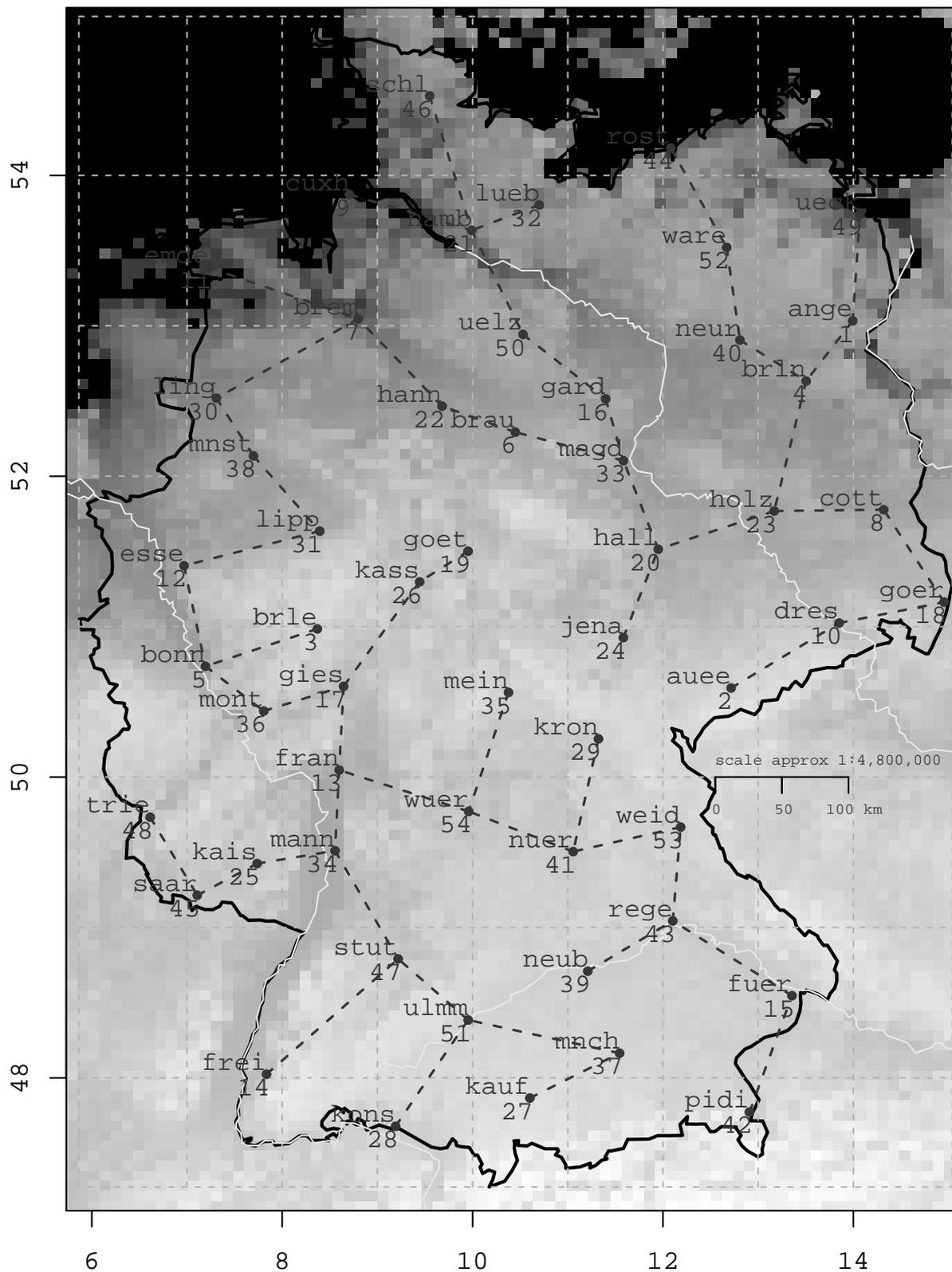
Figure 4.5.3: Tree one of the fitted R-vine embeded into a map of Germany.

R-vine. A further observation arising from the two previously considered figures is, that the R-vine selection procedure most likely chooses variable pairs whose respective observation stations are located closely to each other. Our observations indicate that the spatial arrangement of the observation stations plays an important role with regard to the modeling of the dependence structure.

Finally we are interested in the number of independence copulae respectively the number of elliptical copulae occurring in our R-vine. The upper part of Figure 4.5.4 illustrates the number of non-independence copulae for each tree of the R-vine. The dashed diagonal gives the number of bivariate pairs building the respective tree, i.e. it indicates the maximum number of non-independence copulae that would be possible. From the plot we observe that the number of non-independence copulae decreases quickly with increasing tree number. Starting from tree number 19 a maximum number of ten non-independence copulae isn't exceeded. Contrariwise the lower part of the figure visualizes the share of independence copulae for each tree. We find that more than 50% of the copulae in trees of order ten or higher are independence copulae, which might justify to truncate the R-vine after a certain tree.

The intention of Figure 4.5.5 is it to visualize the number of elliptical copulae composing the R-vine. It shows that all copulae of the first tree are Student-t copulae. Besides we find that the number of elliptical copulae decreases rapidly with increasing tree number. Only a few of the elliptical copulae are Gaussian, most of them are t-copulae. Moreover we observe that the share of non-elliptical copulae is quite high for most of the trees. This observation justifies the application of pair-copula constructions instead of simply using a multivariate Student-t copula.

Motivated by the outcomes of this section we are going to develop some new spatial vine models in the subsequent chapters. Within this framework we will return to our exemplary data set of daily mean temperatures.
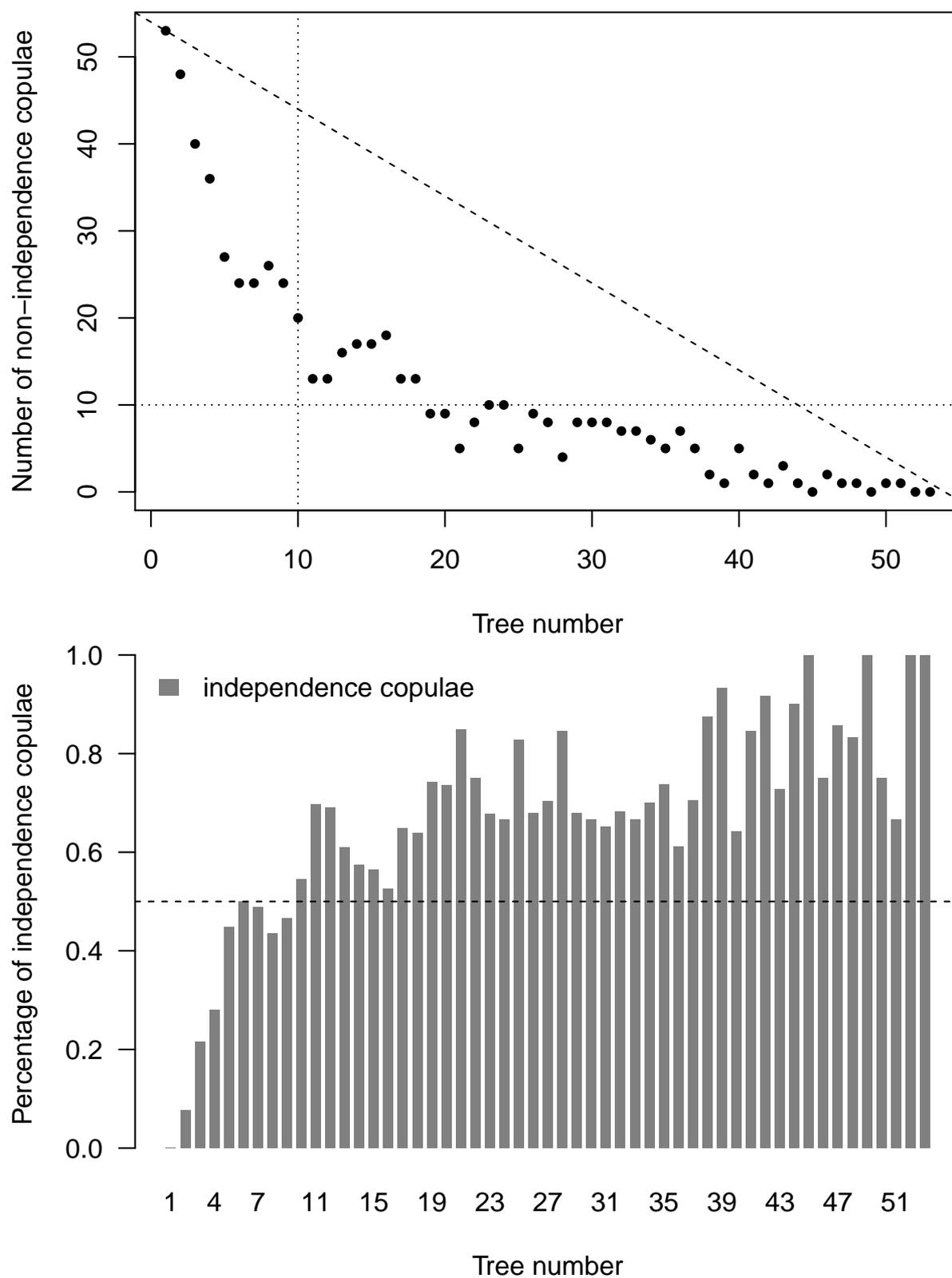
Figure 4.5.4: Visualization of the number of (non-)independence copulae per tree, which occur in the estimated R-vine.
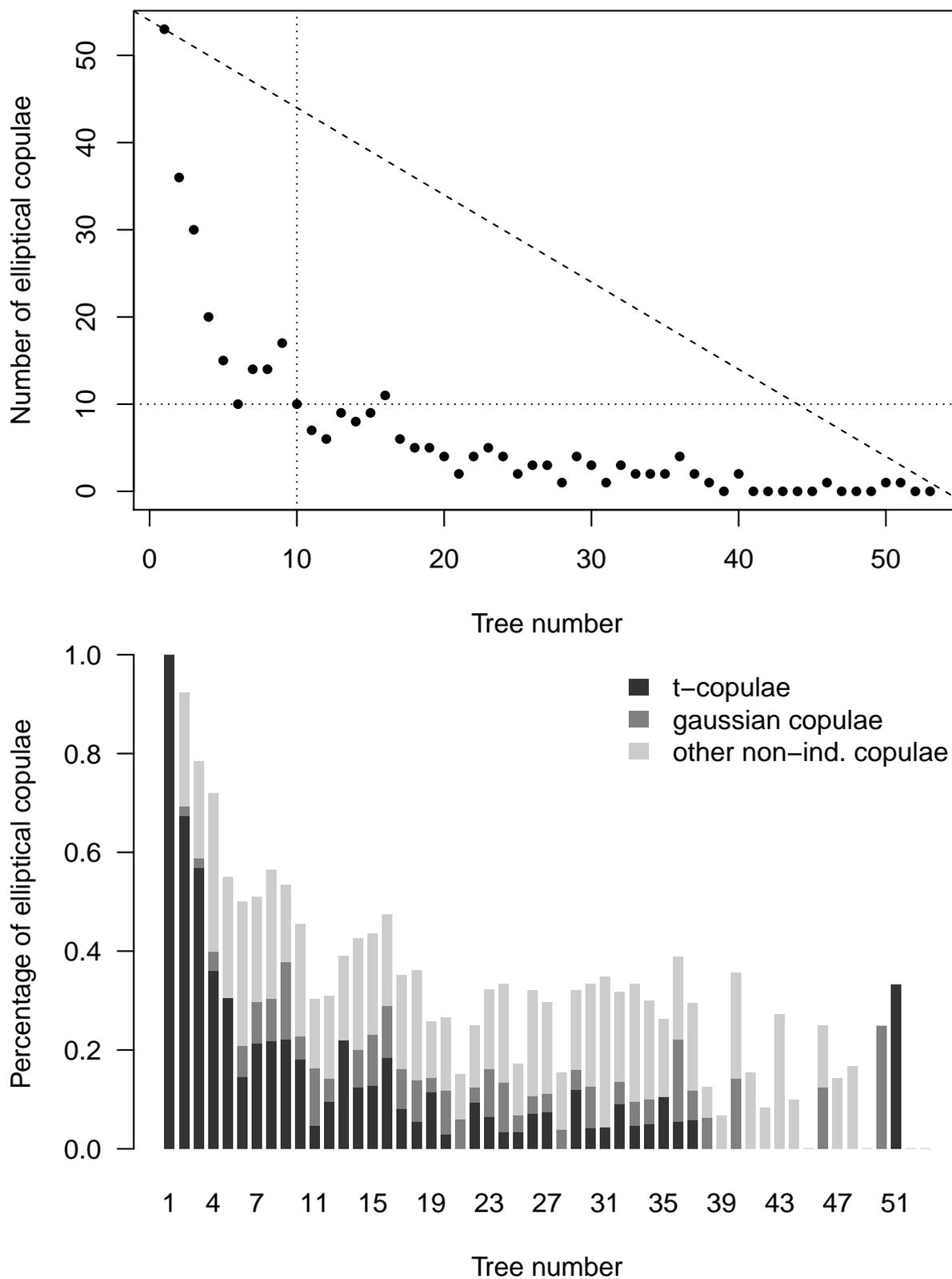
Figure 4.5.5: Visualization of the number of elliptical copulae per tree, which occur in the estimated R-vine.

Chapter 5

Spatial R-vine models

The previous investigations in Chapter 4 showed that for spatial data, the spatial arrangement of the locations where the data stems from plays an important role with regard to the modeling of the dependence structure. Moreover we saw that the number of parameters of an ordinary R-vine model becomes quite high, if a huge number of variables are included into the model. In our case ($d = 54$) we have two consider parameters for 1431 pair-copulae. These observations lead to the idea, to use additional spatial information to reduce the number of parameters and to develop a *spatial R-vine model (SV)*.

5.1 Preliminary analyses

For the purpose of the development of a spatial R-vine model, we first need to introduce some notation. In the following we consider the copula data $\boldsymbol{u}^1, \ldots, \boldsymbol{u}^d$ with $\boldsymbol{u}^s = (u_1^s, \ldots, u_N^s)^\top$, $s = 1, \ldots, d$, i.e. we consider time series of length N for d different observation stations. Elevation, longitude and latitude for an arbitrary observation station s are denoted by $x_{\mathrm{elev},s}$, $x_{\mathrm{long},s}$ and $x_{\mathrm{lat},s}$. These quantities allow to calculate an

$$\text{unsigned distance } d_{i,j}$$

and an

$$\text{unsigned elevation difference } e_{i,j}$$

for each pair of observation stations (i, j) with $1 \leq i < j \leq d$.

For an arbitrary edge e of the subsequently developed spatial R-vine, the corresponding bivariate copula is denoted as

$$C_{i,j;\mathcal{D}_e},$$

where $1 \leq i < j \leq d$ compose the conditioned set $\mathcal{C}_e = \{i, j\}$ and $\mathcal{D}_e$ is the conditioning set according to Definition 4.2. Then the respective first and second copula parameters are denoted as

$$\theta_{i,j|\mathcal{D}_e} \qquad \text{and} \qquad \nu_{i,j|\mathcal{D}_e},$$

where the second parameter is only needed if $C_{i,j;\mathcal{D}_e}$ stems from a copula family with two parameters (e.g. Student-t copula). The corresponding Kendall's τ is given as

$$\tau_{i,j} \quad \text{respectively} \quad \tau_{i,j|\mathcal{D}_e},$$

depending on whether it is calculated directly from the data or based on transformed variables in the trees $\mathcal{T}_2, \mathcal{T}_3, \ldots, \mathcal{T}_{d-1}$ of the R-vine.

Now that the necessary notation is introduced, we return to the mean temperature data set ($d = 54$) to conduct further investigations on the spatial dependencies of the given variables. To this end we are interested in a potential relationship of the correlation and the distance respective the elevation difference of observation station pairs. These relations are investigated in Figure 5.1.1.

For all $d(d-1)/2 = 1431$ possible station pairs (i, j), $1 \leq i < j \leq 54$, the empirical Kendall's τ's $\widehat{\tau}_{i,j}$ are calculated, to quantify the correlations of these pairs. Moreover, the respective pairwise distances $d_{i,j}$ and elevation differences $e_{i,j}$ are calculated. As the empirical Kendall's τ's $\widehat{\tau}_{i,j}$ are restricted to $(-1, 1)$, and since we want to take a linear model into consideration, where the response should live on the whole real line, we apply the Fisher z-transform given in (2.7.1) to transform from $(-1, 1)$ to $(-\infty, \infty)$. Furthermore we apply the natural logarithm to the distances and elevation differences.

The upper part of Figure 5.1.1 is a plot of the Fisher z-transformed Kendall's τ's against the logarithmized distances. Here, a distinct linear relationship is observed. The lower part of the figure illustrates the respective plot against the logarithmized elevation differences. Also in this case there seems to be some kind of linear relationship, but it is not that distinct, as it was in the case of the pairwise distances.

The straight gray lines in both plots depict the regression line corresponding to the particular linear relationship. The horizontal lines help to identify the values $0.2, 0.3, \ldots, 0.8$ of Kendall's τ, whereas the vertical lines indicate the three distances of 50, 100 and 200 kilometers respectively the three elevation differences of 50, 100 and 200 meters.

Recalling our investigations of R-vines on the mean temperature data, we saw that primarily pairs exhibiting short distances respectively high correlations enter the R-vine model in the lower trees. This yields that the relevant pairs are represented by the upper parts of the plots in Figure 5.1.1, where we observe higher fluctuation around the regression lines. This proposes a tree-wise investigation of the pairs occurring in an R-vine, in order to detect significant relationships. This investigation is going to be performed later on.

In order to justify the application of linear models to model the empirical Kendall's τ's $\widehat{\tau}_{i,j}$, we fit the linear model

$$\mathrm{F}_z(\tau_{i,j}) = \alpha + \beta_d \ln(d_{i,j}) + \beta_e \ln(e_{i,j}) + \varepsilon_{i,j}, \quad \varepsilon_{i,j} \overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(0, \sigma^2\right), \quad 1 \leq i < j \leq 54.$$

We obtain highly significant parameter estimates $\widehat{\alpha} = 2.2596$, $\widehat{\beta}_d = -0.3074$ and $\widehat{\beta}_e = -0.0066$, and a pretty high $R^2 = 0.9412$.

To see that the assumptions of a linear model are fulfilled, we investigate the standardized residuals obtained from fitting the above stated model to the data. The Q-Q-plot and the histogram in Figure 5.1.2 yield, that the normality assumption is justified. From the scatter plot of the standardized residuals against the empirical Kendall's τ's we observe random scatter around zero, but increasing volatility with increasing $\widehat{\tau}_{i,j}$, which was already observed from the regression plots in Figure 5.1.1. If we would restrict us once again only to high correlations, the assumption of constant variance might be adequate.

So far we figured out, that the Kendall's τ's corresponding to spatially arranged observation station pairs might be modeled by means of distance and elevation difference.
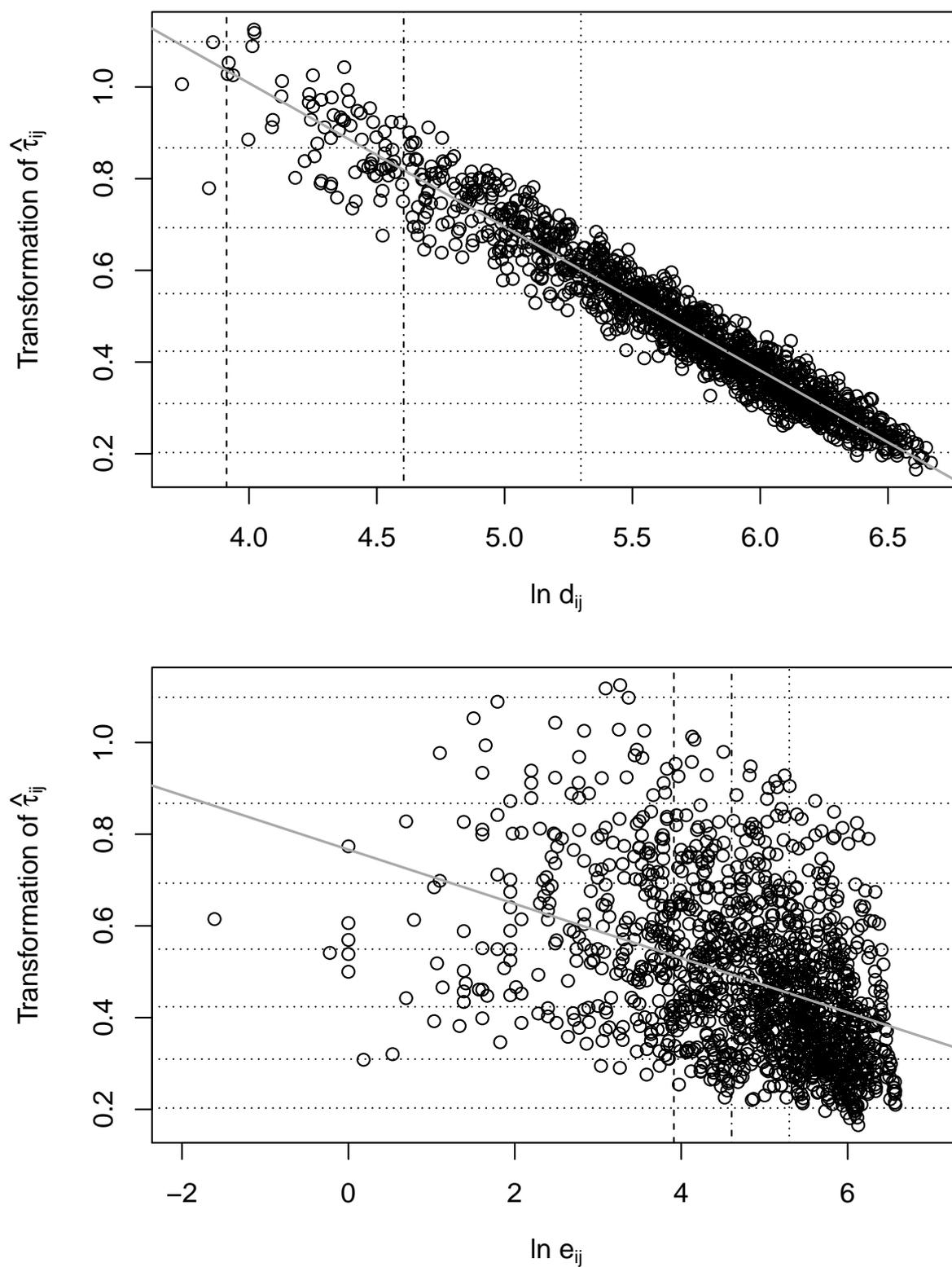
Figure 5.1.1: Relationship of $F_z(\widehat{\tau}_{i,j})$ with $\ln(d_{i,j})$ and $\ln(e_{i,j})$, respectively.
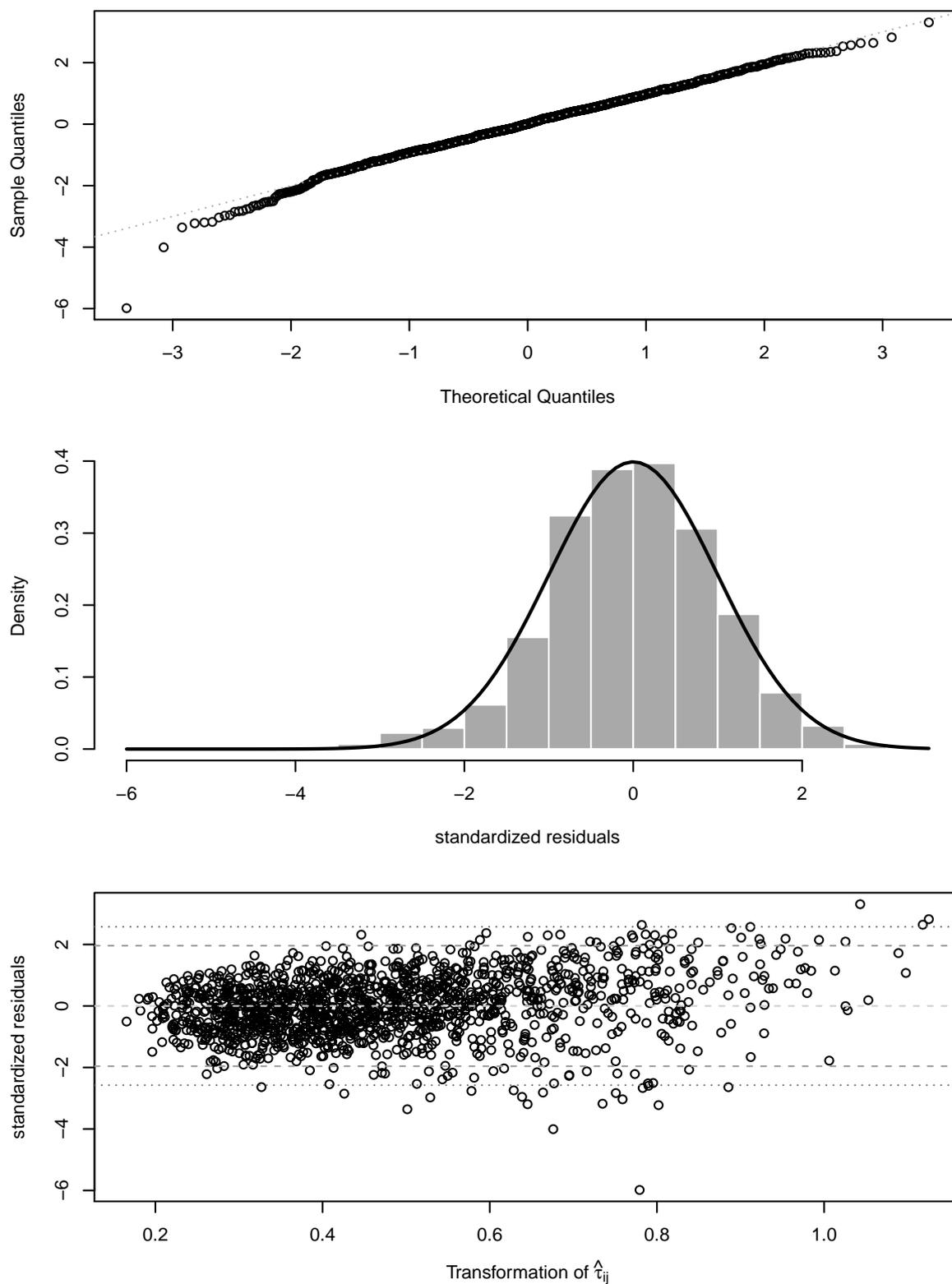
Figure 5.1.2: Analysis plots for linear regression of $F_z(\widehat{\tau}_{i,j})$ against $\ln(d_{i,j})$ and $\ln(e_{i,j})$. Top: Q-Q-plot. Middle: Histogram with standard normal density superimposed. Bottom: Standardized residuals against $F_z(\widehat{\tau}_{i,j})$ and 95% as well as 99% confidence interval (dashed and dotted lines, respectively).

In the following we aim to apply these findings to reduce the number of parameters which are needed in an R-vine model. Hence a tree-wise analysis of an R-vine model fitted to the mean temperature data might lead to a deeper insight into the relationship of the R-vine parameters and available spatial information. For these investigations we consider an R-vine truncated after tree ten, which allows for Gaussian, Student-t, Clayton, Gumbel and Frank copulae, as well as rotated versions thereof, where the copula families are selected according to the Akaike information criterion. Moreover we modify our notation intuitively, in order to indicate dependence on the edge $e \in \mathcal{E}_l$ and sometimes we add $l \leq 10$ as a superscript to underline the dependence on the tree number.

Table 5.1 summarizes the structure of the R-vine which we are going to investigate in more detail in the following, where the families are coded according to Table 4.4. On the one hand we observe that the copula family wich occurs most for the trees one to nine is the Student-t family. On the other hand the number of other copula families increases with the tree number. In tree ten the Gumbel family is the dominating one.

l	# Φ	# t	# C	# G	# F	$\min(\widehat{\tau}^l_{i(e),j(e)\|\mathcal{D}_e})$	$\max(\widehat{\tau}^l_{i(e),j(e)\|\mathcal{D}_e})$	$\overline{\widehat{\nu}^l_{i(e),j(e)\|\mathcal{D}_e}}$
1	0	53	0	0	0	0.591	0.809	7.659
2	1	38	1	7	5	-0.153	0.317	9.908
3	1	35	4	6	5	-0.222	0.356	10.837
4	2	23	5	13	7	-0.180	0.300	11.873
5	1	21	9	9	9	-0.154	0.278	11.843
6	5	20	5	12	6	-0.114	0.285	12.967
7	6	15	10	10	6	-0.193	0.239	13.378
8	5	18	4	7	12	-0.098	0.193	14.888
9	7	15	6	8	9	-0.066	0.279	14.373
10	3	10	8	16	7	-0.128	0.246	14.105
Sum	**31**	**248**	**52**	**88**	**66**			

Table 5.1: Summary of the structure of the truncated R-vine under consideration.

Figure 5.1.3 plots both the Kendall's τ's $\widehat{\tau}^l_{i(e),j(e)\|\mathcal{D}_e}$ arising in the truncated R-vine and the degrees of freedom $\widehat{\nu}^l_{i(e),j(e)\|\mathcal{D}_e}$ of the t-copulae which occur in the R-vine against the respective tree number l. Recall that the Kendall's τ's in tree number two and higher are calculated based on transformed variables.

From the figure and from Table 5.1 we observe that the strong dependencies are already captured in tree one and that the correlations in higher trees scatter mostly between -0.2 and 0.3, i.e. negative dependencies occur as well. In the upper panel of Figure 5.1.3 t-copulae are highlighted, to be able to detect patterns in the relationship of Kendall's τ and the tree number. No such relation can be detected. Moreover we cannot detect any pattern in the Kendall's τ's, with respect to the tree number.

For the degrees of freedom we discover a quadratic trend with regard to the tree number. For higher tree numbers the degrees of freedom seem to be higher, at least in the mean. In the next section we will use this finding to reparametrize the second copula parameters $\nu^l_{i(e),j(e)\|\mathcal{D}_e}$ together for all trees $l = 1, \ldots, 10$.

It remains to specify relations between the first copula parameters $\theta^l_{i(e),j(e)\|\mathcal{D}_e}$ and the
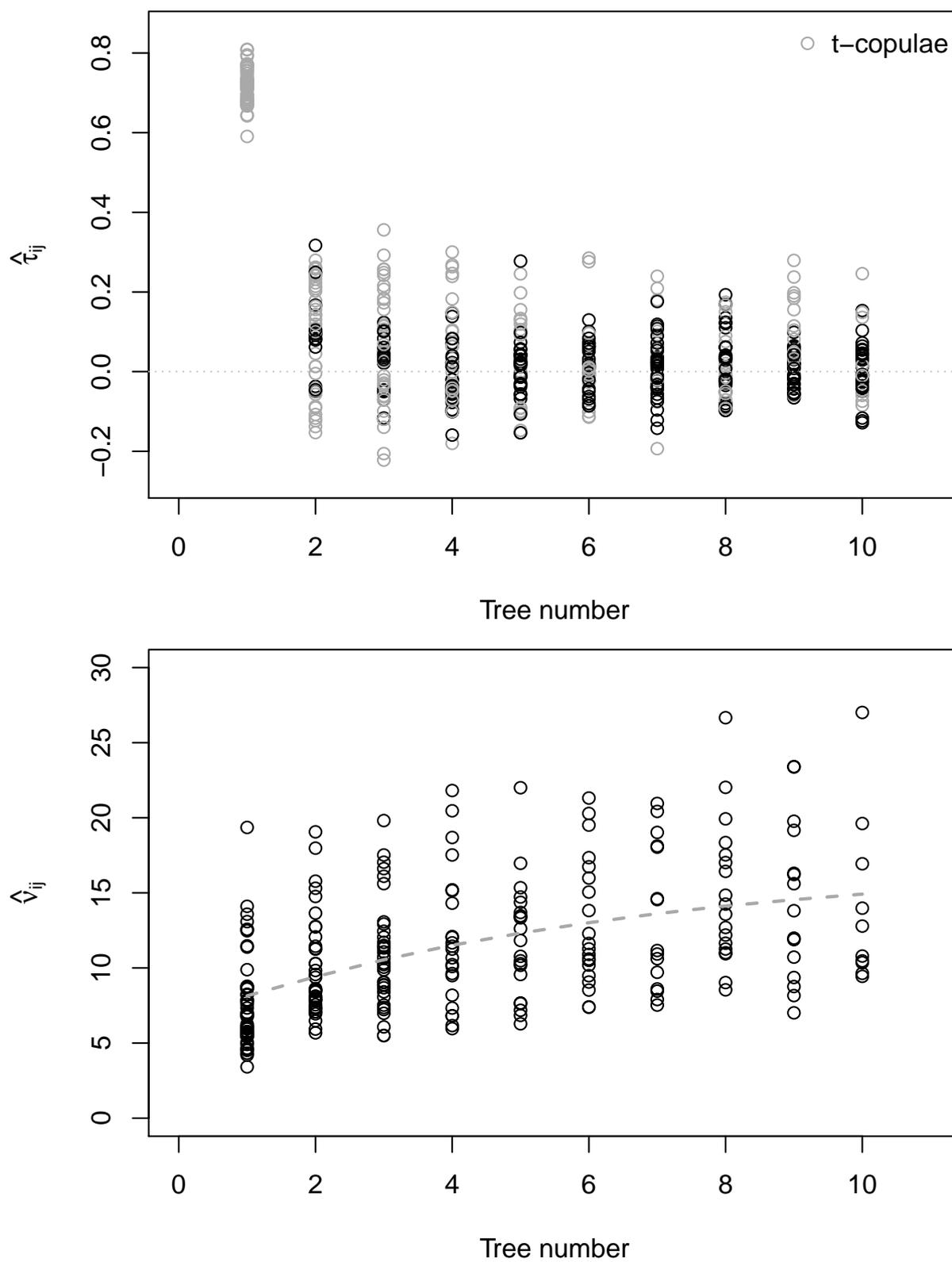
Figure 5.1.3: Plots of $\widehat{\tau}^l_{i(e),j(e)|\mathcal{D}_e}$ and $\widehat{\nu}^l_{i(e),j(e)|\mathcal{D}_e}$, $e \in \mathcal{E}_l$, against the respective tree number $l = 1, \ldots, 10$.

corresponding distances $d_{i(e),j(e)}$ respectively elevation differences $e_{i(e),j(e)}$, distinguishing which tree $l \leq 10$ the edge e stems from. We already figured out, that the Kendall's τ's $\tau^l_{i(e),j(e)|\mathcal{D}_e}$ might be modeled by means of the predictors $d_{i(e),j(e)}$ and $e_{i(e),j(e)}$. Moreover we know that there exist relations between the copula parameters $\theta^l_{i(e),j(e)|\mathcal{D}_e}$ and the Kendall's τ's $\tau^l_{i(e),j(e)|\mathcal{D}_e}$, depending on the copula family $t_{i(e),j(e)|\mathcal{D}_e}$. Hence we have got to investigate possible dependencies of the Fisher z-transformed Kendall's τ's on the distances and elevation differences, separately for each tree, in case we find such relations, we are able to transform back to the copula parameter level.

Before we continue with the tree-wise analysis, we briefly address the relationship between Kendall's τ and the model parameter. For a selection of important copula families, the respective relations have already been stated in the examples of Section 2.3. For our purposes we need a transformation that transforms a Kendall's τ to the respective model parameter. The expression

$$\theta^l_{i(e),j(e)|\mathcal{D}_e} = \mathrm{T}_{\tau \to \theta}\left(\tau^l_{i(e),j(e)|\mathcal{D}_e}; t_{i(e),j(e)|\mathcal{D}_e}\right),$$

denotes the desired transformation.

To continue the tree-wise analysis, we start with an investigation of the first tree (it holds $\mathcal{D}_e = \emptyset$), i.e. let $e \in \mathcal{E}_1$ be an arbitrary edge of $\mathcal{T}_1$. For all 53 edges $e \in \mathcal{E}_1$, we plot the respective Fisher z-transformed Kendall's τ $\mathrm{F}_z(\widehat{\tau}_{i(e),j(e)})$ both against $\ln(d_{i(e),j(e)})$ and against $\ln(e_{i(e),j(e)})$. Figure 5.1.4 shows these plots, from which we conclude that a linear relationship seems appropriate in both cases, since the slope of the regression line is significant.

For edges $e \in \mathcal{E}_l$ of trees $\mathcal{T}_l$ with $l > 1$, $\mathcal{D}_e$ is non-empty. This means that the distances and elevation differences between elements $k \in \mathcal{D}_e$ and $i(e)$ respectively $j(e)$ may also be considered as additional predictors for the parameters $\theta_{i(e),j(e)|\mathcal{D}_e}$. Since there is no ordering of the elements of $\mathcal{D}_e$, there is no justification that the potential predictors corresponding to different elements of $\mathcal{D}_e$ should be treated in a different way. Therefore we define the means

$$\overline{d_{i(e),\mathcal{D}_e}} := \frac{1}{l-1} \sum_{k \in \mathcal{D}_e} d_{i(e),k},$$

$$\overline{e_{i(e),\mathcal{D}_e}} := \frac{1}{l-1} \sum_{k \in \mathcal{D}_e} e_{i(e),k},$$

$$\overline{d_{j(e),\mathcal{D}_e}} := \frac{1}{l-1} \sum_{k \in \mathcal{D}_e} d_{j(e),k},$$

$$\overline{e_{j(e),\mathcal{D}_e}} := \frac{1}{l-1} \sum_{k \in \mathcal{D}_e} e_{j(e),k},$$

and use them as further predictors in our models, i.e. we are going to investigate six plots for all trees $\mathcal{T}_l$ with $l > 1$.

For Tree $\mathcal{T}_2$ it holds that $\mathcal{D}_e = \{k\}$ for all edges $e \in \mathcal{E}_2$, i.e. the above defined means equal the single distances respectively elevation differences that enter the respective sum.
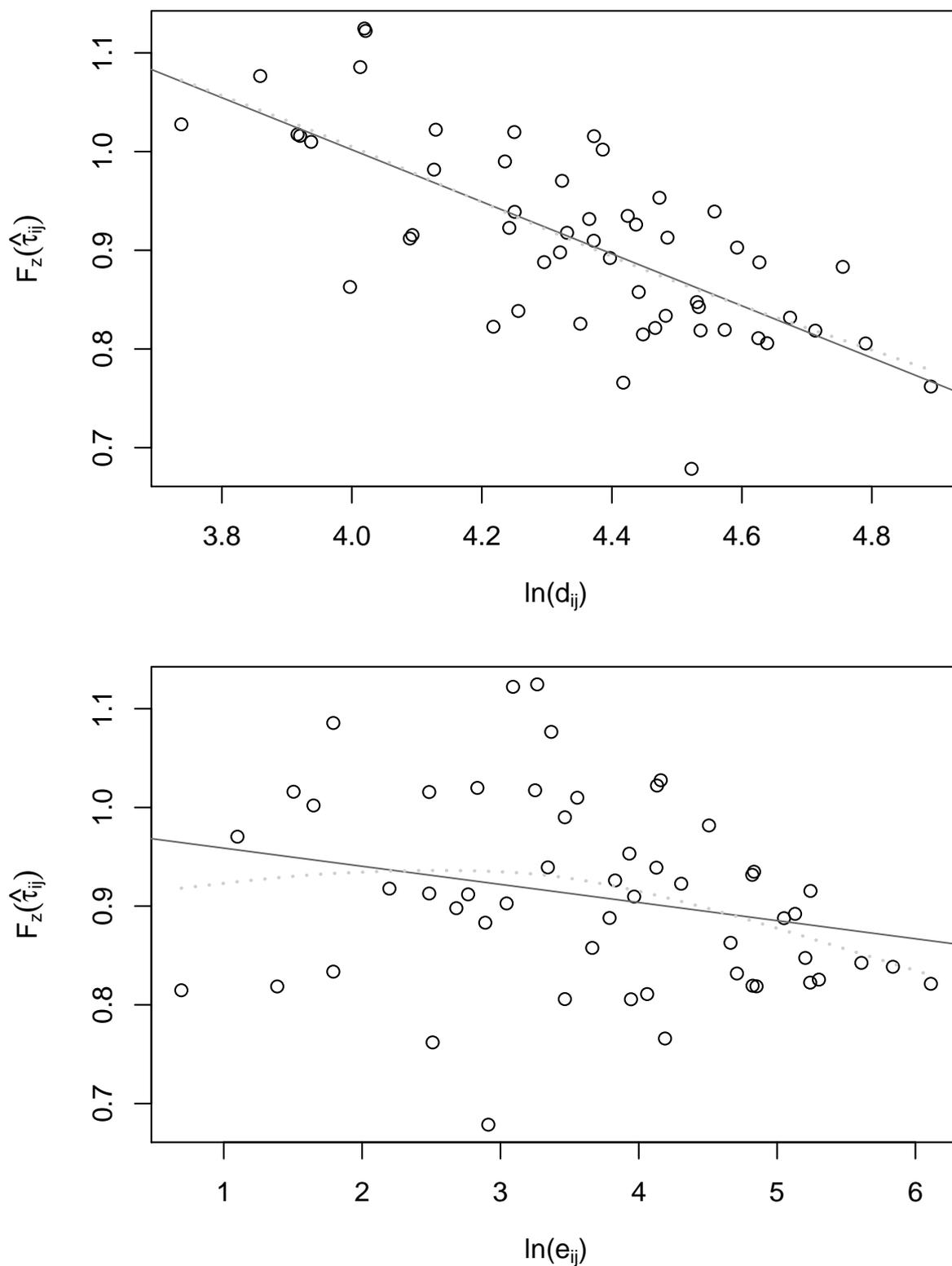
Figure 5.1.4: Scatter plots of $F_z(\widehat{\tau}_{i(e),j(e)})$ against $\ln(d_{i(e),j(e)})$ and $\ln(e_{i(e),j(e)})$, respectively, where $e \in \mathcal{E}_1$ (Tree 1).

The six plots of $\mathrm{F}_z(\widehat{\tau}_{i(e),j(e)|\mathcal{D}_e})$ against $\ln\left(d_{i(e),j(e)}\right)$, $\ln\left(e_{i(e),j(e)}\right)$, $\ln\left(d_{i(e),k}\right)$, $\ln\left(e_{i(e),k}\right)$, $\ln\left(d_{j(e),k}\right)$ and $\ln\left(e_{j(e),k}\right)$ are given in Figure 5.1.5 and yield a linear relationship of $\mathrm{F}_z(\widehat{\tau}_{i(e),j(e)|\mathcal{D}_e})$ on each of the six predictors. The plot against $\ln\left(d_{i(e),j(e)}\right)$ shows a negative slope, whereas the other five plots show positive slopes.

Figures 5.1.6 to 5.1.8, corresponding to the trees $\mathcal{T}_3$, $\mathcal{T}_4$ and $\mathcal{T}_5$, show again a significant negative slope for the predictor $\ln\left(d_{i(e),j(e)}\right)$. The remaining plots mostly do not show any significant slope. These findings suffice us, that we proceed to the next section, where we set up different tree-wise models for the copula parameters, which include different subsets of the predictors examined above.

5.2 Model formulation

In this section we are going to propose several reparametrizations of an R-vine model, based on the spatial predictors examined in the previous section.

Let us first of all assume a polynomial *reparametrization for the second copula parameters* $\widetilde{\nu}^l_{i(e),j(e)|\mathcal{D}_e}$ of the copulae $C_{i(e),j(e);\mathcal{D}_e}$, $e \in \mathcal{E}_l$, $l = 1,\dots,10$, wherever second parameters occur. For this purpose we model the logarithm of the parameters in dependence of the tree number as

$$\ln\nu^l_{i(e),j(e)|\mathcal{D}_e} = \beta^\nu_0 + \beta^\nu_1 l + \beta^\nu_2 l^2 + \dots + \beta^\nu_q l^q + \varepsilon_{i(e),j(e)}, \quad \varepsilon_{i(e),j(e)} \overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(0,\sigma^2\right), \quad (5.2.1)$$
$$e \in \mathcal{E}_l, \quad l = 1,\dots,10,$$

where the degree q of the polynomial is determined by adding summands $\beta^\nu_r l^r$ to the model, as long as β^ν_r is significant at a level of 5% for the data set under consideration. We set $q = r - 1$, with β_r being the first non-significant parameter added to the model.

Having determined q, the reparametrization for the degrees of freedom (nu) is defined via

$$\widetilde{\nu}^l_{i(e),j(e)|\mathcal{D}_e} = \exp\left\{(1, l, l^2, \dots, l^q) \cdot \boldsymbol{\beta}^{\text{SV}}_\nu\right\}, \quad e \in \mathcal{E}_l, \quad l = 1,\dots,10,$$

where the respective parameters are summarized in the vector

$$\boldsymbol{\beta}^{\text{SV}}_\nu = \left(\beta^\nu_0, \beta^\nu_1, \beta^\nu_2, \dots, \beta^\nu_q\right)^\top \in \mathbb{R}^{n^\nu_{\text{par}}},$$

where

$$n^\nu_{\text{par}} = r = q + 1.$$

Now we are going to propose several different reparametrizations for the first parameters $\widetilde{\theta}^l_{i(e),j(e)|\mathcal{D}_e}$ of the bivariate copulae $C_{i(e),j(e);\mathcal{D}_e}$ of the trees $l = 1,\dots,10$, where $e \in \mathcal{E}_l$.

The first reparametrization which we are going to take into consideration is the reparametrization which involves all available predictors. That is why we call it *full reparametrization* (full). It is given by

$$\widetilde{\theta}^l_{i(e),j(e)|\mathcal{D}_e} = g^{\text{full}}_l\left(d_{i(e),j(e)}, e_{i(e),j(e)}, \overline{d_{i(e),\mathcal{D}_e}}, \overline{e_{i(e),\mathcal{D}_e}}, \overline{d_{j(e),\mathcal{D}_e}}, \overline{e_{j(e),\mathcal{D}_e}}|\boldsymbol{\beta}_{\text{full},l}, t_{i(e),j(e)|\mathcal{D}_e}\right)$$
$$= \mathrm{T}_{\tau\to\theta}\left(\mathrm{F}^{-1}_z\left(h^{\text{full}}_l(e|\boldsymbol{\beta}_{\text{full},l})\right); t_{i(e),j(e)|\mathcal{D}_e}\right), \quad e \in \mathcal{E}_l, \quad l = 1,\dots,10,$$

where

$$h^{\text{full}}_1\left(e|\boldsymbol{\beta}_{\text{full},1}\right) = \beta^{\text{full}}_{1,0} + \beta^{\text{full}}_{1,1}\ln\left(d_{i(e),j(e)}\right) + \beta^{\text{full}}_{1,2}\ln\left(e_{i(e),j(e)}\right), \quad e \in \mathcal{E}_1,$$
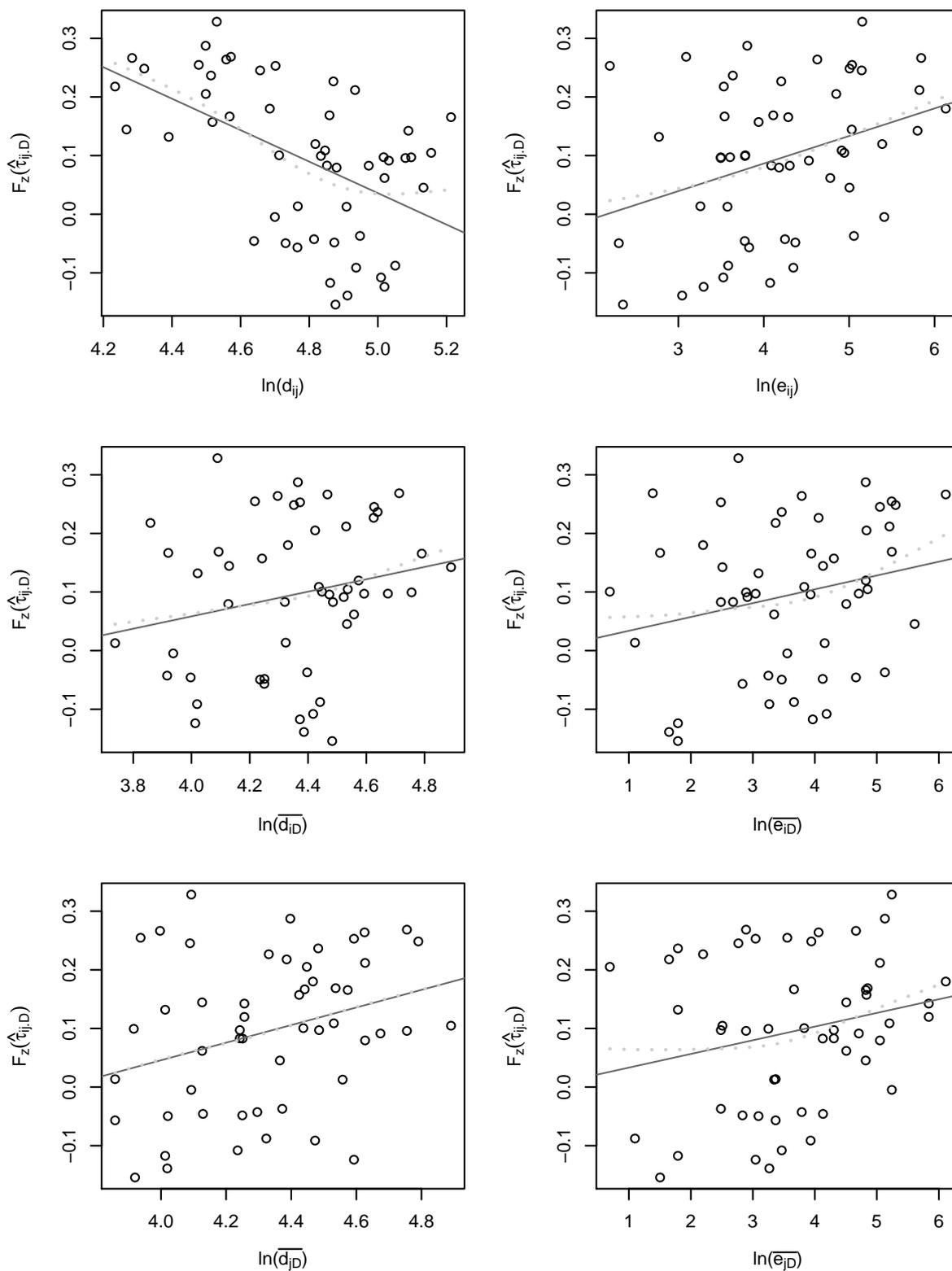
Figure 5.1.5: Scatter plots of $F_z(\widehat{\tau}_{i(e),j(e)|\mathcal{D}_e})$ against $\ln\left(d_{i(e),j(e)}\right)$, $\ln\left(e_{i(e),j(e)}\right)$, $\ln\left(\overline{d_{i(e),\mathcal{D}_e}}\right)$, $\ln\left(\overline{e_{i(e),\mathcal{D}_e}}\right)$, $\ln\left(\overline{d_{j(e),\mathcal{D}_e}}\right)$ and $\ln\left(\overline{e_{j(e),\mathcal{D}_e}}\right)$, respectively, where $e \in \mathcal{E}_2$ (Tree 2).
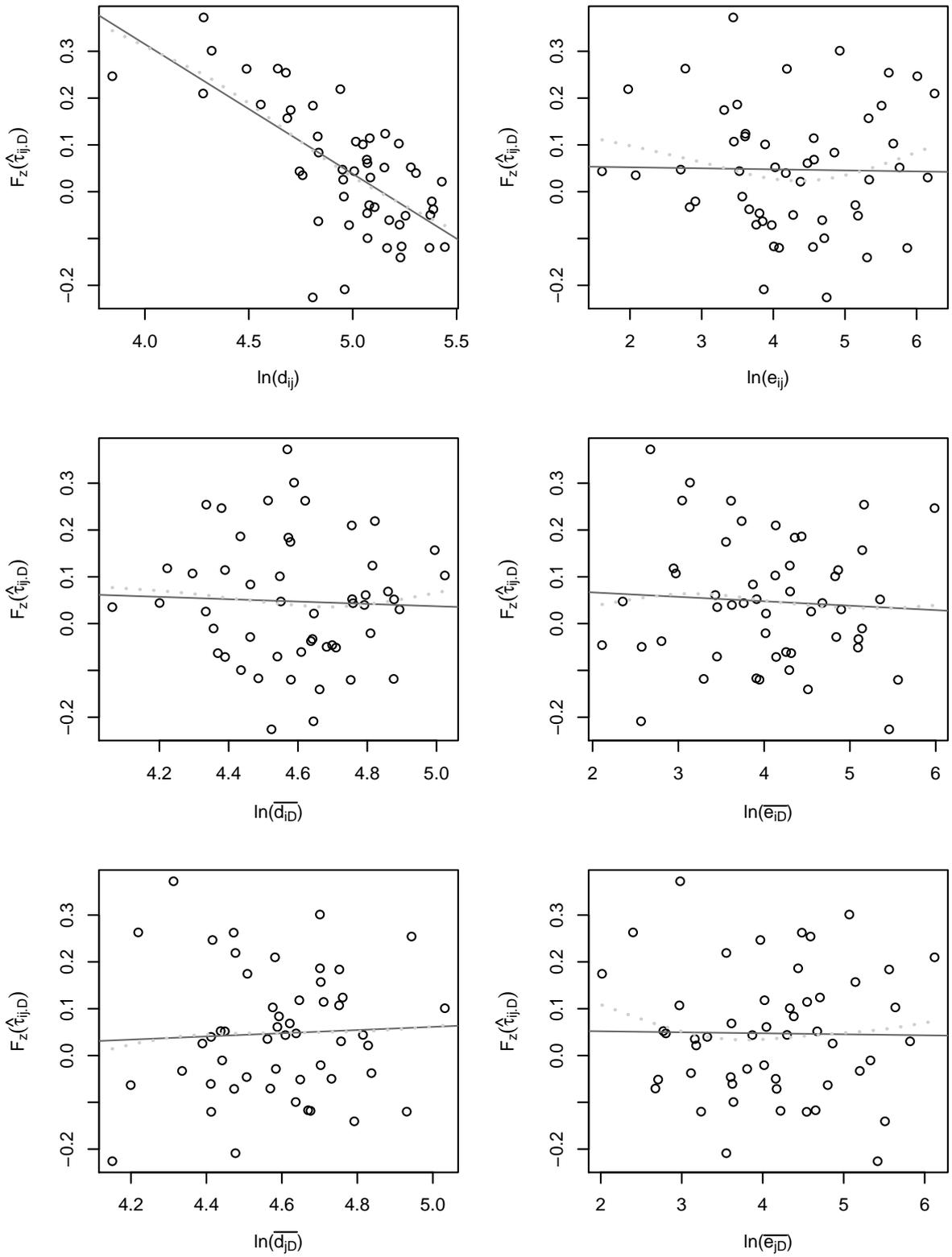
Figure 5.1.6: Scatter plots of $F_z\!\left(\widehat{\tau}_{i(e),j(e)|\mathcal{D}_e}\right)$ against $\ln\left(d_{i(e),j(e)}\right)$, $\ln\left(e_{i(e),j(e)}\right)$, $\ln\left(\overline{d_{i(e),\mathcal{D}_e}}\right)$, $\ln\left(\overline{e_{i(e),\mathcal{D}_e}}\right)$, $\ln\left(\overline{d_{j(e),\mathcal{D}_e}}\right)$ and $\ln\left(\overline{e_{j(e),\mathcal{D}_e}}\right)$, respectively, where $e \in \mathcal{E}_3$ (Tree 3).
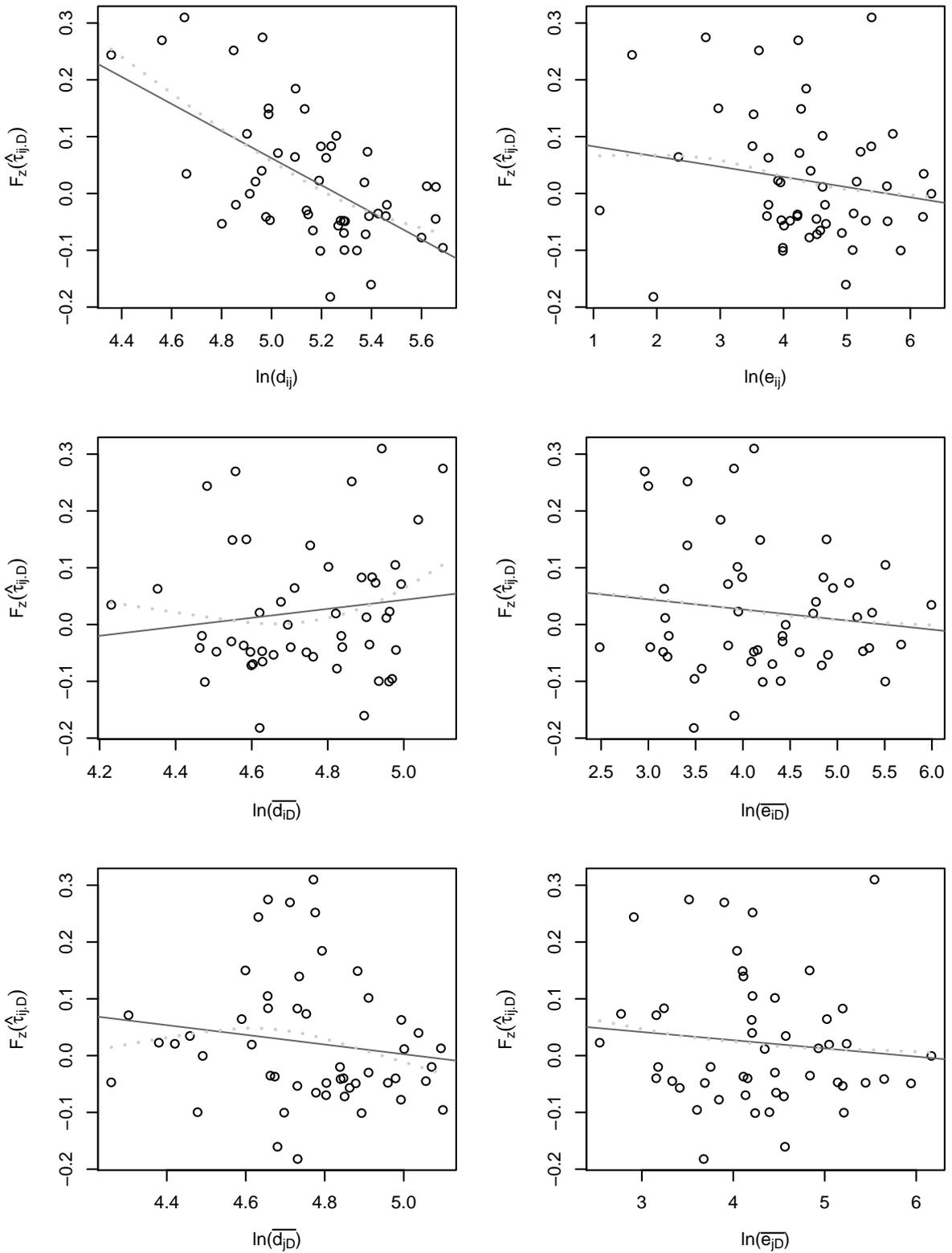
Figure 5.1.7: Scatter plots of $F_z(\widehat{\tau}_{i(e),j(e)|\mathcal{D}_e})$ against $\ln\big(d_{i(e),j(e)}\big)$, $\ln\big(e_{i(e),j(e)}\big)$, $\ln\big(\overline{d_{i(e),\mathcal{D}_e}}\big)$, $\ln\big(\overline{e_{i(e),\mathcal{D}_e}}\big)$, $\ln\big(\overline{d_{j(e),\mathcal{D}_e}}\big)$ and $\ln\big(\overline{e_{j(e),\mathcal{D}_e}}\big)$, respectively, where $e \in \mathcal{E}_4$ (Tree 4).
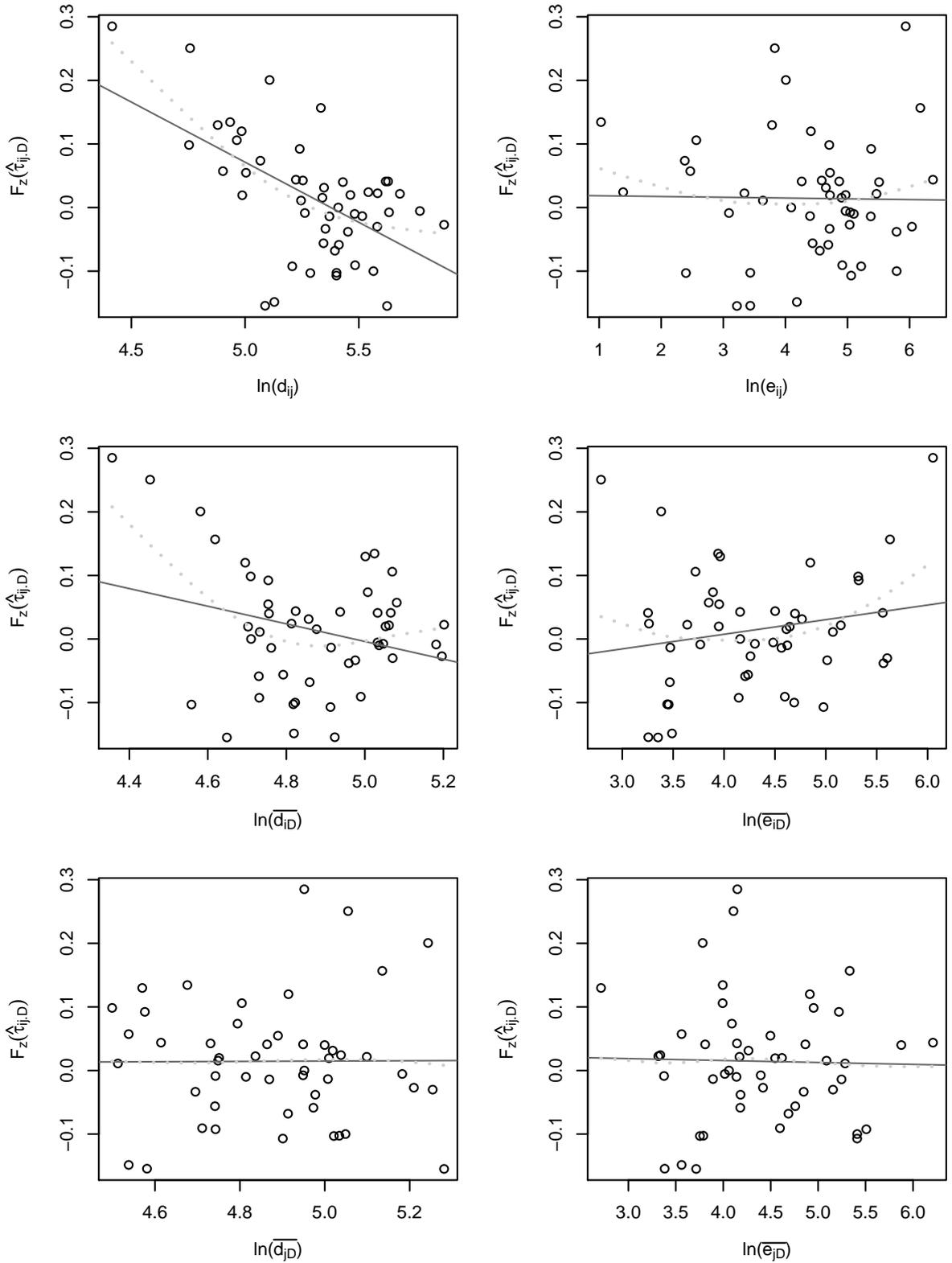
Figure 5.1.8: Scatter plots of $F_z(\widehat{\tau}_{i(e),j(e)|\mathcal{D}_e})$ against $\ln\left(d_{i(e),j(e)}\right)$, $\ln\left(e_{i(e),j(e)}\right)$, $\ln\left(\overline{d_{i(e),\mathcal{D}_e}}\right)$, $\ln\left(\overline{e_{i(e),\mathcal{D}_e}}\right)$, $\ln\left(\overline{d_{j(e),\mathcal{D}_e}}\right)$ and $\ln\left(\overline{e_{j(e),\mathcal{D}_e}}\right)$, respectively, where $e \in \mathcal{E}_5$ (Tree 5).

with

$$\boldsymbol{\beta}_{\text{full},1} = \left(\beta_{1,0}^{\text{full}}, \beta_{1,1}^{\text{full}}, \beta_{1,2}^{\text{full}}\right)^{\top} \in \mathbb{R}^3,$$

and where

$$\begin{aligned} h_l^{\text{full}} \left(e|\boldsymbol{\beta}_{\text{full},l}\right) &= \beta_{l,0}^{\text{full}} + \beta_{l,1}^{\text{full}} \ln\left(d_{i(e),j(e)}\right) + \beta_{l,2}^{\text{full}} \ln\left(e_{i(e),j(e)}\right) \\ &\quad + \beta_{l,3}^{\text{full}} \ln\left(\overline{d_{i(e),\mathcal{D}_e}}\right) + \beta_{l,4}^{\text{full}} \ln\left(\overline{e_{i(e),\mathcal{D}_e}}\right) \\ &\quad + \beta_{l,5}^{\text{full}} \ln\left(\overline{d_{j(e),\mathcal{D}_e}}\right) + \beta_{l,6}^{\text{full}} \ln\left(\overline{e_{j(e),\mathcal{D}_e}}\right), \quad e \in \mathcal{E}_l, \quad l = 2,\dots,10, \end{aligned}$$

with

$$\boldsymbol{\beta}_{\text{full},l} = \left(\beta_{l,0}^{\text{full}}, \beta_{l,1}^{\text{full}}, \dots, \beta_{l,6}^{\text{full}}\right)^{\top} \in \mathbb{R}^7, \quad l = 2,\dots,10.$$

All parameters of the full reparametrization are summarized in the vector

$$\boldsymbol{\beta}_{\text{full}}^{\text{SV}} = \left(\boldsymbol{\beta}_{\text{full},1}^{\top}, \dots, \boldsymbol{\beta}_{\text{full},10}^{\top}\right)^{\top} \in \mathbb{R}^{n_{\text{par}}^{\text{full}}},$$

where $n_{\text{par}}^{\text{full}} = 3 + 7 \cdot 9 = 66$.

All other reparametrizations which we are going to consider include only a part of the predictors included in the full reparametrization. The *distance reparametrization* (dist) for example includes all possible distances. It is given by

$$\begin{aligned} \widetilde{\theta}_{i(e),j(e)|\mathcal{D}_e}^{l} &= g_l^{\text{dist}}\left(d_{i(e),j(e)}, \overline{d_{i(e),\mathcal{D}_e}}, \overline{d_{j(e),\mathcal{D}_e}}|\boldsymbol{\beta}_{\text{dist},l}, t_{i(e),j(e)|\mathcal{D}_e}\right) \\ &= \mathrm{T}_{\tau \to \theta}\left(\mathrm{F}_z^{-1}\left(h_l^{\text{dist}}(e|\boldsymbol{\beta}_{\text{dist},l})\right); t_{i(e),j(e)|\mathcal{D}_e}\right), \quad e \in \mathcal{E}_l, \quad l = 1,\dots,10, \end{aligned}$$

where

$$h_1^{\text{dist}}\left(e|\boldsymbol{\beta}_{\text{dist},1}\right) = \beta_{1,0}^{\text{dist}} + \beta_{1,1}^{\text{dist}} \ln\left(d_{i(e),j(e)}\right), \quad e \in \mathcal{E}_1,$$

with

$$\boldsymbol{\beta}_{\text{dist},1} = \left(\beta_{1,0}^{\text{dist}}, \beta_{1,1}^{\text{dist}}\right)^{\top} \in \mathbb{R}^2,$$

and where

$$\begin{aligned} h_l^{\text{dist}}\left(e|\boldsymbol{\beta}_{\text{dist},l}\right) &= \beta_{l,0}^{\text{dist}} + \beta_{l,1}^{\text{dist}} \ln\left(d_{i(e),j(e)}\right) \\ &\quad + \beta_{l,2}^{\text{dist}} \ln\left(\overline{d_{i(e),\mathcal{D}_e}}\right) + \beta_{l,3}^{\text{dist}} \ln\left(\overline{d_{j(e),\mathcal{D}_e}}\right), \quad e \in \mathcal{E}_l, \quad l = 2,\dots,10, \end{aligned}$$

with

$$\boldsymbol{\beta}_{\text{dist},l} = \left(\beta_{l,0}^{\text{dist}}, \beta_{l,1}^{\text{dist}}, \beta_{l,2}^{\text{dist}}, \beta_{l,3}^{\text{dist}}\right)^{\top} \in \mathbb{R}^4, \quad l = 2,\dots,10.$$

The parameters of the distance reparametrization are summarized as

$$\boldsymbol{\beta}_{\text{dist}}^{\text{SV}} = \left(\boldsymbol{\beta}_{\text{dist},1}^{\top}, \dots, \boldsymbol{\beta}_{\text{dist},10}^{\top}\right)^{\top} \in \mathbb{R}^{n_{\text{par}}^{\text{dist}}},$$

where $n_{\text{par}}^{\text{dist}} = 2 + 4 \cdot 9 = 38$.

We are going to reduce the number of parameters even more, i.e. we are going to investigate the following *d0 reparametrization* (d0), which involves only the distances between the pairs which are indicated by the conditioned set of the respective copulae.

$$\begin{aligned} \widetilde{\theta}_{i(e),j(e)|\mathcal{D}_e}^{l} &= g_l^{\text{d0}}\left(d_{i(e),j(e)}|\boldsymbol{\beta}_{\text{d0},l}, t_{i(e),j(e)|\mathcal{D}_e}\right) \\ &= \mathrm{T}_{\tau \to \theta}\left(\mathrm{F}_z^{-1}\left(h_l^{\text{d0}}(e|\boldsymbol{\beta}_{\text{d0},l})\right); t_{i(e),j(e)|\mathcal{D}_e}\right), \quad e \in \mathcal{E}_l, \quad l = 1,\dots,10, \end{aligned}$$

where

$$h_l^{\mathrm{d0}}\left(e|\boldsymbol{\beta}_{\mathrm{d0},l}\right) = \beta_{l,0}^{\mathrm{d0}} + \beta_{l,1}^{\mathrm{d0}} \ln\left(d_{i(e),j(e)}\right), \quad e \in \mathcal{E}_l, \quad l = 1,\ldots,10,$$

with

$$\boldsymbol{\beta}_{\mathrm{d0},l} = \left(\beta_{l,0}^{\mathrm{d0}}, \beta_{l,1}^{\mathrm{d0}}\right)^{\top} \in \mathbb{R}^2, \quad l = 1,\ldots,10.$$

The summary of the respective parameters is given by

$$\boldsymbol{\beta}_{\mathrm{d0}}^{\mathrm{SV}} = \left(\boldsymbol{\beta}_{\mathrm{d0},1}^{\top}, \ldots, \boldsymbol{\beta}_{\mathrm{d0},10}^{\top}\right)^{\top} \in \mathbb{R}^{n_{\mathrm{par}}^{\mathrm{d0}}},$$

where $n_{\mathrm{par}}^{\mathrm{d0}} = 2 \cdot 10 = 20$.

Furthermore we take the two reparametrizations called *elevation reparametrization* (elev) and *e0 reparametrization* (e0) into consideration. These are constructed in analogy to the distance and d0 reparametrization, respectively. Replacement of the distances in the distance and d0 reparametrization by the respective elevation differences yields the elevation and e0 reparametrization.

The last reparametrization of interest gets the name *d0+e0 reparametrization* (de). It is again a reparametrization which includes both, distances and elevation differences, but only those which are given by the pairs indicated by the conditioned set of the respective copulae. The reparametrization is given by

$$\begin{aligned}\widetilde{\theta}_{i(e),j(e)|\mathcal{D}_e}^{l} &= g_l^{\mathrm{de}}\left(d_{i(e),j(e)}, e_{i(e),j(e)}|\boldsymbol{\beta}_{\mathrm{de},l}, t_{i(e),j(e)|\mathcal{D}_e}\right) \\ &= \mathrm{T}_{\tau\to\theta}\left(\mathrm{F}_z^{-1}\left(h_l^{\mathrm{de}}(e|\boldsymbol{\beta}_{\mathrm{de},l})\right); t_{i(e),j(e)|\mathcal{D}_e}\right), \quad e \in \mathcal{E}_l, \quad l = 1,\ldots,10,\end{aligned}$$

where

$$h_l^{\mathrm{de}}\left(e|\boldsymbol{\beta}_{\mathrm{de},l}\right) = \beta_{l,0}^{\mathrm{de}} + \beta_{l,1}^{\mathrm{de}} \ln\left(d_{i(e),j(e)}\right) + \beta_{l,2}^{\mathrm{de}} \ln\left(e_{i(e),j(e)}\right), \quad e \in \mathcal{E}_l, \quad l = 1,\ldots,10,$$

with

$$\boldsymbol{\beta}_{\mathrm{de},l} = \left(\beta_{l,0}^{\mathrm{de}}, \beta_{l,1}^{\mathrm{de}}, \beta_{l,2}^{\mathrm{de}}\right)^{\top} \in \mathbb{R}^3, \quad l = 1,\ldots,10.$$

The occurring parameters are summarized as

$$\boldsymbol{\beta}_{\mathrm{de}}^{\mathrm{SV}} = \left(\boldsymbol{\beta}_{\mathrm{de},1}^{\top}, \ldots, \boldsymbol{\beta}_{\mathrm{de},10}^{\top}\right)^{\top} \in \mathbb{R}^{n_{\mathrm{par}}^{\mathrm{de}}}$$

where $n_{\mathrm{par}}^{\mathrm{de}} = 3 \cdot 10 = 30$.

Based on the above reparametrizations, six different *spatial R-vine models* arise. In order to be able to estimate the respective model parameters, we have to specify the corresponding likelihood, since parameter estimation is done via maximum-likelihood estimation. Applying the reparametrization (nu) for the degrees of freedom $\widetilde{\nu}_{i(e),j(e)|\mathcal{D}_e}^{l}$ and selecting a reparametrization (r) for $\widetilde{\theta}_{i(e),j(e)|\mathcal{D}_e}^{l}$, out of the six reparametrizations (full), (dist), (d0), (elev), (e0) or (de), the likelihood is given through

$$\mathcal{L}_{\mathrm{SV}}\left(\boldsymbol{\beta}_r^{\mathrm{SV}}, \boldsymbol{\beta}_\nu^{\mathrm{SV}} \mid \boldsymbol{u}^1, \ldots, \boldsymbol{u}^d\right) = \prod_{t=1}^{N}\prod_{l=1}^{10}\prod_{e \in \mathcal{E}_l} c_{i(e),j(e);\mathcal{D}_e}\left(\widetilde{u}_t^{i(e)}, \widetilde{u}_t^{j(e)}; \widetilde{\theta}_{i(e),j(e)|\mathcal{D}_e}^{l}, \widetilde{\nu}_{i(e),j(e)|\mathcal{D}_e}^{l}, t_{i(e),j(e)|\mathcal{D}_e}\right),$$

where the transformed variables $\widetilde{u}_t^{i(e)}$ and $\widetilde{u}_t^{j(e)}$ are calculated according to

$$\widetilde{u}_t^{i(e)} = F_{i(e)|\mathcal{D}_e}(u_t^{i(e)} \,|\, \boldsymbol{u}_t^{\mathcal{D}_e}),$$
$$\widetilde{u}_t^{j(e)} = F_{j(e)|\mathcal{D}_e}(u_t^{j(e)} \,|\, \boldsymbol{u}_t^{\mathcal{D}_e}),$$

with

$$\boldsymbol{u}_t^{\mathcal{D}_e} := \left\{ u_t^s : s \in \mathcal{D}_e \right\}.$$

Numerical maximization of the log-likelihood

$$\ell_{\mathrm{SV}}\left(\boldsymbol{\beta}_{\mathrm{r}}^{\mathrm{SV}}, \boldsymbol{\beta}_{\nu}^{\mathrm{SV}} \,|\, \boldsymbol{u}^1, \ldots, \boldsymbol{u}^d\right) = \ln \mathcal{L}_{\mathrm{SV}}\left(\boldsymbol{\beta}_{\mathrm{r}}^{\mathrm{SV}}, \boldsymbol{\beta}_{\nu}^{\mathrm{SV}} \,|\, \boldsymbol{u}^1, \ldots, \boldsymbol{u}^d\right) \tag{5.2.2}$$

yields the maximum-likelihood estimates (mle)

$$\widehat{\boldsymbol{\beta}}_{\mathrm{mle}}^{\mathrm{SV}} = \left(\widehat{\boldsymbol{\beta}}_{\mathrm{r}}^{\mathrm{SV}}, \widehat{\boldsymbol{\beta}}_{\nu}^{\mathrm{SV}}\right)^{\top} \in \mathbb{R}^{(n_{\mathrm{par}}^{\mathrm{r}} + n_{\mathrm{par}}^{\nu})}.$$

In the rest of Chapter 5, we will address the above specified spatial R-vine models according to the selection of the reparametrization as full, distance, d0, elevation, e0, respectively d0+e0 model.

5.3 A spatial R-vine model for mean temperature

In this section we are going to investigate a spatial R-vine model for the mean temperature data set. To this end we aim to select a suitable reparametrization of an R-vine copula on this data set, which is truncated after tree ten and allows for Gaussian, Student-t, Clayton, Gumbel and Frank copulae, as well as rotated versions thereof. We are going to select the reparametrization among the six reparametrizations presented in the previous section.

5.3.1 Model selection

The first step of our model selection is the determination of the degree q for the reparametrization of the degrees of freedom. The analysis of the lower plot of Figure 5.1.3 in Section 5.1 makes us expect that a degree equal to two might be suitable. Table 5.2 confirms this expectation. It summarizes the parameter estimates and their estimated standard errors, Wald statistics and p-values of a fit of the linear model (5.2.1) in the case $q = 2$. The respective numbers for the case $q = 3$ are given in Table 5.3. Whereas some of the parameter estimates in Table 5.3 are insignificant at a significance level of $\alpha = 5\%$, all estimates in Table 5.2 are significant, i.e. the degree q is chosen to be 2 and it holds $n_{\mathrm{par}}^{\nu} = 3$. The estimates in Table 5.2 can be used as start parameters for $\boldsymbol{\beta}_{\nu}$, when the log-likelihood (5.2.2) of the reparametrized R-vine model is going to be maximized.

One principal purpose of our reparametrization is the reduction of the number of parameters. Whereas the R-vine on the mean temperature data truncated after tree ten is parametrized by 733 parameters, the numbers of parameters of the six models presented in the previous section are considerably smaller. The respective numbers are summarized

	Estimate	Std. Error	t-value	p-value
β_0^ν	1.8424	0.0645	28.55	0.0000
β_1^ν	0.1848	0.0324	5.71	0.0000
β_2^ν	-0.0112	0.0032	-3.54	0.0005

Table 5.2: Parameter estimates, estimated standard errors, Wald statistics and p-values of the polynomial model for the second parameters $\nu_{i(e),j(e)|\mathcal{D}_e}^l$, $e \in \mathcal{E}_l$, $l = 1, \ldots, 10$ ($R^2 = 0.2949$, $R_{\text{adj}}^2 = 0.2891$).

	Estimate	Std. Error	t-value	p-value
β_0^ν	1.7350	0.1059	16.39	0.0000
β_1^ν	0.2929	0.0904	3.24	0.0014
β_2^ν	-0.0369	0.0203	-1.82	0.0707
β_3^ν	0.0017	0.0013	1.28	0.2021

Table 5.3: Parameter estimates, estimated standard errors, Wald statistics and p-values of a polynomial model for the second parameters $\nu_{i(e),j(e)|\mathcal{D}_e}^l$, $e \in \mathcal{E}_l$, $l = 1, \ldots, 10$, with non-significant parameter ($R^2 = 0.2996$, $R_{\text{adj}}^2 = 0.2910$).

in Table 5.4. The last row of the table gives the total number of parameters needed, if the reparametrization according to the respective model given in the header of the table is chosen.

Table 5.5 gives the cumulated number of parameters for the different models. It shows how the parameter numbers could be reduced further, if a lower truncation level would be chosen.

To be able to compare the explanatory power of the models under consideration, we fit the six linear models

$$F_z(\tau_{i(e),j(e)|\mathcal{D}_e}) = h_l^{\text{full}}(e|\boldsymbol{\beta}_{\text{full},l}) + \varepsilon_{i(e),j(e)}, \quad \varepsilon_{i(e),j(e)} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2), \quad e \in \mathcal{E}_l,$$

$$F_z(\tau_{i(e),j(e)|\mathcal{D}_e}) = h_l^{\text{dist}}(e|\boldsymbol{\beta}_{\text{dist},l}) + \varepsilon_{i(e),j(e)}, \quad \varepsilon_{i(e),j(e)} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2), \quad e \in \mathcal{E}_l,$$

$$F_z(\tau_{i(e),j(e)|\mathcal{D}_e}) = h_l^{\text{d0}}(e|\boldsymbol{\beta}_{\text{d0},l}) + \varepsilon_{i(e),j(e)}, \quad \varepsilon_{i(e),j(e)} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2), \quad e \in \mathcal{E}_l,$$

$$F_z(\tau_{i(e),j(e)|\mathcal{D}_e}) = h_l^{\text{elev}}(e|\boldsymbol{\beta}_{\text{elev},l}) + \varepsilon_{i(e),j(e)}, \quad \varepsilon_{i(e),j(e)} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2), \quad e \in \mathcal{E}_l,$$

	full	distance	d0	elevation	e0	d0+e0
n_{par}^ν	3	3	3	3	3	3
Tree 1	3	2	2	2	2	3
Tree 2	7	4	2	4	2	3
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
Tree 10	7	4	2	4	2	3
Sum	**69**	**41**	**23**	**41**	**23**	**33**

Table 5.4: Number of parameters in case of the 'full', 'distance', 'd0', 'elevation', 'e0' and 'd0+e0' model.

# trees	full	distance	d0	elevation	e0	d0+e0
1	6	5	5	5	5	6
2	13	9	7	9	7	9
3	20	13	9	13	9	12
4	27	17	11	17	11	15
5	34	21	13	21	13	18
6	41	25	15	25	15	21
7	48	29	17	29	17	24
8	55	33	19	33	19	27
9	62	37	21	37	21	30
10	69	41	23	41	23	33

Table 5.5: Cumulated number of parameters for the 'full', 'distance', 'd0', 'elevation', 'e0' and 'd0+e0' model.

$$F_z(\tau_{i(e),j(e)|\mathcal{D}_e}) = h_l^{e0}\left(e|\boldsymbol{\beta}_{e0,l}\right) + \varepsilon_{i(e),j(e)}, \quad \varepsilon_{i(e),j(e)} \overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(0,\sigma^2\right), \quad e \in \mathcal{E}_l \quad \text{and}$$

$$F_z(\tau_{i(e),j(e)|\mathcal{D}_e}) = h_l^{de}\left(e|\boldsymbol{\beta}_{de,l}\right) + \varepsilon_{i(e),j(e)}, \quad \varepsilon_{i(e),j(e)} \overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(0,\sigma^2\right), \quad e \in \mathcal{E}_l,$$

separately for all trees $\mathcal{T}_l$, $l = 1, \ldots, 10$. The respective R^2 and R^2_{adj} are summarized in Table 5.6 and Table 5.7 for comparison. Furthermore, they are visualized in Figure 5.3.1. Note moreover, that the estimates of the above model fittings can be used as start parameters for $\boldsymbol{\beta}_r$, when (5.2.2) is maximized in order to fit a spatial R-vine model on the mean temperature data.

Tree	full	distance	d0	elevation	e0	d0+e0
1	0.5507	0.4992	0.4992	0.0608	0.0608	0.5507
2	0.6935	0.6275	0.2642	0.1259	0.1136	0.3633
3	0.6069	0.5643	0.4638	0.0045	0.0003	0.4683
4	0.6099	0.5964	0.3603	0.0336	0.0307	0.3693
5	0.4862	0.4415	0.3271	0.0733	0.0002	0.3402
6	0.4269	0.4094	0.2610	0.0559	0.0294	0.2616
7	0.3895	0.2977	0.1875	0.0719	0.0626	0.2807
8	0.5540	0.4929	0.1902	0.0477	0.0004	0.2125
9	0.6132	0.5645	0.3810	0.0214	0.0002	0.4038
10	0.2987	0.2290	0.1631	0.0771	0.0024	0.1671

Table 5.6: R^2 for the 'full', 'distance', 'd0', 'elevation', 'e0' and 'd0+e0' model.

From the plots of Figure 5.3.1 we observe clearly, that the models which include only elevation differences and no distances perform considerably bad, i.e. the explanatory power of the elevation predictors appears to be only marginal. On the other hand we see that the explanatory power of the distance model seems to be close to the explanatory power of the full model. In terms of the R^2_{adj} the distance model is even superior compared to the full model for two trees. The performance of the d0 and the d0+e0 model cannot keep up with the performance of the distance and the full model. Due to the fact that the
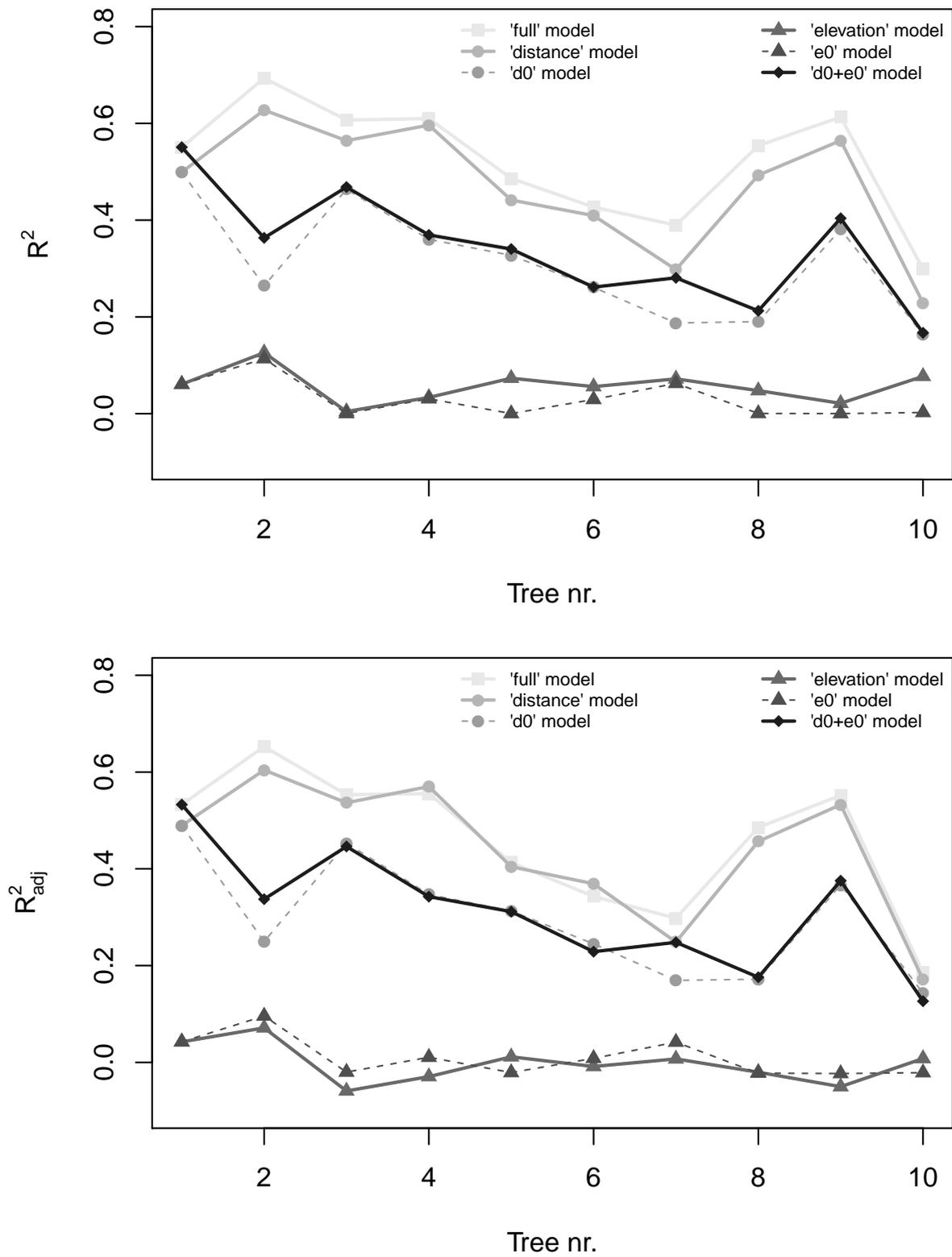
Figure 5.3.1: Tree-wise R^2 and R^2_{adj} for different kinds of spatial R-Vine models.

Tree	full	distance	d0	elevation	e0	d0+e0
1	0.5328	0.4894	0.4894	0.0424	0.0424	0.5328
2	0.6526	0.6042	0.2495	0.0713	0.0958	0.3374
3	0.5532	0.5365	0.4529	-0.0591	-0.0201	0.4462
4	0.5555	0.5701	0.3470	-0.0295	0.0105	0.3425
5	0.4128	0.4043	0.3128	0.0115	-0.0211	0.3115
6	0.3430	0.3691	0.2449	-0.0085	0.0083	0.2287
7	0.2979	0.2487	0.1694	0.0071	0.0418	0.2480
8	0.4854	0.4567	0.1718	-0.0204	-0.0223	0.1759
9	0.5521	0.5326	0.3666	-0.0502	-0.0231	0.3754
10	0.1850	0.1712	0.1432	0.0079	-0.0213	0.1264

Table 5.7: R^2_{adj} for the 'full', 'distance', 'd0', 'elevation', 'e0' and 'd0+e0' model.

selection of the distance model in comparison to the full model yields a reduction in the parameters from 69 to 41, and that the distance model performs nearly as good as the full model, we decide to choose the distance model as our *final model*.

5.3.2 Model fit

Based on the selected reparametrization, the distance reparametrization (dist), we are now ready to fit our model via maximum likelihood estimation.

For this purpose we have to weaken the copula family specification of the truncated R-vine under consideration in terms of family rotation, since the reparametrized parameter $\widetilde{\theta}^l_{i(e),j(e)|\mathcal{D}_e}$ may change its sign during the numerical optimization procedure. The final parameter estimates will determine the rotation of the corresponding families.

Finally model fitting is performed by maximization of the log-likelihood (5.2.2), where the distance model (dist) is selected for (r). For the mean temperature data set we obtain the parameter estimates $\widehat{\boldsymbol{\beta}}^{\mathrm{SV}}_{\mathrm{mle}} = (\widehat{\boldsymbol{\beta}}^{\mathrm{SV}}_{\mathrm{dist}}, \widehat{\boldsymbol{\beta}}^{\mathrm{SV}}_{\nu})^{\top} \in \mathbb{R}^{(n^{\mathrm{dist}}_{\mathrm{par}}+n^{\nu}_{\mathrm{par}})}$, which are summarized in Table 5.8. Additionally the start values, which were used for the maximum-likelihood estimation, are given.

A supplementary visualization of the estimates and start values, which cuts of the intercepts $\beta^{\mathrm{dist}}_{1,0}$, $\beta^{\mathrm{dist}}_{8,0}$ as well as β^{ν}_0, is presented in Figure 5.3.2. It is intended to help to detect possibly existing trends in the parameter estimates. To be able to do so, the estimates of the distance models' intercepts $\beta^{\mathrm{dist}}_{l,0}$, $l = 1, \ldots, 10$, and the ones corresponding to the predictors $d_{i(e),j(e)}$, $\overline{d_{i(e),\mathcal{D}_e}}$ and $\overline{d_{j(e),\mathcal{D}_e}}$, are connected by lines, respectively. Obviously there is no simple trend for the intercepts, but it may be possible to find adequate linear or quadratic reparametrizations of the remaining parameters $\beta^{\mathrm{dist}}_{l,1}$, $\beta^{\mathrm{dist}}_{l,2}$ and $\beta^{\mathrm{dist}}_{l,3}$ across all ten trees $l = 1, \ldots, 10$, in order to further reduce the number of parameters. This is left open and we are contented with our model at the moment.

Furthermore we note from the table and the figure that the used start values lie pretty close to their respective final estimates. This fact is also reflected in the small difference of the initial value of the log-likelihood 54533.74 and the final maximum log-likelihood $\ell_{\mathrm{dist}}\left(\widehat{\boldsymbol{\beta}}^{\mathrm{SV}}_{\mathrm{dist}}, \widehat{\boldsymbol{\beta}}^{\mathrm{SV}}_{\nu} \mid \boldsymbol{u}^1, \ldots, \boldsymbol{u}^d\right) = 54650.11$. Hence the calculation of start values according

no.	par	start	mle	no.	par	start	mle	no.	par	start	mle
1	$\beta_{1,0}^{\mathrm{dist}}$	2.056	1.965	15	$\beta_{5,0}^{\mathrm{dist}}$	0.011	-0.014	29	$\beta_{8,2}^{\mathrm{dist}}$	0.233	0.185
2	$\beta_{1,1}^{\mathrm{dist}}$	-0.264	-0.245	16	$\beta_{5,1}^{\mathrm{dist}}$	-0.289	-0.295	30	$\beta_{8,3}^{\mathrm{dist}}$	0.201	0.207
3	$\beta_{2,0}^{\mathrm{dist}}$	0.095	0.165	17	$\beta_{5,2}^{\mathrm{dist}}$	0.114	0.125	31	$\beta_{9,0}^{\mathrm{dist}}$	-0.403	-0.393
4	$\beta_{2,1}^{\mathrm{dist}}$	-0.405	-0.346	18	$\beta_{5,3}^{\mathrm{dist}}$	0.201	0.203	32	$\beta_{9,1}^{\mathrm{dist}}$	-0.226	-0.223
5	$\beta_{2,2}^{\mathrm{dist}}$	0.238	0.200	19	$\beta_{6,0}^{\mathrm{dist}}$	-0.303	-0.308	33	$\beta_{9,2}^{\mathrm{dist}}$	0.121	0.136
6	$\beta_{2,3}^{\mathrm{dist}}$	0.208	0.169	20	$\beta_{6,1}^{\mathrm{dist}}$	-0.227	-0.191	34	$\beta_{9,3}^{\mathrm{dist}}$	0.204	0.184
7	$\beta_{3,0}^{\mathrm{dist}}$	0.198	0.227	21	$\beta_{6,2}^{\mathrm{dist}}$	0.187	0.165	35	$\beta_{10,0}^{\mathrm{dist}}$	-0.133	-0.162
8	$\beta_{3,1}^{\mathrm{dist}}$	-0.334	-0.297	22	$\beta_{6,3}^{\mathrm{dist}}$	0.124	0.107	36	$\beta_{10,1}^{\mathrm{dist}}$	-0.122	-0.100
9	$\beta_{3,2}^{\mathrm{dist}}$	0.131	0.088	23	$\beta_{7,0}^{\mathrm{dist}}$	-0.458	-0.463	37	$\beta_{10,2}^{\mathrm{dist}}$	0.035	0.018
10	$\beta_{3,3}^{\mathrm{dist}}$	0.197	0.196	24	$\beta_{7,1}^{\mathrm{dist}}$	-0.205	-0.204	38	$\beta_{10,3}^{\mathrm{dist}}$	0.123	0.122
11	$\beta_{4,0}^{\mathrm{dist}}$	-0.298	-0.293	25	$\beta_{7,2}^{\mathrm{dist}}$	0.140	0.165	39	β_{0}^{ν}	1.842	1.618
12	$\beta_{4,1}^{\mathrm{dist}}$	-0.376	-0.412	26	$\beta_{7,3}^{\mathrm{dist}}$	0.174	0.148	40	β_{1}^{ν}	0.185	0.247
13	$\beta_{4,2}^{\mathrm{dist}}$	0.308	0.304	27	$\beta_{8,0}^{\mathrm{dist}}$	-1.129	-0.946	41	β_{2}^{ν}	-0.011	-0.016
14	$\beta_{4,3}^{\mathrm{dist}}$	0.168	0.213	28	$\beta_{8,1}^{\mathrm{dist}}$	-0.196	-0.191				

Table 5.8: Start values and maximum-likelihood estimates of the spatial R-vine model for the mean temperature data.
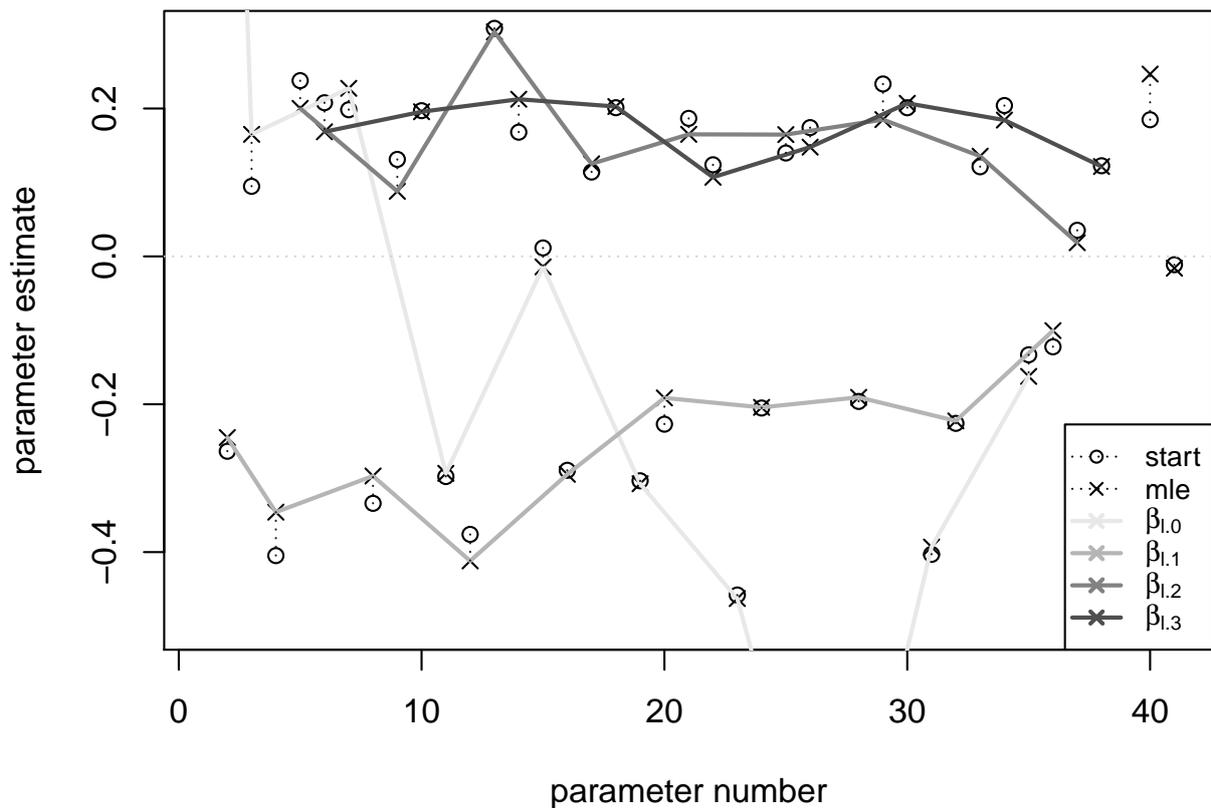
Figure 5.3.2: Visualization of the maximum-likelihood estimates of the spatial R-vine model for the mean temperature data.

to the previous subsection yields a model which is already close to the finally estimated one.

Figure 5.3.3 illustrates the differences between the Kendall's τ's occurring in our newly estimated spatial R-vine model, which are calculated as

$$\widehat{\tau}_{i(e),j(e)|\mathcal{D}_e}^{\text{dist}} = \text{F}_z^{-1}\left(h_l^{\text{dist}}(e|\widehat{\boldsymbol{\beta}}_{\text{dist},l})\right), \quad e \in \mathcal{E}_l, \quad l = 1,\ldots,10,$$

and the respective Kendall's τ's

$$\widehat{\tau}_{i(e),j(e)|\mathcal{D}_e}, \quad e \in \mathcal{E}_l, \quad l = 1,\ldots,10,$$

of the original R-vine model on which the development of the spatial R-vine model was performed.

The upper part of Figure 5.3.3 is a plot of $\widehat{\tau}_{i(e),j(e)|\mathcal{D}_e}^{\text{dist}}$ against $\widehat{\tau}_{i(e),j(e)|\mathcal{D}_e}$, $e \in \mathcal{E}_l$, $l = 1,\ldots,10$. If the correlations of the bivariate building blocks of our model were completely explained by our distance model, all points of the plot should lie on the angle bisector. That is of course not the case, but the points scatter closely around this line, which supports the adequacy of our reparametrization. It is however conspicuous that our model tends to estimate $\widehat{\tau}_{i(e),j(e)|\mathcal{D}_e}^{\text{dist}}$ close to zero, where negative Kendall's τ's $\widehat{\tau}_{i(e),j(e)|\mathcal{D}_e}$ occur.

In addition Figure 5.3.3 plots

$$\left|\widehat{\tau}_{i(e),j(e)|\mathcal{D}_e}^{\text{dist}} - \widehat{\tau}_{i(e),j(e)|\mathcal{D}_e}\right|, \quad e \in \mathcal{E}_l, \quad l = 1,\ldots,10,$$

against the respective tree number $l = 1,\ldots,10$, in its lower part. Moreover the tree-wise shares of differences smaller than 0.1 are given. We observe that the spatial R-vine model doesn't differ that much from the original R-vine model for the first tree. However about 23% of the correlations in tree two show a difference of more than 0.1 to the respective R-vine model correlations. For the remaining trees more than 80% of the Kendall's τ's exhibit a precision better than 0.1.

Finally we present an illustration of the dependencies modeled by the spatial R-vine model. Figure 5.3.4 shows all 54 observation stations of the mean temperature data set on which the spatial R-vine model is built and all edges that occur in the ten trees of the R-vine. The magnitude of correlation between station pairs which is modeled by the spatial R-vine model is indicated through edge width and edge color. The thicker and darker the edges are, the higher is the respective correlation. The resulting network gives an impression for which pairs dependencies are modeled and how strong they are and which pairs are treated as conditionally independent by the model. One clearly observes that the strongest dependencies are already captured in tree one, which is silhouetted against all other trees. Moreover, some stations like for instance *Berleburg, Bad-Stünzel* (3) or *Holzdorf (Flugplatz)* (23) play an important role for the modeling of dependencies, whereas the dependency of stations like *Göttingen* (19) or *Rostock-Warnemünde* (44) with other stations is mostly captured by the respective edges of the first tree. Note furthermore, that due to the selection of the first tree and the truncation after tree ten there are no edges between stations which are located relatively close to each other, e.g. stations *Jena (Sternwarte)* (24) and *Kronach* (29).
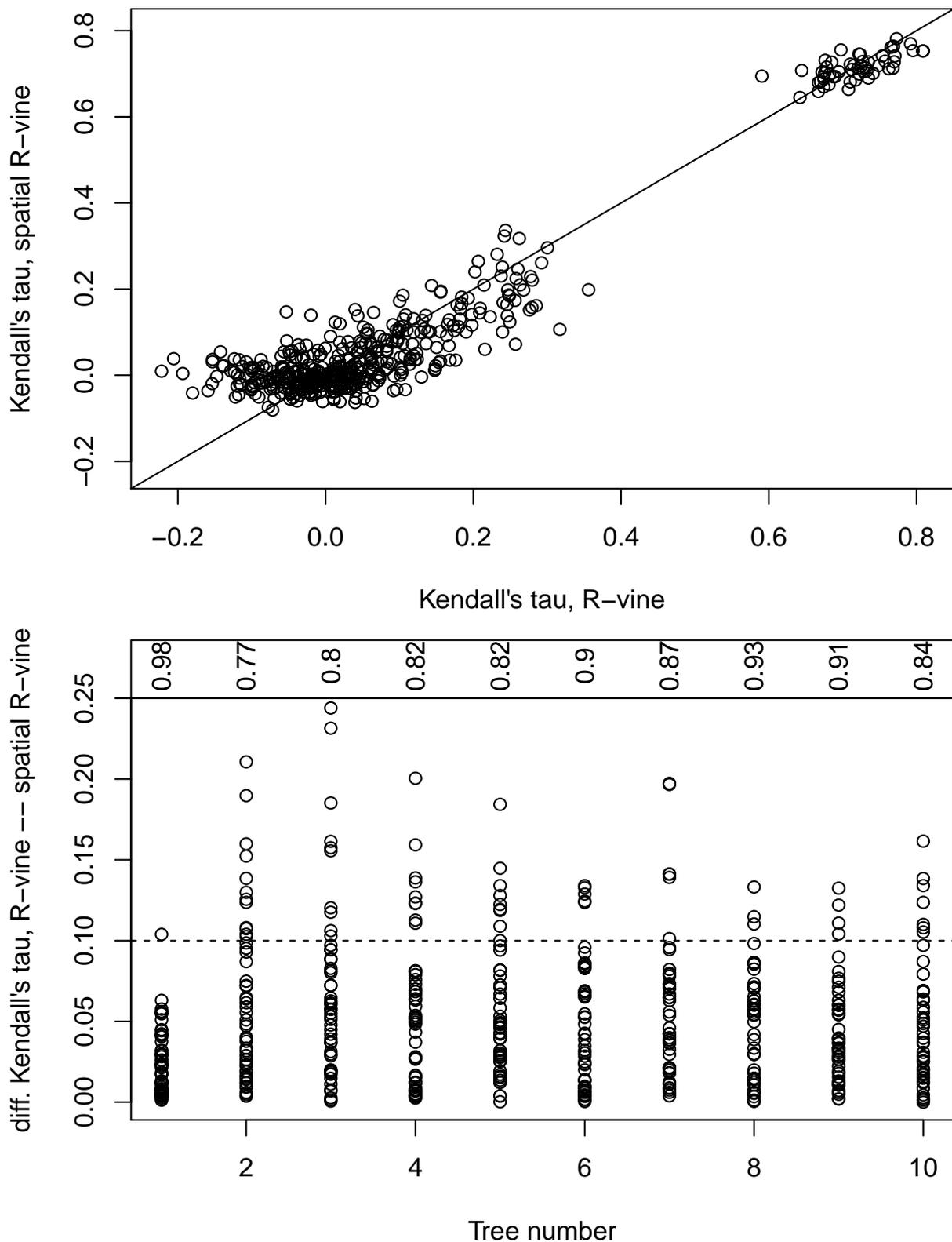
Figure 5.3.3: Upper: Comparison of the Kendall's τ's occurring in the spatial R-vine model and in the original R-vine. Lower: Absolute value of the difference between these two Kendall's τ's sorted by tree number. The tree-wise shares of differences smaller than 0.1 are given.
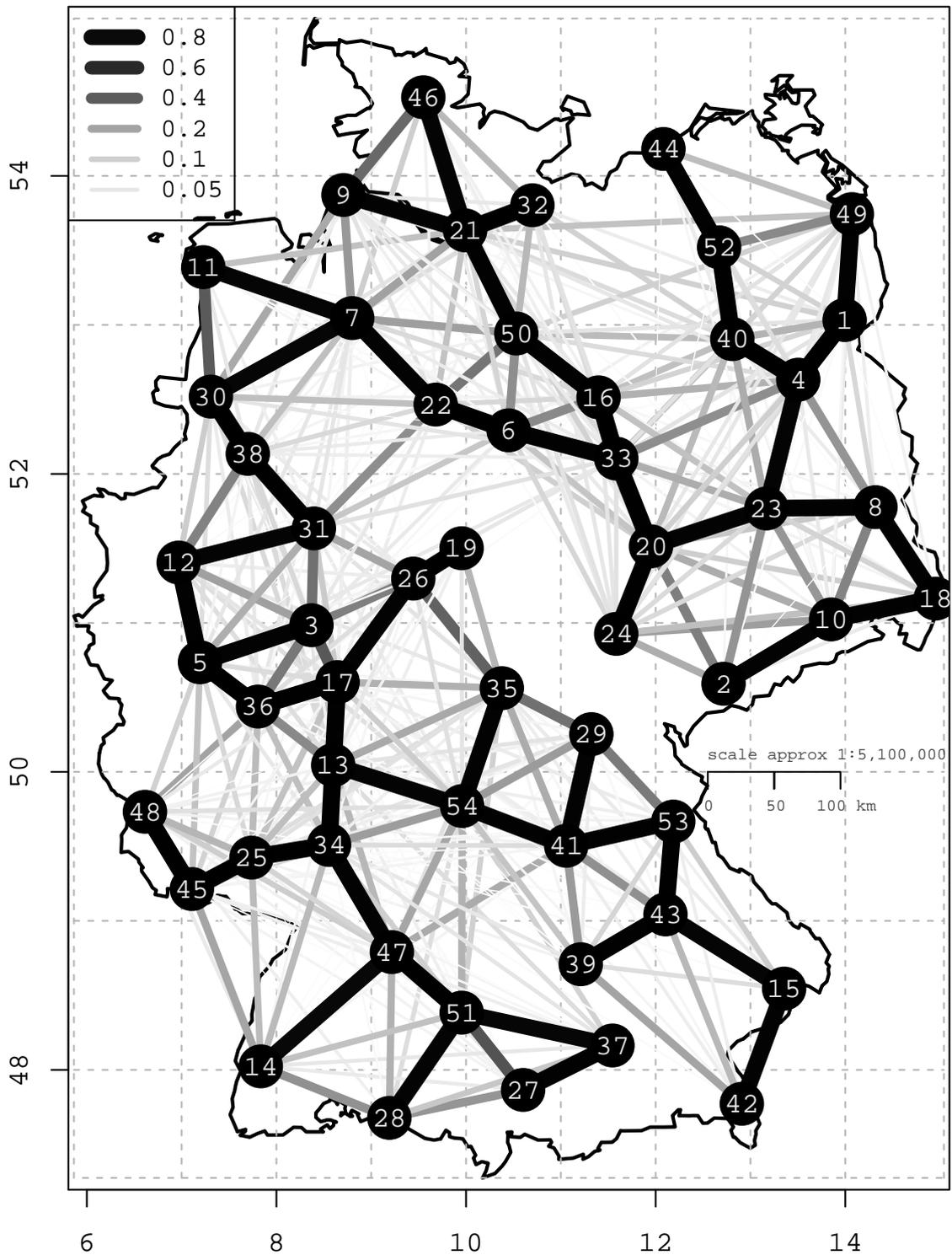
Figure 5.3.4: Visualization of the dependence structure in the estimated spatial R-vine model. The edges of all ten trees of the truncated R-vine are depicted. The thicker and darker the edges are, the higher is the respective correlation.

5.3.3 Prediction

Now that we have selected an adequate spatial R-vine model for the mean temperature data and the respective model parameters are estimated, we aim to predict mean temperatures at new locations based on the model fit. In order to be able to validate the outcome of these predictions, the locations from which we predict are chosen among observation stations where mean temperature data is available for the relevant period. The validation data set which we are going to take into consideration is composed out of the mean temperatures of 24 observation stations across Germany over the period 01/01/2010-12/31/2012. All stations of the validation data set get again their own identification number and a short name. Besides their names and IDs the respective stations' longitudes, latitudes and elevations are supplied in Table 5.9. Moreover the stations' locations are visualized on a map of Germany in Figure 5.3.5. The locations of the 54 stations composing the training data set are indicated by plus signs.

s/ ID	short name	full name	$x_{\text{long},s}$ (longitude)	$x_{\text{lat},s}$ (latitude)	$x_{\text{elev},s}$ (elevation)
55	albs	Albstadt-Badkap	8.98	48.22	759.00
56	alfe	Alfeld	9.80	51.97	143.90
57	arko	Arkona	13.44	54.68	42.00
58	arns	Arnsberg-Neheim	7.98	51.47	159.00
59	augs	Augsburg	10.94	48.43	461.40
60	blan	Blankenrath	7.31	50.04	417.00
61	bork	Borkum-Flugplatz	6.70	53.60	3.00
62	bvoe	Bremervörde	9.14	53.45	10.00
63	buch	Buchen, Kr. Neckar-Odenwald	9.32	49.52	340.00
64	cosc	Coschen	14.73	52.02	40.00
65	ebra	Ebrach	10.50	49.85	346.00
66	ellw	Ellwangen-Rindelbach	10.13	48.99	460.00
67	falk	Falkenberg, Kr.Rottal-Inn	12.73	48.48	472.00
68	gram	Grambek	10.68	53.58	27.00
69	garb	Großer Arber	13.14	49.11	1436.00
70	grue	Grünow	13.94	53.32	55.90
71	hohe	Hoherodskopf/Vogelsberg	9.22	50.51	743.30
72	luec	Lüchow	11.14	52.97	17.00
73	mitt	Mittenwald-Buckelwiesen	11.27	47.48	981.00
74	muel	Müllheim	7.64	47.81	273.00
75	neuh	Neuhaus am Rennweg	11.14	50.50	845.00
76	ohrz	Oberharz am Brocken-Stiege	10.88	51.67	494.00
77	rahd	Rahden-Varl	8.57	52.45	42.00
78	wies	Wiesenburg	12.46	52.12	187.00

Table 5.9: Observation stations (validation data): ID, short name, full name, longitude, latitude and elevation.
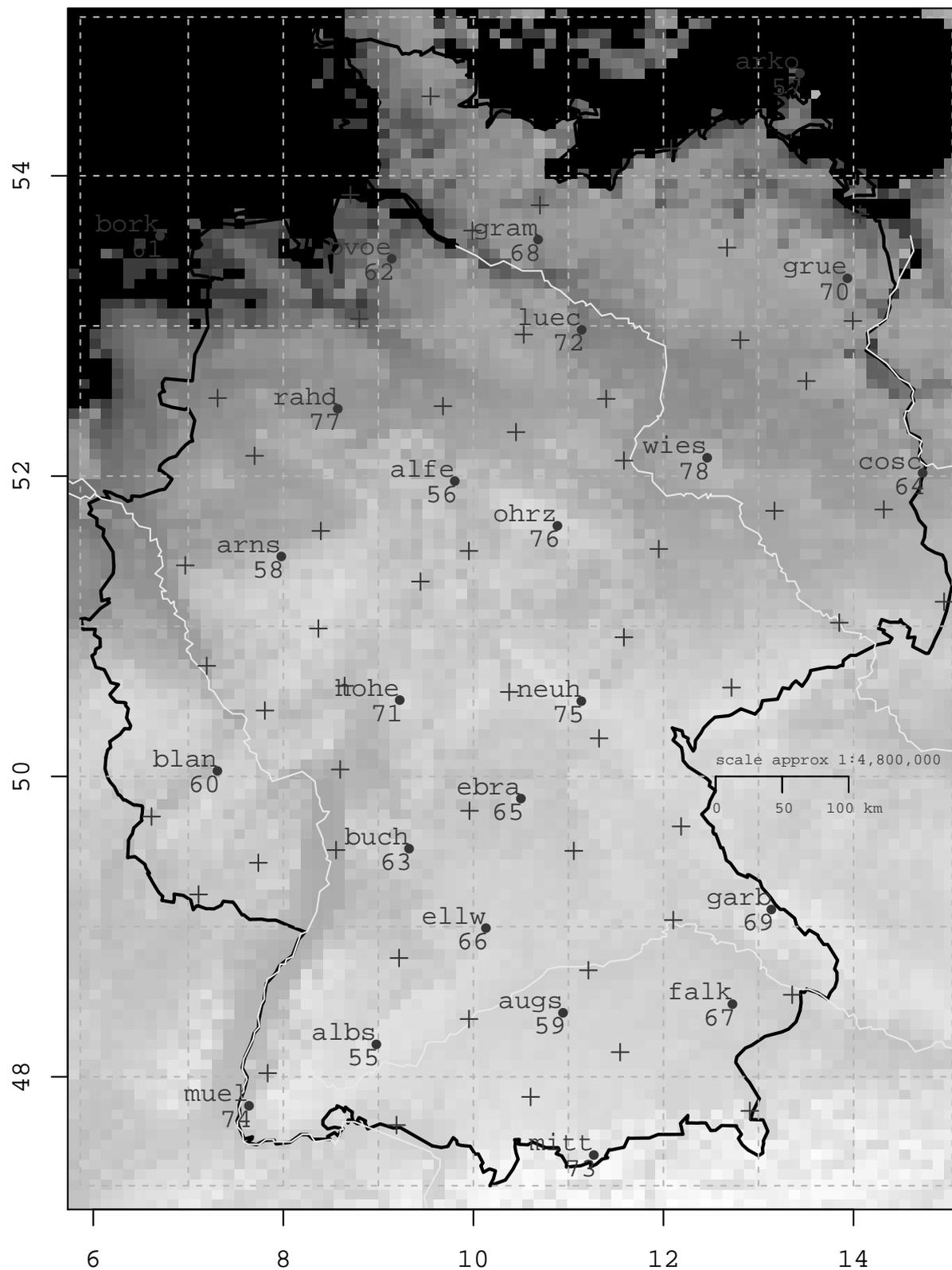
Figure 5.3.5: 24 locations across Germany where we will predict mean temperature time series for the period 01/01/2010-12/31/2012.

We have to keep in mind based on which data the models under consideration were developed and that these models might not be able to predict outside the scope of the training data set. Screening of the metadata supplied and visualized in Table 5.9 and Figure 5.3.5, respectively, yields that prediction might be problematic at some extreme locations which exceed the scope of the locations setting up the training data set. For example the elevations of *Albstadt-Badkap* (55), *Großer Arber* (69), *Hoherodskopf/Vogelsberg* (71), *Mittenwald-Buckelwiesen* (73) and *Neuhaus am Rennweg* (75) lie outside the range of elevations of the training data set, where the maximum elevation is 716 meters. Whereas *Mittenwald-Buckelwiesen* (73) lies comparatively extremely southern, *Arkona* (57) is located extremely northern on an island in the Baltic Sea.

Since the spatial R-vine model is constructed based on data transformed to copula data, predictions from this model will also be on a copula data level. Thus a back transformation to the original level of mean temperatures is needed, which is based on the joint marginal model developed in Chapter 3. Due to the polynomial structures (3.6.1) involved in the model, extreme locations s might yield meaningless values for the aggregated parameters $\widehat{\beta}_0(s)$, $\widehat{\beta}_{\sin}(s)$, $\widehat{\beta}_{\cos}(s)$, $\widehat{\gamma}_1(s)$, $\widehat{\gamma}_2(s)$, $\widehat{\gamma}_3(s)$, $\widehat{\xi}(s)$, $\widehat{\omega}(s)$, $\widehat{\alpha}(s)$ and $\widehat{\nu}(s)$. In order to achieve meaningful predictions, we are going to impose restrictions on the aggregated parameters calculated for new locations s, i.e. aggregated parameters which exceed the lower or the upper bound of a specific interval will be restricted to the value of the respective bound.[1]

Table 5.10 summarizes the aggregated parameters $\widehat{\beta}_0(s)$, $\widehat{\beta}_{\sin}(s)$, $\widehat{\beta}_{\cos}(s)$, $\widehat{\gamma}_1(s)$, $\widehat{\gamma}_2(s)$, $\widehat{\gamma}_3(s)$, $\widehat{\xi}(s)$, $\widehat{\omega}(s)$, $\widehat{\alpha}(s)$ and $\widehat{\nu}(s)$, which are calculated for the validation data set without any restrictions. Additionally the aggregated parameters $\widehat{\lambda}(s)$ and $\widehat{\delta}(s)$ are provided, which are calculated from $\widehat{\beta}_{\sin}(s)$ and $\widehat{\beta}_{\cos}(s)$ according to (3.2.2). Moreover the ranges $[\text{lb}(\widehat{\theta}), \text{ub}(\widehat{\theta})]$ of the respective aggregated parameters of the training data set are given, where we define

$$\text{lb}(\widehat{\theta}) \coloneqq \min_{s=1,\dots,54} \widehat{\theta}(s) \qquad \text{and} \qquad \text{ub}(\widehat{\theta}) \coloneqq \max_{s=1,\dots,54} \widehat{\theta}(s),$$

for all $\widehat{\theta} \in \{\widehat{\beta}_0, \widehat{\beta}_{\sin}, \widehat{\beta}_{\cos}, \widehat{\gamma}_1, \widehat{\gamma}_2, \widehat{\gamma}_3, \widehat{\xi}, \widehat{\omega}, \widehat{\alpha}, \widehat{\nu}, \widehat{\lambda}, \widehat{\delta}\}$. All aggregated parameters which lie outside of the respective range are highlighted. Especially for the stations *Großer Arber* (69), *Mittenwald-Buckelwiesen* (73) and *Neuhaus am Rennweg* (75), most of the aggregated parameters lie outside the range defined by the training data set.

For the purpose of back transformation of the copula data to the original mean temperature level, we are going to restrict the aggregated parameters to the ranges $[\text{lb}(\widehat{\theta}), \text{ub}(\widehat{\theta})]$ specified in the two bottom rows of Table 5.10, due to the insights we recently got. However we do not restrict $\widehat{\beta}_{\sin}(s)$ and $\widehat{\beta}_{\cos}(s)$ directly. Instead $\widehat{\lambda}(s)$ and $\widehat{\delta}(s)$ are restricted, since this kind of restriction yields more meaningful results for the seasonality component. Also $\widehat{\xi}(s)$ is restricted indirectly. Since the marginal model errors are assumed to have zero mean, we set $\widehat{\xi}(s) = -\widehat{\omega}(s)\widehat{\mu}(s)$, where $\widehat{\mu}(s)$ is calculated according to (3.4.1), which ensures the zero mean condition.

---

[1]In order to avoid such restrictions the aggregated parameters of the joint marginal model have to be modeled by means of a different approach which prohibits extreme parameter values for longitudes, latitudes and elevations which exceed the range of the training data set. One possible approach may be the modeling via B-splines.

s	$\widehat{\beta}_0(s)$	$\widehat{\beta}_{\sin}(s)$	$\widehat{\beta}_{\cos}(s)$	$\widehat{\gamma}_1(s)$	$\widehat{\gamma}_2(s)$	$\widehat{\gamma}_3(s)$	$\widehat{\xi}(s)$	$\widehat{\omega}(s)$	$\widehat{\alpha}(s)$	$\widehat{\nu}(s)$	$\widehat{\lambda}(s)$	$\widehat{\delta}(s)$
55	1.1331	-0.3982	-1.1212	1.0139	-0.2915	0.0555	*1.2291*	*1.8187*	*-1.3633*	18.8603	1.1898	*-1.9120*
56	1.2696	-0.3428	-1.2513	1.0076	-0.2650	0.0521	-0.2300	1.2921	0.2372	13.2563	1.2974	-1.8382
57	1.1060	-0.3964	-1.2910	0.9275	-0.1989	0.0601	-0.4835	1.1464	0.5132	10.8552	1.3505	-1.8687
58	1.3027	-0.3379	-1.1879	1.0203	-0.2935	0.0685	-0.1626	1.2714	0.1677	16.5775	1.2350	-1.8479
59	1.3419	-0.3531	-1.4680	0.9939	-0.2675	0.0541	0.6010	1.3746	-0.6002	9.7624	1.5098	-1.8068
60	1.2325	-0.3230	-1.2359	1.0662	*-0.3781*	0.0980	0.4564	1.2558	-0.4321	15.9371	1.2774	-1.8264
61	1.2320	-0.3690	*-1.0216*	0.9812	-0.2718	0.0931	-0.5525	1.1555	0.5995	10.2337	*1.0862*	*-1.9174*
62	1.2402	-0.3732	-1.1578	0.9690	-0.2382	0.0694	-0.5750	1.2086	0.6172	8.8461	1.2164	-1.8826
63	1.3374	-0.3279	-1.3625	1.0422	-0.3396	0.0839	0.2889	1.2429	-0.2767	12.1654	1.4014	-1.8070
64	1.3442	-0.3781	-1.4579	1.0129	-0.2971	0.0752	-0.3109	1.4610	0.3045	*26.9586*	1.5061	-1.8246
65	1.3031	-0.3380	-1.3976	1.0409	-0.3372	0.0799	0.2910	1.2713	-0.2769	13.4318	1.4379	-1.8081
66	1.2928	-0.3317	-1.4079	1.0299	-0.3182	0.0722	0.5655	1.3362	-0.5565	10.8634	1.4464	-1.8022
67	1.3286	-0.3598	-1.5308	1.0012	-0.2843	0.0599	0.6633	1.4178	-0.6703	11.5409	1.5726	-1.8016
68	1.2159	-0.3822	-1.2347	0.9570	-0.2254	0.0664	-0.5396	1.2022	0.5737	8.0906	1.2925	-1.8710
69	*0.5357*	*-3.9182*	*380.0520*	*1.1185*	*-0.4369*	0.0802	*2.6803*	*7.4065*	*-2.9194*	*2653.6833*	*380.0722*	*1.5811*
70	1.2168	-0.3831	-1.3729	0.9838	-0.2687	0.0755	-0.3735	1.2682	0.3824	11.5135	1.4254	-1.8429
71	*0.9411*	-0.3777	-1.1241	1.0703	-0.3667	0.0770	1.0956	*1.7855*	-1.1930	29.5974	1.1859	-1.8950
72	1.2770	-0.3667	-1.2735	0.9850	-0.2454	0.0545	-0.5283	1.3024	0.5580	11.1549	1.3252	-1.8511
73	*1.0289*	*-0.7329*	*8.9112*	*1.0650*	*-0.3714*	0.0701	*1.7411*	*2.4965*	*-2.0938*	*66.0644*	*8.9412*	*1.6529*
74	1.5409	-0.3596	-1.3545	0.9838	-0.2668	0.0653	0.2421	1.2060	-0.2579	10.3756	1.4014	-1.8303
75	*0.8640*	*-0.4699*	*0.2649*	*1.0731*	*-0.3697*	0.0716	*1.3219*	*2.1299*	*-1.5120*	*45.7792*	*0.5394*	*2.6283*
76	1.0285	-0.3496	-1.3067	1.0329	-0.2988	0.0521	0.5286	1.4250	-0.4970	13.8447	1.3526	-1.8323
77	1.3047	-0.3426	-1.2019	1.0059	-0.2662	0.0584	-0.4598	1.2975	0.4812	13.6346	1.2498	-1.8485
78	1.2228	-0.3551	-1.3431	1.0105	-0.2738	0.0514	-0.1025	1.3141	0.1074	13.2741	1.3893	-1.8292
lb	1.0006	-0.4180	-1.5736	0.9173	-0.3683	0.0476	-0.6007	1.0888	-1.2650	7.6031	1.1142	-1.9092
ub	1.5497	-0.3144	-1.0532	1.0575	-0.1817	0.1071	1.1570	1.7336	0.6501	26.0114	1.6202	-1.7962

Table 5.10: Aggregated marginal model parameters calculated for the validation stations and parameter range for training data.

Having imposed these restrictions, the rest of the back transformation of the copula data $\widehat{u}_4^s, \ldots, \widehat{u}_N^s$ for a location s, which result from the respective prediction procedure, is nearly straightforward. Remember, that due to the inclusion of the three autoregression components in the marginal model there are no copula data available for the first three points in time.

The first step is a back transformation of the copula data to the marginal model residuals $\widehat{\varepsilon}_4^s, \ldots, \widehat{\varepsilon}_N^s$. This is achieved by means of the quantile function T_{skew}^{-1} of the skew-t distribution with (restricted) parameters $\widehat{\xi}^r(s)$, $\widehat{\omega}^r(s)$, $\widehat{\alpha}^r(s)$ and $\widehat{\nu}^r(s)$, i.e. we calculate

$$\widehat{\varepsilon}_t^s = T_{\text{skew}}^{-1}\left(\widehat{u}_t^s \mid \widehat{\xi}^r(s), \widehat{\omega}^r(s), \widehat{\alpha}^r(s), \widehat{\nu}^r(s)\right), \quad t = 4, \ldots, N.$$

The only difficulty which arises for the back transformation is due to the autoregression components in the marginal model. In order to be able to calculate the predictions $\widehat{y}_4^s, \ldots, \widehat{y}_N^s$, we have to determine meaningful start values for this time series for $t = 1, 2, 3$. We proceed by predicting $\widehat{y}_1^s$, $\widehat{y}_2^s$ and $\widehat{y}_3^s$ based on three linear models of the form

$$Y_t^s = \theta_0 + \theta_{\text{elev}} x_{\text{elev},s} + \theta_{\text{long}} x_{\text{long},s} + \theta_{\text{lat}} x_{\text{lat},s} + \varepsilon_t^s, \quad \varepsilon_t^s \overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(0, \sigma^2\right),$$

for $t = 1, 2, 3$.

In a last step we have to divide the inital predictions $\widehat{y}_1^s$, $\widehat{y}_2^s$ and $\widehat{y}_3^s$ by their respective weights $\widehat{w}_1$, $\widehat{w}_2$ and $\widehat{w}_3$ which results in $\widetilde{\widehat{y}}_1^s$, $\widetilde{\widehat{y}}_2^s$ and $\widetilde{\widehat{y}}_3^s$. Afterwards the weighted mean temperatures $\widetilde{\widehat{y}}_4^s, \ldots, \widetilde{\widehat{y}}_N^s$ can be calculated successively as

$$\widetilde{\widehat{y}}_t^s = \widehat{\beta}_0^r(s) + \widehat{\lambda}^r(s) \sin\left(\frac{2\pi t}{365.25} + \widehat{\delta}^r(s)\right) + \widehat{\gamma}_1^r(s)\widetilde{\widehat{y}}_{t-1}^s + \widehat{\gamma}_2^r(s)\widetilde{\widehat{y}}_{t-2}^s + \widehat{\gamma}_3^r(s)\widetilde{\widehat{y}}_{t-3}^s + \widehat{\varepsilon}_t^s,$$

where $t = 4, \ldots, N$ and $\widehat{\beta}_0^r(s), \widehat{\lambda}^r(s), \widehat{\delta}^r(s), \widehat{\gamma}_1^r(s), \widehat{\gamma}_2^r(s), \widehat{\gamma}_3^r(s)$ are the restricted aggregated parameter estimates for location s. Finally we obtain the unweighted mean temperatures as

$$\widehat{y}_t^s = \widetilde{\widehat{y}}_t^s \sqrt{\widehat{w}_t}, \quad t = 1, \ldots, N.$$

Thus we know how to transform from copula data back to mean temperatures, but until know nothing was said about how the actual prediction from the spatial R-vine model is conducted. We are going to explain this in the following as general as possible.

In order to predict mean temperatures respectively the corresponding copula data u_t^s of a location s for an arbitrary point in time t, we need to specify the conditional distribution $F_{s|1,\ldots,d}(u_t^s|u_t^1, \ldots, u_t^d)$ of the variable u_t^s conditioned on the known values $u_t^1, \ldots, u_t^d$ constituting the copula data at the point in time t corresponding to the training data set on which the spatial R-vine model is built. The spatial R-vine model specifies the joint distribution of $u_t^1, \ldots, u_t^d$, which is essentially an R-vine distribution. Following, access to the conditional distribution $F_{s|1,\ldots,d}(u_t^s|u_t^1, \ldots, u_t^d)$ can be achieved by extending the R-vine underlying the spatial R-vine model.

If one wants to preserve the structure of the underlying (truncated) R-vine, one has to add the new variable as a leaf to the first R-vine tree. To do so we calculate the Kendall's τ's

$$\widehat{\tau}_{i(e_1),j(e_1)|\mathcal{D}_{e_1}} = \widehat{\tau}_{i(e_1),j(e_1)} = \widehat{\tau}_{r,s} = \mathrm{F}_z^{-1}\left(h_1^{\text{mod}}\left(e_1 = \{r, s\}|\widehat{\boldsymbol{\beta}}_{\text{mod},1}\right)\right)$$

for all d edges $e_1 = \{i(e_1), j(e_1)\} = \{r, s\}$, $r = 1, \ldots, d$, which may be added to the first R-vine tree, where h_1^{mod} is the model function for the first parameters of the first R-vine tree corresponding to the chosen model. The edge e_1^* which yields the biggest Kendall's τ is selected to extend the first R-vine tree. For this edge a copula family $t_{i(e_1^*), j(e_1^*)}$ has to be selected. This selection may be based on the number of the different families constituting the original R-vine. Then the first copula parameter is calculated as

$$\widehat{\theta}_{i(e_1^*), j(e_1^*)|\mathcal{D}_{e_1^*}}^1 = \mathrm{T}_{\tau \to \theta}\left(\widehat{\tau}_{i(e_1^*), j(e_1^*)|\mathcal{D}_{e_1^*}}; t_{i(e_1^*), j(e_1^*)}\right). \tag{5.3.1}$$

If needed the second copula parameter $\widehat{\nu}_{i(e_1^*), j(e_1^*)|\mathcal{D}_{e_1^*}}^1$ is calculated as

$$\widehat{\nu}_{i(e_1^*), j(e_1^*)|\mathcal{D}_{e_1^*}}^1 = h_\nu\left(e_1^*, 1|\widehat{\boldsymbol{\beta}}_\nu^{\mathrm{SV}}\right), \tag{5.3.2}$$

where the function $h_\nu\left(e, l|\boldsymbol{\beta}_\nu^{\mathrm{SV}}\right)$, which depends on the respective edge e and tree number l and is parametrized by $\boldsymbol{\beta}_\nu^{\mathrm{SV}}$, represents the model specification for the second copula parameters.

The rest of the R-vine is extended tree-wise starting from tree number two. For each tree l we have to ensure that the proximity condition is fulfilled after a new edge e_l has been added. For all edges e_l with $j(e_l) = s$ and $\mathcal{D}_{e_l} = \mathcal{D}_{e_{l-1}^*} \cup i(e_{l-1}^*)$ which fulfill the proximity condition, we calculate the respective Kendall's τ's

$$\widehat{\tau}_{i(e_l), j(e_l)|\mathcal{D}_{e_l}} = \mathrm{F}_z^{-1}\left(h_l^{\mathrm{mod}}\left(e_l|\widehat{\boldsymbol{\beta}}_{\mathrm{mod}, l}\right)\right).$$

Again the edge e_l^* with the biggest Kendall's τ is selected and included into the R-vine. As before a copula family $t_{i(e_l^*), j(e_l^*)|\mathcal{D}_{e_l^*}}$ has to be selected and the corresponding parameters $\widehat{\theta}_{i(e_l^*), j(e_l^*)|\mathcal{D}_{e_l^*}}^l$ and $\widehat{\nu}_{i(e_l^*), j(e_l^*)|\mathcal{D}_{e_l^*}}^l$ have to be calculated in analogy to (5.3.1) and (5.3.2), respectively.

For trees exceeding the truncation level $k < d$, arbitrary edges which fulfill the proximity condition can be chosen. The copulae corresponding to these edges are selected to be independence copulae. Thus, no parameters have to be specified for these copulae.

The above described procedure yields an R-vine copula specification corresponding to the variables $u_t^s, u_t^1, \ldots, u_t^d$ with R-vine distribution $\boldsymbol{F}(u_t^s, u_t^1, \ldots, u_t^d)$. Applying Lemma 4.7, this allows to calculate $F_{s|1,\ldots,d}(u_t^s|u_t^1, \ldots, u_t^d)$ iteratively. Following we are able to simulate from the predictive distribution $F_{s|1,\ldots,d}(u_t^s|u_t^1, \ldots, u_t^d)$ by means of the probability integral transform. We simulate $v \sim \mathcal{U}(0, 1)$ and calculate $\check{u}_t^s = F_{s|1,\ldots,d}^{-1}(v|u_t^1, \ldots, u_t^d)$.

If one transforms the copula data $\check{u}_t^s$ resulting from these simulations back to the level of the originally modeled data $\check{y}_t^s$, one can calculate *point predictions* as the mean of the back transformed simulations. Moreover the calculation of *prediction intervals* is possible. The conditional distribution $F_y(y_t^s|y_t^1, \ldots, y_t^d)$ can be approximated by the empirical distribution function calculated from the respective simulated sample.

Omiting all arguments, the conditional densitiy $f_{s|1,\ldots,d}$ corresponding to $F_{s|1,\ldots,d}$ can be obtained by decomposing numerator and denominator of

$$f_{s|1,\ldots,d} = \frac{f_{s,1,\ldots,d}}{f_{1,\ldots,d}},$$

according to (4.1.1) into products of pair copulae and marginal densities. Since the R-vine copula specification corresponding to $f_{s,1,\ldots,d}$ differs from the R-vine copula specification corresponding to $f_{1,\ldots,d}$ only in terms of the additional edges $e_1^*, \ldots, e_{d-1}^*$, this yields

$$f_{s|1,\ldots,d} = \prod_{l=1}^{d-1} c_{i(e_l^*),j(e_l^*);\mathcal{D}_{e_l^*}} \left(F_{i(e_l^*)|\mathcal{D}_{e_l^*}}, F_{j(e_l^*)|\mathcal{D}_{e_l^*}}; \widehat{\theta}^l_{i(e_l^*),j(e_l^*)|\mathcal{D}_{e_l^*}}, \widehat{\nu}^l_{i(e_l^*),j(e_l^*)|\mathcal{D}_{e_l^*}}, t_{i(e_l^*),j(e_l^*)|\mathcal{D}_{e_l^*}} \right),$$
(5.3.3)

where the marginal density f_s vanishes, because we perform our calculations on the copula data level. Due to the fact that it holds $j(e_l^*) = s$ by construction and that we usually consider truncations at a certain level $k < d$, (5.3.3) simplifies to

$$f_{s|1,\ldots,d} = \prod_{l=1}^{k} c_{i(e_l^*),s;\mathcal{D}_{e_l^*}} \left(F_{i(e_l^*)|\mathcal{D}_{e_l^*}}, F_{s|\mathcal{D}_{e_l^*}}; \widehat{\theta}^l_{i(e_l^*),s|\mathcal{D}_{e_l^*}}, \widehat{\nu}^l_{i(e_l^*),s|\mathcal{D}_{e_l^*}}, t_{i(e_l^*),s|\mathcal{D}_{e_l^*}} \right).$$
(5.3.4)

In our case we perform the above presented calculations based on the distance model and on the model (5.2.1) for the second copula parameters. Due to our previous investigations on the structure of the R-vine underlying the spatial R-vine model (see Table 5.1) we select a Student-t copula for every edge which is added to the truncated R-vine.

Eventually we are ready to predict the 24 mean temperature time series over the period 01/01/2010-12/31/2012 corresponding to the 24 observation stations constituting the validation data set. The complete results of these predictions, which are based on 1000 simulations of each time series, are visualized in Figures A.1.1-A.1.12 in Appendix A.1. Since we expect a dependence of the quality of the predictions on the elevation, we arranged the results in the different types of figures and tables according to the respective observation station elevations.

Figures A.1.1-A.1.4 illustrate the predicted time series for the 24 observation stations of the validation data set. For the purpose of comparison we plotted the observed values in black and the prediction in gray. Moreover, the respective 95% prediction intervals are indicated by the light gray area around the point predictions. For most of the observation stations with low elevations we observe that our predictions reflect the realizations pretty good over all three years. The gray lines representing the predictions cover the realizations nearly completely. Only for some of the stations with low elevations a deviance of the predictions from the observed time series can be detected based on these plots. Moreover the respective prediction intervals tend to be narrow. *Arkona* (57), which is a station with an elevation of 42 meters, is an example for a station with low elevation where clear deviations between prediction and realization can be detected. With rising elevation we increasingly observe time periods where the predictions doesn't meet the realizations, up to observation stations like *Großer Arber* (69) which are located in mountainous regions, where the prediction fails systematically.

In order to be able to compare the goodness of the predictions of the 24 different time series we calculated *mean squared errors* according to

$$\mathrm{MSE}(s) := \frac{1}{N} \sum_{t=1}^{N} (y_t^s - \widehat{y}_t^s)^2$$

s	short name	elevation	MSE(s)	s	short name	elevation	MSE(s)
61	bork	3.00	0.57	63	buch	340.00	1.08
62	bvoe	10.00	0.36	65	ebra	346.00	1.01
72	luec	17.00	0.24	60	blan	417.00	0.80
68	gram	27.00	0.19	66	ellw	460.00	0.96
64	cosc	40.00	0.85	59	augs	461.40	0.68
57	arko	42.00	1.87	67	falk	472.00	0.43
77	rahd	42.00	0.28	76	ohrz	494.00	0.92
70	grue	55.90	0.17	71	hohe	743.30	3.65
56	alfe	143.90	0.64	55	albs	759.00	2.06
58	arns	159.00	0.94	75	neuh	845.00	6.58
78	wies	187.00	0.75	73	mitt	981.00	2.85
74	muel	273.00	0.63	69	garb	1436.00	14.69

Table 5.11: Mean squared errors of the mean temperature predictions over the period 01/01/2010-12/31/2012 for the observation stations of the validation data set based on the spatial R-vine model.

for $s = 55, \ldots, 78$, where $N = 1096$. The resulting mean squared errors (ordered according to elevation) are summarized in Table 5.11.

Based on the mean squared errors we select three stations for which we will have a closer look on the results of the predictions. We select the stations *Grünow* (70) and *Großer Arber* (69) which exhibit the lowest and the highest mean squared error, respectively. Moreover we choose the station *Arkona* (57), which is the station with the highest mean squared error among all stations where there was no need of restriction of the aggregated parameters summarized in Table 5.10. The respective predictions are compared in Figure 5.3.6. Whereas the predictions for Grünow meet the observed values pretty good and the prediction intervals are very narrow, we observe noticeable deviations for Arkona. There seems to be more uncertainty in the predictions for Arkona, which is reflected in the comparatively broad prediction intervals, however there are periods in which the predictions are better and periods in which the predictions deviate from reality. From the plot for the observation station Großer Arber we conclude that the temperatures there are more volatile. Most often the predictions are far apart from the realizations and the true temperature observations fall outside the prediction intervals. This observed deviance is a result of a misspecification of the marginal model parameters. We already observed by means of Table 5.10, that our marginal model is not able to capture the temperature trends of stations which lie outside the range of the training data set. In the case of the station Großer Arber especially a misspecification of the parameters of the seasonality component and the shape and scale parameter of the skew-t distribution seem to be problematic, since we observe periods of consequent over- respective underestimation of the true temperatures and since comparatively extreme temperature events are not captured by the predictions.
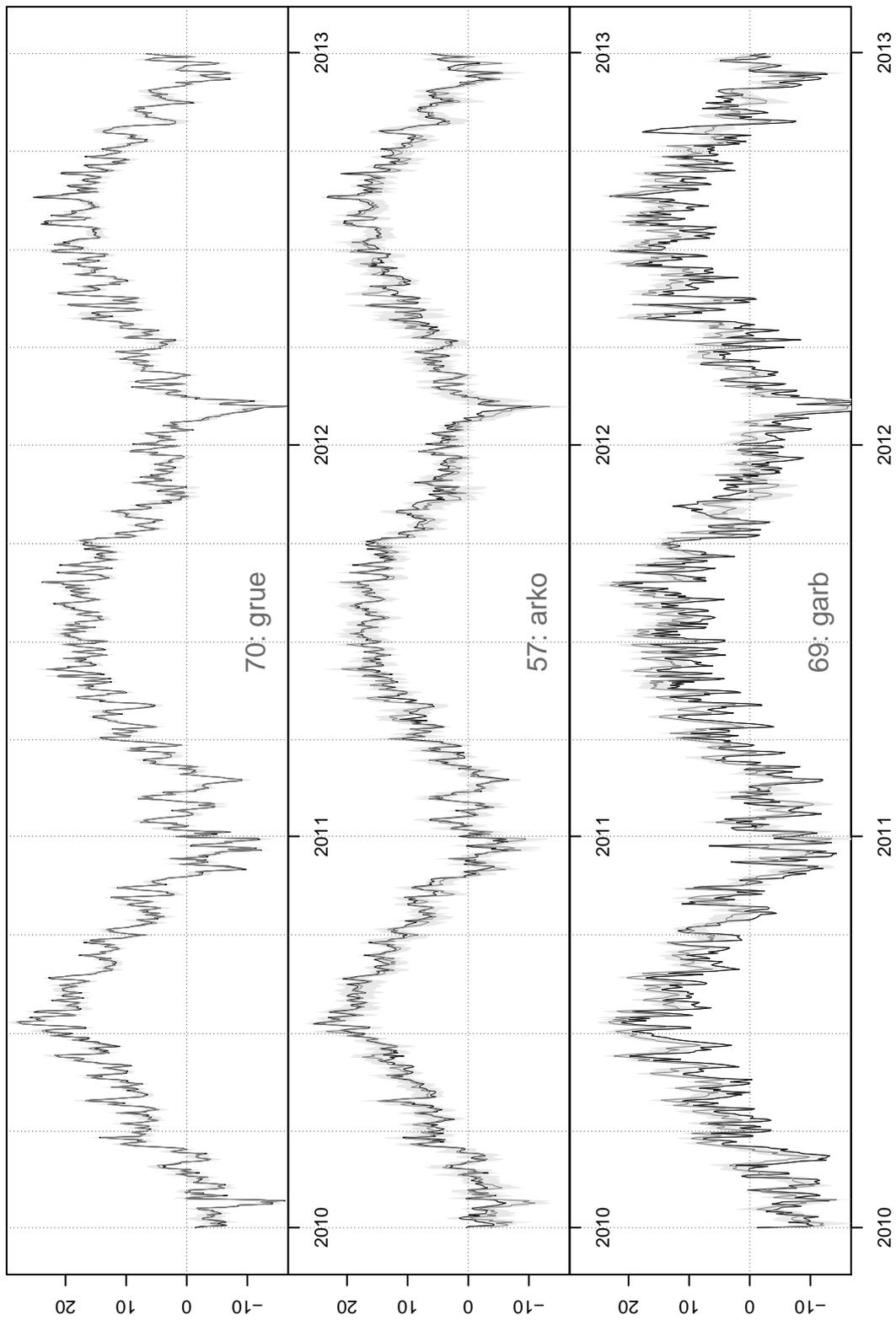
Figure 5.3.6: Prediction of the mean temperatures for the observation stations Grünow (70), Arkona (57) and Großer Arber (69) for the period 01/01/2010-12/31/2012 based on the spatial R-vine model. black line: observed values. dark gray line: prediction. light gray area: 95% prediction intervals.
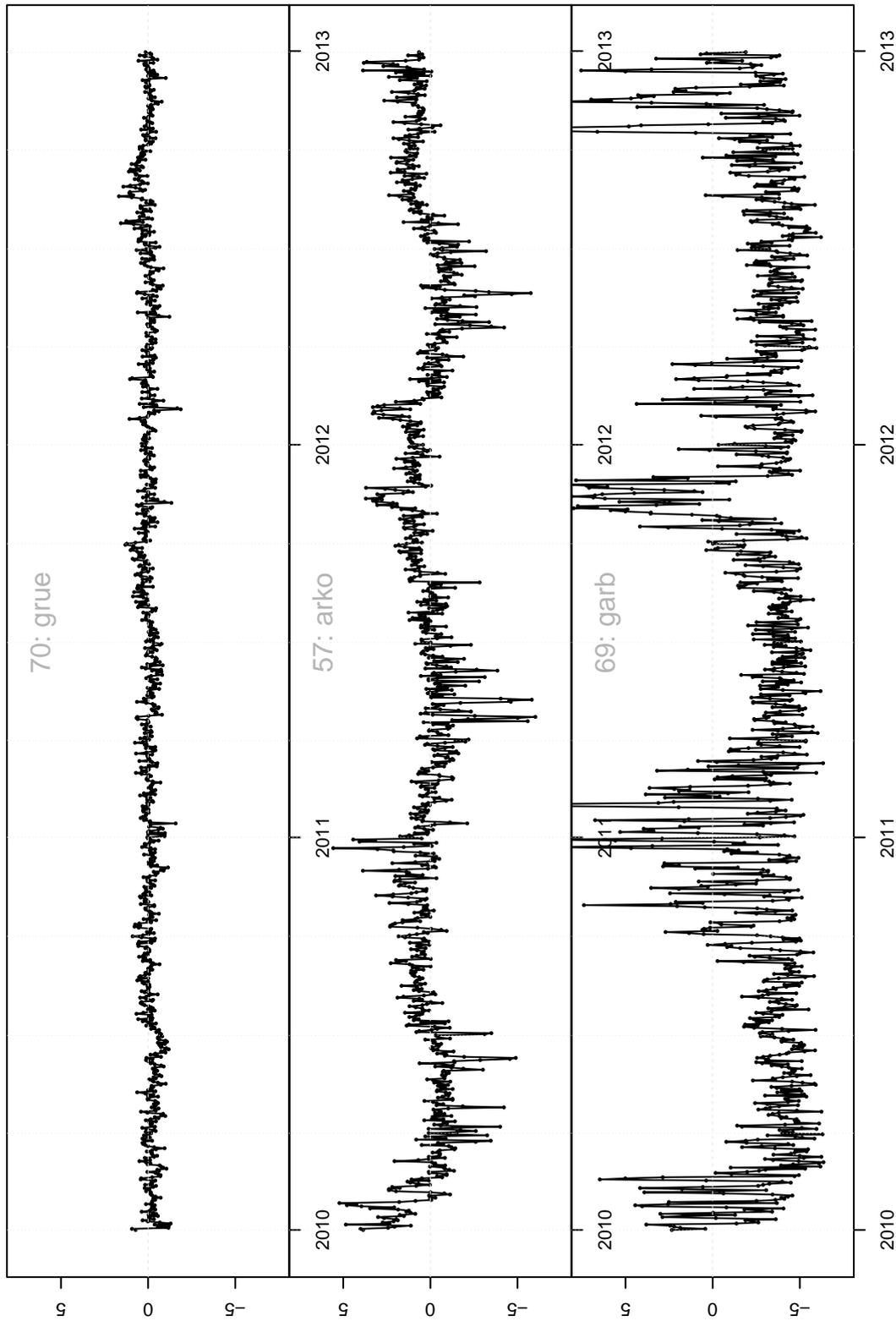
Figure 5.3.7: Prediction errors of the predictions for the observation stations Grünow (70), Arkona (57) and Großer Arber (69) for the period 01/01/2010-12/31/2012 based on the spatial R-vine model.
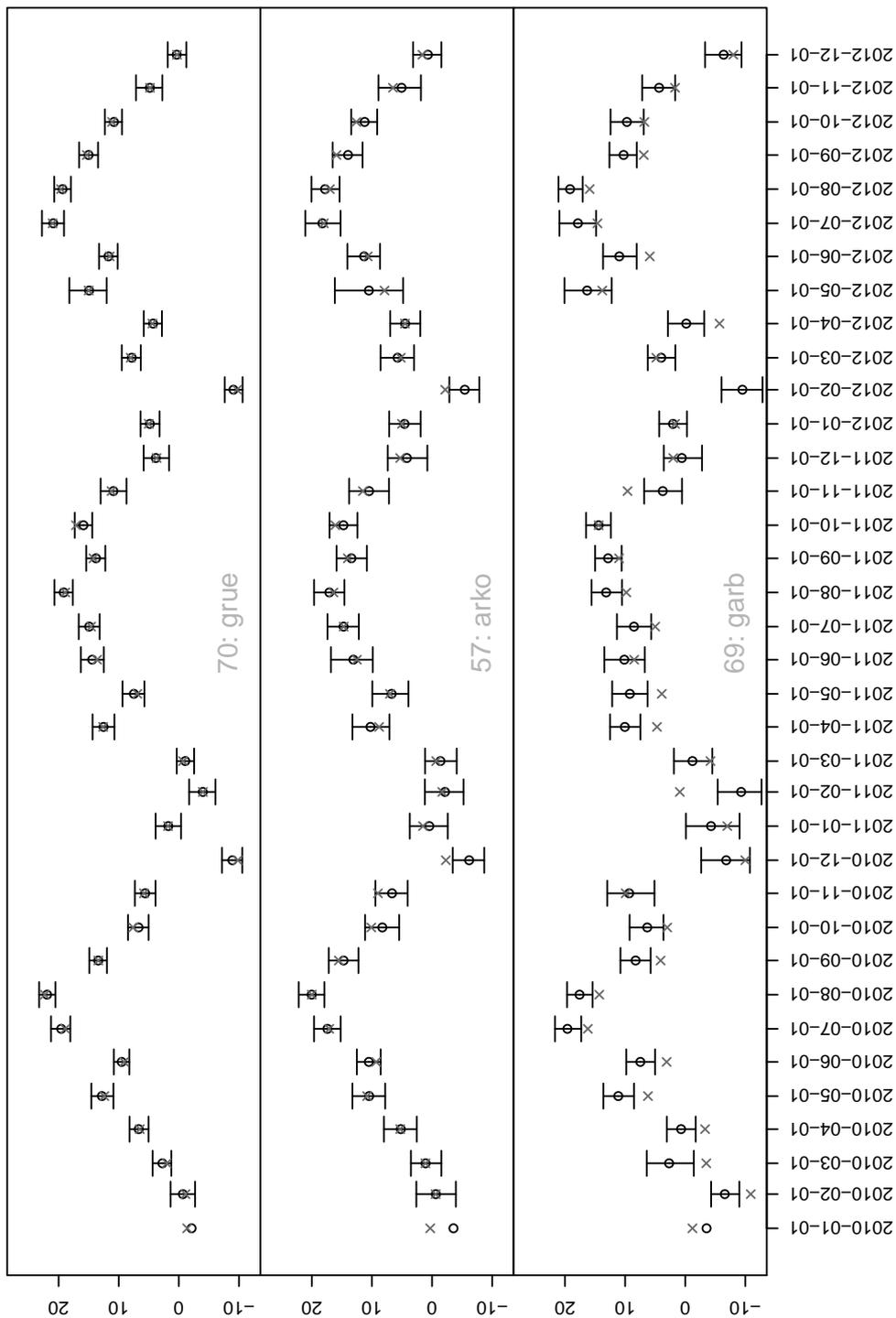
Figure 5.3.8: 95% prediction intervals, point predictions and observed mean temperatures for the first days of each months during the years 2010-2012 for the observation stations Grünow (70), Arkona (57) and Großer Arber (69), each calculated based on 1000 simulations from the predictive distribution of the mean temperatures at the respective observation station (spatial R-vine model).

Figures A.1.5-A.1.8 are plots of the *prediction errors* $(y_t^s - \widehat{y}_t^s)$ for all 24 time series of the validation data set, once again arranged according to increasing elevation. We find that for all observation stations there are single days where the prediction errors are comparatively high. Whereas we observe that most of the prediction errors for stations like *Bremervörde* (62), *Grambek* (68), *Grünow* (70) and *Rahden-Varl* (77) lie within a range of $-2°$C and $2°$C, we see that the errors of stations with elevations higher than 740 meters can get pretty large. Moreover we see that there are stations like *Ebrach* (65) and *Falkenberg, Kr.Rottal-Inn* (67) where the real temperatures tend to be over- respectively underestimated. Furthermore, systematic prediction errors owing to a misspecified seasonality component, are detected for a handful of observation stations, e.g. *Arkona* (57) and *Großer Arber* (69). Figure 5.3.7 highlights the corresponding prediction errors for the three previously selected stations Grünow (70), Arkona (57) and Großer Arber (69).

As an additional tool for the analysis of the predictions based on the spatial R-vine model, we provide the plots given in Figures A.1.9-A.1.12. They depict 95% prediction intervals, point predictions and observed mean temperatures for the first days of each months during the years 2010-2012 for all 24 validation data stations. These monthly snapshots of our predictions and the respective observed values facilitate the evaluation of the predictions compared to an analysis of Figures A.1.1-A.1.3, they give an impression how the complete time series behave over the years 2010-2012.

Let us first remark, that there always is no prediction interval for the first of January 2010, since we had to calculate fix start values for each of the time series, which is due to the inclusion of three autoregressive components in the marginal model. Comparison of the different plots shows that there are stations where the predictions are more uncertain compared to other stations. Besides we see that there are single days like the first of May 2012, where we observe higher uncertainty over large parts of Germany. An overall consideration of Figures A.1.9-A.1.12 yields that for most of the stations only up to two of the considered observations fall outside their respective prediction interval, however, for the highly situated observation stations *Neuhaus am Rennweg* (75) and *Großer Arber* (69) a big share of the observations under consideration lie outside the corresponding prediction interval. Surprisingly nearly all considered observations for the observation station *Mittenwald-Buckelwiesen* (73) which has the second highest elevation fall into their associated prediction interval. The respective plots for the three selected stations Grünow (70), Arkona (57) and Großer Arber (69) are illustrated in Figure 5.3.8.

This first analysis of predictions from our spatial R-vine model yields a good first impression of the prediction capabilities of our model. In most cases our predictions seem to be able to approximate the reality appropriately good. However some shortcomings of our marginal model became clear. With these observations we advance to the next chapter, where we are going to present a different approach to model spatial dependencies based on vine copulae.

Chapter 6

Spatial composite vine models

The previous chapter was a straigthforward approach to model spatial dependencies based on vine copula models by embedding spatial information into the parametrization of an R-vine model. This chapter is going to follow a different approach which composes several local dependency models by means of a composite likelihood, where the modeling of these local dependencies is carried out by C-vines.

6.1   Composite likelihood methods

In order to have the necessary background to be able to understand the models presented in the following, we briefly provide some theoretical aspects on *composite likelihood methods*. For this purpose we refer to Varin, Reid, and Firth (2011), who give a broad overview on these methods.

Composite likelihood methods are inferential statistics methods. The main idea of these methods is the composition of a composite likelihood function as a product of potentially dependent likelihoods corresponding to certain parts of the model under consideration. Due to the fact that the single likelihoods are not necessarily independent, the composite likelihood can be considered as the likelihood of a misspecified model. The usage of such misspecified models is most often motivated computationally.

Let us now formally define the composite likelihood. To this end, let $\boldsymbol{Y}$ be an m-dimensional random vector with multivariate probability density function $f(\boldsymbol{y}; \boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is a p-dimensional parameter vector. Moreover we consider K events $\mathcal{A}_1, \ldots, \mathcal{A}_K$ and weights $w_k > 0$, $k = 1, \ldots, K$, related to these events. Furthermore, let the likelihoods corresponding to the events $\mathcal{A}_k$, $k = 1, \ldots, K$, be denoted by $\mathcal{L}_k(\boldsymbol{\theta} \mid \boldsymbol{y}) \propto f(\boldsymbol{y} \in \mathcal{A}_k; \boldsymbol{\theta})$, $k = 1, \ldots, K$. Then the *composite likelihood* is defined as

$$\mathcal{L}_C(\boldsymbol{\theta} \mid \boldsymbol{y}) = \prod_{k=1}^{K} [\mathcal{L}_k(\boldsymbol{\theta} \mid \boldsymbol{y})]^{w_k}.$$

Following the composite log-likelihood is given and denoted by

$$c\ell(\boldsymbol{\theta} \mid \boldsymbol{y}) := \ln \mathcal{L}_C(\boldsymbol{\theta} \mid \boldsymbol{y}) = \sum_{k=1}^{K} w_k \ln [\mathcal{L}_k(\boldsymbol{\theta} \mid \boldsymbol{y})] = \sum_{k=1}^{K} w_k \ell_k(\boldsymbol{\theta} \mid \boldsymbol{y}).$$

The idea behind the weights w_k, $k = 1, \ldots, K$, is to improve the efficiency of maximum composite likelihood estimation. Their selection depends on the specific problem.

Varin et al. (2011) distinguish between composite conditional and composite marginal likelihoods. Whereas a composite conditional likelihood is composed out of conditional densities, a composite marginal likelihood is based on marginal densities. Since we are going to use a composite marginal likelihood, we focus on this variant in the following.

A straightforward composite marginal likelihood is the product of the univariate marginal densities $f(y_r; \boldsymbol{\theta})$, $r = 1, \ldots, m$. Due to this structure, this kind of likelihood does not allow to access parameters which are related to the dependence of the components of $\boldsymbol{Y}$. In order to do so, multivariate marginal densities have to be taken into consideration. A popular example is the so called pairwise likelihood

$$\mathcal{L}_{\text{pair}} (\boldsymbol{\theta} \mid \boldsymbol{y}) = \prod_{r=1}^{m-1} \prod_{s=r+1}^{m} f(y_r, y_s; \boldsymbol{\theta}),$$

but it is also possible to use products of higher dimensional marginal densities, what we are going to do in the following.

6.2   Composite vine model

Now we are ready to propose our own composite likelihood model in the area of spatial dependence modeling. For this purpose let us again consider the copula data $\boldsymbol{u}^1, \ldots, \boldsymbol{u}^d$ with $\boldsymbol{u}^s = (u_1^s, \ldots, u_N^s)^\top$, $s = 1, \ldots, d$. To model the dependence of this data, we are going to set up a composite likelihood model which is composed out of the densities corresponding to d 4-dimensional C-vines numbered $1, \ldots, d$, i.e. we consider one C-vine for every variable $\boldsymbol{u}^s$, $s = 1, \ldots, d$. The first root node of every C-vine $s = 1, \ldots, d$ complies with the number s of the C-vine. In other words, the C-vines are numbered according to their first root nodes. For each C-vine s we select the variable p_s with the shortest distance to the first root node s as the second root, the variable q_s with the second shortest distance as the third root and the variable r_s with the third shortest distance as the forth variable of the C-vine. Now we have uniquely fixed the tree structure of all d C-vines in the model. Figure 6.2.1 gives a graphical representation of these C-vines.

Moreover, the families of the bivariate copulae corresponding to the edges of the d C-vines have to be specified. They are determined by means of a sequential selection algorithm, for each single C-vine.

Finally, as all necessary components of the C-vines $s = 1, \ldots, d$ are established, we are able to state their likelihoods. They are given as

$$
\begin{aligned}
\mathcal{L}_s \left(\boldsymbol{\theta}_s, \boldsymbol{\nu}_s \mid u_t^s, u_t^{p_s}, u_t^{q_s}, u_t^{r_s}\right) &= c_s \left(u_t^s, u_t^{p_s}, u_t^{q_s}, u_t^{r_s}; \boldsymbol{\theta}_s, \boldsymbol{\nu}_s\right) \\
&= c_{sp_s} \left(u_t^s, u_t^{p_s}; \theta_{sp_s}, \nu_{sp_s}\right) \cdot c_{sq_s} \left(u_t^s, u_t^{q_s}; \theta_{sq_s}, \nu_{sq_s}\right) \\
&\quad \cdot c_{sr_s} \left(u_t^s, u_t^{r_s}; \theta_{sr_s}, \nu_{sr_s}\right) \\
&\quad \cdot c_{p_s q_s; s} \left(\breve{u}_t^{p_s}, \breve{u}_t^{q_s}; \theta_{p_s q_s | s}, \nu_{p_s q_s | s}\right) \cdot c_{p_s r_s; s} \left(\breve{u}_t^{p_s}, \breve{u}_t^{r_s}; \theta_{p_s r_s | s}, \nu_{p_s r_s | s}\right) \\
&\quad \cdot c_{q_s r_s; s p_s} \left(\breve{u}_t^{q_s}, \breve{u}_t^{r_s}; \theta_{q_s r_s | s p_s}, \nu_{q_s r_s | s p_s}\right)
\end{aligned}
$$
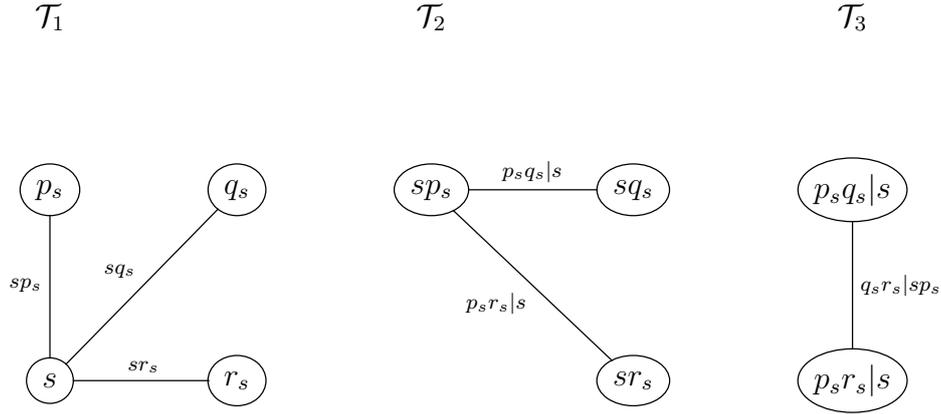
Figure 6.2.1: Structure of all C-vines $s = 1, \ldots, d$ composing the composite likelihood model.

where

$$\breve{u}_t^o = F(u_t^o \mid u_t^s) = \frac{\partial}{\partial u_t^s} C_{os}\left(u_t^o, u_t^s; \theta_{so}, \nu_{so}\right), \quad o = p_s, q_s, r_s,$$

$$\breve{u}_t^o = F(u_t^o \mid u_t^s, u_t^{p_s}) = \frac{\partial}{\partial \breve{u}_t^{p_s}} C_{op_s;s}\left(\breve{u}_t^o, \breve{u}_t^{p_s}; \theta_{p_s o|s}, \nu_{p_s o|s}\right), \quad o = q_s, r_s,$$

and

$$\boldsymbol{\theta}_s = \left(\theta_{sp_s}, \theta_{sq_s}, \theta_{sr_s}, \theta_{p_s q_s|s}, \theta_{p_s r_s|s}, \theta_{q_s r_s|sp_s}\right)^\top,$$

$$\boldsymbol{\nu}_s = \left(\nu_{sp_s}, \nu_{sq_s}, \nu_{sr_s}, \nu_{p_s q_s|s}, \nu_{p_s r_s|s}, \nu_{q_s r_s|sp_s}\right)^\top.$$

Note that the second parameter $\nu_{ij|\ldots}$ of the bivariate copula is only needed in the case of a Student-t copula, for the respective degrees of freedom. Otherwise the bivariate copula does not depend on $\nu_{ij|\ldots}$ and we omit this parameter or set it simply to zero. Note furthermore, that in trees of order one it holds for the parameters, that $\theta_{ij} = \theta_{ji}$ and $\nu_{ij} = \nu_{ji}$, $i = 1, \ldots, d$, $j = 1, \ldots, d$, since each edge $\{i, j\}$ may occur in up to two of the d C-vines composing the model.

In order to be able to set up the composite likelihood, it remains to specify the weights w_s, $s = 1, \ldots, d$. We define them as the reciprocal value of the number counts

$$n_s := \# \{k : s \text{ is included in C-vine } k\}, \quad s = 1, \ldots, d,$$

of C-vines which include $s = 1, \ldots, d$, i.e. we define

$$w_s := \frac{1}{n_s}.$$

All in all, this results in the composite likelihood

$$\mathcal{L}_{\text{CVM}}\left(\boldsymbol{\theta}^{\text{CVM}}, \boldsymbol{\nu}^{\text{CVM}} \mid \boldsymbol{u}^1, \ldots, \boldsymbol{u}^d\right) = \prod_{s=1}^{d} \prod_{t=1}^{N} \left[\mathcal{L}_s\left(\boldsymbol{\theta}_s, \boldsymbol{\nu}_s \mid u_t^s, u_t^{p_s}, u_t^{q_s}, u_t^{r_s}\right)\right]^{w_s}$$

and accordingly in the composite log-likelihood

$$cl_{\mathrm{CVM}}\left(\boldsymbol{\theta}^{\mathrm{CVM}}, \boldsymbol{\nu}^{\mathrm{CVM}} \,|\, \boldsymbol{u}^1, \ldots, \boldsymbol{u}^d\right) = \sum_{s=1}^{d} \sum_{t=1}^{N} w_s \ell_s \left(\boldsymbol{\theta}_s, \boldsymbol{\nu}_s \,|\, u_t^s, u_t^{p_s}, u_t^{q_s}, u_t^{r_s}\right), \qquad (6.2.1)$$

where

$$\ell_s\left(\boldsymbol{\theta}_s, \boldsymbol{\nu}_s \,|\, u_t^s, u_t^{p_s}, u_t^{q_s}, u_t^{r_s}\right) = \ln \mathcal{L}_s\left(\boldsymbol{\theta}_s, \boldsymbol{\nu}_s \,|\, u_t^s, u_t^{p_s}, u_t^{q_s}, u_t^{r_s}\right)$$

and where

$$\boldsymbol{\theta}^{\mathrm{CVM}} = \{\theta : \theta \in \boldsymbol{\theta}_s, s = 1, \ldots, d\},$$
$$\boldsymbol{\nu}^{\mathrm{CVM}} = \{\nu : \nu \in \boldsymbol{\nu}_s, s = 1, \ldots, d\}.$$

We call the above model a (three neighbors) *composite vine model (CVM)*.

6.3　A composite vine model for mean temperature

Eventually we are going to investigate the mean temperature data set with the help of the above model. In a first step we have to select the structure of the composite vine model. This involves amongst others the selection of the 54 C-vines. They are selected as described in Section 6.2. The result of this selection is given in Table 6.1. Moreover we have to determine adequate families for the $6 \cdot 54 = 324$ bivariate copulae corresponding to the edges of the previously chosen C-vines. To this end we allow for Gaussian, Student-t, Clayton, Gumbel and Frank copulae, without specifying in detail if a rotated version of the respective copula is used or not. The exact rotation type will be fixed later on, when the respective parameters are estimated. We achieve families as given in Table 6.2, where the families are coded according to Table 4.4 and where it remains unsettled, if any kind of rotation is used. This results in a parametrization with 464 parameters.

Figure 6.3.1 illustrates the achieved model structure. On the one hand the numbered circles indicate the locations of the 54 observation stations $s = 1, \ldots, 54$ summarized in Table 3.1. Their circumference and color represents the numbers n_s, $s = 1, \ldots, 54$, of C-vines in which the respective nodes are included, i.e. the inverse of the respective weights w_s, $s = 1, \ldots, 54$, according to the scale defined at the left bottom of the figure. On the other hand the lines indicate all C-vine edges that occur in the first trees of the C-vines $s = 1, \ldots, 54$. Dotted lines represent edges that occur only once, whereas dashed lines represent edges that occur in the first tree of the two C-vines whose first root nodes are the nodes at the end of the respective line.

For the mean temperature data set numerical maximization of the composite log-likelihood (6.2.1), whose structure and components where specified above, yields a maximum composite log-likelihood of $cl_{\mathrm{CVM}}\left(\widehat{\boldsymbol{\theta}}^{\mathrm{CVM}}, \widehat{\boldsymbol{\nu}}^{\mathrm{CVM}} \,|\, \boldsymbol{u}^1, \ldots, \boldsymbol{u}^d\right) = 42106.41$. The respective AIC and BIC are given by $\mathrm{AIC}_{\mathrm{CVM}} = -83222.8$ and $\mathrm{BIC}_{\mathrm{CVM}} = -80749.5$.
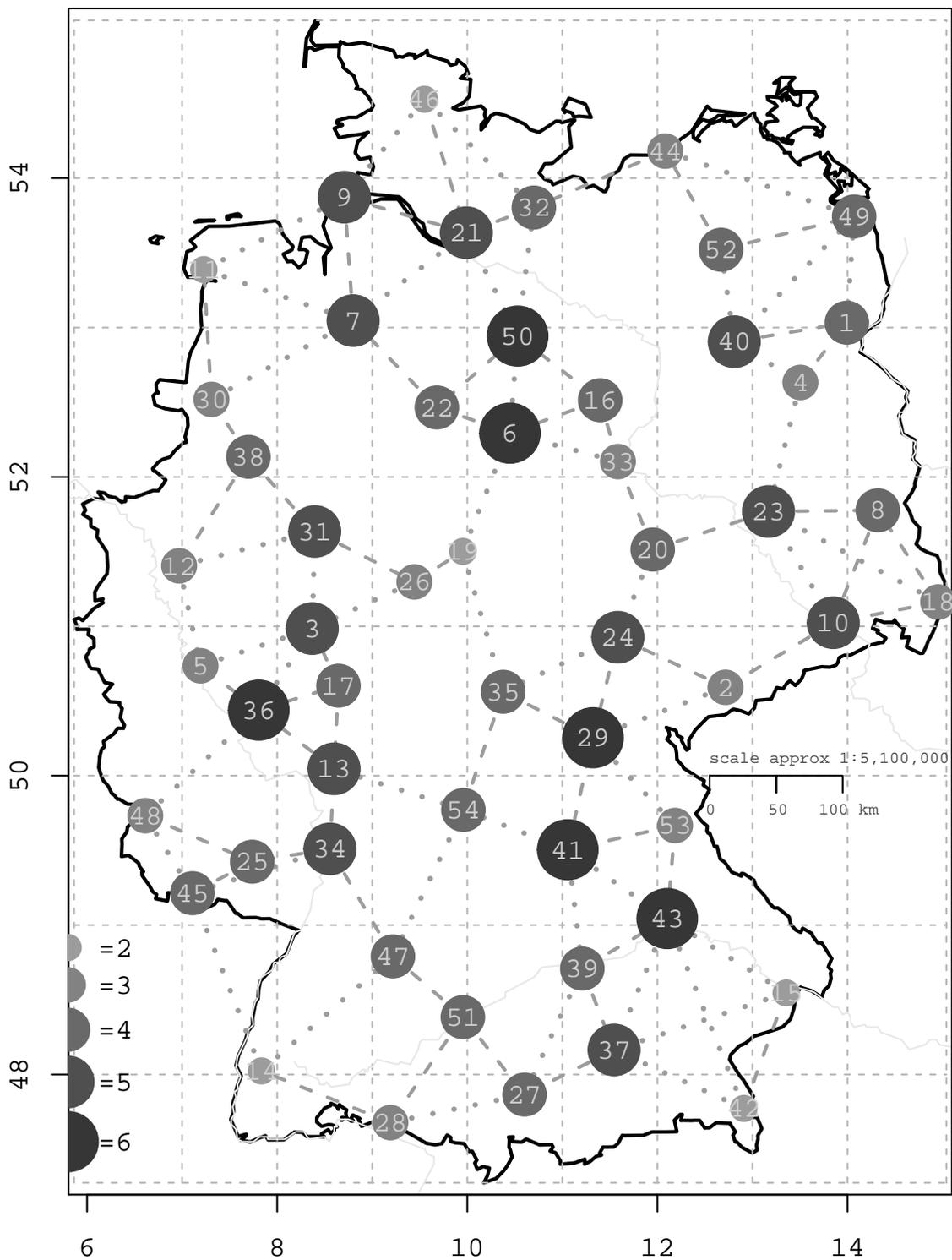
Figure 6.3.1: Visualization of the model structure of a (spatial) composite vine model on the 54 observation stations of our mean temperature data set.

s	p_s	q_s	r_s	s	p_s	q_s	r_s	s	p_s	q_s	r_s	s	p_s	q_s	r_s
1	4	49	40	15	42	43	37	29	35	24	41	43	53	39	41
2	24	10	29	16	33	6	50	30	38	11	7	44	52	32	49
3	17	31	36	17	3	13	36	31	3	38	26	45	25	48	34
4	1	40	23	18	10	8	23	32	21	50	44	46	9	21	32
5	36	12	3	19	26	6	35	33	16	20	6	47	51	34	54
6	22	16	50	20	33	24	23	34	13	25	47	48	45	25	36
7	22	9	21	21	32	50	9	35	29	54	24	49	1	52	40
8	23	18	10	22	6	50	7	36	5	17	13	50	6	16	22
9	21	46	7	23	8	20	10	37	39	27	43	51	47	27	28
10	18	8	2	24	20	29	2	38	30	31	12	52	40	44	49
11	30	7	9	25	45	34	48	39	37	43	41	53	43	41	29
12	5	38	31	26	19	31	3	40	4	52	1	54	41	35	13
13	34	17	36	27	51	37	39	41	53	54	29				
14	28	47	45	28	51	27	14	42	15	37	43				

Table 6.1: Variables composing the 54 C-vines of the (spatial) composite vine model on the 54 observation stations of our mean temperature data set.

s	sp_s	sq_s	sr_s	$p_sq_s\|s$	$p_sr_s\|s$	$q_sr_s\|sp_s$	s	sp_s	sq_s	sr_s	$p_sq_s\|s$	$p_sr_s\|s$	$q_sr_s\|sp_s$	s	sp_s	sq_s	sr_s	$p_sq_s\|s$	$p_sr_s\|s$	$q_sr_s\|sp_s$
1	t	t	t	t	t	t	19	t	t	t	t	t	C	37	t	t	t	t	t	t
2	t	t	t	t	t	t	20	t	t	t	t	t	G	38	t	t	t	t	t	t
3	t	t	t	t	t	C	21	t	t	t	t	t	C	39	t	t	t	t	t	t
4	t	t	t	t	t	G	22	t	t	t	t	t	t	40	t	t	t	C	t	t
5	t	t	t	t	t	t	23	t	t	t	t	t	t	41	t	t	t	t	t	G
6	t	t	t	F	t	t	24	t	t	t	t	t	t	42	t	t	t	t	t	Φ
7	t	t	t	C	F	t	25	t	t	t	t	t	t	43	t	t	t	G	t	t
8	t	t	t	G	t	t	26	t	t	t	t	t	t	44	t	G	t	t	t	t
9	t	t	t	t	t	t	27	t	t	t	t	t	t	45	t	t	t	F	t	t
10	t	t	t	t	t	t	28	t	t	t	t	t	G	46	t	t	G	t	G	t
11	t	t	t	t	F	t	29	t	t	t	F	t	t	47	t	t	t	C	t	t
12	t	t	t	t	t	t	30	t	t	t	t	t	t	48	t	t	t	t	t	t
13	t	t	t	t	G	t	31	t	t	t	t	t	t	49	t	t	t	t	t	t
14	t	t	t	t	t	t	32	t	t	G	t	t	G	50	t	t	t	t	t	t
15	t	t	G	Φ	t	Φ	33	t	t	t	t	t	C	51	t	t	t	t	t	t
16	t	t	t	t	t	t	34	t	t	t	F	t	t	52	t	t	t	t	t	t
17	t	t	t	t	t	t	35	t	t	t	t	t	t	53	t	t	t	t	t	t
18	t	t	t	t	t	t	36	t	t	t	G	t	t	54	t	t	t	t	t	t

Table 6.2: Families of the copulae of the C-vines $s = 1, \ldots, 54$ of the (spatial) composite vine model for our mean temperature data set.

6.4   Spatial composite vine models

Now we are going to enhance our composite vine model in terms of parameter number reduction by utilization of additionally available spatial information on the considered variables. Therefore we follow a similar reparametrization approach as in the chapter about spatial R-vine models (Chapter 5).

This time we have to consider reparametrizations of the parameters $\boldsymbol{\theta}^{\mathrm{CVM}}$ of the composite vine model. Since these reparametrizations differ from the ones in Chapter 5, we are going to state them in detail in the following. Furthermore a reparametrization of the parameters $\boldsymbol{\nu}^{\mathrm{CVM}}$ is needed.

Once again we will initiate our model building process by a tree-wise exploratory analysis of the dependence of the parameters $\boldsymbol{\theta}^{\mathrm{CVM}}$ on the distances and elevation differences occurring in the C-vines composing the composite vine model on the mean temperature data. To this end we again calculate and transform the Kendall's τ's corresponding to the variable pairs specified by the 54 C-vines and plot their transformations against the respective logarithmized distances and logarithmized elevation differences.

The respective plots for the first trees of the C-vines are depicted in Figure 6.4.1. 54 trees $s = 1, \ldots, 54$, each consisting of the three edges $\{s, p_s\}$, $\{s, q_s\}$ and $\{s, r_s\}$, yield 162 points per plot, where a big share of the previously mentioned edges occur twice, i.e. lots of the points of the plot are twofold. These twofold points are highlighted in gray. Both plots, the plot of the Fisher z-transformed Kendall's τ's against the logarithmized distances and the plot against the logarithmized elevation differences, suggest the inclusion of the predictors $d_{s,o}$ and $e_{s,o}$ into the model for the parameters of the first tree, where $o = p_s, q_s, r_s$.

For the second and third trees we additionally have to consider the distances and elevation differences to the conditioning variables of the respective copulae. This time there is an ordering of the conditioning variables, since there is an ordering of the root nodes of the C-vines, i.e. the variables s enter the conditioning set before the variables p_s, $s = 1, \ldots, 54$. Hence we do not have to define means according to our investigations in Section 5.1, i.e. the distances and elevation differences corresponding to the edges of the third trees can be considered separately.

Figure 6.4.2 shows the six scatter plots of the Fisher z-transformed Kendall's τ's $\widehat{\tau}_{p_s,o|s}$ against the logarithmized distances and elevation differences of the station pairs $\{p_s, o\}$, $\{p_s, s\}$ and $\{o, s\}$, where $o = q_s, r_s$. We observe evident linear trends for the distances. The three plots for the elevation differences exhibit only slight slopes of the regression lines and seemingly random broad scatter around this line.

The plots for the third C-vine trees are given in Figure 6.4.3. In accordance with the lower trees, the Fisher z-transformed Kendall's τ's show an evident linear dependence with respect to the logarithmized distances corresponding to the conditioned variables, whereas the regression line slopes of most of the remaining plots are rather negligible.

The above investigations of the mean temperature data set give us an idea, which distances and elevation differences might play an important role in the reparametrization of our composite vine model. In order to be able to select an adequate reparametrization
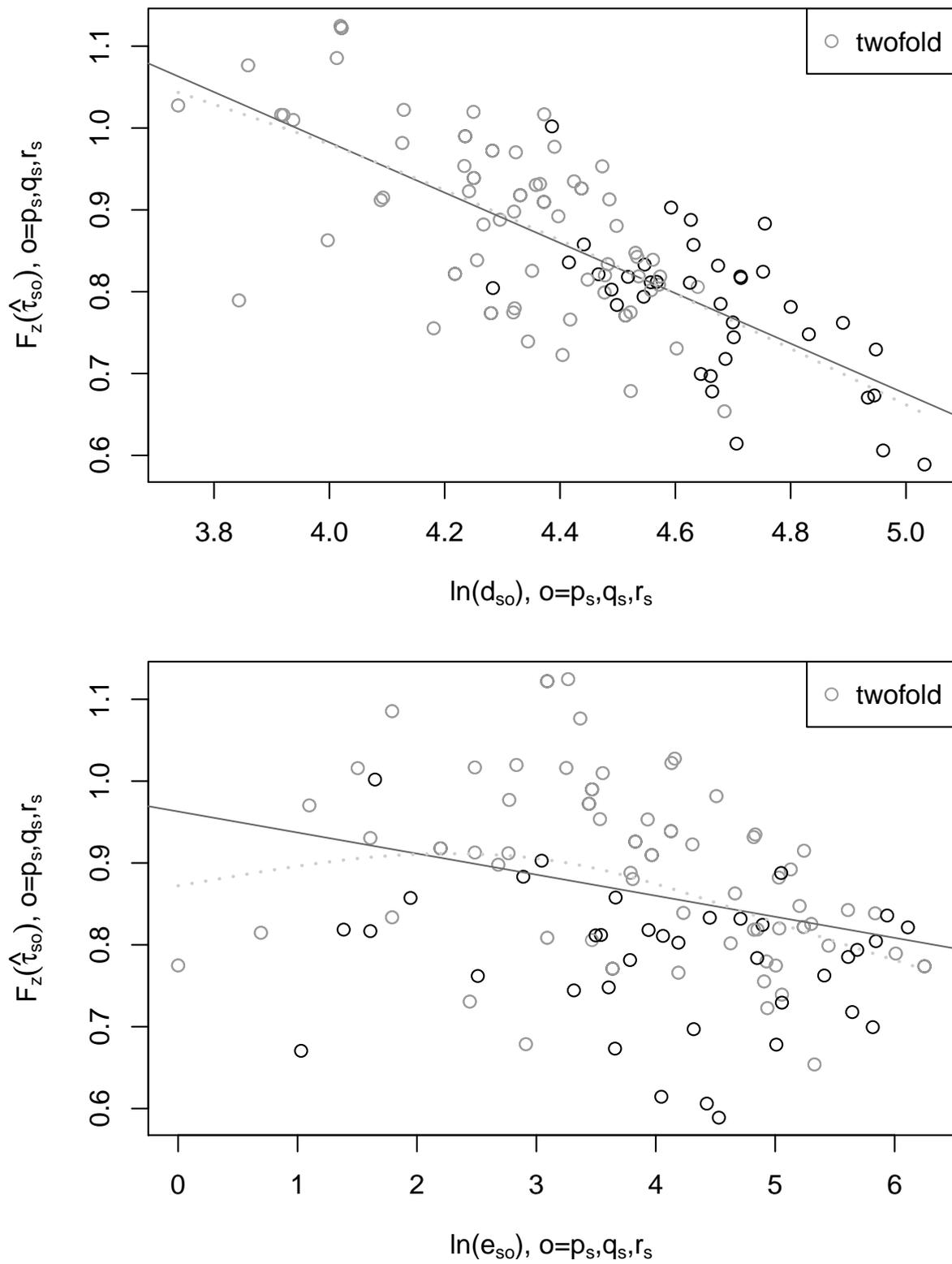
Figure 6.4.1: Scatter plots of $F_z(\widehat{\tau}_{s,o})$ against $\ln(d_{s,o})$ and $\ln(e_{s,o})$, respectively, where $o = p_s, q_s, r_s$ and $s = 1, \ldots, 54$ (Tree 1).
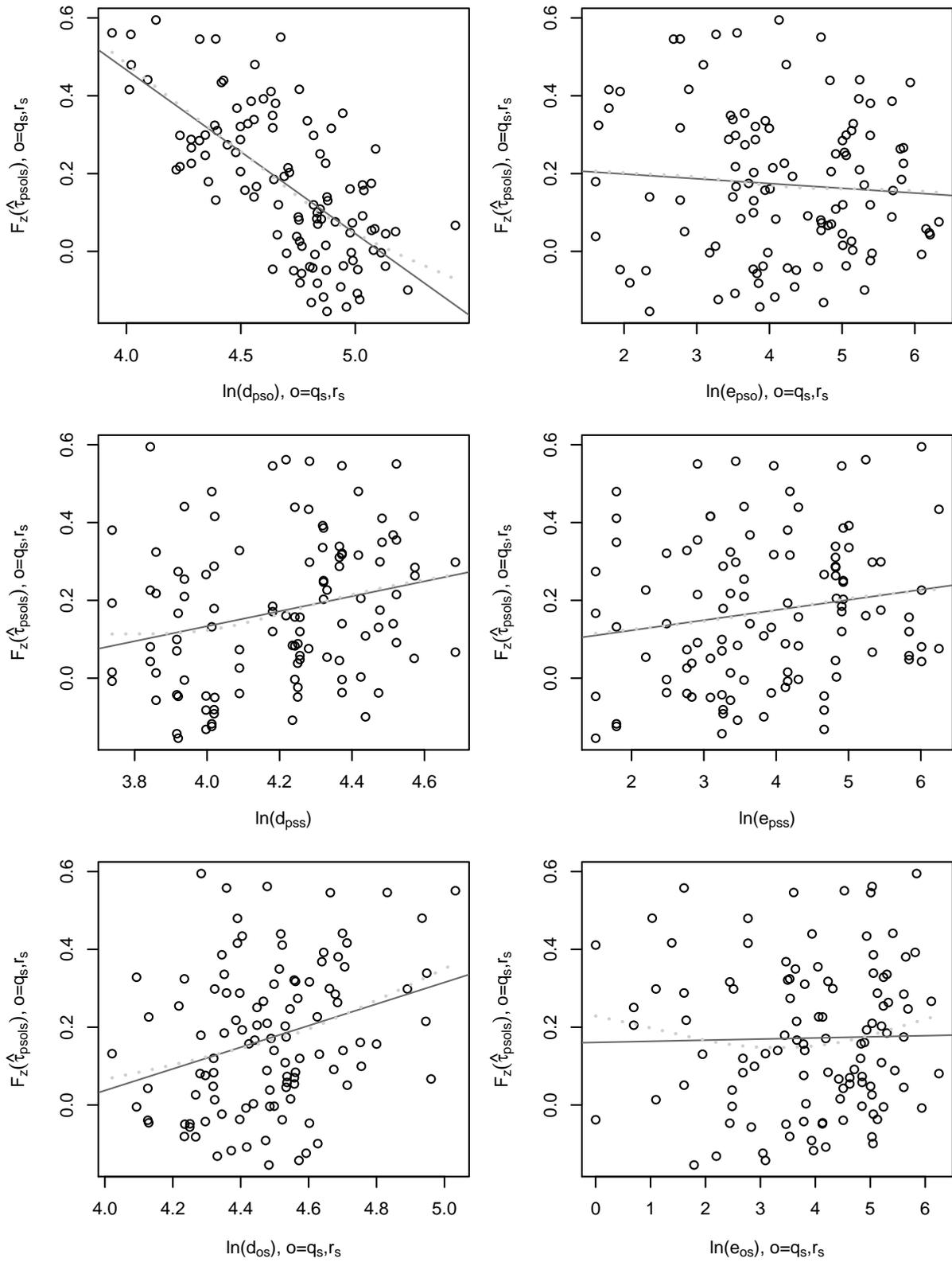
Figure 6.4.2: Scatter plots of $F_z(\widehat{\tau}_{p_s,o|s})$ against $\ln(d_{p_s,o})$, $\ln(e_{p_s,o})$, $\ln(d_{p_s,s})$, $\ln(e_{p_s,s})$, $\ln(d_{o,s})$ and $\ln(e_{o,s})$, respectively, where $o = q_s, r_s$ and $s = 1, \ldots, 54$ (Tree 2).
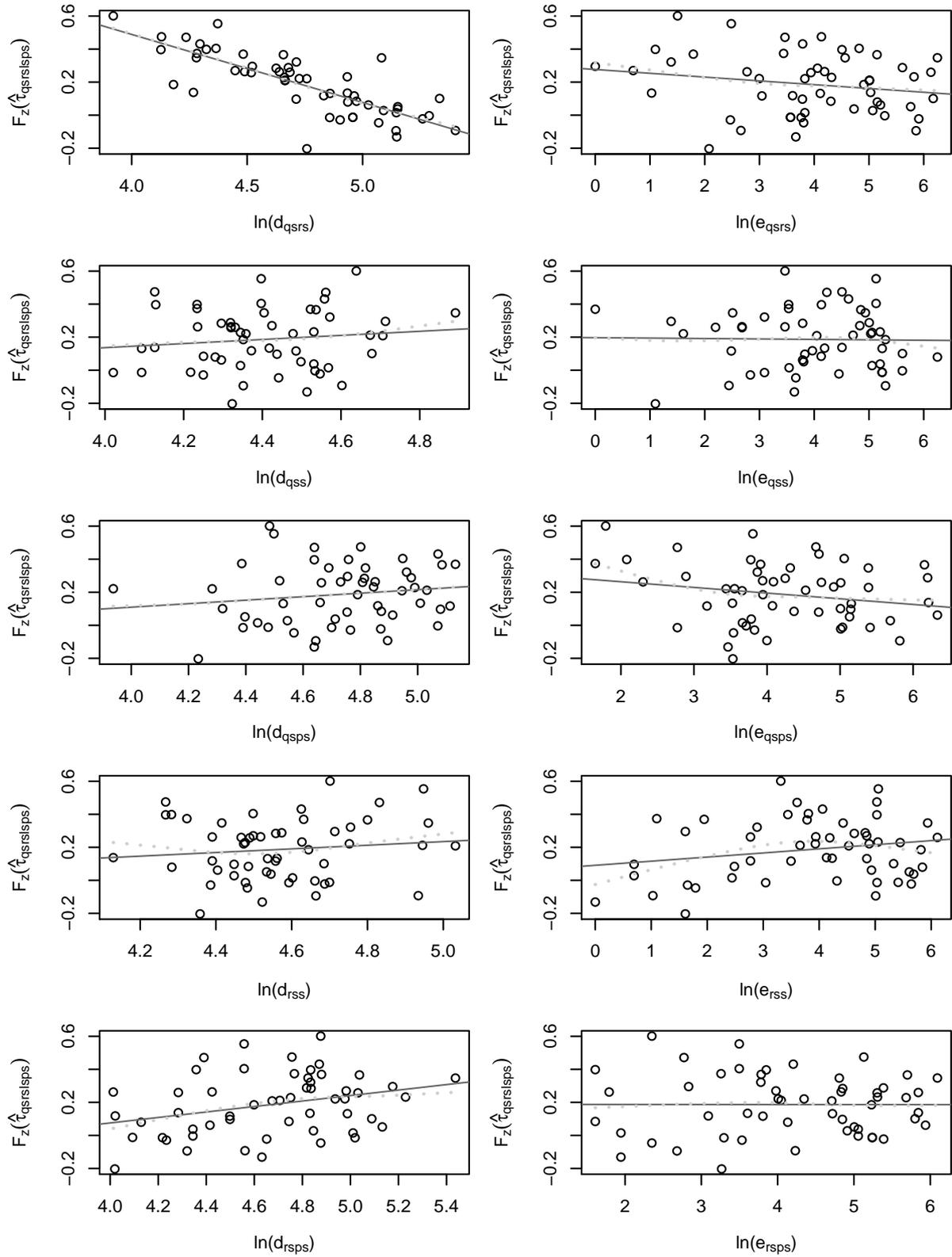
Figure 6.4.3: Scatter plots of $F_z(\widehat{\tau}_{q_s,r_s|sp_s})$ against $\ln(d_{q_s,r_s})$, $\ln(e_{q_s,r_s})$, $\ln(d_{q_s,s})$, $\ln(e_{q_s,s})$, $\ln(d_{q_s,p_s})$, $\ln(e_{q_s,p_s})$, $\ln(d_{r_s,s})$, $\ln(e_{r_s,s})$, $\ln(d_{r_s,p_s})$ and $\ln(e_{r_s,p_s})$, respectively, where $s = 1,\ldots,54$ (Tree 3).

we set up different reparametrizations of the form

$$
\begin{aligned}
\widetilde{\theta}_{io_1} &= \mathrm{T}_{\tau \to \theta} \left(\mathrm{F}_z^{-1} \left(h_1^{\mathrm{mod}}(i, o_1 | \boldsymbol{\beta}_{\mathrm{mod},1}) \right) ; t_{io_1} \right), \\
\widetilde{\theta}_{jo_2|i} &= \mathrm{T}_{\tau \to \theta} \left(\mathrm{F}_z^{-1} \left(h_2^{\mathrm{mod}}(i, j, o_2 | \boldsymbol{\beta}_{\mathrm{mod},2}) \right) ; t_{jo_2|i} \right), \\
\widetilde{\theta}_{ko_3|ij} &= \mathrm{T}_{\tau \to \theta} \left(\mathrm{F}_z^{-1} \left(h_3^{\mathrm{mod}}(i, j, k, o_3 | \boldsymbol{\beta}_{\mathrm{mod},3}) \right) ; t_{ko_3|ij} \right),
\end{aligned}
\tag{6.4.1}
$$

where 'mod' stands for the reparametrization under consideration. Selection of

$$
\begin{aligned}
i &= s, \\
j &= p_s, \\
k &= q_s, \\
o_1 &= p_s, q_s, r_s, \\
o_2 &= q_s, r_s, \quad \text{and} \\
o_3 &= r_s,
\end{aligned}
$$

for all $s = 1, \ldots, 54$, yields formulas for all parameters composing $\boldsymbol{\theta}$. The *model functions* h_l^{mod}, $l = 1, 2, 3$, for different reparametrizations 'mod' will be specified in detail, below. Due to the structural similarity of these reparametrizations to the reparametrizations defined in Chapter 5, we adopt the names applied there.

First of all we consider the *full reparametrization* (full), where the model functions h_l^{mod}, $l = 1, 2, 3$, in (6.4.1) are given as

$$
\begin{aligned}
h_1^{\mathrm{full}} \left(i, o_1 | \boldsymbol{\beta}_{\mathrm{full},1} \right) &= \beta_{1,0}^{\mathrm{full}} + \beta_{1,1}^{\mathrm{full}} \ln \left(d_{i,o_1} \right) + \beta_{1,2}^{\mathrm{full}} \ln \left(e_{i,o_1} \right), \\
h_2^{\mathrm{full}} \left(i, j, o_2 | \boldsymbol{\beta}_{\mathrm{full},2} \right) &= \beta_{2,0}^{\mathrm{full}} + \beta_{2,1}^{\mathrm{full}} \ln \left(d_{j,o_2} \right) + \beta_{2,2}^{\mathrm{full}} \ln \left(d_{j,i} \right) + \beta_{2,3}^{\mathrm{full}} \ln \left(d_{o_2,i} \right) \\
&\quad + \beta_{2,4}^{\mathrm{full}} \ln \left(e_{j,o_2} \right) + \beta_{2,5}^{\mathrm{full}} \ln \left(e_{j,i} \right) + \beta_{2,6}^{\mathrm{full}} \ln \left(e_{o_2,i} \right), \\
h_3^{\mathrm{full}} \left(i, j, k, o_3 | \boldsymbol{\beta}_{\mathrm{full},3} \right) &= \beta_{3,0}^{\mathrm{full}} + \beta_{3,1}^{\mathrm{full}} \ln \left(d_{k,o_3} \right) + \beta_{3,2}^{\mathrm{full}} \ln \left(d_{k,i} \right) + \beta_{3,3}^{\mathrm{full}} \ln \left(d_{k,j} \right) \\
&\quad + \beta_{3,4}^{\mathrm{full}} \ln \left(d_{o_3,i} \right) + \beta_{3,5}^{\mathrm{full}} \ln \left(d_{o_3,j} \right) \\
&\quad + \beta_{3,6}^{\mathrm{full}} \ln \left(e_{k,o_3} \right) + \beta_{3,7}^{\mathrm{full}} \ln \left(e_{k,i} \right) + \beta_{3,8}^{\mathrm{full}} \ln \left(e_{k,j} \right) \\
&\quad + \beta_{3,9}^{\mathrm{full}} \ln \left(e_{o_3,i} \right) + \beta_{3,10}^{\mathrm{full}} \ln \left(e_{o_3,j} \right),
\end{aligned}
$$

and it holds

$$
\begin{aligned}
\boldsymbol{\beta}_{\mathrm{full},1} &= \left(\beta_{1,0}^{\mathrm{full}}, \beta_{1,1}^{\mathrm{full}}, \beta_{1,2}^{\mathrm{full}} \right)^{\top} \in \mathbb{R}^3, \\
\boldsymbol{\beta}_{\mathrm{full},2} &= \left(\beta_{2,0}^{\mathrm{full}}, \ldots, \beta_{2,6}^{\mathrm{full}} \right)^{\top} \in \mathbb{R}^7, \\
\boldsymbol{\beta}_{\mathrm{full},3} &= \left(\beta_{3,0}^{\mathrm{full}}, \ldots, \beta_{3,10}^{\mathrm{full}} \right)^{\top} \in \mathbb{R}^{11},
\end{aligned}
$$

for the corresponding parameter vectors. The full reparametrization involves all available predictors. Its parameters are summarized in the vector

$$
\boldsymbol{\beta}_{\mathrm{full}}^{\mathrm{SCVM}} = \left(\boldsymbol{\beta}_{\mathrm{full},1}^{\top}, \boldsymbol{\beta}_{\mathrm{full},2}^{\top}, \boldsymbol{\beta}_{\mathrm{full},3}^{\top} \right)^{\top} \in \mathbb{R}^{n_{\mathrm{par}}^{\mathrm{full}}},
$$

where $n_{\mathrm{par}}^{\mathrm{full}} = 3 + 7 + 11 = 21$.

Next we give the model functions h_l^{mod}, $l = 1, 2, 3$, for the *distance reparametrization* (dist). They are defined as

$$h_1^{\mathrm{dist}} (i, o_1 | \boldsymbol{\beta}_{\mathrm{dist},1}) = \beta_{1,0}^{\mathrm{dist}} + \beta_{1,1}^{\mathrm{dist}} \ln (d_{i,o_1}),$$

$$h_2^{\mathrm{dist}} (i, j, o_2 | \boldsymbol{\beta}_{\mathrm{dist},2}) = \beta_{2,0}^{\mathrm{dist}} + \beta_{2,1}^{\mathrm{dist}} \ln (d_{j,o_2}) + \beta_{2,2}^{\mathrm{dist}} \ln (d_{j,i}) + \beta_{2,3}^{\mathrm{dist}} \ln (d_{o_2,i}),$$

$$h_3^{\mathrm{dist}} (i, j, k, o_3 | \boldsymbol{\beta}_{\mathrm{dist},3}) = \beta_{3,0}^{\mathrm{dist}} + \beta_{3,1}^{\mathrm{dist}} \ln (d_{k,o_3}) + \beta_{3,2}^{\mathrm{dist}} \ln (d_{k,i}) + \beta_{3,3}^{\mathrm{dist}} \ln (d_{k,j})$$
$$+ \beta_{3,4}^{\mathrm{dist}} \ln (d_{o_3,i}) + \beta_{3,5}^{\mathrm{dist}} \ln (d_{o_3,j}),$$

with

$$\boldsymbol{\beta}_{\mathrm{dist},1} = \left(\beta_{1,0}^{\mathrm{dist}}, \beta_{1,1}^{\mathrm{dist}} \right)^\top \in \mathbb{R}^2,$$

$$\boldsymbol{\beta}_{\mathrm{dist},2} = \left(\beta_{2,0}^{\mathrm{dist}}, \ldots, \beta_{2,3}^{\mathrm{dist}} \right)^\top \in \mathbb{R}^4,$$

$$\boldsymbol{\beta}_{\mathrm{dist},3} = \left(\beta_{3,0}^{\mathrm{dist}}, \ldots, \beta_{3,5}^{\mathrm{dist}} \right)^\top \in \mathbb{R}^6.$$

All possible distances are included. The occurring parameters are summarized as

$$\boldsymbol{\beta}_{\mathrm{dist}}^{\mathrm{SCVM}} = \left(\boldsymbol{\beta}_{\mathrm{dist},1}^\top, \boldsymbol{\beta}_{\mathrm{dist},2}^\top, \boldsymbol{\beta}_{\mathrm{dist},3}^\top \right)^\top \in \mathbb{R}^{n_{\mathrm{par}}^{\mathrm{dist}}},$$

where $n_{\mathrm{par}}^{\mathrm{dist}} = 2 + 4 + 6 = 12$.

The *d0 reparametrization* (d0) is defined by the model functions

$$h_1^{\mathrm{d0}} (i, o_1 | \boldsymbol{\beta}_{\mathrm{d0},1}) = \beta_{1,0}^{\mathrm{d0}} + \beta_{1,1}^{\mathrm{d0}} \ln (d_{i,o_1}),$$

$$h_2^{\mathrm{d0}} (i, j, o_2 | \boldsymbol{\beta}_{\mathrm{d0},2}) = \beta_{2,0}^{\mathrm{d0}} + \beta_{2,1}^{\mathrm{d0}} \ln (d_{j,o_2}),$$

$$h_3^{\mathrm{d0}} (i, j, k, o_3 | \boldsymbol{\beta}_{\mathrm{d0},3}) = \beta_{3,0}^{\mathrm{d0}} + \beta_{3,1}^{\mathrm{d0}} \ln (d_{k,o_3}),$$

where

$$\boldsymbol{\beta}_{\mathrm{d0},1} = \left(\beta_{1,0}^{\mathrm{d0}}, \beta_{1,1}^{\mathrm{d0}} \right)^\top \in \mathbb{R}^2,$$

$$\boldsymbol{\beta}_{\mathrm{d0},2} = \left(\beta_{2,0}^{\mathrm{d0}}, \beta_{2,1}^{\mathrm{d0}} \right)^\top \in \mathbb{R}^2,$$

$$\boldsymbol{\beta}_{\mathrm{d0},3} = \left(\beta_{3,0}^{\mathrm{d0}}, \beta_{3,1}^{\mathrm{d0}} \right)^\top \in \mathbb{R}^2.$$

Only the distances indicated by the conditioned sets of the copulae are involved. The parameters of the d0 reparametrization are summarized as

$$\boldsymbol{\beta}_{\mathrm{d0}}^{\mathrm{SCVM}} = \left(\boldsymbol{\beta}_{\mathrm{d0},1}^\top, \boldsymbol{\beta}_{\mathrm{d0},2}^\top, \boldsymbol{\beta}_{\mathrm{d0},3}^\top \right)^\top \in \mathbb{R}^{n_{\mathrm{par}}^{\mathrm{d0}}},$$

where $n_{\mathrm{par}}^{\mathrm{d0}} = 3 \cdot 2 = 6$.

Again we also consider reparametrizations called *elevation reparametrization* (elev) and *e0 reparametrization* (e0). They are obtained analogous to the distance and d0 reparametrization, by choosing elevation differences instead of distances.

Eventually the *d0+e0 reparametrization* (de) is given by

$$h_1^{\mathrm{de}} (i, o_1 | \boldsymbol{\beta}_{\mathrm{de},1}) = \beta_{1,0}^{\mathrm{de}} + \beta_{1,1}^{\mathrm{de}} \ln (d_{i,o_1}) + \beta_{1,2}^{\mathrm{de}} \ln (e_{i,o_1}),$$

$$h_2^{\mathrm{de}} (i, j, o_2 | \boldsymbol{\beta}_{\mathrm{de},2}) = \beta_{2,0}^{\mathrm{de}} + \beta_{2,1}^{\mathrm{de}} \ln (d_{j,o_2}) + \beta_{2,2}^{\mathrm{de}} \ln (e_{j,o_2}),$$

$$h_3^{\mathrm{de}}\left(i, j, k, o_3|\boldsymbol{\beta}_{\mathrm{de},3}\right) = \beta_{3,0}^{\mathrm{de}} + \beta_{3,1}^{\mathrm{de}} \ln\left(d_{k,o_3}\right) + \beta_{3,2}^{\mathrm{de}} \ln\left(e_{k,o_3}\right),$$

and

$$\boldsymbol{\beta}_{\mathrm{de},1} = \left(\beta_{1,0}^{\mathrm{de}}, \beta_{1,1}^{\mathrm{de}}, \beta_{1,2}^{\mathrm{de}}\right)^{\top} \in \mathbb{R}^3,$$
$$\boldsymbol{\beta}_{\mathrm{de},2} = \left(\beta_{2,0}^{\mathrm{de}}, \beta_{2,1}^{\mathrm{de}}, \beta_{2,2}^{\mathrm{de}}\right)^{\top} \in \mathbb{R}^3,$$
$$\boldsymbol{\beta}_{\mathrm{de},3} = \left(\beta_{3,0}^{\mathrm{de}}, \beta_{3,1}^{\mathrm{de}}, \beta_{3,2}^{\mathrm{de}}\right)^{\top} \in \mathbb{R}^3.$$

It involves the distances and elevation differences indicated by the respective conditioned sets. The summary of its parameters is given by

$$\boldsymbol{\beta}_{\mathrm{de}}^{\mathrm{SCVM}} = \left(\boldsymbol{\beta}_{\mathrm{de},1}^{\top}, \boldsymbol{\beta}_{\mathrm{de},2}^{\top}, \boldsymbol{\beta}_{\mathrm{de},3}^{\top}\right)^{\top} \in \mathbb{R}^{n_{\mathrm{par}}^{\mathrm{de}}}$$

where $n_{\mathrm{par}}^{\mathrm{de}} = 3 \cdot 3 = 9$.

In order to terminate the reparametrization of the composite vine model parameters, a reparametrization of the second copula parameters respectively the degrees of freedom $\boldsymbol{\nu}^{\mathrm{CVM}}$ is required. As before, we investigate the parameters stemming from the composite vine model on the mean temperature data.

Let $\widehat{\nu}_{ij|\ldots}$ denote any arbitrary degrees of freedom parameter of our composite vine model, irrespective of the tree which the parameter stems from. Figure 6.4.4 shows plots of the parameters $\widehat{\nu}_{ij|\ldots}$ against their respective tree number and against the distances $d_{i,j}$ and elevation differences $e_{i,j}$. From the plots we observe, that the average degrees of freedom rise with the tree number. Moreover there seems to be a significant trend of the parameters $\widehat{\nu}_{ij|\ldots}$ with respect to the distances $d_{i,j}$, but there is no evident influence of the elevation differences.

Due to our observations on Figure 6.4.4, we are going to use the reparametrization

$$\widetilde{\nu}_{io_1} = \exp\left(\beta_0^{\nu} + \beta_1^{\nu} \cdot 1 + \beta_2^{\nu} d_{i,o_1}\right),$$
$$\widetilde{\nu}_{jo_2|i} = \exp\left(\beta_0^{\nu} + \beta_1^{\nu} \cdot 2 + \beta_2^{\nu} d_{j,o_2}\right), \tag{6.4.2}$$
$$\widetilde{\nu}_{ko_3|ij} = \exp\left(\beta_0^{\nu} + \beta_1^{\nu} \cdot 3 + \beta_2^{\nu} d_{k,o_3}\right),$$

of the second copula parameters, in the following. The respective parameters are summarized as

$$\boldsymbol{\beta}_{\nu}^{\mathrm{SCVM}} = \left(\beta_0^{\nu}, \beta_1^{\nu}, \beta_2^{\nu}\right)^{\top} \in \mathbb{R}^{n_{\mathrm{par}}^{\nu}},$$

where $n_{\mathrm{par}}^{\nu} = 3$.

A (three neighbors) composite vine model which is reparametrized by reparametrizations of the type (6.4.1) and (6.4.2), will be called (three neighbors) *spatial composite vine model (SCVM)*. In order to estimate the respective model parameters, the composite log-likelihood (6.2.1) parametrized via (6.4.1) and (6.4.2) has to be maximized. This yields the maximum composite likelihood estimates (mcle)

$$\widehat{\boldsymbol{\beta}}_{\mathrm{mcle}}^{\mathrm{SCVM}} = \left(\widehat{\boldsymbol{\beta}}_{\mathrm{mod}}^{\mathrm{SCVM}}, \widehat{\boldsymbol{\beta}}_{\nu}^{\mathrm{SCVM}}\right)^{\top} \in \mathbb{R}^{\left(n_{\mathrm{par}}^{\mathrm{mod}} + n_{\mathrm{par}}^{\nu}\right)}.$$

From now on, we will address the above specified spatial composite vine models according to the selected reparametrization as full, distance, d0, elevation, e0 or d0+e0 model.
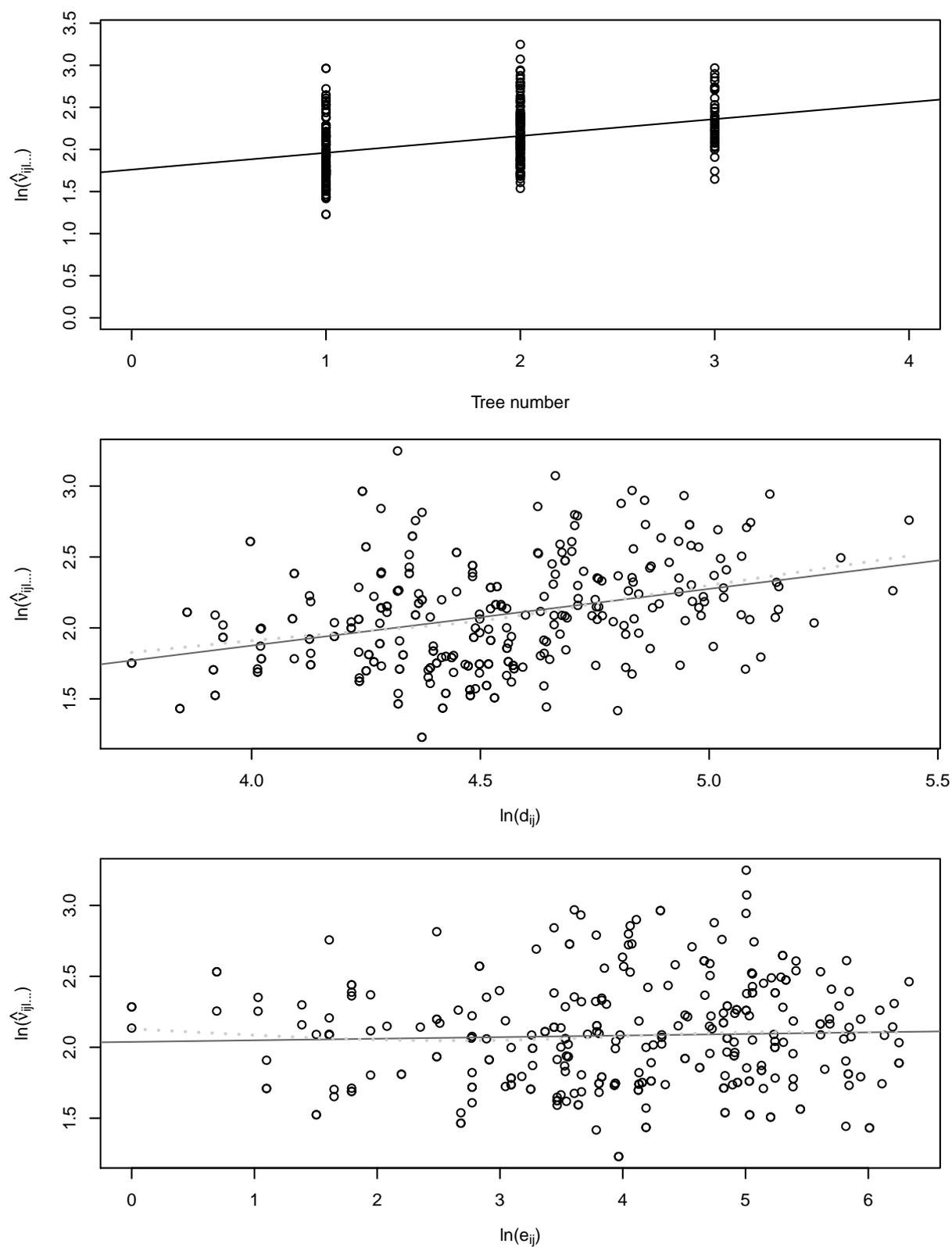
Figure 6.4.4: Plots of $\widehat{\nu}_{ij|\ldots}$ against tree number, $\ln(d_{i,j})$ and $\ln(e_{i,j})$, respectively.

6.5 A spatial composite vine model for mean temperature

This application of the previously specified model is intended to investigate a spatial composite vine model for the mean temperature data set. For this purpose we are going to choose a suitable reparametrization of the composite vine model which was set up in Section 6.3.

6.5.1 Model selection

Table 6.3 summarizes the required number of parameters for the six models specified up to now. Even the full model needs only 24 parameters. We will conduct further analyses in the following, intended to check, if the number of parameters can be reduced even more.

	full	distance	d0	elevation	e0	d0+e0
Tree 1	3	2	2	2	2	3
Tree 2	7	4	2	4	2	3
Tree 3	11	6	2	6	2	3
n_{par}^{ν}	3	3	3	3	3	3
Sum	**24**	**15**	**9**	**15**	**9**	**12**

Table 6.3: Number of parameters for the 'full', 'distance', 'd0', 'elevation', 'e0' and 'd0+e0' model.

In order to get a clue about the explanatory power of the six reparametrizations of the first copula parameters given in Section 6.4, we investigate the following linear models. For all reparametrizations (mod) among the reparametrizations (full), (dist), (d0), (elev), (e0), and (de), we define the linear models

$$\mathrm{F}_z(\tau_{so}) = h_1^{\mathrm{mod}}(s, o|\boldsymbol{\beta}_{\mathrm{mod},1}) + \varepsilon_{so}, \quad \varepsilon_{so} \overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(0, \sigma^2\right), \quad o \in \{p_s, q_s, r_s\},$$

$$\mathrm{F}_z(\tau_{p_so|s}) = h_2^{\mathrm{mod}}(s, p_s, o|\boldsymbol{\beta}_{\mathrm{mod},2}) + \varepsilon_{p_so|s}, \quad \varepsilon_{p_so|s} \overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(0, \sigma^2\right), \quad o \in \{q_s, r_s\},$$

$$\mathrm{F}_z(\tau_{q_sr_s|sp_s}) = h_3^{\mathrm{mod}}(s, p_s, q_s, r_s|\boldsymbol{\beta}_{\mathrm{mod},3}) + \varepsilon_{q_sr_s|sp_s}, \quad \varepsilon_{q_sr_s|sp_s} \overset{\text{i.i.d.}}{\sim} \mathcal{N}\left(0, \sigma^2\right),$$

where $s = 1, \dots, 54$. The R^2 and R^2_{adj} which are obtained from fitting these linear modes are summarized in the tables 6.4 and 6.5. The resulting parameter estimates can be used as start parameters for $\boldsymbol{\beta}_{\mathrm{mod}}$ for the purpose of fitting a spatial composite vine model on the mean temperature data set.

Tree	full	distance	d0	elevation	e0	d0+e0
1	0.5789	0.4873	0.4873	0.0980	0.0980	0.5789
2	0.7349	0.7183	0.4548	0.0714	0.0058	0.4565
3	0.8107	0.7803	0.6334	0.1833	0.0356	0.6341

Table 6.4: R^2 for the 'full', 'distance', 'd0', 'elevation', 'e0' and 'd0+e0' model.

Tree	full	distance	d0	elevation	e0	d0+e0
1	0.5736	0.4841	0.4841	0.0924	0.0924	0.5736
2	0.7191	0.7102	0.4497	0.0447	-0.0035	0.4462
3	0.7667	0.7575	0.6264	0.0983	0.0171	0.6198

Table 6.5: R^2_{adj} for the 'full', 'distance', 'd0', 'elevation', 'e0' and 'd0+e0' model.

Comparison of the values given in Table 6.4 and Table 6.5 again yields bad performance of the elevation and e0 model, which include only elevation differences. For the first trees the d0+e0 respectively the full model is distinctly superior compared to the other models. Considering the R^2_{adj} for the second and third trees, the d0 model outperforms the d0+e0 model. Extension of the d0 model to the distance model yields a considerable improvement in terms of the R^2, respectively R^2_{adj}. Since the distance model performs similarly well as the full model for the trees two and three, we decide to choose a mixed reparametrization (select) of the first copula parameters, which is build up of the (de) reparametrization for the first trees and the (dist) reparametrization for the second and third trees.

The model functions h_l^{select}, $l = 1, 2, 3$, of the reparametrization (select) are given as

$$
\begin{aligned}
h_1^{\mathrm{select}}\left(i, o_1 | \boldsymbol{\beta}_{\mathrm{select},1}\right) &= h_1^{\mathrm{de}}\left(i, o_1 | \boldsymbol{\beta}_{\mathrm{select},1}\right), \\
h_2^{\mathrm{select}}\left(i, j, o_2 | \boldsymbol{\beta}_{\mathrm{select},2}\right) &= h_2^{\mathrm{dist}}\left(i, j, o_2 | \boldsymbol{\beta}_{\mathrm{select},2}\right), \\
h_3^{\mathrm{select}}\left(i, j, k, o_3 | \boldsymbol{\beta}_{\mathrm{select},3}\right) &= h_3^{\mathrm{dist}}\left(i, j, k, o_3 | \boldsymbol{\beta}_{\mathrm{select},3}\right),
\end{aligned}
\tag{6.5.1}
$$

where

$$
\begin{aligned}
\boldsymbol{\beta}_{\mathrm{select},1} &= \left(\beta_{1,0}^{\mathrm{select}}, \beta_{1,1}^{\mathrm{select}}, \beta_{1,2}^{\mathrm{select}}\right)^{\top} \in \mathbb{R}^3, \\
\boldsymbol{\beta}_{\mathrm{select},2} &= \left(\beta_{2,0}^{\mathrm{select}}, \ldots, \beta_{2,3}^{\mathrm{select}}\right)^{\top} \in \mathbb{R}^4, \\
\boldsymbol{\beta}_{\mathrm{select},3} &= \left(\beta_{3,0}^{\mathrm{select}}, \ldots, \beta_{3,5}^{\mathrm{select}}\right)^{\top} \in \mathbb{R}^6.
\end{aligned}
$$

Its parameters are summarized as

$$
\boldsymbol{\beta}_{\mathrm{select}}^{\mathrm{SCVM}} = \left(\boldsymbol{\beta}_{\mathrm{select},1}^{\top}, \boldsymbol{\beta}_{\mathrm{select},2}^{\top}, \boldsymbol{\beta}_{\mathrm{select},3}^{\top}\right)^{\top} \in \mathbb{R}^{n_{\mathrm{par}}^{\mathrm{select}}},
$$

where $n_{\mathrm{par}}^{\mathrm{select}} = 3 + 4 + 6 = 13$.

It remains to analyze the reparametrization (6.4.2) of the second copula parameters. To this end we fit the linear model defined by

$$
\begin{aligned}
\ln(\nu_{so}) &= \beta_0^{\nu} + \beta_1^{\nu} l + \beta_2^{\nu} d_{s,o} + \varepsilon_{so}, \quad o \in \{p_s, q_s, r_s\}, \\
\ln(\nu_{p_s o | s}) &= \beta_0^{\nu} + \beta_1^{\nu} l + \beta_2^{\nu} d_{p_s, o} + \varepsilon_{p_s o | s}, \quad o \in \{q_s, r_s\}, \\
\ln(\nu_{q_s o | s p_s}) &= \beta_0^{\nu} + \beta_1^{\nu} l + \beta_2^{\nu} d_{q_s, o} + \varepsilon_{q_s o | s p_s}, \quad o \in \{r_s\},
\end{aligned}
$$

where $s = 1, \ldots, 54$ and $\varepsilon_{s p_s}, \varepsilon_{s q_s}, \varepsilon_{s r_s}, \varepsilon_{p_s q_s | s}, \varepsilon_{p_s r_s | s}, \varepsilon_{q_s r_s | s p_s} \overset{\mathrm{i.i.d.}}{\sim} \mathcal{N}\left(0, \sigma^2\right)$ for $s = 1, \ldots, 54$. The respective parameter estimates, their standard errors, Wald statistics and p-values are given in Table 6.6. The estimates for β_0^{ν}, β_1^{ν} and β_2^{ν} are significant at a significance level $\alpha = 5\%$. These findings yield that the reparametrization (6.4.2) is adequate for our

	Estimate	Std. Error	t-value	p-value
β_0^ν	0.6745	0.2900	2.33	0.0207
β_1^ν	0.1532	0.0299	5.13	0.0000
β_2^ν	0.2571	0.0677	3.80	0.0002

Table 6.6: Parameter estimates, estimated standard errors, Wald statistics and p-values of the model for the second parameters ν_{ij}.

purposes. As before, the obtained parameter estimates can be used as start values for $\boldsymbol{\beta}_\nu$ within the framework of model fitting.

All in all we have selected a reparametrization according to (6.5.1) and (6.4.2), which is based on $n_{\mathrm{par}}^{\mathrm{select}} + n_{\mathrm{par}}^\nu = 16$ parameters. In the following we will call the resulting model *selected spatial composite vine model.*

6.5.2 Model fit

Next we are going to fit our selected spatial composite vine model to the data. This is conducted by maximization of the composite log-likelihood (6.2.1) which is reparametrized by means of (6.5.1) and (6.4.2). This time we get a maximum composite log-likelihood of $c\ell_{\mathrm{SCVM}}\left(\widehat{\boldsymbol{\beta}}_{\mathrm{select}}^{\mathrm{SCVM}}, \widehat{\boldsymbol{\beta}}_\nu^{\mathrm{SCVM}} \mid \boldsymbol{u}^1, \ldots, \boldsymbol{u}^d\right) = 40938.80$, AIC and BIC are given as $\mathrm{AIC}_{\mathrm{SCVM}} = -81845.6$ and $\mathrm{BIC}_{\mathrm{SCVM}} = -81765.7$. Compared to the composite vine model the composite log-likelihood of the spatial composite vine model shrinked by about 1168. Due to the fact that the number of parameters is reduced considerably, the BIC of the spatial composite vine model is smaller. Comparison based on the AIC would however favor the composite vine model.

The fitting procedure yields the parameter estimates $\widehat{\boldsymbol{\beta}}_{\mathrm{mcle}}^{\mathrm{SCVM}} := (\widehat{\boldsymbol{\beta}}_{\mathrm{select}}^{\mathrm{SCVM}}, \widehat{\boldsymbol{\beta}}_\nu^{\mathrm{SCVM}})^\top \in \mathbb{R}^{16}$, which are given in Table 6.7. The table moreover compares the parameter estimates to their respective start values, which were already addressed in the previous subsection.

no.	par	start	mcle	no.	par	start	mcle	no.	par	start	mcle
1	$\beta_{1,0}^{\mathrm{select}}$	2.300	2.251	7	$\beta_{2,3}^{\mathrm{select}}$	0.297	0.248	13	$\beta_{3,5}^{\mathrm{select}}$	0.120	0.119
2	$\beta_{1,1}^{\mathrm{select}}$	-0.305	-0.299	8	$\beta_{3,0}^{\mathrm{select}}$	0.256	0.704	14	β_0^ν	0.675	0.404
3	$\beta_{1,2}^{\mathrm{select}}$	-0.025	-0.022	9	$\beta_{3,1}^{\mathrm{select}}$	-0.434	-0.405	15	β_1^ν	0.153	0.015
4	$\beta_{2,0}^{\mathrm{select}}$	0.306	0.577	10	$\beta_{3,2}^{\mathrm{select}}$	0.109	0.076	16	β_2^ν	0.257	0.271
5	$\beta_{2,1}^{\mathrm{select}}$	-0.510	-0.480	11	$\beta_{3,3}^{\mathrm{select}}$	0.089	0.016				
6	$\beta_{2,2}^{\mathrm{select}}$	0.222	0.175	12	$\beta_{3,4}^{\mathrm{select}}$	0.115	0.095				

Table 6.7: Start values and maximum composite-likelihood estimates of the spatial composite vine model for the mean temperature data.

6.5.3   Prediction

Again we aim to predict from our newly developed model. The structure of the spatial composite vine model suggests to perform prediction based on a four-dimensional C-vine which is composed out of the variables corresponding to the new location s from which to predict and to its three closest neighbors among the stations $1, \ldots, d$. As before we denote the closest observation station as p_s, the second closest neighbor as q_s and the third closest station as r_s.

For the purpose of prediction we are going to sample from the conditional distribution $F_{s|p_s q_s r_s}(u^s | u^{p_s}, u^{q_s}, u^{r_s})$. It has to be calculated iteratively according to Lemma 4.7. Based on the copulae needed for these calculations, we are going to set up the corresponding C-vine. For convenience we drop the arguments of the conditional distribution functions F in the following, since it is anyway indicated in the indices which arguments to use. Moreover the estimates of the first and the second copula parameters, whose calculation will be presented later on, are denoted by $\widehat{\theta}$ and $\widehat{\nu}$, respectively, the corresponding copula families are denoted by t with respective indices.

The conditional distribution $F_{s|p_s q_s r_s}$ is calculated as

$$F_{s|p_s q_s r_s} = \frac{\partial C_{r_s s; p_s q_s}\left(F_{s|p_s q_s}, F_{r_s|p_s q_s}; \widehat{\theta}_{r_s s|p_s q_s}, \widehat{\nu}_{r_s s|p_s q_s}, t_{r_s s|p_s q_s}\right)}{\partial F_{r_s|p_s q_s}}.$$

This yields that we are going to use $r_s s | p_s q_s$ as edge for the third C-vine tree. The iteration continues with the two equations

$$F_{r_s|p_s q_s} = \frac{\partial C_{q_s r_s; p_s}\left(F_{r_s|p_s}, F_{q_s|p_s}; \widehat{\theta}_{q_s r_s|p_s}, \widehat{\nu}_{q_s r_s|p_s}, t_{q_s r_s|p_s}\right)}{\partial F_{q_s|p_s}},$$

and

$$F_{s|p_s q_s} = \frac{\partial C_{q_s s; p_s}\left(F_{s|p_s}, F_{q_s|p_s}; \widehat{\theta}_{q_s s|p_s}, \widehat{\nu}_{q_s s|p_s}, t_{q_s s|p_s}\right)}{\partial F_{q_s|p_s}},$$

which yields the second tree edges $q_s r_s | p_s$ and $q_s s | p_s$. The arguments needed above are furthermore calculated via

$$F_{q_s|p_s} = \frac{\partial C_{p_s q_s}\left(F_{q_s}, F_{p_s}; \widehat{\theta}_{p_s q_s}, \widehat{\nu}_{p_s q_s}, t_{p_s q_s}\right)}{\partial F_{p_s}} = \frac{\partial C_{p_s q_s}\left(u^{q_s}, u^{p_s}; \widehat{\theta}_{p_s q_s}, \widehat{\nu}_{p_s q_s}, t_{p_s q_s}\right)}{\partial u^{p_s}},$$

$$F_{r_s|p_s} = \frac{\partial C_{p_s r_s}\left(F_{r_s}, F_{p_s}; \widehat{\theta}_{p_s r_s}, \widehat{\nu}_{p_s r_s}, t_{p_s r_s}\right)}{\partial F_{p_s}} = \frac{\partial C_{p_s r_s}\left(u^{r_s}, u^{p_s}; \widehat{\theta}_{p_s r_s}, \widehat{\nu}_{p_s r_s}, t_{p_s r_s}\right)}{\partial u^{p_s}},$$

$$F_{s|p_s} = \frac{\partial C_{p_s s}\left(F_s, F_{p_s}; \widehat{\theta}_{p_s s}, \widehat{\nu}_{p_s s}, t_{p_s s}\right)}{\partial F_{p_s}} = \frac{\partial C_{p_s s}\left(u^s, u^{p_s}; \widehat{\theta}_{p_s s}, \widehat{\nu}_{p_s s}, t_{p_s s}\right)}{\partial u^{p_s}}.$$

The unconditioned distributions F_s, F_{p_s}, F_{q_s}, F_{r_s} involved here are known to be the cumulative distribution functions of $\mathcal{U}(0,1)$ distributed random variables. Following it
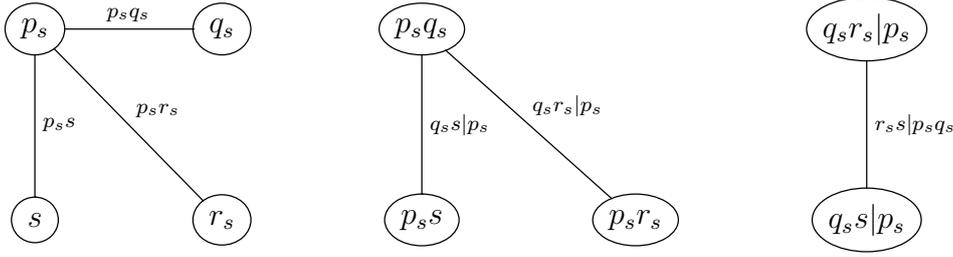
Figure 6.5.1: C-vine structure in order to sample from $F_{s|p_sq_sr_s}(u^s|u^{p_s}, u^{q_s}, u^{r_s})$.

holds that $F_o(u^o) = u^o$ for $o \in \{s, p_s, q_s, r_s\}$. We gain p_sq_s, p_sr_s and p_ss for the three edges of the first C-vine tree. All in all we obtain the C-vine illustrated in Figure 6.5.1.

In order to be able to simulate from $F_{s|p_sq_sr_s}(u^s|u^{p_s}, u^{q_s}, u^{r_s})$, we have to specify the copula families and the respective parameters involved in the calculations above. The first copula parameters are calculated based on the model functions of the selected spatial composite vine model and the respective parameter estimates according to

$$\widehat{\theta}_{p_sq_s} = \mathrm{T}_{\tau \to \theta} \left(\mathrm{F}_z^{-1} \left(h_1^{\mathrm{select}} \left(p_s, q_s | \widehat{\boldsymbol{\beta}}_{\mathrm{select},1} \right) \right) ; t_{p_sq_s} \right),$$

$$\widehat{\theta}_{p_sr_s} = \mathrm{T}_{\tau \to \theta} \left(\mathrm{F}_z^{-1} \left(h_1^{\mathrm{select}} \left(p_s, r_s | \widehat{\boldsymbol{\beta}}_{\mathrm{select},1} \right) \right) ; t_{p_sr_s} \right),$$

$$\widehat{\theta}_{p_ss} = \mathrm{T}_{\tau \to \theta} \left(\mathrm{F}_z^{-1} \left(h_1^{\mathrm{select}} \left(p_s, s | \widehat{\boldsymbol{\beta}}_{\mathrm{select},1} \right) \right) ; t_{p_ss} \right),$$

$$\widehat{\theta}_{q_sr_s|p_s} = \mathrm{T}_{\tau \to \theta} \left(\mathrm{F}_z^{-1} \left(h_2^{\mathrm{select}} \left(p_s, q_s, r_s | \widehat{\boldsymbol{\beta}}_{\mathrm{select},2} \right) \right) ; t_{q_sr_s|p_s} \right),$$

$$\widehat{\theta}_{q_ss|p_s} = \mathrm{T}_{\tau \to \theta} \left(\mathrm{F}_z^{-1} \left(h_2^{\mathrm{select}} \left(p_s, q_s, s | \widehat{\boldsymbol{\beta}}_{\mathrm{select},2} \right) \right) ; t_{q_ss|p_s} \right),$$

$$\widehat{\theta}_{r_ss|p_sq_s} = \mathrm{T}_{\tau \to \theta} \left(\mathrm{F}_z^{-1} \left(h_3^{\mathrm{select}} \left(p_s, q_s, r_s, s | \widehat{\boldsymbol{\beta}}_{\mathrm{select},3} \right) \right) ; t_{r_ss|p_sq_s} \right),$$

where $t_{p_sq_s}$, $t_{p_sr_s}$, t_{p_ss}, $t_{q_sr_s|p_s}$, $t_{q_ss|p_s}$ and $t_{r_ss|p_sq_s}$ denote the respective copula families. Due to the fact that most copulae in our selected spatial composite vine model were determined to be Student-t copulae (see Table 6.2), we choose the Student-t family for all of these families. Following we have to determine a second copula parameter for each of the six Student-t copulae involved. According to the reparametrization (6.4.2), we obtain

$$\widehat{\nu}_{p_sq_s} = \exp \left(\widehat{\beta}_0^\nu + \widehat{\beta}_1^\nu + \widehat{\beta}_2^\nu d_{p_s,q_s} \right),$$

$$\widehat{\nu}_{p_sr_s} = \exp \left(\widehat{\beta}_0^\nu + \widehat{\beta}_1^\nu + \widehat{\beta}_2^\nu d_{p_s,r_s} \right),$$

$$\widehat{\nu}_{p_ss} = \exp \left(\widehat{\beta}_0^\nu + \widehat{\beta}_1^\nu + \widehat{\beta}_2^\nu d_{p_s,s} \right),$$

$$\widehat{\nu}_{q_sr_s|p_s} = \exp \left(\widehat{\beta}_0^\nu + 2\widehat{\beta}_1^\nu + \widehat{\beta}_2^\nu d_{q_s,r_s} \right),$$

$$\widehat{\nu}_{q_ss|p_s} = \exp \left(\widehat{\beta}_0^\nu + 2\widehat{\beta}_1^\nu + \widehat{\beta}_2^\nu d_{q_s,s} \right),$$

$$\widehat{\nu}_{r_ss|p_sq_s} = \exp \left(\widehat{\beta}_0^\nu + 3\widehat{\beta}_1^\nu + \widehat{\beta}_2^\nu d_{r_s,s} \right).$$

Furthermore we are interested in the predictive density $f_{s|p_sq_sr_s}$ corresponding to $F_{s|p_sq_sr_s}$. It is calculated similarly to the conditioned R-vine density in Subsection 5.3.3 of

Chapter 5. We obtain

$$f_{s|p_sq_sr_s} = c_{p_ss}\left(F_{p_s}, F_s; \widehat{\theta}_{p_ss}, \widehat{\nu}_{p_ss}, t_{p_ss}\right)$$
$$\cdot c_{q_ss;p_s}\left(F_{q_s|p_s}, F_{s|p_s}; \widehat{\theta}_{q_ss|p_s}, \widehat{\nu}_{q_ss|p_s}, t_{q_ss|p_s}\right) \qquad (6.5.2)$$
$$\cdot c_{r_ss;p_sq_s}\left(F_{r_s|p_sq_s}, F_{s|p_sq_s}; \widehat{\theta}_{r_ss|p_sq_s}, \widehat{\nu}_{r_ss|p_sq_s}, t_{r_ss|p_sq_s}\right).$$

This concludes the methodology of prediction from spatial composite R-vine models on the copula data level. The copula data obtained from the simulations can be transformed back to the original data level as described in Subsection 5.3.3 of the previous chapter.

The results of the spatial composite R-vine model based predictions of the mean temperatures at the 24 observation stations constituting the validation data set are visualized in Figures A.2.1-A.2.12. Since they show the results of the predictions in analogy to Figures A.1.1-A.1.12 and due to the fact that the results are pretty similar, however not identical, we do not discuss every figure in detail and refer instead to the observations of Subsection 5.3.3 in Chapter 5. The respective plots for the three selected stations Grünow (70), Arkona (57) and Großer Arber (69) are given by Figures 6.5.2-6.5.4.

We again summarize the mean squared prediction errors in Table 6.8. As before *Arkona* (57) has an outstanding high mean squared error among the stations with elevations smaller than 500 meters. The reason why we have problems with the prediction of mean temperatures at Arkona may originate from the location of the observation station. Due to the fact that the observation station Arkona lies on an island in the Baltic Sea, it has an outstanding position compared to the rest of our observation stations and it may be exposed to factors which are not captured by our models. Once again the mean squared errors and the plots for the observation stations *Neuhaus am Rennweg* (75) and *Großer Arber* (69) show obviously that the prediction fails in these cases.

s	short name	elevation	MSE(s)	s	short name	elevation	MSE(s)
61	bork	3.00	0.55	63	buch	340.00	1.21
62	bvoe	10.00	0.38	65	ebra	346.00	0.92
72	luec	17.00	0.30	60	blan	417.00	0.66
68	gram	27.00	0.22	66	ellw	460.00	0.90
64	cosc	40.00	0.92	59	augs	461.40	0.47
57	arko	42.00	2.32	67	falk	472.00	0.41
77	rahd	42.00	0.25	76	ohrz	494.00	0.72
70	grue	55.90	0.19	71	hohe	743.30	3.42
56	alfe	143.90	0.23	55	albs	759.00	2.04
58	arns	159.00	0.98	75	neuh	845.00	6.45
78	wies	187.00	0.99	73	mitt	981.00	2.90
74	muel	273.00	0.68	69	garb	1436.00	15.07

Table 6.8: Mean squared errors of the mean temperature predictions over the period 01/01/2010-12/31/2012 for the observation stations of the validation data set based on the spatial composite vine model.

Whereas the mean squared errors given in the Tables 5.11 and 6.8 allow a first comparison of the different prediction methods respectively models, we are going to perform an extensive model comparison in the following chapter, which is amongst others based on scoring.
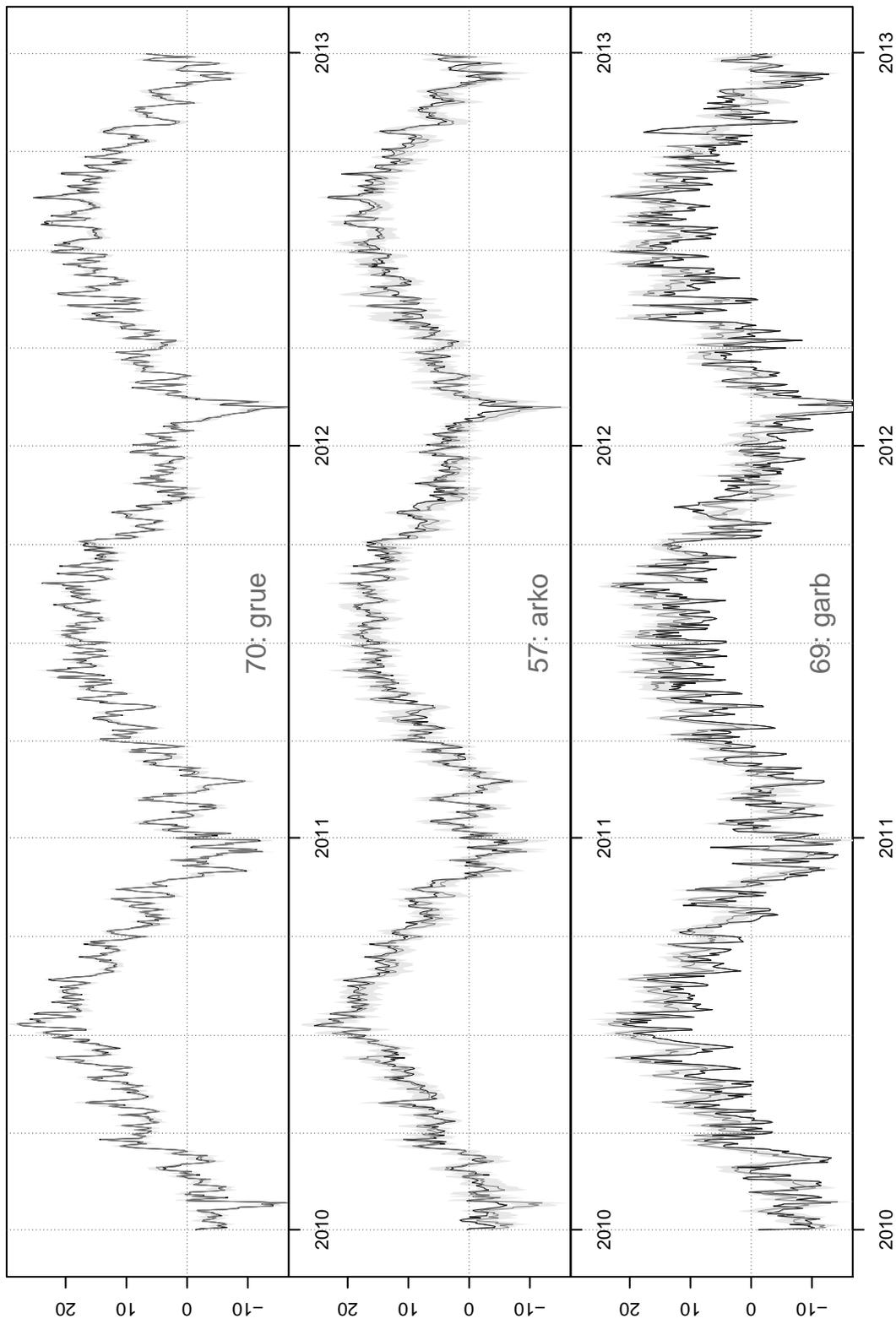
Figure 6.5.2: Prediction of the mean temperatures for the observation stations Grünow (70), Arkona (57) and Großer Arber (69) for the period 01/01/2010-12/31/2012 based on the spatial composite vine model. black line: observed values. dark gray line: prediction. light gray area: 95% prediction intervals.
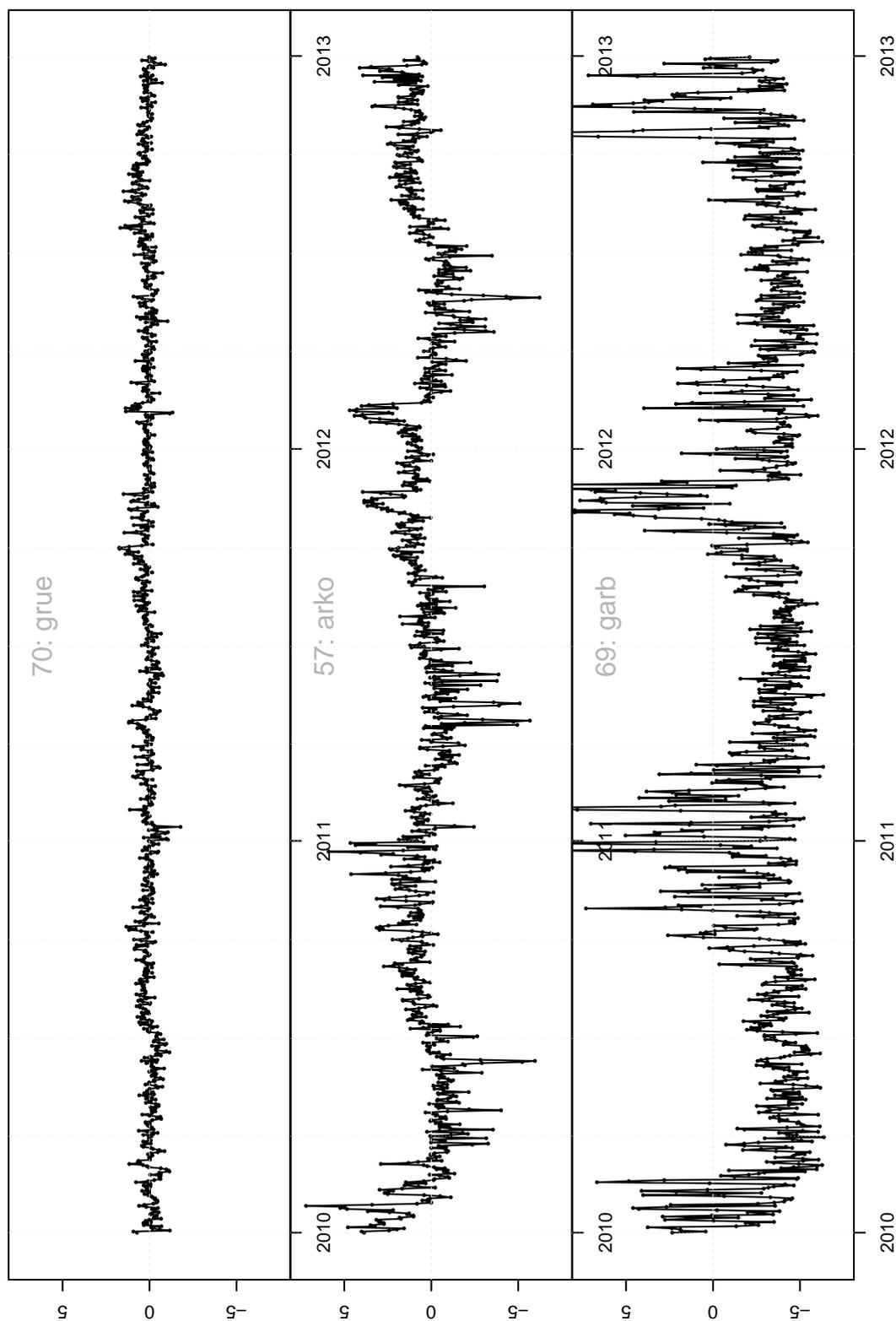
Figure 6.5.3: Prediction errors of the predictions for the observation stations Grünow (70), Arkona (57) and Großer Arber (69) for the period 01/01/2010-12/31/2012 based on the spatial composite vine model.
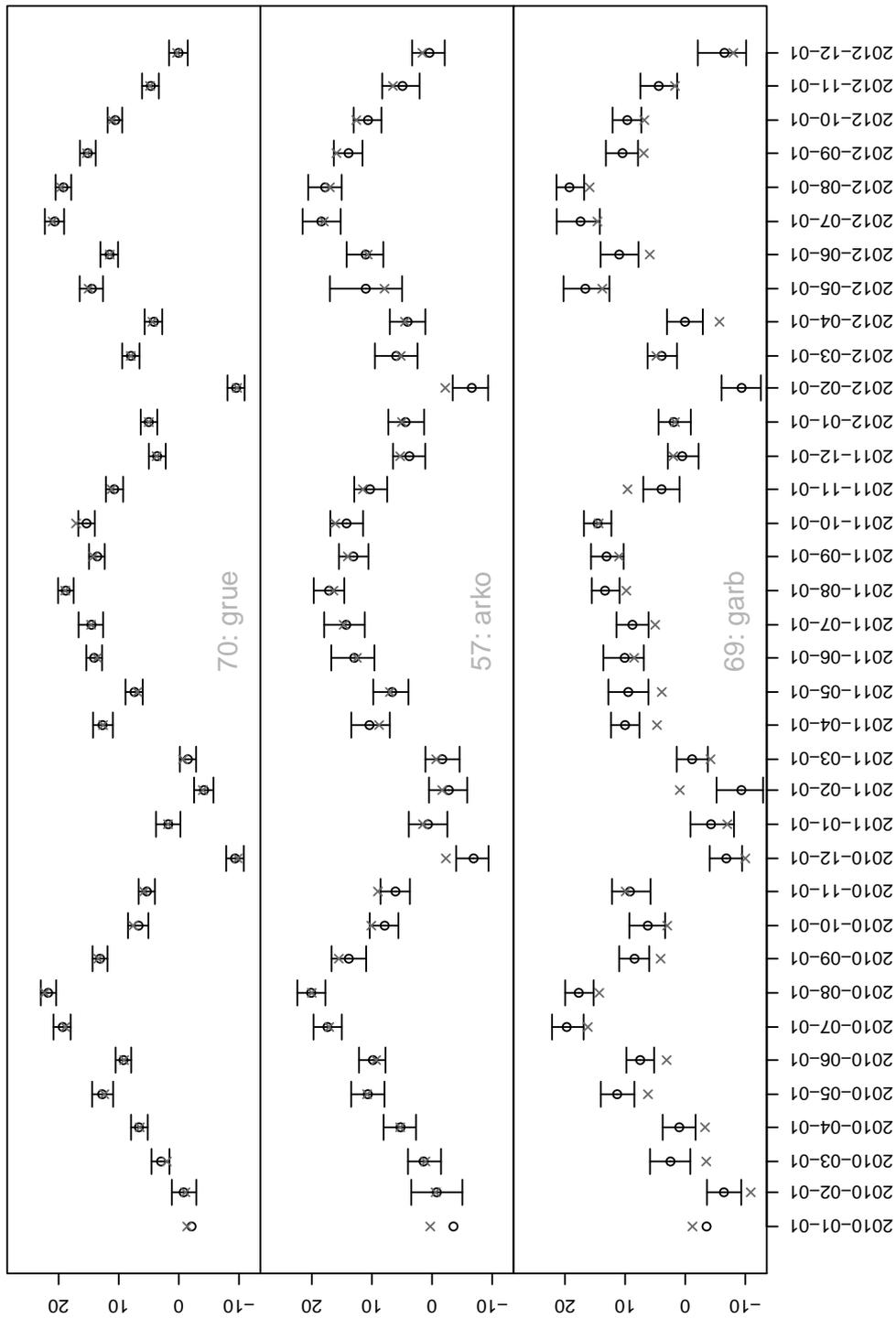
Figure 6.5.4: 95% prediction intervals, point predictions and observed mean temperatures for the first days of each months during the years 2010-2012 for the observation stations Grünow (70), Arkona (57) and Großer Arber (69), each calculated based on 1000 simulations from the predictive distribution of the mean temperatures at the respective observation station (spatial composite vine model).

Chapter 7

Spatial Gaussian model

In the subsequent chapter we are going to compare the predictions of the spatial R-vine model and the spatial composite vine model to each other and to the predictions from a classical semivariogram based model. Before we can conduct this comparison, we have to introduce the classical model, which will be called *spatial Gaussian model* (SG) from now on.

Let $\widetilde{Y}_t^s$ be a real valued random variable, which represents the (weighted) mean temperature at a location s and a point in time t. Let moreover $\widetilde{\boldsymbol{Y}}_t := (\widetilde{Y}_t^1, \ldots, \widetilde{Y}_t^d)^\top \in \mathbb{R}^d$ for all $t = 1, \ldots, N$. The random variables Y_t^s are weighted according to Section 3.5. Then the spatial Gaussian model for the mean temperature data set which we are going to take into consideration is of the form

$$\widetilde{\boldsymbol{Y}}_t = \boldsymbol{\mu}_t + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}_d \left(\mathbf{0}, \Sigma(\boldsymbol{\theta}^{\text{SG}}) \right), \quad t = 1, \ldots, N,$$

where $\boldsymbol{\mu}_t := (\mu_t^1, \ldots, \mu_t^d)^\top \in \mathbb{R}^d$ is a vector of means for all $t = 1, \ldots, N$ and $\Sigma(\boldsymbol{\theta}) \in \mathbb{R}^{d \times d}$ is a positive definite covariance matrix depending on some n_{par}^Σ-dimensional parameter vector $\boldsymbol{\theta}^{\text{SG}}$. The components of the mean vector $\boldsymbol{\mu}_t$ are modeled in analogy to the joint marginal model (3.6.2) as

$$\mu_t^s := g(t, s) = g \left(t, \widetilde{Y}_{t-1}^s, \widetilde{Y}_{t-2}^s, \widetilde{Y}_{t-3}^s, x_{\text{elev},s}, x_{\text{long},s}, x_{\text{lat},s}; \boldsymbol{\beta}_{\text{ls}} \right)$$

$$:= \beta_0(s) + \beta_{\sin}(s) \sin \left(\frac{2\pi t}{365.25} \right) + \beta_{\cos}(s) \cos \left(\frac{2\pi t}{365.25} \right)$$

$$+ \gamma_1(s) \widetilde{Y}_{t-1}^s + \gamma_2(s) \widetilde{Y}_{t-2}^s + \gamma_3(s) \widetilde{Y}_{t-3}^s,$$

i.e. including an intercept, a seasonality component and three auto regression terms, which are thought to capture all temporal dependencies of the variables. The spatial dependencies are modeled by means of the covariance matrix $\Sigma(\boldsymbol{\theta}^{\text{SG}}) = (\Sigma_{i,j}(\boldsymbol{\theta}^{\text{SG}}))_{i,j=1,\ldots,d}$ which in turn is modeled based on a *Gaussian variogram model* (see e.g. Gelfand, Diggle, Fuentes, and Guttorp, 2010, Chapter 3)

$$\gamma(h; \eta, \varsigma, \rho) := \varsigma \left(1 - \exp \left(-\frac{h^2}{\rho^2} \right) \right) + \eta \mathbb{1}_{(0,\infty)}(h),$$

where we call the parameters η, ς and ρ, *nugget effect*, *sill parameter* and *range parameter*, respectively. Then we model

$$\Sigma_{i,j}(\boldsymbol{\theta}^{\text{SG}}) := \sigma^2 - \gamma(d_{i,j}; \eta, \varsigma, \rho),$$

where the parameter vector $\boldsymbol{\theta}^{\mathrm{SG}}$ consists of the four components $\sigma, \eta, \varsigma, \rho$ and $d_{i,j}$ are the pairwise distances between the observation station pairs (i, j), $i, j = 1, \ldots, d$. By doing so we implicitly make a stationarity assumption, i.e. we assume $\mathrm{E}(\varepsilon_t^i) = \mathbf{0}$, $\mathrm{Var}(\varepsilon_t^i) = \sigma^2 < \infty$ and $\mathrm{Cov}(\varepsilon_t^i, \varepsilon_t^j) = \sigma^2 - \gamma(d_{i,j}; \eta, \varsigma, \rho)$ for all $i, j = 1, \ldots, d$ and $t = 1, \ldots, N$.

Eventually the model parameters are estimated in two steps. First the mean vectors $\boldsymbol{\mu}_t$, $t = 1, \ldots, N$, are estimated by means of least-squares estimation of the parameter vector $\boldsymbol{\beta}_{\mathrm{ls}}$. This is done in the same way as for the marginal model in Section 3.6 and we obtain the same estimates $\widehat{\boldsymbol{\beta}}_{\mathrm{ls}}$. Based on these estimates we calculate the residual vectors $\widehat{\boldsymbol{\varepsilon}}_t = \widetilde{\boldsymbol{y}}_t - \widehat{\boldsymbol{\mu}}_t$. In a second step we perform maximum likelihood estimation of the parameters $\boldsymbol{\theta}^{\mathrm{SG}} = (\sigma, \eta, \varsigma, \rho)^\top$. The log-likelihood to be maximized is given as

$$\ell_{\mathrm{SG}}\left(\boldsymbol{\theta}^{\mathrm{SG}} \,\middle|\, \widehat{\boldsymbol{\varepsilon}}_1, \ldots, \widehat{\boldsymbol{\varepsilon}}_N\right) = -\frac{N}{2} \ln\left((2\pi)^d \left|\Sigma(\boldsymbol{\theta}^{\mathrm{SG}})\right|\right) - \frac{1}{2} \sum_{t=1}^N \widehat{\boldsymbol{\varepsilon}}_t^\top \left[\Sigma(\boldsymbol{\theta}^{\mathrm{SG}})\right]^{-1} \widehat{\boldsymbol{\varepsilon}}_t.$$

For the purpose of prediction of the mean temperatures at a new location o we assume that the mean temperatures $\widetilde{Y}_t^o$, $t = 1, \ldots, N$ follow the model specified above. Thus we assume that

$$\begin{pmatrix} \widetilde{Y}_t^o \\ \widetilde{\boldsymbol{Y}}_t \end{pmatrix} = \begin{pmatrix} \mu_t^o \\ \boldsymbol{\mu}_t \end{pmatrix} + \begin{pmatrix} \varepsilon_t^o \\ \boldsymbol{\varepsilon}_t \end{pmatrix}, \quad \begin{pmatrix} \varepsilon_t^o \\ \boldsymbol{\varepsilon}_t \end{pmatrix} \overset{\mathrm{i.i.d.}}{\sim} \mathcal{N}_{d+1}\left(\mathbf{0}, \Sigma^*(\boldsymbol{\theta}^{\mathrm{SG}})\right), \quad t = 1, \ldots, N,$$

where

$$\Sigma^*(\boldsymbol{\theta}) = \left(\begin{array}{c|c} \sigma^2 & \boldsymbol{\sigma}_o^\top \\ \hline \boldsymbol{\sigma}_o & \Sigma(\boldsymbol{\theta}^{\mathrm{SG}}) \end{array}\right) \in \mathbb{R}^{d+1},$$

with $\boldsymbol{\sigma}_o = (\gamma(d_{1,o}; \eta, \varsigma, \rho), \ldots, \gamma(d_{d,o}; \eta, \varsigma, \rho))^\top \in \mathbb{R}^d$. Using some basic results on the conditional distribution of a mulitvariate normal distribution (see e.g. Eaton, 2007, Section 3.4), this yields that

$$\varepsilon_t^o | \boldsymbol{\varepsilon}_t = \boldsymbol{e} \overset{\mathrm{i.i.d.}}{\sim} \mathcal{N}\left(\overline{\mu}(\boldsymbol{\theta}^{\mathrm{SG}}), \overline{\Sigma}(\boldsymbol{\theta}^{\mathrm{SG}})\right), \quad t = 1, \ldots, N,$$

with $\overline{\mu}(\boldsymbol{\theta}^{\mathrm{SG}}) = \boldsymbol{\sigma}_o^\top \left[\Sigma(\boldsymbol{\theta}^{\mathrm{SG}})\right]^{-1} \boldsymbol{e}$ and $\overline{\Sigma}(\boldsymbol{\theta}^{\mathrm{SG}}) = \sigma^2 - \boldsymbol{\sigma}_o^\top \left[\Sigma(\boldsymbol{\theta}^{\mathrm{SG}})\right]^{-1} \boldsymbol{\sigma}_o$.

We perform the prediction in analogy to our vine copula based models by means of simulation. The first step is the repeated simulation of $\breve{\varepsilon}_t^o$ from $\mathcal{N}(\overline{\mu}(\widehat{\boldsymbol{\theta}}^{\mathrm{SG}}), \overline{\Sigma}(\widehat{\boldsymbol{\theta}}^{\mathrm{SG}}))$. Next we calculate $\breve{\mu}_t^o = g(t, \breve{\widetilde{y}}_{t-1}^o, \breve{\widetilde{y}}_{t-2}^o, \breve{\widetilde{y}}_{t-3}^o, x_{\mathrm{elev},o}, x_{\mathrm{long},o}, x_{\mathrm{lat},o}; \widehat{\boldsymbol{\beta}}_{\mathrm{ls}})$, imposing the same restrictions as in the back transformation of the copula data in Subsection 5.3.3 of Chapter 5. Moreover we calculate initial values for $\breve{\widetilde{y}}_t^o$, $t = 1, 2, 3$, in the same was as in Subsection 5.3.3, in order to be able to perform the previous calculations. Then the weighted mean temperatures can be calculated as $\breve{\widetilde{y}}_t^o = \breve{\mu}_t^o + \breve{\varepsilon}_t^o$. The unweighted mean temperatures $\breve{y}_t^o$ are again obtained by multiplication with the square roots of the respective weights $\widehat{w}_t$.
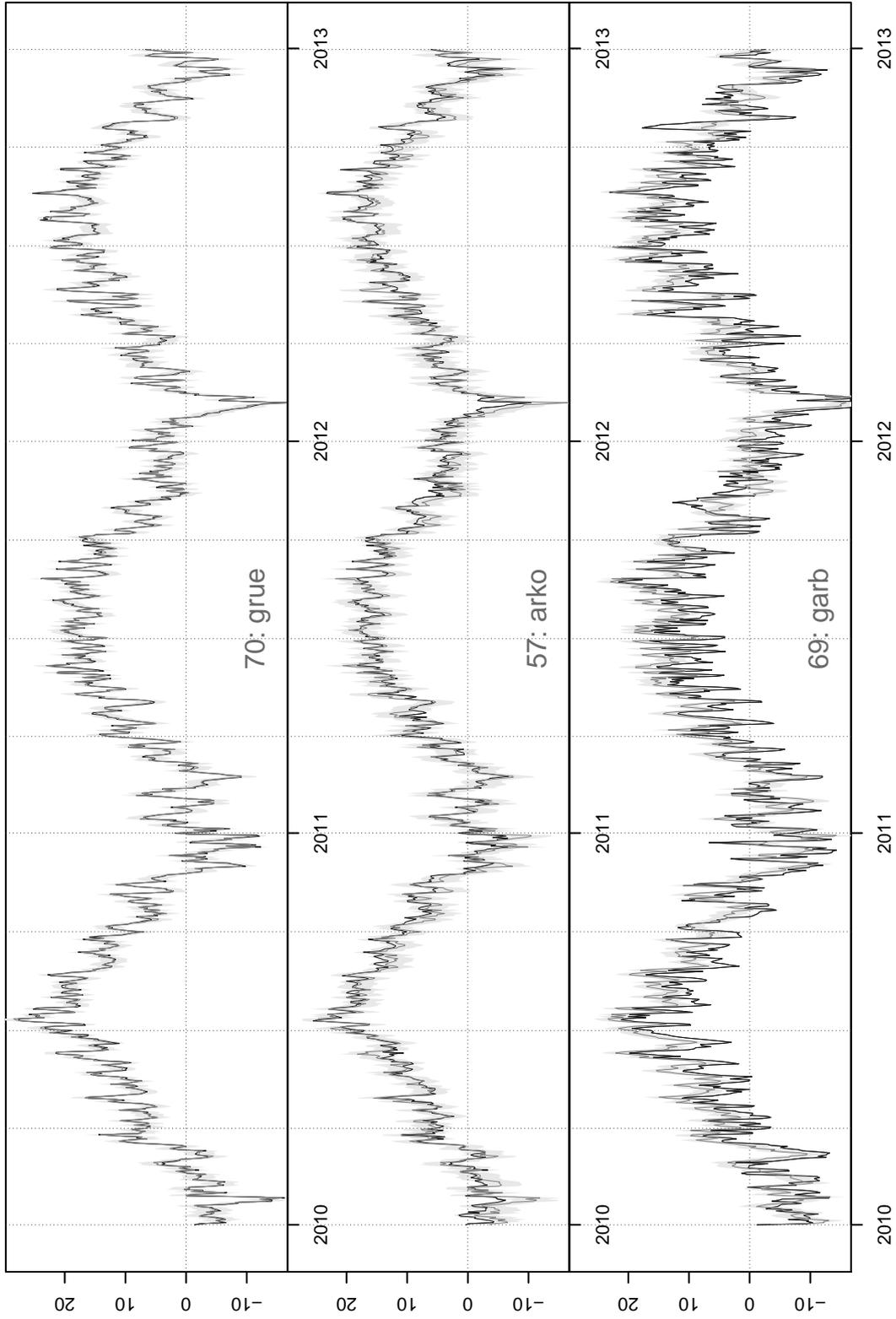
Figure 7.0.1: Prediction of the mean temperatures for the observation stations Grünow (70), Arkona (57) and Großer Arber (69) for the period 01/01/2010-12/31/2012 based on the spatial Gaussian model. black line: observed values. dark gray line: prediction. light gray area: 95% prediction intervals.
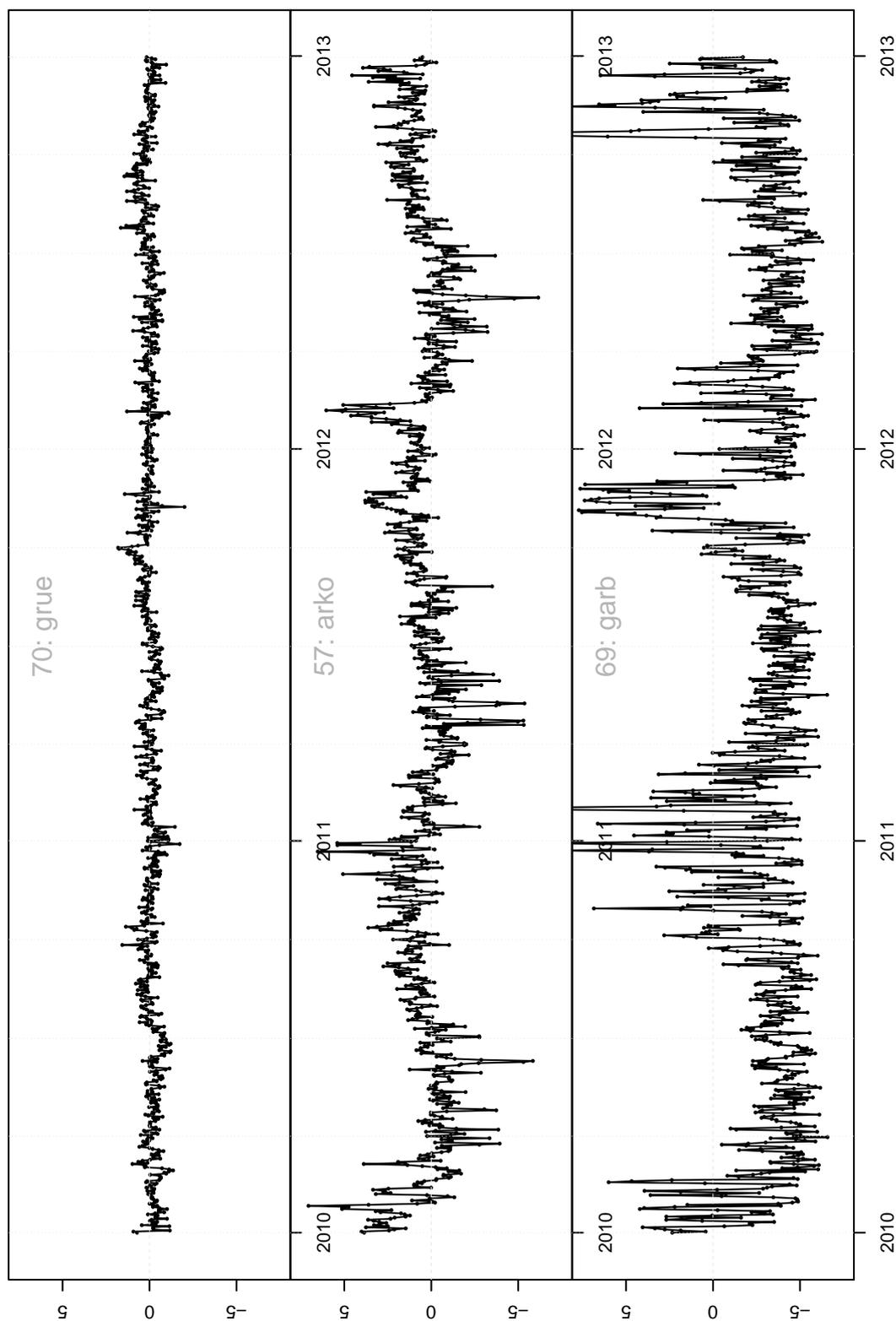
Figure 7.0.2: Prediction errors of the predictions for the observation stations Grünow (70), Arkona (57) and Großer Arber (69) for the period 01/01/2010-12/31/2012 based on the spatial Gaussian model.
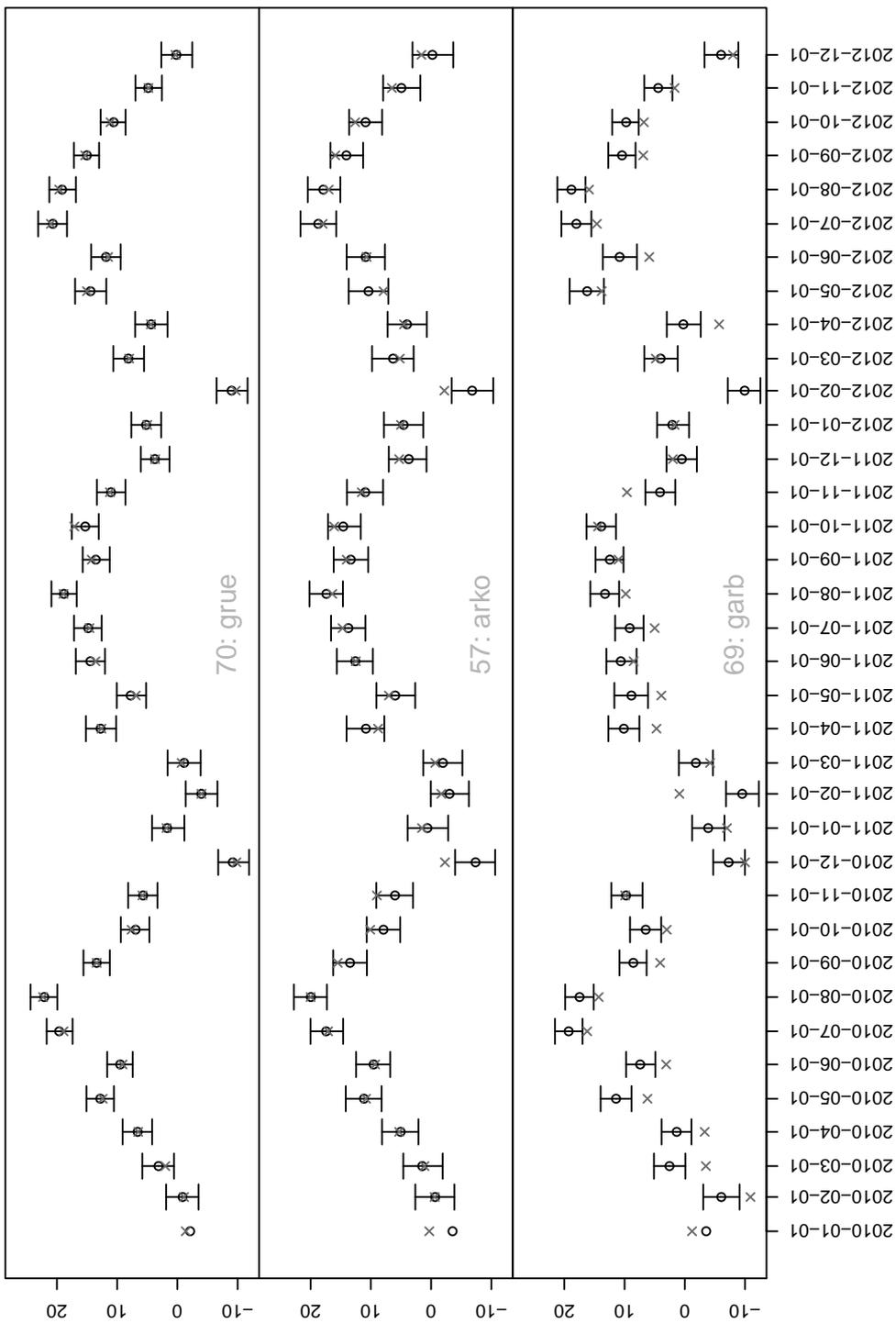
Figure 7.0.3: 95% prediction intervals, point predictions and observed mean temperatures for the first days of each months during the years 2010-2012 for the observation stations Grünow (70), Arkona (57) and Großer Arber (69), each calculated based on 1000 simulations from the predictive distribution of the mean temperatures at the respective observation station (spatial Gaussian model).

The results from the predictions at the 24 observation stations $s = 55, \ldots, 78$ based on the spatial Gaussian model fitted to the mean temperature training data set are illustrated in Figures A.3.1-A.3.12. These figures are quite similar compared to the respective figures for the spatial R-vine and the spatial composite vine model. It is very hard to detect the tiny differences in the predictions based on Figures A.3.1-A.3.4 and their respective figures of the previous chapters. Consideration of the differences depicted in the Figures A.3.5-A.3.8 eases the comparison a little bit. The most eye-catching observation from Figures A.3.9-A.3.12 is that the width of the prediction intervals seem to be more homogeneous across time and space, compared to the prediction intervals obtained for the spatial R-vine and the spatial composite vine model. The respective plots for the three selected stations Grünow (70), Arkona (57) and Großer Arber (69) are given by Figures 7.0.1-7.0.3.

In order to perform a first model comparison and to see that there are differences in the predictions, we compare the mean squared prediction errors of the three different models for all 24 validation stations, which are summarized in Table 7.1. We obtain the smallest mean squared errors for all models for the station *Grünow* (70), where the spatial R-vine model yields the lowest mean squared error. The station *Großer Arber* (69) yields the biggest mean squared errors by far. The predictions from the spatial R-vine model and the spatial composite vine model yield eight respectively six times the smallest mean squared errors. The spatial Gaussian model however yields ten times the best mean squared error. If we calculate overall mean squared errors for all three models over all 24 stations, we obtain $\text{MSE}_{\text{SV}} = 1.800$, $\text{MSE}_{\text{SCVM}} = 1.798$ and $\text{MSE}_{\text{SG}} = 1.788$, which lie pretty close together.

s	short name	MSE(s) SV	MSE(s) SCVM	MSE(s) SG	s	short name	MSE(s) SV	MSE(s) SCVM	MSE(s) SG
61	bork	0.57	*0.55*	0.78	63	buch	1.08	1.21	*1.03*
62	bvoe	0.36	0.38	*0.28*	65	ebra	1.01	0.92	*0.88*
72	luec	*0.24*	0.30	0.25	60	blan	0.80	0.66	*0.35*
68	gram	*0.19*	0.22	0.29	66	ellw	0.96	0.90	*0.78*
64	cosc	*0.85*	0.92	1.08	59	augs	0.68	*0.47*	0.75
57	arko	*1.87*	2.32	2.51	67	falk	0.43	0.41	*0.40*
77	rahd	0.28	*0.25*	0.26	76	ohrz	0.92	0.72	*0.61*
70	grue	*0.17*	0.19	0.21	71	hohe	3.65	*3.42*	3.60
56	alfe	0.64	*0.23*	0.27	55	albs	2.06	*2.04*	2.08
58	arns	*0.94*	0.98	1.30	75	neuh	6.58	6.45	*6.13*
78	wies	0.75	0.99	*0.67*	73	mitt	*2.85*	2.90	2.99
74	muel	*0.63*	0.68	0.79	69	garb	14.69	15.07	*14.65*

Table 7.1: Comparison of the mean squared errors of the mean temperature predictions over the period 01/01/2010-12/31/2012 for the observation stations of the validation data set based on the spatial R-vine model, the spatial composite vine model and the spatial Gaussian model.

The above comparison by means of mean squared errors would lead to the conclusion that the spatial Gaussian model barely yields the best prediction results, however this

method doesn't take the whole predictive distributions into consideration, instead it measures the goodness of the predictions only based on the respective point estimates. Thus we are going to conduct further comparisons in the following chapter, which are based on proper scoring rules.

Chapter 8

Model validation and comparison

In Chapter 5 and Chapter 6 two different kinds of new models for spatial data, the spatial R-vine model (SV) and the spatial composite vine model (SCVM), were introduced. On the one hand this chapter aims to compare these models to corresponding models which do not account for spatial information, on the other hand these models are going to be compared to the spatial Gaussian model, which was presented in Chapter 7.

8.1   Comparison with non-spatial models

We start with the comparison of the spatial R-vine model to a full R-vine model and a truncated R-vine model where the truncation level equals the truncation level ten of the spatial R-vine model. We compare these models based on their maximum log-likelihood, AIC, BIC, number of parameters and computation time, where the computations were performed on a 2.6 GHz AMD Opteron processor. The respective numbers are given in the upper part of Table 9.3 in Chapter 9. Compared to the full R-vine model the likelihood of the spatial R-vine model loses 8064.08 points. This number reduces to 3716.59 if we compare the spatial R-vine model to the truncated R-vine model. However the number of parameters is reduced immensely. Compared to the full R-vine model and the truncated R-vine model the spatial R-vine model needs only 2.3% respectively 5.6% of the numbers of parameters. This results in a considerably reduction of computation time. Whereas the maximum likelihood estimation of the full R-vine model takes more than 50 days, this time is reduced to approximately 3.7 days and only 18 hours for the truncated and the spatial R-vine respectively. Even though the full R-vine model would be ranked highest according to AIC and BIC, the immense differences in computation time suggest that the application of a spatial R-vine model should be favored in practice.

Moreover, Table 9.3 compares our spatial composite vine model with its non-spatial complement, the composite vine model. Whereas the maximum composite-likelihood estimation for the ordinary composite vine model involves the estimation of 495 parameters and 6.1 days are needed for the computation, the respective calculations for the spatial version of the model involve only 16 parameters and take 1.6 hours of computation time. The maximum composite log-likelihoods differ about 1167.61 points. According to the BIC the spatial composite vine model is preferred over the composite vine model.

8.2 Spatial model comparison

Before we are going to compare our three spatial models based on the scoring techniques presented in Section 2.8, we compare the **densities of the predictive distributions** of the copula data for the spatial R-vine model and the spatial composite vine model. Kernel density estimates of these densities, for the first of February, April, June, August, October and December 2010-2012 at the observation station *Arkona* (57), which are calculated based on 1000 simulations from the particular predictive distribution, are illustrated in Figure 8.2.1. Moreover the respective medians and 95% prediction intervals are indicated. We do not observe any kind of temporal dependence, which is in agreement with the modeling assumption that the marginal model should capture all temporal dependencies. On the one hand there are some days like the first of August 2012, where both densities are nearly identical, contrariwise there are many days like the first of February 2012, where the densities differ considerably, however we always observe some kind of similarity in terms of shape, scale and skew.
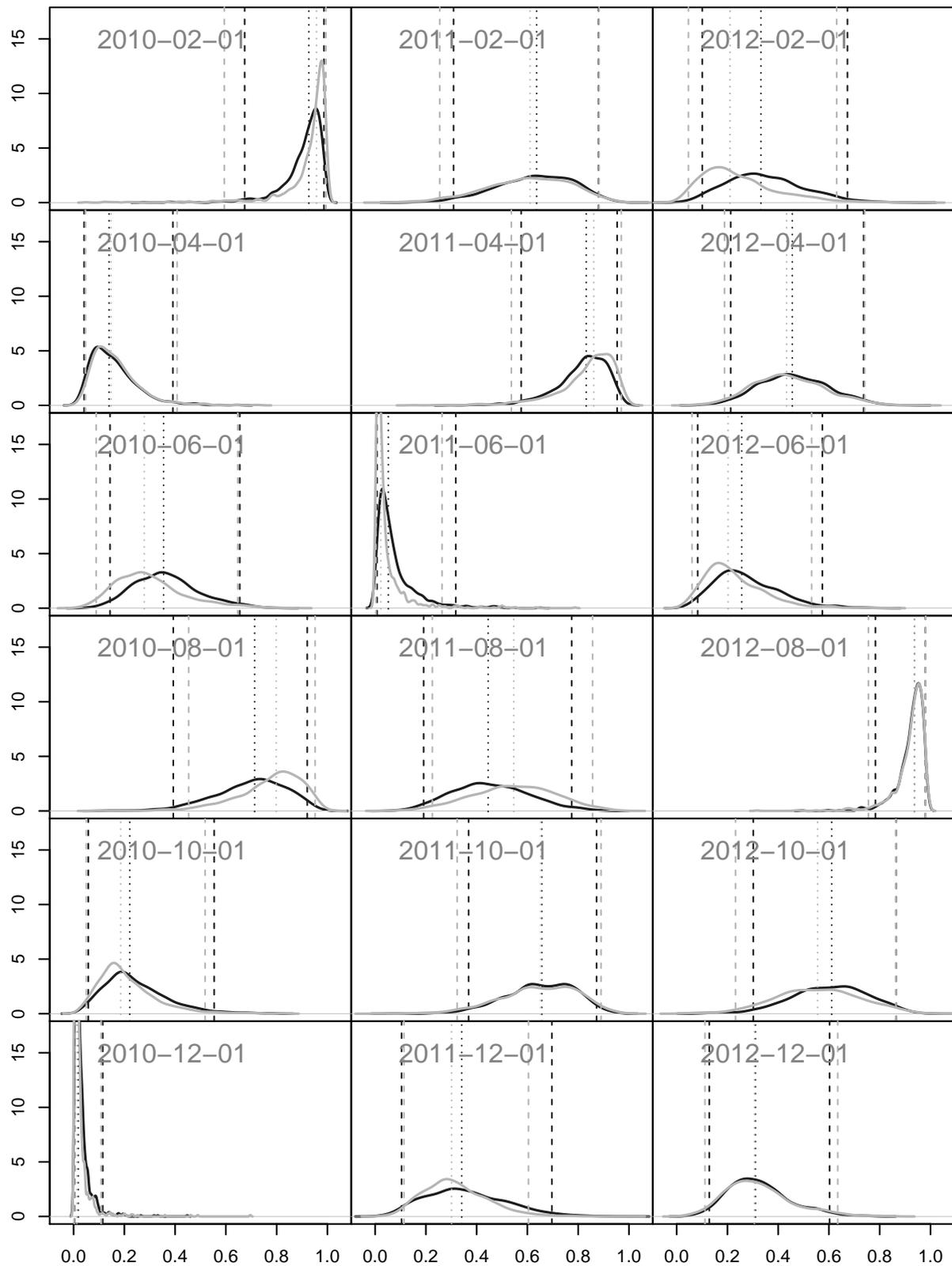
Figure 8.2.1: Comparison of kernel density estimates, each calculated based on 1000 simulations from the predictive distributions arising from the spatial R-vine model (black) and the spatial composite vine model (gray), for the first of February, April, June, August, October and December 2010-2012 at the observation station Arkona (57). The respective medians and 95% prediction intervals are indicated as vertical lines in the corresponding color.

8.2.1 Score based model comparison

For the purpose of further comparison we calculate **continuous ranked probability scores** and **interval scores for** $\alpha = 0.05$. These techniques will allow for an adequate comparative model validation. We consider

- averaged scores (Table 8.1),

- percentaged model outperformance (Figures 8.2.2 and 8.2.3) and

- a new concept called log-score difference plots (Figures 8.2.4-8.2.9, 8.2.10)

in terms of both, continuous ranked probability scores and interval scores. In all figures and tables where we compare results for the different stations of the validation data set, we order the stations according to their elevation, as we did it in the previous chapters.

In order to get a first impression which models provide the best performance in terms of **averaged scores**, we compare the averaged continuous ranked probability scores (CRPS) and interval scores ($IS_{0.05}$) in Table 8.1, where we average over time. Moreover the overall averages are given in the last row of the table. For the purpose of comparison we recall that scores close to zero are preferred, i.e. the overall consideration of the averaged scores in Table 8.1 yields, that we prefer the spatial R-vine model according to the continuous ranked probability scores and the spatial composite vine model according to the interval scoring method. We observe furthermore, that the performance of the different models seems to depend on the elevation of the respective observation stations. The results of Table 8.1 can be summarized as

$$SV \succ SCVM \succ SG \quad \text{in terms of } \overline{CRPS},$$
$$SCVM \succ SG \succ SV \quad \text{in terms of } \overline{IS}_{0.05},$$

where $\succ$ is defined as 'is favored over'.

s	short name	$\overline{\text{CRPS}}$			$\overline{\text{IS}}_{0.05}$		
		SV	SCVM	SG	SV	SCVM	SG
61	bork	*-2.318*	-2.420	-3.321	-4.569	*-4.455*	-6.331
62	bvoe	-2.324	*-2.113*	-2.620	-4.349	*-3.808*	-4.969
72	luec	*-2.212*	-2.454	-2.617	*-3.976*	-4.066	-4.859
68	gram	*-1.840*	-1.904	-2.577	-3.433	*-3.310*	-4.834
64	cosc	*-2.652*	-2.948	-2.874	*-5.274*	-5.532	-5.984
57	arko	*-3.112*	-3.289	-3.383	*-7.827*	-7.987	-8.849
77	rahd	*-2.360*	-2.576	-2.655	*-4.269*	-4.382	-5.034
70	grue	-1.977	*-1.917*	-2.641	-3.593	*-3.271*	-4.908
56	alfe	-3.197	*-2.281*	-2.604	-5.710	*-3.861*	-4.820
58	arns	*-2.246*	-2.474	-2.616	*-4.930*	-5.326	-5.670
78	wies	-2.722	-2.806	*-2.627*	-5.319	-5.663	*-5.242*
74	muel	*-2.220*	-2.434	-2.997	*-4.599*	-4.826	-5.644
63	buch	-2.540	*-2.487*	-2.626	-5.069	-5.036	*-5.016*
65	ebra	*-2.333*	-2.489	-2.606	-5.301	*-4.844*	-5.093
60	blan	-2.988	*-2.565*	-2.623	-5.524	-5.060	*-4.883*
66	ellw	-3.221	-3.204	*-2.613*	-5.819	-5.449	*-4.905*
59	augs	-2.732	-2.662	*-2.572*	-5.135	*-4.502*	-4.794
67	falk	-2.636	-2.678	*-2.614*	-4.721	*-4.548*	-4.818
76	ohrz	-3.657	-3.023	*-2.611*	-6.640	-5.440	*-5.404*
71	hohe	-2.654	-3.436	*-2.562*	-12.573	*-8.854*	-11.258
55	albs	-2.995	-3.094	*-2.634*	-6.577	-6.429	*-6.060*
75	neuh	*-2.164*	-2.678	-2.552	-34.449	-27.331	*-21.725*
73	mitt	-3.487	-3.765	*-3.006*	*-9.154*	-10.223	-11.604
69	garb	-3.209	-3.520	*-2.972*	-50.834	*-47.922*	-54.172
mean		**-2.658**	-2.717	-2.730	-8.735	**-8.005**	-8.620

Table 8.1: Comparison of the continuous ranked probability scores (CRPS) and the interval scores ($\text{IS}_{0.05}$) of the spatial R-vine model (SV), the spatial composite vine model (SCVM) and the spatial Gaussian model (SG) averaged over the period $01/01/2010 - 12/31/2012$ for the observation stations of the validation data set. In the last row of the table the overall averages are given.

The following two figures, Figures 8.2.2 and 8.2.3, compare our three spatial models pairwise by means of continuous ranked probability scores and interval scores, respectively, and illustrate the **percentaged outperformance**[1] of one model compared to the other for all 24 observation stations composing the validation data set. As before, we detect a dependence of the model performance on the elevation. From Figure 8.2.2 we observe 15 outperformances of the spatial R-vine model over the spatial Gaussian model, 16 outperformances of the spatial composite vine model over the spatial Gaussian model and 7 outperformances of the spatial composite vine model over the spatial R-vine model in terms of the continuous ranked probability scores. Figure 8.2.3 compares the outperformance in terms of the interval scores. This time we gain 17 victories of the spatial R-vine model over the spatial Gaussian model, 18 outperformances of the spatial composite vine model over the spatial Gaussian model and the spatial composite vine model yields better results than the spatial R-vine model for 15 stations. We summarize the previous findings as

$$SV \succ SCVM \succ SG \quad \text{in terms of \%CRPS,}$$
$$SCVM \succ SV \succ SG \quad \text{in terms of \%IS,}$$

where $\succ$ is defined as 'is favored over'.

---

[1]For all stations it is counted for how many points in time one model yields a higher score than the other model.
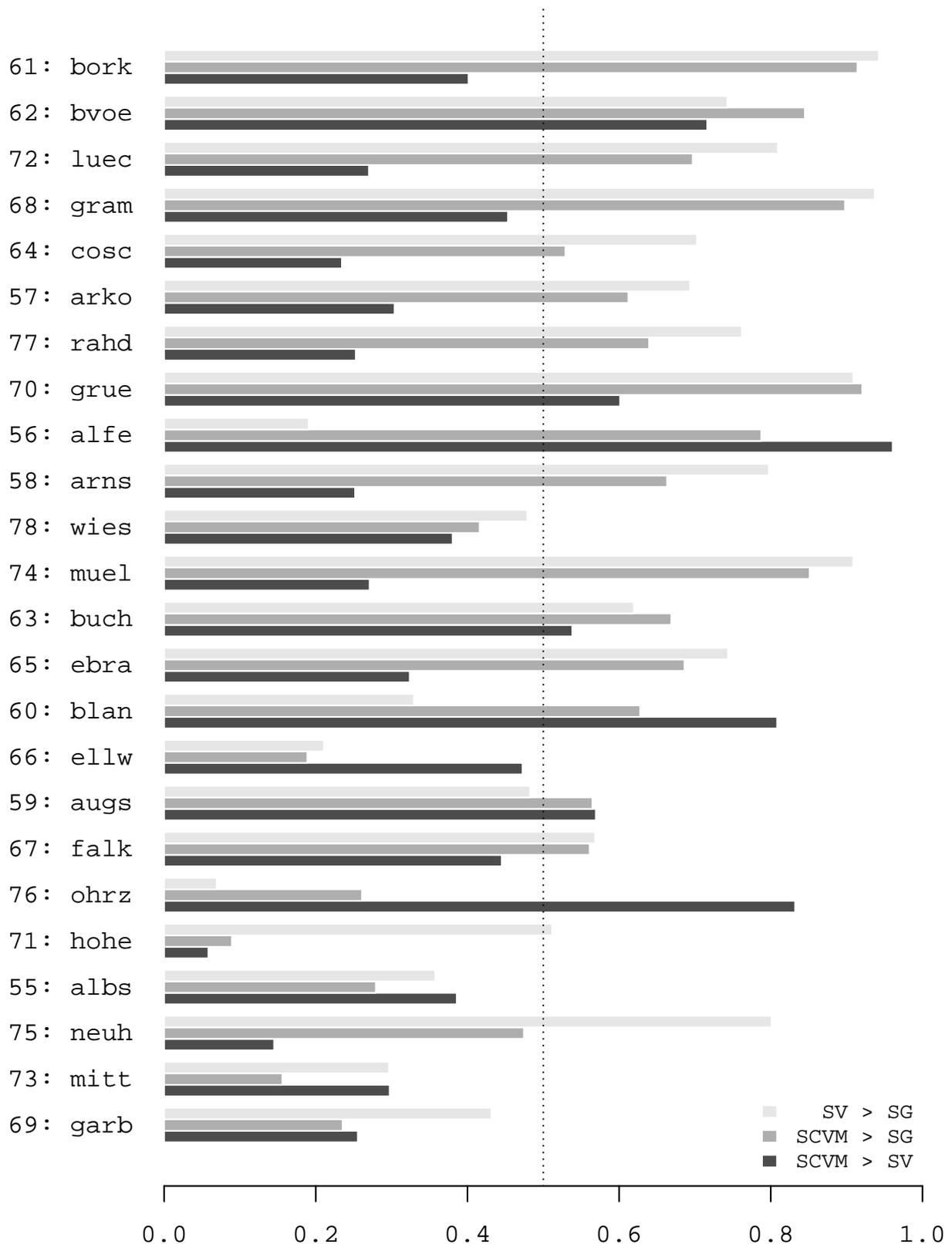
Figure 8.2.2: Percentaged outperformance of the spatial R-vine model over the spatial Gaussian model, the spatial composite vine model over the spatial Gaussian model and the spatial composite vine model over the spatial R-vine model in terms of continuous ranked probability score for all 24 observation stations of the validation data set.
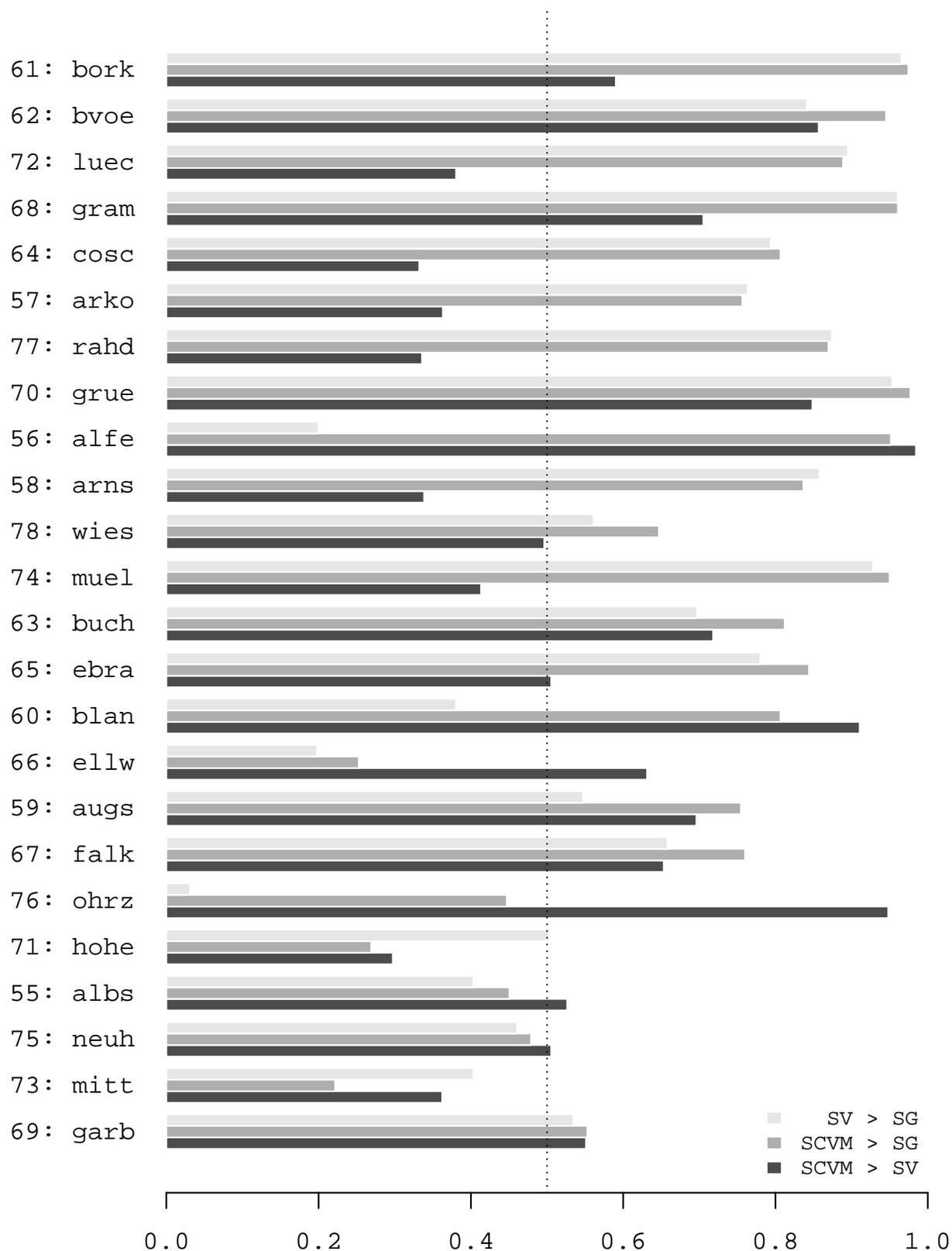
Figure 8.2.3: Percentaged outperformance of the spatial R-vine model over the spatial Gaussian model, the spatial composite vine model over the spatial Gaussian model and the spatial composite vine model over the spatial R-vine model in terms of interval score for all 24 observation stations of the validation data set.

It could be possible that the model outperformance depends on the time, i.e. there may be time intervals in which one model yields better results than the others. In order to be able to detect such kinds of time dependencies, we consider the Figures 8.2.4-8.2.9, as well as Figure 8.2.10. We call the plots given in these figures **log-score difference plots**, since they depict the difference of the logarithmized negatively oriented[2] scores of two models against the respective points in time.

Let us first consider the **log-score difference plots for the continuous ranked probability scores**. Figure 8.2.4 compares the spatial R-vine model and the spatial Gaussian model. With increasing elevation the domination of the black x signs which indicate the outperformance of the spatial R-vine model decreases and we observe predominance of the grey plus signs for many of the stations with high elevations. For example for the station *Oberharz am Brocken-Stiege* (76) a clear outperformance of the spatial Gaussian model is observed. An overall consideration of the 24 plots in Figure 8.2.4 yields, that there are short time intervals (in the winters) where the spatial Gaussian model yields higher scores, whereas the spatial R-vine model tends to perform better for some bigger time intervals. The log-score difference plots in Figure 8.2.5 which compare the spatial composite vine model and the spatial Gaussian model yield similar results as Figure 8.2.4, however we observe differences in the outperformance. The comparison of the spatial composite vine model and the spatial R-vine model in Figure 8.2.6 yields distinct outperformance of the spatial composite vine model for stations like *Bremervörde* (62), *Alfeld* (56), *Blankenrath* (60) and *Oberharz am Brocken-Stiege* (76). Moreover, the spatial R-vine model scores tend to be higher for stations like *Hoherodskopf/Vogelsberg* (71) and *Neuhaus am Rennweg* (75).

Now we take a look at Figures 8.2.7-8.2.9, the respective **log-score difference plots for the interval scores**. We find that the scores for stations like *Hoherodskopf/Vogelsberg* (71), *Neuhaus am Rennweg* (75) and *Großer Arber* (69), which exhibit high elevations, tend to differ more distinctly. For most of the other stations there are always relatively few outliers, but for the big share of the points in time the differences are rather moderate. Nonetheless it is possible to judge the station-wise outperformance in terms of interval scores.

Figure 8.2.10 depicts the **averaged log-score difference plots**, which average the results from Figures 8.2.4, 8.2.5, 8.2.7 and 8.2.8. From the plots for the continuous ranked probability scores we see, that there are short time intervals around the turns of the years, where the spatial Gaussian model consequently yields higher scores than the other two models. Moreover we observe that the differences between the spatial R-vine model and the spatial Gaussian model seem to be higher than the respective differences of the spatial composite vine model and the spatial Gaussian model. The comparison of the spatial R-vine model and the spatial Gaussian model by means of the interval scores yields a balanced picture, whereas the comparison of the spatial composite vine model and the spatial Gaussian model favors the spatial composite vine model.

---

[2]Negatively oriented means that the scores are multiplied by −1 and are therefore positive.
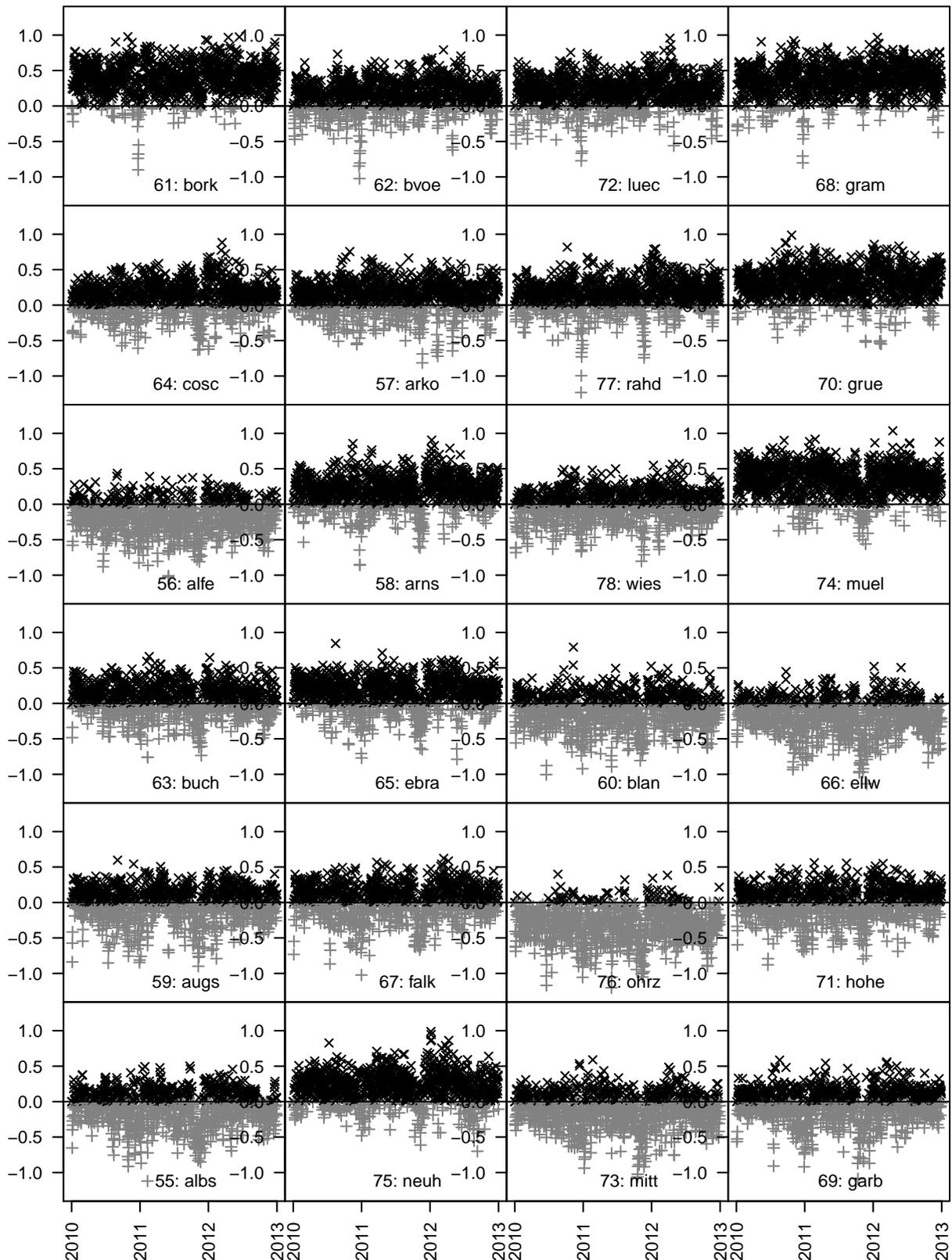
Figure 8.2.4: Station-wise log-score difference plots of the continuous ranked probability scores comparing the spatial R-vine model and the spatial Gaussian model, for all 24 observation stations of the validation data set. Points in time where the spatial R-vine model has the higher score are marked by a black x. In contrast, points in time where the spatial Gaussian model has the higher score are marked by a gray plus sign.
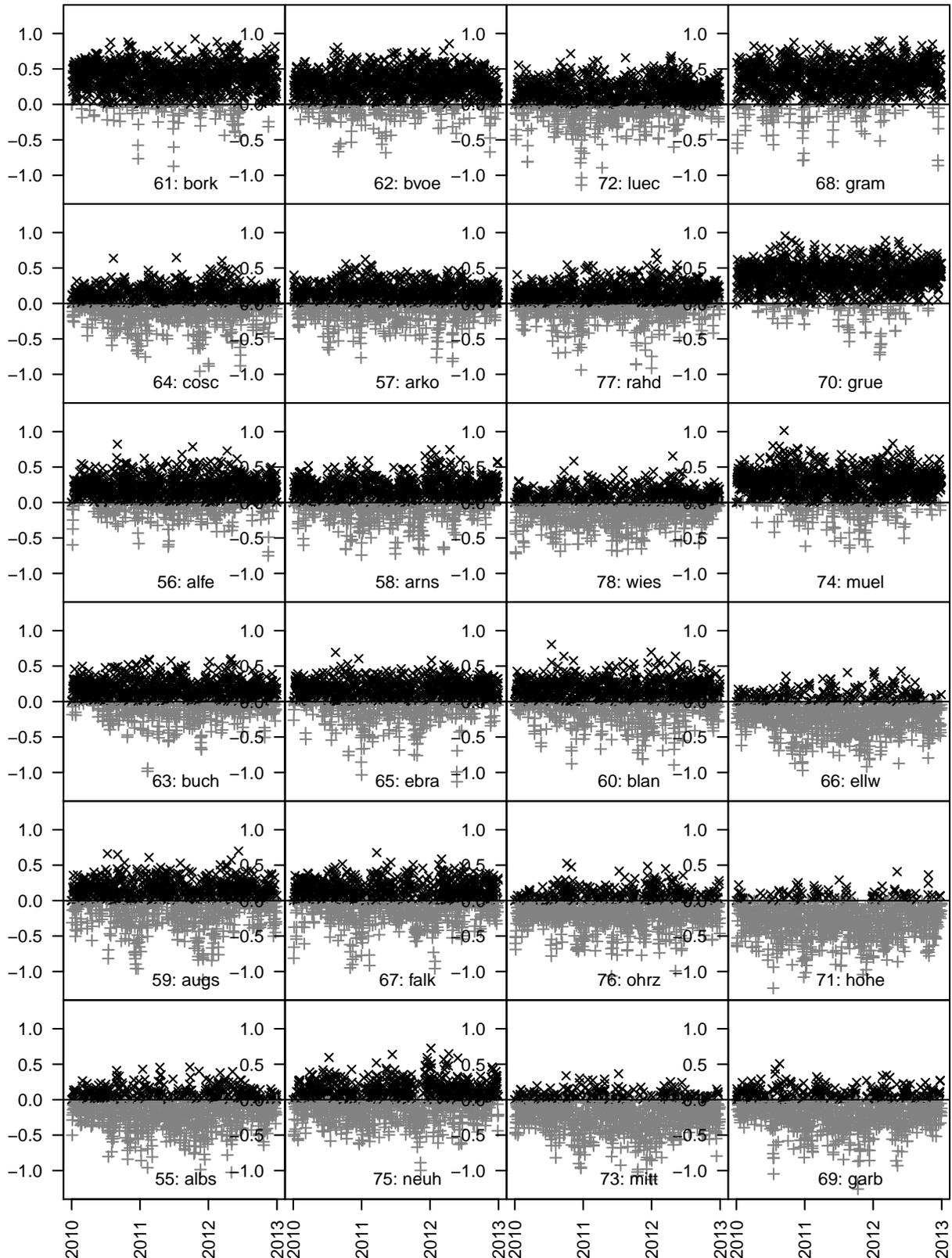
Figure 8.2.5: Station-wise log-score difference plots of the continuous ranked probability scores comparing the spatial composite vine model and the spatial Gaussian model, for all 24 observation stations of the validation data set. Points in time where the spatial composite vine model has the higher score are marked by a black x. In contrast, points in time where the spatial Gaussian model has the higher score are marked by a gray plus sign.
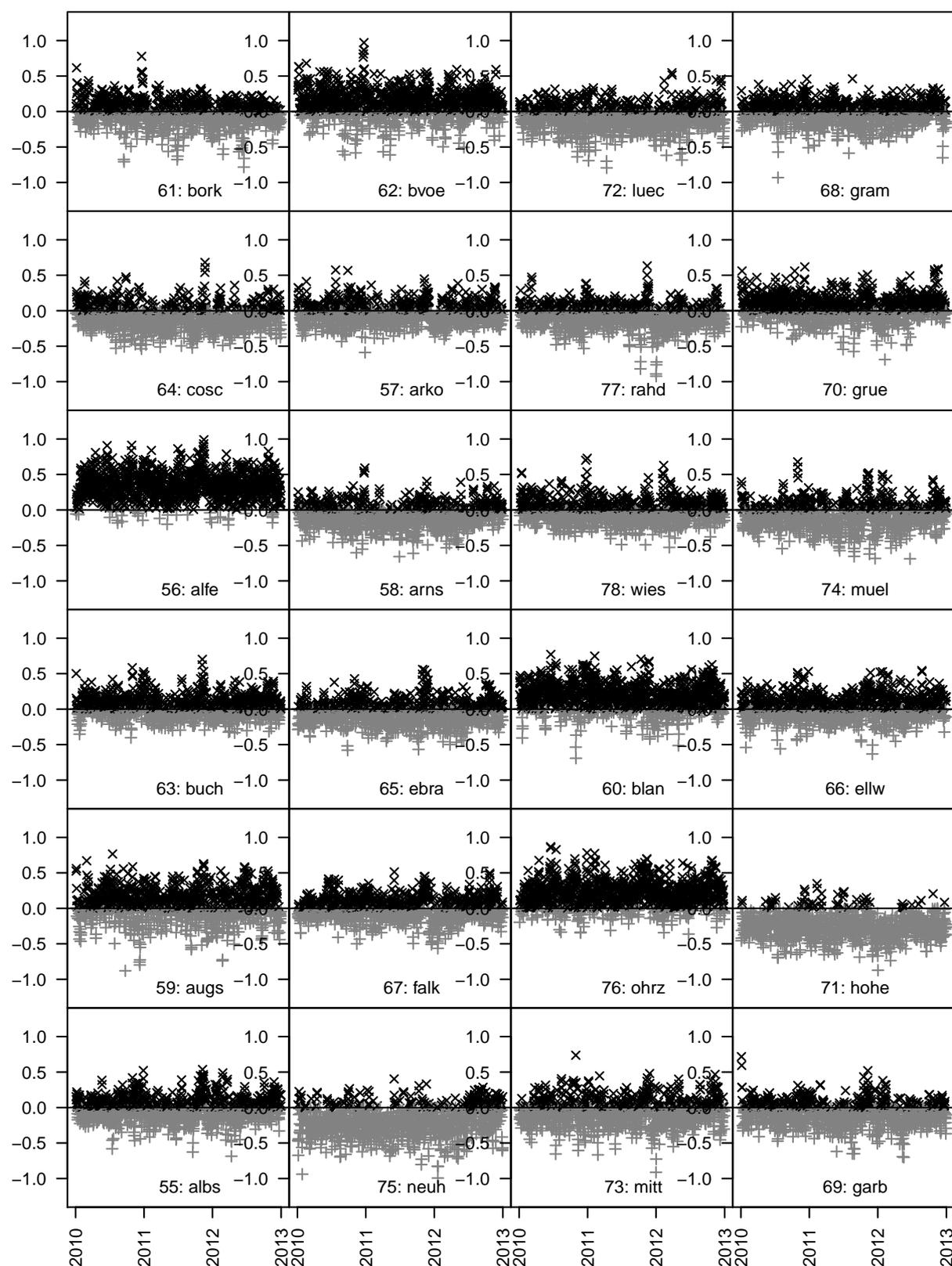
Figure 8.2.6: Station-wise log-score difference plots of the continuous ranked probability scores comparing the spatial composite vine model and the spatial R-vine model, for all 24 observation stations of the validation data set. Points in time where the spatial composite vine model has the higher score are marked by a black x. In contrast, points in time where the spatial R-vine model has the higher score are marked by a gray plus sign.
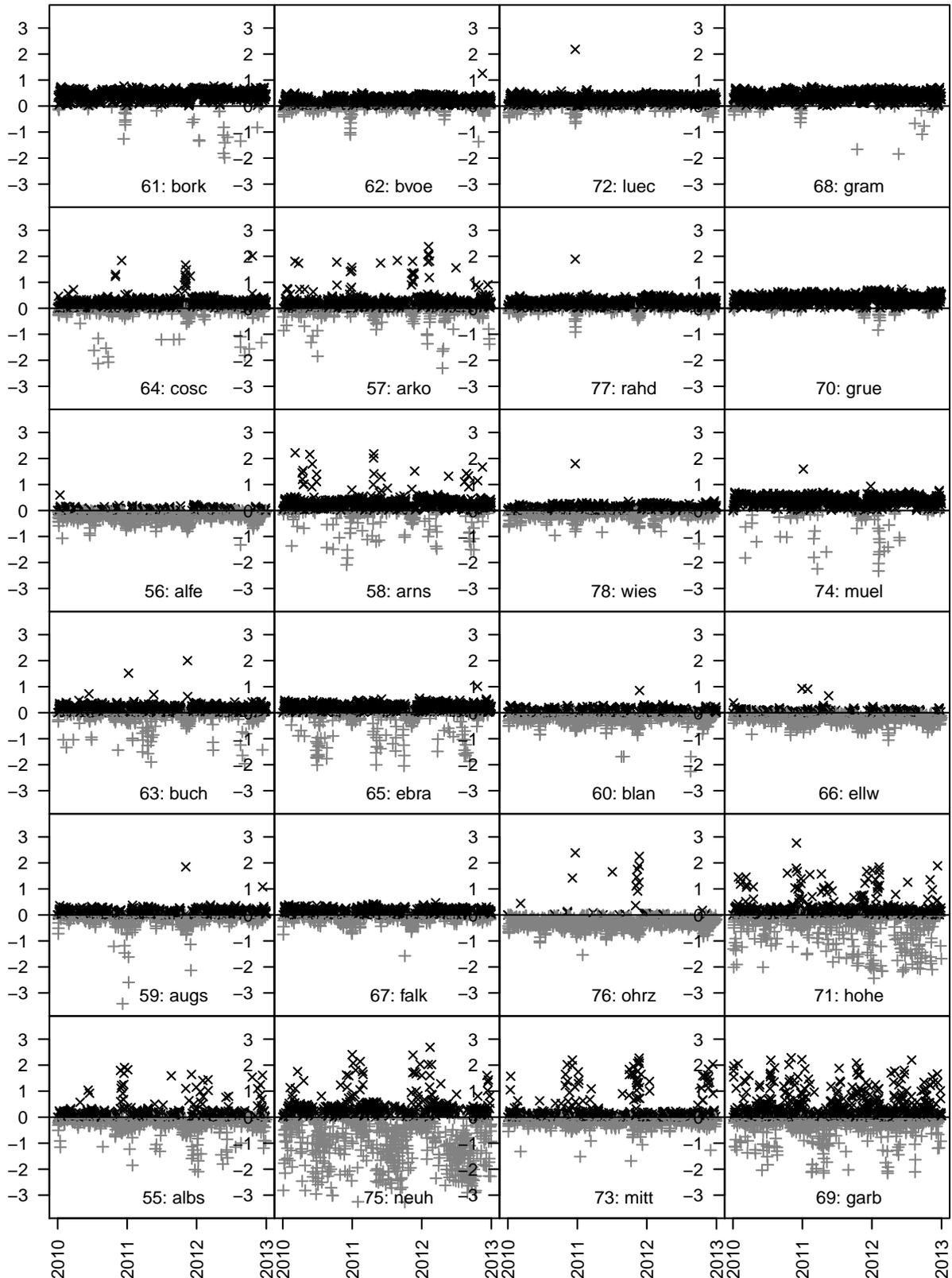
Figure 8.2.7: Station-wise log-score difference plots of the interval scores comparing the spatial R-vine model and the spatial Gaussian model, for all 24 observation stations of the validation data set. Points in time where the spatial R-vine model has the higher score are marked by a black x. In contrast, points in time where the spatial Gaussian model has the higher score are marked by a gray plus sign.
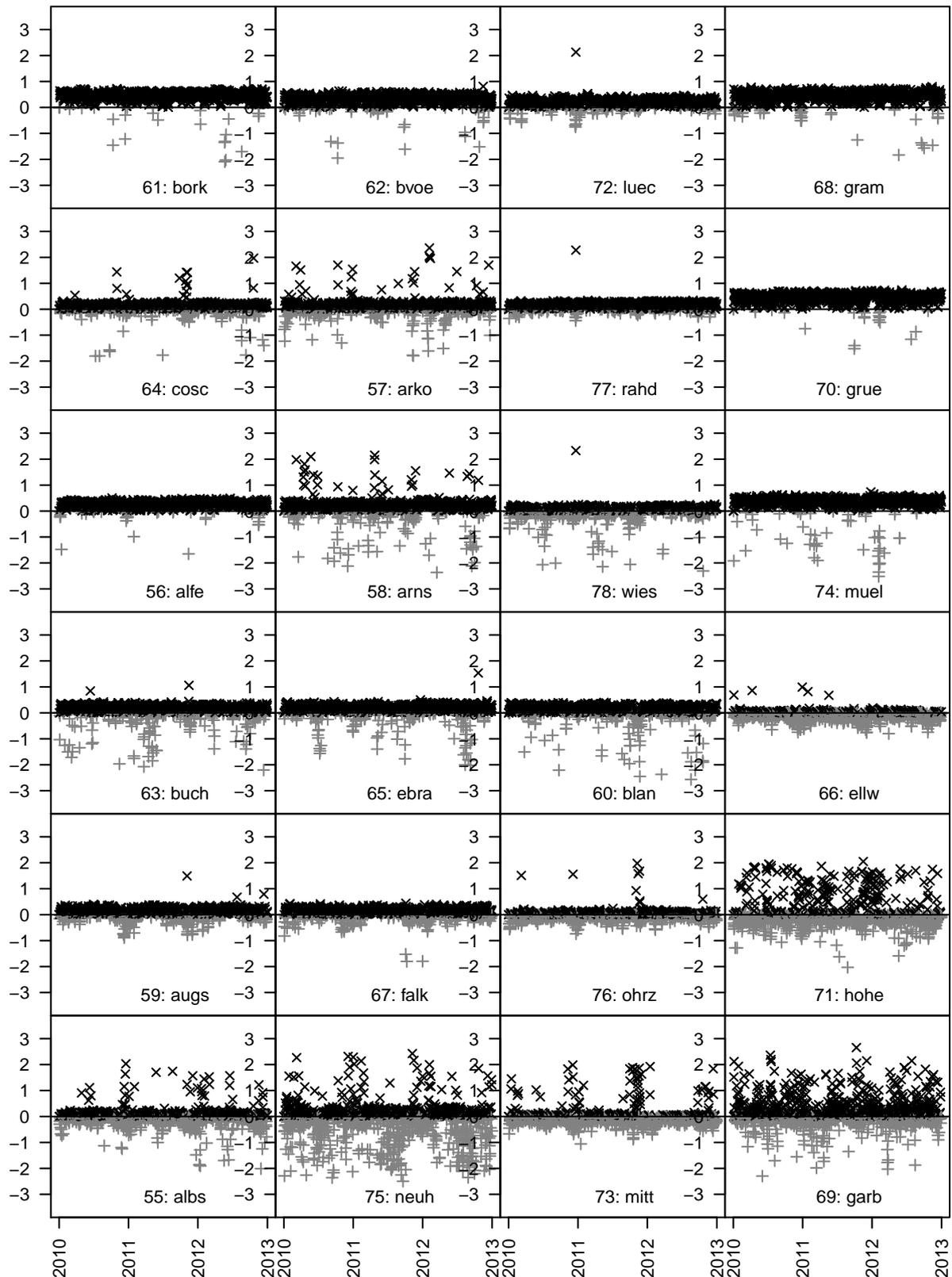
Figure 8.2.8: Station-wise log-score difference plots of the interval scores comparing the spatial composite vine model and the spatial Gaussian model, for all 24 observation stations of the validation data set. Points in time where the spatial composite vine model has the higher score are marked by a black x. In contrast, points in time where the spatial Gaussian model has the higher score are marked by a gray plus sign.
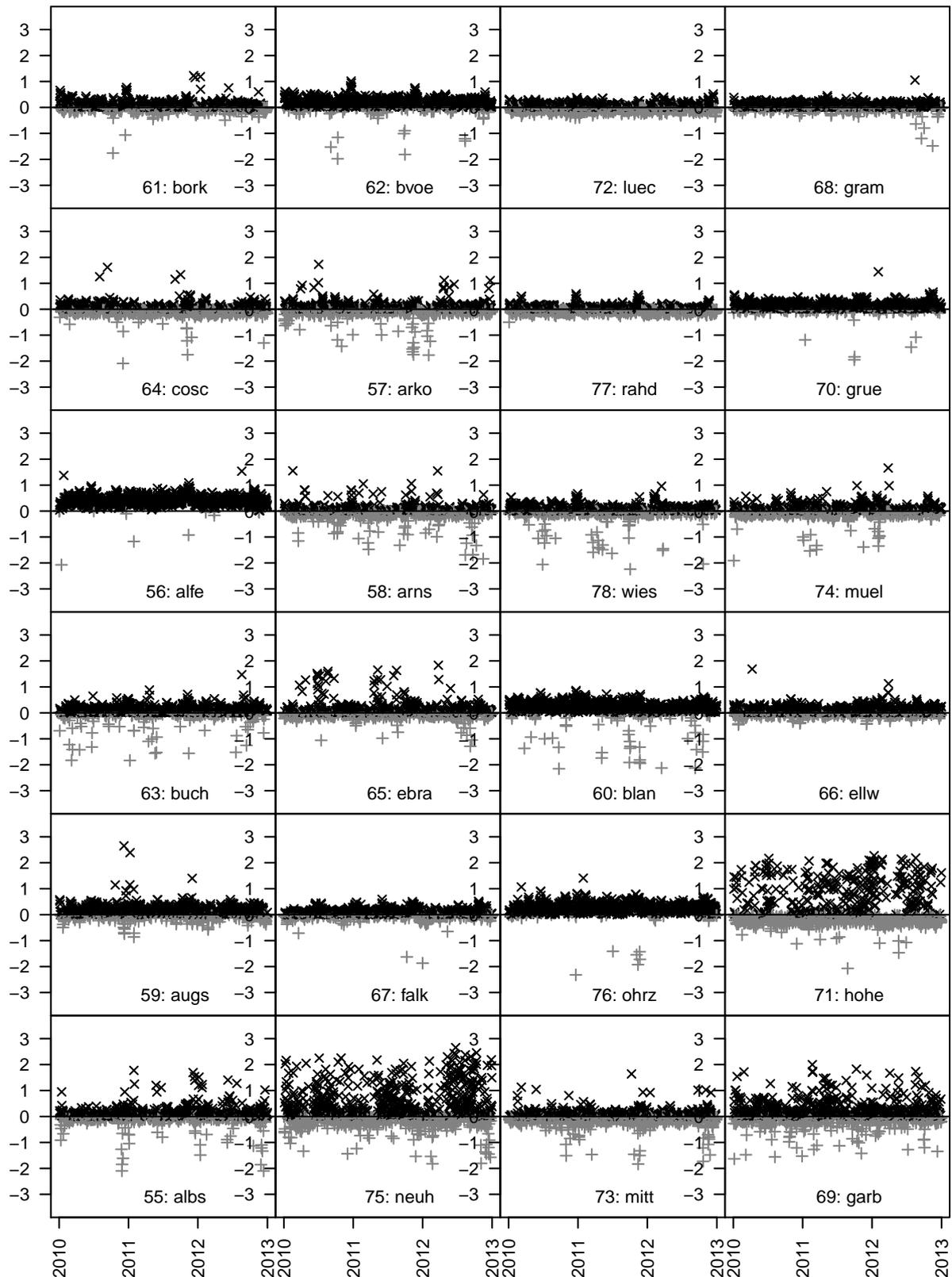
Figure 8.2.9: Station-wise log-score difference plots of the interval scores comparing the spatial composite vine model and the spatial R-vine model, for all 24 observation stations of the validation data set. Points in time where the spatial composite vine model has the higher score are marked by a black x. In contrast, points in time where the spatial R-vine model has the higher score are marked by a gray plus sign.
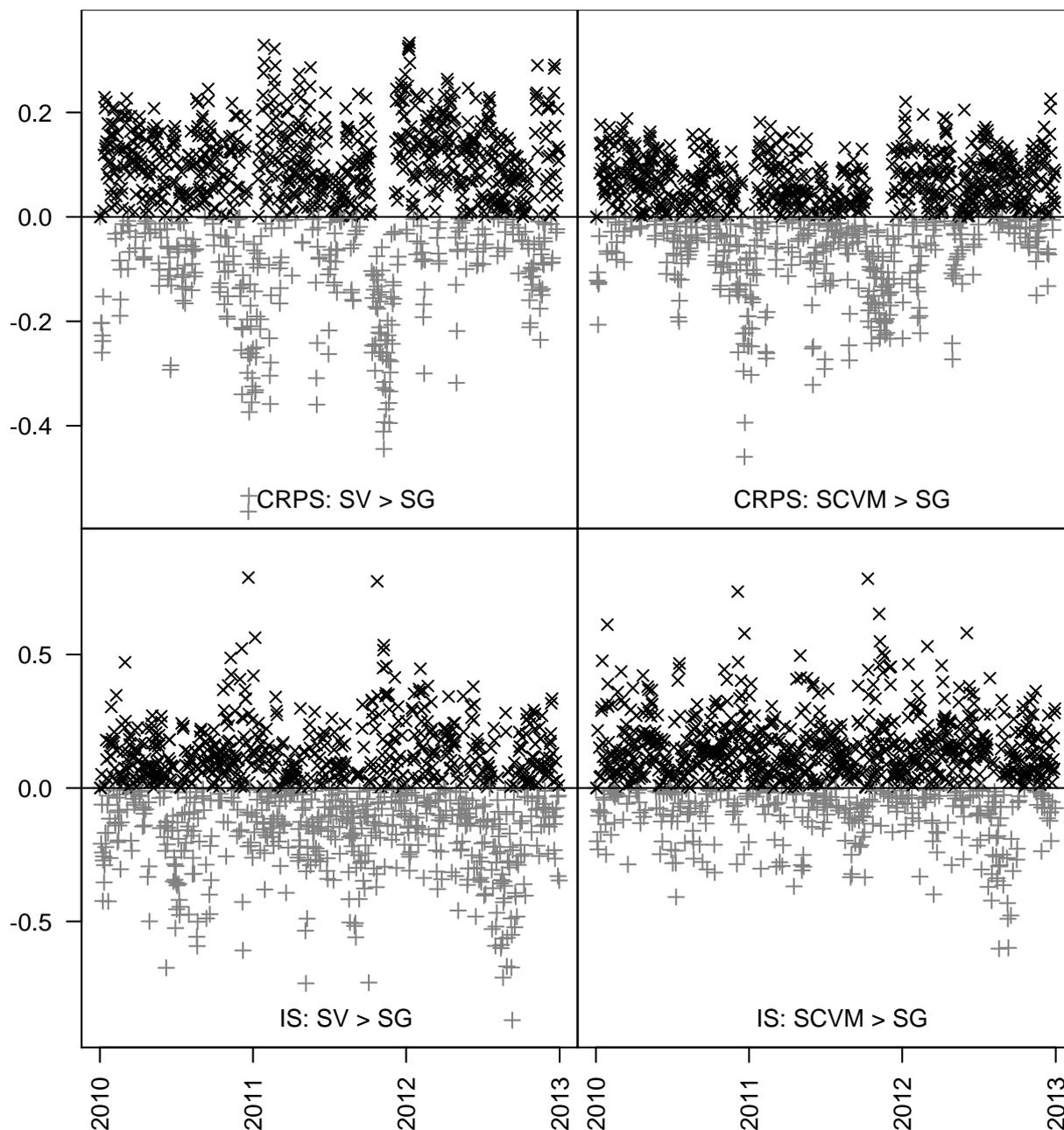
Figure 8.2.10: Log-score difference plots of the averaged continuous ranked probability scores respectively the averaged interval scores comparing the spatial composite vine model and the spatial R-vine model to the corresponding averaged spatial Gaussian model scores (average over all 24 observation stations of the validation data set). Points in time where the first mentioned models have the higher average scores are marked by a black x. On the other hand, points in time where the spatial Gaussian model has the higher average scores are marked by a gray plus sign.

8.2.2   Comprehensive prediction all over Germany

Finally we aim to predict the mean temperatures all over Germany. For this purpose we perform prediction for all pixels on a regular 80×120 grid. For each of these $80 \cdot 120 = 9600$ pixels and for each point in time, at least up to the point in time of interest (e.g. $t_i = 200$), we have to perform several (let's say 100) simulations from the predictive distribution in order to get adequate predictions. It becomes clear, that this procedure is connected with high computational effort. In our case, where we want to predict the mean temperatures on the 19th of July 2010 we have to perform $9600 \cdot 200 \cdot 100 = 192$million simulations and back transformations in total. Comparison of the predictive densities (5.3.4) of the spatial R-vine model and (6.5.2) of the spatial composite vine model yields that the computational effort for a prediction from the spatial R-vine model is usually higher. Due to this fact we decided to restrict us to a prediction based on the spatial composite vine model and to compare it to a prediction from the spatial Gaussian model. The results are visualized in Figures 8.2.11-8.2.15.

Figure 8.2.11 compares the two **level plots of the point predictions of the mean temperatures on the** 80×120 **grid covering Germany for the** 19**th of July** 2010. The locations and temperatures of the 54 stations on which the models are built are indicated by circles, whereas the squares depict the mean temperatures at the 24 observation stations composing the validation data set. We observe comparatively high temperature predictions in the west of Germany and rather low predictions for mountainous areas, which is reasonable. The predictions from both models tend to meet the observed temperatures. In order to ease the analysis of the slight differences of both predictions, we have a look at the **difference of both predictions**, illustrated in Figure 8.2.12. The biggest differences emerge in the west of Germany. The yellow and red areas in the left plot indicate that the temperatures predicted by the spatial composite vine model are one to two degrees higher. Conversely the violet and pink areas indicate that the temperature predictions of the spatial Gaussian model are one to two degrees higher than the predictions of the spatial composite vine model.

The next two figures, Figures 8.2.13 and 8.2.14 depict the 5% **and the** 95% **quantiles of the predictions**, respectively. We observe that most of the realized temperatures at the observation stations fall into the 90% prediction interval. Moreover we are interested in the **lengths of these prediction intervals**. For this purpose we plotted the differences between the 5% and the 95% quantiles in Figure 8.2.15. We find that the prediction intervals for the spatial composite vine model are pretty short for locations close to the observation stations of the training data set (circles). Their length increases however the further one moves away from theses stations, i.e. predictions outside the range of the training data set are more uncertain. We observe however that the prediction intervals in the south-east of Germany tend to be bigger than the ones in the north-west of Germany. The respective plot for the spatial Gaussian model yields a more balanced picture. There is no big variation in the length of the prediction intervals over the area which is covered by the model. Naturally we observe increasing prediction interval length, the further we move away from this area.
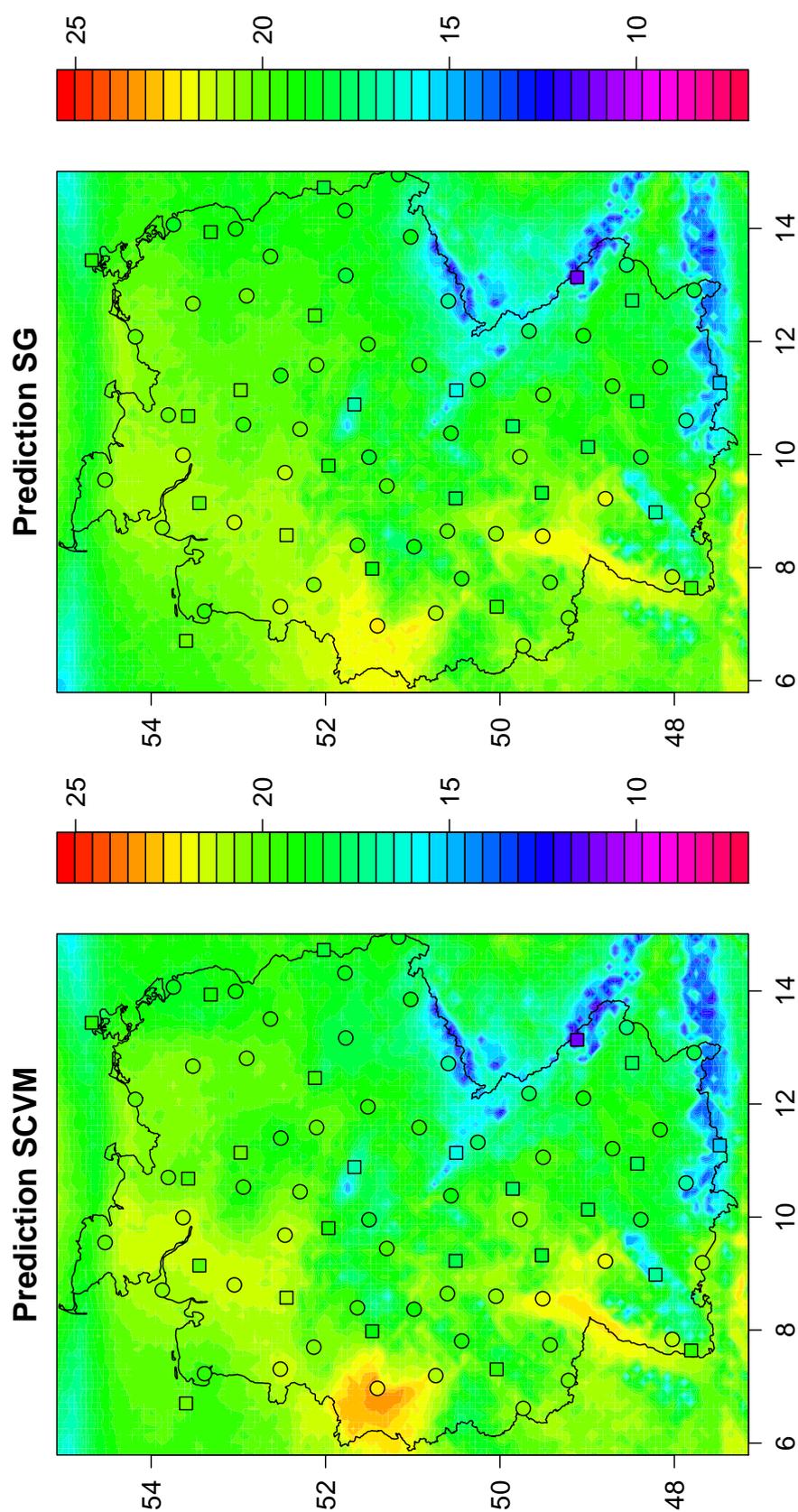
Figure 8.2.11: Level plots comparing predictions based on the spatial composite vine model and the spatial Gaussian model on a 80×120 grid covering Germany on the 19th of July 2010. Whereas the circles represent the mean temperatures at the 54 observation stations of the training data set, the squares depict the mean temperatures at the 24 stations of the validation data set.
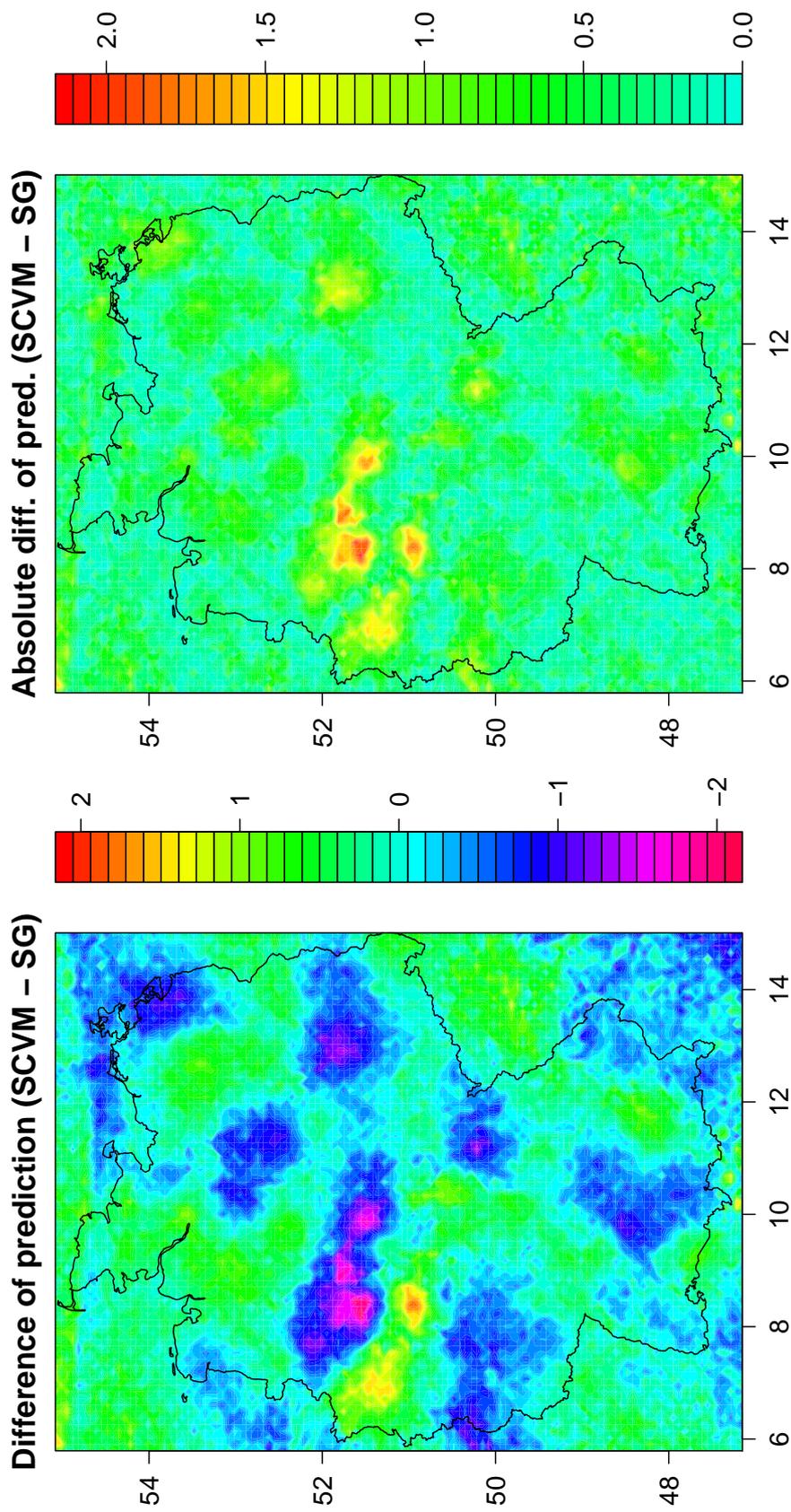
Figure 8.2.12: Level plots of (the absolute value of) the difference between the predictions based on the spatial composite vine model and the spatial Gaussian model on a 80×120 grid covering Germany on the 19th of July 2010.
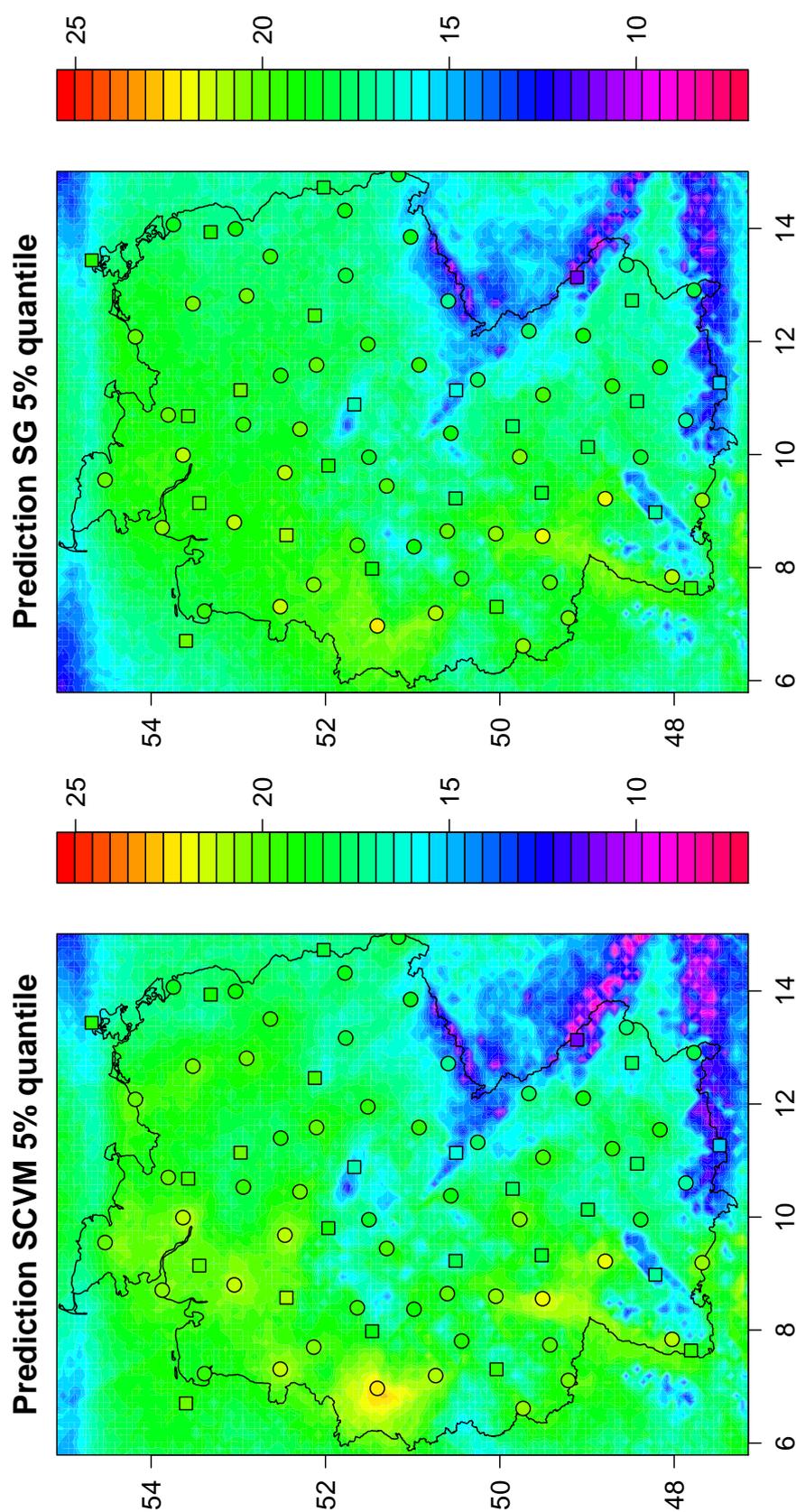
Figure 8.2.13: Level plots comparing the 5% quantiles of predictions based on the spatial composite vine model and the spatial Gaussian model on a 80×120 grid covering Germany on the 19th of July 2010. Whereas the circles represent the mean temperatures at the 54 observation stations of the training data set, the squares depict the mean temperatures at the 24 stations of the validation data set.
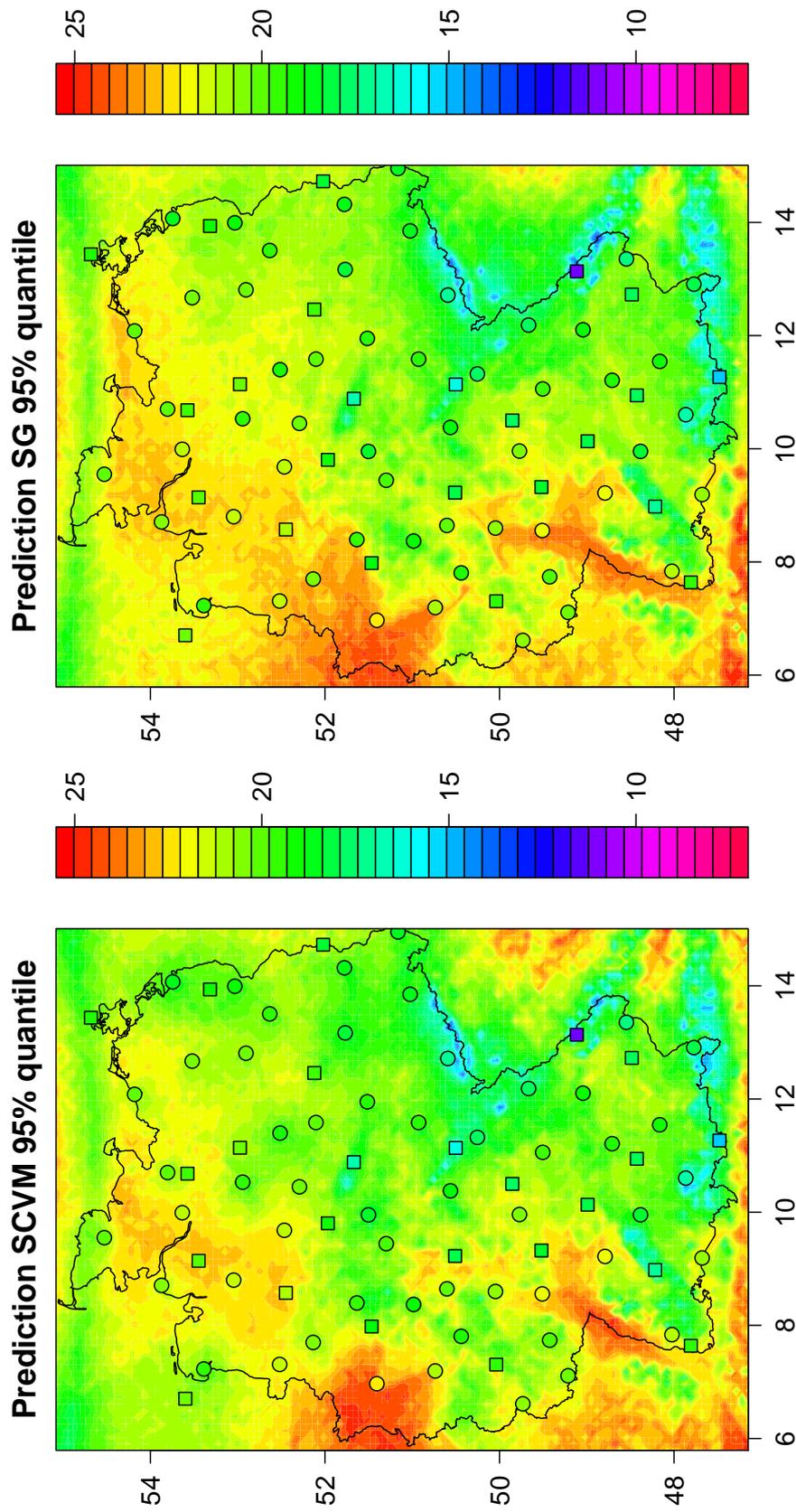
Figure 8.2.14: Level plots comparing the 95% quantiles of predictions based on the spatial composite vine model and the spatial Gaussian model on a 80×120 grid covering Germany on the 19th of July 2010. Whereas the circles represent the mean temperatures at the 54 observation stations of the training data set, the squares depict the mean temperatures at the 24 stations of the validation data set.
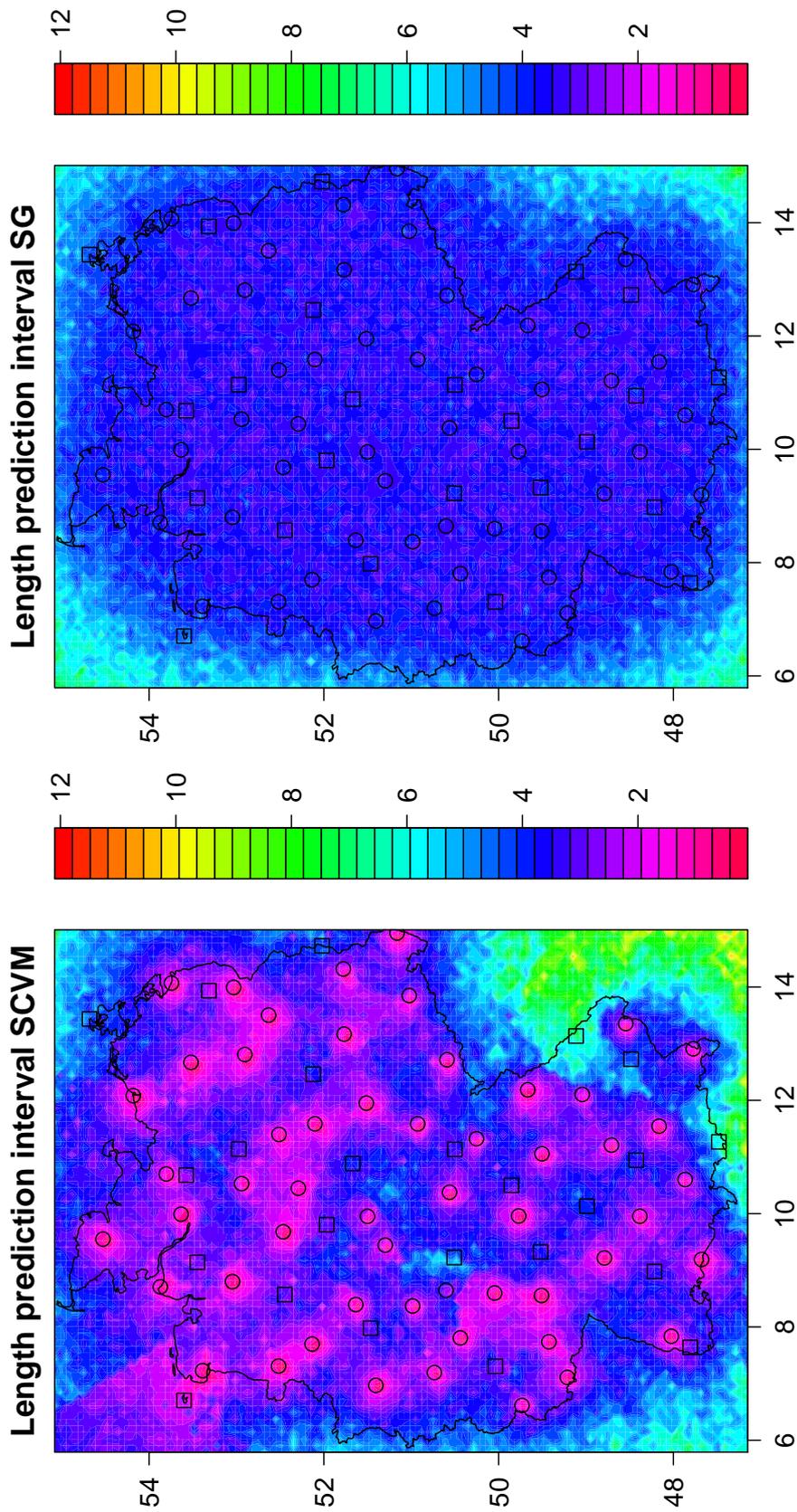
Figure 8.2.15: Level plots comparing the lengths of the 90% prediction intervals based on the spatial composite vine model and the spatial Gaussian model on a 80×120 grid covering Germany on the 19th of July 2010. The circles represent the 54 observation stations of the training data set, the squares mark the locations of the 24 stations of the validation data set.

For the purpose of comparison we provide Figures A.4.1-A.4.5 in Appendix A.4, which are the equivalent figures to Figures 8.2.11-8.2.15 for the 20th of January 2010.

We terminate our model comparison at this point and advance to a summary of the results of this thesis and the drawn conclusions.

Chapter 9

Conclusions and outlook

Finally we aim to summarize the results of the investigations performed in this thesis and give a brief outlook on possible model improvements and future fields of interest. The modeling process in this thesis evolved starting from a marginal model to the presentation of two new vine copula based models for spatial dependencies, the spatial R-vine model and the spatial composite vine model, and ended in an evaluation and comparison of these models. Furthermore, prediction methods based on the presented models were introduced. The fundament of all modeling was a data set of daily mean temperatures at 54 selected observation stations across Germany.

The marginal model is intended to capture the marginal trends and distributions of the single variables whose dependencies are modeled subsequently. Besides an intercept term, components which capture the seasonal fluctuations and the temporal dependencies of the data are included into the regression model of the marginals. Moreover we found that a skew-t distribution is an appropriate distribution for the error terms of the regression model, since we observed a skew and heavy tails for the residuals calculated from the data. Due to the fact that the final goal of the thesis is the prediction from the developed models, it was reasonable to model all marginals jointly, depending on elevation, longitude and latitude. An investigation of the parameter estimates obtained from separate marginal model fits led to the replacement of the model parameters by polynomials of these quantities. Further investigations showed, that the eventually developed marginal model, which is summarized in the subsequent box, is able to eliminate any seasonal fluctuations and time dependencies in the data to a reasonable extent. However we found that our marginal model can lead to inadequate prediction results if the location from which to predict lies outside the range of the data based on which we developed our models. Obviously, this is due to the inclusion of the polynomial components into our model. As already addressed in Chapter 5, there is room for an improvement of our marginal model, for instance by applying B-splines instead of the polynomials. On the other hand it might be reasonable to select the training data observation stations such that their locations are evenly spread over the area from which one wants to predict.

175

joint marginal model (JMM):

$$\widetilde{Y}_t^s = g\left(t, \widetilde{Y}_{t-1}^s, \widetilde{Y}_{t-2}^s, \widetilde{Y}_{t-3}^s, x_{\mathrm{elev},s}, x_{\mathrm{long},s}, x_{\mathrm{lat},s}; \boldsymbol{\beta}_{\mathrm{ls}}\right) + \varepsilon_t^s$$

$$\varepsilon_t^s \sim \mathrm{skew}\text{-}t\left(\xi(s), \omega(s), \alpha(s), \nu(s)\right)$$

$$t = 1, \ldots, N, s = 1, \ldots, d$$

$$g(t, s) := g\left(t, \widetilde{Y}_{t-1}^s, \widetilde{Y}_{t-2}^s, \widetilde{Y}_{t-3}^s, x_{\mathrm{elev},s}, x_{\mathrm{long},s}, x_{\mathrm{lat},s}; \boldsymbol{\beta}_{\mathrm{ls}}\right)$$

$$:= \beta_0(s) + \beta_{\sin}(s)\sin\left(\frac{2\pi t}{365.25}\right) + \beta_{\cos}(s)\cos\left(\frac{2\pi t}{365.25}\right)$$

$$+ \gamma_1(s)\widetilde{Y}_{t-1}^s + \gamma_2(s)\widetilde{Y}_{t-2}^s + \gamma_3(s)\widetilde{Y}_{t-3}^s$$

$$\beta_0(s) := \beta_{00} + \sum_{j=1}^{1}\beta_{01j}x_{\mathrm{elev},s}^j + \sum_{l=1}^{1}\beta_{03l}x_{\mathrm{lat},s}^l$$

$$\beta_{\sin}(s) := \beta_{\sin 0} + \sum_{j=1}^{4}\beta_{\sin 1j}x_{\mathrm{elev},s}^j + \sum_{k=1}^{1}\beta_{\sin 2k}x_{\mathrm{long},s}^k + \sum_{l=1}^{6}\beta_{\sin 3l}x_{\mathrm{lat},s}^l$$

$$\beta_{\cos}(s) := \beta_{\cos 0} + \sum_{j=1}^{6}\beta_{\cos 1j}x_{\mathrm{elev},s}^j + \sum_{k=1}^{2}\beta_{\cos 2k}x_{\mathrm{long},s}^k + \sum_{l=1}^{1}\beta_{\cos 3l}x_{\mathrm{lat},s}^l$$

$$\gamma_1(s) := \gamma_{10} + \sum_{j=1}^{1}\gamma_{11j}x_{\mathrm{elev},s}^j + \sum_{k=1}^{2}\gamma_{12k}x_{\mathrm{long},s}^k + \sum_{l=1}^{6}\gamma_{13l}x_{\mathrm{lat},s}^l$$

$$\gamma_2(s) := \gamma_{20} + \sum_{j=1}^{1}\gamma_{21j}x_{\mathrm{elev},s}^j + \sum_{k=1}^{2}\gamma_{22k}x_{\mathrm{long},s}^k + \sum_{l=1}^{6}\gamma_{23l}x_{\mathrm{lat},s}^l$$

$$\gamma_3(s) := \gamma_{30} + \sum_{k=1}^{4}\gamma_{32k}x_{\mathrm{long},s}^k + \sum_{l=1}^{7}\gamma_{33l}x_{\mathrm{lat},s}^l$$

$$\xi(s) := \xi_0 + \sum_{j=1}^{1}\xi_{1j}x_{\mathrm{elev},s}^j + \sum_{k=1}^{2}\xi_{2k}x_{\mathrm{long},s}^k + \sum_{l=1}^{1}\xi_{3l}x_{\mathrm{lat},s}^l$$

$$\omega(s) := \exp\left\{\omega_0 + \sum_{j=1}^{3}\omega_{1j}x_{\mathrm{elev},s}^j + \sum_{k=1}^{1}\omega_{2k}x_{\mathrm{long},s}^k + \sum_{l=1}^{6}\omega_{3l}x_{\mathrm{lat},s}^l\right\}$$

$$\alpha(s) := \alpha_0 + \sum_{j=1}^{4}\alpha_{1j}x_{\mathrm{elev},s}^j + \sum_{k=1}^{2}\alpha_{2k}x_{\mathrm{long},s}^k + \sum_{l=1}^{1}\alpha_{3l}x_{\mathrm{lat},s}^l$$

$$\nu(s) := \exp\left\{\nu_0 + \sum_{j=1}^{2}\nu_{1j}x_{\mathrm{elev},s}^j + \sum_{k=1}^{2}\nu_{2k}x_{\mathrm{long},s}^k + \sum_{l=1}^{4}\nu_{3l}x_{\mathrm{lat},s}^l\right\}$$

An extensive analysis of the structure of an (truncated[1]) R-vine copula fitted to the data led to a first new model for spatial dependencies, the spatial R-vine model. The investigation of the relationship between the Kendall's τ values occurring in the R-vine copula and the distances and elevation differences which can be associated to these Kendall's τ's proposed different kinds of tree-wise reparametrizations of the first copula parameters. We found that the explanatory power of the elevation differences is comparatively small, whereas the station distances are able to explain the respective correlations to a big extent. This induced the selection of a reparametrization which accounts for all distances between the observation stations which are associated to the respective bivariate copulae of the R-vine copula specification. Moreover a reparametrization of the second copula parameters $\widetilde{\nu}^l_{i(e),j(e)|\mathcal{D}_e}$ depending on the tree number $l = 1, \ldots, 10$ is applied. All in all the selected reparametrizations led to a reduction from 733 parameters (truncated R-vine copula) to 41 parameters. They are summarized subsequently.

---

spatial R-vine model (SV):

$$\widetilde{\theta}^l_{i(e),j(e)|\mathcal{D}_e} = g^{\mathrm{dist}}_l \left(d_{i(e),j(e)}, \overline{d_{i(e),\mathcal{D}_e}}, \overline{d_{j(e),\mathcal{D}_e}} | \boldsymbol{\beta}_{\mathrm{dist},l}, t_{i(e),j(e)|\mathcal{D}_e} \right)$$
$$= \mathrm{T}_{\tau \to \theta} \left(\mathrm{F}^{-1}_z \left(h^{\mathrm{dist}}_l(e|\boldsymbol{\beta}_{\mathrm{dist},l}) \right) ; t_{i(e),j(e)|\mathcal{D}_e} \right), \quad e \in \mathcal{E}_l, \quad l = 1, \ldots, 10$$

$$h^{\mathrm{dist}}_1 (e|\boldsymbol{\beta}_{\mathrm{dist},1}) = \beta^{\mathrm{dist}}_{1,0} + \beta^{\mathrm{dist}}_{1,1} \ln \left(d_{i(e),j(e)} \right), \quad e \in \mathcal{E}_1$$
$$h^{\mathrm{dist}}_l (e|\boldsymbol{\beta}_{\mathrm{dist},l}) = \beta^{\mathrm{dist}}_{l,0} + \beta^{\mathrm{dist}}_{l,1} \ln \left(d_{i(e),j(e)} \right)$$
$$+ \beta^{\mathrm{dist}}_{l,2} \ln \left(\overline{d_{i(e),\mathcal{D}_e}} \right) + \beta^{\mathrm{dist}}_{l,3} \ln \left(\overline{d_{j(e),\mathcal{D}_e}} \right), \quad e \in \mathcal{E}_l, \quad l = 2, \ldots, 10$$

$$\widetilde{\nu}^l_{i(e),j(e)|\mathcal{D}_e} = \exp \left\{ (1, l, l^2) \cdot \boldsymbol{\beta}^{\mathrm{SV}}_\nu \right\}, \quad e \in \mathcal{E}_l, \quad l = 1, \ldots, 10$$

---

[1]We considered a truncation after tree ten.

Our second spatial dependency model is built on the theory of composite likelihood methods. In a first step the model structure, a composition of C-vine copulae, is set up, which results in the definition of a composite vine model. For each variable u_t^s under consideration we determined the three closest[2] variables denoted as $u_t^{p_s}$, $u_t^{q_s}$ and $u_t^{r_s}$, to set up the four dimensional C-vines composing the model. The resulting components $\mathcal{L}_s$ of the composite likelihood $\mathcal{L}_{\text{CVM}}$ and the respective weights w_s are given in the following box.

---

composite vine model (CVM):

$$\mathcal{L}_{\text{CVM}}\left(\boldsymbol{\theta}^{\text{CVM}}, \boldsymbol{\nu}^{\text{CVM}} \mid \boldsymbol{u}^1, \ldots, \boldsymbol{u}^d\right) = \prod_{s=1}^{d} \prod_{t=1}^{N} \left[\mathcal{L}_s\left(\boldsymbol{\theta}_s, \boldsymbol{\nu}_s \mid u_t^s, u_t^{p_s}, u_t^{q_s}, u_t^{r_s}\right)\right]^{w_s}$$

with

$$\begin{aligned}
\mathcal{L}_s\left(\boldsymbol{\theta}_s, \boldsymbol{\nu}_s \mid u_t^s, u_t^{p_s}, u_t^{q_s}, u_t^{r_s}\right) &= c_s\left(u_t^s, u_t^{p_s}, u_t^{q_s}, u_t^{r_s}; \boldsymbol{\theta}_s, \boldsymbol{\nu}_s\right) \\
&= c_{sp_s}\left(u_t^s, u_t^{p_s}; \theta_{sp_s}, \nu_{sp_s}\right) \\
&\cdot c_{sq_s}\left(u_t^s, u_t^{q_s}; \theta_{sq_s}, \nu_{sq_s}\right) \\
&\cdot c_{sr_s}\left(u_t^s, u_t^{r_s}; \theta_{sr_s}, \nu_{sr_s}\right) \\
&\cdot c_{p_s q_s; s}\left(\breve{u}_t^{p_s}, \breve{u}_t^{q_s}; \theta_{p_s q_s|s}, \nu_{p_s q_s|s}\right) \\
&\cdot c_{p_s r_s; s}\left(\breve{u}_t^{p_s}, \breve{u}_t^{r_s}; \theta_{p_s r_s|s}, \nu_{p_s r_s|s}\right) \\
&\cdot c_{q_s r_s; sp_s}\left(\breve{u}_t^{q_s}, \breve{u}_t^{r_s}; \theta_{q_s r_s|sp_s}, \nu_{q_s r_s|sp_s}\right), \quad t = 1, \ldots, N
\end{aligned}$$

$$\breve{u}_t^o = F(u_t^o \mid u_t^s) = \frac{\partial}{\partial u_t^s} C_{os}\left(u_t^o, u_t^s; \theta_{so}, \nu_{so}\right), \quad o = p_s, q_s, r_s$$

$$\breve{u}_t^o = F(u_t^o \mid u_t^s, u_t^{p_s}) = \frac{\partial}{\partial \breve{u}_t^{p_s}} C_{op_s; s}\left(\breve{u}_t^o, \breve{u}_t^{p_s}; \theta_{p_s o|s}, \nu_{p_s o|s}\right), \quad o = q_s, r_s$$

$$d = \# \text{ of components}$$
$$N = \# \text{ of observations}$$

and weights

$$w_s := \frac{1}{n_s}, \quad \text{with} \quad n_s := \#\{k : s \text{ is included in C-vine } k\}, \ s = 1, \ldots, d$$

---

[2]in terms of station distance

We subsequently investigated a fit of the resulting composite vine model to the data in a similar way as for the spatial R-vine model. Again different reparametrizations were compared. This time we decided to account additionally to the distances for the elevation differences in the joint reparametrization of the first copula parameters of the first C-vine trees. The second and third tree copula reparametrizations account furthermore only for the occurring distances as predictors. The reparametrization of the second copula parameters depends on the tree number $l = 1, \ldots, 10$ and the distances indicated by the respective conditioned sets. Whereas the composite vine model is based on 495 parameters, this number could be reduced to 16 for the reparametrized composite vine model, the spatial composite vine model. It is summarized in the following box.

---

spatial composite vine model (SCVM):

$$\widetilde{\theta}_{so_1} = \mathrm{T}_{\tau \to \theta} \left(\mathrm{F}_z^{-1} \left(h_1^{\mathrm{select}}(s, o_1 | \boldsymbol{\beta}_{\mathrm{select},1}) \right) ; t_{so_1} \right)$$

$$\widetilde{\theta}_{p_s o_2 | s} = \mathrm{T}_{\tau \to \theta} \left(\mathrm{F}_z^{-1} \left(h_2^{\mathrm{select}}(s, p_s, o_2 | \boldsymbol{\beta}_{\mathrm{select},2}) \right) ; t_{p_s o_2 | s} \right)$$

$$\widetilde{\theta}_{q_s r_s | s p_s} = \mathrm{T}_{\tau \to \theta} \left(\mathrm{F}_z^{-1} \left(h_3^{\mathrm{select}}(s, p_s, q_s, r_s | \boldsymbol{\beta}_{\mathrm{select},3}) \right) ; t_{q_s r_s | s p_s} \right), \qquad s = 1, \ldots, d$$

$$h_1^{\mathrm{select}}(s, o_1 | \boldsymbol{\beta}_{\mathrm{select},1}) = \beta_{1,0}^{\mathrm{select}} + \beta_{1,1}^{\mathrm{select}} \ln(d_{s,o_1}) + \beta_{1,2}^{\mathrm{select}} \ln(e_{s,o_1})$$

$$h_2^{\mathrm{select}}(s, p_s, o_2 | \boldsymbol{\beta}_{\mathrm{select},2}) = \beta_{2,0}^{\mathrm{select}} + \beta_{2,1}^{\mathrm{select}} \ln(d_{p_s,o_2}) + \beta_{2,2}^{\mathrm{select}} \ln(d_{p_s,s}) + \beta_{2,3}^{\mathrm{select}} \ln(d_{o_2,s})$$

$$h_3^{\mathrm{select}}(s, p_s, q_s, r_s | \boldsymbol{\beta}_{\mathrm{select},3}) = \beta_{3,0}^{\mathrm{select}} + \beta_{3,1}^{\mathrm{select}} \ln(d_{q_s,r_s}) + \beta_{3,2}^{\mathrm{select}} \ln(d_{q_s,s}) + \beta_{3,3}^{\mathrm{select}} \ln(d_{q_s,p_s})$$

$$+ \beta_{3,4}^{\mathrm{select}} \ln(d_{r_s,s}) + \beta_{3,5}^{\mathrm{select}} \ln(d_{r_s,p_s})$$

$$\widetilde{\nu}_{so_1} = \exp\left(\beta_0^{\nu} + \beta_1^{\nu} \cdot 1 + \beta_2^{\nu} d_{s,o_1} \right)$$

$$\widetilde{\nu}_{p_s o_2 | s} = \exp\left(\beta_0^{\nu} + \beta_1^{\nu} \cdot 2 + \beta_2^{\nu} d_{p_s,o_2} \right)$$

$$\widetilde{\nu}_{q_s r_s | s p_s} = \exp\left(\beta_0^{\nu} + \beta_1^{\nu} \cdot 3 + \beta_2^{\nu} d_{q_s,r_s} \right)$$

$$o_1 = p_s, q_s, r_s$$

$$o_2 = q_s, r_s$$

For the purpose of model evaluation and comparison a third spatial model was introduced, the spatial Gaussian model, which models the spatial dependencies based on a Gaussian variogram model and assumes a multivariate Gaussian distribution of the mean temperatures.

---

spatial Gaussian model (SG):

$$\widetilde{\boldsymbol{Y}}_t = \boldsymbol{\mu}_t + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \overset{\text{i.i.d.}}{\sim} \mathcal{N}_d\left(\boldsymbol{0}, \Sigma(\boldsymbol{\theta}^{\text{SG}})\right), \quad t = 1, \ldots, N$$

$$\mu_t^s := g(t,s) = g\left(t, \widetilde{Y}_{t-1}^s, \widetilde{Y}_{t-2}^s, \widetilde{Y}_{t-3}^s, x_{\text{elev},s}, x_{\text{long},s}, x_{\text{lat},s}; \boldsymbol{\beta}_{\text{ls}}\right)$$

$$:= \beta_0(s) + \beta_{\sin}(s)\sin\left(\frac{2\pi t}{365.25}\right) + \beta_{\cos}(s)\cos\left(\frac{2\pi t}{365.25}\right)$$

$$+ \gamma_1(s)\widetilde{Y}_{t-1}^s + \gamma_2(s)\widetilde{Y}_{t-2}^s + \gamma_3(s)\widetilde{Y}_{t-3}^s \qquad \text{(compare marginal model)}$$

$$\gamma(h; \eta, \varsigma, \rho) := \varsigma\left(1 - \exp\left(-\frac{h^2}{\rho^2}\right)\right) + \eta\mathbb{1}_{(0,\infty)}(h) \qquad \text{(Gaussian variogram model)}$$

$$\Sigma_{i,j}(\boldsymbol{\theta}^{\text{SG}}) := \sigma^2 - \gamma(d_{i,j}; \eta, \varsigma, \rho)$$

$$\widehat{\boldsymbol{\varepsilon}}_t := \widetilde{\boldsymbol{y}}_t - \widehat{\boldsymbol{\mu}}_t(\widehat{\boldsymbol{\beta}}_{\text{ls}})$$

An overview and comparison of all dependency models under consideration is given by Tables 9.1-9.3. Amongst others, the likelihood-functions corresponding to the models summarized above are given.[3,4]

---

[3] $\widetilde{u}_t^{i(e)} = F_{i(e)|\mathcal{D}_e}(u_t^{i(e)} \,|\, \boldsymbol{u}_t^{\mathcal{D}_e})$, $\widetilde{u}_t^{j(e)} = F_{j(e)|\mathcal{D}_e}(u_t^{j(e)} \,|\, \boldsymbol{u}_t^{\mathcal{D}_e})$, $\boldsymbol{u}_t^{\mathcal{D}_e} := \{u_t^s : s \in \mathcal{D}_e\}$
[4] Reparametrized parameters are indicated by a tilde.

short name	full name	model type	trunc. level	model is based on spatial information	data level	parameters
RV	full R-vine	vine copula	–	no	copula data	$\boldsymbol{\theta}^{\mathrm{RV}}, \boldsymbol{\nu}^{\mathrm{RV}}$
TRV	truncated R-vine	truncated vine copula	10	no	copula data	$\boldsymbol{\theta}^{\mathrm{TRV}}, \boldsymbol{\nu}^{\mathrm{TRV}}$
SV	spatial R-vine	reparametrized truncated vine copula	10	yes	copula data	$\boldsymbol{\beta}_{\mathrm{dist}}^{\mathrm{SV}}, \boldsymbol{\beta}_{\nu}^{\mathrm{SV}}$
CVM	composite vine model	composite likelihood model	–	(for model structure)	copula data	$\boldsymbol{\theta}^{\mathrm{CVM}}, \boldsymbol{\nu}^{\mathrm{CVM}}$
SCVM	spatial composite vine model	reparametrized composite likelihood model	–	yes	copula data	$\boldsymbol{\beta}_{\mathrm{select}}^{\mathrm{SCVM}}, \boldsymbol{\beta}_{\nu}^{\mathrm{SCVM}}$
SG	spatial Gaussian model	semivariogram based	–	yes	residuals	$\boldsymbol{\theta}^{\mathrm{SG}}$

Table 9.1: Model overview: model type, used truncation level, usage of spatial information, data level on which the models are fitted and the model parameters.

short name	full name	likelihood-function
RV	full R-vine	$\mathcal{L}_{\mathrm{RV}}\left(\boldsymbol{\theta}^{\mathrm{RV}}, \boldsymbol{\nu}^{\mathrm{RV}} \mid \boldsymbol{u}^1, \ldots, \boldsymbol{u}^d\right) = \prod_{t=1}^{N} \prod_{l=1}^{d-1} \prod_{e\in\mathcal{E}_l} c_{i(e),j(e);\mathcal{D}_e}\left(\widetilde{u}_t^{i(e)}, \widetilde{u}_t^{j(e)}; \theta_{i(e),j(e)\mid\mathcal{D}_e}, \nu_{i(e),j(e)\mid\mathcal{D}_e}, t_{i(e),j(e)\mid\mathcal{D}_e}\right)$
TRV	truncated R-vine	$\mathcal{L}_{\mathrm{TRV}}\left(\boldsymbol{\theta}^{\mathrm{TRV}}, \boldsymbol{\nu}^{\mathrm{TRV}} \mid \boldsymbol{u}^1, \ldots, \boldsymbol{u}^d\right) = \prod_{t=1}^{N} \prod_{l=1}^{10} \prod_{e\in\mathcal{E}_l} c_{i(e),j(e);\mathcal{D}_e}\left(\widetilde{u}_t^{i(e)}, \widetilde{u}_t^{j(e)}; \theta_{i(e),j(e)\mid\mathcal{D}_e}, \nu_{i(e),j(e)\mid\mathcal{D}_e}, t_{i(e),j(e)\mid\mathcal{D}_e}\right)$
SV	spatial R-vine	$\mathcal{L}_{\mathrm{SV}}\left(\boldsymbol{\beta}^{\mathrm{SV}}_{\mathrm{dist}}, \boldsymbol{\beta}^{\mathrm{SV}}_{\nu} \mid \boldsymbol{u}^1, \ldots, \boldsymbol{u}^d\right) = \prod_{t=1}^{N} \prod_{l=1}^{10} \prod_{e\in\mathcal{E}_l} c_{i(e),j(e);\mathcal{D}_e}\left(\widetilde{u}_t^{i(e)}, \widetilde{u}_t^{j(e)}; \widetilde{\theta}^l_{i(e),j(e)\mid\mathcal{D}_e}, \widetilde{\nu}^l_{i(e),j(e)\mid\mathcal{D}_e}, t_{i(e),j(e)\mid\mathcal{D}_e}\right)$
CVM	composite vine model	$\mathcal{L}_{\mathrm{CVM}}\left(\boldsymbol{\theta}^{\mathrm{CVM}}, \boldsymbol{\nu}^{\mathrm{CVM}} \mid \boldsymbol{u}^1, \ldots, \boldsymbol{u}^d\right) = \prod_{s=1}^{d}\prod_{t=1}^{N}\left[\mathcal{L}_s\left(\boldsymbol{\theta}_s, \boldsymbol{\nu}_s \mid u_t^s, u_t^{p_s}, u_t^{q_s}, u_t^{r_s}\right)\right]^{w_s}$
SCVM	spatial composite vine model	$\mathcal{L}_{\mathrm{SCVM}}\left(\boldsymbol{\beta}^{\mathrm{SCVM}}_{\mathrm{select}}, \boldsymbol{\beta}^{\mathrm{SCVM}}_{\nu} \mid \boldsymbol{u}^1, \ldots, \boldsymbol{u}^d\right) = \prod_{s=1}^{d}\prod_{t=1}^{N}\left[\mathcal{L}_s\left(\widetilde{\boldsymbol{\theta}}_s, \widetilde{\boldsymbol{\nu}}_s \mid u_t^s, u_t^{p_s}, u_t^{q_s}, u_t^{r_s}\right)\right]^{w_s}$
SG	spatial Gaussian model	$\mathcal{L}_{\mathrm{SG}}\left(\boldsymbol{\theta}^{\mathrm{SG}} \mid \widehat{\boldsymbol{\varepsilon}}_1, \ldots, \widehat{\boldsymbol{\varepsilon}}_N\right) = \prod_{t=1}^{N} \frac{1}{\sqrt{(2\pi)^d |\Sigma(\boldsymbol{\theta}^{\mathrm{SG}})|}} \exp\left\{-\frac{1}{2}\widehat{\boldsymbol{\varepsilon}}_t^{\top} [\Sigma(\boldsymbol{\theta}^{\mathrm{SG}})]^{-1} \widehat{\boldsymbol{\varepsilon}}_t\right\}$

Table 9.2: Model overview: likelihood-functions corresponding to the models summarized above.

short name	full name	maximum (composite) log-likelihood	AIC	BIC	#par	computation time
RV	full R-vine	62714.19	-121856.4	-112932.3	1786	~ 54 days
TRV	truncated R-vine	58366.70	-115267.4	-111604.8	733	~ 3.7 days
SV	spatial R-vine	54650.11	-109218.2	-109013.4	41	~ 18 h
CVM	composite vine model	42106.41	-83222.8	-80749.5	495	~ 6.1 days
SCVM	spatial composite vine model	40938.80	-81845.6	-81765.7	16	~ 1.6 h
SG	spatial Gaussian model	-50007.76	100023.5	100043.5	4	~ 5.3 min

Table 9.3: Model overview: maximum (composite) log-likelihood, AIC, BIC, number of parameters needed, elapsed computation time.

The prediction results from all three models were quite similar and reasonable, as long as the location from which we aimed to predict lay within the range of the training data. In order to rank the different model predictions, we calculated and compared continuous ranked probability scores and interval scores. In terms of both scores we observed an outperformance of our two vine copula based spatial dependency models over the spatial Gaussian model. Direct comparison of the spatial R-vine model and the spatial composite vine model yielded an outperformance of the spatial R-vine model with regard to the continuous ranked probability scores and an outperformance of the spatial composite vine model with respect to the interval scores.

Whereas the prediction from our spatial R-vine model uses the information of all 54 observation stations of the training data set, the spatial composite vine model based predictions are only conditioned on the observations of three neighboring stations. These differences in the model structure naturally have an effect with regard to computation time. Therefore, further investigations of the performance of spatial composite vine models which consider different numbers of neighboring stations are of big interest. Moreover an application of our modeling approaches to other types of data sets is desirable, which implicitly requires the development of new marginal models. Especially an investigation of data sets where asymmetries of bivariate dependencies are observed should stand in the focus of further work on this topic.

Appendix A

Outsourced figures

A.1   Predictions: spatial R-vine model
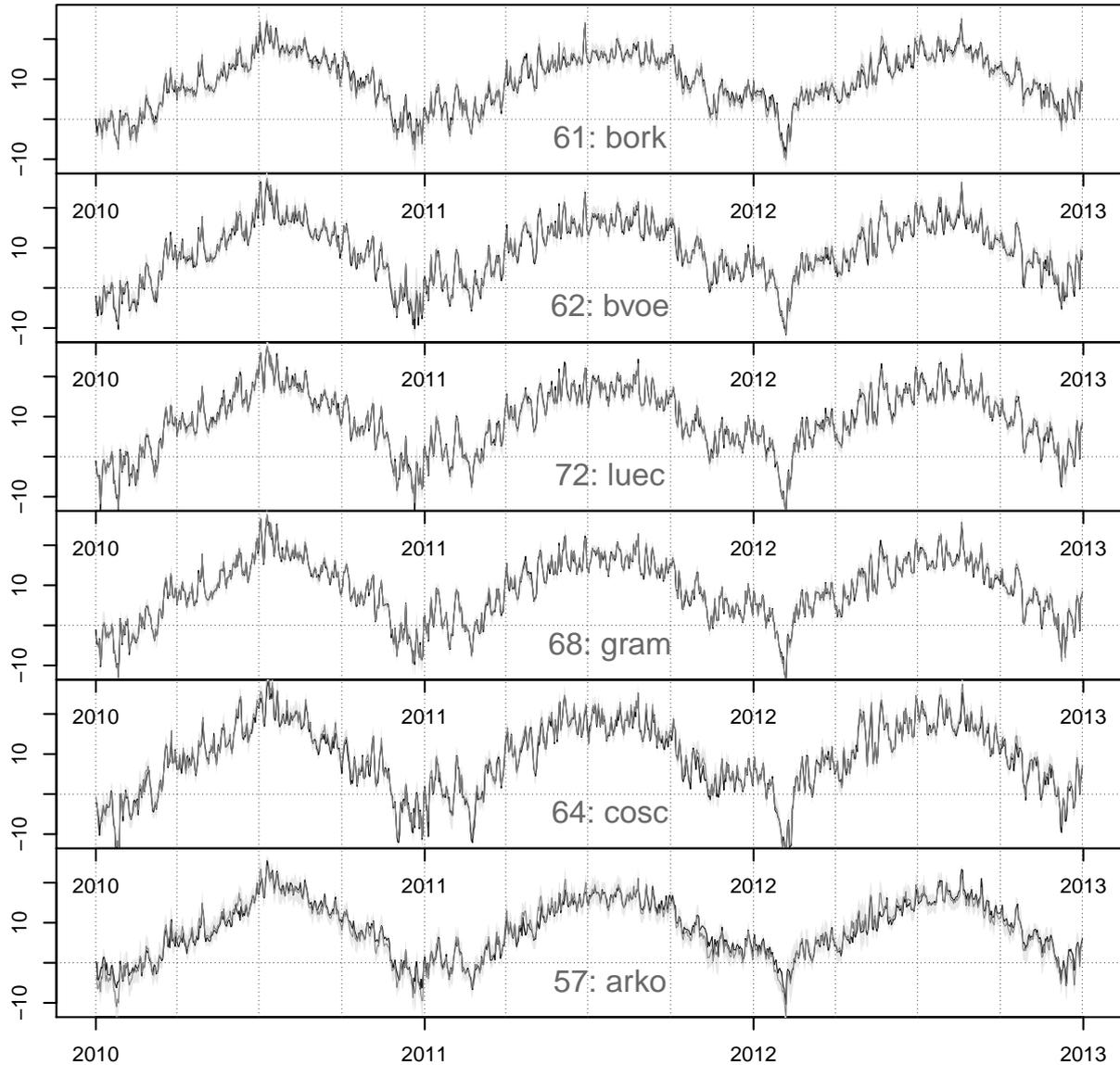
Figure A.1.1: Prediction of the mean temperatures for the observation stations 61, 62, 72, 68, 64 and 57 for the period 01/01/2010-12/31/2012 based on the spatial R-vine model. black line: observed values. dark gray line: prediction. light gray area: 95% prediction intervals.
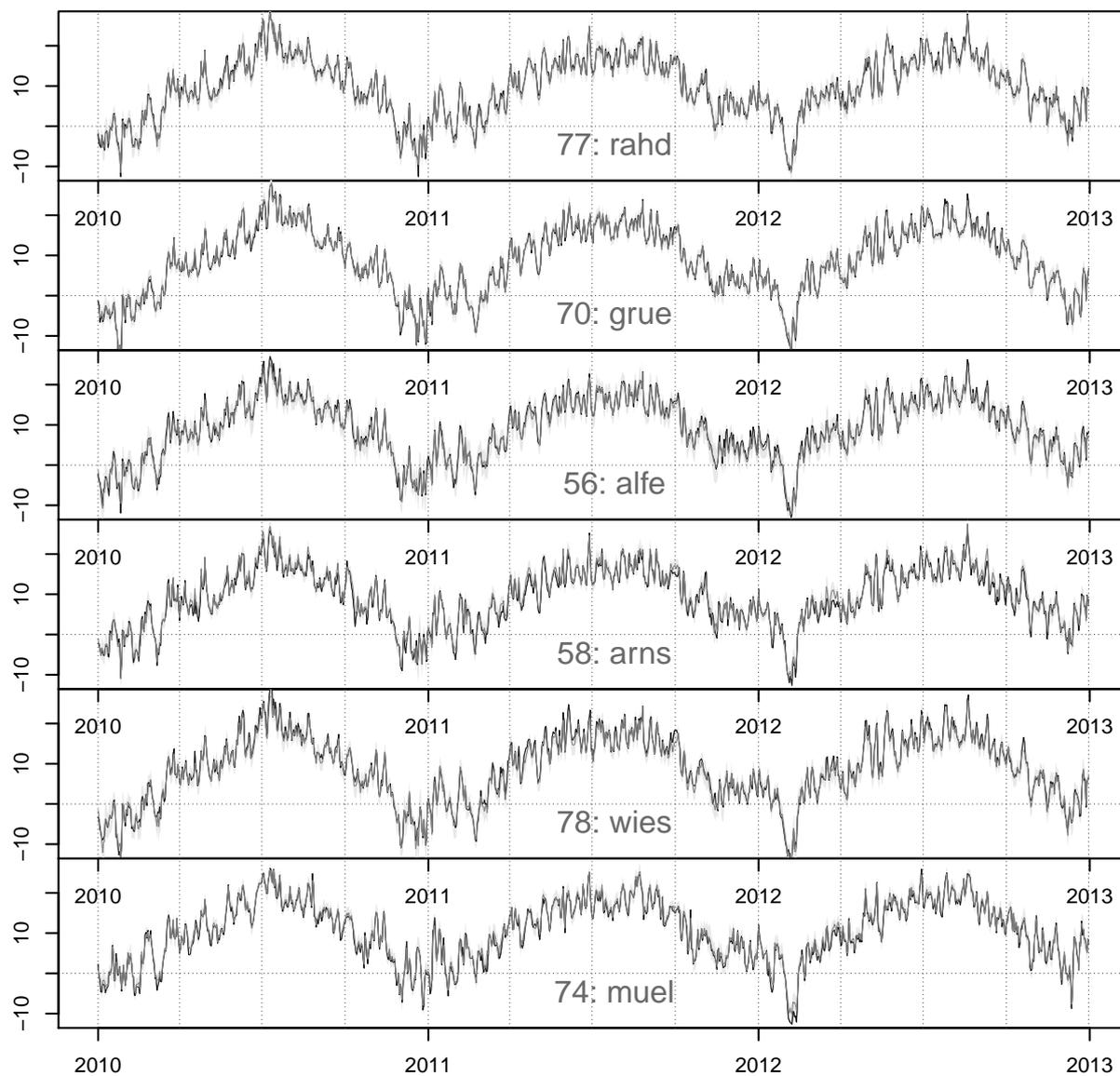
Figure A.1.2: Prediction of the mean temperatures for the observation stations 77, 70, 56, 58, 78 and 74 for the period 01/01/2010-12/31/2012 based on the spatial R-vine model. black line: observed values. dark gray line: prediction. light gray area: 95% prediction intervals.
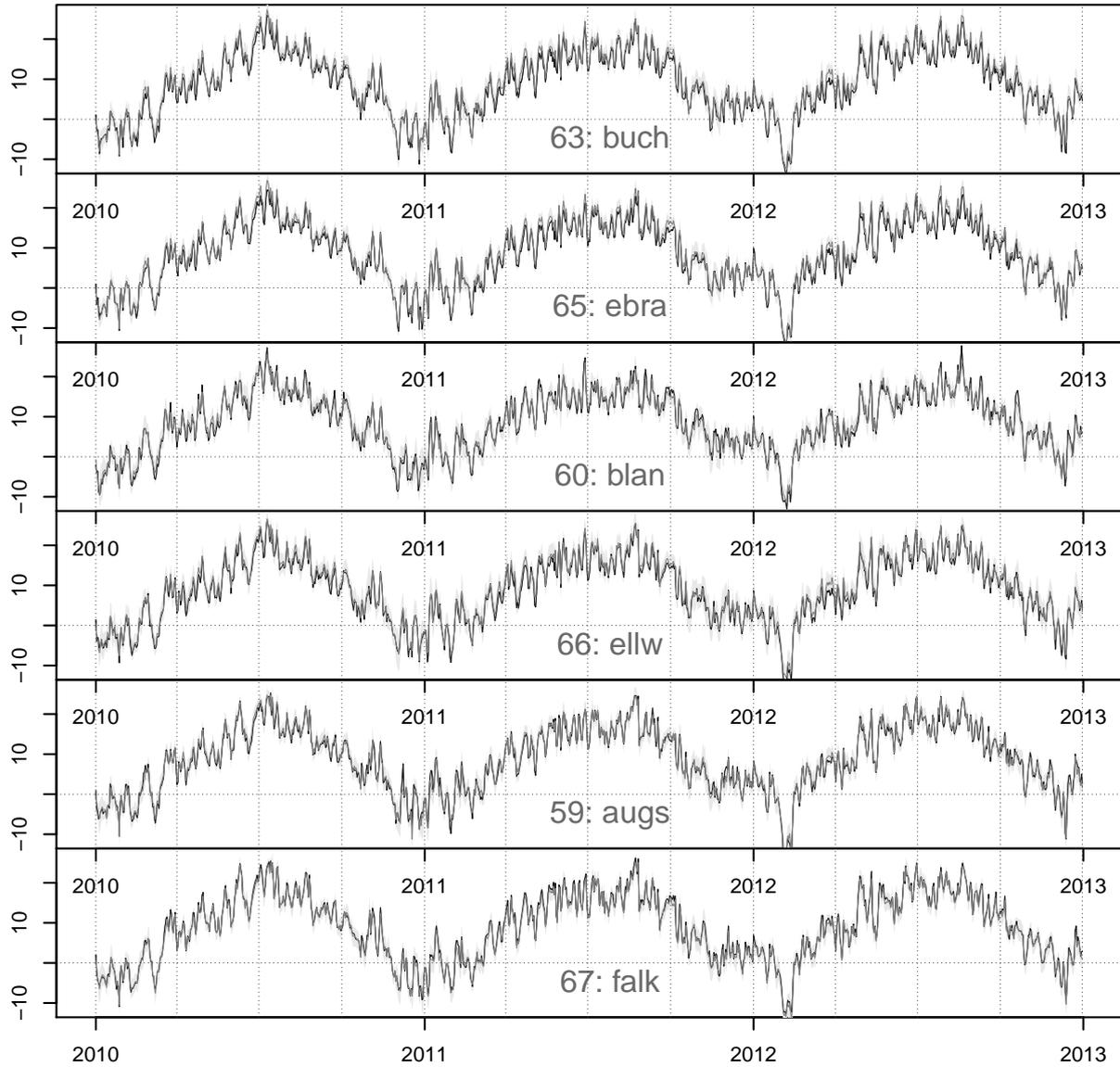
Figure A.1.3: Prediction of the mean temperatures for the observation stations 63, 65, 60, 66, 59 and 67 for the period 01/01/2010-12/31/2012 based on the spatial R-vine model. black line: observed values. dark gray line: prediction. light gray area: 95% prediction intervals.
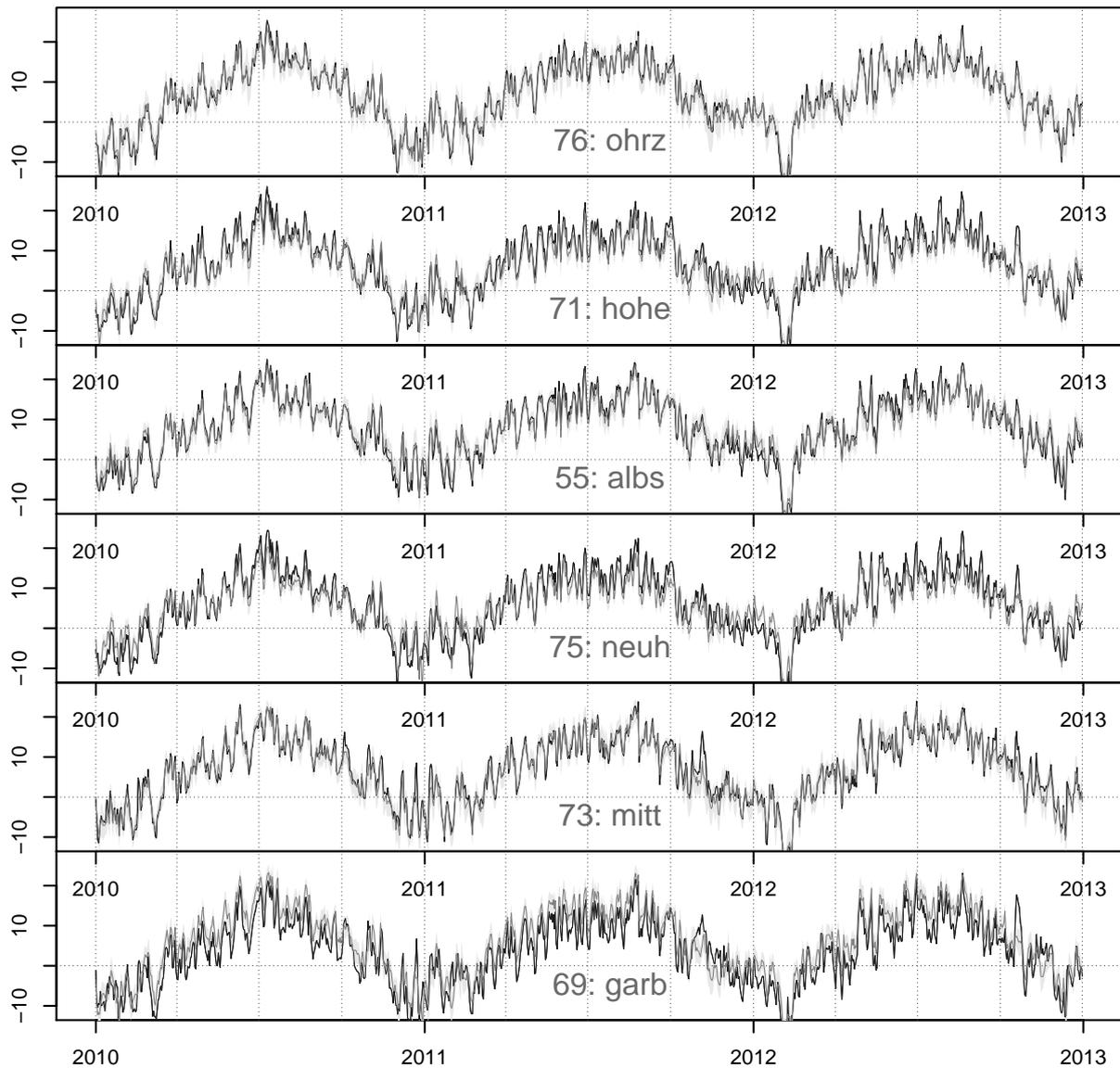
Figure A.1.4: Prediction of the mean temperatures for the observation stations 76, 71, 55, 75, 73 and 69 for the period 01/01/2010-12/31/2012 based on the spatial R-vine model. black line: observed values. dark gray line: prediction. light gray area: 95% prediction intervals.
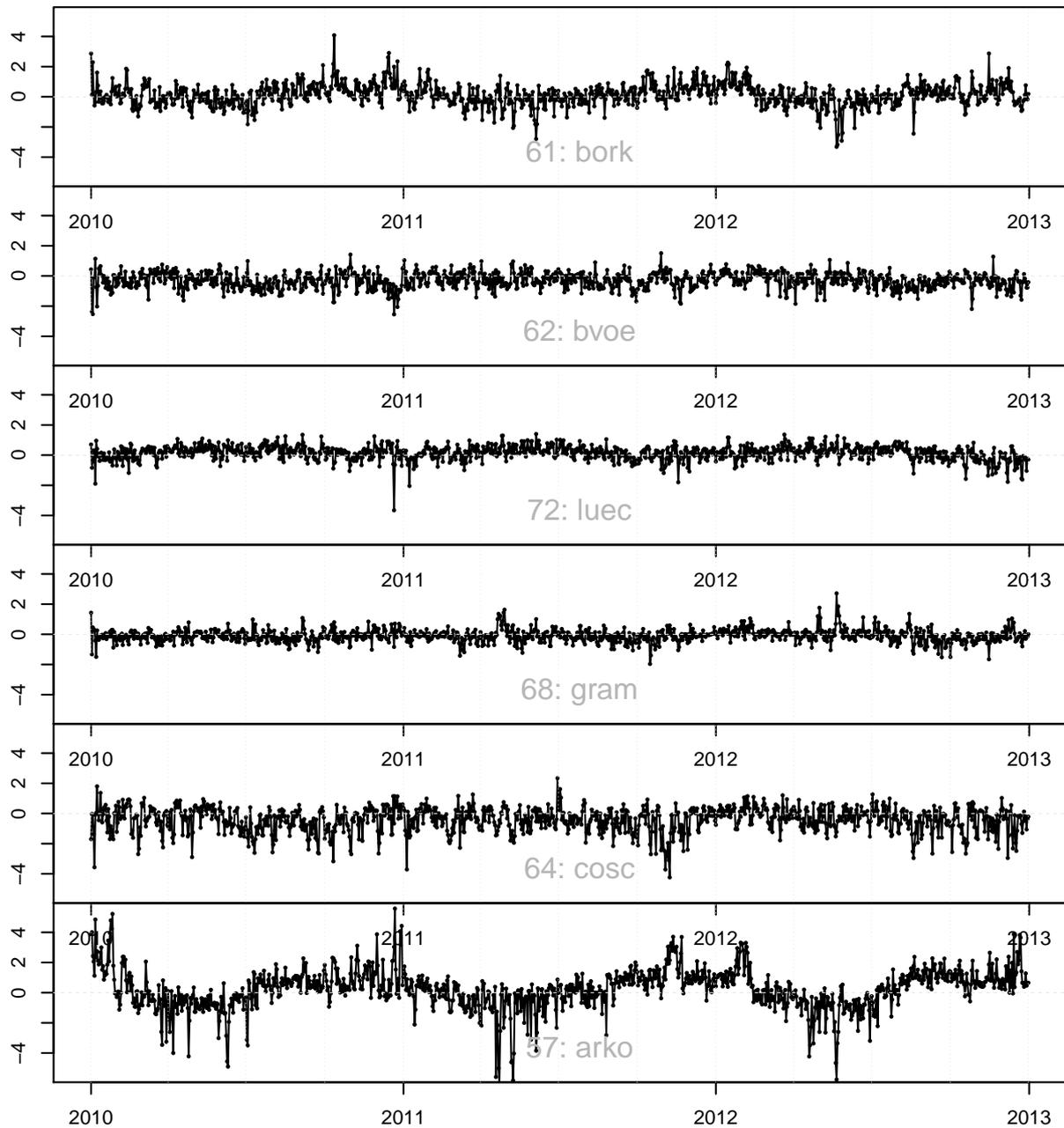
Figure A.1.5: Prediction errors of the predictions for the observation stations 61, 62, 72, 68, 64 and 57 for the period 01/01/2010-12/31/2012 based on the spatial R-vine model.
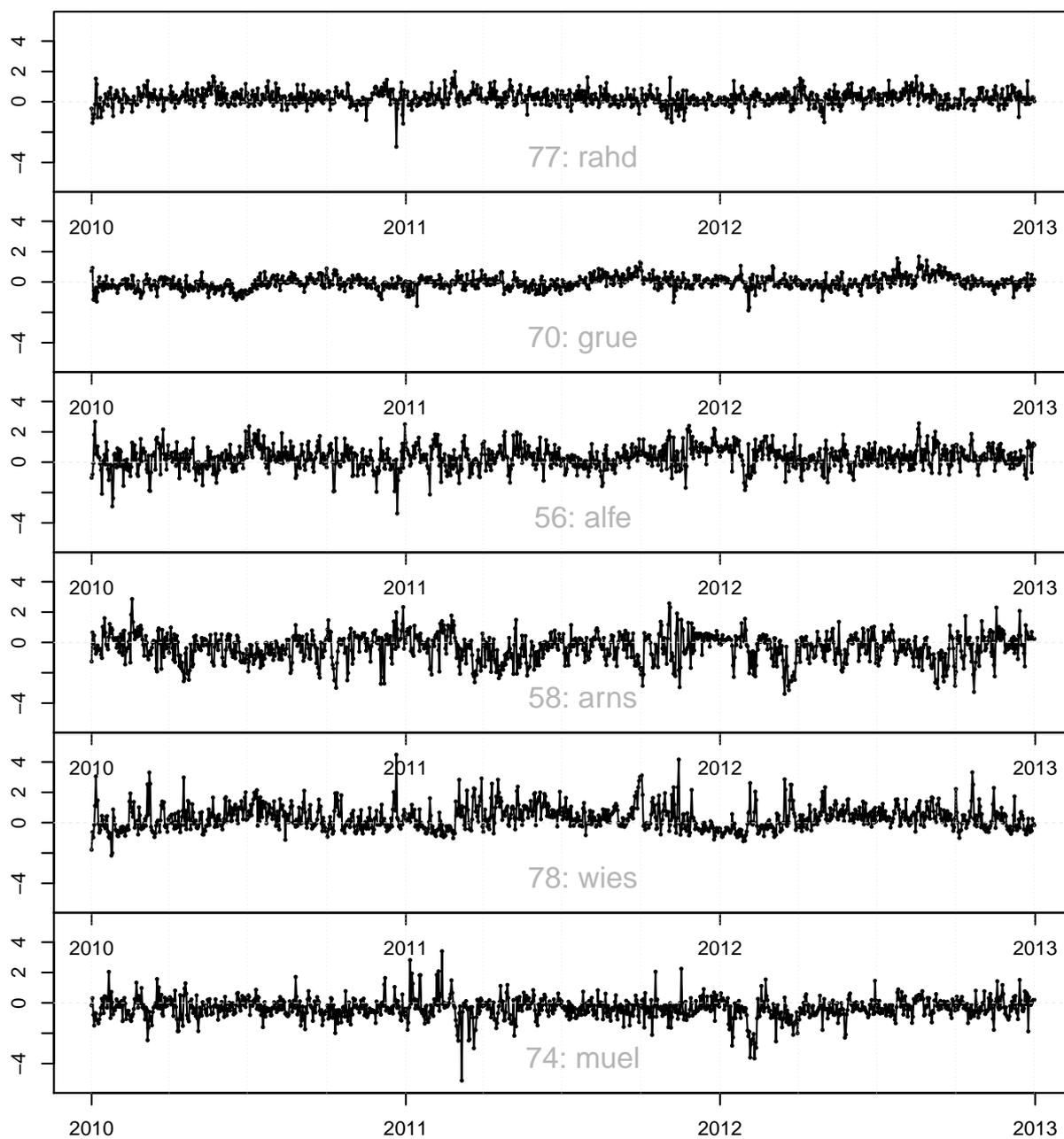
Figure A.1.6: Prediction errors of the predictions for the observation stations 77, 70, 56, 58, 78 and 74 for the period 01/01/2010-12/31/2012 based on the spatial R-vine model.
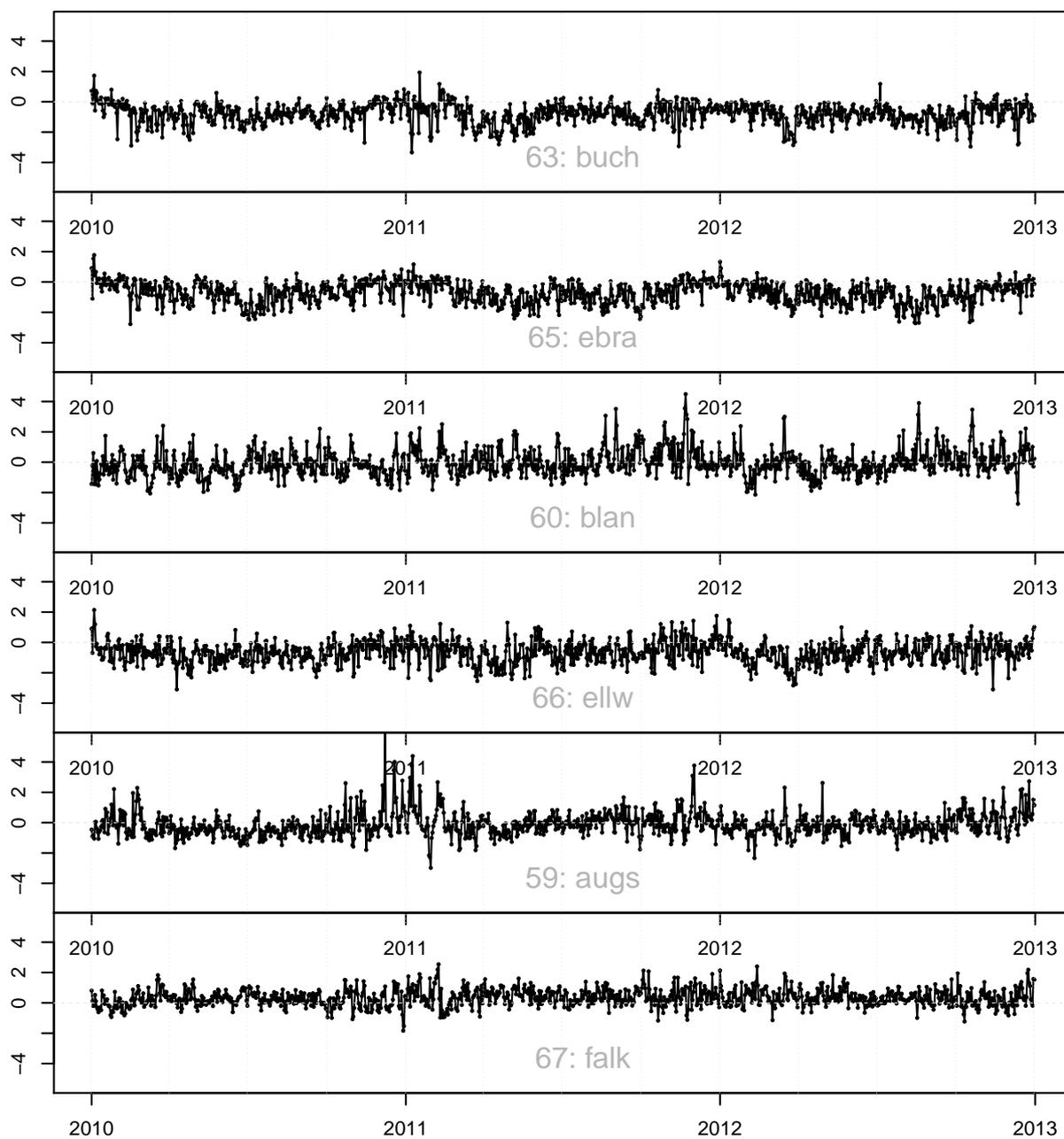
Figure A.1.7: Prediction errors of the predictions for the observation stations 63, 65, 60, 66, 59 and 67 for the period 01/01/2010-12/31/2012 based on the spatial R-vine model.
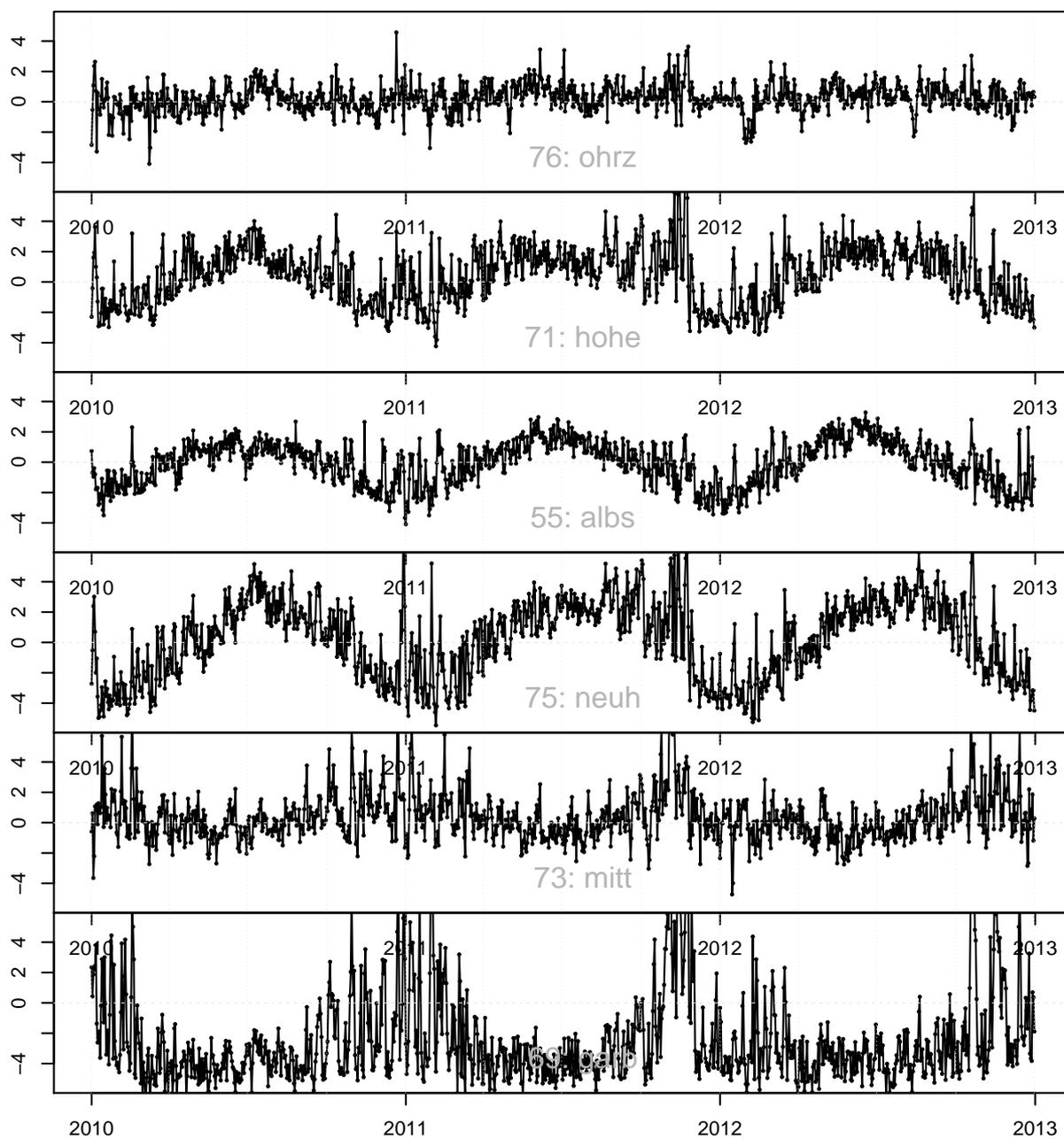
Figure A.1.8: Prediction errors of the predictions for the observation stations 76, 71, 55, 75, 73 and 69 for the period 01/01/2010-12/31/2012 based on the spatial R-vine model.
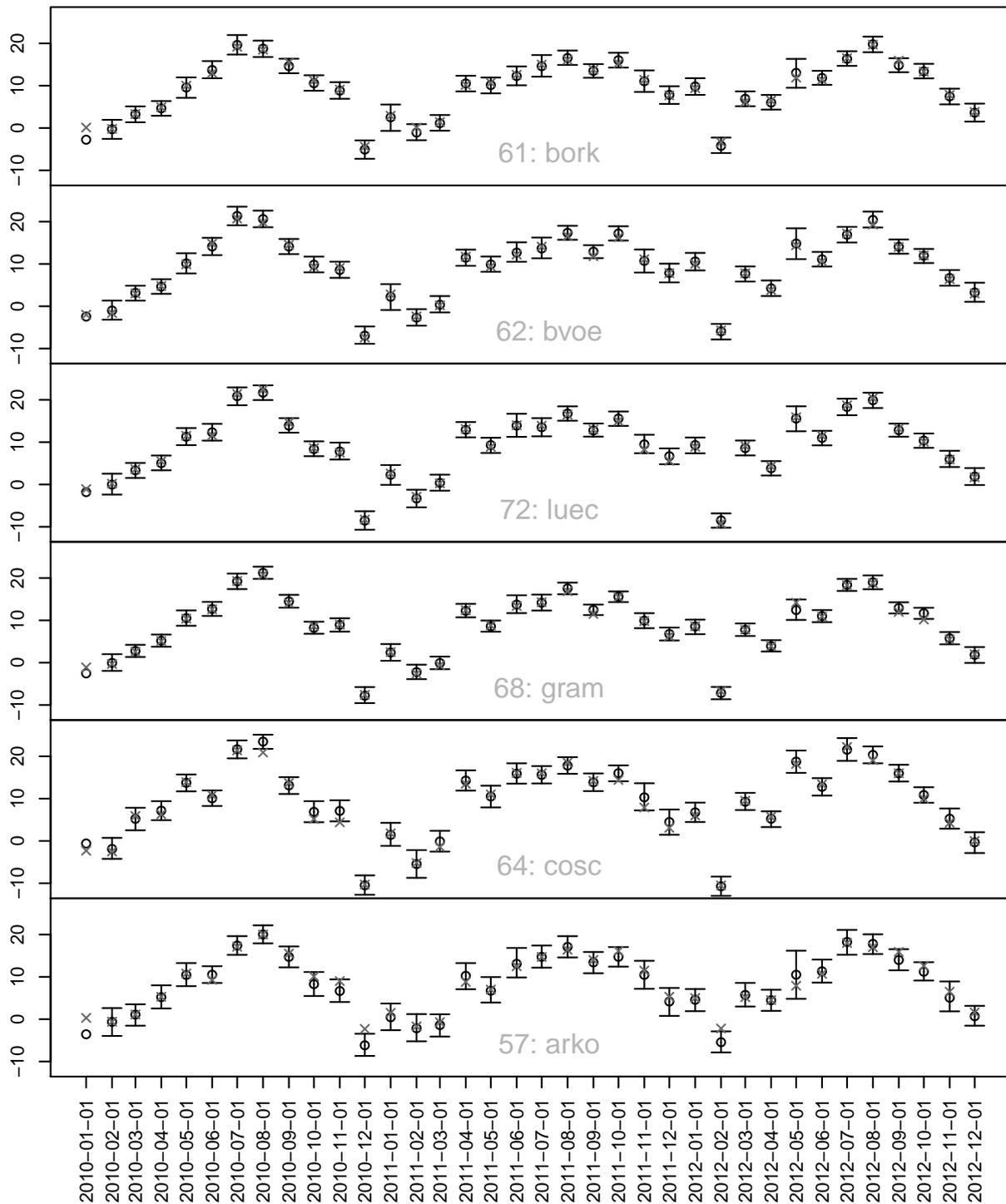
Figure A.1.9: 95% prediction intervals, point predictions and observed mean temperatures for the first days of each months during the years 2010-2012 for the observation stations 61, 62, 72, 68, 64 and 57, each calculated based on 1000 simulations from the predictive distribution of the mean temperatures at the respective observation station (spatial R-vine model).
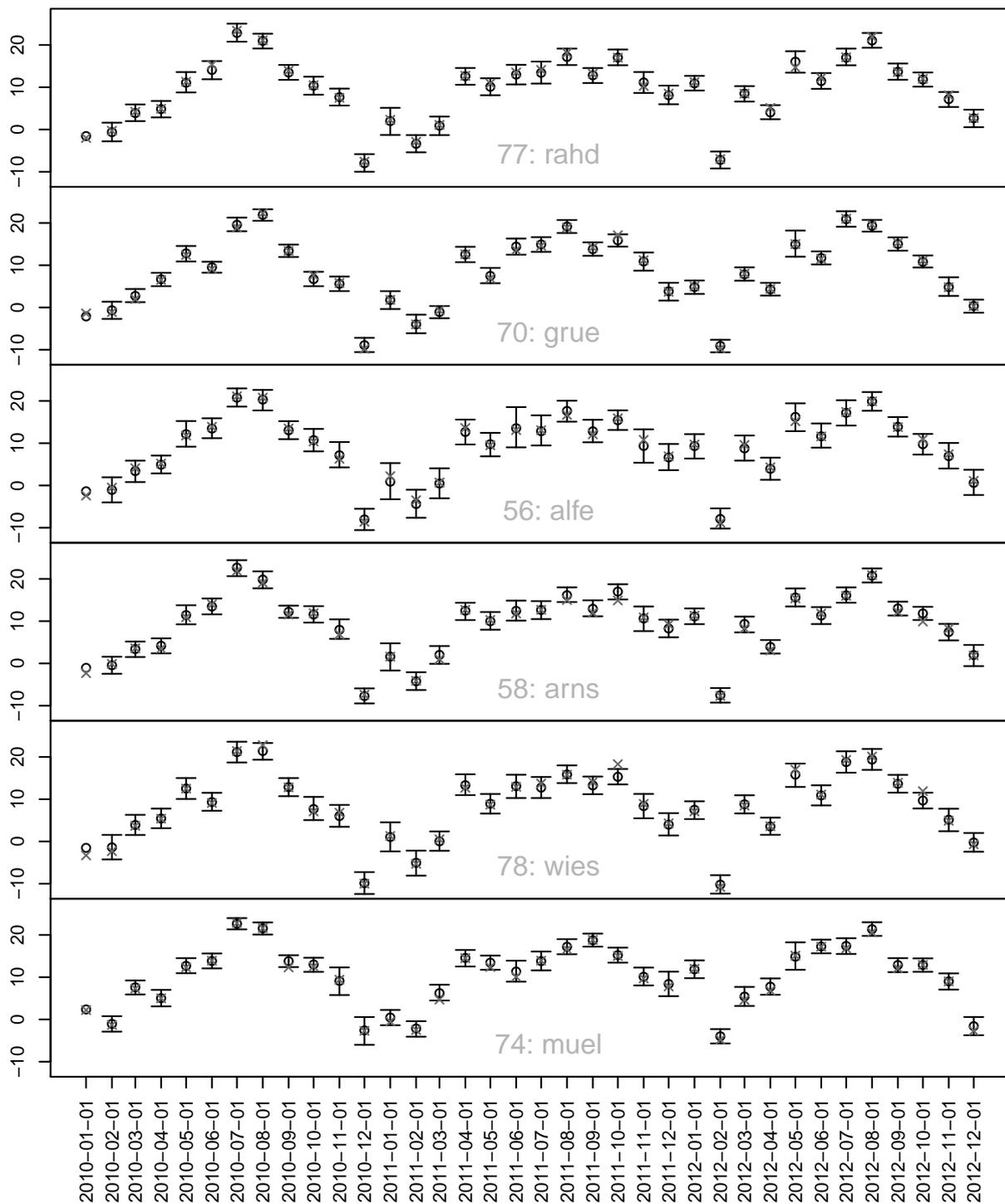
Figure A.1.10: 95% prediction intervals, point predictions and observed mean temperatures for the first days of each months during the years 2010-2012 for the observation stations 77, 70, 56, 58, 78 and 74, each calculated based on 1000 simulations from the predictive distribution of the mean temperatures at the respective observation station (spatial R-vine model).
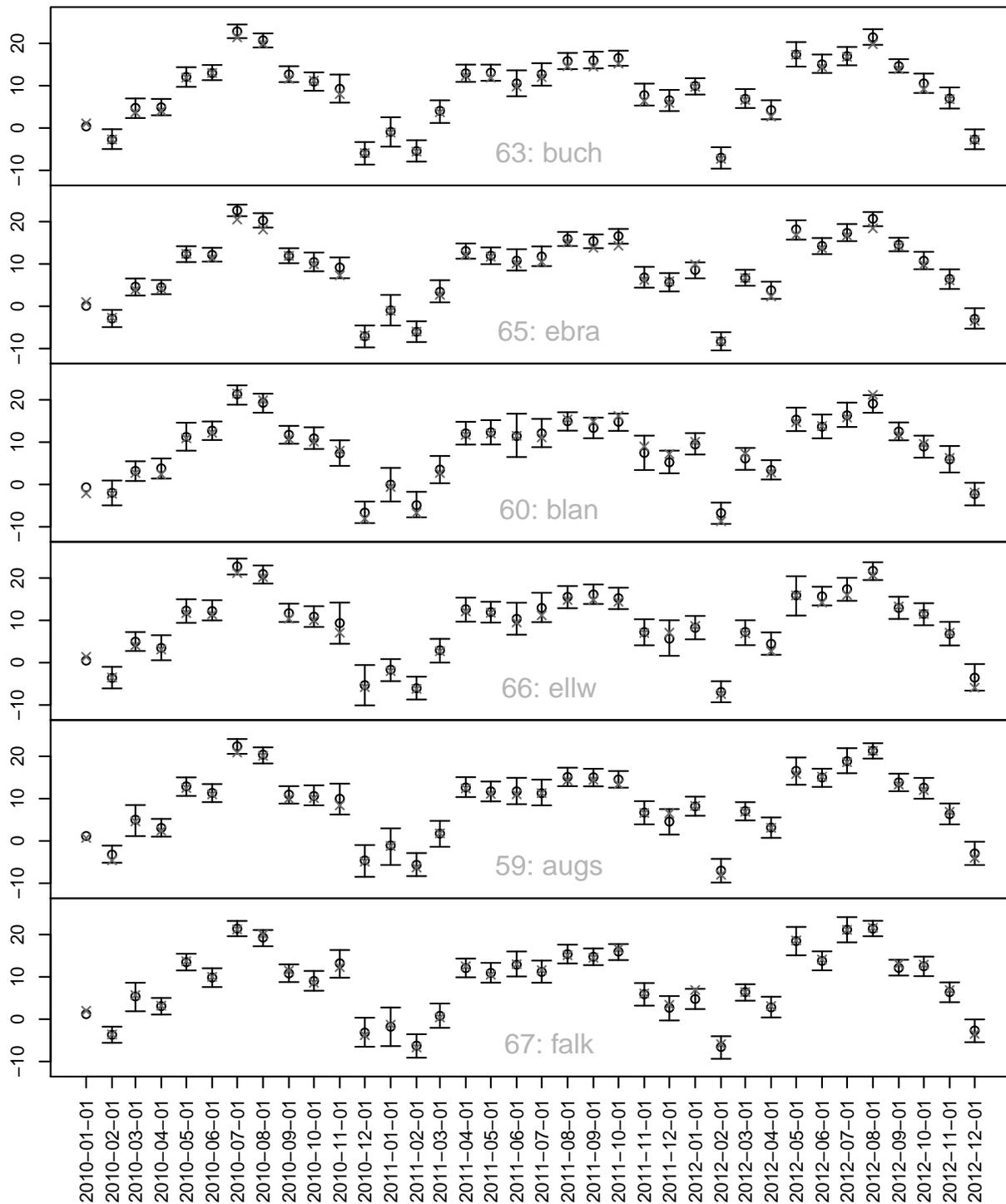
Figure A.1.11: 95% prediction intervals, point predictions and observed mean temperatures for the first days of each months during the years 2010-2012 for the observation stations 63, 65, 60, 66, 59 and 67, each calculated based on 1000 simulations from the predictive distribution of the mean temperatures at the respective observation station (spatial R-vine model).
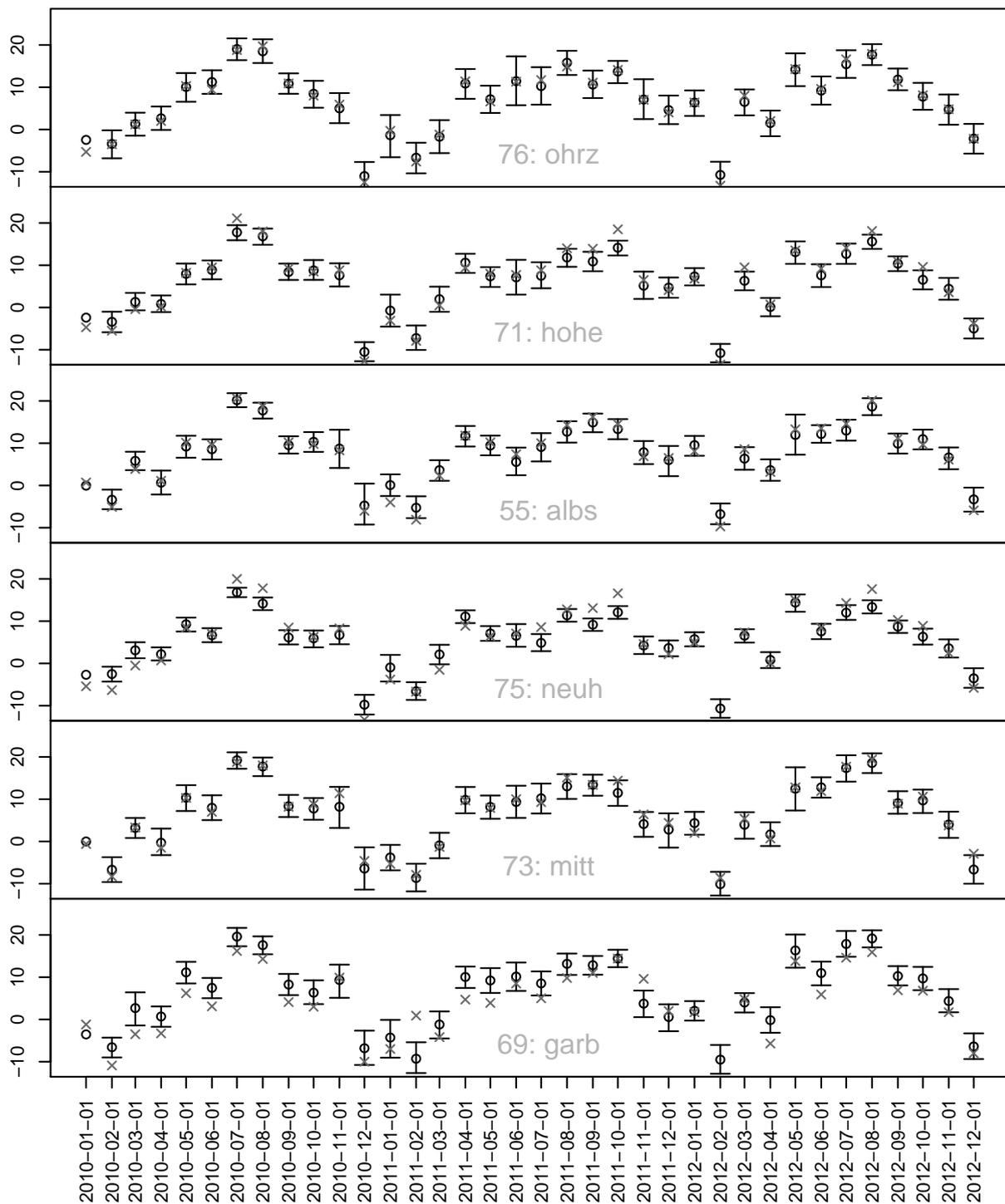
Figure A.1.12: 95% prediction intervals, point predictions and observed mean temperatures for the first days of each months during the years 2010-2012 for the observation stations 76, 71, 55, 75, 73 and 69, each calculated based on 1000 simulations from the predictive distribution of the mean temperatures at the respective observation station (spatial R-vine model).

A.2   Predictions: spatial composite vine model
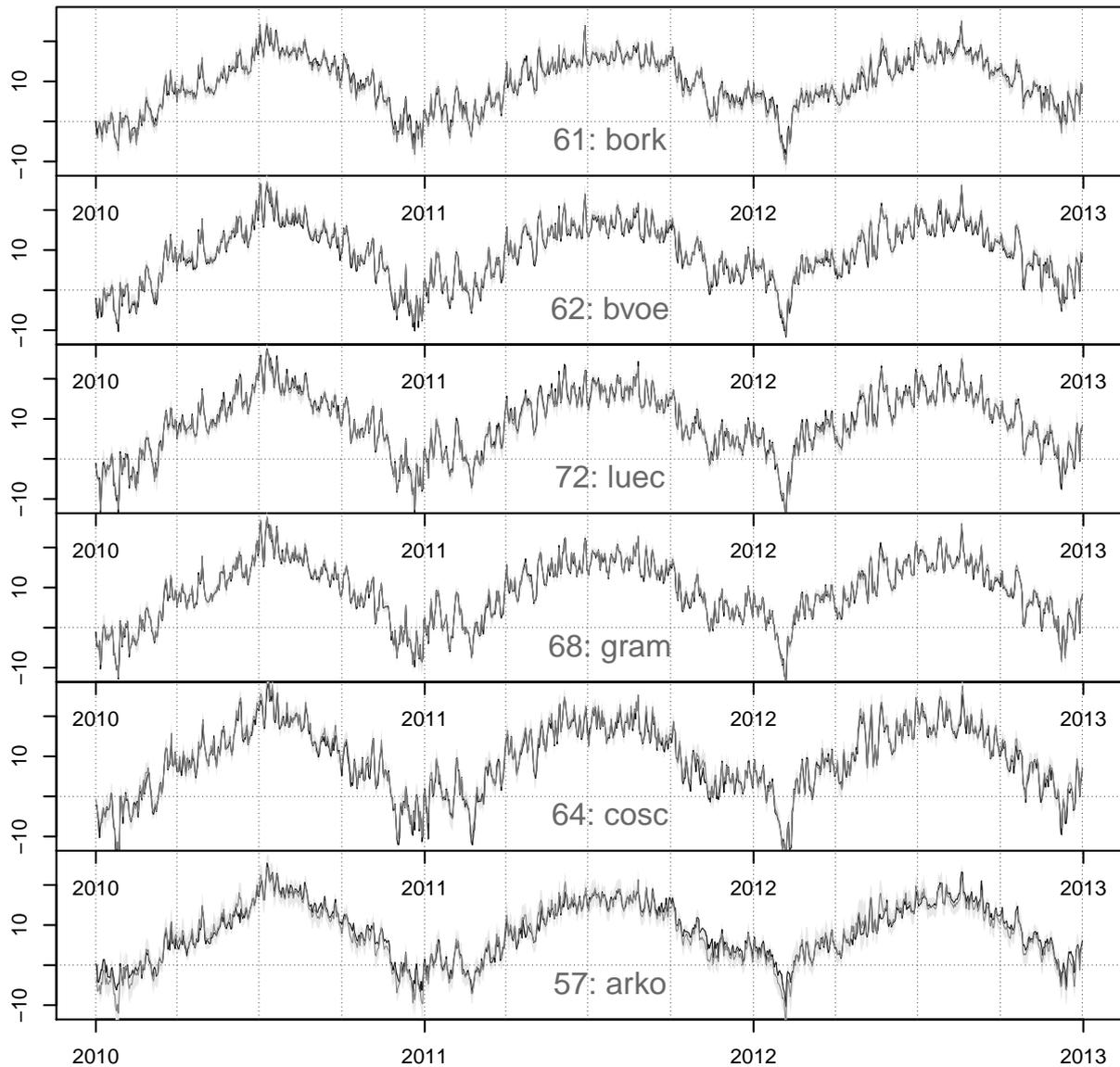
Figure A.2.1: Prediction of the mean temperatures for the observation stations 61, 62, 72, 68, 64 and 57 for the period 01/01/2010-12/31/2012 based on the spatial composite vine model. black line: observed values. dark gray line: prediction. light gray area: 95% prediction intervals.
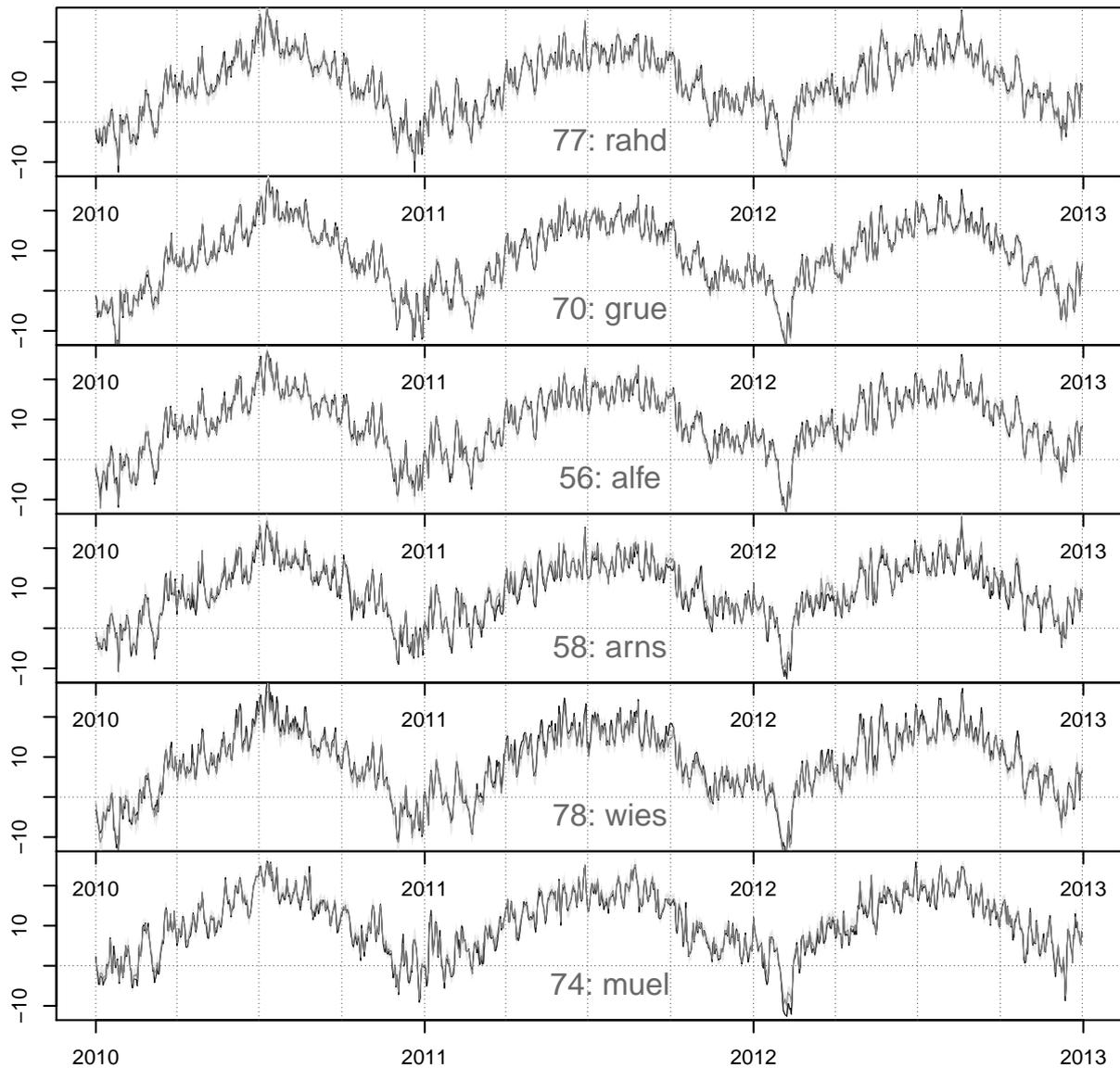
Figure A.2.2: Prediction of the mean temperatures for the observation stations 77, 70, 56, 58, 78 and 74 for the period 01/01/2010-12/31/2012 based on the spatial composite vine model. black line: observed values. dark gray line: prediction. light gray area: 95% prediction intervals.
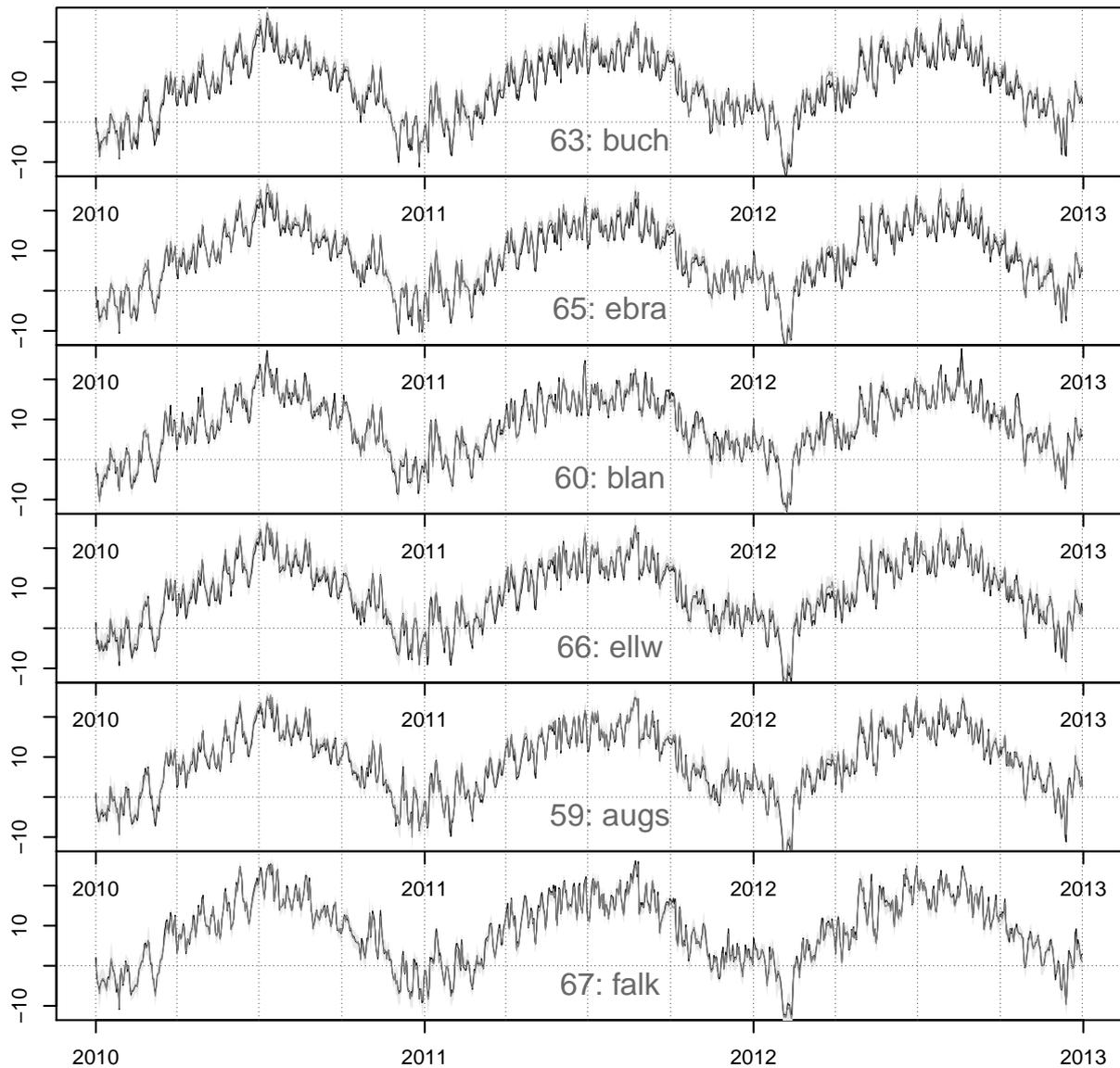
Figure A.2.3: Prediction of the mean temperatures for the observation stations 63, 65, 60, 66, 59 and 67 for the period 01/01/2010-12/31/2012 based on the spatial composite vine model. black line: observed values. dark gray line: prediction. light gray area: 95% prediction intervals.
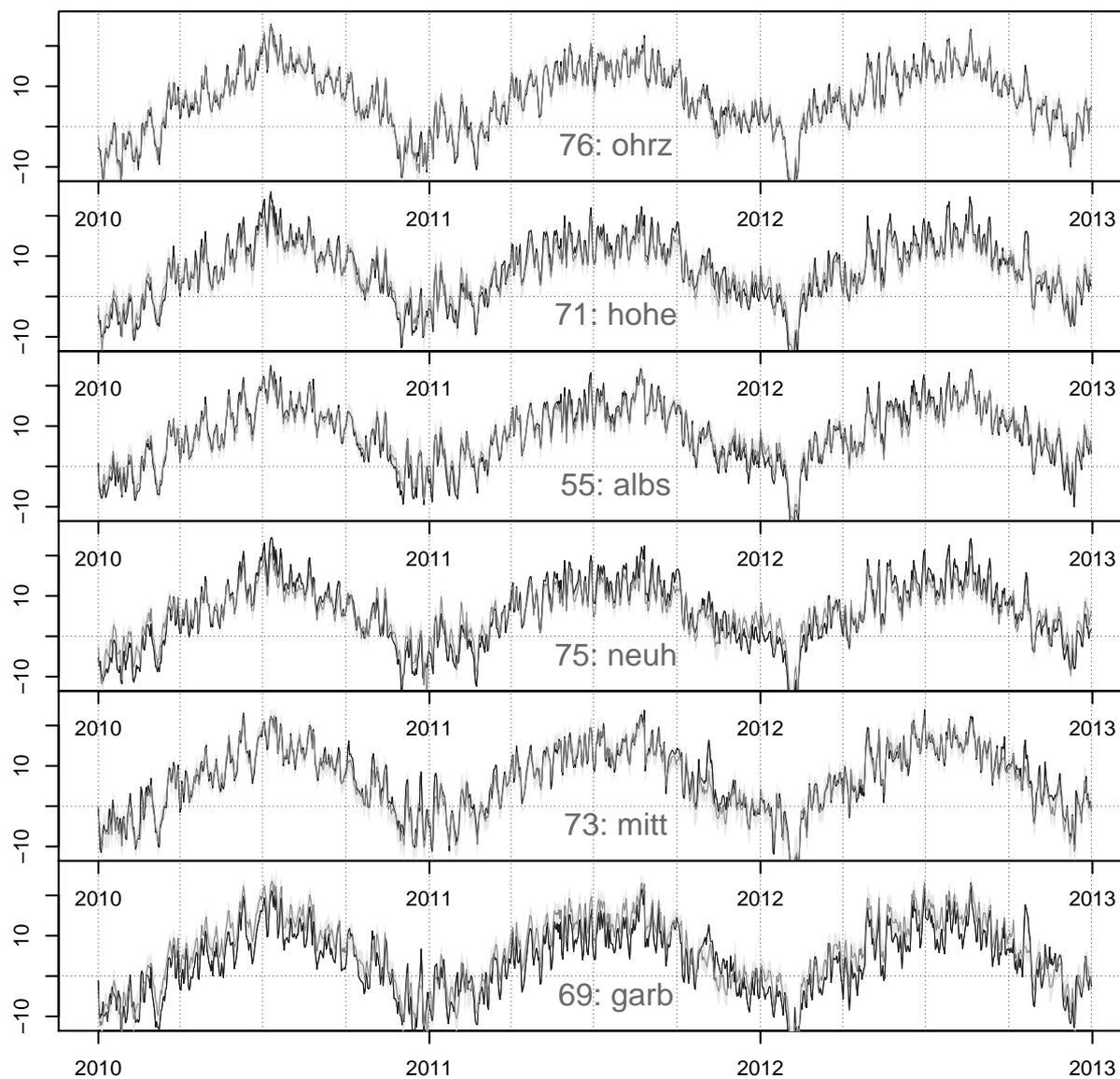
Figure A.2.4: Prediction of the mean temperatures for the observation stations 76, 71, 55, 75, 73 and 69 for the period 01/01/2010-12/31/2012 based on the spatial composite vine model. black line: observed values. dark gray line: prediction. light gray area: 95% prediction intervals.
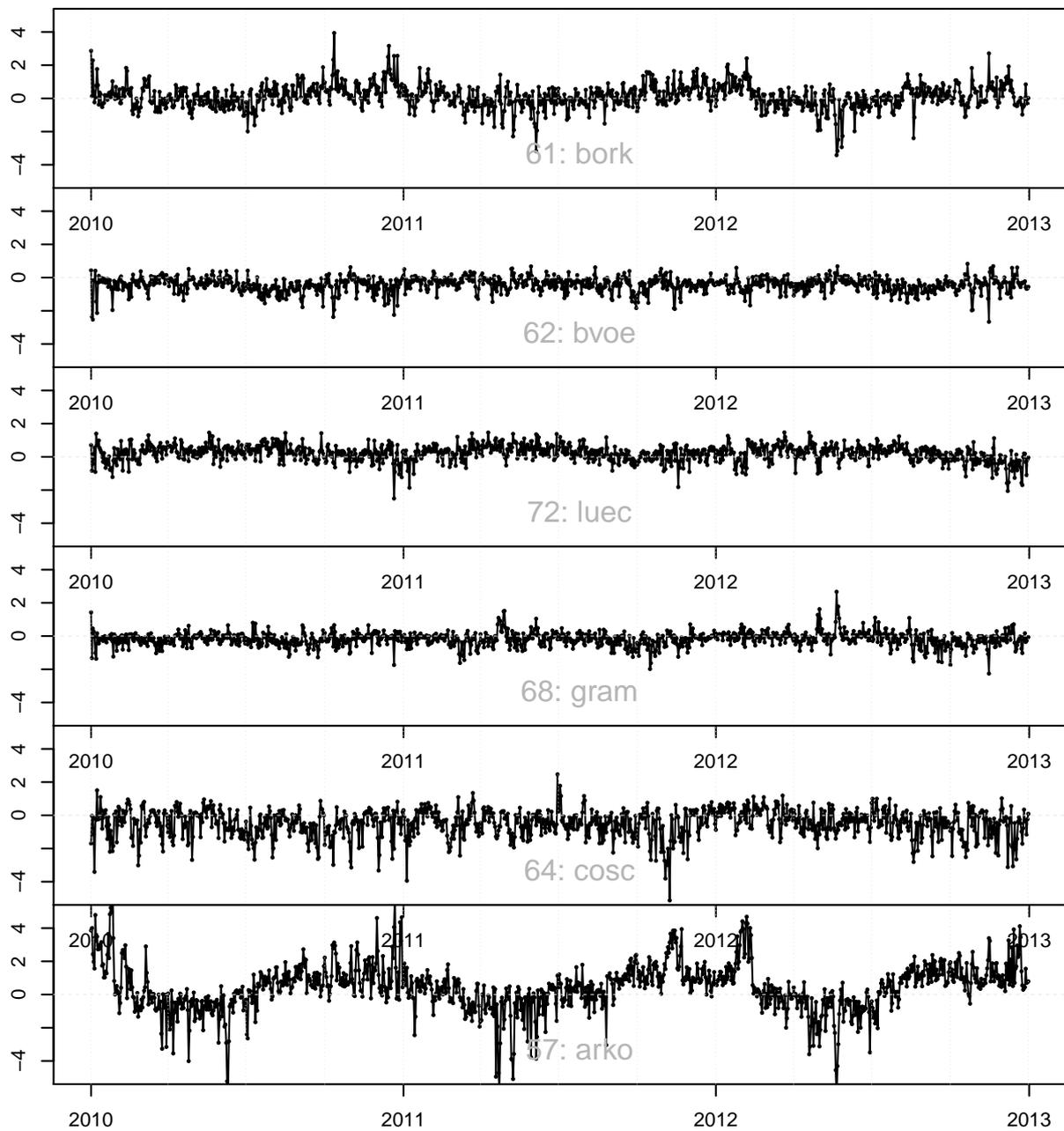
Figure A.2.5: Prediction errors of the predictions for the observation stations 61, 62, 72, 68, 64 and 57 for the period 01/01/2010-12/31/2012 based on the spatial composite vine model.
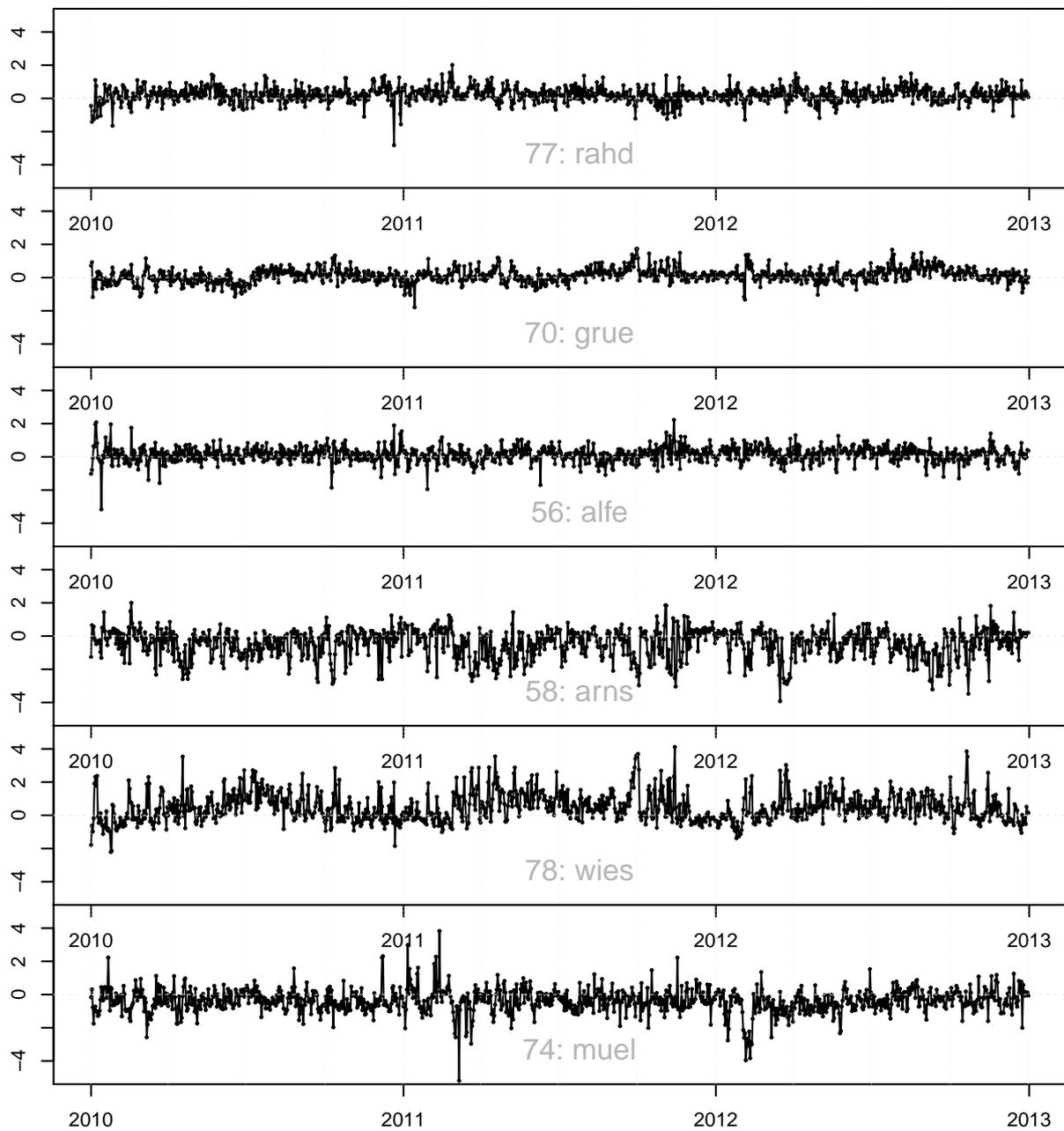
Figure A.2.6: Prediction errors of the predictions for the observation stations 77, 70, 56, 58, 78 and 74 for the period 01/01/2010-12/31/2012 based on the spatial composite vine model.
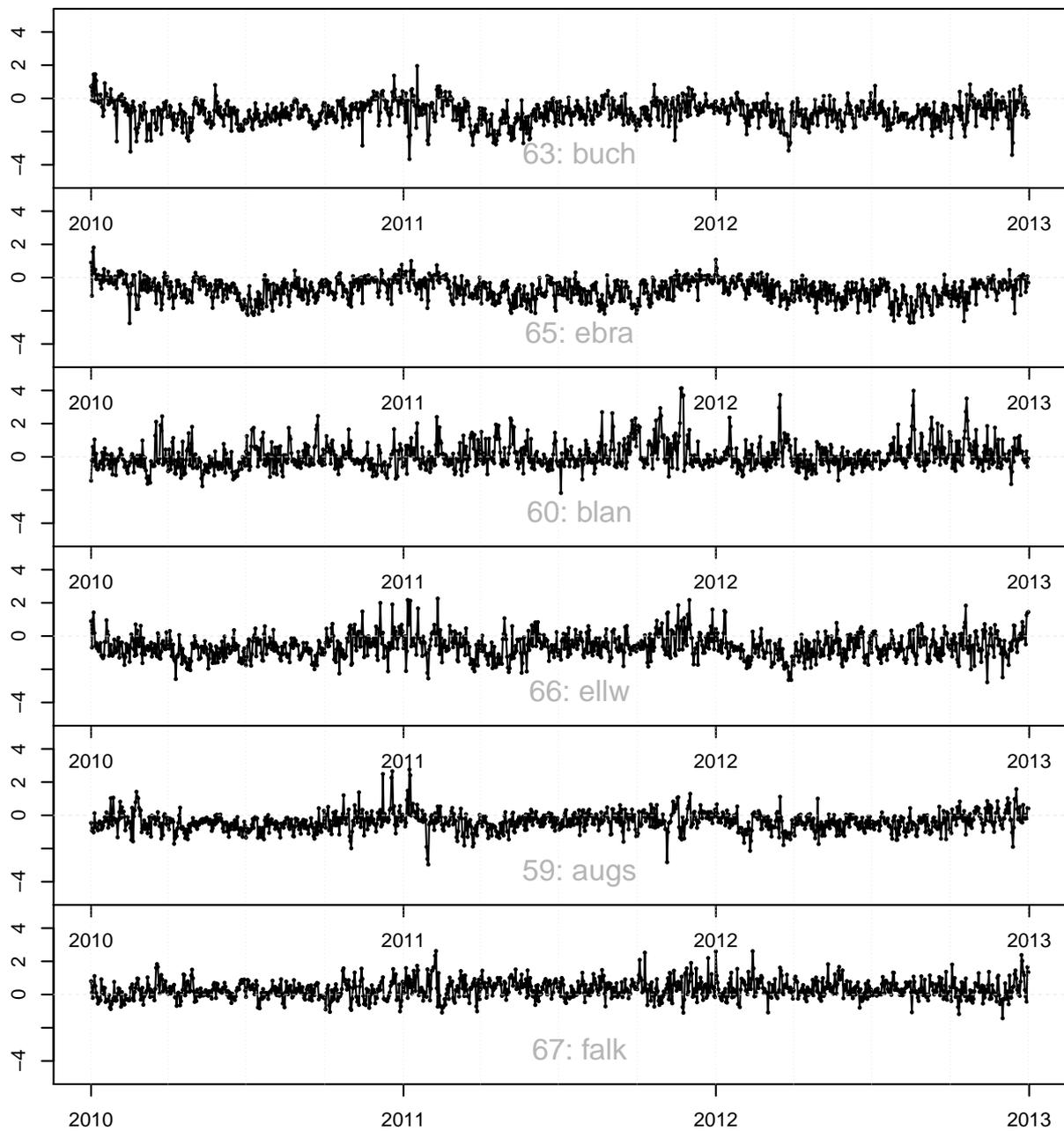
Figure A.2.7: Prediction errors of the predictions for the observation stations 63, 65, 60, 66, 59 and 67 for the period 01/01/2010-12/31/2012 based on the spatial composite vine model.
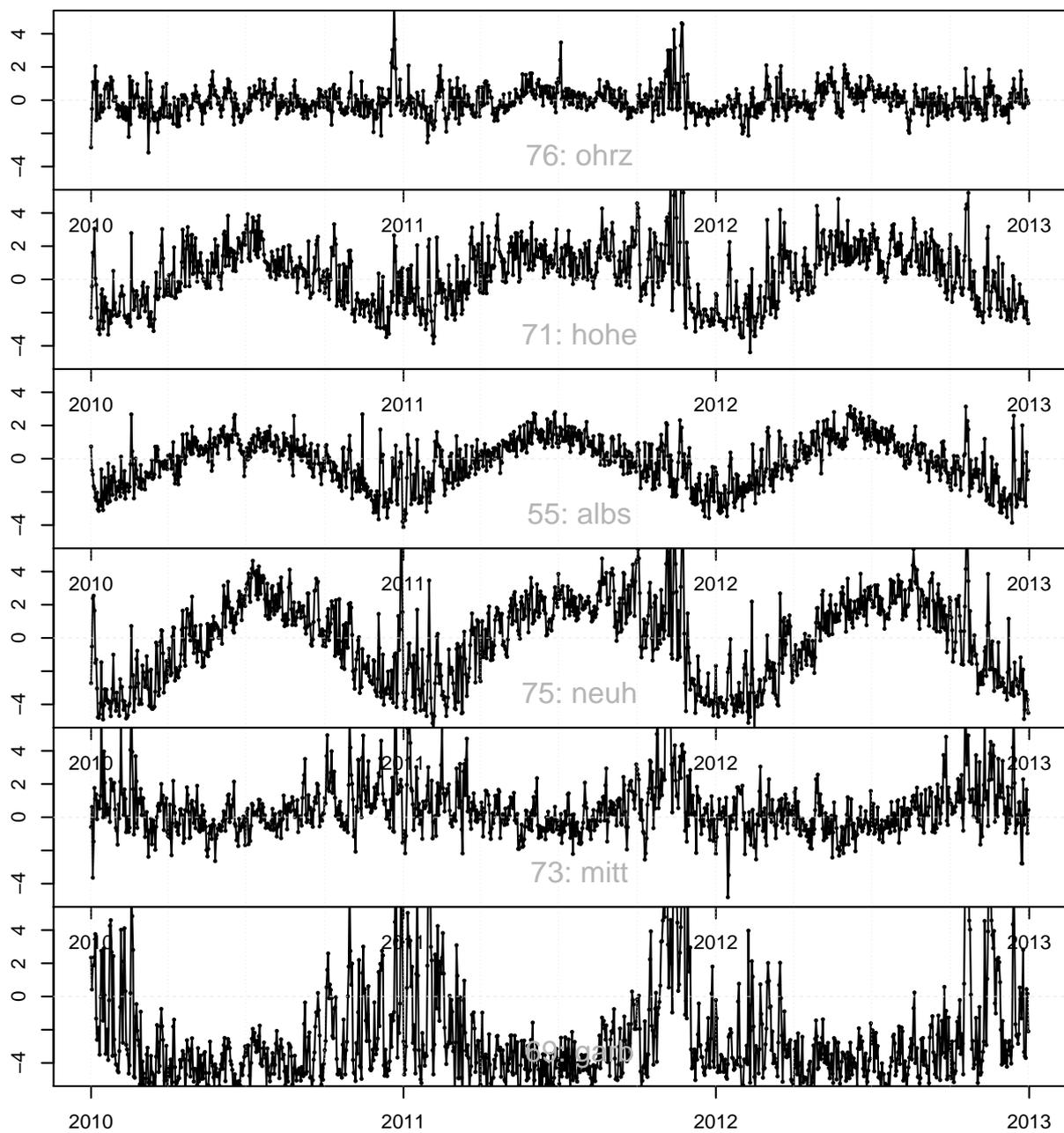
Figure A.2.8: Prediction errors of the predictions for the observation stations 76, 71, 55, 75, 73 and 69 for the period 01/01/2010-12/31/2012 based on the spatial composite vine model.
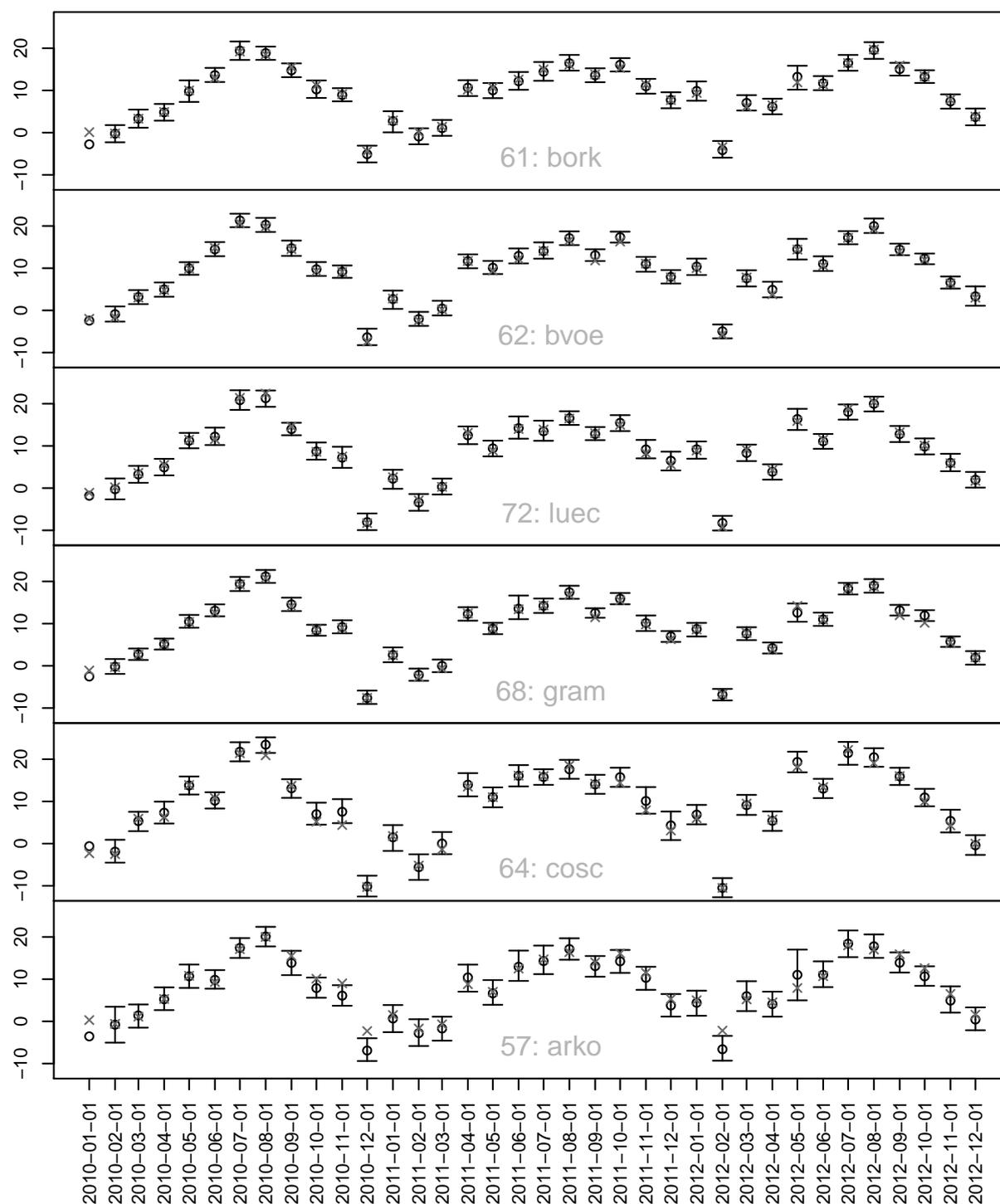
Figure A.2.9: 95% prediction intervals, point predictions and observed mean temperatures for the first days of each months during the years 2010-2012 for the observation stations 61, 62, 72, 68, 64 and 57, each calculated based on 1000 simulations from the predictive distribution of the mean temperatures at the respective observation station (spatial composite vine model).
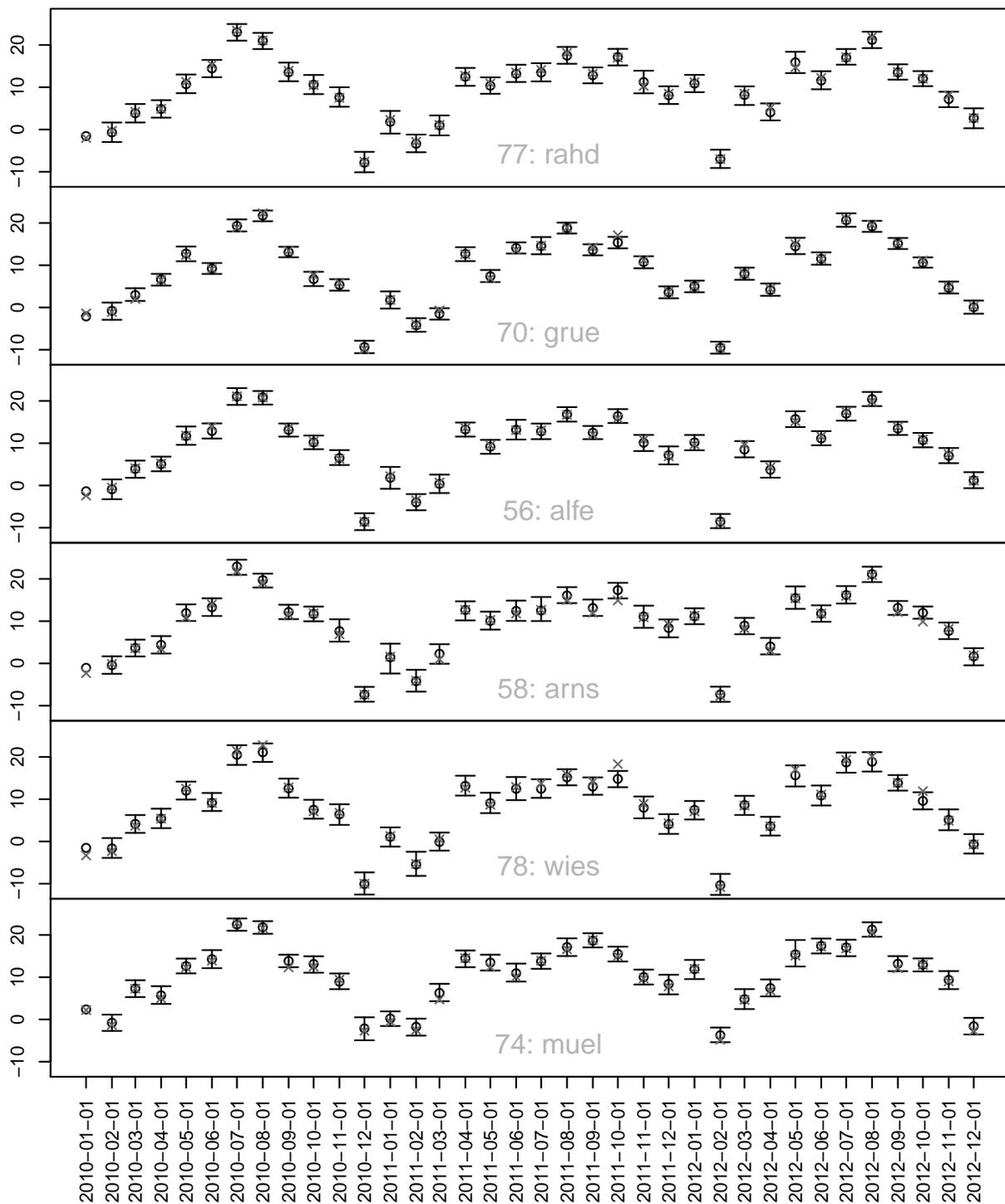
Figure A.2.10: 95% prediction intervals, point predictions and observed mean temperatures for the first days of each months during the years 2010-2012 for the observation stations 77, 70, 56, 58, 78 and 74, each calculated based on 1000 simulations from the predictive distribution of the mean temperatures at the respective observation station (spatial composite vine model).
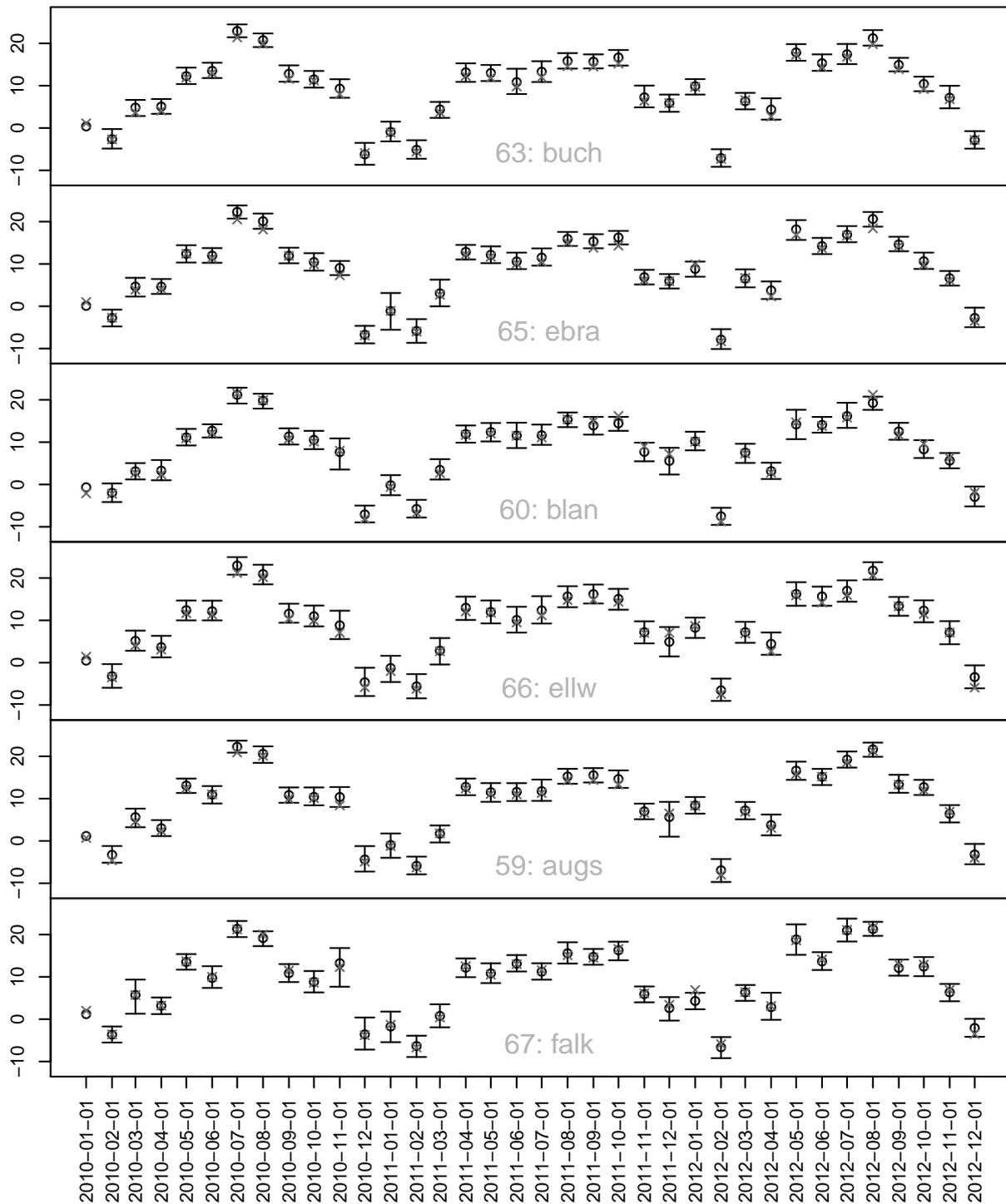
Figure A.2.11: 95% prediction intervals, point predictions and observed mean temperatures for the first days of each months during the years 2010-2012 for the observation stations 63, 65, 60, 66, 59 and 67, each calculated based on 1000 simulations from the predictive distribution of the mean temperatures at the respective observation station (spatial composite vine model).
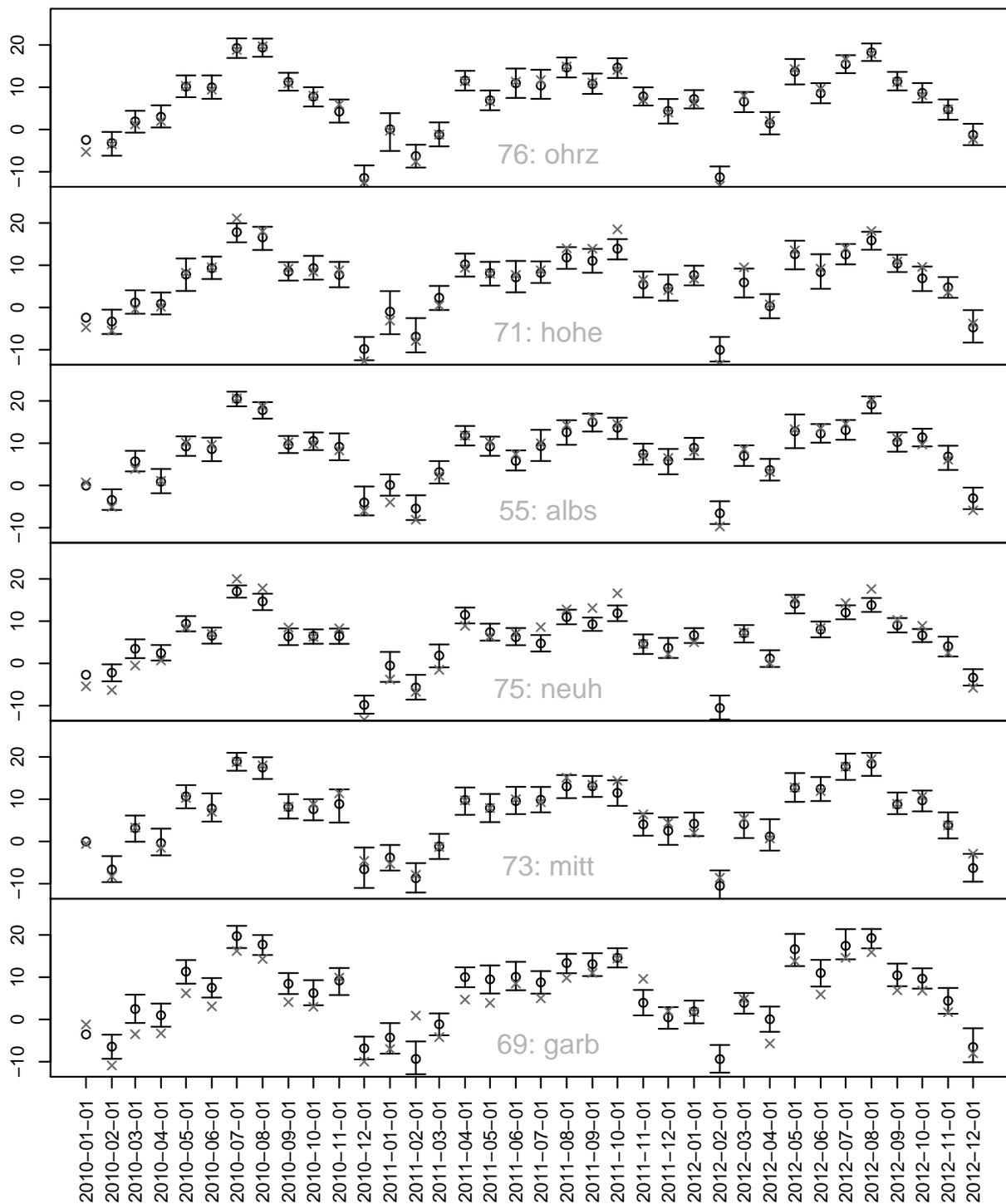
Figure A.2.12: 95% prediction intervals, point predictions and observed mean temperatures for the first days of each months during the years 2010-2012 for the observation stations 76, 71, 55, 75, 73 and 69, each calculated based on 1000 simulations from the predictive distribution of the mean temperatures at the respective observation station (spatial composite vine model).

A.3   Predictions: spatial Gaussian model
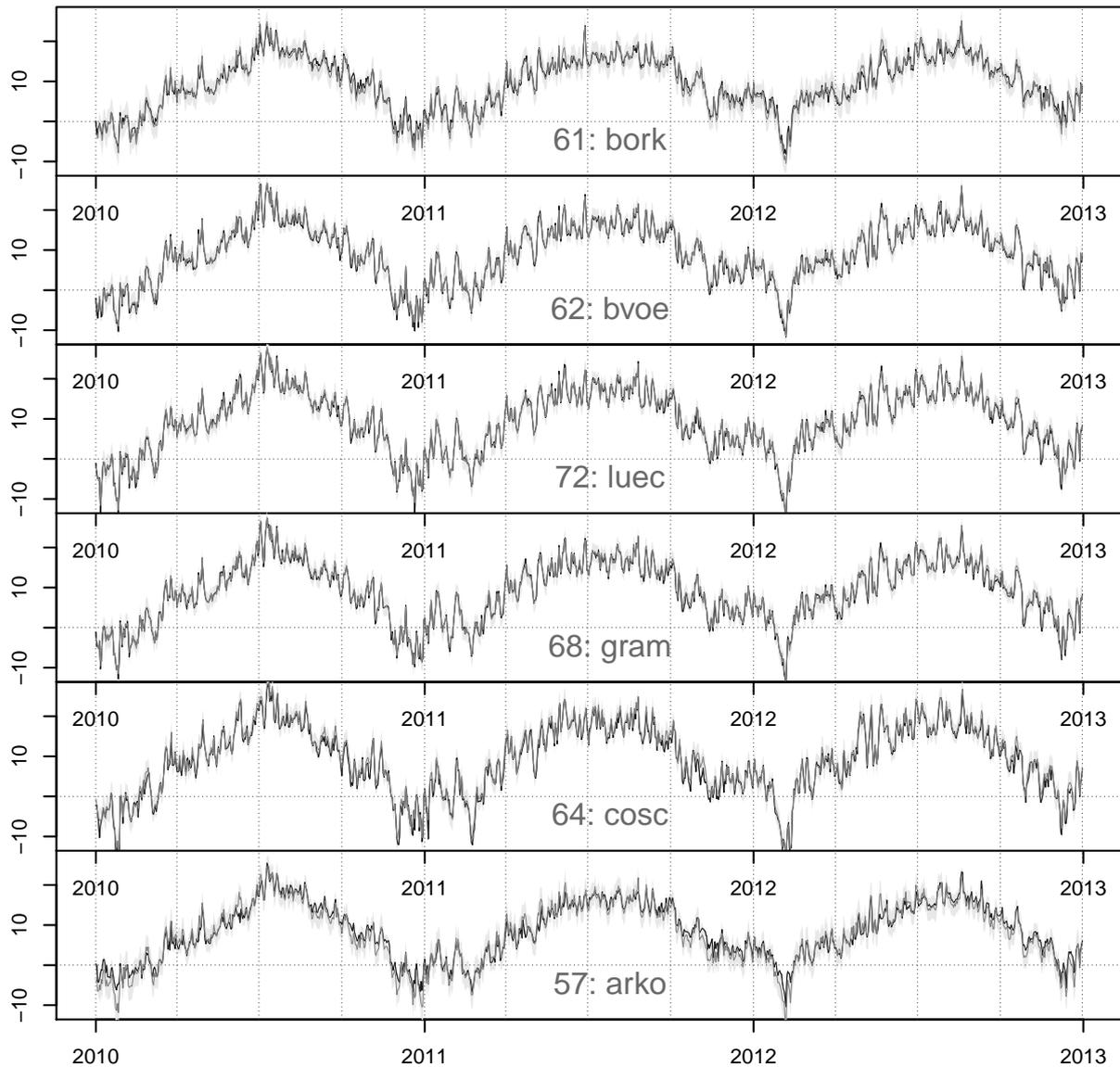
Figure A.3.1: Prediction of the mean temperatures for the observation stations 61, 62, 72, 68, 64 and 57 for the period 01/01/2010-12/31/2012 based on the spatial Gaussian model. black line: observed values. dark gray line: prediction. light gray area: 95% prediction intervals.
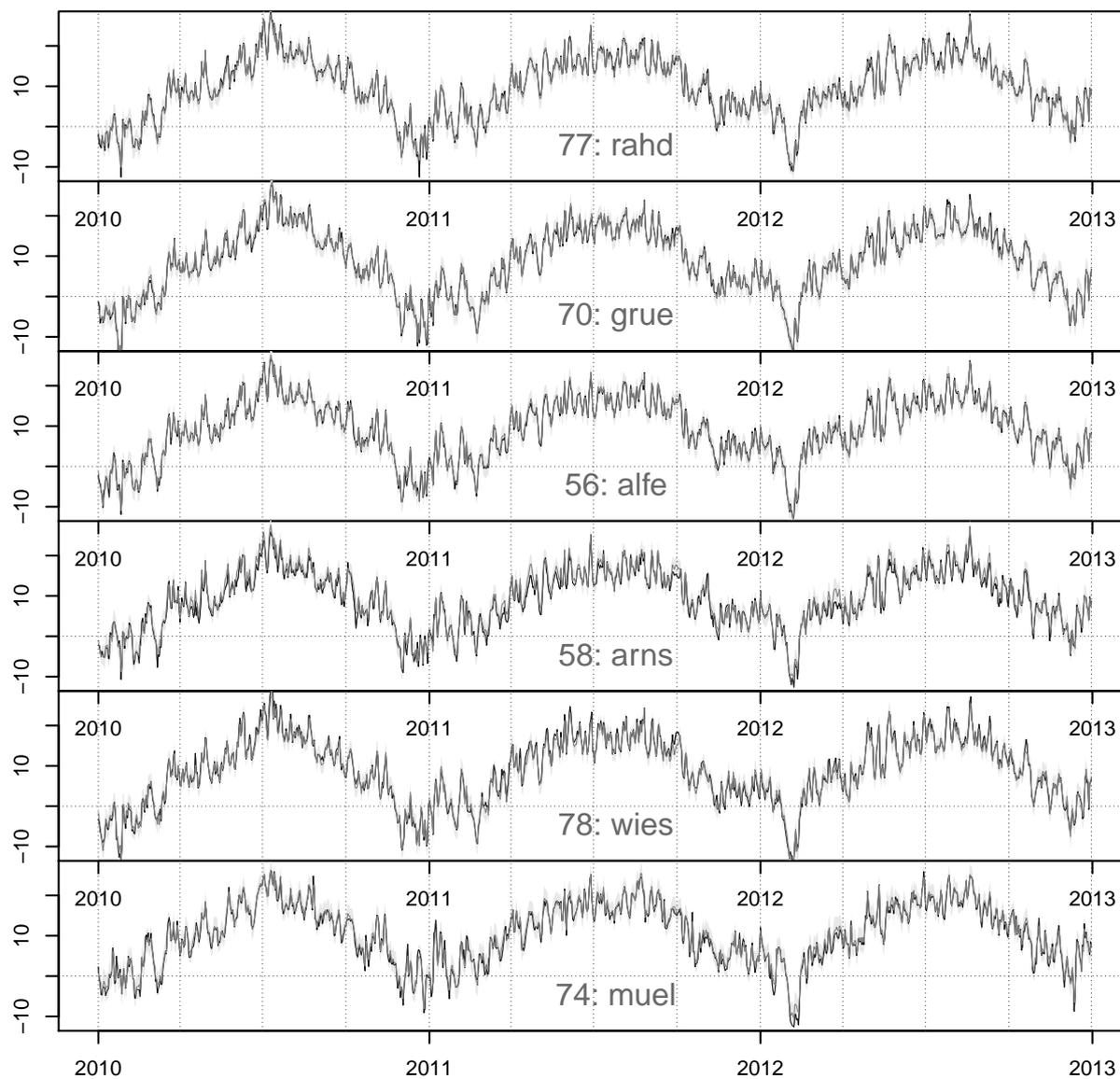
Figure A.3.2: Prediction of the mean temperatures for the observation stations 77, 70, 56, 58, 78 and 74 for the period 01/01/2010-12/31/2012 based on the spatial Gaussian model. black line: observed values. dark gray line: prediction. light gray area: 95% prediction intervals.
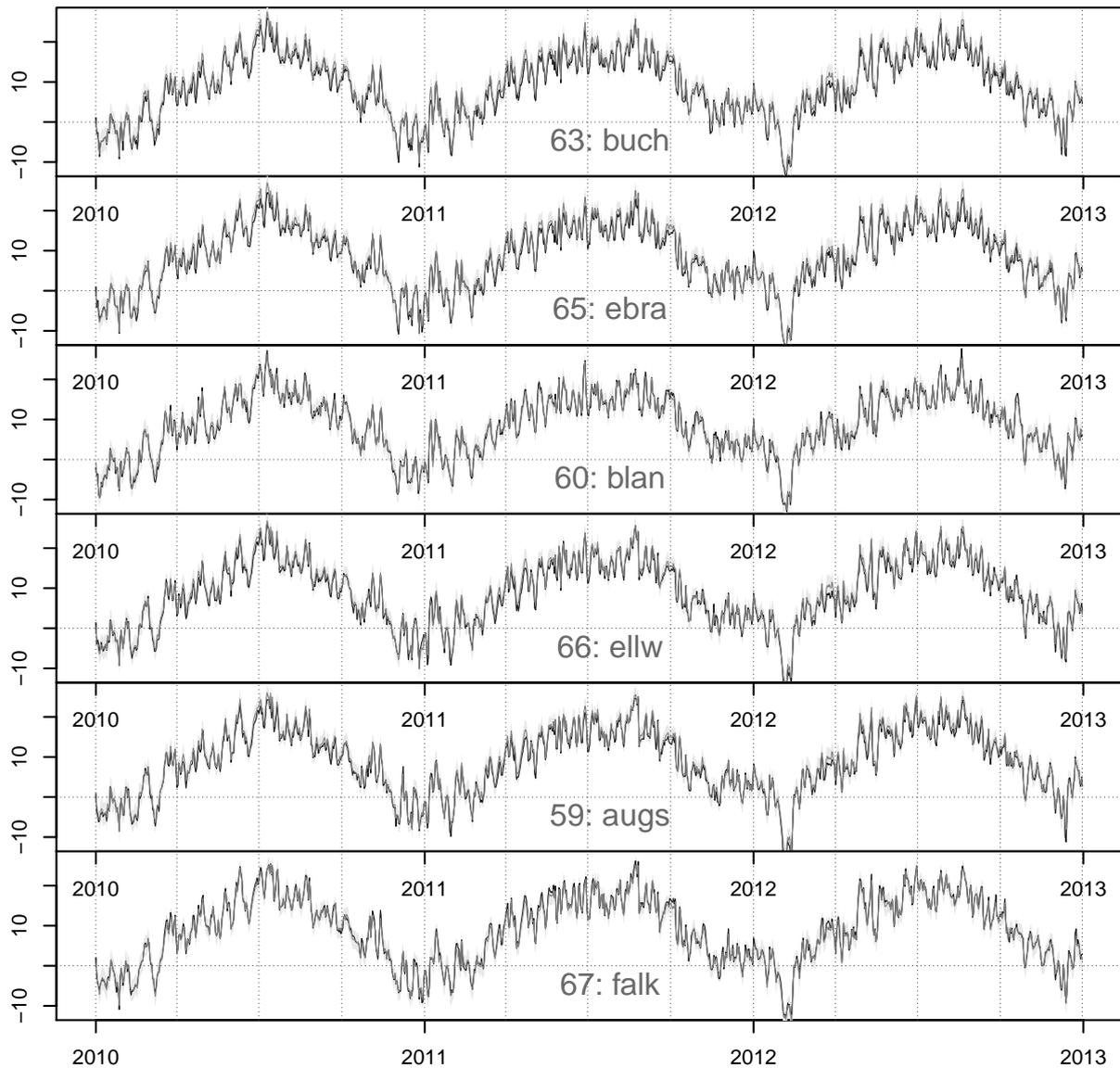
Figure A.3.3: Prediction of the mean temperatures for the observation stations 63, 65, 60, 66, 59 and 67 for the period 01/01/2010-12/31/2012 based on the spatial Gaussian model. black line: observed values. dark gray line: prediction. light gray area: 95% prediction intervals.
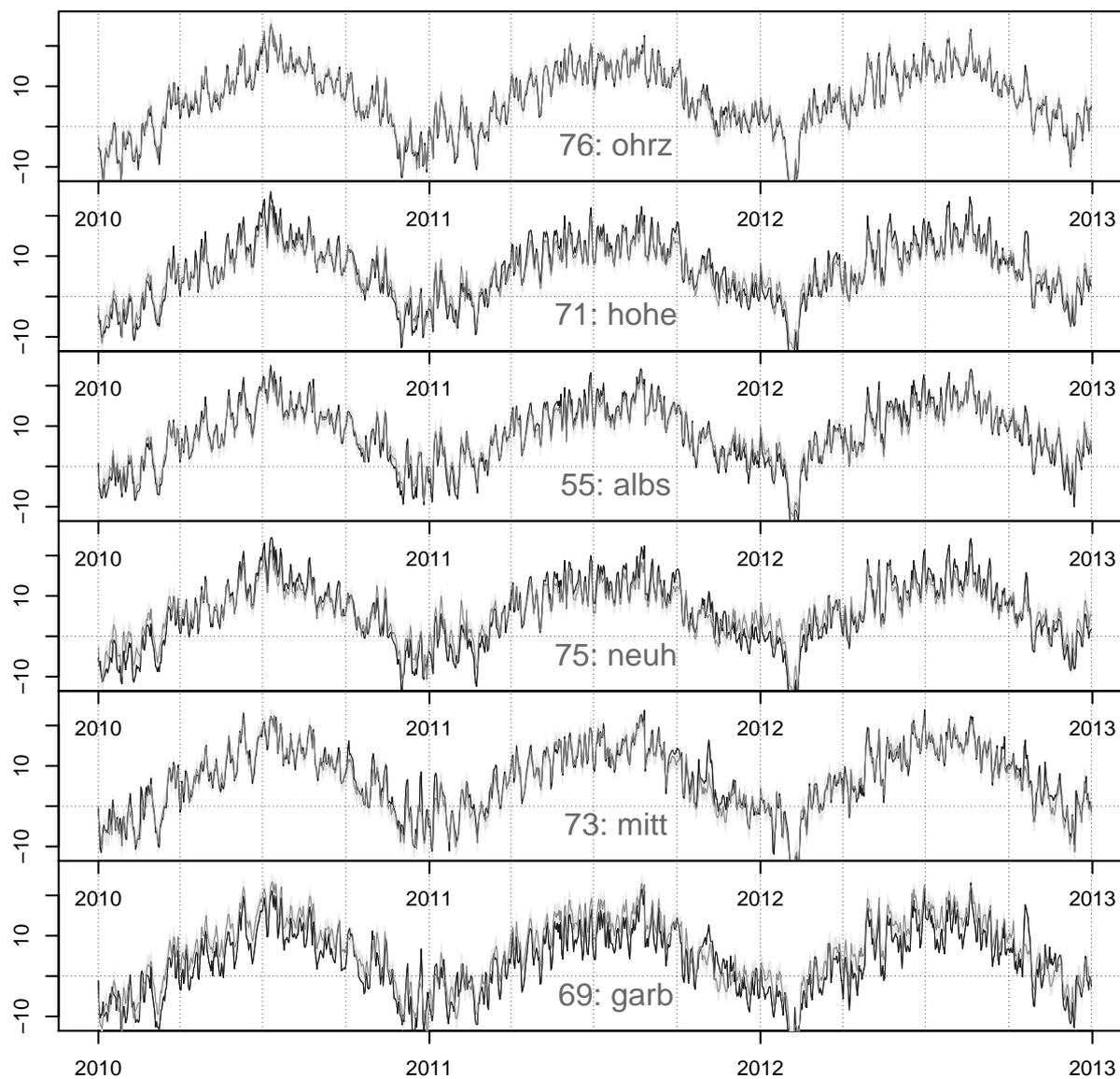
Figure A.3.4: Prediction of the mean temperatures for the observation stations 76, 71, 55, 75, 73 and 69 for the period 01/01/2010-12/31/2012 based on the spatial Gaussian model. black line: observed values. dark gray line: prediction. light gray area: 95% prediction intervals.
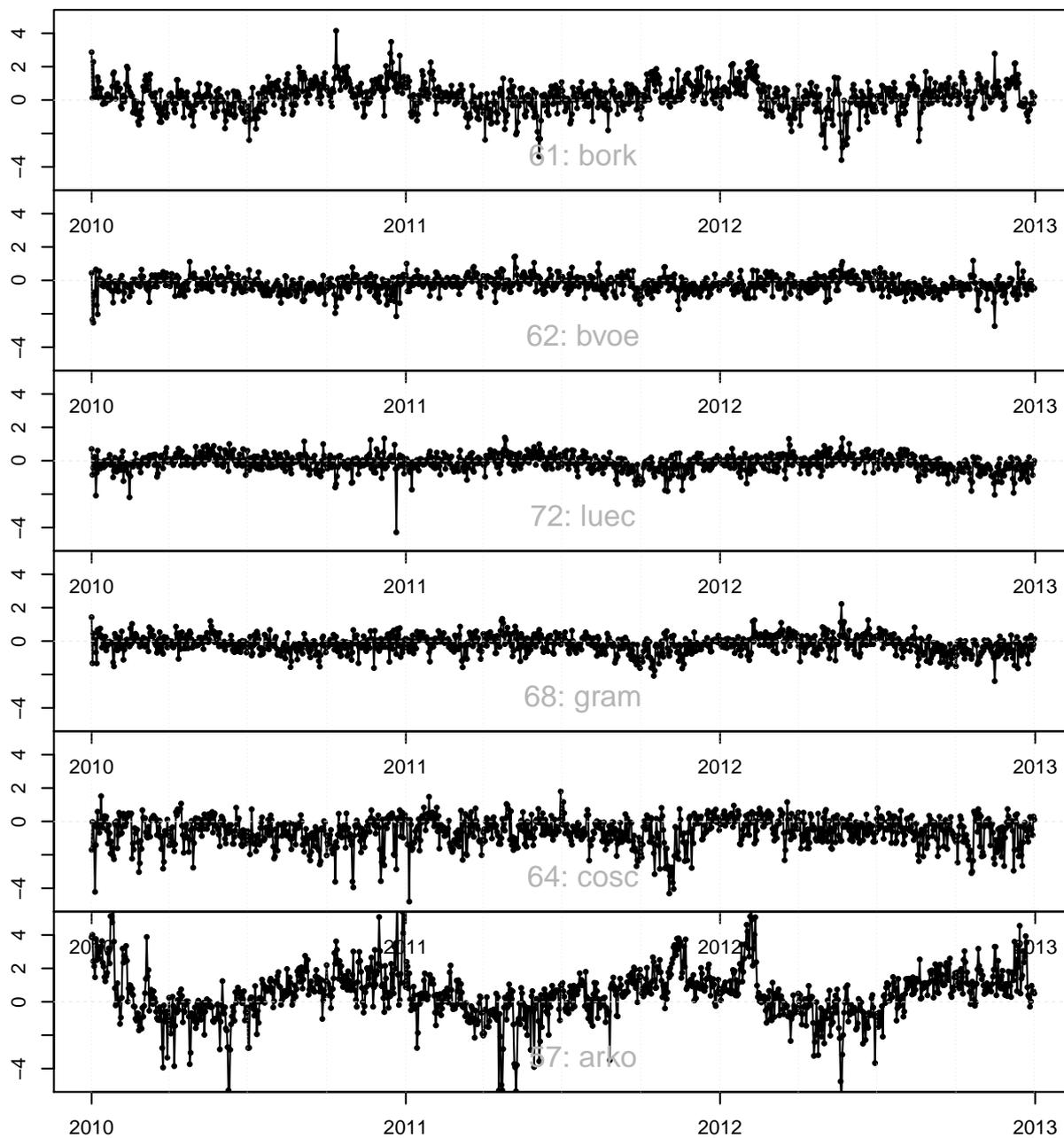
Figure A.3.5: Prediction errors of the predictions for the observation stations 61, 62, 72, 68, 64 and 57 for the period 01/01/2010-12/31/2012 based on the spatial Gaussian model.
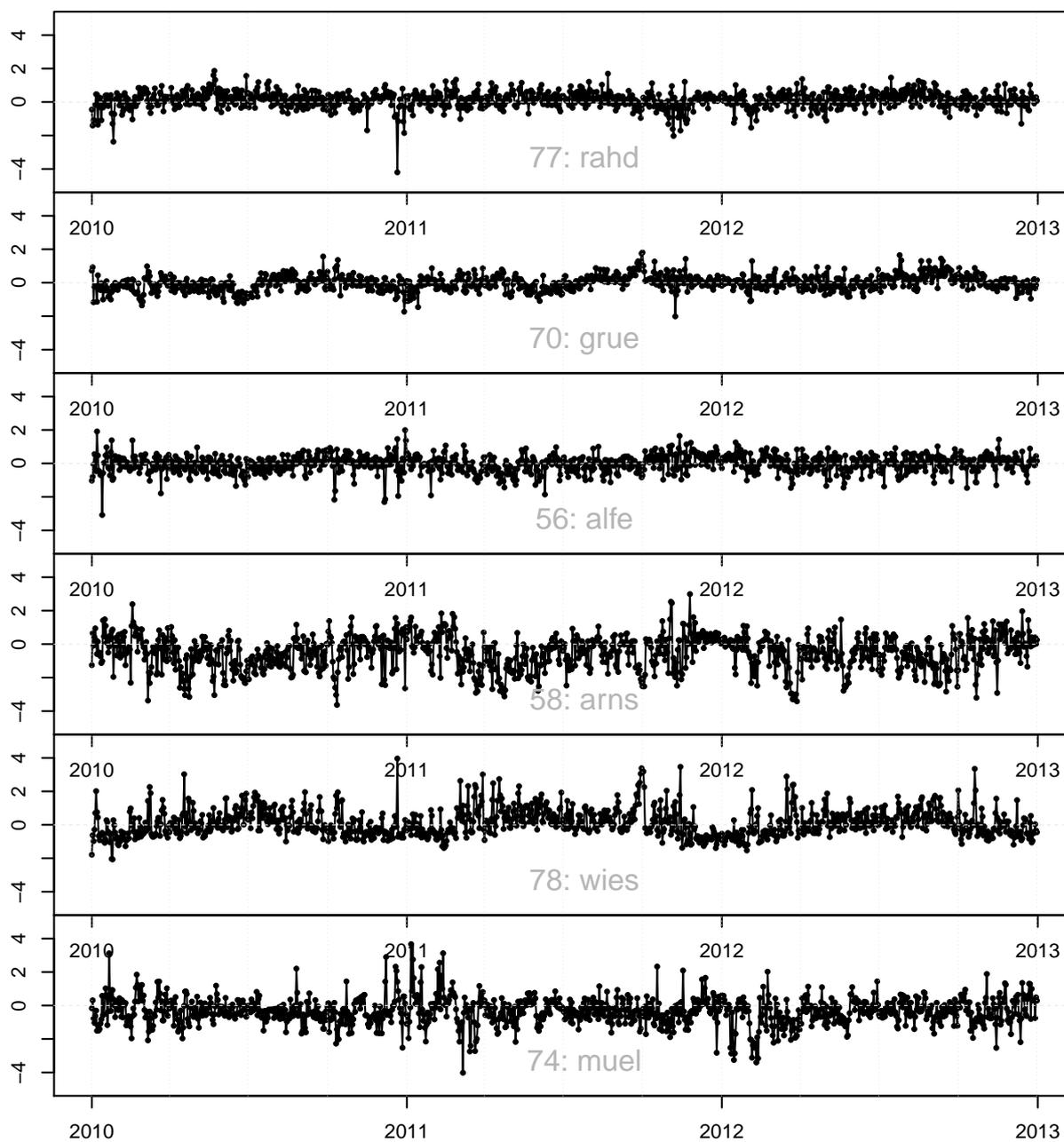
Figure A.3.6: Prediction errors of the predictions for the observation stations 77, 70, 56, 58, 78 and 74 for the period 01/01/2010-12/31/2012 based on the spatial Gaussian model.
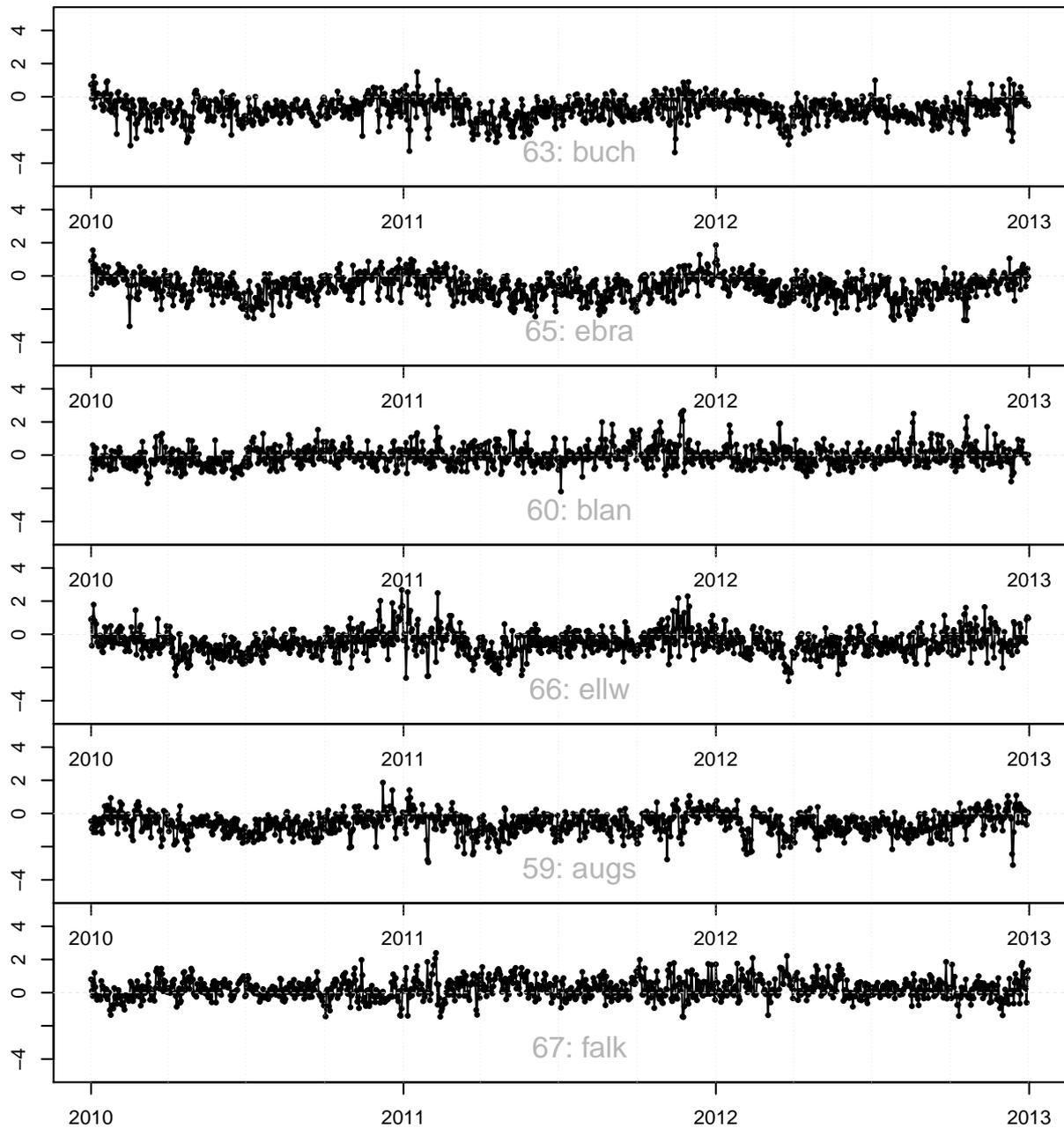
Figure A.3.7: Prediction errors of the predictions for the observation stations 63, 65, 60, 66, 59 and 67 for the period 01/01/2010-12/31/2012 based on the spatial Gaussian model.
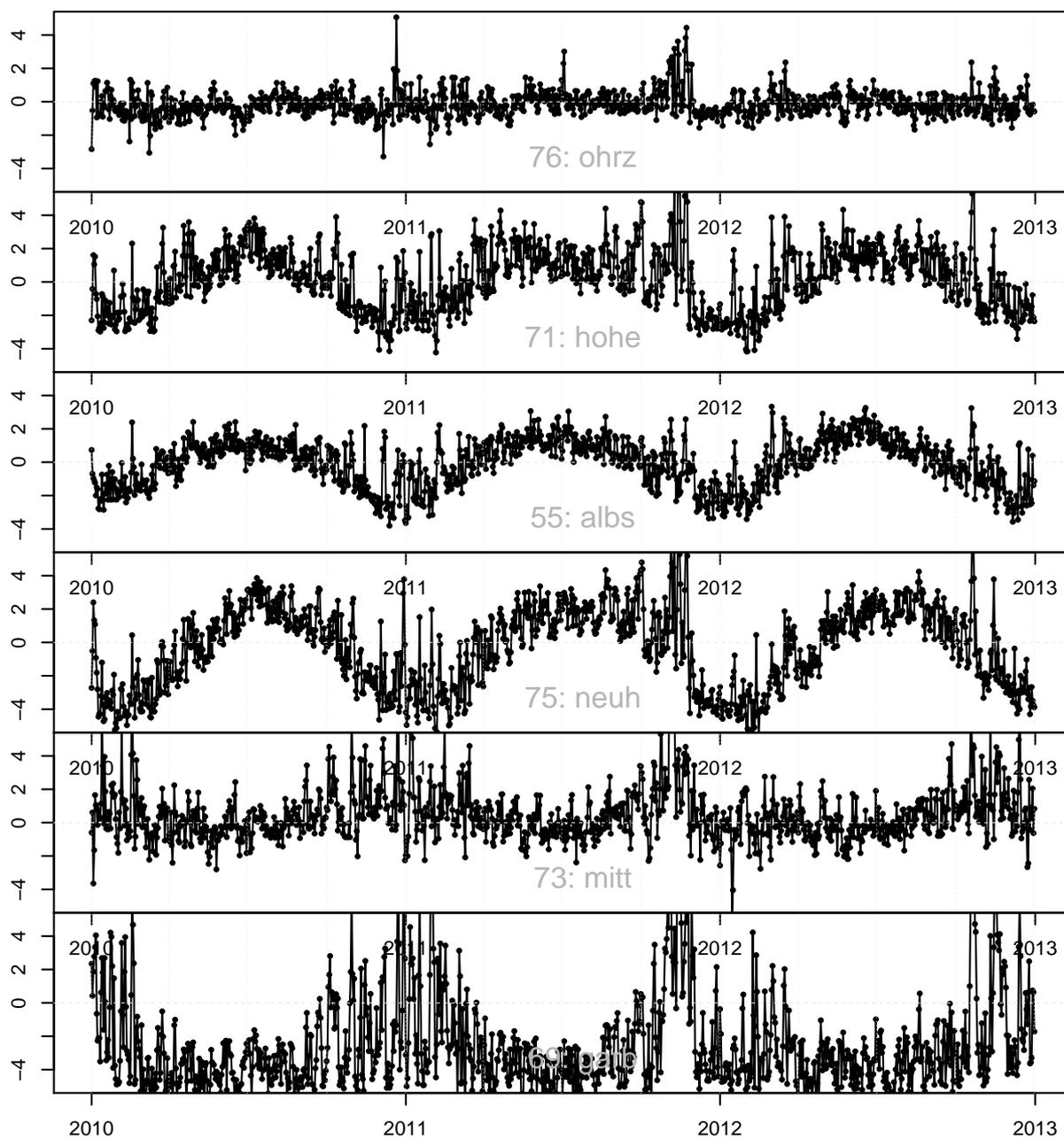
Figure A.3.8: Prediction errors of the predictions for the observation stations 76, 71, 55, 75, 73 and 69 for the period 01/01/2010-12/31/2012 based on the spatial Gaussian model.
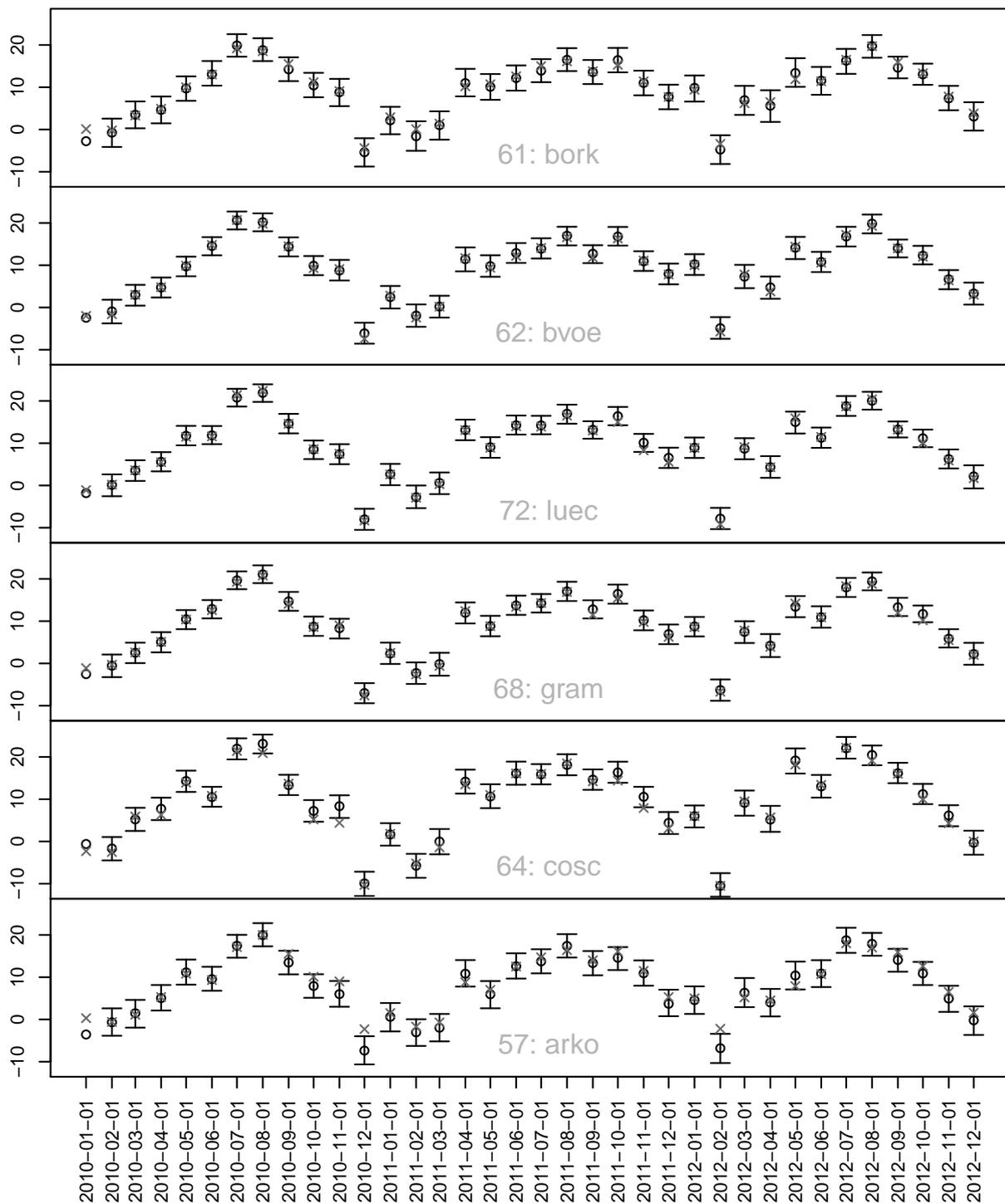
Figure A.3.9: 95% prediction intervals, point predictions and observed mean temperatures for the first days of each months during the years 2010-2012 for the observation stations 61, 62, 72, 68, 64 and 57, each calculated based on 1000 simulations from the predictive distribution of the mean temperatures at the respective observation station (spatial Gaussian model).
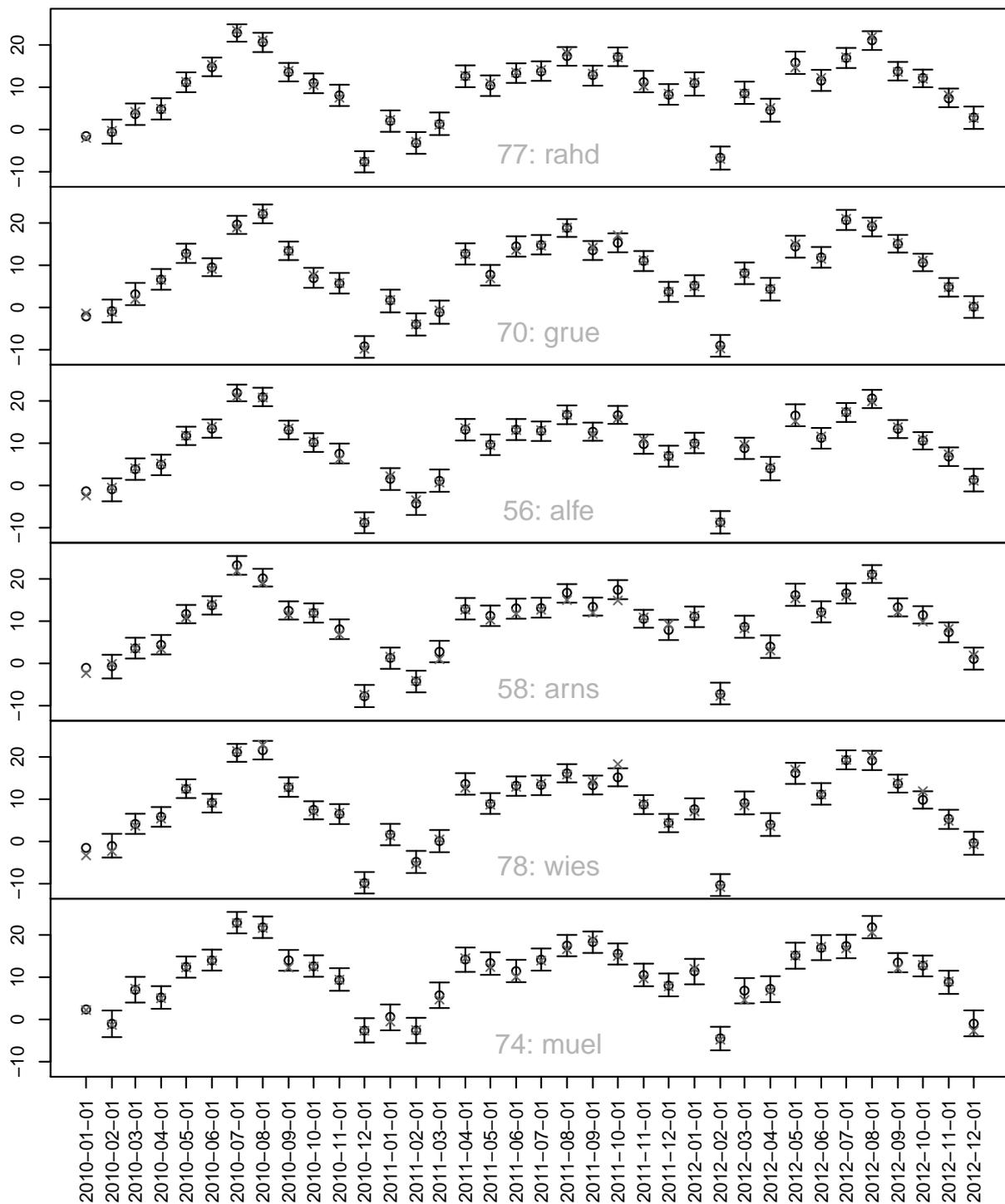
Figure A.3.10: 95% prediction intervals, point predictions and observed mean temperatures for the first days of each months during the years 2010-2012 for the observation stations 77, 70, 56, 58, 78 and 74, each calculated based on 1000 simulations from the predictive distribution of the mean temperatures at the respective observation station (spatial Gaussian model).
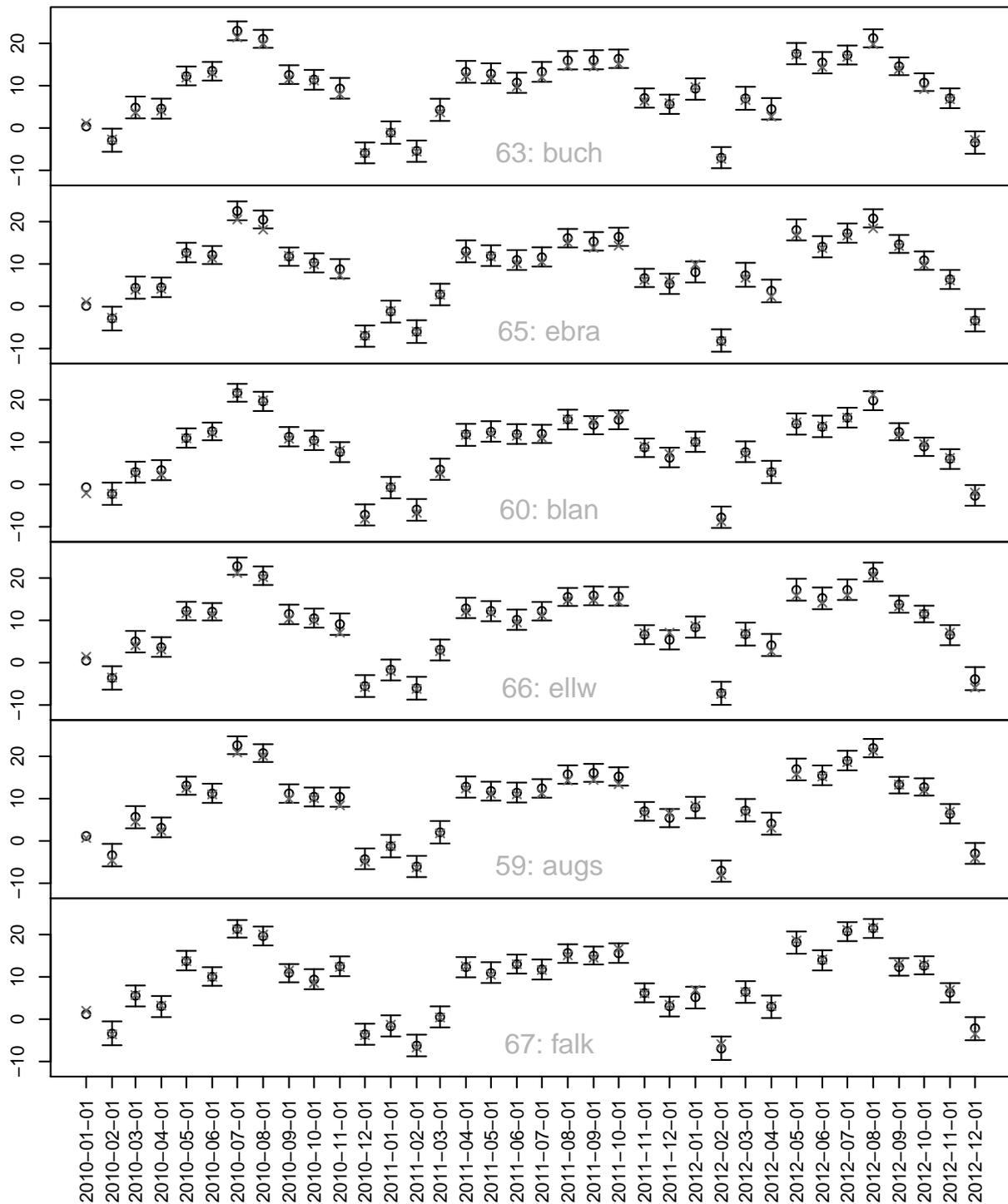
Figure A.3.11: 95% prediction intervals, point predictions and observed mean temperatures for the first days of each months during the years 2010-2012 for the observation stations 63, 65, 60, 66, 59 and 67, each calculated based on 1000 simulations from the predictive distribution of the mean temperatures at the respective observation station (spatial Gaussian model).
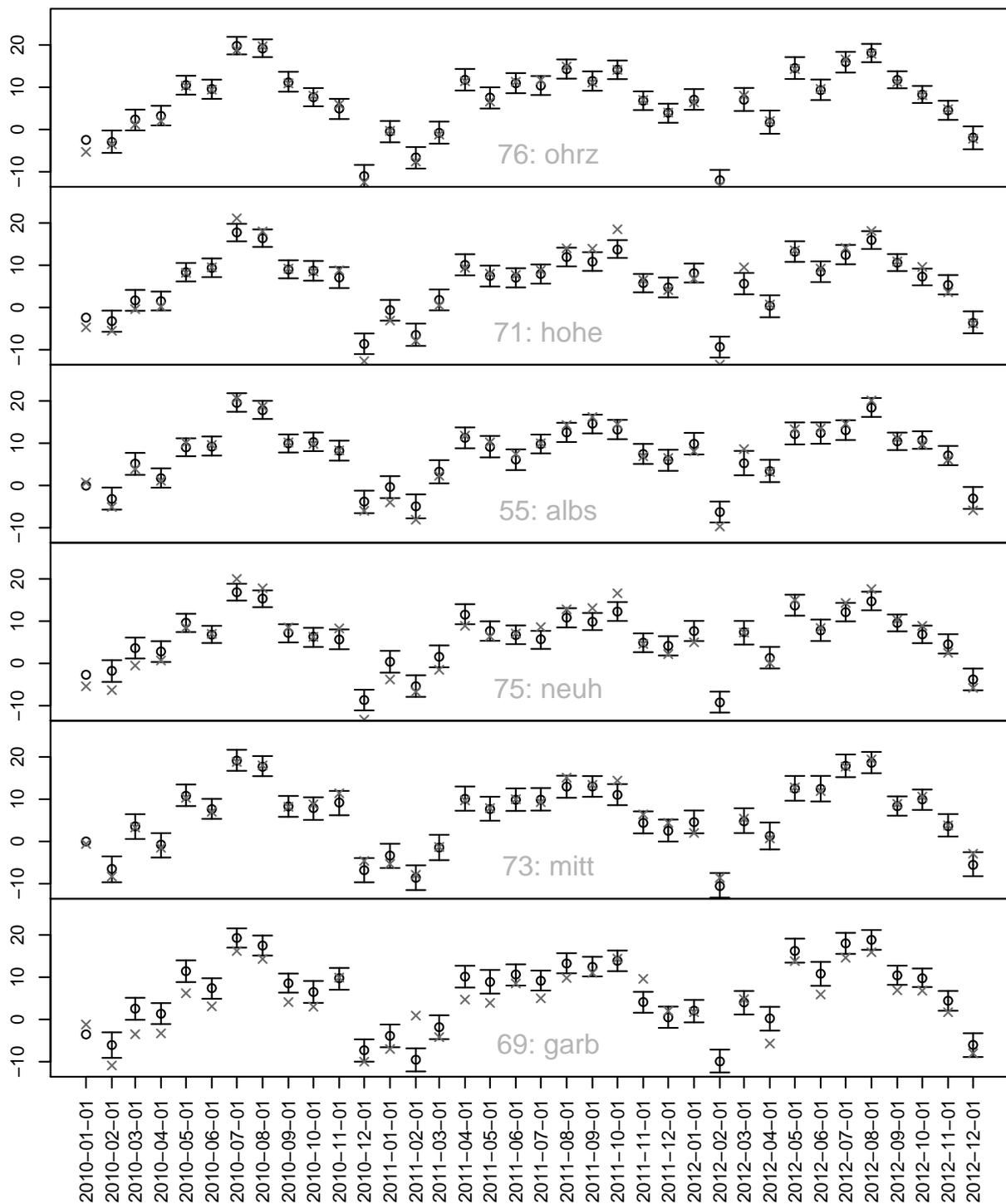
Figure A.3.12: 95% prediction intervals, point predictions and observed mean temperatures for the first days of each months during the years 2010-2012 for the observation stations 76, 71, 55, 75, 73 and 69, each calculated based on 1000 simulations from the predictive distribution of the mean temperatures at the respective observation station (spatial Gaussian model).

A.4   Predictions for the 20th of January 2010
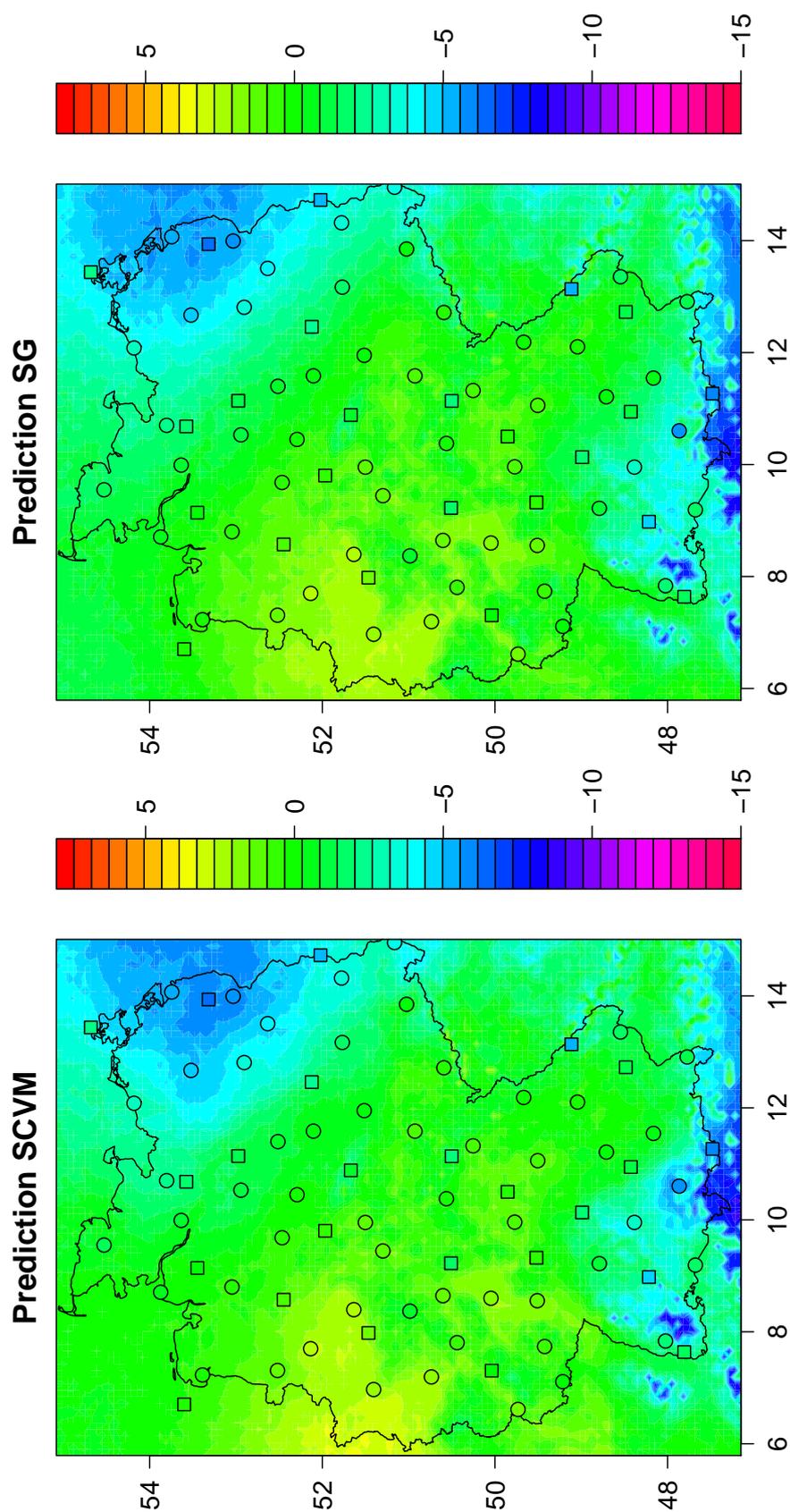
Figure A.4.1: Level plots comparing predictions based on the spatial composite vine model and the spatial Gaussian model on a 80 × 120 grid covering Germany on the 20th of January 2010. Whereas the circles represent the mean temperatures at the 54 observation stations of the training data set, the squares depict the mean temperatures at the 24 stations of the validation data set.
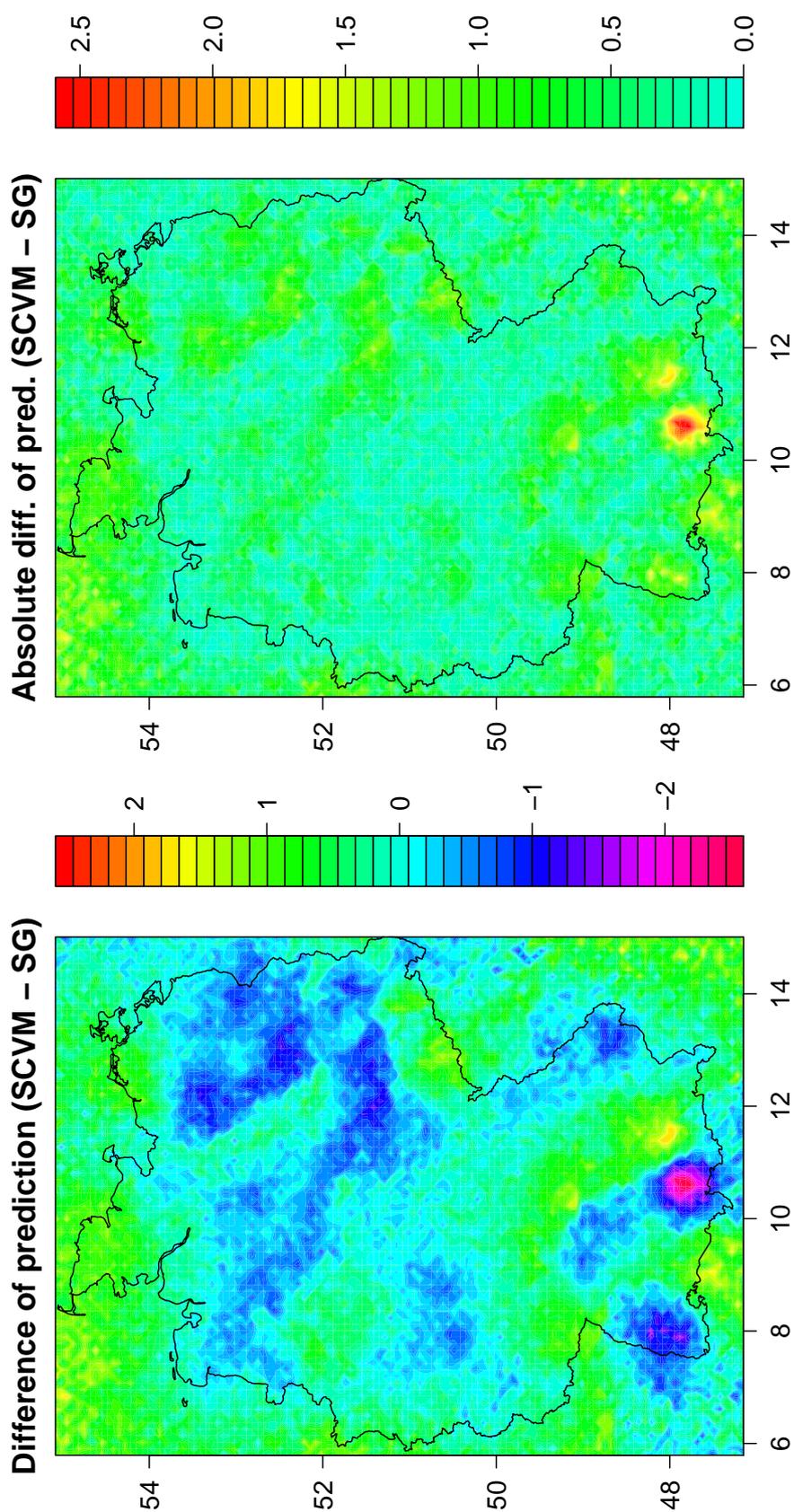
Figure A.4.2: Level plots of (the absolute value of) the difference between the predictions based on the spatial composite vine model and the spatial Gaussian model on a 80×120 grid covering Germany on the 20th of January 2010.
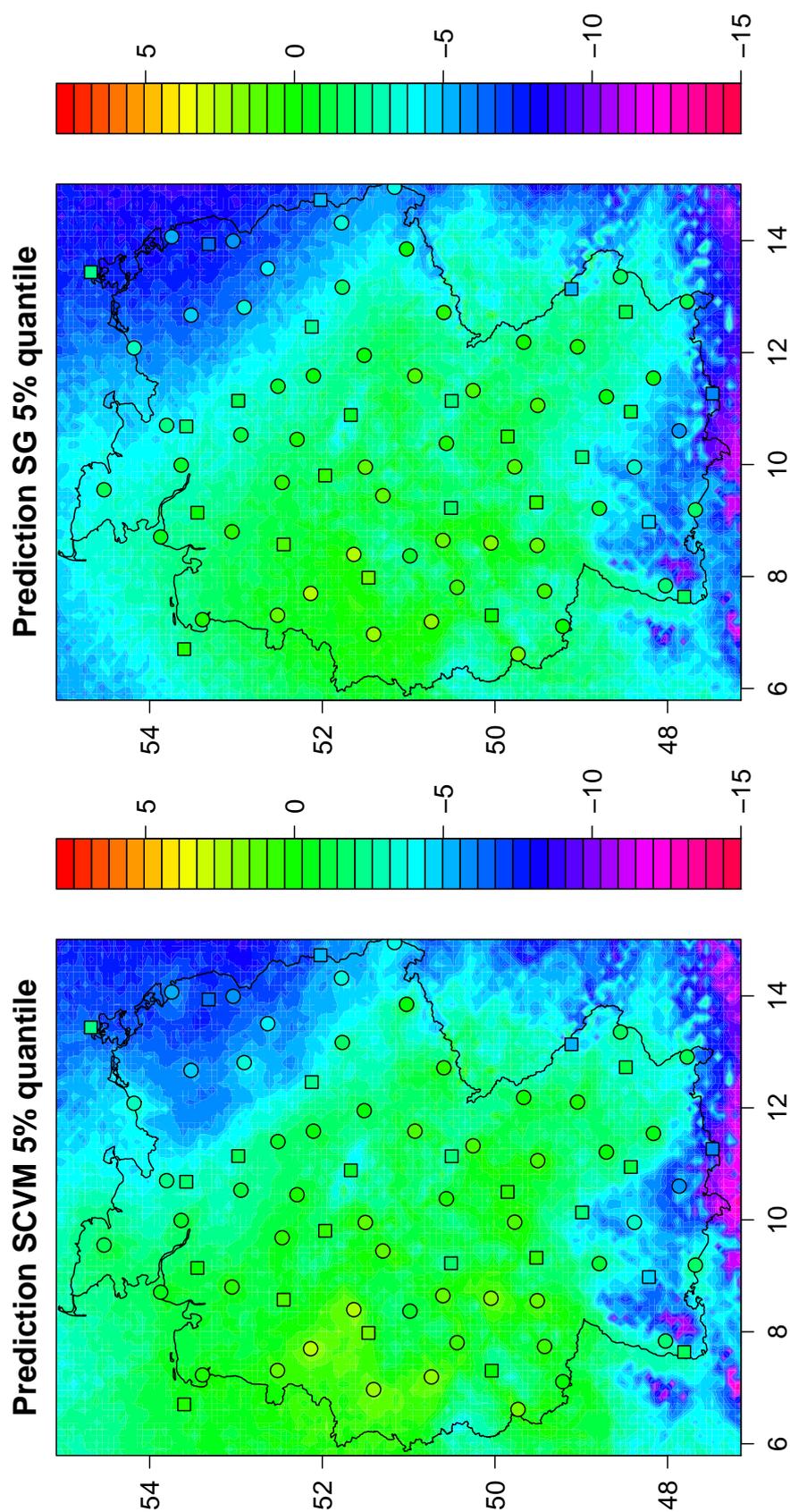
Figure A.4.3: Level plots comparing the 5% quantiles of predictions based on the spatial composite vine model and the spatial Gaussian model on a 80 × 120 grid covering Germany on the 20th of January 2010. Whereas the circles represent the mean temperatures at the 54 observation stations of the training data set, the squares depict the mean temperatures at the 24 stations of the validation data set.
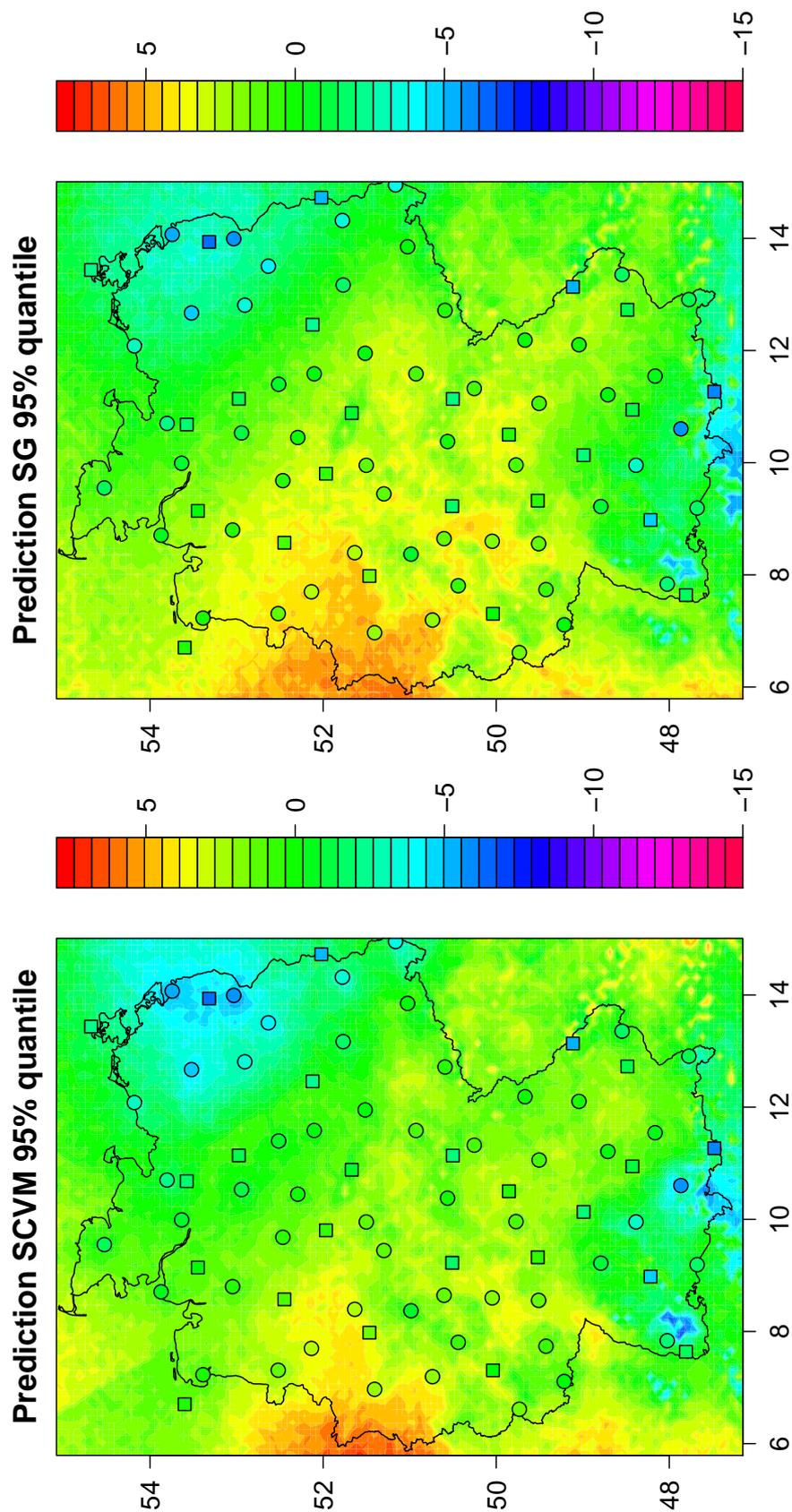
Figure A.4.4: Level plots comparing the 95% quantiles of predictions based on the spatial composite vine model and the spatial Gaussian model on a 80 × 120 grid covering Germany on the 20th of January 2010. Whereas the circles represent the mean temperatures at the 54 observation stations of the training data set, the squares depict the mean temperatures at the 24 stations of the validation data set.
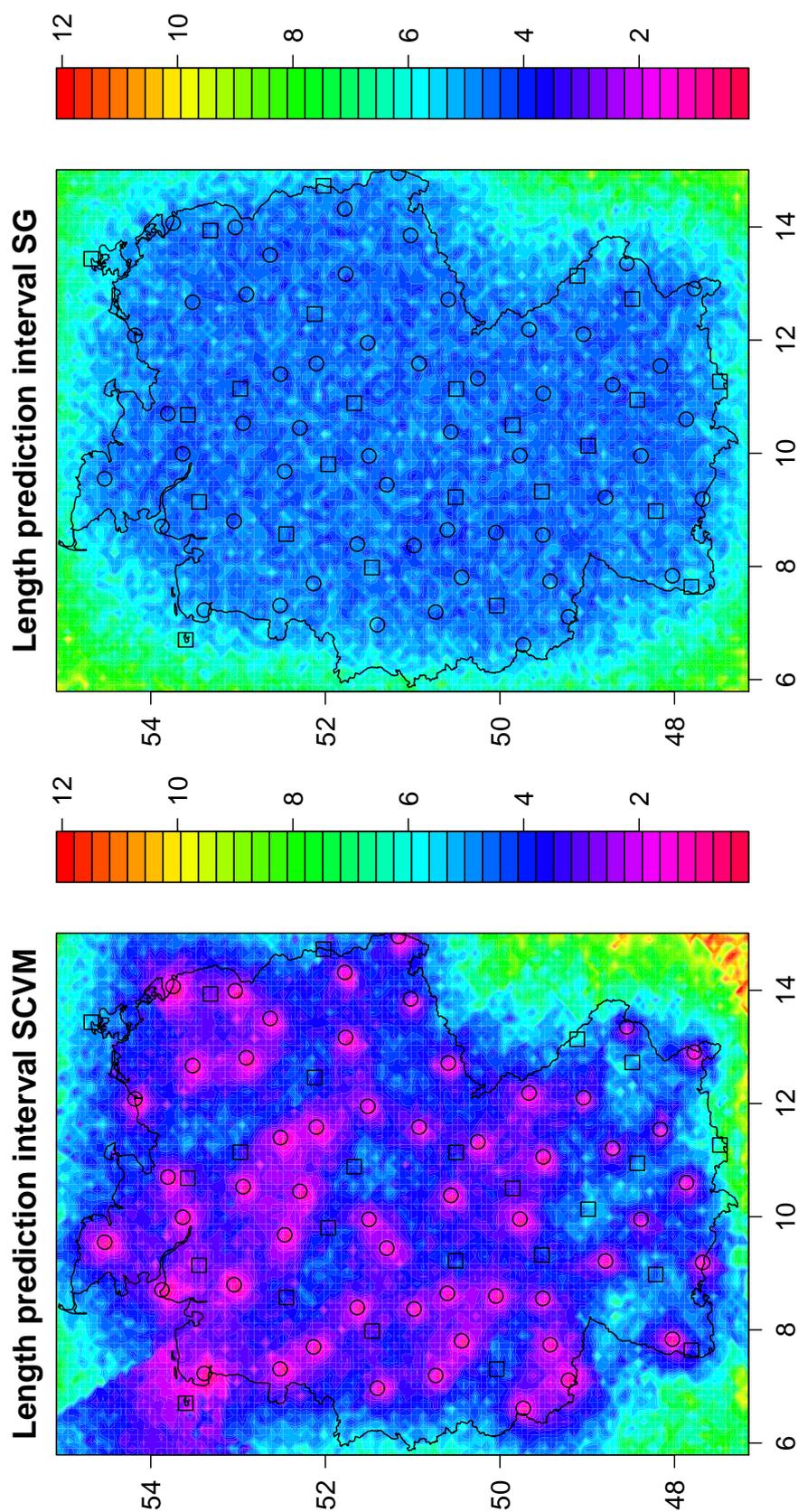
Figure A.4.5: Level plots comparing the lengths of the 90% prediction intervals based on the spatial composite vine model and the spatial Gaussian model on a 80 × 120 grid covering Germany on the 20th of January 2010. The circles represent the 54 observation stations of the training data set, the squares mark the locations of the 24 stations of the validation data set.

List of Tables

List of Figures

Bibliography

Aas, K., C. Czado, A. Frigessi, and H. Bakken (2009). Pair-copula constructions of multiple dependence. *Insurance: Mathematics and Economics 44*(2), 182–198.

Azzalini, A. and A. Capitanio (2003). Distributions generated by perturbation of symmetry with emphasis on a multivariate skew t-distribution. *Journal of the Royal Statistical Society: Series B (Statistical Methodology) 65*(2), 367–389.

Bedford, T. and R. M. Cooke (2001). Probability density decomposition for conditionally dependent random variables modeled by vines. *Annals of Mathematics and Artificial Intelligence 32*, 245–268.

Bedford, T. and R. M. Cooke (2002). Vines - a new graphical model for dependent random variables. *The Annals of Statistics 30*(4), 1031–1068.

Brechmann, E. C. and U. Schepsmeier (2013). Modeling dependence with C- and D-Vine Copulas: The R package CDVine. *Journal of Statistical Software 52*(3), 1–27.

Clayton, D. G. (1978). A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. *Biometrika 65*(1), 141–151.

Cressie, N. A. C. (1991). *Statistics for Spatial Data.* Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. John Wiley & Sons.

Cressie, N. A. C. and C. Wikle (2011). *Statistics for Spatio-Temporal Data.* Wiley Series in Probability and Statistics. John Wiley & Sons.

Czado, C. and T. Schmidt (2011). *Mathematische Statistik.* Statistik und ihre Anwendungen. Springer.

Czado, C. and J. Stöber (2012). Pair copula constructions. In J. Mai and M. Scherer (Eds.), *Simulating Copulas: Stochastic Models, Sampling Algorithms, and Applications*, Series in Quantitative Finance, pp. 185–230. Imperial College Press.

Demarta, S. and A. J. McNeil (2005). The t copula and related copulas. *International Statistical Review 73*(1), 111–129.

Diestel, R. (2010). *Graph Theory* (4th ed.), Volume 173 of *Graduate Texts in Mathematics*. Heidelberg: Springer.

Dißmann, J., E. C. Brechmann, C. Czado, and D. Kurowicka (2013). Selecting and estimating regular vine copulae and application to financial returns. *Computational Statistics & Data Analysis 59*, 52–69.

Dißmann, J. F. (2010). Statistical inference for regular vines and application. Diploma thesis, Technische Universität München.

Durbin, J. (1960). Estimation of parameters in time-series regression models. *Journal of the Royal Statistical Society. Series B (Methodological) 22*(1), 139–153.

Eaton, M. L. (2007). *Multivariate Statistics: A Vector Space Approach*, Volume 53 of *Lecture Notes–Monograph Series*. Beachwood, Ohio, USA: Institute of Mathematical Statistics.

Embrechts, P., F. Lindskog, and A. McNeil (2003). Modelling dependence with copulas and applications to risk management. In S. Rachev (Ed.), *Handbook of Heavy Tailed Distributions in Finance*, pp. 329–384. Elsevier.

Fang, H.-B., K.-T. Fang, and S. Kotz (2002). The meta-elliptical distributions with given marginals. *Journal of Multivariate Analysis 82*(1), 1–16.

Fischer, M. (2011). Multivariate copulae. In D. Kurowicka and H. Joe (Eds.), *Dependence Modeling: Vine Copula Handbook*, pp. 19–36. Singapore: World Scientific.

Fisher, R. A. (1915). Frequency distribution of the values of the correlation coefficients in samples from an indefinitely large population. *Biometrika 10*(4), 507–521.

Frank, M. (1979). On the simultaneous associativity of $F(x,y)$ and $x + y - F(x,y)$. *Aequationes mathematicae 19*, 194–226.

Gelfand, A. E., P. J. Diggle, M. Fuentes, and P. Guttorp (2010). *Handbook of Spatial Statistics*. Chapman & Hall/CRC Handbooks of Modern Statistical Methods. CRC Press: Boca Raton.

Genest, C. and A.-C. Favre (2007). Everything you always wanted to know about copula modeling but were afraid to ask. *Journal of Hydrologic Engineering 12*(4), 347–368.

Genest, C. and J. MacKay (1986). The joy of copulas: Bivariate distributions with uniform marginals. *The American Statistician 40*(4), 280–283.

Gneiting, T. and A. E. Raftery (2007). Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association 102*(477), 359–378.

Gumbel, . J. (1960). Distributions des valeurs extrêmes en plusieurs dimensions. *Publ. Inst. Statist. Univ. Paris 9*, 171–173.

Hersbach, H. (2000). Decomposition of the continuous ranked probability score for ensemble prediction systems. *Weather and Forecasting 15*(5), 559–570.

Joe, H. (1993). Parametric families of multivariate distributions with given margins. *Journal of Multivariate Analysis 46*(2), 262–282.

Joe, H. (1996). Families of m-variate distributions with given margins and $m(m-1)/2$ bivariate dependence parameters. In L. Rüschendorf, B. Schweizer, and M. D. Taylor (Eds.), *Distributions with fixed marginals and related topics*, Volume 28 of *Lecture Notes - Monograph Series*, pp. 120–141. Institute of Mathematical Statistics.

Joe, H. (2001). *Multivariate models and dependence concepts* (1st ed.), Volume 73 of *Monographs on Statistics and Applied Probability*. Chapman & Hall/CRC.

Joe, H. (2011). Tail dependence in vine copulae. In D. Kurowicka and H. Joe (Eds.), *Dependence Modeling: Vine Copula Handbook*, pp. 165–187. Singapore: World Scientific.

Joe, H., H. Li, and A. K. Nikoloulopoulos (2010). Tail dependence functions and vine copulas. *Journal of Multivariate Analysis 101*(1), 252–270.

Kendall, M. G. (1970). *Rank Correlation Methods* (4th ed.). London: Griffin.

Kurowicka, D. and R. Cooke (2003). A parameterization of positive definite matrices in terms of partial correlation vines. *Linear Algebra and its Applications 372*, 225–251.

Kurowicka, D. and R. Cooke (2005). Distribution-free continuous bayesian belief nets. In *Modern Statistical And Mathematical Methods in Reliability*, Series on Quality, Reliability, and Engineering Statistics, pp. 309–322. World Scientific Publishing Company, Incorporated.

Kurowicka, D. and R. Cooke (2006). *Uncertainty analysis with high dimensional dependence modelling*. Wiley Series in Probability and Statistics. John Wiley & Sons, Ltd.

Ljung, G. M. and G. E. P. Box (1978). On a measure of lack of fit in time series models. *Biometrika 65*(2), 297–303.

Matheron, G. (1962). *Traité de Géostatistique Appliquée, Tome I.* Number 14 in Mémoires du Bureau de Recherches Géologiques et Minières. Editions Technip.

Nelsen, R. B. (2006). *An Introduction to Copulas* (2nd ed.). Springer Series in Statistics. New York: Springer.

Pachali, M. (2012). Modeling dependence among meteorological measurements and tree ring data. Diploma thesis, Technische Universität München.

R Development Core Team (2012). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. ISBN 3-900051-07-0.

Schepsmeier, U. (2010). Maximum likelihood estimation of c-vine pair-copula constructions based on bivariate copulas from different families. Diploma thesis, Technische Universität München.

Schepsmeier, U., J. Stöber, and E. C. Brechmann (2013). *VineCopula: Statistical inference of vine copulas.* Technische Universität München. R package version 1.1.

Simmons, L. (1990). Time-series decomposition using the sinusoidal model. *International Journal of Forecasting 6*(4), 485–495.

Sklar, A. (1959). Fonctions de répartition à n dimensions et leures marges. In *Publications de l'Institut de Statistique de L'Université de Paris, 8*, pp. 229–231. Institut Henri Poincaré.

Varin, C., N. Reid, and D. Firth (2011). An overview of composite likelihood methods. *Statistica Sinica 21*(1), 5–42.