Modelling Extreme Wind Speeds

Tilman Payer

Dissertation

an der Fakultät für Mathematik, Informatik und Statistik

der Ludwig–Maximilians–Universität München

23. November 2006

Modelling Extreme Wind Speeds

Dissertation

an der Fakultät für Mathematik, Informatik und Statistik

der Ludwig–Maximilians–Universität München

zur Erlangung des Grades Doctor rerum naturalium (Dr. rer. nat.)

vorgelegt von

Tilman Payer

eingereicht am 23. November 2006

1. Berichterstatter:	Prof. Dr. Helmut Küchenhoff
2. Berichterstatter:	Prof. Dr. Leonhard Held
3. Berichterstatter:	Prof. Dr. Jonathan Tawn

Tag des Rigorosum: 8. März 2007

Acknowledgements

The outcome of this thesis would not have been possible in the current form without the advise and support of many people helping me on my way. I want to thank particularly Helmut Küchenhoff for giving me the opportunity to work on this project, for many helpful discussions, feedback and guidance; his guidance was very important for me both in situations related to my statistical work of this thesis and to other aspects of academic life like teaching skills in tutorials. I am very grateful to Janet Heffernan, who gave me a path through multivariate extreme values and shared her knowledge with me how to progress with my projects; her expertise, patience in explaining me related background knowledge and the many discussions with her were very important for the outcome of this work. I am very glad that I was given the opportunity to work with Jonathan Tawn. He was giving me very much encouraging feedback on my work and was suggesting so many improvements in situations of doubt; his creativity and expertise was giving my working on the project a dimension I can not be grateful enough.

During my thesis I was given the opportunity to work within a funded project with people from Lancaster University. Leonhard Held made me aware of this opportunity and I am grateful for his encouragement; I also want to thank him for many helpful discussions at several stages of my academic life. Rob Henderson was giving me important guidance and advice in Lancaster I am very thankful for.

I want to thank Adam Butler, Mark Latham, Caroline Keef, Stefan Krieger and Holger Gnann for sharing their knowledge and opinion with me on statistical and non-statistical subjects. I also want to thank Carolin Strobl, Sandro Scheid and Stefan Pilz for many interesting discussions; Stefan Pilz was carefully reading a preliminary version of my thesis and giving me important feedback.

I spend most of my study and research time at Munich's Department of Statistics,

which I will always remember as a warm, friendly and creative place. Many people have made this place so special for me both from a statistical and from a personal point of view, and although I am thankful to all, the list would be too long to state here. I am particularly grateful for advice, help and support from Brigitte Maxa, Christa Jürgensonn, and Stefanie Tullius.
I am very glad to have spent time at the Department of Mathematics and Statistics in Lancaster. Many members of it made it such an exciting, friendly and inspiring place to work; I hope they will forgive me mentioning only Helen Shaw, Julia Tawn, Cathy Thomson, and Christian Cable. I am very thankful for the many friendships I made in Lancaster and joint activities which I enjoyed a lot.
I also want to thank the German Railway and German Meteorological Service (DWD) for providing me with the data and the related applicational problem to work with. I further want to thank the Klaproth-Stiftung for supporting my project in the initial stage of the project. I am indepted to the EU Marie Curie Foundation giving me the possibility and financial support to work on my project with people from Lancaster University, which was both for my work and personally a more than positive experience.
Finally I want to thank LLenalia for her patience, understanding, support and encouragement in situations where I really needed them and for the many wonderful moments we had together. I am more than grateful to my family always providing me security and support to complete this work; the outcome of this thesis would not have been possible without the continuous encouragement of my parents Rotraut and Dieter.

Zusammenfassung

Da sehr starke Windereignisse das Entgleisen einiger Hochgeschwindigkeitszüge verursachen können, ist Kenntnis über das Verhalten vom Windprozeß in extremen Bereichen notwendig. Die Windrichtung relativ zur Fahrtrichtung des Zuges spielt eine entscheidende Rolle für die Stabilität des Zuges, so daß diesem Aspekt Rechnung getragen werden muß. Zunächst wird das Sturmverhalten an einer Wetterstation betrachtet. Ein Extremwertmodell für Windgeschwindigkeiten, das auch die Windrichtung berücksichtigt, wird sowohl auf Rohdaten als auch auf modifizierte Daten, die die Kraft des Windes in eine bestimmte Richtung repräsentieren, angewendet. Extreme Quantile und Überschreitungswahrscheinlichkeiten werden geschätzt und zugehörige Konfidenzintervalle bestimmt. Ein gängiges Problem mit Winddaten ist, daß pro Zeitintervall nur die größte Beobachtung aller Richtungen registriert wird, während Beobachtungen in allen anderen Richtungen des selben Zeitintervalls unbeachtet bleiben. Um Modellschätzungen zu verbessern schlagen wir ein Modell vor, das diesem Problem Rechnung trägt. Anhand einer Simulationsstudie werden die Eigenschaften des neuen Modells in unterschiedlichen Situationen untersucht. Dabei wird das Verhalten des neuen Modells mit dem eines herkömmlichen Modells verglichen und auf der Basis des mittleren quadratischen Fehlers extremer Quantile beurteilt. Sowohl in der Simulationsstudie als auch bei nachfolgender Anwendung auf reale Winddaten zeigt das neue Modell wünschenswerte Eigenschaften.
Daraufhin wird ein kürzlich vorgestelltes multivariates Extremwertmodell betrachtet, das ein breites Spektrum verschiedener Abhängigkeitsstrukturen erlaubt und deshalb für viele Anwendungen sehr geeignet ist. Da der Abhängigkeitsgrad dieses Modells von mehreren Größen bestimmt wird, ist eine exakte Quantifizierung der Abhängikeitsstärke nicht einfach. Zur Beurteilung der Abhängigkeit betrachten wir deshalb visuelle Kenngrößen, deren Verhalten in einer Simulationsstudie untersucht

wird. Das multivariate Extremwertmodell wird im weiteren auf Winddaten zweier Wetterstationen unter Berücksichtigung der Windrichtung angewendet. Mit diesem Modell lassen sich Aussagen über das gemeinsame Windverhalten beider Stationen machen. Es trägt somit zur Beurteilung bei, ob Sturmereignisse eher lokal oder über weitere Teile einer Bahnstrecke auftreten.

Abstract

Very strong wind gusts can cause derailment of some high speed trains so knowledge of the wind process at extreme levels is required. Since the sensitivity of the train to strong wind occurrences varies with the relative direction of a gust this aspect has to be accounted for. We first focus on the wind process at one weather station. An extreme value model accounting at the same time for very strong wind speeds and wind directions is considered and applied to both raw data and component data, where the latter represent the force of the wind in a chosen direction. Extreme quantiles and exceedance probabilities are estimated and we give corresponding confidence intervals. A common problem with wind data, called the masking problem, is that per time interval only the largest wind speed over all directions is recorded, while occurrences in all other directions remain unrecorded for this time interval. To improve model estimates we suggest a model accounting for the masking problem. A simulation study is carried out to analyse the behaviour of this model under different conditions; the performance is judged by comparing the new model with a traditional model using the mean square error of high quantiles. Thereafter the model is applied to wind data. The model turns out to have desirable properties in the simulation study as well as in the data application.
We further consider a multivariate extreme value model recently introduced; it allows for a broad range of dependence structures and is thus ideally suited for many applications. As the dependence structure of this model is characterised by several components, quantifying the degree of dependence is not straight forward. We therefore consider visual summary measures to support judging the degree of dependence and study their behaviour and usefulness via a simulation study. Subsequently, the new multivariate extreme value model is applied to wind data of two gauging stations where directional aspects are accounted for. Therefore this model allows

for statements about the joint wind behaviour at the two stations. This knowledge gives insight whether storm events are likely to be jointly present at larger parts of a railway track or rather occur localized.

Contents

Chapter 1

Introduction

Technical innovations often face the risks of failure posed by natural forces. Not properly accounting for such risks can have catastrophic consequences. One of these forces is storms and strong wind events, and there are many applications where a solid knowledge of the present storm behaviour is essential. Examples include large buildings, hang bridges, and other design structures. An application we are particularly interested in is modern high speed trains, which are built with light materials to reach a very high velocity. Both, the increased speed and lower weight of these trains, reduce their stability and they are consequently more sensitive to strong storms. An important aspect in this context is the direction of the wind. The risk of derailment is highest if the wind direction is perpendicular to the motion of the train and vanishes if it is parallel to the rails.

To judge the risk of a strong wind event with potential of causing derailment, knowledge of the wind process at extreme levels is required. The most promising approach is extreme value statistics. It joins two important pieces together: an extensive mathematical theory providing distributions derived for this problem, and information contained in collected data properly exploited by statistical techniques adapted to this situation. This combined tool then allows us to make judgement in regions where data are too scarce to provide solid empirical information or are even beyond any observations made so far.

Extreme value techniques have been applied and proven to be successful in very

many different areas. Temperature and rainfall data are two examples. It is invaluable for designing sea-defences protecting against coastal flooding or in developing off-shore designs. Applications to air-pollution are often related to pollution standards (Küchenhoff and Thamerus, 1996). But also non-environmental applications are common as, for example, to assess portfolio risks, material-strength, or even to assist in judging whether exceptional sport events may have been achieved by the support of drugs (Robinson and Tawn, 1995).

Applications to storm events are also common. The wind behaviour in Germany is dominated by storms resulting from a differential in pressure and by thunderstorms. For applications to complex dynamic wind systems such as tropical storms see for example Casson and Coles (1998) or Walshaw (2000). Walshaw (1991) discusses univariate extreme value statistics applied to wind gusts, and Walshaw and Anderson (2000) consider how to incorporate information about average wind speeds to improve knowledge about gust behaviour. An important feature considered in this thesis is directionality of extreme winds. Coles and Walshaw (1994) suggest a model for extreme wind speeds allowing for smooth directional variation in extreme wind occurrences.

We analyse wind data provided by the Deutscher Wetterdienst DWD (German weather service) for different gauging stations nearby a railway track from Hannover to Würzburg (see Firgure 1.1). They consist of 22 years of daily maxima and the mean wind direction of the hour they occurred within. Additionally we have for the weather station Würzburg data of ten minutes maxima and their corresponding directions. These gauging stations were chosen since the railway track they are nearby has north-south orientation. So with the dominant wind direction being west, gusts roughly hit the train perpendicular to its direction of movement and therefore pose the highest risk.

Knowledge of the storm behaviour is essential for the Deutsche Bahn (German Rail) to make effective safety decisions. These may include building wind walls at certain exposed or dangerous parts of the track, the installation of a wind warning system at the track, or if necessary, trains need to reduce speed at certain parts of the track or at times with a high storm risk. Methods discussed in this thesis, however, are not

restricted to high speed trains, as already mentioned. For example, knowledge of directionality of storm events may lead to considerable cost savings when planning a large building. Depending on the application at hand, it may be more appropriate to work directly with the wind data or transform them, in a way to be made more precise, to represent the force of the wind at a specific point of a train or a design structure exposed in a certain direction.

When analysing the wind behaviour of just one weather station, the random process may be assumed as bivariate with wind speed and wind direction as components. It is, however, more convenient to break up this bivariate random variable into the univariate random variable direction and the conditional variable wind speed given a certain direction. The latter one is often more complex to model, and mostly interest is just in extreme wind speeds for a given direction. The directional distribution may be estimated empirically, as there is usually enough data, so joint probabilities are easily calculated from the two distributions. When considering two or more stations, we are also interested in the dependence structure and thus apply multivariate methods.

The thesis is structured as follows. In the subsequent sections of this chapter we give a short overview of univariate extreme value theory and practical aspects relevant for this thesis. Thereafter, the data we have are considered, and we introduce some of the notation used in this work. Chapter 2 is a reviewed and modified version of Payer and Küchenhoff (2004). In that chapter we discuss the application of a model employing the k largest values and accounting for directionality in the context of high speed trains. A common problem with wind data, called the masking problem, is discussed as well. This problem arises from the way of recording wind data by registering information about wind speeds only in the direction where the maximum occurred. In Chapter 3 we suggest an approach to account for this problem by including the knowledge in other directions to be no larger than the biggest wind speed observed in this day. In Chapter 4 the conditional multivariate extreme value model introduced by Heffernan and Tawn (2004) is considered. It provides much greater flexibility in modelling the dependence structure at extreme levels than earlier models did. The dependence structure is, however, determined by several parameters and a residual distribution, making it difficult to state the degree

of dependence in a simple number. Furthermore, direct comparisons to earlier models is not straight forward. We analyse different visual measures summarizing the dependence structure. Due to the great flexibility of the conditional multivariate extreme value model, it is ideally suited for jointly analysing the wind behaviour at extreme levels for two stations accounting simultaneously for direction. This approach is considered in Chapter 5. In comparison with a sector by sector analysis, this approach has the advantage of considerably reducing the number of parameters as well as allowing information from neighbouring directions to be employed.

Most of the programs used in this thesis were coded up by the author himself. Some codes of Jan Heffernan and Stuart Coles are employed and extended. The statistical software used is R, and in some cases requiring high computational performance the program C was incorporated using the GNU compiler gcc.

1.1 Some univariate techniques for extreme events

We give a short overview of univariate extremes relevant to the work presented in this thesis. Since interest is in rare events often outside the range of data, extreme value theory and statistics is based on parametric distributions. First, the maximum of independent and identically distributed random variables and possible limit distributions are considered, forming the backbone of extreme value theory. How to access this theory for practical purpose and further statistical aspects are discussed thereafter. For a more complete introduction we refer to Embrechts, Klüppelberg and Mikosch (1997), Coles (2001), Beirlant, Geogebeur, Segers and Teugels (2004), de Haan and Ferreira (2006), Leadbetter, Lindgren and Rootzén (1983), and Resnick (1987).

1.1.1 Theoretical results

Let $X_1, \ldots, X_n$ be independent and identically distributed (iid) random variables where X_i, $i = 1, \ldots, n$, has distribution function F with upper endpoint $x_F = \sup\{x \in I\!R : F(x) < 1\}$. The classical approach to extremes is in considering the

distribution of the maximum

$$M_n = \max\{X_1, \ldots, X_n\}, \quad n \in I\!N,$$

that is $P(M_n \leq x) = F^n(x)$. In many applications F is not known exactly, and consequently, the exact distribution of M_n is unknown. To find a more general result it is natural to consider the asymptotic behaviour of M_n. As $n \to \infty$, however, F^n is a degenerate distribution converging to the upper end point x_F. Thus the growth of M_n has to be adjusted properly to avoid degeneracy. In analogy to the central limit theorem, an apparent choice is a linear transformation $(M_n - b_n)/a_n$ with sequences of coefficients $b_n \in I\!R$ and $a_n > 0$. The key result, attributed to Fisher and Tippet (1928) and Gnedenko (1943), states that if there exist sequences of constants $a_n > 0$ and b_n, such that, as $n \to \infty$,

$$P\left(\frac{M_n - b_n}{a_n} \leq x\right) \to G(x) \tag{1.1}$$

for some non-degenerate distribution G, then G belongs to one of the three families

$$\begin{array}{lrcl}
\text{Gumbel}: & \Lambda(x) & = & \exp\left\{-\exp(-[(x-b)/a])\right\}, \quad -\infty < x < \infty \\
\text{Fréchet}: & \Phi_\alpha(x) & = & \begin{cases} 0 & x \leq b \\ \exp(-[(x-b)/a]^{-\alpha}) & x > b, \quad \alpha > 0 \end{cases} \\
\text{Weibull}: & \Psi_\alpha(x) & = & \begin{cases} \exp(-(-[(x-b)/a])^{\alpha}), & x < b, \quad \alpha > 0 \\ 1 & x \geq b, \end{cases}
\end{array}$$

where convergence is in distribution. Collectively the three distributions are referred to as *extreme value family*. The Weibull occurring in the extreme value family is a reversed version of the usually considered standard Weibull. It is worth mentioning that for each member G of the above families $G(ax + b) = G^*(x)$ with $a > 0$ and $b \in I\!R$, G and G^* are belonging to the same family. Furthermore, if G belongs to one of the above families, then for any positive n there exist $a_n > 0$ and b_n so that $G^n(a_n x + b_n) = G(x)$ holds. The latter property is unique to the three extreme value families and often referred to as max-stability.

In many applications the iid assumption underlying the above stated limiting result is not satisfied. Leadbetter et al. (1983) give conditions under which the limiting

distribution of the maximum of a strictly stationary time series is still one of the extreme value families. Two of these conditions are referred to as D and D'. Condition D insures dependency to be negligible for two variables being separated far enough in time. So for a stationary sequence $(\tilde{X}_i)_{i\geq 1}$ with marginal distribution F satisfying D its maximum, $\tilde{M}_n$, has the limiting distribution $P((\tilde{M}_n - b_n)/a_n < x) \to \tilde{G}(x)$ as $n \to \infty$. The relation of this limiting distribution with its counterpart based on iid random variables is $\tilde{G} = G^\delta$ where $\delta \in [0, 1]$ is called the *extremal index* accounting for the reduction of independent information. Condition D' focuses on the short-term dependence behaviour of a sequence and together with D states the limiting distribution of the maximum to behave like an iid sequence.

1.1.2 Statistical aspects

Working with three different limit distributions is not of practical advantage, as it requires a choice to be made in advance. A unification of these three types into a single family (von Mises, 1954; Jenkinson, 1955), known as the *generalized extreme value distribution* (GEV), is given by

$$G(x) = G_{(\xi,\mu,\sigma)}(x) = \exp\left\{-\left[1 + \xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-1/\xi}\right\} \tag{1.2}$$

whenever $\{x : 1 + \xi(x-\mu)/\sigma > 0\}$ with $\mu \in \mathbb{R}$ and $\sigma > 0$ being location and scale parameters, respectively. The shape parameter $\xi \in \mathbb{R}$ determines whether or not the distribution has an upper bound. The former is true whenever $\xi < 0$, which corresponds to a Weibull distribution with upper endpoint $x_F = \mu - \sigma/\xi$, while there is no upper limit for $\xi > 0$, which is of Fréchet type and $\xi = 0$, being interpreted as $\xi \to 0$ yielding a Gumbel distribution.

The sample size n is finite in any application. For large n it is natural to assume the limiting distribution arising from equation (1.2) to be a reasonable approximation. Focusing directly on the distribution of the maximum may be represented as

$$P(M_n \leq x) \approx G((x - b_n)/a_n) = G^*(x),$$

where G and G^* are two different members of the same family given by (1.2) differing in location and scale. Using G^* does not require knowledge of a_n and b_n, so it can be directly fitted to a series of maxima.

In applications the sample size is assumed to be large enough for the limit distribution to serve as a good approximation to the true one. Consequently, the limit distribution is supposed to hold exactly. Common methods of estimation are maximum likelihood and Bayesian estimation (Coles and Powell, 1996; Stephenson and Tawn, 2004). A frequently mentioned competitor is the probability weighted moments (PWM) estimator, where estimation is based on equating these modified theoretical moments to their empirical counterparts. Hosking (1985) shows that for small sample sizes the PWM method is superior to maximum likelihood. However, it is not applicable for $\xi \geq 1$, as the expectation and higher moments do not exist, so the parameter space is a priori restricted to $(-\infty, 1)$. Coles and Dixon (1999) clarify that the supposed superiority for small samples of the estimators based on probability weighted moments is due to this restriction of the parameter space. A major drawback of PWM is that it does not allow for extension to more complex problems like including covariable information (Smith, 1990). Since covariable information is essential to the present work, PWM is not a possible choice. We will use maximum likelihood and methods derived from it.

Prescott and Walden (1980) discuss maximum likelihood estimation for the GEV, giving exact expressions for the calculation of the expected Fisher information matrix. Many authors, however, suggest the observed Fisher information to produce better results (Smith, 1990). With the range of the distribution depending on the parameters of the GEV, common regularity conditions underlying maximum likelihood theory are not satisfied. However, Smith (1985) shows that for $\xi > -1/2$ the asymptotic theory underlying maximum likelihood is still applicable. In particular the asymptotic normality for parameter estimates holds. Distributions with $\xi \leq -1/2$ relate to a very short upper tail, which are rather an exception in environmental applications.

Extreme value statistics is usually considered if interest is in very rare events possibly outside the range of data observed so far. Having estimated the parameters of the

GEV, calculation of the probability of any event is immediate from equation (1.2). Another very common application is to consider a high quantile $x_p = G^{-1}(p)$ given by

$$x_p = \mu - \frac{\sigma}{\xi}\{1 - [-\log(p)]^{-\xi}\},$$

where, for estimation of the quantile, parameters are replaced by their estimates. Often the maximum is taken over an interval corresponding to one year, so a common way to state quantiles is in terms of years. In this context quantiles are called *return levels* and the J-year return-level with $p = 1 - 1/J$ is given by $x^{(J)} = x_{1-1/J}$.

Extreme events are rare in nature, and consequently the amount of observations to draw conclusions from is small. So it is necessary to make the best use of available information. In many applications there is much more data recorded than just the maximum. There are two common approaches to include additional information. One approach is to not only consider the maximum, but also other order statistics. The other possibility is to consider all values which are extreme in the sense of exceeding a certain high value.

An extension of (1.2) is given by considering the k largest values with $k \in \{1, 2, \dots\}$. Let $x^{(1)} \geq x^{(2)} \geq \cdots \geq x^{(n)}$ be the ordered values of a sample of size n. Then the asymptotic distribution of the k largest order statistics has for $\mathbf{x} = (x^{(1)}, \dots, x^{(k)})$ the density given by

$$\begin{aligned} g^{(k)}(\mathbf{x}) &= \sigma^{-k} \exp\left\{ -\left[1 + \xi\left(\frac{x^{(k)} - \mu}{\sigma}\right)\right]^{-1/\xi} - \right. \\ &\qquad \left. - \left(1 + \frac{1}{\xi}\right) \sum_{l=1}^{k} \log\left[1 + \xi\left(\frac{x^{(l)} - \mu}{\sigma}\right)\right] \right\} \end{aligned} \tag{1.3}$$

whenever $\{x^{(l)} : 1 + \xi(x^{(l)} - \mu)/\sigma > 0, l = 1, \dots, k\}$. For $k = 1$ this reduces to the density of the GEV given in (1.2). Statistical aspects are discussed by Smith (1986) and Tawn (1988).

An alternative approach is to consider values which are extreme in the sense of exceeding a large, specified value. The *generalized Pareto distribution* (short GPD)

arises from the conditional distribution of exceedances X of a high threshold u_X. If X is in the domain of attraction of an extreme value distribution given by (1.2), then as u_X approaches x_F

$$P(X \leq x | X > u_X) \approx 1 - \left[1 + \xi \frac{x - u_X}{\beta}\right]_+^{-1/\xi} \quad \text{for } x > u_X,$$

where $\beta > 0$ and $\xi \in \mathbb{R}$ are a scale and a shape parameter, respectively, and $m_+ := \max\{0, m\}$ (this is made precise in Pickands (1975)). The case $\xi = 0$ is interpreted as $\xi \to 0$, resulting in the well–known exponential distribution. There is a strong relation between the GPD and GEV; in particular, the shape parameters ξ of the two distributions coincide. The scale parameter of the GPD relates to parameters of the GEV by $\beta = \beta(u_X) = \sigma + \xi(u_X - \mu)$, where $\beta(u_X)$ stresses the dependence on the choice of threshold. Statistical application of the GPD is discussed by Davison and Smith (1990).

Another approach, which directly focuses on the upper tail of the distribution F of X, is to use the GEV for all values exceeding a certain high value u. Let us assume that (1.1) holds for some large n, and we further assume the existence of some u close to the upper endpoint x_F so that

$$P(M_n \leq a_n x + b_n) = \{F(a_n x + b_n)\}^n \approx G^+(x)$$

holds for each x satisfying $a_n x + b_n > u$, where G^+ is given by (1.2). Then it follows that $\{F(y)\}^n \approx G^+((y - b_n)/a_n) = G(y)$ for $y > u$, and consequently $F(y) \approx G^{1/n}(y)$ holds if $y > u$. G and $G^{1/n}$ have the same shape parameter ξ, while the parameters μ_* and σ_* of $G^{1/n}$ are given by $\mu_* = \mu + (n^{-\xi} - 1)\sigma/\xi$ and $\sigma_* = n^{-\xi}\sigma$. Advantage of this approach is the direct focus on GEV parameters, so return-level calculation is immediate after parameters are estimated. This approach is used in Chapter 3.

The alternatives to just considering maxima are capable of better exploiting information within the data leading to a higher precision of estimates. However, the number of order statistics should not be too high or the threshold too low, as this may invalidate the asymptotic assumptions justifying the use of these approaches.

1.2 Description of the data

The data being analysed in this thesis are from the meteorological gauging stations Würzburg and Hannover, which constitute the two endpoints of a railway track for high speed trains. A schematic map of Germany and the track are shown in Figure 1.1. It can be seen from the map that the orientation of the track is in north south direction. Recent analyses for German wind data based on ten-minutes averages have been carried out by Kasperski (2002).

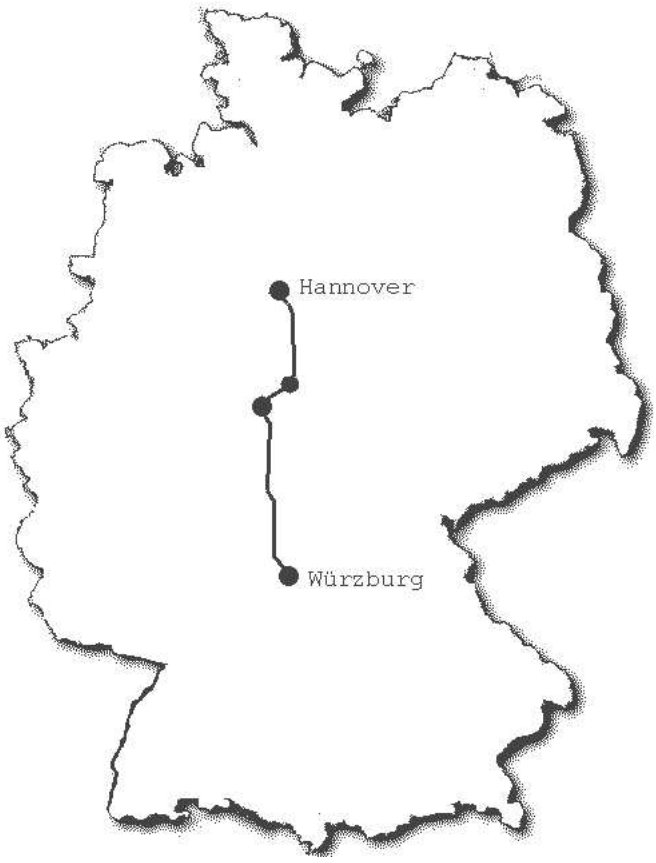

Figure 1.1: Schematic map of Germany and the position of the railway track.

The data consist of 22 years of daily maximum wind speeds corresponding to a 2 - 3 seconds gust and the time of day they occurred. The recording period is

from 1 January 1976 to 31 December 1997. The wind-direction of the maximum itself is not available, but the average of the wind-direction within each hour has been recorded with an accuracy of 10°. Thus we take the average wind-direction of the hour in which the maximum occurred as its direction. Analyses of data from shorter time intervals for one year have shown that the hourly average of the wind-direction constitutes a reliable measurement for the exact direction corresponding to the maximum.

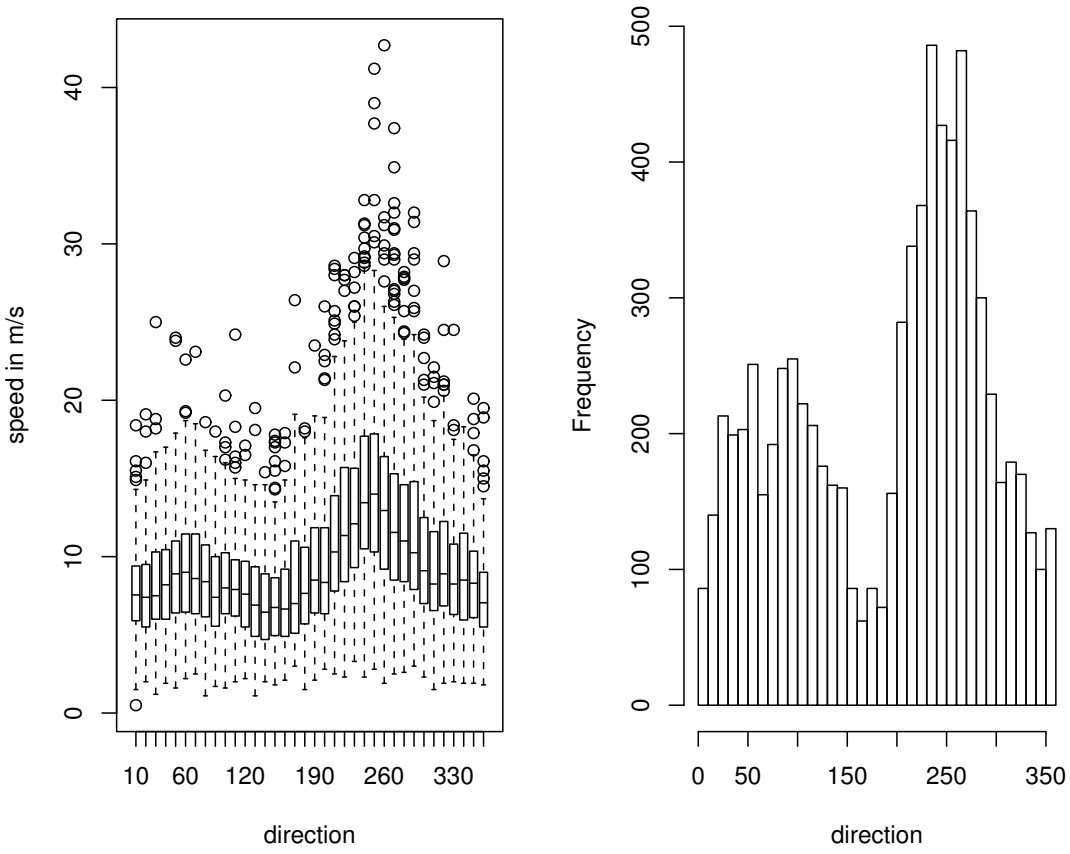

Figure 1.2: (Würzburg) Boxplots of the wind speeds of all observations within the 22 years for different directions (left); histogram of directions (right).

To get an impression of the data we consider plots for the 7892 observations from Würzburg left after removing 144 missing values. Figure 1.2 shows boxplots of the wind speeds of daily maxima for all directions as well as a histogram reflecting corresponding frequencies of directions. The angle $\phi = (0) = 360$ given in degrees is defined as the direction north, and angles are recorded clockwise. The graphs indicate the wind process to change in a smooth fashion over directions. The strongest gusts as well as the highest frequency of gusts are in western direction, but also the opposite direction produces high wind events. Therefore the main wind occurrences are approximately perpendicular to the track. For the weather station Hannover there are 7909 data after removing missing values. For both stations, wind speeds are given in metres/second (m/s) with an accuracy of 0.1 m/s for an effective height of 10 m above ground level.

For a period of ten years, from 1 January 1993 to 31 December 2002, we have wind maxima and their corresponding direction of ten-minutes intervals for the Würzburg station. With missing values and after deletion of some obvious mis-recordings, the data set consists of 486267 observations.

Let the pair (R, Φ) describe the daily maximum wind speed R having direction Φ, where $R = [0, \infty)$ and $\Phi \in \Omega \subset (0, 2\pi]$. In applications Ω is a finite subset of $(0, 2\pi]$. Our data, for example, partition $(0, 2\pi]$ into 36 equally spaced sectors, which are referred to by their center-points in degree, that is $\{10°, 20°, \ldots, 360°\}$, so that with $b = 2\pi/360°$ we have $\Omega = \{b{\cdot}10°, \ldots, b{\cdot}360°\}$ represented in radians. Angles in radians and degrees are used according to convenience, and we drop the degree-symbol if the unit is clear from the context. The conditional random variable $R_\phi = (R|\Phi = \phi)$ is used to describe the conditional distribution of wind speed given wind direction ϕ.

In this thesis we consider two types of data depending on the context and application. The first type simply employs the raw data as recorded. The second type is to consider the force of the wind in a specified direction. For a given wind event R_ϕ in direction ϕ, the power of the wind in direction α is $\tilde{R}_\alpha = R_\phi \cos(\phi - \alpha)$. To distinguish the two approaches we refer to the second as component or resolved data. So while the first type of data reflects the nature of the wind process, the component data focus on the power of the wind, which may be more appropriate

for some applications within an engineering context.

Chapter 2

Modelling extreme wind speeds at a German weather station

In this chapter we consider modelling extreme wind speeds for one weather station, Würzburg. A key factor is to take directionality into account. In the context of high speed trains this knowledge about extreme wind speeds in different directions is crucial. Both, the fact that the trains reach very high velocities as well as lower weights due to the use of light materials to reach this goal, reduce the stability of the train and make them more sensitive to strong wind events which are not parallel to the rails. Therefore models to describe directional behaviour of extremes are necessary.

The problem we face is to assess the risk of derailment caused by extreme gusts. Several factors like speed of the train, track curves, and others have an influence on this risk. One apparent and important factor is the wind speed itself. As the stability of the train to wind varies with relative wind direction this variable has to be taken into account as well. For the analysis of wind speeds allowing for directional variation we apply a model proposed by Coles and Walshaw (1994). It uses the k largest order statistics of every year to estimate the parameters of the generalized extreme value distribution (GEV), the asymptotic distribution of annual maxima. The parameters of the GEV vary according to harmonic terms with direction. Incorporation of this functional relationship allows transfer of information over directions, so precision of estimates can be improved in comparison with a sector by sector analysis.

We consider two approaches. The first simply employs the raw data as they are recorded. The second uses wind speeds resolved to component data to reflect the power of the wind. We discuss both methods and their different interpretation and usefulness in the context of probabilistic assessment.

After having estimated the model, there are two possibilities of probabilistic assessment we look at. The first one is the classical approach, where extreme quantiles, often referred to as return-levels, are calculated; here, the exceedance probability is fixed and the corresponding wind speed is calculated. The second possibility is to fix a critical wind speed value and calculate its probability of being exceeded. The first approach is sensible if we are interested in the wind speeds we must expect to face in order, for example, to think about measures like wind protection. The second is favourable if we know the wind speed which leads to derailment of the train at a particular point of the track.

To get an impression of the precision of either, the return-level or the exceedance probability, confidence intervals are calculated. Two methods are commonly applied: the so called delta method, which yields symmetric intervals, and the profile likelihood method, allowing for asymmetric intervals. We discuss both methods.

All analyses and conclusions are based on the assumption that the applied model using harmonic terms is an appropriate choice and that the data used are enough to yield a good approximation to the applied model, which is justified by asymptotic arguments. The model's performance is investigated through a simulation study. Then for one particular choice confidence intervals of extreme quantiles are used to judge the adequacy.

By just recording the maximum of a certain time interval (say, a day), for analysing extremes over directions there is always the problem of extremes in other directions than the maximum being missed in the resulting data set. This problem, often referred to as 'masking', is partially alleviated by using components. We therefore compare daily maxima and components with maxima of ten-minutes intervals for two subsequent years.

2.1 Theoretical Background

2.1.1 Model for extreme wind speeds

In investigating processes at extreme levels it is common practice to employ parametric models which have an asymptotic justification. The classical approach is to consider the maximum of a large iid sample, which in the case of non-degeneracy converges to the GEV distribution given in (1.2). We analyse the annual maximum of wind speeds, which may be regarded as the maximum of 365 daily maxima. As the asymptotic theory is still valid under mild dependence conditions (Leadbetter et al., 1983), the slight deviation from the independence assumption is not essential. We therefore assume the GEV to be an appropriate model for annual maxima.

Taking only the maximum value of each year is apparently a high loss of information; as, additionally, in most applications only data from a few years are available, the precision of resulting estimates is low. Exploiting the information of other high values leads to a generalization of the GEV which is the limiting distribution of the k largest order statistics. This distribution is characterized by the same parameters as the GEV.

The random variable R_ϕ is defined as the wind speed R given that it occurred in direction $\phi \in \Omega \subset (0, 2\pi]$, while we denote the corresponding outcome by r_ϕ. Furthermore, we denote the order statistics for a given direction ϕ by $r_{\phi j}^{(1)} \geq r_{\phi j}^{(2)} \geq r_{\phi j}^{(3)}, \ldots,$ where $j = 1, \ldots, N$ denotes the time interval or year considered. Then the joint density of $\mathbf{r}_{\phi j} = (r_{\phi j}^{(1)}, \ldots, r_{\phi j}^{(k)})$ for $\{r_{\phi j}^{(l)} : 1 + \xi_\phi (r_{\phi j}^{(l)} - \mu_\phi)/\sigma_\phi > 0, l = 1, \ldots, k\}$ is

$$h_{\phi j}^{(k)}(\mathbf{r}_{\phi j}) = \sigma_\phi^{-k} \exp\left\{ - \left[1 + \xi_\phi \left(\frac{r_{\phi j}^{(k)} - \mu_\phi}{\sigma_\phi} \right) \right]^{-1/\xi_\phi} - \right.$$
$$\left. - \left(1 + \frac{1}{\xi_\phi} \right) \sum_{l=1}^{k} \log \left[1 + \xi_\phi \left(\frac{r_{\phi j}^{(l)} - \mu_\phi}{\sigma_\phi} \right) \right] \right\}. \qquad (2.1)$$

Since we assume the wind process to vary smoothly over directions, we model the dependence of the parameters on direction ϕ by a continuous function. This requires a flexible function allowing for a broad range of possible variations, as well as it needs to satisfy circular boundary conditions. The functional relationship taken here is given by harmonic terms having the form

$$\tau_c(\phi) = a_c + \sum_{t=1}^{n_c} b_{ct} \cos(t\phi - w_{ct}), \tag{2.2}$$

with τ_c, $c = 0, 1, 2$, corresponding to the parameters ξ_ϕ, μ_ϕ, and σ_ϕ. For the model to be well defined the restrictions $b_{ct} \geq 0$ and $0 < w_{ct} \leq 2\pi$ are imposed while $a_c \in I\!R$. With the parameters of interest, namely ξ_ϕ, μ_ϕ, and σ_ϕ, being restated accordingly by a_c, b_{ct}, and w_{ct}, n_c is the number of harmonic terms necessary to account for the variation in direction. The model is therefore determined by a total number of $3 + 2\sum_{c=0}^{2} n_c$ parameters. Let N be the number of intervals, say years, and k denote the number of order statistics for a subset $\Omega \subset (0, 2\pi]$, then, assuming the wind speeds to be independent over different directions, the logarithm of the likelihood is

$$l(\boldsymbol{\vartheta}) = \sum_{\phi \in \Omega} \sum_{j=1}^{N} \log h_{\phi j}^{(k)}(\boldsymbol{r}_{\phi j}), \tag{2.3}$$

with $h_{\phi j}^{(k)}$ being the joint density given in (2.1). After substituting the parameters of the density by harmonic terms as given in (2.2) usual maximization procedures will supply parameter estimates of a_c, b_{ct}, and w_{ct}. Related standard errors are calculated from the observed Hessian $\mathbf{H}_O = -\nabla^2 l(\boldsymbol{\vartheta})$ evaluated at $\boldsymbol{\vartheta} = \hat{\boldsymbol{\vartheta}}$, where $\boldsymbol{\vartheta}$ denotes the vector of all parameters a_c, b_{ct}, and w_{ct}.

The alternative approach is to use component data, which implies a processing of data before analysing them. For each direction α the data consist of all values $\tilde{R}_\alpha = R_\phi \cos(\alpha - \phi)$ whenever $|\alpha - \phi|(\text{modulo } \pi) < \pi/2$ holds and 0 otherwise; R_ϕ represents a gust in direction ϕ. From these values the k largest ones of any direction contribute to the likelihood in the usual way. The dependence induced by the processing procedure does not alter the validity of using maximum likelihood estimation assuming independence as in (2.3) to obtain an asymptotically consistent estimate of $\boldsymbol{\vartheta}$; but the dependence needs, however, to be accounted for when calcu-

lating the standard errors of parameter estimates. Let $l(\boldsymbol{\vartheta})$ denote the logarithm of the likelihood as given in (2.3) stressing the dependence on parameters. Then, by applying an approximation using Taylor series expansion, the covariance matrix of $\hat{\boldsymbol{\vartheta}}$ becomes

$$\operatorname{cov}(\hat{\boldsymbol{\vartheta}}) \approx \mathbf{H}^{-1}\mathbf{V}\mathbf{H}^{-1}, \tag{2.4}$$

where $\mathbf{H} = -E(\nabla^2 l(\boldsymbol{\vartheta}))$ and $\mathbf{V} = \operatorname{cov}(\nabla l(\boldsymbol{\vartheta}))$; ∇ and ∇^2 denote gradient and Hessian, respectively. Dependence across directions invalidates the equality $\mathbf{H} = \mathbf{V}$. To estimate the covariance matrix the following method may be applied: let $h_{\phi j}^{(k)}$ denote the density of the k largest order statistics in year j; with the annual contributions $\mathbf{u}_j(\boldsymbol{\vartheta}) = \nabla \sum_{\phi\in\Omega} \log h_{\phi j}^{(k)}(\mathbf{r}_{\phi j})$ being independent and identically distributed random variables, the score vector can be restated as $\nabla l(\boldsymbol{\vartheta}) = \sum_{j=1}^{N} \mathbf{u}_j(\boldsymbol{\vartheta}) = \sum_{j=1}^{N} \left[\nabla \sum_{\phi\in\Omega} \log h_{\phi j}^{(k)}(\mathbf{r}_{\phi j})\right]$ and therefore its corresponding covariance matrix is given by

$$\mathbf{V} = \operatorname{cov}(\nabla l(\boldsymbol{\vartheta})) = N\mathbf{V}_{\mathbf{u}_j},$$

where $\mathbf{V}_{\mathbf{u}_j} = \operatorname{cov}(\mathbf{u}_j(\boldsymbol{\vartheta}))$. An apparent estimator of $\mathbf{V}_{\mathbf{u}_j}$ is

$$\hat{\mathbf{V}}_{\mathbf{u}_j} = \frac{1}{N}\sum_{j=1}^{N} \mathbf{u}_j(\hat{\boldsymbol{\vartheta}})\mathbf{u}_j(\hat{\boldsymbol{\vartheta}})'.$$

Substitution of $\mathbf{V}_{\mathbf{u}_j}$ by $\hat{\mathbf{V}}_{\mathbf{u}_j}$ and consequently $\mathbf{V}$ by $\hat{\mathbf{V}}$ as well as replacing the expected Fisher information matrix $\mathbf{H}^{-1}$ by its observed counterpart yields, when applying (2.4), an estimate of the desired covariance matrix.

2.1.2 Probabilistic assessment via quantiles

Traditionally, quantiles $G(r_p) = p$ or an equivalent formulation, frequently used in the context of extreme value statistics, return-levels $G(r^{(J)}) = 1 - 1/J$,

$$r_\phi^{(J)} = \mu_\phi - \frac{\sigma_\phi}{\xi_\phi}\{1 - [-\log(1 - 1/J)]^{-\xi_\phi}\}, \tag{2.5}$$

are the quantities of interest. There are two methods of calculating confidence intervals of return-levels, which are commonly applied. A detailed treatment of both methods in the simple case of non-directional modelling may be found at Coles

(2001). The first one, often referred to as delta method, is to construct a symmetric interval by employing the asymptotic normality of the estimated return-level; the corresponding variance is calculated via an approximation based on Taylor series expansion,

$$V_{r_\phi^{(J)}} \approx \mathbf{d}'\mathbf{V}_{\boldsymbol{\vartheta}}\mathbf{d}. \tag{2.6}$$

In (2.6) we have $\mathbf{d} = \nabla\varpi_\phi^{(J)}(\boldsymbol{\vartheta})$, with $\varpi_\phi^{(J)}(\boldsymbol{\vartheta}) = r_\phi^{(J)}$ denoting the return-level given in (2.5) stressing dependence on the vector of parameters $\boldsymbol{\vartheta}$.

The alternative approach to calculate confidence intervals is the so called method of profile likelihood, which is derived from a likelihood ratio test. We first express one parameter, say the constant a_1 of the harmonic term of μ_ϕ, as a function of the return-level $r_\phi^{(J)}$ and all remaining parameters. Using (2.5) and (2.2) this is

$$a_1 = r_\phi^{(J)} + \frac{\tau_{2\phi}}{\tau_{0\phi}}\left\{1 - [-\log(1-1/J)]^{-\tau_{0\phi}}\right\} - (\tau_{1\phi} - a_1), \tag{2.7}$$

where $\tau_{c\phi}$, $c = 0, 1, 2$, are the parameters according to (2.2) at the point ϕ. Maximization of the likelihood (2.3) after substitution of a_1, and maximizing over a reasonable range of return-level-candidates $r_\phi^{(J)}$ for every $\phi \in \{10, \ldots, 360\}$ yields, after comparison with the required quantiles of the χ^2-distribution, the desired confidence bands.

2.1.3 Probabilistic assessment of exceedances

While the preceding paragraph focuses on calculating extreme quantiles, the strategy here is to determine the exceedance probability for a given critical value, which in the subsequent application is the critical wind speed. Let v_{crit} be this critical value, then from (1.2) we get the probability of v_{crit} being exceeded in any one year by $P(R > v_{crit}) = 1 - G_{(\xi,\mu,\sigma)}(v_{crit}) =: \nu(\boldsymbol{\vartheta})$. In practice, however, the parameters are replaced by their estimates, which are subject to sampling error, and so, in turn, is the estimated probability of an annual maximum above the critical value. To assess the precision of the estimated exceedance probability, confidence bounds or bands are desirable.

One way to calculate confidence intervals is via the delta method. Using approximation (2.6) by replacing $\mathbf{d}$ with $\mathbf{d} = \nabla\nu(\boldsymbol{\vartheta})$ yields the variance of the exceedance probability. Because of the approximate normality, it is again straightforward to calculate confidence bounds. It is worth mentioning that this way of determining confidence intervals may result in a negative lower interval bound; the lower interval bound is set equal zero in this situation. When using component data the covariance matrix $\mathbf{V}_{\boldsymbol{\vartheta}}$ is replaced by (2.4).

It is also possible to apply the profile likelihood method to gain confidence intervals for the exceedance probability. Let $p = 1/J$ and replace $r_{\phi}^{(J)}$ with v_{crit}, then equation (2.7) can be restated as

$$a_1 = v_{crit} + \frac{\tau_{2\phi}}{\tau_{0\phi}} \left\{1 - [-\log(1-p)]^{-\tau_{0\phi}}\right\} - (\tau_{1\phi} - a_1). \tag{2.8}$$

For the calculation of the confidence intervals we now require p to vary across a reasonable interval and maximize the likelihood at each step. The profile likelihood intervals become the more asymmetric the more extreme the values are they are calculated for.

2.2 Simulation study

To investigate the performance of the model a simulation study is carried out. Due to the complex structure of the model direct simulation of the distribution of the r largest order statistics is not feasible. An alternative approach is to use the largest values of a distribution which is easy to simulate from and which has the same upper tail as the distribution of daily maxima. This can be achieved by employing the max-stability property of the GEV, restated here as

$$F(x) = G^{1/n}(x), \tag{2.9}$$

which yields a distribution F being again of extreme value type with a change in the parameters μ_{ϕ} and σ_{ϕ}, while ξ_{ϕ} remains the same. By taking $n = 365$ in (2.9), we do not assume the distribution F to be the distribution of daily maxima but it has the same upper tail as the latter. So for values of x near the upper endpoint we

can replace simulated values from the distribution of daily maxima by those of F. The simulation procedure then works as follows. We assume the true parameters to be the estimated values described in Section 2.3, see Table 2.3. As the data are discretized to 10°, we will have $\phi \in \{10, \ldots, 360\}$. For any direction ϕ the parameter values of ξ_ϕ, μ_ϕ, and σ_ϕ are re-calculated from the given parameters a_c, b_{ct}, and w_{ct} using (2.2). Thereafter, n values are simulated from the distribution F for each direction $\phi \in \{10, \ldots, 360\}$ constituting one block, which corresponds to one year; 22 blocks each of which has length $n = 365$ are simulated and joint together; thus simulated data correspond to the number of observed wind data. Finally, the r largest values for each combination of block and direction are extracted and used for model estimation.

As already mentioned, important quantities in applications are extreme quantiles. It is therefore sensible to judge the model by its return-levels. A natural approach is to first calculate the return-levels from simulated data for every point within 10° and 360°; then compute (pointwise) corresponding confidence bands for them; and finally check (again pointwise) whether or not the true values are lying within the confidence bounds. We simulate 200 samples.

In this simulation study we take the delta method yielding symmetric confidence intervals. In the following we use a model having a constant for the parameter ξ_ϕ, and one and four harmonic terms to describe variation in σ_ϕ and μ_ϕ, respectively; this model is abbreviated (0,4,1)-model in the subsequent. After having simulated data for any direction using the method described above, the parameters of a (0,4,1)-model and return-levels for 10, 50, 100, and 1000 years with corresponding 95%-confidence bands based on the delta method are estimated using maximum likelihood. Due to high computational costs the simulation size being 200 is rather small. However, we can recognize basic features and get an impression of the model's performance from this number of replicates.

The results are shown in Table 2.1; for every direction and any return-level the table states the number of values smaller than the lower interval bound in the upper line, while the number of cases exceeding the upper bound are given in the second line. There seems to be a slight systematic pattern of some neighbouring

ϕ	10°	20°	30°	40°	50°	60°	70°	80°	90°	100°	110°	120°	130°	140°	150°	160°	170°	180°	190°
10y	3	1	2	3	4	5	4	3	7	4	5	5	2	1	3	1	3	4	3
	8	7	9	10	8	7	8	8	7	8	7	6	7	7	7	5	4	7	7
50y	1	1	2	3	4	5	6	4	4	5	4	2	2	0	2	2	3	3	1
	8	8	8	10	9	8	9	6	7	9	6	5	6	8	7	4	4	5	7
100y	1	1	1	4	4	5	5	4	4	4	3	2	2	1	2	2	2	2	1
	8	9	8	10	9	11	10	7	7	9	8	8	6	7	8	5	5	5	5
1000y	1	1	1	3	4	4	4	3	1	2	2	2	1	0	1	2	2	1	1
	7	9	9	8	11	11	11	6	8	9	9	9	11	10	9	6	8	8	7

ϕ	200°	210°	220°	230°	240°	250°	260°	270°	280°	290°	300°	310°	320°	330°	340°	350°	360°
10y	1	0	0	0	1	2	3	3	2	2	4	2	3	5	7	6	6
	8	8	9	8	9	14	9	9	7	5	5	7	6	4	5	5	6
50y	1	1	0	0	0	0	1	2	2	3	4	2	3	3	4	4	3
	7	11	10	10	10	12	9	9	9	7	6	7	7	8	9	9	7
100y	2	1	0	0	0	0	1	2	2	3	3	2	3	3	3	3	2
	9	11	11	10	11	11	10	9	9	7	7	6	7	9	10	10	9
1000y	0	1	0	0	0	0	1	1	1	2	2	1	1	2	3	3	2
	7	9	10	9	11	9	9	10	8	7	8	7	7	10	9	10	10

Table 2.1: Results of Simulation study with 200 repetitions; number of cases being smaller than the lower confidence bound are shown in the upper line, while those exceeding the upper interval limit are seen in the lower line: for 10, 50, 100, and 1000 year return-level.

directions to have more values outside the required interval than others. However, as repeated simulations show the opposite phenomenon, we assume it to be random. A striking fact, in contrast, is that most points outside the interval are above the upper bound of the interval, and only a very little amount being smaller than the required bounds. This might be addressed by considering confidence intervals based on profile likelihood, which exhibit an asymmetric shape for extreme return-periods. More precisely, those plots show that the upper bound of the interval has a greater distance to the maximum likelihood point estimate than its lower counterpart. Due to the symmetry of intervals by application of the delta method, the number of points lying outside the interval must be higher for larger values.

2.3 Analysis of German wind data

The model described above is now applied to daily maximum wind speed data of the gauging station of Würzburg. To get an impression of the data, Figure 1.2 shows boxplots of the wind speeds of daily maxima for all directions as well as a histogram reflecting corresponding frequencies of directions. There is a clear pattern supporting the choice of a model for wind speeds which varies smoothly over directions.

A convenient feature of the k largest order distribution is its capability to incorporate different numbers of order statistics for different years or, as in our case, for different directions. The former case often arises when analysing data where just annual maxima are known for the first years, while in the later ones complete data are available; both data may then be analysed at the same time contributing to the same likelihood function. In the present case, the number of order statistics varies with different directions and over years. We restrict the number of largest observations k to be at most five, so each direction within each year contributes by $r \leq 5$ values per year. Table 2.2 shows the number of least available order statistics in any of the 22 years for each direction. For example, taking direction 20°: in each year there are at least two observations recorded. Directions indicated by NA are those having at least one year with no observation being made at all. For the subsequent analysis is based on at most the five largest values in each direction and year, those being five or greater are both indicated by ≥ 5. Data from directions indicated by NA are excluded from the analysis. The model is then estimated for

10°	NA	130°	r=1	250°	$r = 3$
20°	r=2	140°	r=2	260°	$r \geq 5$
30°	r=3	150°	r=2	270°	$r \geq 5$
40°	$r \geq 5$	160°	r=1	280°	$r \geq 5$
50°	r=4	170°	NA	290°	$r \geq 5$
60°	r=4	180°	NA	300°	r=2
70°	r=1	190°	NA	310°	r=2
80°	r=2	200°	r=1	320°	r=3
90°	r=3	210°	r=3	330°	r=1
100°	r=3	220°	r=4	340°	NA
110°	r=1	230°	$r \geq 5$	350°	NA
120°	r=2	240°	$r \geq 5$	360°	r=1

Table 2.2: Number of least available order statistics for each direction in any year for the Würzburg data; directions indicated by NA have at least one year without any one observation.

different numbers of harmonic terms for each parameter.

Model discrimination is carried out by employing a likelihood ratio test with a significance level of 5% using a forward selection procedure. As the location parameter is usually most sensitive, model selection starts with a (0,1,0)-model. Separately for each of the parameters ξ, μ, and σ one harmonic term is added, and the maximum change in log-likelihood is taken to yield the improved model if this change is significant according to a likelihood ratio test. The procedure terminates when none of the three models proposed results in a significant change in the log-likelihood. This favours a (0,4,1)-model, our final choice. Estimated parameters and related standard errors are given in Table 2.3. The shape parameter ξ is estimated by -0.197 with a standard error of 0.011. This gives a clear indication that the Gumbel model ($\xi = 0$) is not an appropriate choice in our case.

As we are investigating extreme events, return-levels are the quantities we are interested in. To assess precision of the estimation confidence-bands are calculated additionally. The two alternative possibilities are, as described in preceding parts, those based on the delta method and those using the profile likelihood method. Fig-

		raw data (0,4,1)	components (1,2,1)
$\hat{\xi}_\phi$	$\hat{a}_0$	-0.197 (0.011)	-0.106 (0.023)
	$\hat{b}_{01}$	NA	0.061 (0.022)
	$\hat{w}_{01}$	NA	1.014 (0.517)
$\hat{\mu}_\phi$	$\hat{a}_1$	14.451 (0.150)	20.061 (0.451)
	$\hat{b}_{11}$	4.843(0.157)	6.668 (0.380)
	$\hat{w}_{11}$	4.542 (0.047)	4.493 (0.049)
	$\hat{b}_{12}$	4.686 (0.212)	2.677 (0.278)
	$\hat{w}_{12}$	2.579 (0.036)	2.451 (0.078)
	$\hat{b}_{13}$	1.086 (0.168)	NA
	$\hat{w}_{13}$	0.797 (0.167)	NA
	$\hat{b}_{14}$	0.589 (0.172)	NA
	$\hat{w}_{14}$	3.619 (0.288)	NA
$\hat{\sigma}_\phi$	$\hat{a}_2$	3.733 (0.06)	2.705 (0.186)
	$\hat{b}_{21}$	0.890 (0.074)	0.793 (0.137)
	$\hat{w}_{21}$	4.536 (0.116)	4.603 (0.141)

Table 2.3: Estimated parameters for the (0,4,1)-model in case of raw data, and the (1,2,1)-model in case of component data; the number of harmonic terms are according to $(\xi_\phi, \mu_\phi, \sigma_\phi)$ for the gauging station Würzburg; standard errors are given in parenthesis.

ure 2.1 shows a plot of the 100-year return-level for the Würzburg data together with a 95%-profile likelihood confidence band. The equivalent graph with confidence intervals based on the delta method is shown in Figure 2.2. In both plots we super-imposed separate estimates based on data of that direction only (points). The strong variation of these points highlights the improvement of the harmonic model by allowing the transfer of information over directions over a sector by sector analysis.

The alternative approach is to use the component data described in Section 2.1.1. In this case we have $r = 5$ for any year and direction. Estimation results are given in Table 2.3. Now the shape parameter ξ depends on direction and is estimated by $-0.106 + 0.061\cos(\phi - 1.014)$. Again the estimation of the general level of ξ given by $a_0 = -0.106$ with standard error 0.023 shows that the Gumbel model cannot be

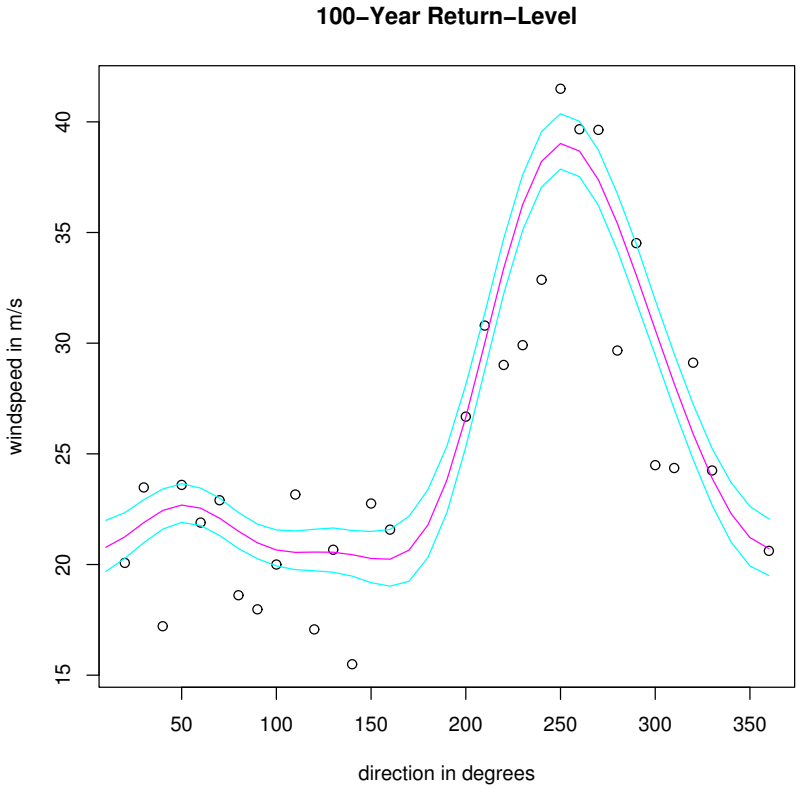

Figure 2.1: Plot of the (0,4,1)-model of Würzburg: ML–estimates for the 100-year return-level and 95%-profile likelihood confidence bands; points are estimated return-levels based on data of that direction only.

applied for all directions in our data.

For comparison of the two methods we have calculated the 100-year return-levels and corresponding confidence bands, see Figure 2.3. When using the likelihood ratio test for model selection using component data the reference distribution needs to be adjusted in order to account for dependencies across directions; for its calculation see Coles and Walshaw (1994). Applying this model selection procedure yields a (1,2,1)-model. One can see a higher overall level of the 100-year return-level for the latter model. This is due to the different definition of the problem, since in this case the components are analysed.

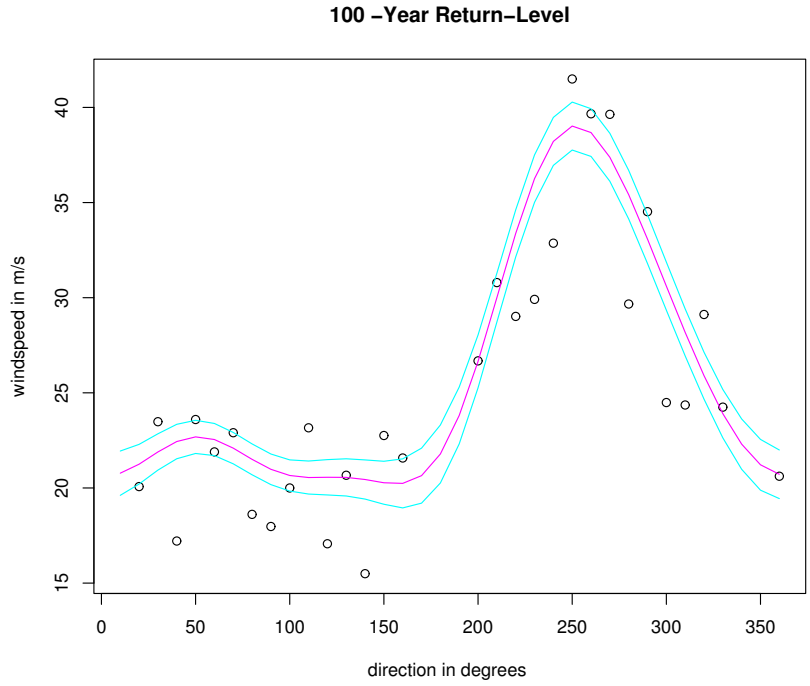

Figure 2.2: Plot of the (0,4,1)-model of Würzburg: ML–estimates for the 100-year return-level and 95%-confidence bands by the delta method; points are estimated return-levels based on data of that direction only.

If a critical wind speed value is known then interest is in calculating its probability of being exceeded. For these probabilities confidence intervals can be calculated by the delta method or by the method of profile likelihood as described in Section 2.1. Applying the delta method over all directions, resulting confidence intervals from models based on non-processed data are shown in Figure 2.4, while those based on component data are given in Figure 2.5. By just using non-processed data, profile-likelihood intervals for the exceedance probabilities of a critical value 38 m/s can be calculated for a fixed direction. Taking the direction where the highest wind speeds occurred, 260°, the profile-likelihood intervals for the two critical wind speeds 32 m/s and 42.7 m/s are shown in Figures 2.6 and 2.7 respectively. These plots indicate, that the profile-likelihood confidence intervals are getting the more asymmetric the

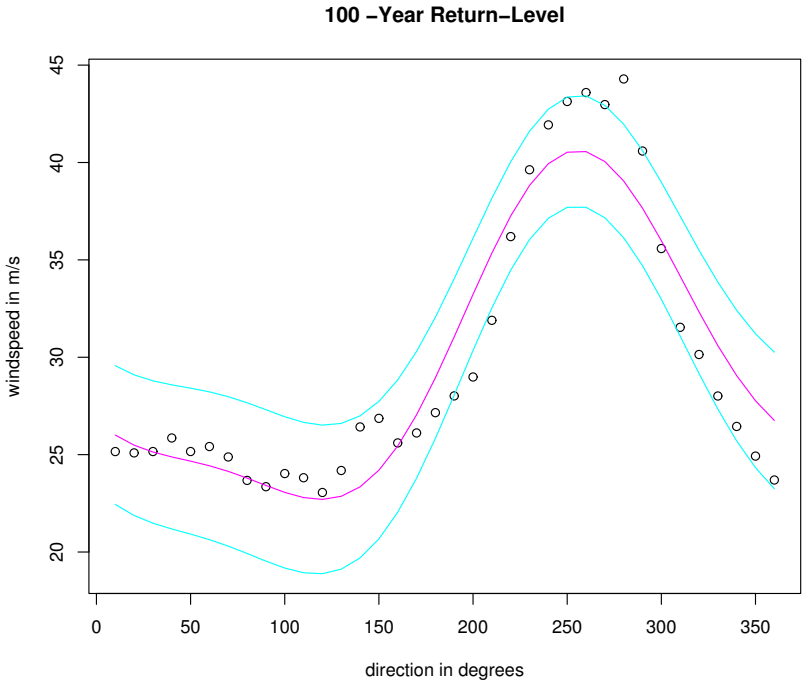

Figure 2.3: Plot of the (1,2,1)-harmonic model of Würzburg: ML–estimates for the 100-year return-level using component data and 95%-confidence bands by the delta method; points are estimated return-levels based on data of that direction only.

greater the critical value is. So for large critical values the approximation by a symmetric interval, such as is the case in the application of the delta method, is questionable. Unfortunately, this approximation is also anti-conservative leaving its applicant possibly expecting himself in a safer position than he actually is. For this reason we prefer the application of intervals based on the profile-likelihood method.

2.4 Aspects of masking

A common problem with wind data when considering directions is masking of gusts (Coles and Walshaw, 1994; Moriarty and Templeton, 1983). This problem is easiest understood by an example: there is a very strong gust from, say east, and at the

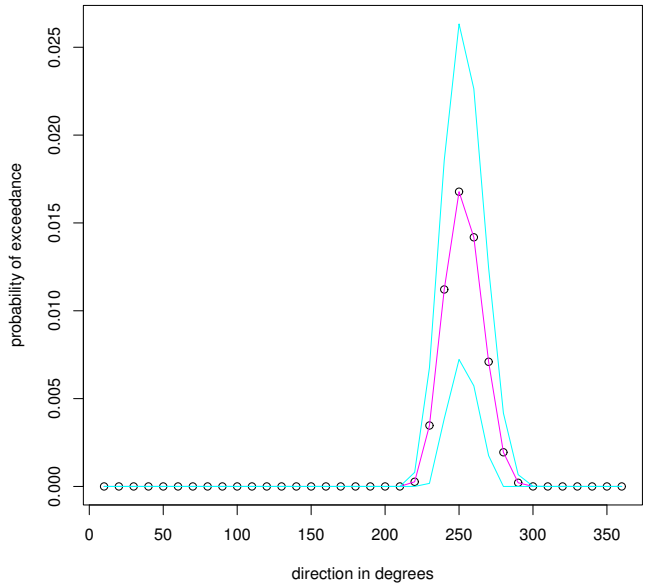

Figure 2.4: Plot of the exceedance probability of the critical wind speed 38m/s as a function of direction using the (0,4,1)-model based on unprocessed data.

same day a slightly stronger one from west. If maxima are recorded daily, the data contain the one from west, but that one of east, which might rank among the greatest ones of this direction, is lost. In this case we say that the gust from east was masked by the one from west and this may cause biased estimates. An immediate consequence of masking is the down-shift of many recorded values compared to the true, unknown ones. A reasonable assumption is therefore expecting return-levels to be underestimated. Moriarty and Templeton (1983) found in their analysis of directional sectors, using annual maxima only, that in many directions calculated return-levels rather overestimate the true values; they argue, that the most extreme values of the whole observation period in most directions are not masked, but lower ones of other years are, a consequence of which is a larger estimate of the scale parameter. Considering equation (2.5), a larger scale parameter, in turn, results in a larger return-level. In our case the effect is not clear.

We consider two different types of data: the non-processed data and derived com-

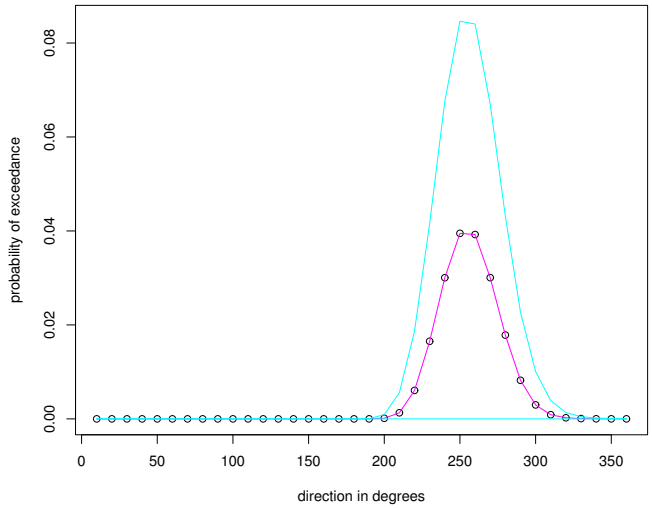

Figure 2.5: Plot of the exceedance probability of the critical wind speed 38m/s as a function of direction using the (1,2,1)-model based on component data.

ponent data. To assess the masking effect a comparison with maxima of shorter time intervals is desirable. For Würzburg, data of maxima within each ten–minutes interval are available for a small number of years. For these maxima also their precise direction is recorded in contrast to the data of daily maxima used before, where just the average direction of the hour the maximum occurred is known and therefore substituted.

We examine data of two years and compare the different types of data. Considering for each direction the largest observation of the original data and their counterparts of ten-minutes maxima clearly reveals the presence of the masking effect (see Figures 2.8 a) and b)). A considerable number of daily recorded maxima lie well below their ten-minutes counterparts. This effect will be even stronger for higher order statistics and may cast doubt on the reliability of using model (2.1) based on non-processed daily maxima. A possibility to overcome this problem is suggested in the next chapter. In 1994 there are a few maxima of the raw data exceeding the ten-minutes data but are equivalent in size to the maximum of the latter in a neighbouring direction; this effect is due to the substitution of the maximum's direction by the average of

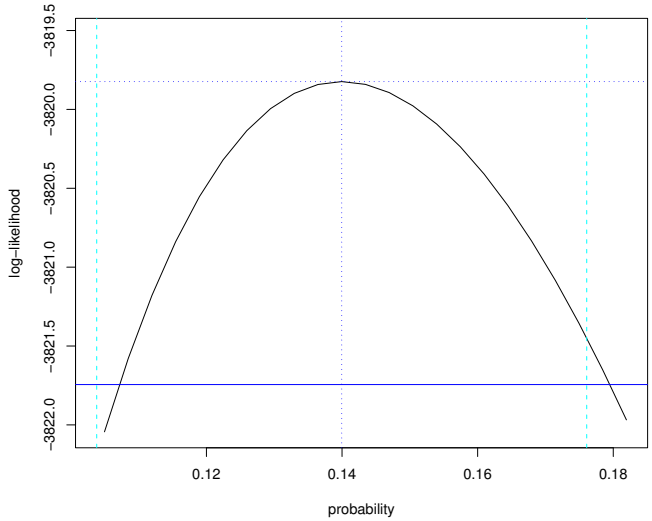

Figure 2.6: Plot of the exceedance probability of the critical wind speed 32m/s using the (0,4,1)-model; profile likelihood with horizontal line indicating the interval limits and dashed vertical lines indicating the corresponding interval based on the delta method .

its hourly direction used for daily maxima. It also indicates that the effect of this substitution is little.

Though ten-minutes values are sometimes slightly exceeded by corresponding components, Figure 2.8 d) shows acceptable agreement between their largest values and the corresponding ones of ten-minutes recordings. This is also true for 2.8 c), but less obvious due to the slight directional shift of the largest observation of daily maxima resulting from the substitution. Finally, comparing the component data with components of ten-minutes recordings shows quite good agreement (see Figures 2.8 e) and f)). In summary, while the non-processed data are heavily affected by masking the component data show much better agreement with those of far shorter time intervals.

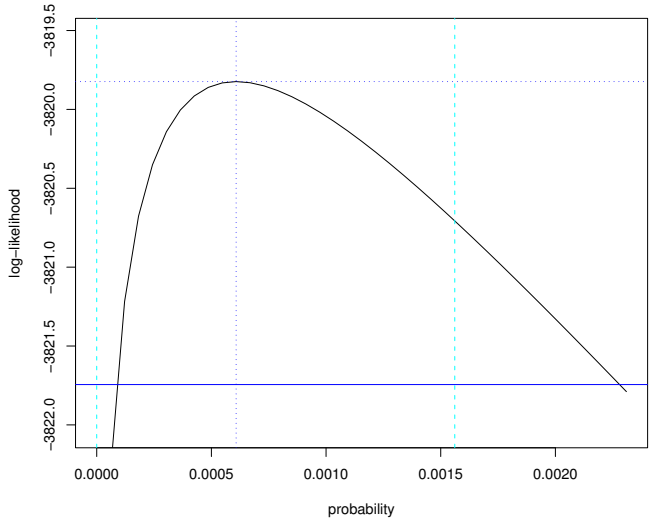

Figure 2.7: Plot of the exceedance probability of the critical wind speed 42.7 m/s using the (0,4,1)-model; profile likelihood with horizontal line indicating the interval limits and dashed vertical lines indicating the corresponding interval based on the delta method.

2.5 Discussion

To model extreme wind behaviour we applied a model extending the annual extreme value approach by employing the largest order statistics. Another possibility would be to consider exceedances of a suitably high threshold (Pandey, 2002; Pandey, Van Gelder and Vrijling, 2001). We have discussed a directional model for extreme value data and two quantities derived therefrom. The first one is using quantiles as commonly applied in the classical approach, while the second one is based on the exceedance probability. As estimates of the model parameters are subject to sampling variation, so are these quantities themselves. A natural way to account for this uncertainty is to calculate confidence intervals providing us with the precision of the estimate under consideration. The two most important methods of calculating intervals, the delta method and the profile likelihood method, are dealt with in detail.

The model has been applied to two types of data. In both cases, a fixed number of order statistics in each direction were extracted for parameter estimation. In the

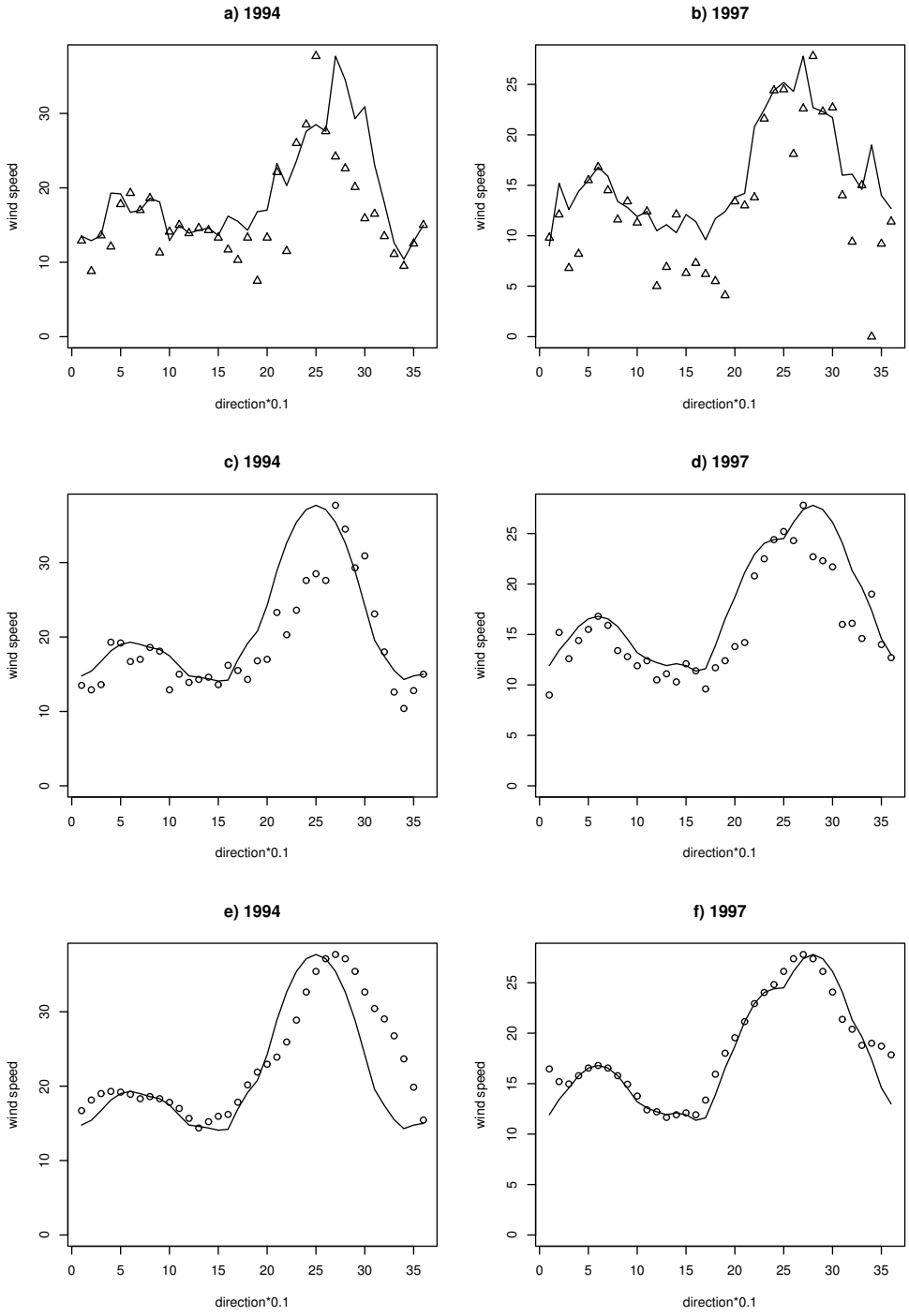

Figure 2.8: The largest observations for each direction are shown both for the ten-minutes data (line) and the original data (triangles) for the year 1994 in a) and for 1997 in b). In c) and d) the maxima of components (line) are compared with ten-minutes maxima (points) for 1994 and 1997. For both years the maxima of components (line) are compared with components of ten-minutes maxima in e) and f).

first approach data from different directions without being processed are used. One has to be cautious when taking this approach. One problem is, that in a number of directions there are only very few observations, so the asymptotics may not hold to justify both the application of the extreme value model used and the assumption of the estimates being normally distributed. A further problem is possible dependencies, both between data of neighbouring directions and successive data. Unlike many other analysis working with hourly observations, the present one uses daily maxima - a much longer period - so the problem of dependence within the data is not that critical when applying maximum-likelihood. Finally, the study of masking has shown that a considerable number of occurrences are missed and not available for model estimation. All these problems cast doubt on the reliability of this model based on unprocessed daily maxima.

By using components and calculating the variance adjusted for this situation, directional dependencies are accounted for. Furthermore, the problem of scarcity of data in some directions is not present any more, so asymptotic arguments apply in the usual way. This advantage is on expense of straightforward calculation: components have to be computed as well as the variance needs to be adjusted; additionally, a reference distribution for model discrimination via a likelihood ratio test has to be calculated, which implies a high amount of additional computing cost. The study of the masking effect has shown that components of daily maxima may serve as an acceptable substitute for data of far shorter time intervals, which further supports the application of models based on these data.

Considering the most extreme observation of the whole investigation period, which is 42.7 m/s in direction 260°, is around the upper limit of the confidence band of the 1000 year return-level when looking at the ordinary model, and far away of the estimate of the corresponding 100-year return-level. This again makes the practical applicability of using a model with non-processed data questionable. A comparison of the largest observation with the 100-year return-level using the component model proves this model to supply far more plausible and reliable results.

For model estimation we have used the method of maximum likelihood. An alternative approach is Bayesian estimation. A review of these methods may be found in

Coles and Powell (1996), Coles and Tawn (1996) give an approach for determining the prior distribution based on expert knowledge, Van Gelder (1996) shows how historical knowledge can be incorporated, and an application to wind data is Walshaw (2000). Furthermore, the calculation of confidence intervals of extreme quantiles by Bayesian methods also yields asymmetric intervals; the results are comparable to those of profile-likelihood when non-informative priors are used.

In many situations of extreme wind speeds a Gumbel distribution is applied (Cook, Harris and Whiting, 2003). This is justified by the fact, that the Weibull distribution often serves as a good approximation to the wind data at hand. Pre-analyses have shown that the Weibull distribution does not well describe the distribution of our data. Furthermore, the application of the generalized extreme value distribution, which includes the Gumbel as a special case, clearly excludes the latter distribution and supports the application of the more general family in our situation. Coles and Pericchi (2003) show that even in cases where a Gumbel is justifiable on statistical grounds its application, instead of the more general GEV, is a risky strategy.

The analyses carried out in this chapter can be used for probabilistic assessment at a track of the German rail. Since there are no wind measurements available it is assumed that the wind at the track differs from that at a close weather station by a constant factor. This factor is determined by meteorological methods. The critical wind speed perpendicular to the track is determined by technical considerations; its probability of being exceeded has then to be estimated. These exceedance probabilities can be estimated by the methods described in this study. Furthermore, we can give confidence intervals for this probabilities. These local estimates serve as a sensible input for an overall measure of the whole track, which adds up functions of the estimated quantities over all points. So the methods give substantial improvement of overall measures compared to those employing empirical quantiles of the wind speed distribution.

Chapter 3

A model for the masking problem

3.1 Introduction

A well-known complication when analysing extreme wind data taking directionality into account is referred to as masking problem. The analysis in Chapter 2 suggests that masking has an impact on resulting model estimates for the data we have analysed, and a visual comparison with data from shorter time intervals in Section 2.4 gives further insight to this phenomenon. The masking problem is due to the recording mechanism. It occurs when the recording interval is large, and just the maximum wind speed of the interval and its direction is recorded. In this case, there might have been a strong gust which ranks among the largest in its own direction, but it is not recorded, since the maximum over the interval was in another direction. Consequently, the recorded data do not necessarily contain the true largest ones in all directions.

Assuming that we have additionally to the observed time interval maxima knowledge about wind events within sub-intervals, we can include this information to reduce the masking effect. In the present case we assume observations to be daily wind speed maxima together with their corresponding directions, and additionally we have the empirical distribution of directions of ten-minutes maxima.

Directional models for extreme wind speeds are commonly based on the largest observations in each direction for a fixed time period (see Chapter 2) or on exceedances

of high directional thresholds. These observations represent maxima over all directions of each time interval. This, however, implies that wind speeds of all other directions in the same time interval are equal or less than the observed one. The approach suggested here is to additionally include this information on other directions to increase the accuracy of estimated return-levels.

The approach presented requires knowledge of the occurrence distribution over directions in sub-intervals. For some gauging stations this knowledge is available at least for a short time period. In this case the empirical directional occurrence distribution may be taken as the true one. In cases of weather stations without knowledge about the directional occurrence distribution in sub-intervals, corresponding data of a nearby weather station might be employed. This is, however, just reasonable, if the model is not too sensitive to deviations from the correct distribution.

In the following we describe the new approach considered here and a classical equivalent, and describe how corresponding quantiles are calculated. A simulation study is carried out to compare these two models under different conditions: ideal conditions, mis-specification of directional probabilities, and serial correlation. Performance of the models is judged by their mean square errors. Thereafter the two models are applied to real data.

3.2 The model

Let the pair (R, Φ) denote the daily maximum wind speed R having direction Φ, where $R \in [0, \infty)$, $\Phi \in \Omega \subset (0, 2\pi]$, and $R_\phi = (R|\Phi = \phi)$ describes the daily maximum wind speed given its in direction ϕ. To describe the distribution in a sub-interval of time we use the pair (S, Θ), where S refers to the wind speed in the sub-interval and Θ its direction. As the directional sectors for the sub-intervals are in the present situation the same as for days, we have $\Theta \in \Omega$. We use the directional occurrence probability of daily maxima $p_\Phi(\phi) = P(\Phi = \phi)$, and for the sub-interval data $p_\Theta(\theta) = P(\Theta = \theta)$, which in the present context refer to the probability of the whole sector they identify.

Daily Maxima Model

For a gust R_ϕ in direction ϕ exceeding a certain high threshold u_ϕ we assume the generalized extreme value distribution $G_\phi(r) = G_{(\xi_\phi,\mu_\phi,\sigma_\phi)}(r)$ given by (1.2) to be an appropriate model for extreme wind speeds. We assume the parameters ξ_ϕ, μ_ϕ, and σ_ϕ to vary smoothly over directions $\phi \in (0, 2\pi]$. Thus a functional relationship of parameters on direction is imposed on the distribution of wind speeds to account for this variation. The relationship is given by harmonic terms.

We start by considering the likelihood of a model just using daily maxima. Let u_ϕ be a high threshold in direction ϕ, and let $p_\Phi(\phi)$ be the probability of a daily maximum occurring in direction ϕ; then equation (1.2) provides an approximation above u_ϕ to the distribution of R in direction Φ

$$\begin{cases} g(r,\phi) &= g_\phi(r)p_\Phi(\phi), & r \geq u_\phi \\ G(u_\phi,\phi) &= G_\phi(u_\phi)p_\Phi(\phi), & r < u_\phi, \end{cases} \tag{3.1}$$

where $g_\phi(r) = g(r|\phi)$ is the density corresponding to $G_\phi(r)$; the index ϕ in the previous density and distribution function is used as an abbreviation for the GEV-parameters $(\xi_\phi, \mu_\phi, \sigma_\phi)$ in direction ϕ for the distribution given in (1.2). The likelihood is now constructed under the assumption of independence over directions and over days. Define $\{r_{exc,\phi}\}$ and $\{r_{bel,\phi}\}$ to be the sets of all pairs (r, ϕ) with r above and below the threshold u_ϕ, respectively, then the log-likelihood may be written as

$$\begin{aligned} l_D &= \sum_{\phi\in\Omega} \log\left(\prod_{\{r_{exc,\phi}\}} g(r,\phi) \prod_{\{r_{bel,\phi}\}} G(u_\phi,\phi) \right) \\ &= \sum_{\phi\in\Omega} \log\left((G(u_\phi,\phi))^{N_{r_{bel},\phi}} \prod_{\{r_{exc,\phi}\}} g(r,\phi) \right), \end{aligned} \tag{3.2}$$

where $N_{r_{bel},\phi}$ denotes the number of daily maxima in direction ϕ below the threshold u_ϕ. If there is no interest in the distribution $p_\Phi(\phi)$, or $p_\Phi(\phi)$ is assumed to be known,

then the above likelihood is proportional to

$$\begin{aligned} l_D &\propto \sum_{\phi\in\Omega} \log\left(\prod_{\{r_{exc,\phi}\}} g_\phi(r) \prod_{\{r_{bel,\phi}\}} G_\phi(u_\phi) \right) \\ &= \sum_{\phi\in\Omega} \log\left((G_\phi(u_\phi))^{N_{r_{bel},\phi}} \prod_{\{r_{exc,\phi}\}} g_\phi(r) \right). \end{aligned} \tag{3.3}$$

Directional variation in parameters is allowed for by including equation (2.2).

In extreme value analysis interest is commonly in high quantiles often termed return-levels. The period corresponding to a return level is usually stated on an annual scale. The daily maxima distribution for each direction corresponds to the number of days of observed daily maxima in that direction, which varies over years. To transform to an annual scale, we assume a Poisson distribution with parameter according to the annual average number of observations $\bar{N}_\phi = n_\phi/N_y$ in direction ϕ, where n_ϕ is the total number of observations in direction ϕ, and N_y is the total number of years. The annual distribution derived from the daily maxima model is therefore given by

$$\begin{aligned} D_\phi(x) &= \sum_{i=0}^{m} \frac{e^{-\bar{N}_\phi}(\bar{N}_\phi)^i}{i!}(G_\phi(x))^i \\ &\approx \sum_{i=0}^{\infty} \frac{e^{-\bar{N}_\phi}(\bar{N}_\phi)^i}{i!}(G_\phi(x))^i = \exp\left\{-\bar{N}_\phi(1-G_\phi(x))\right\}, \end{aligned}$$

where m is the number of days per year and the approximation is reasonable as $\bar{N}_\phi << m$. Return-levels or quantiles are calculated in the usual way by inversion of the above equation. So the quantile corresponding to $D_\phi(x_\phi^{(q)}) = q$ is

$$x_\phi^{(q)} = \mu_\phi + \left\{ \left[-\log\{1 + \frac{\log(q)}{\bar{N}_\phi}\} \right]^{-\xi_\phi} - 1 \right\} \frac{\sigma_\phi}{\xi_\phi}. \tag{3.4}$$

Subinterval Model

In the following we build up the likelihood for a directional extreme value model accounting for masked observations. The key idea here is to incorporate the knowledge of wind speeds in directions other than the actual observation, which need to be less than or equal to the interval maximum. We further assume that values in sub-intervals are independent and the wind-speed distribution of S_θ varies with direction θ. We use S_ϕ to mean $S_\theta|_{\theta=\phi}$ when referring to the wind-speed distribution in sub-intervals considered in the direction ϕ where the observed interval maximum occurred. The contribution of the observed interval maximum r, having occurred in direction ϕ, together with the unobserved occurrences in all directions and other sub-intervals of the same time-interval to the likelihood is

$$\begin{cases} g^*_\phi(r)\prod_{i=1}^{n-1} P(S_i \le r), & r \ge u_\phi \\ G^*_\phi(u_\phi)\prod_{i=1}^{n-1} P(S_i \le r), & r < u_\phi, \end{cases} \tag{3.5}$$

where n is the number of sub-intervals in a day or larger time-interval. Here G^*_ϕ, with density g^*_ϕ, corresponds to the upper tail of the sub-interval distribution; again we use G^*_ϕ as an abbreviation for $G^*_\theta|_{\theta=\phi}$ and it is given by (1.2) with $G^*_\theta = G_{(\xi^*_\theta,\mu^*_\theta,\sigma^*_\theta)}$. Also $P(S_i \le r)$ represents the probability of the maximum of the sub-interval, which is known to be no greater than the maximum of the whole interval. If we assume that the distribution of directions, $h_\Theta(\theta)$, or the relative frequency of directional sectors, $p_\Theta(\theta)$, is known, and sub-interval maxima are independent, then

$$\begin{aligned} P(S_i \le r) &= \int_0^{2\pi}\int_{z=0}^{r} f^*(z|\theta)h_\Theta(\theta)dzd\theta \\ &= \int_0^{2\pi} F^*(r|\theta) \quad h_\Theta(\theta) \quad d\theta \approx \sum_{\theta\in\Omega} F^*(r|\theta)p_\Theta(\theta), \end{aligned} \tag{3.6}$$

where $f^*(r|\theta)$ and $F^*(r|\theta)$ are, respectively, the density and distribution function of the sub-interval wind speed given direction θ. With interest in the upper tail, we replace F^* by G^* in the previous equation. More precisely, for an observation r in direction ϕ, we use $G^*_\theta(r)$ if r exceeds the directional threshold u_θ and $G^*_\theta(u_\theta)$ if r is not greater than u_θ.

Partition the data as $\{r_\phi\} = \{r_{exc,\phi}\} \cup \{r_{bel,\phi}\}$, then the likelihood is given by

$$
\begin{aligned}
l_T = l_T(\boldsymbol{\vartheta}) = -\sum_{\phi\in\Omega} \Big\langle & \\
& n_{exc,\phi} \log(\sigma^*_\phi) + (1 + 1/\xi^*_\phi) \sum_{\{r_{exc,\phi}\}} \log(1 + \xi^*_\phi \frac{r - \mu^*_\phi}{\sigma^*_\phi}) + \sum_{\{r_{exc,\phi}\}} \left[1 + \xi^*_\phi \frac{r - \mu^*_\phi}{\sigma^*_\phi}\right]^{-1/\xi^*_\phi} \\
& + \sum_{\{r_{bel,\phi}\}} \left[1 + \xi^*_\phi \frac{u_\phi - \mu^*_\phi}{\sigma^*_\phi}\right]^{-1/\xi^*_\phi} \\
& - (n-1) \sum_{\{r_\phi\}} \log\left(\sum_{\theta\in\Omega} G^*_\theta(\max\{r, u_\theta\}) p_\Theta(\theta)\right) \Big\rangle, \qquad (3.7)
\end{aligned}
$$

where $n_{exc,\phi}$ is the number of elements of the set $\{r_{exc,\phi}\}$. The second line of the above likelihood represents all observations above the threshold, the third accounts for those observations below the threshold, while the last line includes the information of unobserved occurrences in the sub-intervals. Harmonic terms given by (2.2) are again employed to allow for directional variation in the model parameters; the corresponding parameter vector of all harmonic terms is denoted by $\boldsymbol{\vartheta}$.

To calculate return-levels on an annual scale for the subinterval model we transform its distribution in two steps: first we transform to a daily scale and thereafter to an annual scale. To transform the subinterval model to daily scale, we have to take into account, that the number of sub-interval occurrences in each direction varies from interval to interval. We assume the number of daily observations in direction θ to follow a Poisson distribution with parameter $Np_\Theta(\theta)$, where N is the number of sub-intervals per time-interval (e.g. one day). Then the daily distribution in direction θ can be approximated by

$$
\begin{aligned}
H_\theta(x) &= \sum_{i=0}^{N} \frac{e^{-Np_\Theta(\theta)} (Np_\Theta(\theta))^i}{i!} (G^*_\theta(x))^i \\
&\approx \sum_{i=0}^{\infty} \frac{e^{-Np_\Theta(\theta)} (Np_\Theta(\theta))^i}{i!} (G^*_\theta(x))^i = \exp\left\{-Np_\Theta(\theta)(1 - G^*_\theta(x))\right\}.
\end{aligned}
$$

With interest in the annual maxima distribution $[H_\theta(x)]^m$, where m is the number of days per year, the above approximation yields

$$[H_\theta(x)]^m \approx \exp\left\{-Nmp_\Theta(\theta)(1-G^*_\theta(x))\right\}.$$

Return-levels or quantiles are calculated again by inversion of the above equation. Let $\lambda_\theta := Nmp_\Theta(\theta)$ and G^*_θ follow a GEV with parameters ξ^*_θ, σ^*_θ, and μ^*_θ, then the quantile corresponding to $[H_\theta(x^{(q)}_\theta)]^m = q$ is

$$x^{(q)}_\theta = \mu^*_\theta + \left\{\left[-\log\{1+\frac{\log(q)}{\lambda_\theta}\}\right]^{-\xi^*_\theta} - 1\right\}\frac{\sigma^*_\theta}{\xi^*_\theta}. \tag{3.8}$$

3.3 Comparison of the two models

We consider now the behaviour and features of the new model. A natural way is to compare the classical approach based on (3.3) with the new one given by (3.7). As the objective of extreme value analysis is commonly the estimation of high quantiles, it is sensible to take this quantity for comparison. We choose the mean square error to measure the performance of an estimator. Define m_q to be the ratio of both mean square errors, that is

$$m_q := \frac{\mathrm{mse}(\hat{x}^{(q)}_{l_T})}{\mathrm{mse}(\hat{x}^{(q)}_{l_D})},$$

where $\hat{x}^{(q)}_i$ is a high quantile estimated by maximum likelihood with the likelihood indexed by $i = \{l_T, l_D\}$. This ratio will then quantify superiority or inferiority of the new model according to whether or not m_q is smaller or greater than one.

For a comparison of the two models considered, we simulate data and judge model performance according to the ratio of mean square errors m_q. The data are simulated from an extreme value distribution with some parameters varying over directions. The simulation size is fifteen years with each day consisting of 144 observations. Thereafter, parameters of the daily maxima model and the subinterval model are estimated. While the daily maxima model is allowed to have as many parameters as sensible according to a selection procedure based on likelihood ratio tests, the subinterval model is restricted to have the same structure as is used for the simula-

tion.

We first study how the subinterval model performs under the conditions assumed for it, that is independence for wind speeds and independence for wind directions, where the occurrence distribution is taken to be uniform over directions. The assumption of exact knowledge of the underlying occurrence distribution, $p_\Theta(\theta)$, as well as the independence assumption may not hold for real data. We therefore consider separately sensitivity to mis-specification of the directional distribution, $p_\Theta(\theta)$, and robustness of the model in the presence of serial correlation.

3.3.1 Model comparison under ideal conditions

The simulation is carried out under idealized circumstances. We simulate the subinterval data to be independent of one another and that all distributions remain the same over time. Furthermore, the true distribution of occurrence numbers is taken to be uniform so $p_\Theta(\theta)$ is constant over directions θ.

For the subinterval model with likelihood l_T we choose an exponential link to ensure positivity, that is $\sigma_\theta = \exp(\tau(\theta))$ with τ defined in (2.2), which is also employed for the simulated data. In the following the notation (n_ξ, n_μ, n_σ)-harmonic terms is used to describe the number of harmonic terms n_i for parameter i. We first simulate 15 years with each day consisting of 144 ten-minutes data from a (0,1,0)-harmonic model having the parameters $\boldsymbol{\vartheta} = (-0.15, 15, 3, \pi, \log(3.5))$; so $\xi = -0.15$ and $\sigma = \log(3.5)$ are constant, while $\mu(\theta) = 15 + 3\cos(\theta - \pi)$ changes with direction θ. The sub-interval occurrence distribution of directions, $p_\Theta(\theta)$, is taken to be uniform.

Both models just use daily maxima and associated directions as input-data. For the comparisons we estimate a daily maxima model based on l_D, where a forward-selection procedure is applied to find the model fitting best. From this model, estimated quantiles are derived for the subsequent comparison. The subinterval model uses the same data, but requires the additional information of $p_\Theta(\theta)$. The true distribution for directional frequencies is used here for the subinterval model, that is $p_\Theta(\theta)$ is uniform. Furthermore, we just allow the subinterval model to have

Direction	$x^{(5)}$	$x^{(15)}$	$x^{(100)}$
90°			
$\text{mse}_T/\text{mse}_D$	0.340	0.354	0.380
180°			
$\text{mse}_T/\text{mse}_D$	0.395	0.270	0.236
270°			
$\text{mse}_T/\text{mse}_D$	0.323	0.310	0.324
360°			
$\text{mse}_T/\text{mse}_D$	0.020	0.087	0.280

Table 3.1: Relative efficiency m_q under ideal conditions.

a (0,1,0)-harmonic structure. The choice of thresholds u_θ is based on the generated sub-interval data; we choose u_θ to be the upper 0.001-quantile, where its calculation is based on all simulated sub-interval data in direction θ. Return-levels for four selected directions are considered (an example of which is Figure 3.1), and three particular return-levels (5-years,15-years, and 100-years), imposed as vertical lines, are chosen for the comparison. The simulation size is 500.

Table 3.1 shows the efficiency m_q defined previously. All values are considerably smaller than one, so there is apparent superiority of the subinterval model over the daily maxima model. By symmetry of the harmonic term with respect to the strongest wind direction $\pi = 180°$ and constant $p_\Theta(\theta)$, one might expect the two directions 90° and 270° to be identical. However, the differences are rather small and can be attributed to randomness.

The results also indicate that accuracy in directions tending to have smaller values and therefore produce just a few daily maxima (360° in the current case) is highly increased by the subinterval model. The superiority in directions with few observations might also be due to some sensitivity of quantiles based on equation (3.4). There occurred a few missing values for the 5-year return-level in direction 360°, which were eliminated from the calculation. Since the location parameter μ_θ is smallest for this direction the resulting number of occurrences is the lowest over all

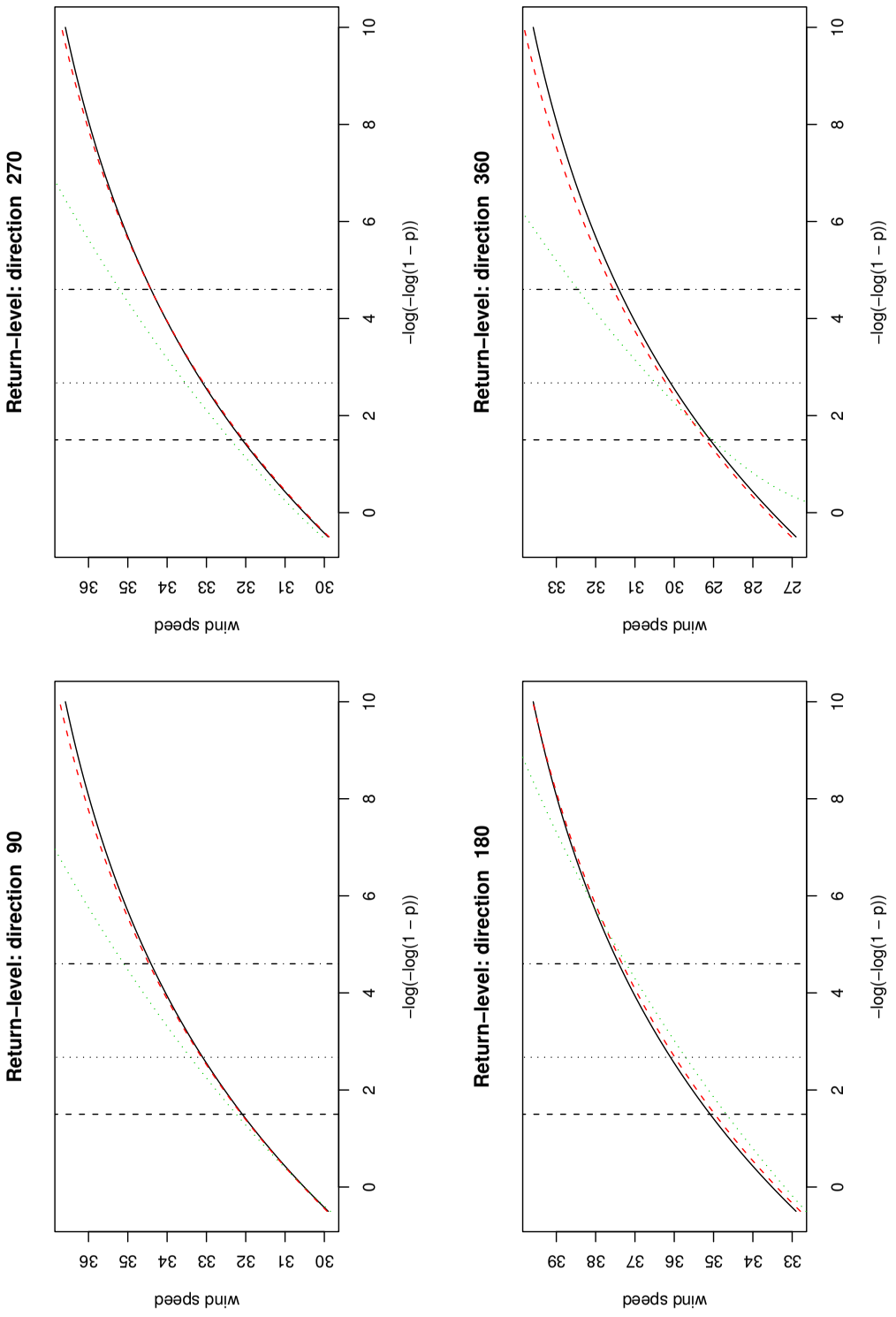

Figure 3.1: Return-levels using a 0.001-threshold for four different directions: true model (solid), daily maxima model (dotted), subinterval model (dashed); vertical lines correspond to return-levels for five years (left), 15 years (middle), and 100 years (right).

Direction	Squared Bias			Variance		
	$x^{(5)}$	$x^{(15)}$	$x^{(100)}$	$x^{(5)}$	$x^{(15)}$	$x^{(100)}$
90°						
TenMin	0.001	0.000	0.002	0.011	0.017	0.047
Daily	0.002	0.000	0.003	0.032	0.048	0.127
180°						
TenMin	0.000	0.000	0.003	0.010	0.016	0.046
Daily	0.000	0.000	0.000	0.026	0.059	0.206
270°						
TenMin	0.001	0.000	0.002	0.011	0.016	0.044
Daily	0.003	0.001	0.008	0.034	0.051	0.133
360°						
TenMin	0.001	0.001	0.001	0.013	0.017	0.046
Daily	0.337	0.034	0.001	0.358	0.176	0.169

Table 3.2: Squared bias and variance under ideal conditions.

directions. Note that for quantiles based on equations (3.4) and (3.8)

$$0 < 1 + \frac{\log(q)}{\lambda} < 1$$

needs to hold; the right inequality is naturally satisfied whenever $\lambda > 0$ and $q \in (0, 1)$; the left inequality requires $\lambda > -\log(q)$; for a 5-year return-level $\lambda > -\log(1 - 1/5) = 3.347/15$, so at least four observations are required to fulfill the constraint. The efficiency in 360° for a 15-year return-level is still far away from all other values, while the 100-year return-level is similar to those of other directions. So even when the constraint is fulfilled, return-levels based on (3.4) may be interpreted with caution when both the occurrence number and the return-period are small.

We now consider the bias and variance contributions to the mean square errors; this is shown in Table 3.2 where, for reasons of comparison, the bias is squared. For the daily maxima model in direction 360° and small return-levels, both bias and variance are large for reasons discussed above. For all other directions considered here, the major part leading to the size of the mean square error is due to the variance and

not to the bias contribution. This is a surprising result. With the masking of data possibly leaving some large observations in several directions un-recorded, this may lead to the assumption of derived estimates being biased.

3.3.2 Wrong directional probabilities

The directional occurrence distribution, $p_\Theta(\theta)$, plays an important role in the subinterval model. As $p_\Theta(\theta)$ may not always be known in real applications, it is natural to think of substituting it with a corresponding one from a neighbouring station, assuming that differences are little. This is, however, just sensible if the model is not too sensitive to departures from the true distribution. We therefore analyse the robustness of the subinterval model when the directional distribution, $p_\Theta(\theta)$, is not correctly specified.

As the *true* model for directional frequency, $p_\Theta(\theta)$, we use again a uniform distribution. From this model data are simulated. To analyse model departures, we define the $p_\Theta(\theta)$ that we use in the likelihood to follow a von Mises distribution vM(η, κ), where η specifies location while κ is a scale parameter (Mardia and Jupp, 2000; Fisher, 1993). Our choice is the von Mises distribution vM(π, κ) with different $\kappa = 0$, 0.1, 0.3, 0.5, 0.7, 0.9, 1.0. When $\kappa = 0$ we have a uniform distribution, while a bigger κ corresponds to a more concentrated distribution with a higher density at its mode $\eta = \pi$; Figure 3.2 shows the densities of this different distributions. The distribution of wind speeds is taken to be the same (0,1,0)-harmonic GEV–model used above.

Figure 3.3 shows in each plot the change of relative efficiency with the change of κ for the different directions 90°, 180°, 270°, and 360°. For $\kappa = 0$ the relative efficiency reflects the correctly specified directional distribution, while the departure of correct specification increases with κ. The three different plots differ in the value of the quantile, which were chosen as before.

For the shortest return-period, 5 years, m_q is almost the same for the two directions 90° and 270°, approximately constant and considerably smaller than one for all κ values. A possible explanation for this constant good performance is that for this

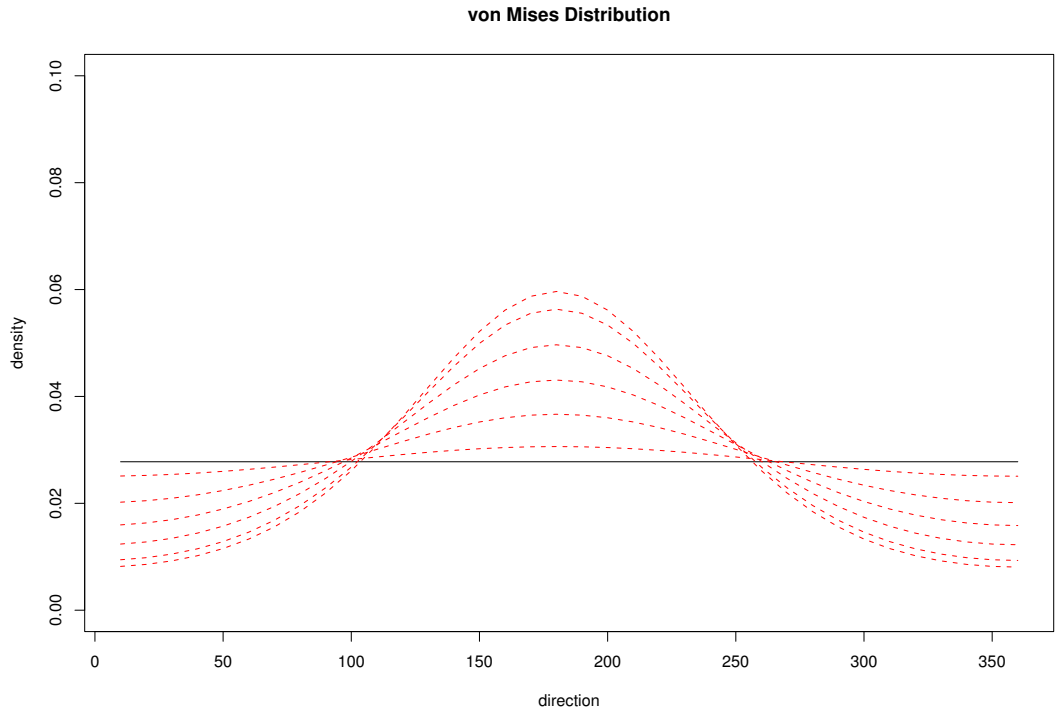

Figure 3.2: Von Mises distribution over (0,360°] with κ =0, 0.1, 0.3, 0.5, 0.7, 0.9, 1. κ=0 is the uniform distribution and the higher κ the more concentrated the distribution.

two directions the occurrence-probability is almost unchanged (compare Figure 3.2). The relative efficiency of the strongest direction 180° is below one for κ smaller than 0.6. The opposite direction 360° is close to 0, as both squared bias and variance of the daily maxima model (see Figure 3.4) are very high for reasons discussed above.

The 15-year return-level corresponds to the period the simulated data represent. Apart from m_q in direction 180°, which starts exceeding 1 for a κ around 0.7, all relative efficiency curves keep below this boundary for all κ values. The values of m_q for the weakest wind speed direction 360° are smaller than the corresponding values of m_q for the strongest wind speed direction 180° for all values of κ. However, a closer inspection at the bias - variance plots, see Figure 3.5, gives more insight to this phenomenon. For the subinterval model the squared bias for 180° is smaller than for direction 360° for all κs. However, the squared bias and variance for the

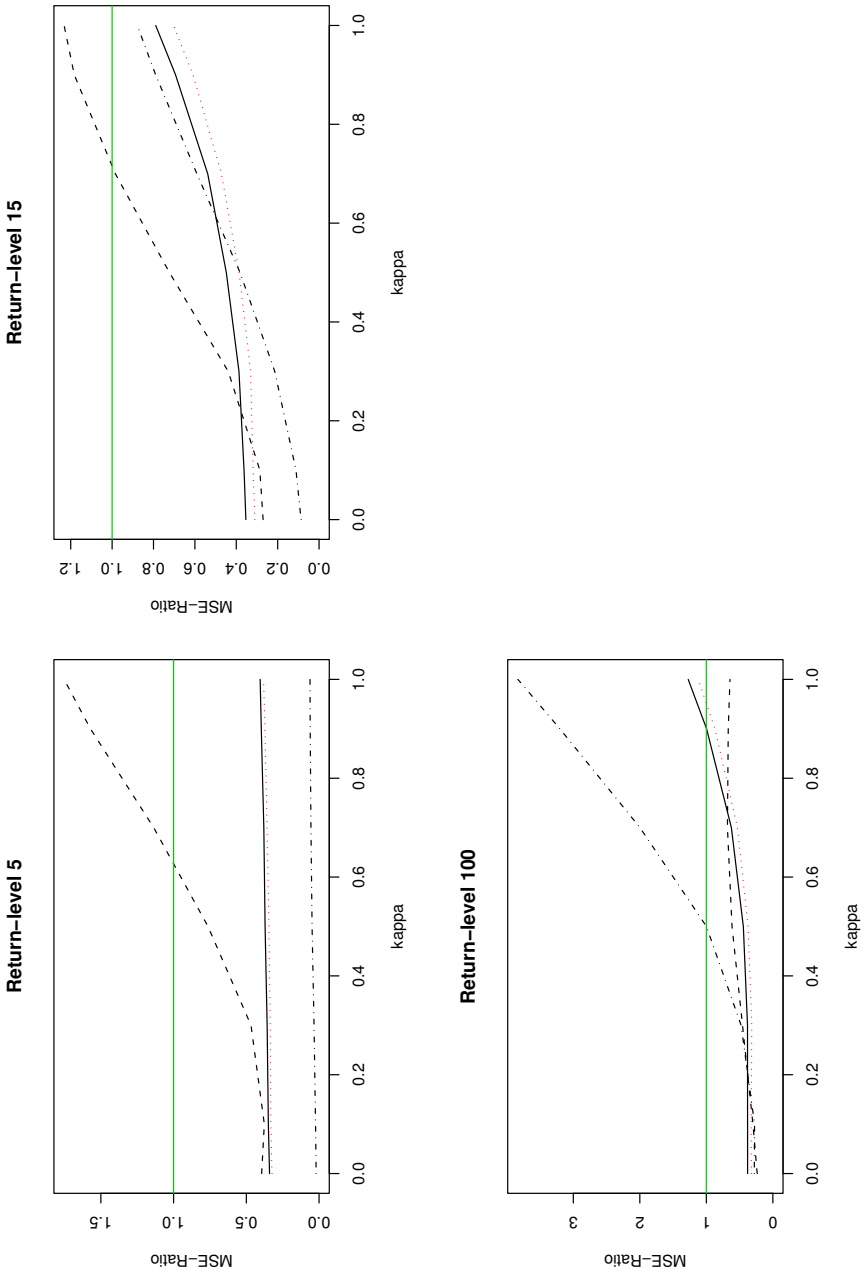

Figure 3.3: Relative efficiency m_q for mis-specified $p_\Theta(\theta)$ with κ =0, 0.1, 0.3, 0.5, 0.7, 0.9, 1, for four different directions (solid: 90°, dashed: 180°, dotted: 270°, dash-dotted: 360°) in each plot, and three different return-levels.

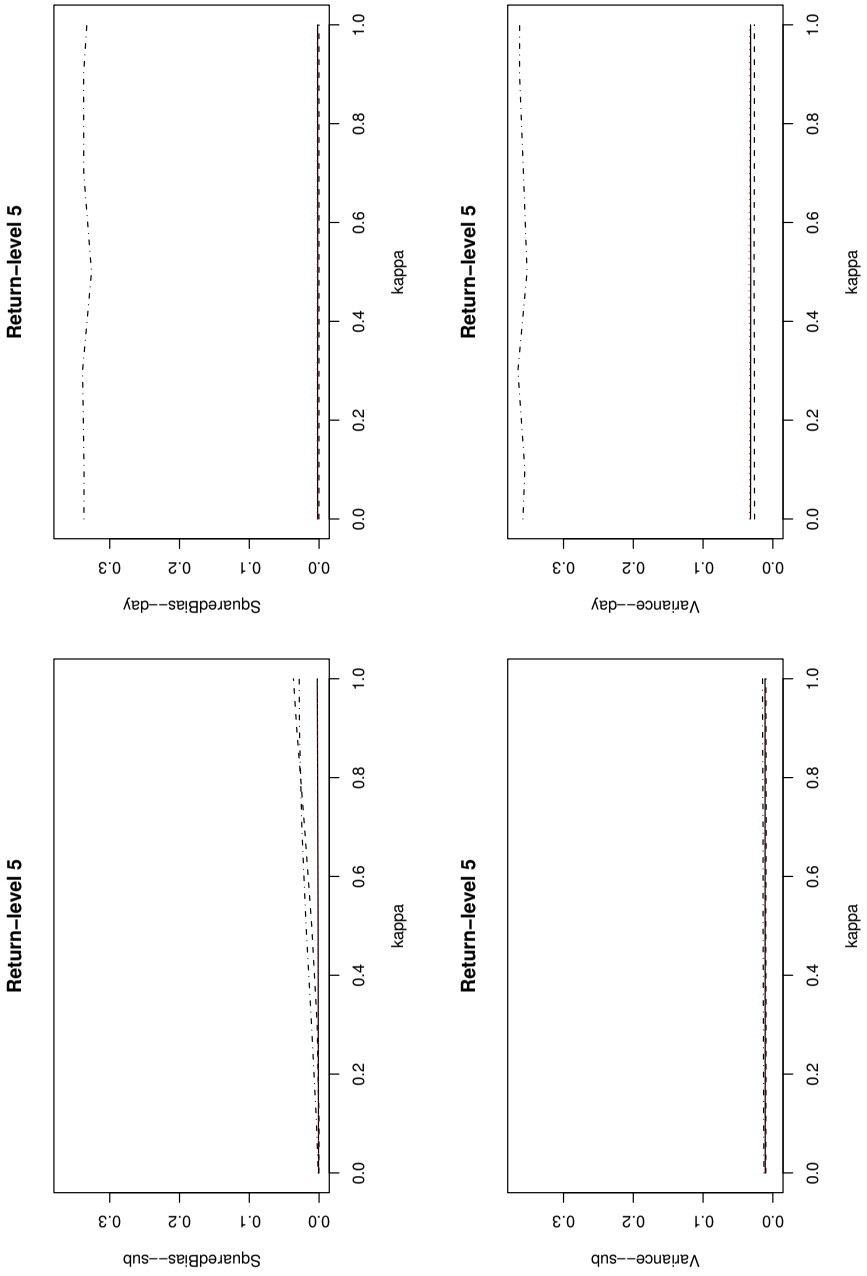

Figure 3.4: Squared bias (top row) and variance (bottom row) for κ =0, 0.1, 0.3, 0.5, 0.7, 0.9, 1, for four different directions (solid: 90°, dashed: 180°, dotted: 270°, dash-dotted: 360°) in each plot for a 5-year return-level; subinterval model (left) and daily maxima model (right).

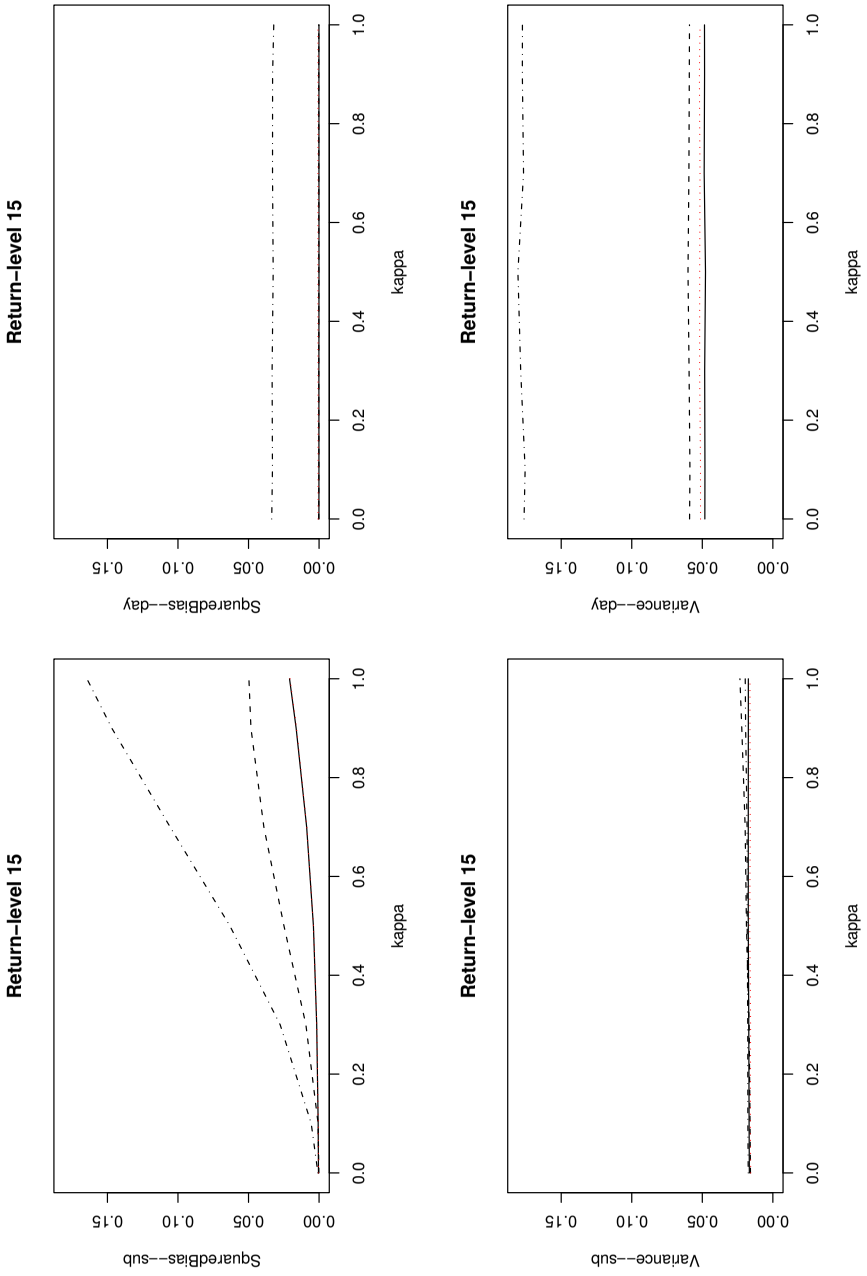

Figure 3.5: Squared bias (top row) and variance (bottom row) for κ =0, 0.1, 0.3, 0.5, 0.7, 0.9, 1, for four different directions (solid: 90°, dashed: 180°, dotted: 270°, dash-dotted: 360°) in each plot for a 15-year return-level; subinterval model (left) and daily maxima model (right).

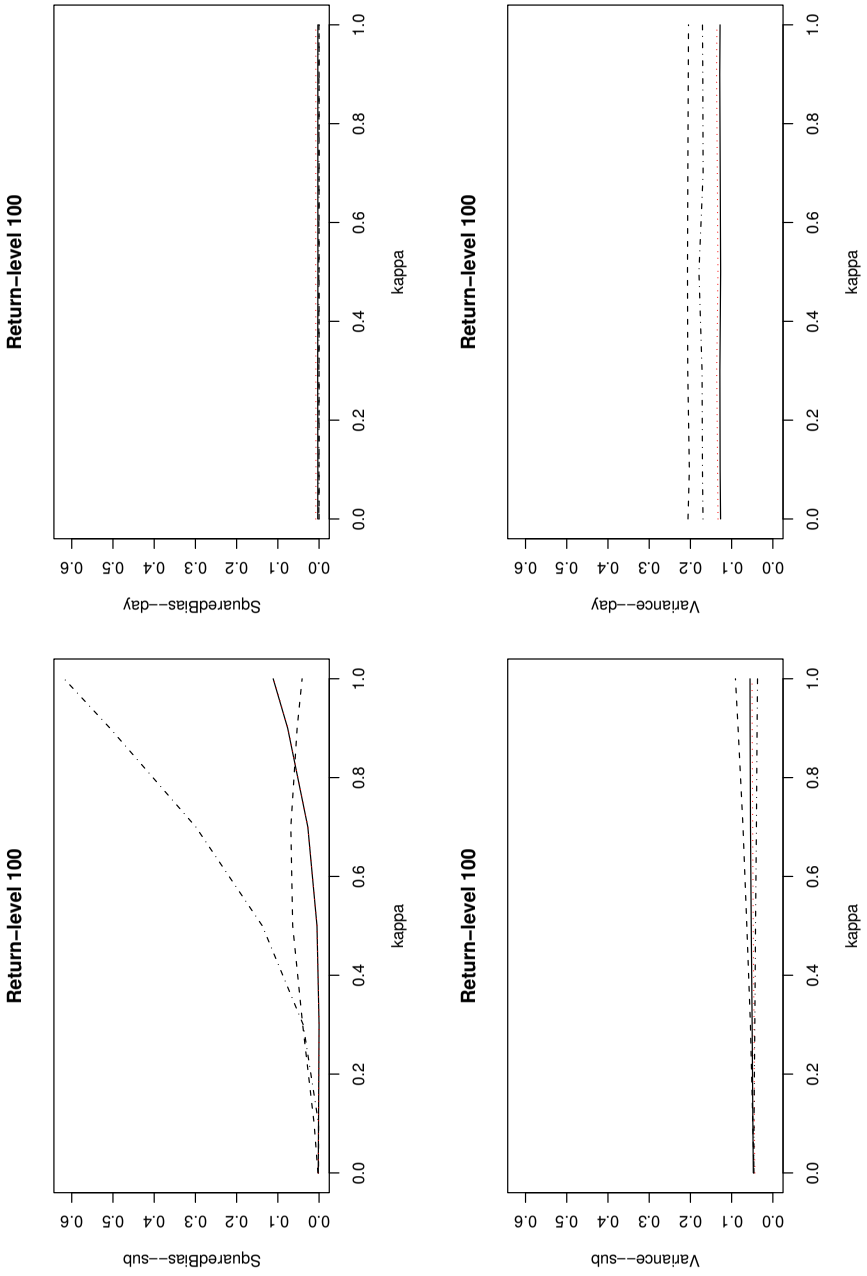

Figure 3.6: Squared bias (top row) and variance (bottom row) for κ =0, 0.1, 0.3, 0.5, 0.7, 0.9, 1, for four different directions (solid: 90°, dashed: 180°, dotted: 270°, dash-dotted: 360°) in each plot for a 100-year return-level; subinterval model (left) and daily maxima model (right).

daily maxima model are considerably bigger for 360° than for 180° leading to the surprising effect when comparing the efficiency.

We consider now the 100-year return-level, which is probably the most important one in real applications. The ratio of mean square errors for direction 180° is permanently below one while the opposite direction 360° exceeds this value for a κ around 0.5. The two directions 90° and 270° perform again quite well apart from for very high values for κ.

In conclusion, the subinterval model is not very sensitive to departures of correctly specified $p_\Theta(\theta)$. A closer look at Figure 3.2 suggests, that superiority of the daily maxima model is just present in cases where model departures are strong. Thus, the values for κ chosen here do rather highlight the often still superior behaviour of the subinterval model under extreme model departures.

3.3.3 Serial correlation

The assumption of independence is not valid in most applications, serial correlation is present instead. As we use independence as a working assumption for the subinterval model, it is necessary to study the behaviour of the model when this assumption is violated. We therefore simulate data, which exhibit dependence from one observation to the next both in direction and in speed. We first consider the directional distribution and thereafter the conditional wind speed distribution. The performance of the subinterval model is again judged by comparison with the daily maxima model based on the ratio of mean square errors.

Let us consider dependence for the direction. We allow a change from one direction just to a neighbouring one each with a probability p^Θ. Let $p_{i+1|i}(\theta_{i+1}|\theta_i) = P(\Theta_{i+1} = \theta_{i+1}|\Theta_i = \theta_i)$ then the probabilities for possible states of Θ_{i+1} are

$$p_{i+1|i}(\theta_i + 10^\circ|\theta_i) = p_{i+1|i}(\theta_i - 10^\circ|\theta_i) = \frac{1}{2}(1 - p_{i+1|i}(\theta_i|\theta_i)) = p^\Theta, \tag{3.9}$$

where circular boundary conditions have to be taken into account. The smaller p^Θ

the less likely the process is to leave its current state inducing higher directional dependence. With first simulating a starting directional value for Θ_1 from a discrete uniform distribution over $\{10, \ldots, 360\}$ the subsequent simulation of $\Theta_2, \Theta_3, \ldots$ is straight forward.

The conditional distribution of wind speeds $S|\Theta$ is modelled as a time series with marginal distribution $G_{(\xi_\theta, \mu_\theta, \sigma_\theta)}$ and with Markov dependence structure given by a bivariate normal copula; a detailed treatment of copulas is given by Nelsen (1999) and Joe (1997). For simulation of dependence in wind speeds, the dependence structure of a simple normal autoregressive model is employed. More precisely, consider

$$Z_i = \alpha Z_{i-1} + \epsilon_i, \qquad i = 1, 2, \ldots, \tag{3.10}$$

where $\epsilon_i \sim N(0, (1 - \alpha^2))$ and $\alpha \in (0, 1)$. Having simulated a starting value for Z_0 from a standard normal and ϵ_i from a normal distribution with mean zero and variance $(1 - \alpha^2)$, recursive calculation yields $Z_1, Z_2, \ldots$. Let Ψ denote the distribution function of a standard normal distribution and $G^{-1}_{(\xi_\theta, \mu_\theta, \sigma_\theta)}(x)$ the quantile function of the extreme value distribution in direction θ. Then after simulation of Z_i and Θ_i, we first transform the Z_i to standard uniform margins $U_i = \Psi(Z_i)$, and in a second step transform back to the required extreme value distribution

$$S_i = G^{-1}_{(\xi_\theta, \mu_\theta, \sigma_\theta)}(U_i) = G^{-1}_{(\xi_\theta, \mu_\theta, \sigma_\theta)}(\Psi(Z_i)). \tag{3.11}$$

Increasing dependence is reflected by a larger value of α.

In order to carry out a comparison of the subinterval model with the daily maxima model, we employ again the mean square error of a certain return-level. This requires knowledge of the true return level. The dependence introduced by the simulation scheme requires calculation of return-levels to be adjusted to account for the reduction in effectively independent information. A well-known measure of dependence at extreme levels is the extremal index. For an application of the extremal index to directional extremes we refer to Robinson and Tawn (1997). In the present case we employ a representation of the extremal index suggested by O'Brien (1987)

Direction	$(0.3, 0.8)$	$(0.2, 0.95)$	$(0.18, 0.99)$	$(p^{\Theta}(Z), 0.99)$
90°				
$\text{mse}_T/\text{mse}_D$	0.496	0.565	0.586	0.528
180°				
$\text{mse}_T/\text{mse}_D$	0.656	0.727	0.576	0.390
270°				
$\text{mse}_T/\text{mse}_D$	0.517	0.552	0.634	0.493
360°				
$\text{mse}_T/\text{mse}_D$	0.518	0.524	0.687	0.503

Table 3.3: Relative efficiency for $x^{(100)}$ in the presence of serial correlation

adapted to the directional nature of our data, given by

$$\delta_\theta = P(\max_{j=2,\ldots,J}\{S_j I(\Theta_j = \theta)\} < u_\theta | S_1 > u_\theta, \Theta_1 = \theta) \tag{3.12}$$

with u_θ being a high value in direction θ, and I is the indicator function. Equation (3.12) states the conditional probability, that given any one occurrence exceeds its directional threshold u_θ, none of the subsequent $J-1$ values exceed u_θ if they occur in direction θ. The true return-levels are calculated using equation (3.8) adjusted for dependence by taking $\lambda_\theta = (Nmp_\theta)\delta_\theta$. In the subsequent application δ_θ turned out to vary considerably, so we used a very high number of simulated values first to compute the extremal index. In equation (3.12) we choose $J = 144$ corresponding to a period of one day.

We consider three different combinations of parameters for directional dependence and wind speed dependence, where dependence is increased for both simultaneously. The chosen values for the pair (p^{Θ}, α) are (0.3, 0.8), (0.2, 0.95), and (0.18, 0.99). We simulate 500 times repeatedly from a 15-year period using the same distribution for wind speeds as above. The threshold for calculation of the extremal index is chosen to be identical with the threshold u_θ used for subsequent model estimation.

Table 3.3 shows the ratio of mean square errors of the 100-year return-levels for the three different degrees of dependence described previously. All cases show clear

superiority of the subinterval model over the daily maxima approach. There is slight suggestion of m_q rising with dependence, but it is neither strong nor present in all directions.

From studying real wind data, it is not an uncommon phenomenon that directional dependence increases with strong wind events. We therefore consider a further simulation mechanism given by a latent Gaussian prossess $\{Z_t\}$, given by (3.10), with S_t and Θ_t both determined by Z_t and hence dependent. Specifically for $\Theta_t|Z_t$ we take the random walk of (3.9) with $p_i^{\Theta}(Z_i) = (1 - \Psi(Z_i)) \cdot 0.4 + 0.1$; so for large Z_t, p^{Θ} is small and hence the Θ_t value is more likely to be equal to Θ_{t-1}. Also $S_t|(\Theta_t, Z_t)$ has the form (3.11). This simulation scheme reflects increasing directional dependence with increasing quantiles of the wind speed distribution. For the process generating the Z_i an $\alpha = 0.99$ was chosen. Resulting ratios of mean square errors, shown in the last column of Table 3.3, confirm the sub-interval model performing well under this changed conditions.

In summary, the simulation study has shown superiority of the subinterval model over a daily maxima approach. Its performance is best under iid conditions and a correctly specified occurrence distribution. For deviations from ideal conditions, the model still shows good results suggesting it to be reasonably robust. In the simulation study we have just used one particular choice for the wind speed distribution. To get further insight it would be useful to analyse a range of different models.

3.4 Application to data

We now apply the two models considered to wind data at hand. These data exhibit seasonality with the strongest gusts occurring during the winter period. As our aim is to present clearly the features of the new model, we avoid seasonality and restrict our analysis to the winter period including the months November, December, and January.

We consider two data situations: a data set of daily maxima where additionally a ten-minutes data set for a shorter observation period is available; another data set of daily maxima without additional subinterval information. For the daily maxima

model we employ in both situations a forward selection procedure using likelihood ratio tests, as successive days appear to be roughly independent. In the case of the subinterval model we just suggest a discrimination procedure in the situation where sub-interval information is available which is based on bootstrap methods.

We first consider wind data from Würzburg, where ten-minutes data are available for a period of ten years. The occurrence distribution of directions for the subinterval model, $p_\Theta(\theta)$, is taken to be the empirical distribution of the directions of ten-minutes data. Selection of the subinterval model is based on a bootstrap procedure, which is described in the following. As sub-interval data are available it is sensible to make use of them. For selecting the number of harmonic terms in the subinterval model, we use a block bootstrap with replacement, taking blocks of whole days from the ten-minutes data. The blocks are joined to give the equivalent of ten years of data, this being the same period as the observed ten-minutes data. Though the simulation study did not show the model to be very sensitive to departures from a correctly specified $p_\Theta(\theta)$ and serial correlation, this approach allows us to re-estimate $p_\Theta(\theta)$ from the bootstrap sample so that it preserves the dependence between sub-intervals within the data. As we restrict the analysis just to the winter season, each year consists of three months. The subinterval model is then estimated for a selected number of harmonic terms. This process of simulating and estimating is repeated 100 times to give 100 estimates for each parameter included in the model. We now explain the selection process, where the parameters of one harmonic term are either jointly discarded or kept together in the model.

As we assume the model to correctly describe the underlying process, a harmonic term actually present in the underlying process will not just have a positive amplitude, but also a unique location parameter, which is fixing the position of the harmonic term. So by repeated block bootstrap simulation, estimates of the location parameter should be highly concentrated around one certain value. In contrast, if these estimates show a different behaviour, like a high scattering over the whole range $(0, 2\pi]$, this harmonic term is not likely to be present in the underlying process. So we use the variation of the location parameter to judge on whether to keep the harmonic term in the model or not. In difficult cases the size of the amplitude parameter may be considered as well.

We start the selection procedure by restricting the maximum number of harmonic terms for each parameter to be three. Then, having simulated and estimated repeatedly, the largest harmonic term of each parameter is considered. Applying the selection criteria just described, we decide on whether or not the harmonic term is retained in the model. This selection process yields a (2,2,2)-harmonic model in the present case, which is the model we continue working with.

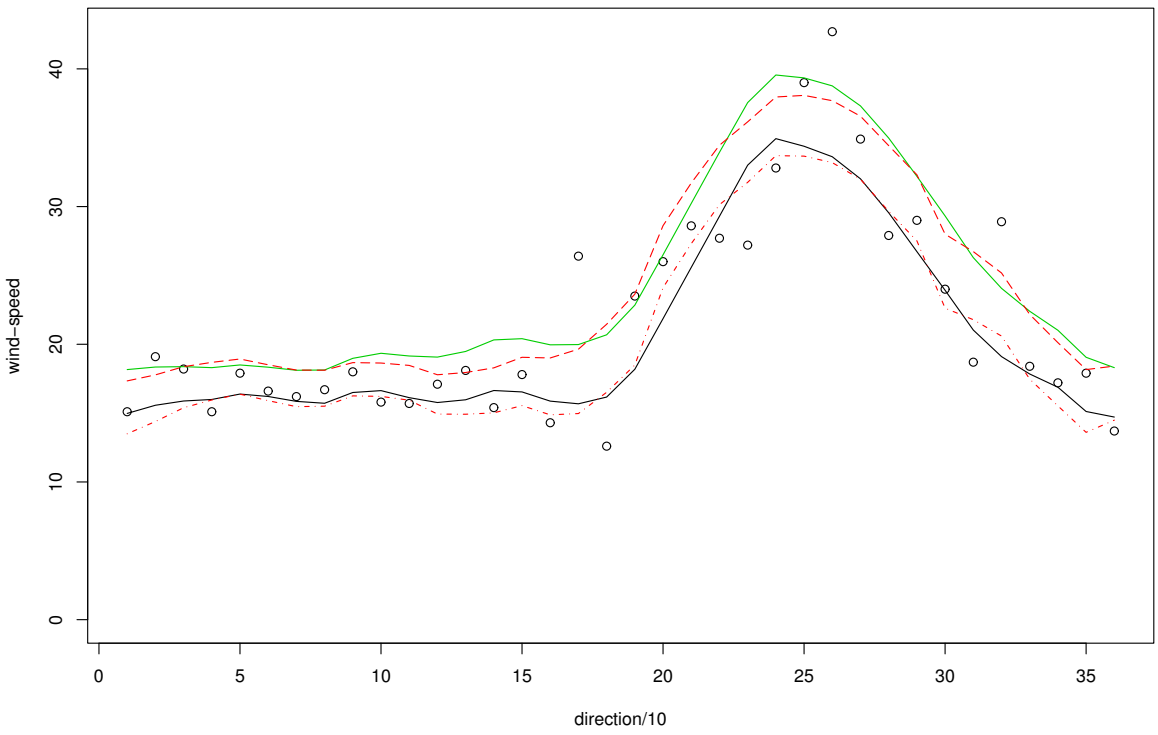

Figure 3.7: (Würzburg) Return-level estimates based on the (2,2,2)-harmonic subinterval model: 100-year return-levels (upper solid line) and 22-year return-levels (lower solid line); return-level estimates for the daily maxima model: 100-year return-levels (dashed line) and 22-year return-levels (dashed-dotted line).

The selected subinterval model is applied to the 22 years of daily wind-data of

Würzburg restricted to the winter period. The occurrence distribution of directions, $p_\Theta(\theta)$, is taken to be the empirical distribution from the ten years of ten-minutes data. Figure 3.7 shows estimates of 100-year and 22-year return-levels for the subinterval model as upper and lower solid lines respectively, and the largest observation for each direction over the 22 years observation period as circles. As well in the graphs are the corresponding return-levels from the daily maxima model, which was found to be a (0,2,1)-harmonic model by the forward selection procedure; with 100-year return-levels dashed, and 22-year return-levels dotted-dashed. In general, both methods seem to describe the behaviour of extreme wind speeds well. The 100-year return-level of the subinterval model is closer to the largest observation (42 m/s; 260°) than the corresponding one of the daily maxima model, indicating slight improvement.

Let us consider data from Hannover. The directional distributions of daily data for Würzburg and Hannover are given in the histograms in Figure 3.8. The two histograms show reasonable similarity making it likely that their corresponding directional ten-minutes distributions are not differing too much. Thus we estimate the subinterval model for Hannover taking the ten-minutes directional occurrence distribution from Würzburg, and the 22 years of daily maxima from Hannover. A reasonable subinterval model in terms of number of parameters, but still a flexible choice, is the (1,3,2)-harmonic subinterval model, which is used for Hannover. Applying this model, estimates of the 100-year return-level are given as upper solid line in Figure 3.9, as well as the 22-year return-level (lower solid line) corresponding to the observation period. Observed maxima of the daily data for each direction are super-imposed in the same plot as circles. For comparison, we include return-levels based on the same daily data from the (1,2,1)-harmonic daily maxima model found by the forward selection procedure. Corresponding 100-year and 22-year return-levels are included as dashed and dashed-dotted lines in the plot.

A comparison of the 22-year return-levels with maxima of the observation period shows the sub-interval model to reflect the structure much better than the daily maxima model. The stronger wind events in eastern directions (around 90°) are well captured. Especially, when considering the 100-year return-level, the large observation in direction 270° appears not to be so unlikely as for the daily maxima

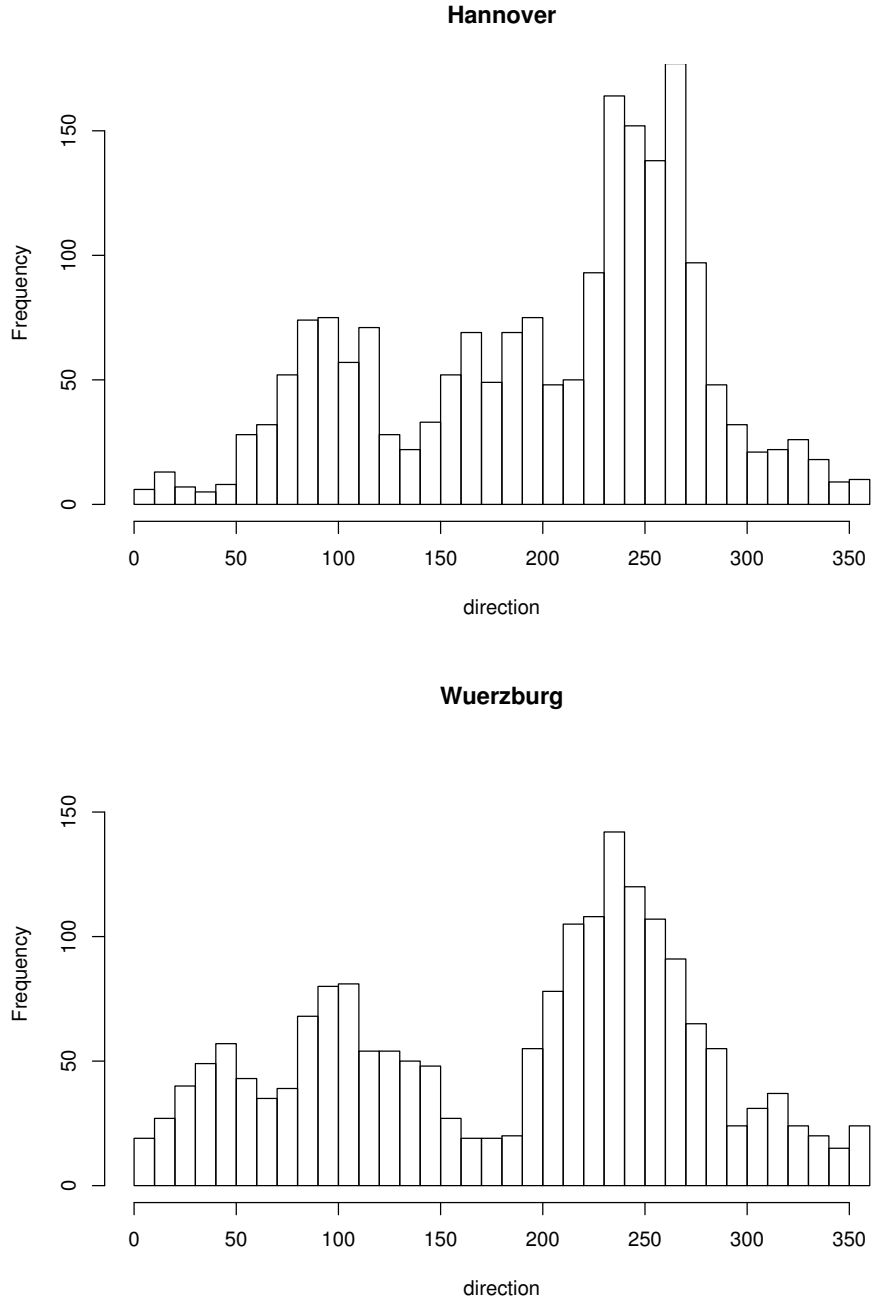

Figure 3.8: Histograms of wind directions of daily maxima for Hannover (top) and Würzburg (bottom).

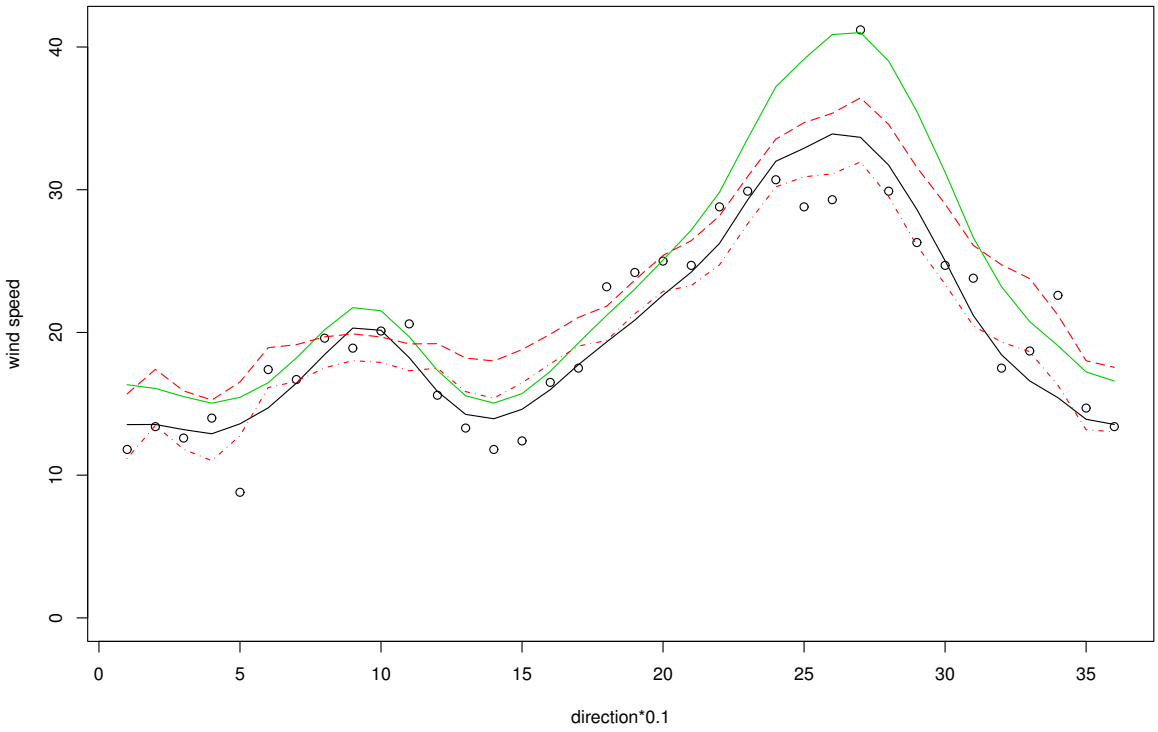

Figure 3.9: (Hannover) Return-level estimates based on the (1,3,2)-harmonic subinterval model based on $\hat{p}_\Theta(\theta)$ of Würzburg: 100-year return-levels (upper solid line) and 22-year return-levels (lower solid line); return-level estimates for the daily maxima model: 100-year return-levels (dashed line) and 22-year return-levels (dashed-dotted line).

model, being much more in agreement with meteorological judgement.

Considering again the histograms of Figure 3.8, the daily directional distribution appears to be shifted roughly by ten degree. We therefore re-estimate the model changing the occurrence distribution of directions by 10°, $p_\Theta^{\text{Hannover}}(\theta) = p_\Theta^{\text{Würzburg}}(\theta - 10°)$; results are given in the Figure 3.10. The estimated 100-year return-level in direction 270° appears now to be slightly above the largest observation in the same direction. Estimates, however, did not change very much, a finding which is in agreement with the results of our simulation study of Section 3.3.2 into the sensitivity of the sub-

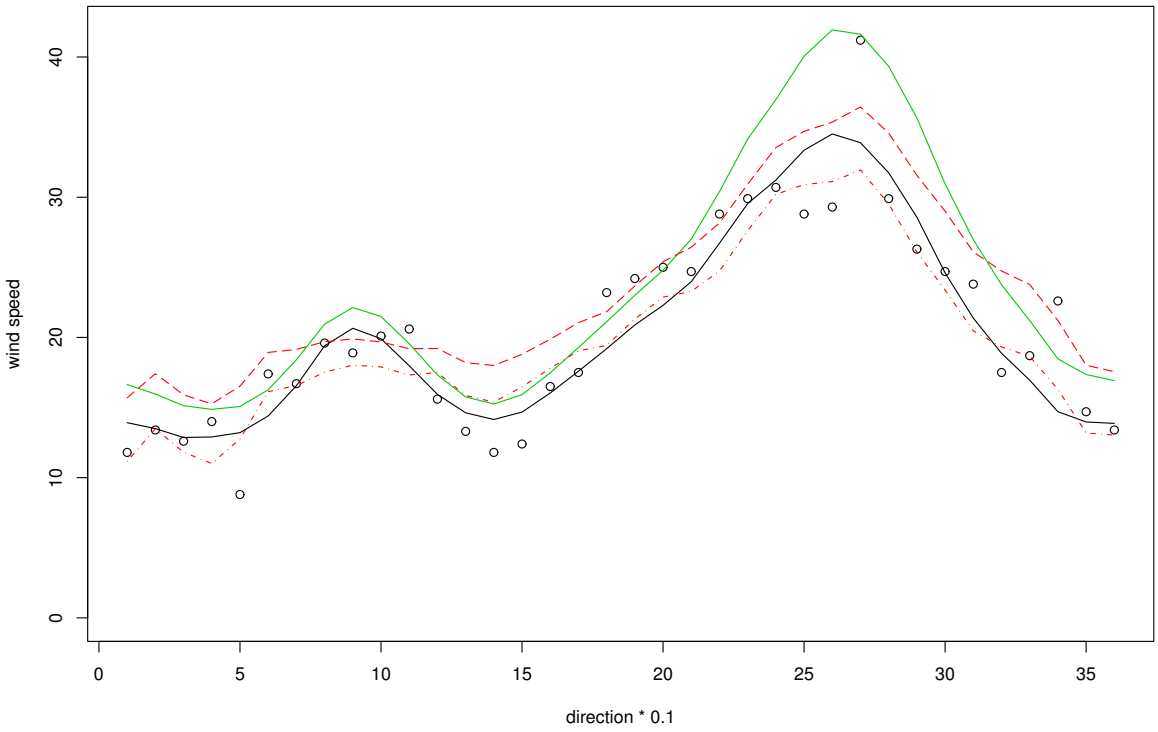

Figure 3.10: (Hannover) Return-level estimates based on the (1,3,2)-harmonic subinterval model based on $\hat{p}_\Theta(\theta - 10°)$ of Würzburg: 100-year return-levels (upper solid line) and 22-year return-levels (lower solid line); return-level estimates for the daily maxima model: 100-year return-levels (dashed line) and 22-year return-levels (dashed-dotted line).

interval model's modelling assumptions.

3.5 Discussion

In this chapter we have considered an approach to treat the masking problem. This problem is due to the recording mechanism of the data, where just the maximum wind speed of an interval is recorded, but no information about the wind behaviour in other directions within this interval is registered.

While daily maxima approaches model only the data actually observed, our approach additionally includes for all remaining directions that occurrences were no larger than the maximum wind speed of that interval. We assume that knowledge of the occurrence probability of each direction for sub-intervals can be approximated reasonably. This is motivated by the fact that at many weather stations such data exist, at least for short time periods. Here, we have daily maxima for 22 years for two stations, and additionally a shorter data set of ten-minutes maxima for one of the stations.

The performance of the approach suggested here to account for masked data is based on a comparison with the daily maxima approach. Mean square errors of high quantiles are used for this comparison via a simulation study. The quantiles we consider are adjusted to the situation of varying numbers of occurrences within intervals. The simulation study has been carried out to analyse the model behaviour under ideal conditions, but also to assess its robustness to deviations of the assumptions occurring with real data.

In the simulation study we have first considered ideal conditions. The new approach shows considerable improvement over the daily maxima approach for all directions. Another result worth noting is that mostly there is a small bias for the daily maxima approach; it just gets large in cases of small return-levels and directions with a low occurrence frequency, which may be attributed to the adjusted return-levels used. The main contribution to the mean square error in case of the daily maxima approach is the variance.

We analysed the robustness of the model in cases of departures from a correctly specified occurrence distribution. The model has turned out to be robust against such departures. This is a very important feature as in a real application the period of available sub-interval data may be short. Furthermore, if no sub-interval data are available for the weather station considered those of a nearby station may be taken instead.

We have also considered the impact of serial correlation, which real data for short

time-intervals exhibit. There is still clear superiority of the new approach, although slightly less strong than for ideal conditions.

Application to real data confirm the results of the simulation study. For the data of Würzburg there is a slight improvement. In case of the data of Hannover, the new model has been applied using the occurrence distribution of Würzburg; this approach seems reasonable due to the results on robustness in the simulation study. Especially for Hannover, the new approach seems to pick up much better the general structure when compared with real data than its daily maxima counterpart. Furthermore, the largest recorded observation for that station and corresponding return-level estimates based on the new approach are much more in agreement with meteorological judgement.

For clarity of presentation, we have restricted the analysis to data from the winter season, which is producing the strongest storm events. Seasonality may, for example, be incorporated as covariates on the parameters (Coles, 2001). Our simulation study has just used one particular model for the wind-speed distribution. Additional analysis with different models might be useful to get further insight.

Chapter 4

Visual summary measures for the conditional model for multivariate extreme values

In previous chapters we have considered the wind behaviour at one weather station only. This included directions and extreme wind speeds. Although this problem may naturally be regarded as bivariate, we have actually split up this into two components, the wind direction and the wind speed given a certain direction. The focus has mainly been on the latter component, which has been treated as univariate. In the following we concentrate on multivariate extreme values. The margins still have to follow a univariate extreme value distribution, but additionally the dependence structure between variables has to be taken into account.

The classical approach to multivariate extremes has turned out to provide a range of possibilities to allow for modelling this dependence structure. This range of dependence structure entirely covers cases where the dependence structure remains the same when moving further into the tails, which is often termed asymptotic dependence. In many situations, however, this type of dependence structure is not present in the underlying process, and a broader class is required. In the Chapters 4 and 5 we consider a recently introduced model for multivariate extreme values that is overcoming this and other restrictions faced by earlier models.

The conditional approach for multivariate extreme values can describe the behaviour of variables having different forms of dependence including asymptotic dependence and asymptotic independence; the latter comprises the three cases of positive and negative extremal dependence, and near extremal independence. An additional advantage is that not all variables need to jointly become large. This feature is important for the approach discussed in Chapter 5. Furthermore, the joint probability of an event falling into a specified region at extreme levels can be calculated for a big variety of regions.

Though, having all these advantages, direct comparison with well known models is not straight forward. The dependence structure of the conditional model is defined pairwise; for each of these pairs we consider one variable is conditioned on the other one which is getting large, and vice versa. So for a given pair of variables we have two dependence statements describing the behaviour of each variable when the other is large. Each of these statements is, in turn, given by two or three parameters and a residual distribution. It is therefore desirable to describe the joint dependence structure of pairs of variables.

In this study we compare visual methods based on different failure regions in order to judge the dependence structure. Performance of these visual methods is investigated via simulation of bivariate normal data using a range of different correlations. We start by giving a short overview of earlier multivariate extreme value approaches and then describe the conditional multivariate extreme value model.

4.1 Introduction

Multivariate extensions of the univariate approach are less straight forward. For example the way of ordering multivariate observations is not obvious (Barnett, 1976). A common approach is to consider componentwise maxima $\mathbf{M}_n = (M_{1n}, \ldots, M_{dn})$ of n iid replicates of $\mathbf{X} = (X_1, \ldots, X_d)$. Usually, marginal and dependence aspects are treated separately. Therefore all marginal distributions are estimated and then transformed to a common distribution. The choice of common distribution is not essential, although some choices turn out to be more convenient than others. Thereafter the dependence structure is analysed.

The classical approach of multivariate extreme values is often presented on unit Fréchet margins, given by $P(Z \leq z) = \exp(-1/z)$ for $z \geq 0$. In the bivariate case the componentwise normalized maxima after transformation to Fréchet margins $P(M_1/n \leq z_1, M_2/n \leq z_2)$ can be shown (Resnick, 1987) to converge to a bivariate distribution function given by

$$G(z_1, z_2) = \exp\left\{-\int_0^1 \max\{sz_1^{-1}, (1-s)z_2^{-1}\} dH(s)\right\}, \tag{4.1}$$

where H is a non-negative measure, which satisfies integral constraints to ensure the marginal distributions are Fréchet distributed, but is arbitrary otherwise. Inherent in the structure given by model (4.1) is the assumption of a constant dependence structure over all extreme levels, which is termed asymptotic dependence. As in the univariate case threshold methods were developed for multivariate applications to make better use of the information at hand (de Haan, 1985; Coles and Tawn, 1991; Coles and Tawn, 1994). In the classical case, the dependence structure at extreme levels is assessed and used for extrapolation. These models allow for a variety of possible 'failure regions', see Coles and Tawn (1994). If the dependence structure of the physical process under consideration does not remain the same over all levels within the tails then application of a model given by (4.1) is likely to yield misleading results.

Ledford and Tawn suggested a generalization of the above model (Leford and Tawn, 1996; Leford and Tawn, 1997). One property of their model in the bivariate case, for variables (Y_1, Y_2) having Gumbel margins, is

$$P((Y_1, Y_2) \in t + D) = \exp(-t/\eta) P((Y_1, Y_2) \in D) \tag{4.2}$$

for an extreme set D which is large in all components and a scalar $t > 0$. They term $\eta \in (0, 1]$ the coefficient of tail dependence and the model given by (4.2) is capable of accounting for asymptotic dependence and asymptotic independence. The coefficient of tail dependence allows for changes in the dependence structure over different levels within the joint tail region, with $\eta = 1/2$ representing near independence, while $\eta \in (0, 1/2)$ and $\eta \in (1/2, 1)$ represent negative and positive extremal

dependence within the class of asymptotic independence, respectively. If $\eta = 1$ the class is asymptotically dependent and the dependence structure is not changing over different levels t. Containing this great flexibility, the model has, however, the disadvantage of requiring all components to become large at the same time. A model overcoming this drawback is summarized subsequently.

For the conditional multivariate extreme value model introduced by Heffernan and Tawn (2004) separation of marginal and dependence aspects is retained. So prior to analysing the dependence structure, the margins of $\mathbf{X} = (X_1, \ldots, X_d)$ are estimated using the semi-parametric model

$$\hat{F}_{X_i}(x) = \begin{cases} 1 - \{1 - \tilde{F}_{X_i}(u_{X_i})\}\{1 + \xi_i(x - u_{X_i})/\beta_i\}_+^{-1/\xi_i} & \text{for } x > u_{X_i}, \\ \tilde{F}_{X_i}(x) & \text{for } x \leq u_{X_i}, \end{cases} \tag{4.3}$$

consisting of the generalized Pareto above a high marginal threshold u_{X_i}, and using the empirical distribution function $\tilde{F}_{X_i}$ below the threshold. All margins are then transformed to standard Gumbel, $P(Y \leq y) = \exp(-\exp(-y))$ for real y, simplifying the presentation of the dependence structure.

Let $\mathbf{Y} = (Y_1, \ldots, Y_d)$ denote a vector with margins following a standard Gumbel distribution, and define $\mathbf{Y}_{-i}$ to be the vector of all but the i–th component. The conditional limiting distribution of all but the i–th component, given Y_i gets large, is then given by

$$\lim_{y_i \to \infty} P(\mathbf{Y}_{-i} \leq \mathbf{a}_{|i}(y_i) + \mathbf{b}_{|i}(y_i)\mathbf{z}_{|i} | Y_i = y_i) = G_{|i}(\mathbf{z}_{|i}), \tag{4.4}$$

where the components of $\mathbf{a}_{|i}(\cdot)$ and $\mathbf{b}_{|i}(\cdot)$ are normalizing functions, and the only assumption on $G_{|i}$ is to have non-degenerate margins. An alternative representation of equation (4.4) is given by

$$\lim_{y_i \to \infty} P(\mathbf{Z}_{|i} \leq \mathbf{z}_{|i} | Y_i = y_i) = G_{|i}(\mathbf{z}_{|i}). \tag{4.5}$$

where $\mathbf{Z}_{|i}$ are standardized variables given by

$$\mathbf{Z}_{|i} = \frac{\mathbf{Y}_{-i} - \mathbf{a}_{|i}(y_i)}{\mathbf{b}_{|i}(y_i)} \tag{4.6}$$

and $G_{|i}$ has non-degenerate marginal distributions. An important feature of the model is that in the limit, the $\mathbf{Z}_{|i}$ and Y_i are conditionally independent given Y_i is large.

There is no unique form to which the normalizing functions $\mathbf{a}_{|i}(\cdot)$ and $\mathbf{b}_{|i}(\cdot)$ are restricted to. Heffernan and Tawn (2004) suggest a functional relationship which has a natural structure and is supported by a broad range of parametric model examples. Their suggestion is to use

$$\begin{aligned} \mathbf{a}_{|i}(y) &= \mathbf{a}_{|i} y + I_{\{\mathbf{a}_{|i}=0, \mathbf{b}_{|i}<0\}}\{\mathbf{c}_{|i} - \mathbf{d}_{|i}\log(y)\} \\ \mathbf{b}_{|i}(y) &= y^{\mathbf{b}_{|i}}, \end{aligned} \tag{4.7}$$

where I denotes the indicator function. The components of $\mathbf{a}_{|i}$, $\mathbf{b}_{|i}$, $\mathbf{c}_{|i}$, and $\mathbf{d}_{|i}$ are constants with $0 \leq a_{j|i} \leq 1$, $-\infty < b_{j|i} < 1$, $-\infty < c_{j|i} < \infty$, and $0 \leq d_{j|i} \leq 1$ for $j \neq i$. They contain information about the dependence structure which can be categorized into asymptotic dependence ($a_{j|i} = 1$, $b_{j|i} = 0$), positive extremal dependence (either $a_{j|i} \in (0,1)$ or $a_{j|i} = 0, b_{j|i} > 0$), near independence ($a_{j|i} = 0$, $b_{j|i} \leq 0$, $d_{j|i} = 0$) or negative extremal dependence ($a_{j|i} = 0$, $b_{j|i} < 0$, $d_{j|i} \in (0,1)$, $c_{j|i} \in I\!R$). Applications of the model then proceed by assuming equation (4.4) to hold exactly for y_i above some threshold u_{Y_i}.

As no specific structure is required for $G_{|i}$ a non-parametric model is adopted for it. We take the empirical distribution of $G_{|i}$, which can be calculated from equation (4.6) with parameters of the functions $\mathbf{a}_{|i}(\cdot)$ and $\mathbf{b}_{|i}(\cdot)$ being replaced by their estimates. Estimation of the parameters of $\mathbf{a}_{|i}(\cdot)$ and $\mathbf{b}_{|i}(\cdot)$ is based on the standardized residuals given by equation (4.6) and assuming the $\mathbf{Z}_{|i}$ to have finite first two moments

$$\begin{aligned} \boldsymbol{\mu}_{|i}(y) &= \mathbf{a}_{|i}(y) + \boldsymbol{\mu}_{|i}\mathbf{b}_{|i}(y) \\ \boldsymbol{\sigma}_{|i}(y) &= \boldsymbol{\mu}_{|i}\mathbf{b}_{|i}, \end{aligned}$$

where the components of $\boldsymbol{\mu}_{|i}$ and $\boldsymbol{\sigma}_{|i}$ are constants. All parameters are then estimated by a pseudolikelihood approach in which an objective function is maximised with respect to all parameters $a_{j|i}$, $b_{j|i}$, $c_{j|i}$, $d_{j|i}$, $\mu_{j|i}$, and $\sigma_{j|i}$ in order to obtain point estimates of $a_{j|i}$, $b_{j|i}$, $c_{j|i}$, and $d_{j|i}$ while the $\mu_{j|i}$ and $\sigma_{j|i}$ are nuisance parameters. Technically this is carried out by falsely assuming all margins of $\mathbf{Z}_{|i}$ for all i to follow a Gaussian distribution and that all contributions to the objective function are independent; so we assume independence between all margins of each conditional distribution as well as independence between all conditional distributions. Although the margins of $G_{|i}$ may actually not follow a Gaussian distribution this approach exploits the consistency property of maximum likelihood estimates for the parameters $a_{j|i}$, $b_{j|i}$, $c_{j|i}$, and $d_{j|i}$, which we are interested in. The independence assumption of the margins of $G_{|i}$ may not hold, although it is often present in common theoretical examples; however, this approach yields consistent point estimators for the parameters $a_{j|i}$, $b_{j|i}$, $c_{j|i}$, and $d_{j|i}$ of the margins of the conditional distribution. The false independence assumption between different conditional distributions can be shown in the case of a Gaussian error distribution to be an approximation to the joint likelihood function (Heffernan and Tawn, 2004).

4.2 Simulation study

Although the multivariate conditional model is flexible and allows for different forms of dependence, it is not always easy to quantify the degree of dependence. The dependence structure is characterized by the parameter functions $a(\cdot)$, $b(\cdot)$ and the distribution of the residuals Z. It is therefore desirable to have a scalar quantity, or a visual summary, to assess the magnitude of dependence. One quantity already introduced by Coles, Heffernan and Tawn (1999) is $\bar{\chi} = 2\eta - 1$ or equivalently η. To give further assistance in assessing the degree of dependence we introduce a range of visual summaries. To see how $\bar{\chi}$ and the visual methods perform, we carry out a simulation study. The choice of distributions to simulate from should include a wide range of degrees of dependence, and its behaviour should also be well understood. A possible candidate is the bivariate normal distribution, which we apply using a range of different correlations. We first describe the simulation design, thereafter discuss resulting parameter estimates, and finally suggest visual summaries of dependence

and consider their performance when applied to the simulated bivariate normal data.

4.2.1 Parameter estimates and dependence

To study the performance of the conditional multivariate model and its behaviour for different degrees of dependence, we carry out a simulation study using a bivariate normal distribution with different correlations covering the range from negative dependence via independence to positive dependence. The choice of correlation coefficients in the subsequent study is $\rho = 0.9$, 0.75, 0.5, 0.25, 0, -0.25, -0.5, -0.75, -0.9. We simulate 20 000 iid replicates in each case. Both the marginal thresholds u_X and the dependence thresholds u_Y are taken to be the 95%-marginal quantile.

Estimates of the shape parameter are significantly different from $\xi = 0$ at a 5%-level, although it is well known that the limit of a normal distribution is of Gumbel type. This disagreement with theoretical results is due to the slow convergence of the normal distribution; the thresholds would have needed to be much higher and the sample size by far larger to yield estimates agreeing with the null hypothesis $\xi = 0$. However, the aim of the study is to see how the model and resulting measures for dependence behave for sample sizes approximately in the order of a potential application.

Having simulated the data, the parameters of the model given by (4.7) are estimated; after estimation of the margins, Heffernan and Tawn (2004) suggest a two step procedure for estimation of the dependence parameters: first using a pseudo-likelihood (subsequently referred to just as likelihood) approach with parameter functions $a_{j|i}(y) = a_{j|i}y$ and $b_{j|i}(y) = y^{b_{j|i}}$ as given in (4.7), then in a second step, if both $a_{j|i} = 0$ and $b_{j|i} < 0$ hold, re-estimating $a_{j|i}(y)$ by using $c_{j|i} - d_{j|i}\log(y)$ instead. In some cases of our simulation study we detected local instead of global maxima of the likelihood in applying the procedure just described. To avoid this problem we calculate both likelihoods and use information criteria for discrimination between them; in our case the Akaike information criterion is applied. Table 4.1 shows the estimated parameters from the simulated normal data.

ρ	a	b	c	d	η	$\bar{\chi}$	$\bar{\chi}_o$
0.9	0.752	0.557	0	0	0.899	0.799(0.764,0.964)	0.817 (0.641,1.020)
	0.724	0.580	0	0			
0.75	0.460	0.630	0	0	0.844	0.687(0.589,0.801)	0.744(0.477,0.889)
	0.430	0.560	0	0			
0.5	0.359	0.567	0	0	0.796	0.592(0.337,0.612)	0.377(0.274,0.658)
	0.310	0.298	0	0			
0.25	0.075	0.128	0	0	0.598	0.196(0.073,0.499)	0.336(0.084,0.443)
	0.089	-0.056	0	0			
0	0	-0.249	5.066	1	0.425	-0.150(-0.218,0.356)	0.105(-0.068,0.209)
	0	-0.143	0.566	0.164			
-0.25	0	-0.146	0.433	0.508	0.386	-0.229(-1.000,0.268)	-0.204(-0.234,-0.009)
	0	-0.214	0.377	0.424			
-0.5	0	-0.369	1.188	1	0	-1(-1,NA)	-0.403(-0.417,-0.232)
	0	-0.278	0.232	0.720			
-0.75	0	-0.224	0.123	0.876	NA	NA	-0.530(-0.599, -0.481)
	0	-0.427	-1.909	0.274			
-0.9	0	-0.299	-3.075	0.258	NA	NA	-0.712 (-0.745,-0.673)
	0	-0.401	0.011	0.926			

Table 4.1: Estimated parameters of the conditional model for different correlations ρ and estimates for η, $\bar{\chi}$, and $\bar{\chi}_o$.

Figure 4.1 shows the value of the profile log-likelihood for data with different correlations and for a range of different b values. For illustration we applied a simplification by using the likelihood functions including the parameters μ, σ, a (where μ and σ are the nuisance parameters used for estimation discussed in Section 4.1) whenever $b > 0$ and μ, σ, c, d otherwise. The vertical dotted line indicates the resulting estimate, when the procedure is allowed to stop after having found a maximum for positive b. For small negative correlated data the procedure without applying an information criterion works quite well, as the likelihood increases when positive values of b are getting smaller resulting in estimates with $a = 0$ and $b < 0$. In contrast, for correlations in between -0.45 and -0.55, and stronger negative cases, the likelihood has a local maximum at a positive value of b a consequence of which is the estimation procedure terminating with $b > 0$. However, as these plots demonstrate, a far higher likelihood is achieved for negative b.

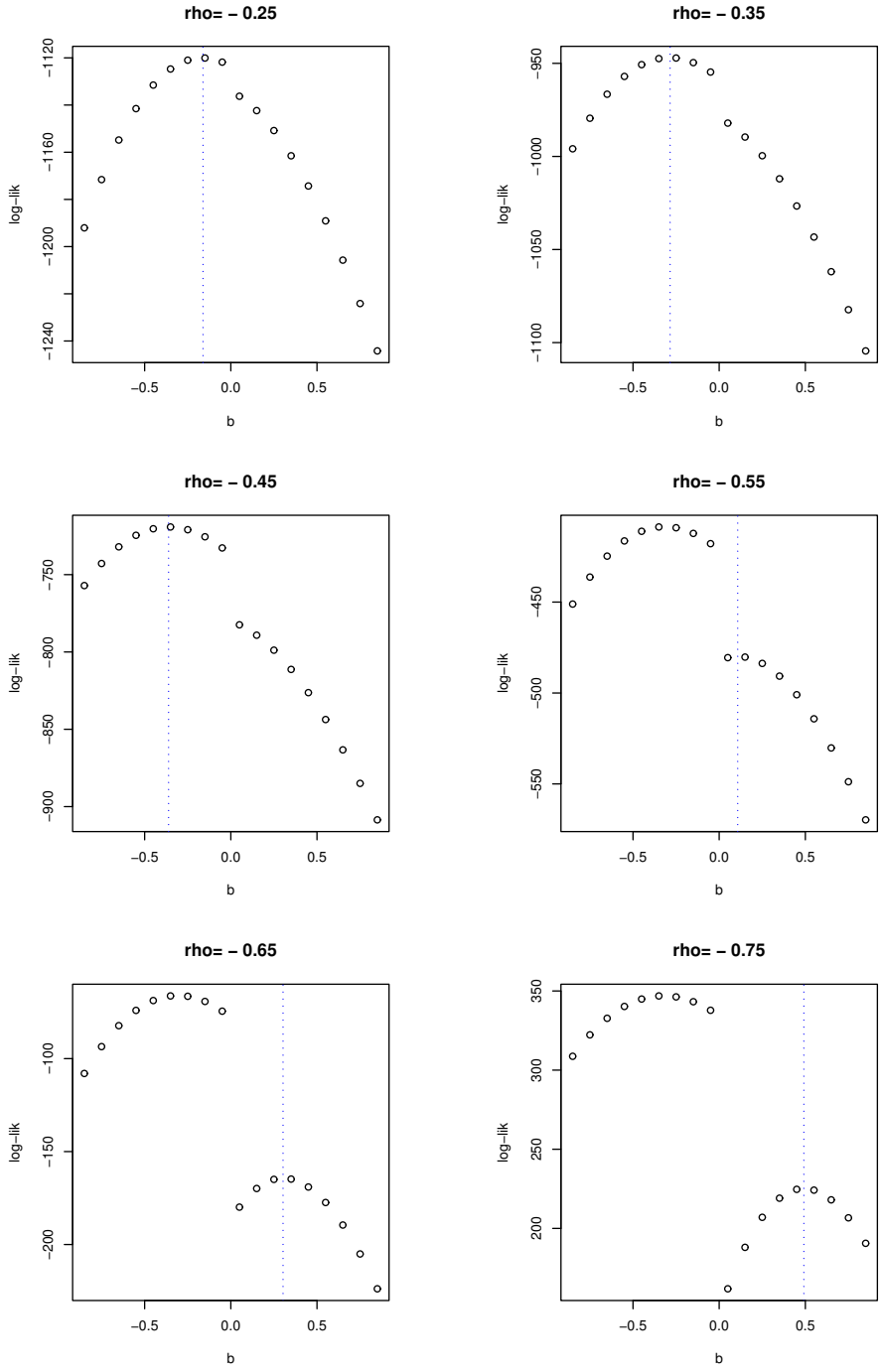

Figure 4.1: (Modified) Profile log-likelihood using a model based on normalizing functions given by (4.7) applied to normal data having a range of different negative correlations.

One feature of the conditional model is that (in the limit) residuals are independent of the conditioning variable. Analysis of the residuals in the present case did not show any dependence on the conditioning variable. The residuals were regressed on the conditioning variable on the original scale and on their quantiles (which are equally spaced). In both cases all coefficients of the regression analysis undertaken had non-significant slope-parameters. Furthermore, simulated normal data and simulated values under the conditional model have shown good agreement in the tail regions.

4.2.2 Dependence summaries based on failure regions

In this section we look at how different forms of dependence are summarized using four chosen *failure regions*. Let $x = (x_1, x_2)$ denote a point in $\mathbb{R}^2$, these failure regions are defined as

$$\begin{array}{lcl} a) A_{1,v} & = & \{(x_1, x_2) : x_1 + x_2 > v\} \\ b) A_{2,v} & = & \{(x_1, x_2) : (x_1^2 + x_2^2)^{1/2} > v\} \\ c) A_{max,v} & = & \{(x_1, x_2) : \max(x_1, x_2) > v\} \\ d) A_{min,v} & = & \{(x_1, x_2) : \min(x_1, x_2) > v\}. \end{array} \tag{4.8}$$

While a) – c) are derived from well known distance measures, d) seems a sensible choice for extremes as it measures whether a point is large in both components.

After model estimation, we can simulate data from the conditional model above the dependence threshold u_Y. The procedure is as follows. Simulate y_i^* from a standard Gumbel distribution given it exceeds its threshold u_{Y_i}, and calculate $y_{j|i}^* = a_{j|i}(y_i^*) + b_{j|i}(y_i^*)Z_{j|i}$ with $i, j = 1, 2$, $i \neq j$, where $Z_{j|i}$ is sampled with replacement from the set of residuals independently of the value y_i^*. This yields pairs $(y_1^*, y_{2|1}^*)$ given y_1^* is large and pairs $(y_{1|2}^*, y_2^*)$ given y_2^* is large. To be able to calculate probabilities of sets of interest we first consider how to split up the space of points exceeding at least one threshold. Let $\{C_1, C_2\}$ be a disjoint partitioning of the space of points (y_1, y_2) exceeding at least one threshold, and define $C_1 = \{(y_1, y_2) : y_1 > u_{Y_1}, y_2 < (u_{Y_2} - u_{Y_1}) + y_1\}$ and C_2 similar by interchanging the indices 1 and 2. Then the

probability of A_v can be stated as

$$P(A_v) = \sum_{i=1}^{2} P(A_v \cap C_i | Y_i > u_{Y_i}) P(Y_i > u_{Y_i}). \tag{4.9}$$

By using the pairs of simulated values $(y_1^*, y_{2|1}^*)$ and $(y_{1|2}^*, y_2^*)$ it is then possible to estimate the probabilities $P(A_v \cap C_i | Y_i > u_{Y_i})$, $i = 1, 2$, using their empirical counterparts, while $P(Y_i > u_{Y_i})$ is calculated directly from the Gumbel distribution. Multiplication of these probabilities and summing up over $i = 1, 2$ yields an estimate for $P(A_v)$ whenever $v > k$, where k is the value so that all points in A_v are above the threshold u_{Y_i}. To estimate $P(A_v)$ for $v < k$ additionally requires calculation of the probability for an event falling below the threshold; as data are dense in this region, this probability is evaluated empirically.

Coefficient of tail dependence

We now consider the coefficient of tail dependence η, which can be calculated by re-arranging equation (4.2) as

$$\eta = \frac{-t}{\log(P(Y \in t + A_{min,v})) - \log(P(Y \in A_{min,v}))}. \tag{4.10}$$

The set $A_{min,v}$ given by (4.8d) is equivalent to D in equation (4.2); obviously $t + A_{min,v}$ is the same set as $A_{min,v+t}$ and we use both representations as convenient. The procedure described in relation with equation (4.9) is used to calculate the probabilities $P(A_{min,v})$ and $P(t + A_{min,v})$. For each set of simulated bivariate normal data considered we now use the model estimated in Section 4.2.1. To estimate η we simulate 10 000 values for each of the margins and the corresponding value of the other component from the estimated model, and substitute the $P(Y \in \cdot)$ terms in (4.10) by their empirical counterparts. The value v is chosen to be the dependence threshold $u_{Y_1} = u_{Y_2}$. In the case of strong positive dependence the choice of t is not crucial. However t needs to be large enough for the differences of probabilities in (4.10) to be well estimated by its empirical counterparts, but should not be too large to make estimates for $P(Y \in t + A_{min,v})$ unreliable. The latter point will have a considerable impact for negatively dependent data, as joint exceedances in this region become rare. We therefore recommend to try a number of t values

and choose a value from within an interval where η appears to be approximately constant. In this study t is taken to be 1. For normal data the theoretical value of η is in the limit $(1 + \rho_{ij})/2$. Rearranging terms yields $\bar{\chi} = 2\eta - 1$, which is in the limit ρ_{ij}, where the term $\bar{\chi}$ follows Coles et al. (1999); therefore $\bar{\chi}$ allows direct comparison with ρ. Table 4.1 shows results for the same bivariate normal data estimated previously. Data in the joint upper region, the set which $\bar{\chi}$ is based on, are getting scarce with rising negative dependence. In the case of strong negative dependence, there are no data left in this region which explains the NA values of $\bar{\chi}$ if $\rho \leq -0.75$. The 95%-confidence intervals for $\bar{\chi}$ shown in Table 4.1 are based on 200 bivariate normal samples each of the same size as the original sample and re-estimation of $\bar{\chi}$.

Since estimation of $\bar{\chi}$ is unreliable or even impossible for a range of cases with negative dependence, we alternatively calculate this dependence measure without using the conditional model. Instead we use a non–parametric transformation to Gumbel margins of the data although the transformation could also have been achieved by using the marginal parameter estimates. To ensure having enough data for estimation, the joint exceedance regions are chosen to be large quantiles fixed in advance. Let (Y_1, Y_2) denote a bivariate random variable with Gumbel margins. Applying the transformation $U = \min\{(Y_1, Y_2)\}$ it is seen that $P(U > v)$ corresponds to the probability of (Y_1, Y_2) falling into the set $A_{min,v}$, where we assume v to be a large quantile of the distribution U. Using the distribution of U we fix the probability of falling into the regions $A_{min,v}$ and $t + A_{min,v} = A_{min,v+t}$ by taking v and $v + t$ to be the upper 98% and 99% quantile, respectively. Having these regions fixed, calculation of t is straightforward from the quantiles of U corresponding to $A_{min,v}$ and $t + A_{min,v}$. Calculation of η_o or $\bar{\chi}_o$ is immediate from equation (4.10), where the index "o" is used to distinguish the estimates from those based on the conditional model above. It can be seen from Table 4.1 that $\bar{\chi}$ is (in terms of confidence intervals) superior to $\bar{\chi}_o$ for correlations around or greater than 0.5. However, due to its construction, $\bar{\chi}_o$ supplies estimates for all values of ρ and yields better estimates compared to $\bar{\chi}$ for any ρ below 0.25. Negative dependence is however permanently underestimated. This bias is a consequence of the slow convergence in the normal case to its limit; for illustration of the convergence of $\bar{\chi}$ in the normal case see Coles et al. (1999).

Visual summary measures

The procedure described in relation with equation (4.9) is used again to calculate the probability $P(A_v)$. In order to establish graphs based on this procedure we evaluate $P(A_v)$ for a range of values v on Gumbel scale according to the different failure regions A_v given by (4.8) a) – d). The calculation of the probability of the region A_v is based on 100 000 simulated values above the dependence threshold u_{Y_i}. Since we are interested in the behaviour of the tails we consider $\log(1/P(A_v))$ rather than $P(A_v)$ in the graphs. To provide a reference for comparison, values of $\log(1/P(A_v))$ corresponding to perfect dependence and independence are additionally superimposed.

We now consider the graphs based on different failure regions as described above, see Figures 4.2 to 4.5. All graphs were calculated for a range of different values v on Gumbel scale above the threshold; for a smaller range of values the visual summaries were also evaluated below the threshold to get an impression how model based estimation agrees with those additionally based on the empirical distribution. The dotted vertical line corresponds to the value v of the particular failure region at the dependence thresholds (u_{Y_1}, u_{Y_2}). The dashed line indicates independence of the two variables considered, while the dotted line represents perfect positive dependence. Note that while for most graphs the dependence line is below the independence line for v greater than the threshold, in case of A_{max} the opposite is true.

As can be seen the graphs show good discrimination for positive dependence for any choice of failure region given in (4.8) a) – d) above. One apparent drawback of A_{max} and A_2 is their lack in discriminating between data having different degree of negative dependence including independence; this drawback makes their application questionable. A_{min} and A_1 do not exhibit this lack, and both of them seem appropriate to judge strength of dependence. We therefore focus on these two alternatives.

Although A_{min} has the potential to discriminate well between different degrees of dependence, the plots indicate problems when correlation is negative. The reason for this behaviour is that A_{min} is based on the upper joint tail region, and resulting estimates for negatively related data are based on very few values. One possibility

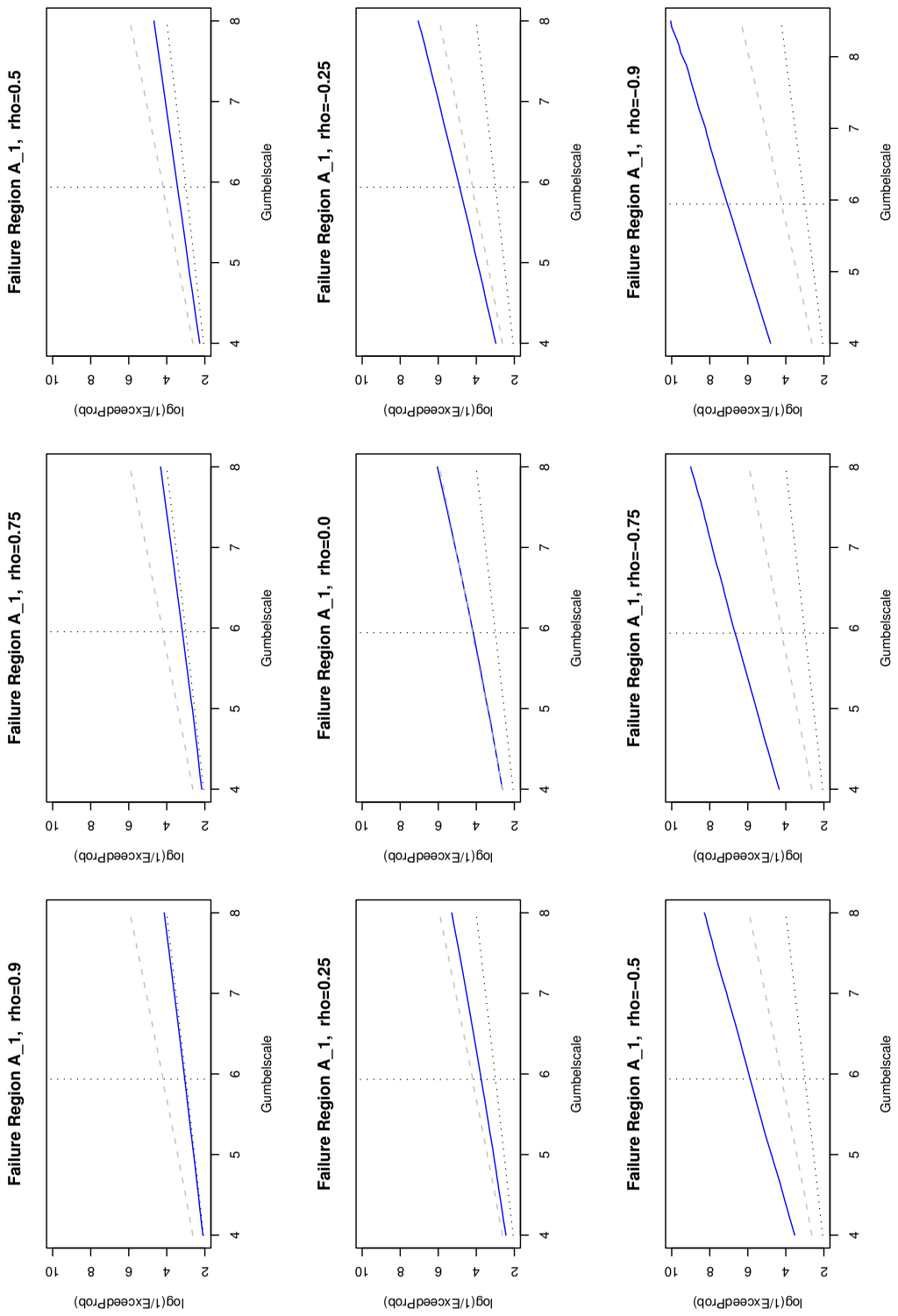

Figure 4.2: Graphs show application of the failure region A_1 to normal data with correlations (row-wise) ρ= 0.9 , 0.75, 0.5, 0.25, 0, -0.25, -0.5, -0.75, -0.9. Lines: solid (estimate), dotted (perfect dependence), dashed (independence).

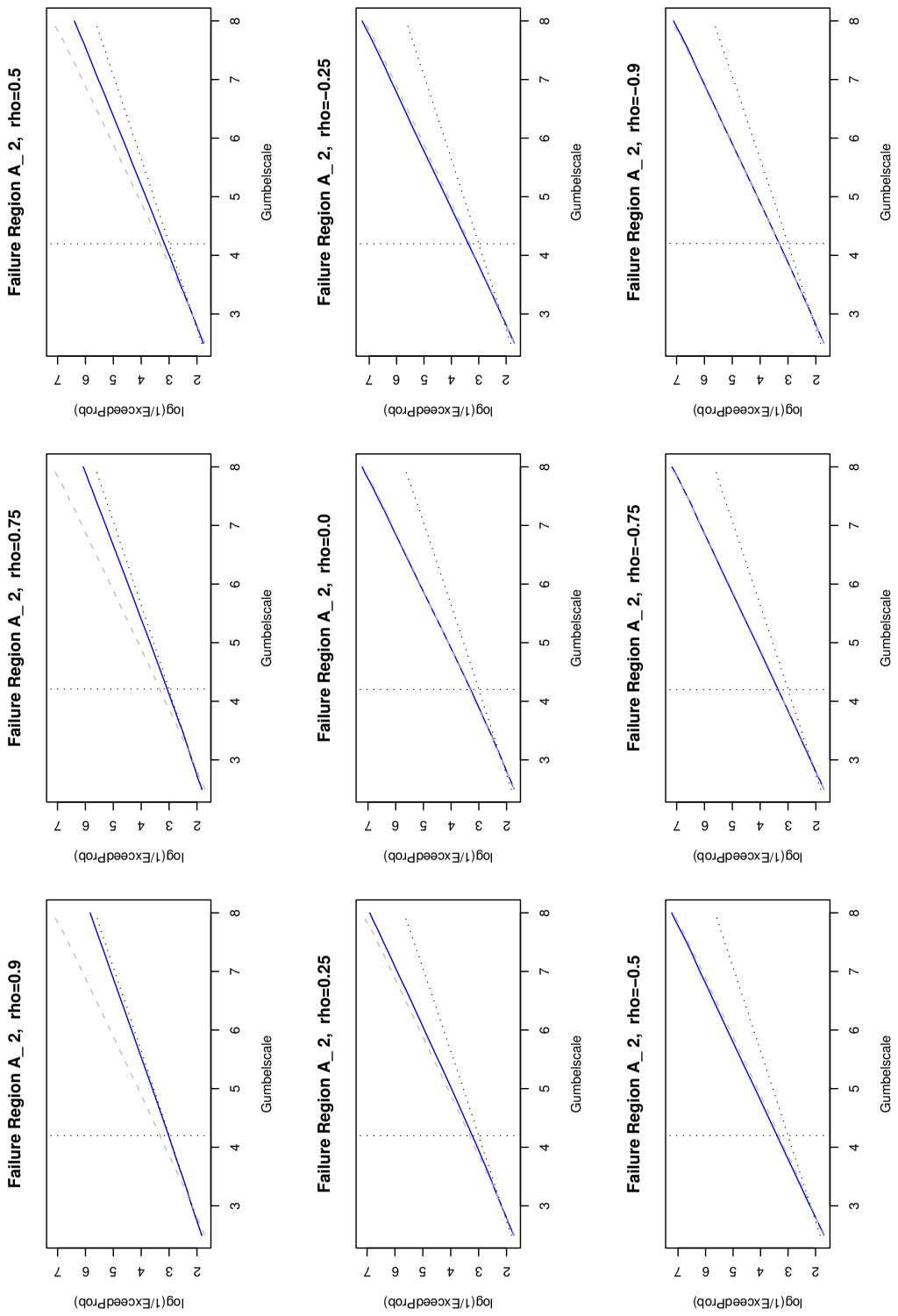

Figure 4.3: Graphs show application of the failure region A_2 to normal data with correlations (row-wise) ρ= 0.9, 0.75, 0.5, 0.25, 0, -0.25, -0.5, -0.75, -0.9. Lines: solid (estimate), dotted (perfect dependence), dashed (independence).

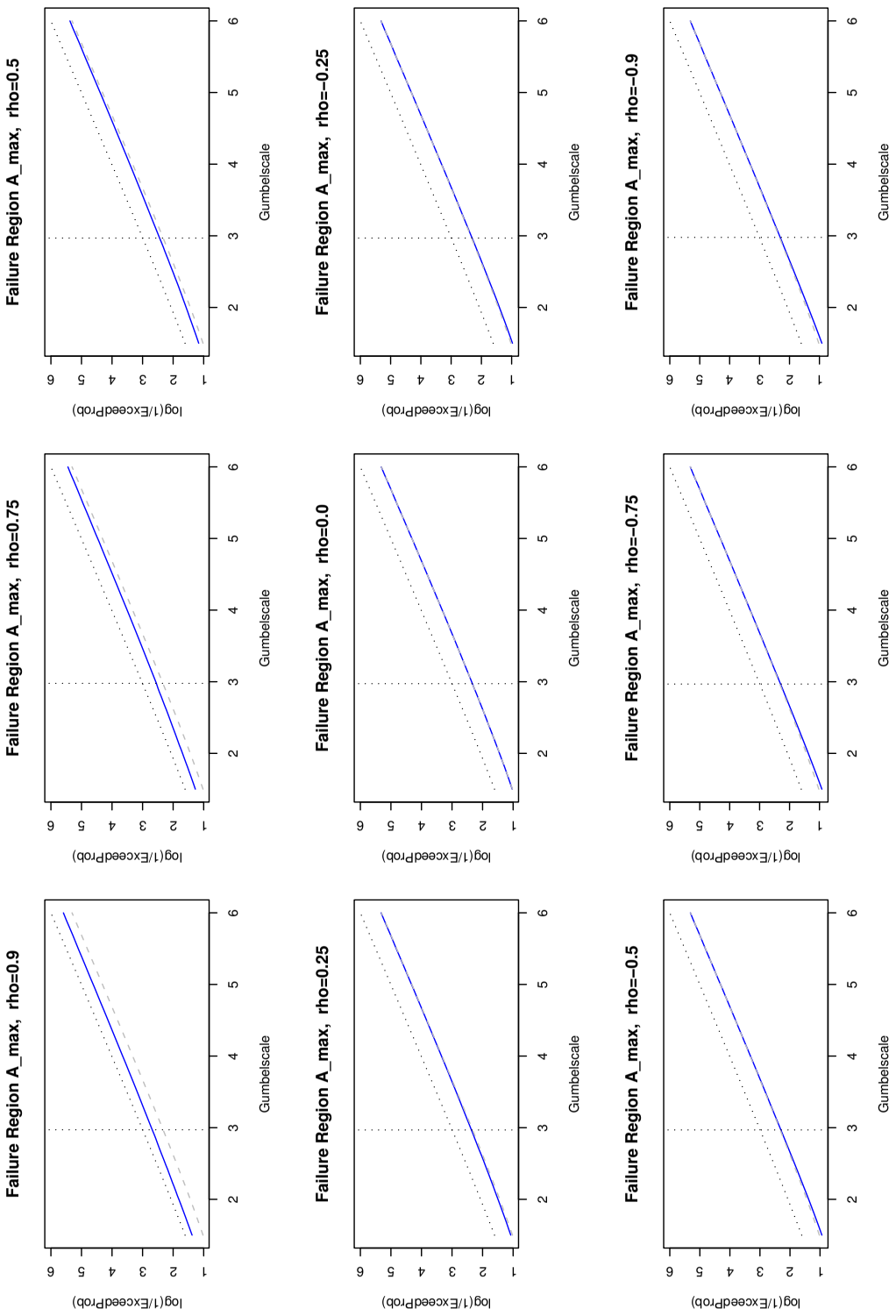

Figure 4.4: Graphs show application of the failure region A_{max} to normal data with correlations (row-wise) ρ= 0.9, 0.75, 0.5, 0.25, 0, -0.25, -0.5, -0.75, -0.9. Lines: solid (estimate), dotted (perfect dependence), dashed (independence).

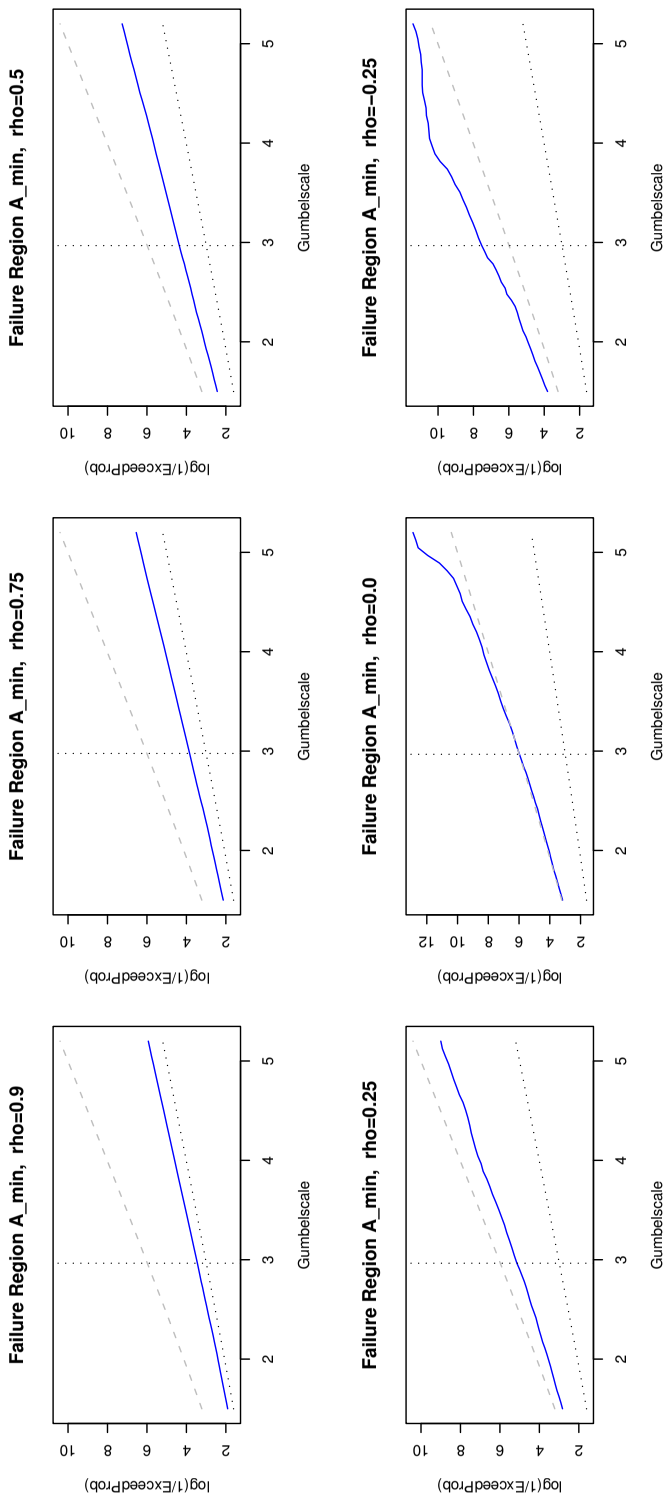

Figure 4.5: Graphs show application of the failure region A_{min} to normal data with correlations (row-wise) ρ= 0.9, 0.75, 0.5, 0.25, 0, -0.25,(-0.5, -0.75, -0.9). Lines: solid (estimate), dotted (perfect dependence), dashed (independence).

to alleviate this problem is to restrict v to smaller values to ensure there are enough data for estimation. As we are mostly interested in the behaviour of the upper tail this approach is not sensible. Considering just the non-negative case, estimated curves appear to discriminate approximately linear between different degrees of dependence making interpretation much easier.

There is a close link between $\bar{\chi}$ and visual measures based on A_{min}. Both of them are based on probabilities of sets including joint exceedances only. So as $\bar{\chi}$ is not a good measure in the case of negative dependence, due to scarcity of joint extremes, similar arguments hold for a visual procedure based on A_{min}. Furthermore, once having calculated one of both dependence measures, the other one is not likely to contain much additional information.

We consider now the A_1 measure. Its slight disadvantage is the crossing of dependence and independence line at a low value (not shown in the graph), which makes interpretation in this region difficult. However, in most cases, the dependence threshold is above this value, and the tail is the region we are primarily interested in. Measures based on A_1 show good discrimination for all possible correlations. The reason is that they take account of all extreme regions and not just those being jointly extreme. Therefore a considerable amount of data are incorporated for estimation at all reasonable levels. So in contrast to the A_{min}, which agrees with traditional approaches in that all components have to become jointly extreme, A_1 much better ties in with the idea of the conditional approach in that additionally regions are considered where not all components are extreme at the same time.

4.3 Wind speed application

In this section we apply the conditional multivariate model to wind data of three different weather stations: Hannover, Göttingen, and Würzburg, which are in the stated order lying approximately on one line in north-south direction with the distance Göttingen-Würzburg being roughly double the distance Hannover-Göttingen. As we expect towns close together to show stronger dependence than those far apart, we consider how well estimated dependences reflect the distances of towns. The data

consist of daily maximum wind gusts for 22 successive years. No declustering of the data is carried out here, as this procedure may slightly bias estimates of the dependence structure; for example choosing componentwise maxima of blocks can yield values of two components resulting from different events. However, non-declustering should be taken into account when considering standard errors which may be underestimated due to non-independence of data, while point estimates themselves are not materially affected. However, as we are working with daily data, serial correlation is rather weak.

	β	ξ
Hannover	3.923	-0.105
Würzburg	4.469	-0.081
Göttingen	3.691	-0.112

Table 4.2: Parameter estimates of the marginal GPD with scale and shape paramters β and ξ for three different weather stations.

In each case a marginal threshold of 15 m/s, corresponding for Hannover, Würzburg, and Göttingen to quantiles 0.84, 0.83, and 0.92, respectively, and a dependence threshold with exceedance probability 0.2 were found appropriate by inspection of corresponding mean exceedance plots and residual plots, respectively. Estimates of the GPD marginal parameters are shown in Table 4.2.

	a	b	η	$\bar{\chi}$
Wü|Ha	0.802	0.411	0.927	0.855
Gö|Ha	0.824	0.366	0.958	0.916
Ha|Wü	0.670	0.463		
Gö|Wü	0.764	0.326	0.946	0.892
Ha|Gö	0.835	0.507		
Wü|Gö	0.836	0.408		

Table 4.3: Parameter estimates of the conditional multivariate extreme value model, and estimates for η and $\bar{\chi}$.

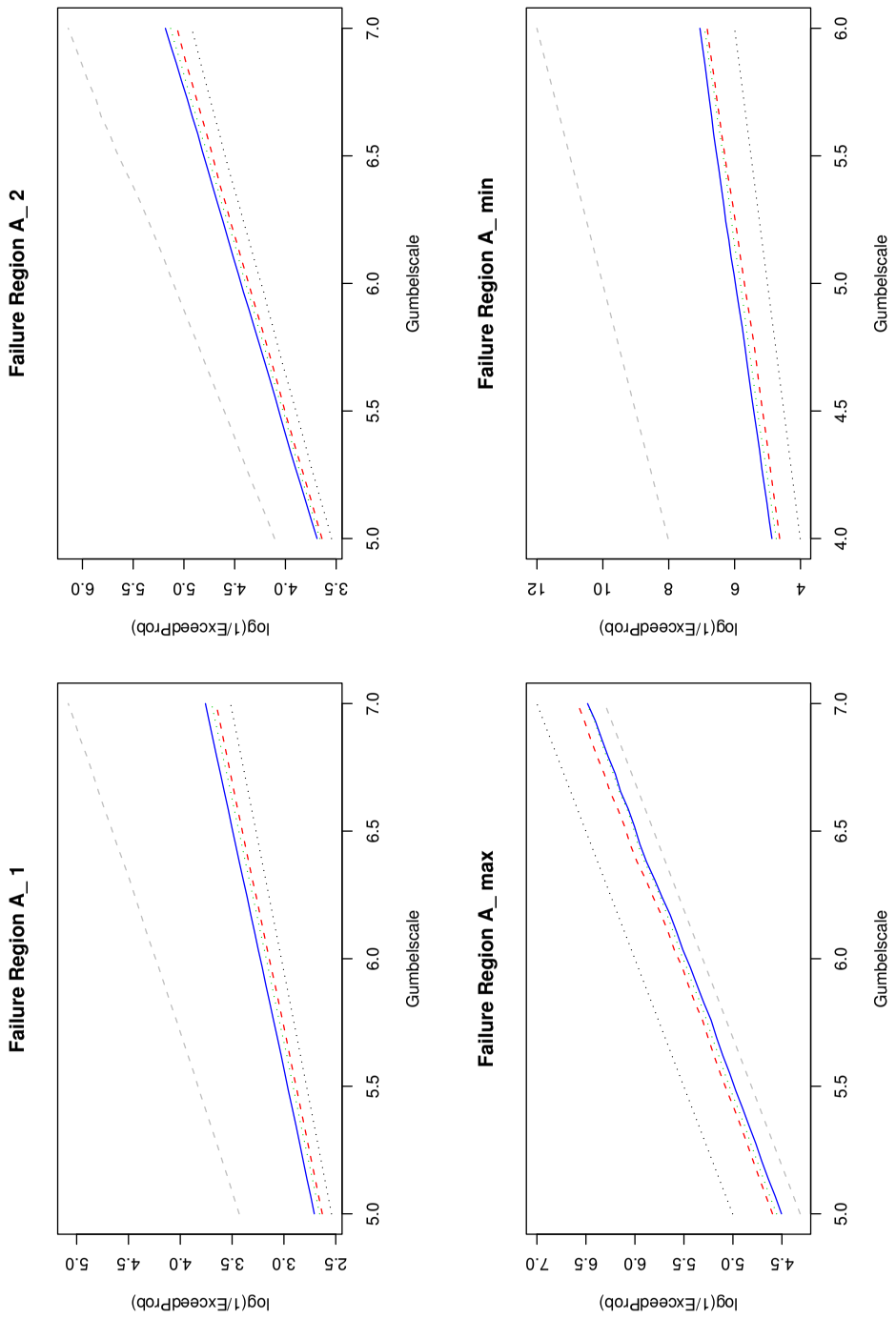

Figure 4.6: Plot of the visual summary using different failure regions applied to wind data: Hannover-Würzburg (solid), Würzburg-Göttingen (dashed), Hannover-Göttingen (dotted), and covering dotted lines for dependence (black) and independence(grey).

Parameter estimates of the conditional multivariate model are found in Table 4.3. The estimate of parameter a indicates strong dependence, though direct, exact interpretation just by parameter estimates of a and b is difficult. Also shown are estimates for η and $\bar{\chi}$; their magnitudes well represent the order of distances between pairs of towns.

Figure 4.6 shows plots of the wind data for the different failure regions evaluated above the dependence threshold. 100 000 values are used for the simulation above the dependence threshold to assess failure probabilities for each region. The solid line corresponds to the largest distance, Hannover-Würzburg, the dotted line is for Würzburg-Göttingen, while the dashed line describes the dependence of Hannover-Göttingen. The covering dotted and dashed lines again represent perfect dependence and independence, respectively. All plots well represent the order of the distances of pairs of towns.

A comparison with the estimates of $\bar{\chi}$ confirm the observations drawn in the simulation study. Since dependence appears to be strongly positive across all levels shown in the plot, all four measures perform satisfactory. A drawback of the A_2 measure is its crossing of lines of stronger and weaker dependent data at a value around three, making interpretation more difficult. Graphs based on failure regions A_1 and A_{min} show desirable discrimination between different degrees of dependence highlighting well the behaviour in the upper tail.

4.4 Discussion

The conditional approach for multivariate extreme values overcomes a number of drawbacks earlier models face. Its dependence structure is, however, determined by two normalizing functions and the distribution of the residuals, and assessing its dependence is therefore not easy. In this study we have looked at summaries of the dependence structure. The scalar measure $\bar{\chi}$, forming a link to earlier approaches, has been considered in a simulation study, which has shown satisfactory behaviour for positive correlation. However, in case of negative dependence, estimates of $\bar{\chi}$ are not reliable or for large negative ρ not obtainable. So for the normal data the estimator $\bar{\chi}_o$ performs considerably better than $\bar{\chi}$ for negative correlation, but is

biased when negative dependence is present.

An alternative to the scalar dependence measure presented in this study is visual summaries of dependence. These make use of four different failure regions, and their ability to discriminate dependence has been analysed in detail using a simulation study and applying them to a set of data. Since we are primarily interested in the behaviour in the tails, the plots are organized so as to highlight this region.

In the case of positive correlation all visual measures can discriminate between different degrees of dependence. These findings are also confirmed by application of these measures to wind data of three stations having different distances. For these data all of the visual measures represent smaller distances by higher dependence, a result being intuitively reasonable. Furthermore, results of the visual summaries appear to be in good agreement with estimates of the scalar measure $\bar{\chi}$.

For the failure regions A_2 and A_{max} these visual measures are unable to distinguish between different degrees of non-positive dependence; the remaining two have the capability to discriminate in all cases. Among these latter two, visual summaries based on the failure region A_{min} turn out to yield very unreliable estimates when based on data with small positive or small negative dependence. This is a consequence of scarceness of data in the joint tail region and far extrapolation is therefore not possible in these cases. For the same reason estimation of the probability of A_{min} in case of strongly negative correlated data is based on virtually no joint exceedances invalidating a visual summary based on this failure region. All these problems are not present when applying the graphs to the failure region A_1. Although separation between different degrees of dependence is not exactly linear it appears to discriminate well for the whole range of dependences. Furthermore, extrapolation far into the tails shows good and reliable behaviour.

A common question when maximizing a likelihood is whether the applied procedure indeed returns the overall maximum of the objective function. Though negative dependence is not the most common case in applications, resulting estimation procedures may find a local instead of a global maximum of the likelihood function in these situations. This situation has occurred in our study using simulated data hav-

ing intermediate or strong negative correlation. Standard techniques can be applied to avoid this problem.

It is worth mentioning that in the present study visual plots for negatively correlated data still indicate negative dependence when the estimates of the parameters resulted in $a = 0$ and $b > 0$, which would give rise to other conclusions. This is a consequence of the visual summaries taking the distribution of the residuals into account, which have in those cases most of their mass in the negative part. The plots may thus provide a helpful tool to make a check on estimation results.

We finally want to mention that although the simulated data are symmetric, we have not exploited this feature in the estimation procedure. Restricting estimation to this symmetric structure may, however, improve the results. In applications symmetry is not always a valid assumption and we have therefore not restricted estimation to this special case in the present study.

Chapter 5

Directional dependence in extremes for two stations

5.1 Introduction

For a number of applications of extreme wind speeds directionality plays an important role. One application we are particularly interested in is related to some high speed trains for which an extreme gust may cause derailment and the corresponding risk depends on the angle between gust direction and direction of the rails. An important question here is whether extreme wind events occur rather localized and independent of points at some distance, or if some sort of joint dependence is present.

The two towns Hannover and Würzburg constitute the end points of a railway track for high speed trains. The orientation of this track is in north-south direction. The risk is highest for gusts perpendicular to the motion of the train, thus easterly and westerly storm events are of particular interest. With the highest wind speeds occurring roughly from west, we mainly focus on this direction. North is be denoted by $360°(=0°)$ and directions are defined clockwise, so that west is $270°$.

For the weather stations of the two towns, daily maximum wind speeds are available for the years 1976-1997. In the following we just consider data for days available at both stations. In the current application interest is in the force of the wind, so it is sensible to resolve all data appropriately to all directions. The wind component $\tilde{R}_\alpha$

is defined here as $\tilde{R}_\alpha = R_\phi \cos(\alpha - \phi)$ for all directions α and observed wind speeds R_ϕ in direction ϕ; note that $\tilde{R}_\alpha \in \mathbb{R}$ can take on negative values being interpreted as the same absolute wind force in the opposite direction.

With interest in the joint behaviour of extreme wind events at both stations, it is sensible to first consider our data. The scatterplot in Figure 5.1 shows component wind speeds of Würzburg in direction 270° plotted against the corresponding ones of Hannover in direction 270°. For extreme levels at the upper end, the data exhibit a positive relation between the two stations for this particular choice of combination of directions, suggesting the presence of an underlying positive dependence.

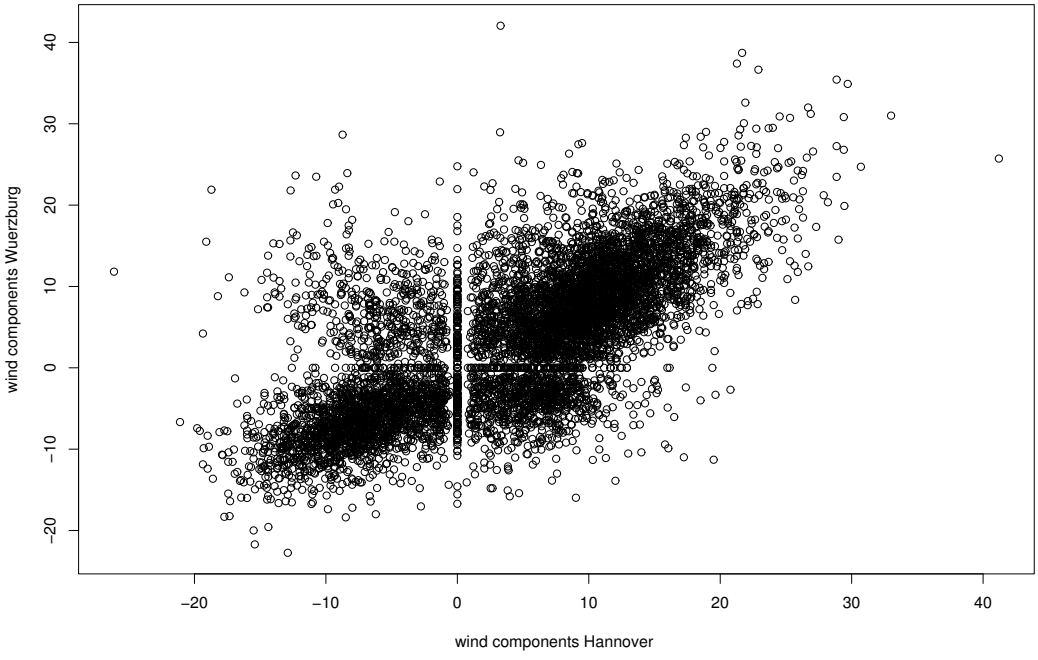

Figure 5.1: Scatterplot of wind components in direction 270° for both stations Hannover and Würzburg.

Let us consider a simplified version of the track consisting of two stations only. Then

it is natural to look at the probability of observing an extreme event at the track. In mathematical terms this is given by

$$P((R_{\phi_1} > x) \cup (R_{\phi_2} > y)) = \\ P(R_{\phi_1} > x) + P(R_{\phi_2} > y) - P(R_{\phi_2} > y | R_{\phi_1} > x) P(R_{\phi_1} > x). \quad (5.1)$$

So to consider, whether at least at one of the two stations we observe a wind speed of, say $x = y = 40$m/s, we use equation (5.1). For the calculation of the above probability we need, however, to be able to calculate the conditional probability at the right hand side of equation (5.1), which comprises the joint dependence between the two variables of interest, and this will vary with ϕ_1 and ϕ_2.

To investigate the dependence structure of the wind process at extreme levels we make use of the conditional multivariate extreme value model introduced in the previous chapter. The advantage of this particular model in the current application is its high flexibility and the possibility to incorporate all forms of dependence, including both positive and negative extremal dependence. As it is natural to assume the wind process to vary smoothly over directions, we use functions to describe model parameters which vary continuously over directions. All margins are transformed to the double exponential distribution allowing for a smooth transition over different forms of dependence. In contrast to a separate sector by sector analysis this further allows us to employ neighbouring information to improve on estimates. Additionally, the number of parameters can be reduced considerably.

5.2 Model definition

To analyse the dependence structure it is convenient to have a standard distribution for all margins. The choice of common margins is to some extent arbitrary, in many cases for mathematical convenience, presentational transparency or easier interpretation. The conditional extreme value model introduced in the previous section was presented using Gumbel margins as suggested by Heffernan and Tawn (2004).

Due to the asymmetry of the Gumbel distribution there is actually two different forms to describe the dependence structure with a difference in the number of parameters. Having two different forms of the dependence structure is rather inconvenient when considering dependence of wind speeds over directions, especially when the number of parameters and their interpretation is changing. We therefore suggest to use the double exponential distribution, having the same upper tail as the Gumbel distribution, but being symmetric at the same time. The double exponential distribution is given by

$$P(Y \leq y) = 0.5 \exp(y)\mathbf{1}_{\{y \leq 0\}} + (1 - 0.5 \exp(-y))\mathbf{1}_{\{y > 0\}},$$

where $\mathbf{1}_{\{\cdot\}}$ denotes the indicator function. By having the same upper tail and symmetry we can use the modified version

$$\mathbf{a}^*_{|i}(y) = \mathbf{a}_{|i} y \qquad \mathbf{b}^*_{|i}(y) = y^{\mathbf{b}_{|i}} \tag{5.2}$$

of parametrisation (4.7) for the whole range of y. That is, we do not need to use different dependence functions for positive and negative dependence. We have now $-1 < a_{i|j} \leq 1$ with negative values of $a_{i|j}$ corresponding to negative dependence.

We start building up a global model by first considering separate models for different directions and then join these together. We begin by conditioning on one variable, that is the wind speed for a fixed direction ϕ_1 at the first station, and determine the dependence structure of wind events in all directions at the other station, $\phi_2 \in (0, 2\pi]$. Let (Y_{ϕ_1}, Y_{ϕ_2}) be a pair of random variables after transformation to double exponential margins with each component representing the wind speed in direction ϕ_i at station $i = 1, 2$. Then the conditional distribution of Y_{ϕ_2} given an extreme event at the first station in direction ϕ_1 is assumed to follow

$$\lim_{y_{\phi_1} \to \infty} P(Y_{\phi_2} \leq a_{\phi_1}(\phi_2) \cdot y_{\phi_1} + z_{\phi_2|\phi_1} \cdot y_{\phi_1}^{b_{\phi_1}(\phi_2)} | Y_{\phi_1} = y_{\phi_1}) = G_{\phi_2|\phi_1}(z_{\phi_2|\phi_1}), \tag{5.3}$$

where $a_{\phi_1}(\phi_2)$ and $b_{\phi_1}(\phi_2)$ are parametric functions which, in the terminology of (5.2), would be read for each fixed pair of directions (ϕ_1, ϕ_2) as $a_{\phi_2|\phi_1}$ and $b_{\phi_2|\phi_1}$. According to the continuous nature of $\phi \in (0, 2\pi]$ the parameters as components as

given by (5.2) have now to be represented by continuous cyclic functions.

The limiting distribution of $Z_{\phi_2|\phi_1}$, $G_{\phi_2|\phi_1}$, in (5.3) is assumed to be non-degenerate. The estimation procedure further requires specification of the mean $\mu_{\phi_2|\phi_1}$ and standard deviation $\sigma_{\phi_2|\phi_1}$ of the limiting distribution of $Z_{\phi_2|\phi_1}$. As the wind process can be assumed to vary smoothly over directions, it is natural to model all parameters by continuous functions satisfying circular boundary conditions.

5.2.1 Conditioning on a single direction

To model the distribution $Y_{\phi_2}|(Y_{\phi_1} = y_{\phi_1})$, where $\phi_2 \in (0, 2\pi]$ while ϕ_1 is fixed, we account for directional variation by using m harmonic terms given by

$$\eta_{\zeta,\phi_1}(\phi_2) = \gamma_\zeta + \sum_{j=1}^{m} \beta_{\zeta,j} \cos(j\phi_2 - \omega_{\zeta,j}) \tag{5.4}$$

with $\zeta \in \{a, b, \mu, \sigma\}$. For reasons of identifiability we let $\gamma_\zeta \in \mathbb{R}$, but restrict the amplitude parameters $\beta_{\zeta,j} \geq 0$ and location parameters $\omega_{\zeta,j} \in (0, 2\pi]$. Instead of using (5.4), we employ the modified version

$$\eta_{\zeta,\phi_1}(\phi_2) = \gamma_\zeta + \sum_{j=1}^{m} \beta_{\zeta,j} \cos(j\phi_2 - \omega_{\zeta,j} - \phi_1) \tag{5.5}$$

to account for variation purely induced by changing ϕ_1. To ensure that the range conditions of the individual parameters are satisfied, a further transformation or link function ρ is suggested to each of the harmonic terms, for example $\mu_{\phi_1}(\phi_2) = \rho_\mu(\eta_{\mu,\phi_1}(\phi_2))$ and the other parameters similarly. Derived from common likelihood techniques, the objective function for a fixed direction ϕ_1 is given by

$$-\sum_{\phi_2 \in \Omega_2} \left\{ n_{\phi_1} \log(\sigma_{\phi_1}(\phi_2)) + b_{\phi_1}(\phi_2) \sum_{i=1}^{n_{\phi_1}} \log(y_{\phi_1,i}) + \frac{1}{2} \sum_{i=1}^{n_{\phi_1}} \left[\frac{y_{\phi_2|\phi_1,i} - a_{\phi_1}(\phi_2) y_{\phi_1,i} - \mu_{\phi_1}(\phi_2) y_{\phi_1,i}^{b_{\phi_1}(\phi_2)}}{\sigma_{\phi_1}(\phi_2) \cdot y_{\phi_1,i}^{b_{\phi_1}(\phi_2)}} \right]^2 \right\}, \tag{5.6}$$

where Ω_2 is a finite subset of $(0, 2\pi]$ and $y_{\phi_1,i}$, $i = 1, \ldots, n_{\phi_1}$, are exceedances of a high dependence threshold u_{ϕ_1} in direction ϕ_1.

5.2.2 Global model, extension to conditioning on all directions

To extend the model to all directions $\phi_1 \in (0, 2\pi]$, we need to allow each of the parameters in expression (5.5) of all harmonic terms to vary over directions ϕ_1. These parameter functions are denoted by $\psi_\vartheta(\phi_1)$ with $\vartheta \in \{\gamma_\zeta, \beta_\zeta, \omega_\zeta\}$ and $\zeta \in \{a, b, \mu, \sigma\}$. As we assume the wind process to vary smoothly over directions, we restrict ourselves to functions ψ_ϑ being continuous and satisfying circular boundary conditions. Abbreviating terms to simplify notation we may restate function (5.5) as, for example,

$$\eta_a(\phi_1, \phi_2) = \psi_{\gamma_a}(\phi_1) + \sum_{j=1}^{m} \psi_{\beta_{a,j}}(\phi_1) \cos(j\phi_2 - \psi_{\omega_{a,j}}(\phi_1) - \phi_1) \tag{5.7}$$

and similarly for μ, σ, and b. As before we then use the link functions ρ to transform the harmonic functions to satisfy the appropriate range conditions, for example $a_{\phi_1,\phi_2} = \rho(\eta_a(\phi_1, \phi_2))$. The objective function then takes the form

$$-\sum_{\phi_1 \in \Omega_1} \sum_{\phi_2 \in \Omega_2} \left(n_{\phi_1} \log(\sigma_{\phi_1,\phi_2}) + b_{\phi_1,\phi_2} \sum_{i=1}^{n_{\phi_1}} \log(y_{\phi_1,i}) + \frac{1}{2} \sum_{i=1}^{n_{\phi_1}} \left[\frac{y_{\phi_2|\phi_1,i} - a_{\phi_1,\phi_2} y_{\phi_1,i} - \mu_{\phi_1,\phi_2} y_{\phi_1,i}^{b_{\phi_1,\phi_2}}}{\sigma_{\phi_1,\phi_2} \cdot y_{\phi_1,i}^{b_{\phi_1,\phi_2}}} \right]^2 \right), \tag{5.8}$$

with Ω_1 being a finite subset of $(0, 2\pi]$.

5.3 Implementation of the model

Commonly, all margins are estimated separately. However, in the present case, pooling information over directions can be used to improve on estimates of the margins.

We employ the model already discussed in Chapter 2. The model suggests a distribution for annual maxima by assuming model parameters to vary smoothly over directions. To improve estimates the k largest order statistics in each direction are used. After having chosen thresholds in each direction, standard transformations yield parameters of the corresponding GPD; together with the exceedance probability of the thresholds and the empirical distribution, this is everything required for the model given by (4.3). To analyse the dependence structure, margins are then transformed to the double exponential distribution.

We approach the global dependence model by first considering the conditional model $Y_{\phi_2}|(Y_{\phi_1} = y_{\phi_1})$ separately for $\phi_2 \in \Omega_2$. The data at hand are recorded over directions within 36 equally–spaced intervals, so $\Omega_2 = \{10, \ldots, 360\}$ given in degrees. To satisfy range conditions we suggested the use of link functions ρ of functions given by (5.5). In the following we choose

- the identity link for the parameters $\mu \in I\!R$
- an exponential transformation for $\sigma \in (0, \infty)$, that is $\sigma = \exp(\eta)$
- a modified logit link for $a \in [-1, 1]$ taking the form

$$\text{mod.logit}(\eta) = 2\left(\frac{\exp(\eta)}{1 + \exp(\eta)}\right) - 1$$

- a modification of an exponential transform for $b \in (-\infty, 1)$ given by $\text{mod.exp}(\eta) = 1 - \exp(-\eta)$

We consider now the choice of functions ψ_ϑ used in (5.7). Plots of ψ_ϑ based on separate estimation, that is conditioning on a single direction ϕ_1 only, carried out for all conditioning directions ϕ_1, again suggest harmonic terms to be a good choice to capture directional variation in ψ_ϑ. In the case of intercept and amplitude parameters harmonic terms itself seem to be most appropriate. We therefore use

$$\psi_\vartheta(\phi_1) = \lambda_\vartheta + \sum_{l=1}^{p} \kappa_{l,\vartheta} \cos(l\phi_1 + \nu_{l,\vartheta}) \tag{5.9}$$

for $\vartheta \in \{\gamma_\zeta, \beta_\zeta\}$ and $\zeta \in \{a, b, \mu, \sigma\}$, and restrict $\kappa_{l,\vartheta} \geq 0$, $\nu_{l,\vartheta} \in (0, 2\pi]$ while $\lambda_\vartheta \in I\!R$. To accommodate the periodic nature of ω_ζ, we use a modified version of (5.9) given by

$$\psi_\omega(\phi_1) = \left(\lambda_\omega + \sum_{l=1}^{p} \kappa_{l,\omega} \cos(l\phi_1 + \nu_{l,\omega}) \right) \bmod(2\pi), \tag{5.10}$$

where mod is the usual modulo function, giving the value of $\psi_\omega(\phi_1)$ up to a shift of $k \cdot 2 \cdot \pi$, where k is an integer. For reasons of identifiability we therefore constrain $\lambda_\omega \in (0, 2\pi]$.

Parameters are then estimated by maximizing expression (5.8) using a gradient-based optimization procedure; an analytic version of the gradient can be found in Appendix A. Since the model may contain a high number of parameters, a good choice of starting values is advisable to speed up the optimization process. A possibility to get these is using least square estimates from separate estimation.

5.4 Bootstrap

An important part of every statistical analysis is to assess the fit of an estimated model and the sampling variation of estimated parameters. Since the objective function (5.8) is not a proper likelihood, the corresponding asymptotic theory is not applicable in the present case.

More flexible but computationally expensive alternatives are methods based on bootstrap (Davison and Hinkley, 1997). As we are interested in extreme events, and want to allow for variation due to the resolving process and uncertainty of the dependence model, standard bootstrap methods are not applicable. More precisely, using simple non-parametric re-sampling from the original data set does not allow for more extreme values than those observed in the data set. Using a parametric version, on the other hand, will not be capable of maintaining the inherent dependence structure, as we do not have a fully parametric model for the dependence structure.

We therefore suggest a semi-parametric bootstrap, being a mixture of non-parametric and parametric, which is in a similar fashion to the one used by Heffernan and Tawn (2004). Once a model has been estimated from the data at hand, we are in possession of both a marginal model for components based on (4.3), which we refer to by (M), and a model describing the dependence structure.

In step one we sample with replacement from the original data. These data are then resolved to components for both sites ϕ_j, $j = 1, 2$. Having the component data, observations are ranked for each margin separately, so that for each observation $\tilde{R}^{+}_{\phi_j,t}$, $t = 1, \ldots, n$, we have its corresponding rank $L_{\phi_j,t} \in \{1, \ldots, n\}$, where n is the total number of daily data at hand.

In the second step, we re-sample from the estimated model (M) in each direction ϕ_j exactly the same number of data, as we obtained in step one, which we denote by $\tilde{R}^{*}_{\phi_j,i}$, $i = 1, \ldots, n$. Then we match for each pair of directions (ϕ_1, ϕ_2) the pair of parametrically sampled values according to the ranks obtained in step one, that is $(\tilde{R}^{*}_{\phi_1,L_{\phi_1,t}}, \tilde{R}^{*}_{\phi_2,L_{\phi_2,t}})$. This will keep up the dependence structure of the component data in terms of ordered size (ranks), though the functional relationship induced by components may not exactly be maintained. Additionally values obtained by this simulation procedure may be more extreme than those observed in the actual data set making it appropriate for analysing extreme values.

The model is then re-estimated for the new data-set based on pairs $(\tilde{R}^{*}_{\phi_1,L_{\phi_1,t}}, \tilde{R}^{*}_{\phi_2,L_{\phi_2,t}})$ using the procedure described previously. Repeating this procedure several times yields full sampling distributions of the parameters. Model selection based on the bootstrap is discussed in a later section.

5.5 Return-level estimation

Common quantities of interest in extreme value applications are high quantiles often referred to as return-levels. A related concept often employed is known as the return-period. The return-period is the average number of repetitions of a process required to produce a value at least as high as the return-level. Estimates of return-levels may be compared with corresponding observations to judge model fit.

In the present case, we are interested in extreme quantiles in direction ϕ_2 of station two, given an extreme event occurred at station one in direction ϕ_1. More precisely, we consider

$$P(\tilde{R}_{\phi_2} \leq r_{\phi_2,p} | \tilde{R}_{\phi_1} > r_{\phi_1}) = p, \quad r_{\phi_1} > u_{\phi_1}, \tag{5.11}$$

where u_{ϕ_1} is the dependence threshold in direction ϕ_1 above which we assume the dependence model to be valid. Equation (5.11) provides great flexibility to adjust for quantities of interest. For example, by adjusting p we find the quantile r_{ϕ_2} corresponding to an event at the first station known to be greater than r_{ϕ_1}. Apart from high quantiles, $p = 0.5$ yields the median at the second station with a particular extreme event for $\tilde{R}_{\phi_1}$. On the other hand, for a given known design value r_{ϕ_1} we can find the corresponding distribution of $\tilde{R}_{\phi_2}$ using (5.11). We may for example wonder, what is the probability of a train passing the second station first but facing an extreme event leading to derailment at station one. These and other quantities can be derived from equation (5.11).

To estimate the distribution given in (5.11) we proceed as follows. First we simulate $Y^*_{\phi_1}$ from a double exponential distribution. Given the simulated value $Y^*_{\phi_1}$ exceeds its dependence threshold u_{ϕ_1}, we additionally sample with replacement a value Z^* from the empirical distribution of $G_{\phi_2|\phi_1}(z)$, and compute $Y^*_{\phi_2}$ using (5.2) with parameters being replaced by their estimates. Thereafter, both margins are back-transformed to their original margins. Repeating this procedure N times yields pairs $(\tilde{R}^*_{\phi_1,j}, \tilde{R}^*_{\phi_2,j})$, $j = 1, \ldots, N$, from which we can empirically calculate the probabilities of interest.

A further important use of (5.11) is assessing the fit of the model. We therefore calculate return-levels for each pair of directions (ϕ_1, ϕ_2) based on the procedure described above. Model judgement is then based on the comparison of model estimates and empirical values. Using the bootstrap procedure proposed in the previous section, confidence intervals can be calculated to support assessing the fit of the model.

5.6 Application of the model to the data

In this section we apply the wind data of the two stations to the conditional model described above. As outlined before, the data at both stations are first resolved to components. Thereafter, distributions of the margins are estimated and their margins are transformed to the double exponential distribution. Based on these transformed data the dependence structure using the conditional approach described above is estimated. ϕ_1 refers to the considered direction at Hannover, while ϕ_2 denotes the corresponding one for Würzburg.

5.6.1 Model selection

An important part of every statistical analysis is to explore which parameters of a model are important to describe the underlying process producing the data. Since the objective function given by (5.8) is not a proper likelihood, standard likelihood methods for model selection are not applicable here. We therefore use the bootstrap procedure introduced above on which parameter selection is based on. We start using a model including many harmonic terms and subsequently discard those which are not relevant. We limit the number of harmonic terms to four. That is, for each η_ζ, $\zeta \in \{\mu, \sigma, a, b\}$, in (5.7) we start with $m = 4$. Furthermore, for each of the ψ_ϑ in (5.7) we also allow four harmonic terms. So the total number of parameters in the starting-model is 324, compared with 5184 parameters if all combinations of directions are estimated separately.

The first harmonic term has one oscillation and therefore picks up the basic structure of variation over directions. The higher the harmonic term the more it is oscillating, and it is natural to think that a harmonic term of higher order accounts for fine and subtle adjustments rather than reflecting the rough, basic structure. For this reason we start by considering the highest harmonic terms to decide on whether to discard them or not. More precisely, we repeat the whole bootstrap five times (as there is in each case four harmonic terms and one intercept), and after each bootstrap consider the highest harmonic term for each ψ_ϑ (36 in the present case) and either retain it in or reject it from the model. Note, that the approach here is selecting harmonic terms and not their individual parameters. So parameters of one harmonic term are either jointly discarded or together kept in the model.

A natural way to decide on whether a harmonic term is to be kept in the model is by checking if its amplitude is significantly different from zero. However, to check this is impossible with the current procedure as the amplitude values are for reasons of uniqueness bound to be non-negative so that no interval based on the bootstrap procedure can overlap zero; in fact with zero being the value on the boundary of the parameter space, it will almost never take on zero itself. As the bootstrap procedure is not a (purely) parametric one, but mainly based on resampling from original data, it is neither possible to simulate from a parametric distribution with a certain null hypothesis to test against.

Two possible approaches are as follows. The first one is based on the idea above to check if the amplitude is close to zero. We therefore set a small boundary-value (for example 0.01) and discard the considered harmonic term if a certain percentage of the bootstrap estimates are smaller than this bound. If this is the case, the corresponding amplitude parameter cannot be significantly greater than this bound, and its contribution to the model is not likely to be substantial. Of course the choice of this critical value is to some degree subjective, but apart of this the selection process is straight forward with an obvious decision rule.

The second possibility is the selection procedure already used in Chapter 3 by considering the variation of the location parameter of each harmonic term. If bootstrap estimates of the location parameter are highly concentrated around a certain value, the corresponding harmonic term is kept in the model; otherwise, if the location parameter shows, for example, high scattering over $(0, 2\pi]$ it is removed from the model. This approach is based on the assumption that the true underlying process is correctly described by the model. A harmonic term actually present in the underlying process has a fixed position, so bootstrap estimates will exhibit a rather high concentration around a certain value. Other behaviour of the bootstrap estimates rather suggests this harmonic term not to be present in the underlying process.

The selection procedure applied here is based on the second approach. We use plots of (circular-adjusted) kernel-density estimates of the location parameters and graphs where the location parameters are plotted against the corresponding amplitude pa-

rameter. In cases of difficulty, additionally to the variation of the location parameter the actual size of the amplitude is used to decide on its importance. Again subjectivity is an issue here; however, each individual judgement can reveal features of the data, which are not seen when just applying a straight forward decision rule.

For intercept parameters λ_ϑ in equation (5.9) of the terms ψ_γ and ψ_β, judgement based on whether confidence intervals include zero can be employed, as their range is allowed to be all over the real line. The intercepts of location parameters λ_ω in (5.10), ψ_ω, are restricted to $(0, 2\pi]$, and thus the second method based on its variation is appropriate.

Three typical examples from the graphs of the bootstrap are given in Figure 5.2. The plots of every row belong together with the left graph showing the kernel-density estimate of a chosen location parameter, while the right graph is showing the same estimates plotted against their corresponding amplitude. The first row exhibits a pronounced bimodal density of parameter estimates with modes having roughly a distance of π. From the corresponding plot on the right hand side it can be seen that any amplitude value can come from either of the two modes. It does not appear very sensible that one harmonic term and the one resulting from a shift of π (which is equivalent to a reflection of the term on its horizontal axes) are the same likely to be in the model. This harmonic term is therefore discarded from the model. The second example in the middle row shows almost a uniform distribution over the whole range of $(0, 2\pi]$; every value of the location parameter based on the bootstrap is roughly the same likely and therefore it is not sensible to be kept within the model. The final row shows a density highly concentrated roughly around π; the same can be seen from the right-hand plot, showing no change of location with different values for the size of the amplitude. This parameter is kept in the model.

The model-selection procedure finally yields a (3,4,1,2)-model with 76 parameters, that is three harmonic terms to describe the variation in μ, four for σ, one for a, and two for b. These are given in terms of the ψ_ϑ, $\vartheta \in \{\gamma, \beta, \omega\}$, as stated in (5.7). The ψ_ϑ-terms, in turn, are given by harmonic terms defined in (5.9) and (5.10), which are determined by the estimated parameters λ, κ, and ν. Table 5.1 shows the results and estimates for the final model.

To get an impression of the validity of the model, we compare estimates of the four parameters μ, σ, a, and b based on the model with separate estimates. By separate estimates we mean estimates which are individually obtained for each sector-combination (ϕ_1, ϕ_2). Figure 5.3 shows these separate estimates (points) for all directions $\phi_2 \in \{10, \ldots, 360\}$ and conditioning direction $\phi_1 = 90°$. The solid line is based on the final model as given by Table 5.1. For comparison we use the (4,4,4,4)-model with 324 parameters shown as dashed line. The plot shows that the μ-parameter is estimated quite well, and the a and b estimates are also close to their separately estimated counterparts. The σ-parameter is not captured exactly by the model, but the basic behaviour of the separate estimates is clearly exhibited. For ϕ_1=180° (see Figure 5.4) the adaptation to the separate model is less good but its general overall features are well reflected apart from estimates for b, which however have a smaller variation than it is the case for ϕ_1=90°; furthermore, the (4,4,4,4)-model with a much higher number of parameters is just slightly better than our final model. Similar behaviour is present for directions ϕ_1=270° and 360° (see Figures 5.5 and 5.6), where the final model ranges from reflecting the basic features of the separate estimates to an almost perfect description. For these directions the final model does not seem to appear worse than the (4,4,4,4)-model.

5.6.2 Calculation of return-levels

After having selected a model as outlined above, we consider now return-level calculation based on equation (5.11). Thus we compute pairs of values $(\tilde{R}^*_{\phi_1,j}, \tilde{R}^*_{\phi_2,j})$, $j = 1, \ldots, N$, which in turn result from simulating and calculating pairs $(Y^*_{\phi_1,j}, Y^*_{\phi_2,j})$ and back-transform them to the original margins. Given a simulated value $Y^*_{\phi_1,j}$ exceeds a high specified value the resulting $Y^*_{\phi_2,j}$ depends both on the estimated parameters for $a_{\phi_2|\phi_1}$ and $b_{\phi_2|\phi_1}$ and the stochastic residual $Z_{\phi_2|\phi_1}$. To calculate $Y^*_{\phi_2,j}$ we therefore need to sample from the distribution of $Z_{\phi_2|\phi_1}$. However, this distribution is just known empirically, which limits the number of possible values to simulate from.

We consider now a possibility to extend the number of values for $Z_{\phi_2|\phi_1}$ which to simulate from. The approach suggested here is to check for possible inclusion of

term	noht	parameter values of harmonic terms
ψ_{γ_μ}	0	-0.00079
$\psi_{\beta_{\mu,1}}$	3	0.5040, 0.4496, 1.3130, 0.0723, 0.1548, 0.0410, 1.9370
$\psi_{\omega_{\mu,1}}$	2	6.1479, 0.9422, 0.0839, 0.1914, 2.1965
$\psi_{\beta_{\mu,2}}$	0	0.0099
$\psi_{\omega_{\mu,2}}$	0	0.9989
$\psi_{\beta_{\mu,3}}$	2	0.0791, 0.04407, 0.9321, 0.0138, 3.8333
$\psi_{\omega_{\mu,3}}$	1	5.4787, 0.3366, 0.6283
ψ_{γ_σ}	4	0.1081, 0.0623, 1.0657, 0.1056, 5.8802, 0.0170, 0.6690, 0.0168, 3.0884
$\psi_{\beta_{\sigma,1}}$	0	0.0134
$\psi_{\omega_{\sigma,1}}$	0	3.1090
$\psi_{\beta_{\sigma,2}}$	2	0.0816, 0.1721, 4.4536, 0.0318, 2.4764
$\psi_{\omega_{\sigma,2}}$	0	1.2293
$\psi_{\beta_{\sigma,3}}$	0	0.0073
$\psi_{\omega_{\sigma,3}}$	0	2.6375
$\psi_{\beta_{\sigma,4}}$	1	0.0212, 0.0128, 4.3316
$\psi_{\omega_{\sigma,4}}$	2	0.9010, 0.3750 ,0.6121, 0.7189, 1.4789
ψ_{γ_a}	1	-0.0238, 0.0232 ,1.1076
$\psi_{\beta_{a,1}}$	2	1.0203, 0.7547, 4.3332, 0.2359, 2.1837
$\psi_{\omega_{a,1}}$	3	6.0740,0.8031, 3.2805, 0.1691, 0.2368, 0.1496, 2.3879
ψ_{γ_b}	2	0.3172, 0.0722, 3.6360, 0.1755, 2.7258
$\psi_{\beta_{b,1}}$	0	0.0155
$\psi_{\omega_{b,1}}$	0	3.1408
$\psi_{\beta_{b,2}}$	1	0.0589, 0.1706, 1.1950
$\psi_{\omega_{b,2}}$	0	1.168

Table 5.1: The final model: First column shows the term considered, the second the number of harmonic terms (noht) to describe the term in the first column (with 0 being just an intercept parameter), while the last column gives the estimates of this harmonic terms.

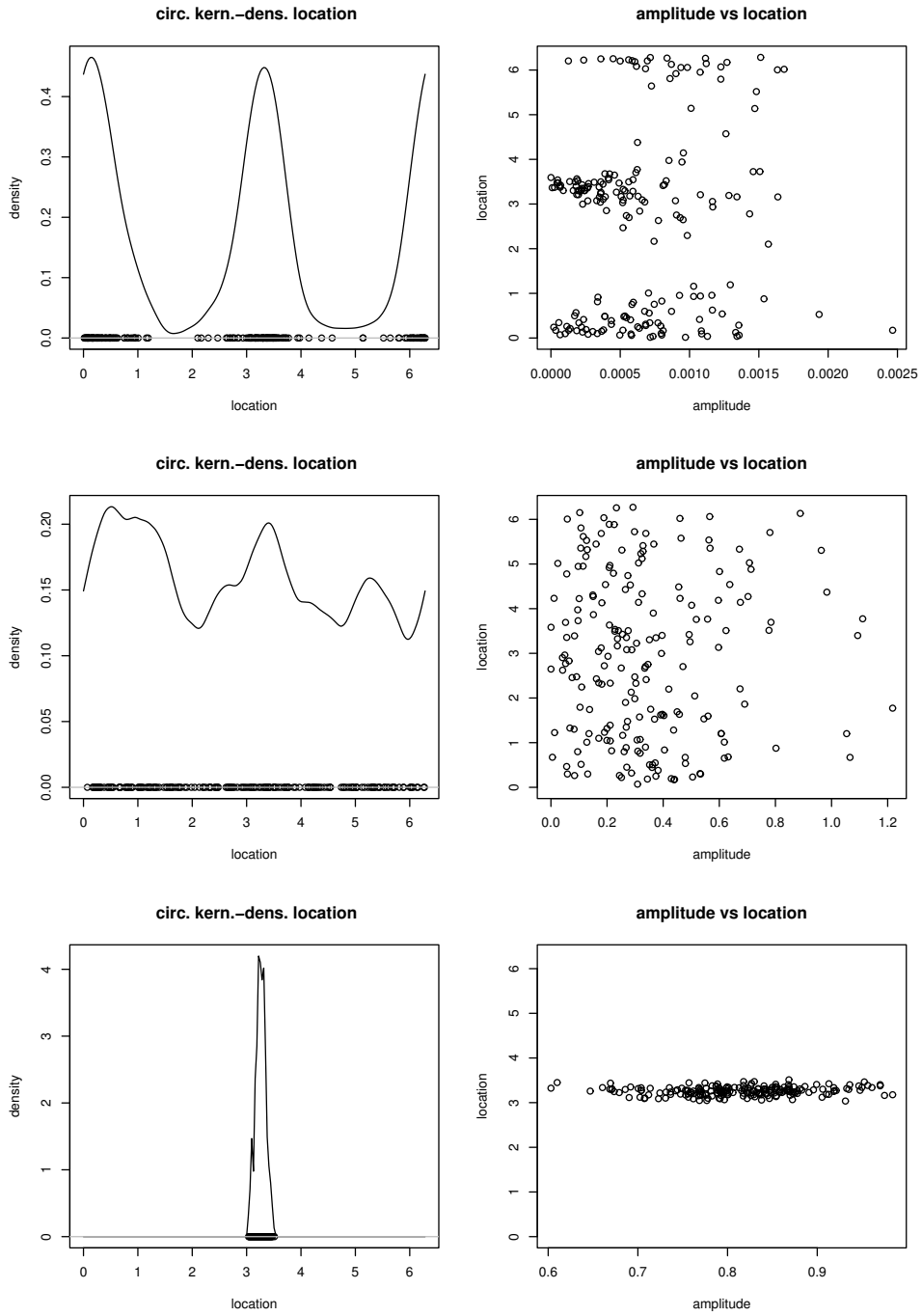

Figure 5.2: Three typical examples from bootstrap procedure; each row is for the same harmonic term with the left plot showing the density estimate of the location parameter and the right plot the very same location parameter against its corresponding amplitude parameter.

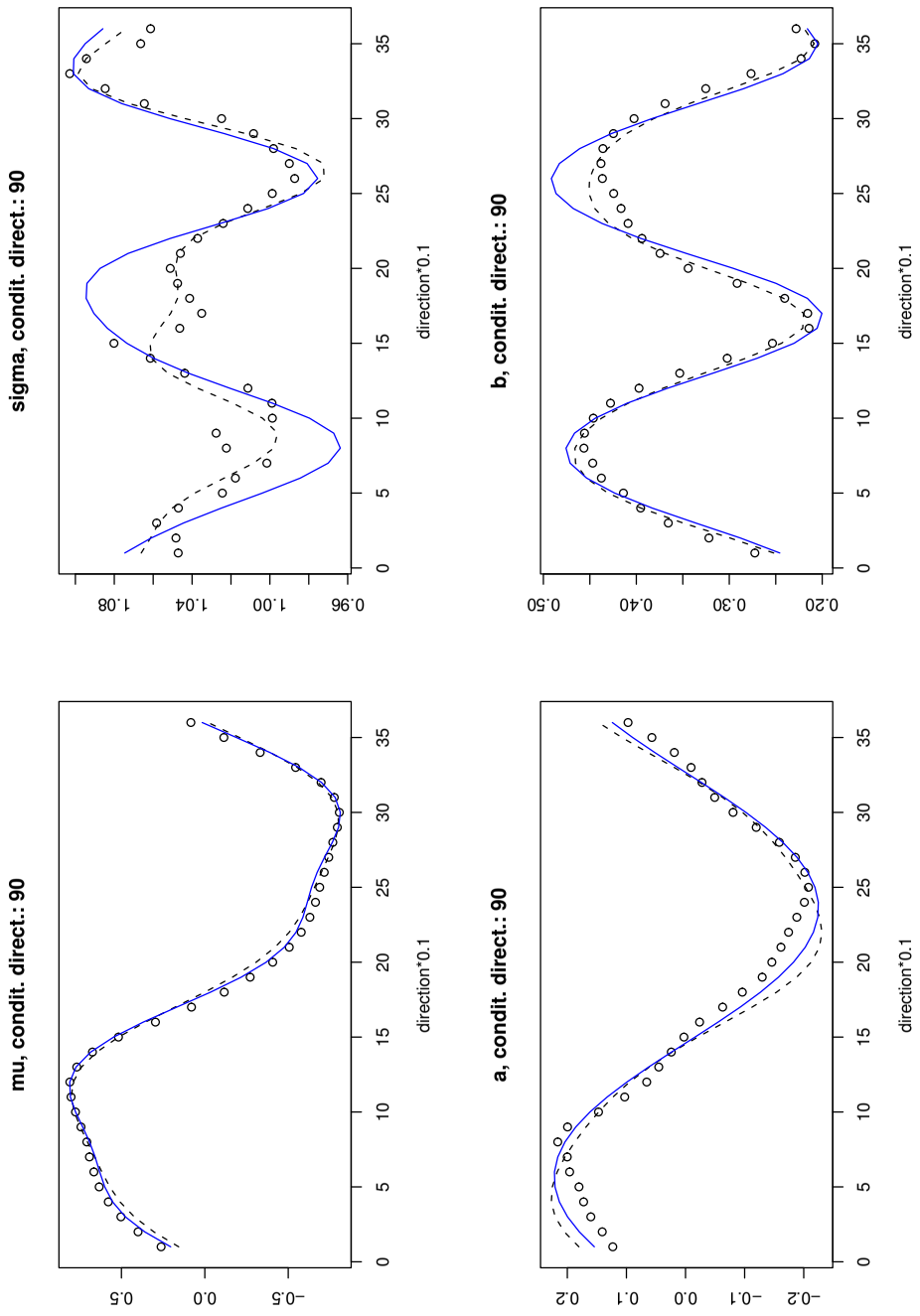

Figure 5.3: The final model (solid line) given that the observation in ϕ_1=90° exceeded its dependence threshold showing parameters μ, σ, a, and b for all directions ϕ_2 at the other station. The points show individual estimates when each sector combination is considered separately. For comparison the full (4,4,4,4)-model was included (dashed line).

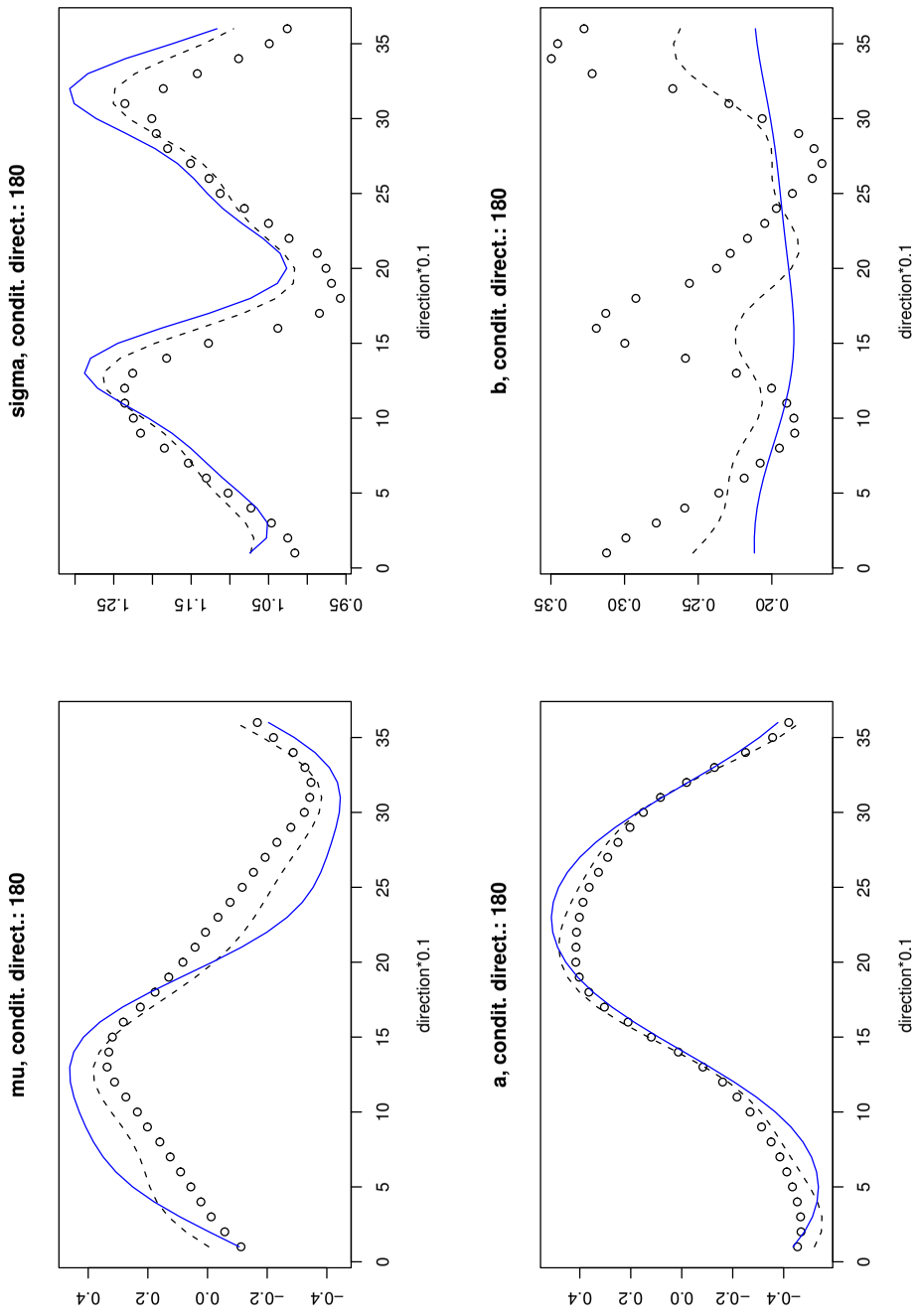

Figure 5.4: The final model (solid line) given that the observation in ϕ_1=180° exceeded its dependence threshold showing parameters μ, σ, a, and b for all directions ϕ_2 at the other station. The points show individual estimates when each sector combination is considered separately. For comparison the full (4,4,4,4)-model was included (dashed line).

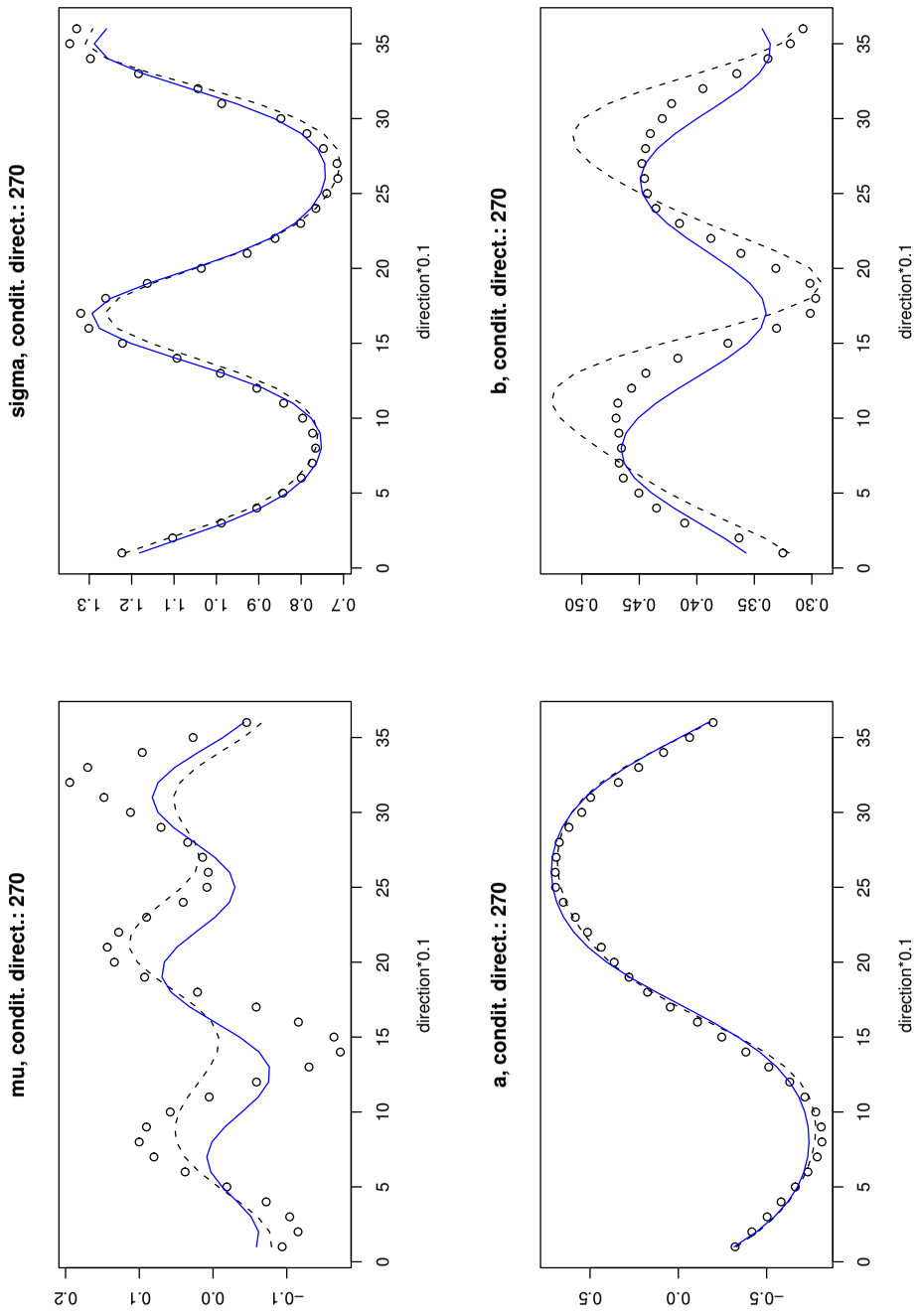

Figure 5.5: The final model (solid line) given that the observation in ϕ_1=270° exceeded its dependence threshold showing parameters μ, σ, a, and b for all directions ϕ_2 at the other station. The points show individual estimates when each sector combination is considered separately. For comparison the full (4,4,4,4)-model was included (dashed line).

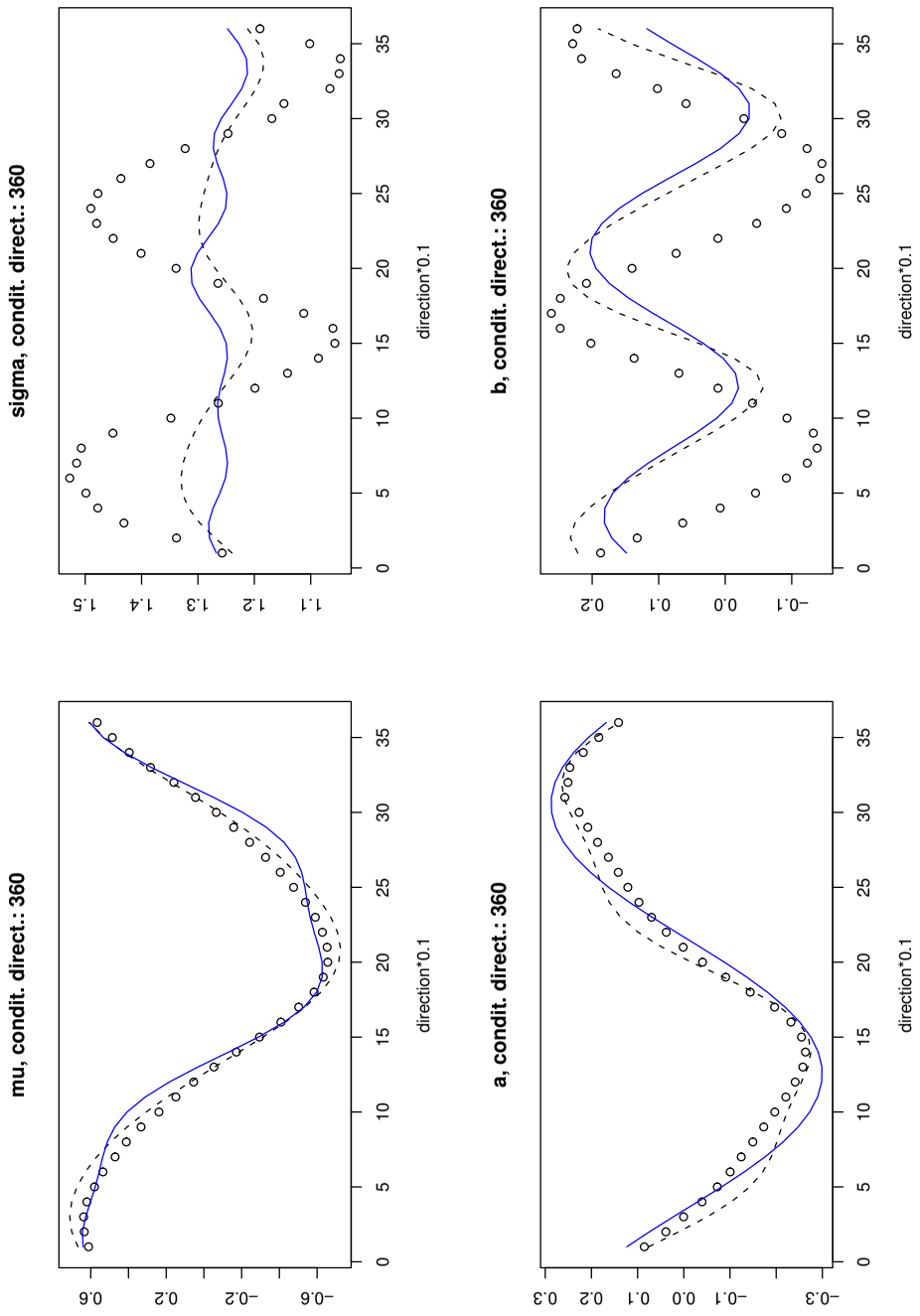

Figure 5.6: The final model (solid line) given that the observation in ϕ_1=360° exceeded its dependence threshold showing parameters μ, σ, a, and b for all directions ϕ_2 at the other station. The points show individual estimates when each sector combination is considered separately. For comparison the full (4,4,4,4)-model was included (dashed line).

residuals of neighbouring directions which are, due to local proximity, supposed to be sufficiently similar in distribution. The procedure applied is as follows. We first normalize all residuals in all directions ϕ_1 and ϕ_2,

$$Z^s_{\phi_2|\phi_1,j} = (Z_{\phi_2|\phi_1,j} - \mu_{\phi_2|\phi_1})/\sigma_{\phi_2|\phi_1}.$$

Thereafter we consider the distribution of $Z^{+1}_{\phi_2^*|\phi_1^*}$ which consists of joining all $Z^s_{\phi_2^*|\phi_1^*}$ where $(\phi_1^* \times \phi_2^*) \in \{\phi_1 - 10°, \phi_1, \phi_1 + 10°\} \times \{\phi_2 - 10°, \phi_2, \phi_2 + 10°\}$. Let the empirical distribution of $Z^s_{\phi_2|\phi_1}$ consist of m observed values. Then we can simulate m values from the distribution of $Z^{+1}_{\phi_2^*|\phi_1^*}$ and re-arrange these values in an increasing order. Repeating this simulation process several times allows for calculating pointwise confidence intervals on the ordered values (we used 95%−confidence intervals based on 200 simulations). If the ordered values of the observed $Z^s_{\phi_2|\phi_1}$ keep within this confidence limits, the distributions are not significantly different from one another and we can use the set $Z^{+1}_{\phi_2^*|\phi_1^*}$ to simulate from. In this case we consider a further extension using $Z^{+2}_{\phi_2^*|\phi_1^*}$ where $(\phi_1^* \times \phi_2^*) \in \{\phi_1 - 20°, \ldots, \phi_1 + 20°\} \times \{\phi_2 - 20°, \ldots, \phi_2 + 20°\}$ and repeat the above calculation of confidence intervals and compare with the distribution of $Z^s_{\phi_2|\phi_1}$. We continue this procedure by considering $Z^{+t}_{\phi_2^*|\phi_1^*}$ with $(\phi_1^* \times \phi_2^*) \in \{\phi_1 - t \cdot 10°, \ldots, \phi_1 + t \cdot 10°\} \times \{\phi_2 - t \cdot 10°, \ldots, \phi_2 + t \cdot 10°\}$, $t = 3, 4, \ldots$, until the first time that the ordered $Z^s_{\phi_2|\phi_1}$ do not keep inside the calculated confidence interval. Of course for all t the values $\phi_j \pm t \cdot 10°$, $j = 1, 2$, are adapted to the circular nature insuring to be within $\{10°, \ldots, 360°\}$. We then take the largest set $Z^{+t}_{\phi_2^*|\phi_1^*}$ where $Z^s_{\phi_2|\phi_1}$ kept within the intervals to simulate residuals from; this set is denoted by $Z^{+\max}_{\phi_2^*|\phi_1^*}$. Having simulated (with replacement) a value $Z^{s*}_{\phi_2^*|\phi_1^*,j} \in Z^{+\max}_{\phi_2^*|\phi_1^*}$, $j = 1, \ldots, N$, it is back-transformed using $Z^*_{\phi_2|\phi_1,j} = \sigma_{\phi_2|\phi_1} \cdot Z^{s*}_{\phi_2^*|\phi_1^*,j} + \mu_{\phi_2|\phi_1}$.

As described above we then simulate a value $Y^*_{\phi_1,j} > v_{r_{\phi_1}}$, where $v_{r_{\phi_1}} > u_{\phi_1}$ is the value r_{ϕ_1} given in (5.11) transformed to double exponential scale. Using $Y^*_{\phi_1,j}$ and $Z^*_{\phi_2|\phi_1,j}$, we calculate $Y^*_{\phi_2,j}$ via

$$Y^*_{\phi_2,j} = a_{\phi_2|\phi_1} \cdot Y^*_{\phi_1,j} + Z^*_{\phi_2|\phi_1,j} \cdot {Y^*_{\phi_1,j}}^{b_{\phi_2|\phi_1}},$$

where the model parameters $a_{\phi_2|\phi_1}$ and $b_{\phi_2|\phi_1}$ are replaced by their estimates. Thus, a pair $(Y^*_{\phi_1,j}, Y^*_{\phi_2,j})$ is obtained, which is then back-transformed to its original margins yielding the desired pair $(\tilde{R}^*_{\phi_1,j}, \tilde{R}^*_{\phi_2,j})$.

We consider now the procedure just described in the application at hand using the model results of the previous section. Figure 5.7 illustrates the steps of the selection process of the normalized residuals. In case of the first three extension steps (top line plots and bottom left plot) using neighbouring residuals, namely $Z^{+h}_{\phi_2^*|\phi_1^*}$, $h = 1, 2, 3$, the observed $Z^s_{\phi_2|\phi_1}$ keep well within the confidence limits. The fourth extension (bottom right plot), in contrast, exhibits significant difference between the distributions of $Z^s_{\phi_2|\phi_1}$ and $Z^{+4}_{\phi_2^*|\phi_1^*}$. Consequently the residuals are simulated from the largest set not in disagreement with $Z^s_{\phi_2|\phi_1}$, that is $Z^{+\max}_{\phi_2^*|\phi_1^*} = Z^{+3}_{\phi_2^*|\phi_1^*}$.

Having simulated the residuals and values $Y^*_{\phi_1}$ the corresponding $Y^*_{\phi_2}$ result immediately. The top left plot of Figure 5.8 shows for a simulation size of $n = 5000$ the results for exceedances of the value $v_{r_{\phi_1}}$ corresponding to $r_{\phi_1} = 21$ m/s of the conditioning variable in direction $\phi_1 = 270°$. Included lines highlight the conditional distribution of Y_{ϕ_2} given $Y_{\phi_1} > v_{r_{\phi_1}}$ for the combination $(\phi_1, \phi_2) = (270°, 250°)$; the middle line represents the mean response of Y_{ϕ_2} given y_{ϕ_1}, surrounded first by 0.25- and 0.75-quantiles, followed by 0.1- and 0.9-quantiles and finally by 0.005- and 0.995-quantiles. Both, points and lines, clearly indicate a positive relation between conditioning and conditioned variable. Back-transformation to original margins is shown in the bottom left plot of Figure 5.8 represented by black circles. For comparison, corresponding data points where super-imposed indicated by crosses. The plot also indicates that the model provides a good description of the data. When interest is entirely in the marginal conditional distribution of $\tilde{R}_{\phi_2}$, kernel density plots are an appropriate choice of representation; for the current combination of directions this is given in the top-right plot.

Equation (5.11) can also be applied by fixing a certain value p and r_{ϕ_1}, and thereby considering different quantiles of the distribution $\tilde{R}_{\phi_2}$. Lets assume, for example, that at the first station we know that the wind speed in direction $\phi_1 = 270°$ exceeds 23 m/s, we may want to know the median wind speed at station two. In the top-left plot of Figure 5.9, based on model results, this is given for all directions by the solid line, using a simulation size of $n = 250000$. Its empirical counterpart is superimposed by circles. The same figure also shows results for higher quantiles. For most of the quantiles there is good agreement between empirical and model results.

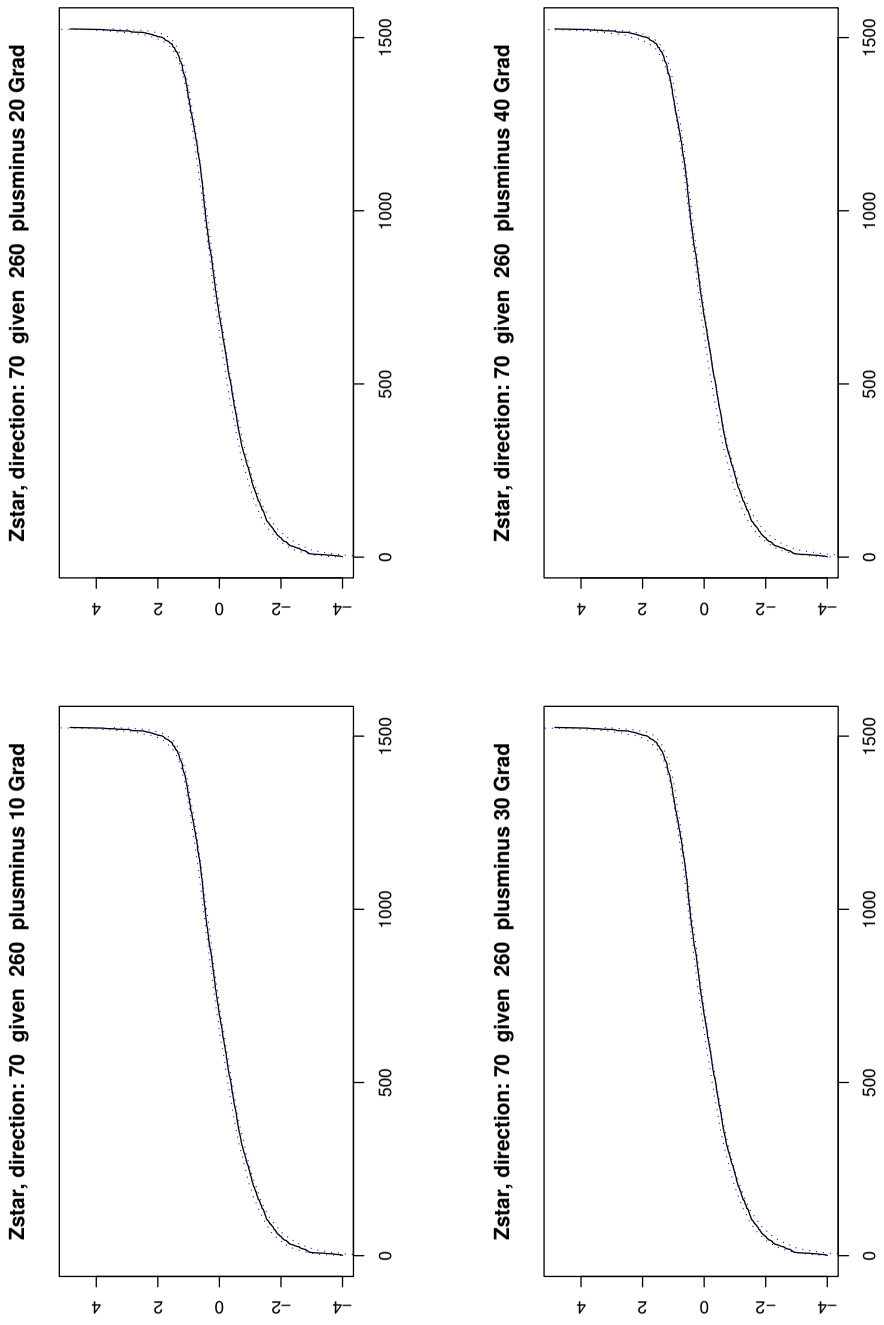

Figure 5.7: Ordered standardised residuals $Z^s_{\phi_1|\phi_2}$ (solid line) for (ϕ_1, ϕ_2)=(260°,70°) with 95%−confidence intervals (dotted line) based on the sets $Z^{+1}_{\phi_2^*|\phi_1^*}$ (top left), $Z^{+2}_{\phi_2^*|\phi_1^*}$ (top right), $Z^{+3}_{\phi_2^*|\phi_1^*}$ (bottom left), and $Z^{+4}_{\phi_2^*|\phi_1^*}$ (bottom right).

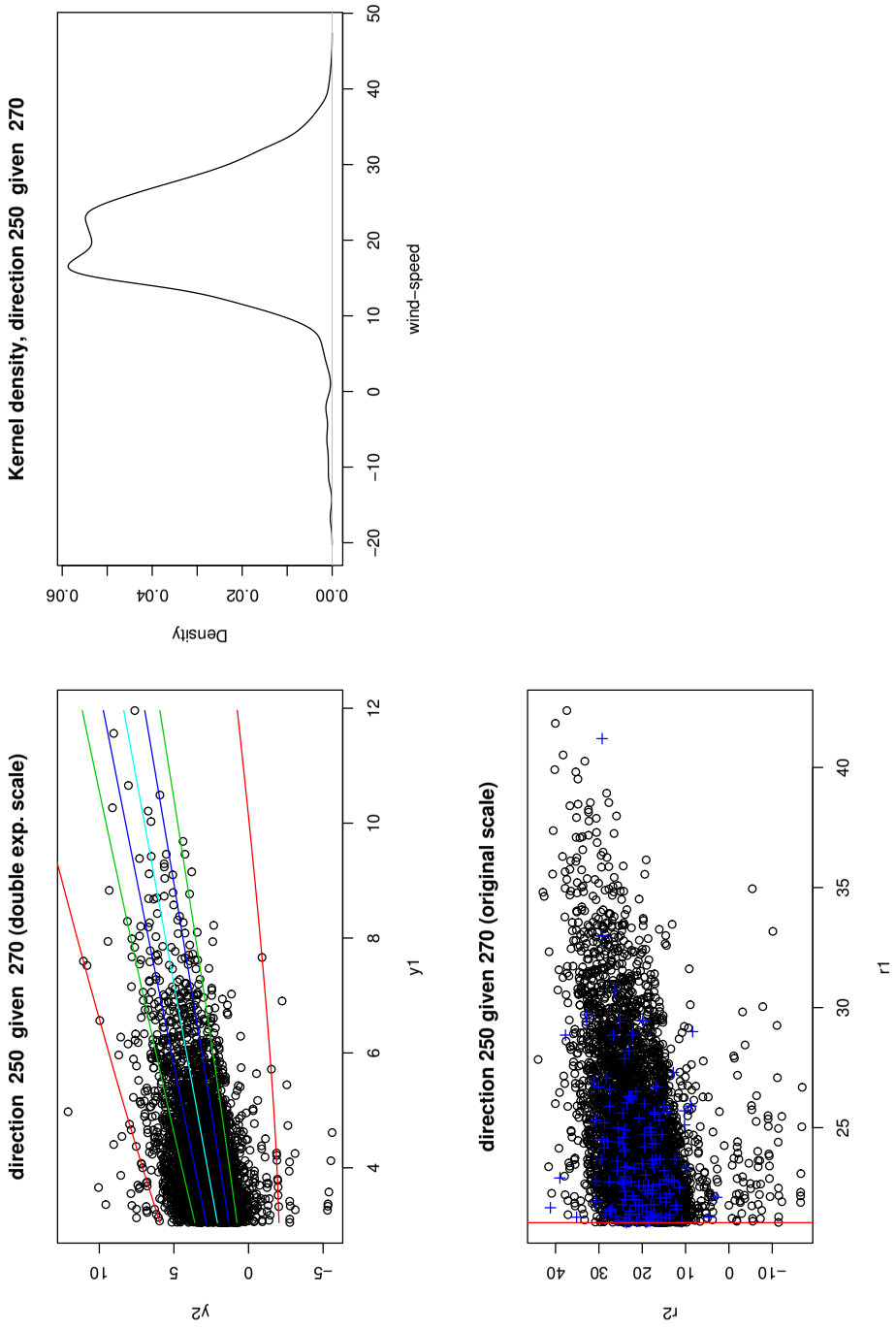

Figure 5.8: Top left: simulated values (circles) with lines representing different quantiles from final model; bottom left: simulated values (circles) on original scale and corresponding original data (crosses); top right: kernel density estimate for the conditional distribution.

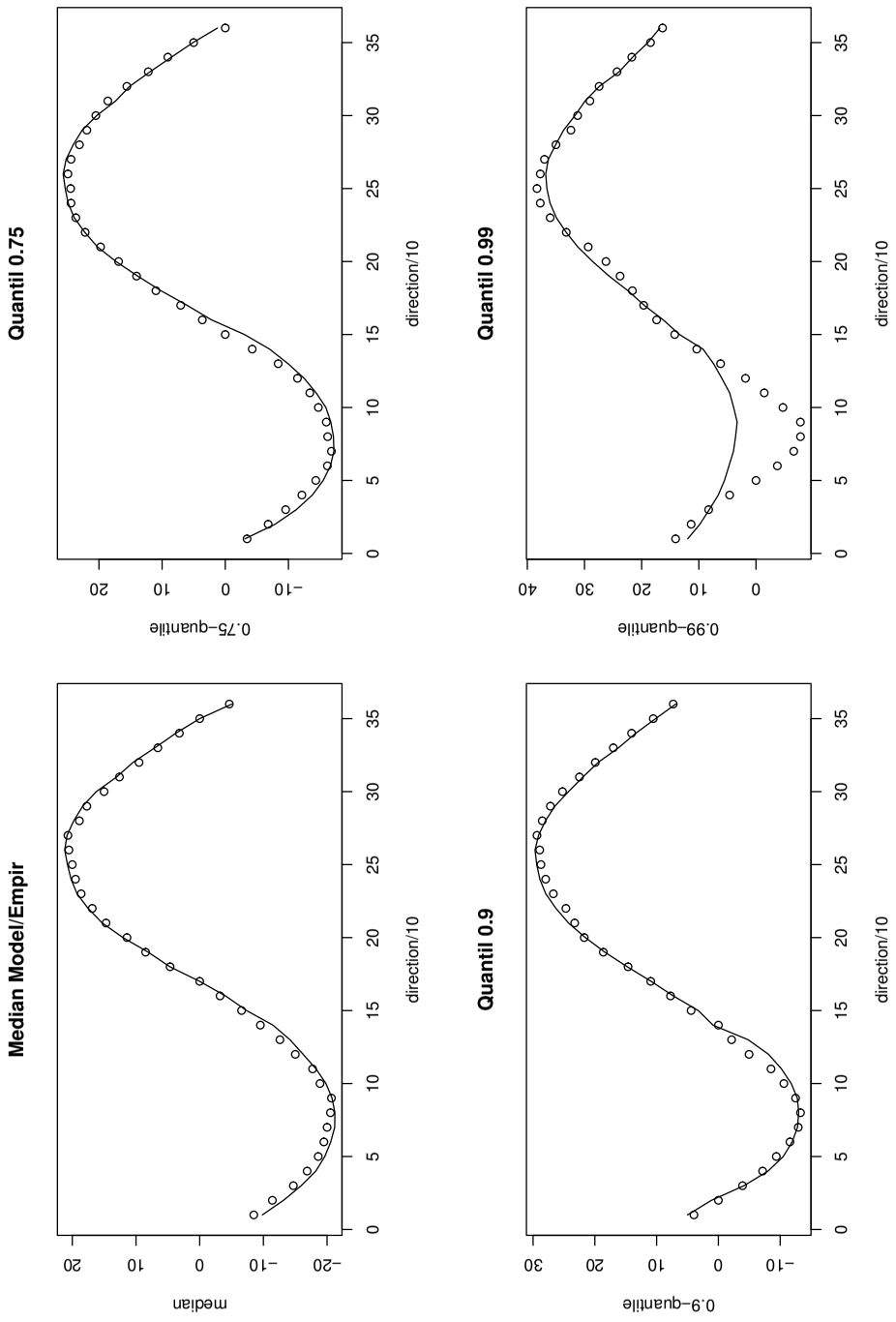

Figure 5.9: Different quantiles for the conditional distribution with lines for the model and circles based on data.

A slight departure of this can be observed for easterly directions (around 90°) for the 99%−quantile in the bottom right plot. It should be kept in mind, however, that the number of empirical observations exceeding this quantile, is comparatively small; the smooth appearance of these empirical points is to be attributed to the component nature of the data and should not be misinterpreted as resulting from a high number of exceedances entering the calculation of the quantile. These results again suggest a good fit of the model to the data.

The dominant wind direction at the two stations is west. It is therefore natural to closer examine these directions. Figure 5.9 shows different quantiles given a high wind speed in a western direction at the first station. Roughly half of the wind components in western direction of the second station reach 20 m/s or more, every tenth occurrence reaches almost 30 m/s or more, while one out of hundred events can be close to 40 m/s or above. In contrast, quantiles in easterly directions tend to be negative. As negative speeds are defined as the same wind speed in the opposite direction, this fact clarifies the non-independent wind behaviour at the two stations.

The plots of Figure 5.9 show the necessity for a model accommodating a range of different dependencies to describe the changing behaviour of the wind process over directions. Clearly, the conditional model provides this necessary flexibility. With adaptations to the conditional model suggested in this paper, smooth changes of the wind process over directions can be captured by the model; these smooth changes are even possible, and in the present case necessary, at the transition from positive to negative dependence.

While the plots provide examples within the range of the data to examine the fit of the model, it easily extends to ranges where no data are available and in which it is the only possibility to make sensible judgement from. This can supply important information for decisions on how to run high speed trains in the presence of extreme wind occurrences. With the north-south orientation of the track, wind forces from east and west pose the biggest risk to the train. The joint behaviour, that is the occurrence of high wind components at the second station in westerly direction when they are present at the first station, makes it very likely, that strong wind events are also present in between this two stations. A possible consequence may be lower

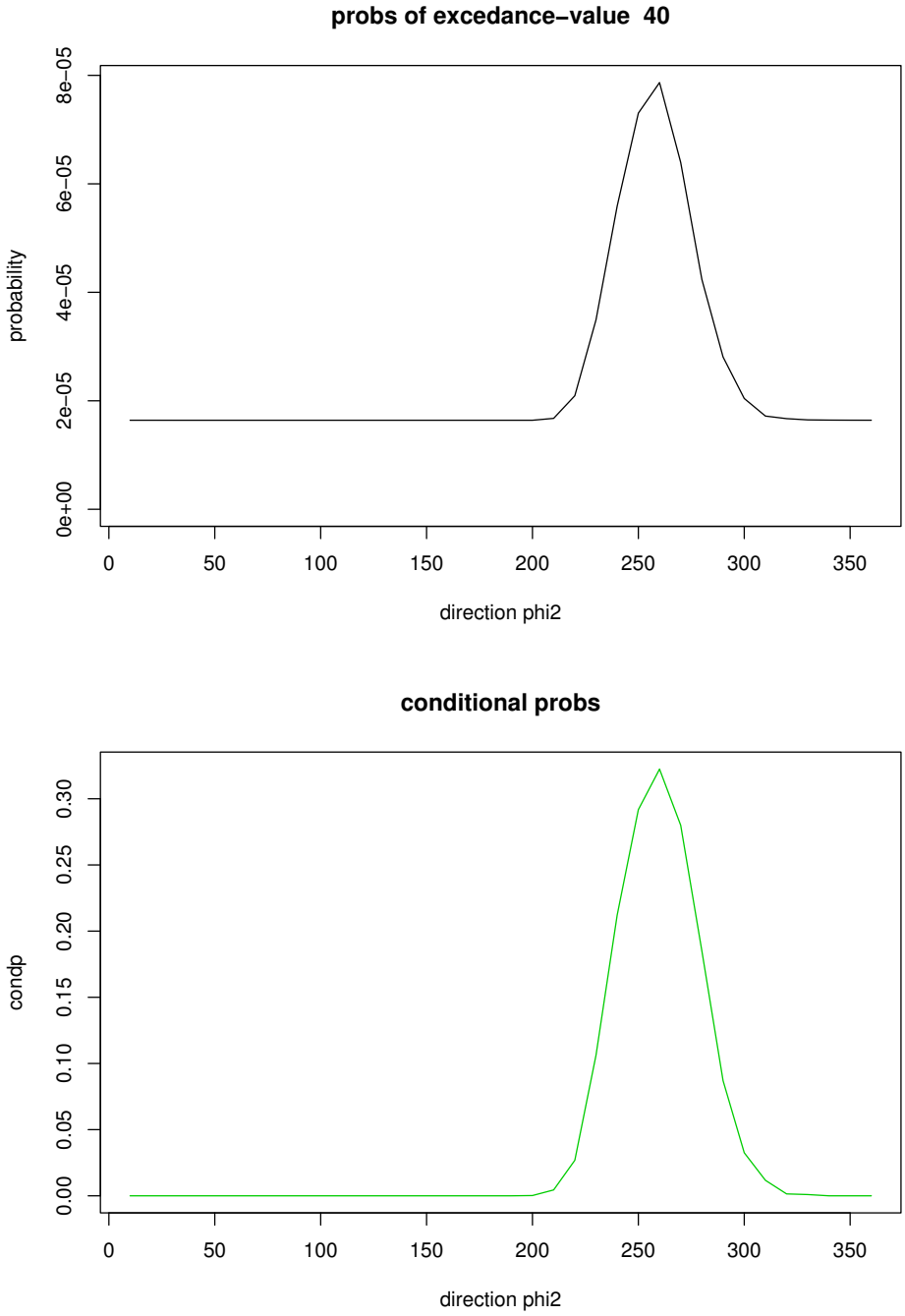

Figure 5.10: Top: Shows the probability for one day of observing a wind speed exceeding at least at one station 40m/s over all directions ϕ_2 for fixed ϕ_1=270°. Bottom: Shows the corresponding conditional probabilities over directions.

train-speeds at certain points or over the whole track when extreme wind conditions are present.

Let us consider again equation (5.1) for a simplified track consisting just of two stations. For both stations we consider the event of exceeding the wind speed 40m/s, that is $x = y = 40$, when ϕ_1=270°. The probability for observing this event at least at one station is given in top of Figure 5.10, while the corresponding conditional probabilities required in the calculation are given in the bottom. The probabilities correspond to a one-day period. With the major wind direction being west and fixing ϕ_1 to be in this direction, the probability of observing such a strong gust at least at one station is largest in western directions. With a strong wind event in direction ϕ_1=270°, the conditional probability clearly exhibits the dependence of a big event at the second station to occur in a similar direction than the one at the first station did.

5.7 Discussion

An important question in assessing the risk of storm events is whether strong gusts occur rather localized or if they are present over extended parts of the track. This requires knowledge of the dependence structure of the wind process at extreme levels. A crucial aspect in the context of high speed trains is the wind direction, which has to be taken into account. As we are interested in the force of the wind, resolved components are analysed. The dependence structure changes with the combination of directions considered. Thus a flexible model is required accommodating a wide range of dependence structures.

When simplifying the track to just two stations, a necessary component to calculate the probability of at least one station facing an extreme storm is a conditional probability comprising the dependence structure. The conditional approach for multivariate extremes is thus a natural candidate. Its high flexibility in incorporating different types of dependence supports this choice, as this is feature is required in the present application.

The Gumbel distribution is originally employed in the conditional method as com-

mon margins for the dependence function. It consists of two different parametrisations differing in the number of parameters. As we require smooth transition from positive to negative extremal dependence, we take the double exponential distribution for the common margins instead. This choice allows to have the same parameters for all types of dependence.

To enable transfer of information over directions, a functional relationship accounting for directional variation of the parameters is imposed. This functional relationship depends on both directions of the two stations, which is handled in two stages. In both stages harmonic terms or adaptations of them are taken, as these terms turn out to be very flexible and satisfy circular boundary conditions.

The objective function is derived from common likelihood methods, but is not a real likelihood in itself. Thus inference has been based on a bootstrap procedure. This procedure has been adapted to requirements essential for extreme values and component data. Model selection based on this bootstrap procedure can not be carried out in the usual way; we have proposed a method applicable in the present situation based on the variation of bootstrap estimates of the location parameter.

We have considered return-level estimation, which is a common quantity of interest in extreme value statistics. Return-level estimation was based on simulation. As this requires estimation from the empirical distribution function of the residual distribution, we suggest possibilities to extend the number of values of the residual distribution to simulate from by considering neighbouring directions.

A way to assess the fit of the model is by comparison of model based quantities with the corresponding ones of the data. The return-levels are a natural choice for this comparison. In the present study there is good agreement between model-based return-levels and their equivalent based on the data. This suggests that the model provides a good basis for judgement when considering levels beyond the data.

Chapter 6

Summary

In many situations strong wind gusts pose a severe risk for the system under consideration to fail. These situations include design structures and also the stability of high speed trains, which we are particularly interested in. The direction of strong gusts plays an important role so the analysis of storm events need to account for this feature. Extreme value theory provides us with the necessary tools allowing us to make judgement at extreme levels and even beyond the range of the data. Although directionality of strong wind events plays an important role in many applications, it has not found very much attention in most analysis. The aim of this thesis is to develop extreme value models taking directionality of strong wind events into account.

We have first considered univariate extreme value theory and related statistical aspects relevant for the present work. The data analysed consist of 22 years of daily maxima for the towns Würzburg and Hannover, which are located at each endpoint of a highspeed track. The track was chosen since the dominant wind directions, west and east, are perpendicular to the train's motion thus posing the highest risk for derailment. We have another data set consisting of ten-minutes data for Würzburg from a shorter observation period, which gives insight into the wind behaviour within a day. Two types of data are considered which are of interest in wind applications: raw data and component data representing the force of the wind in a certain direction.

Though the wind process may naturally be represented as bivariate with components wind speed and wind direction, it is easier to consider them separately by splitting them up into the wind direction and the wind speed given its direction; this allows the conditional distribution wind speed given its direction to be treated as univariate. In Chapter 2 we analyse daily wind data of Würzburg using an extreme value model employing the k largest order statistics in each direction. Variation in direction is accounted for by allowing the distributional parameters to vary with directions according to a functional relationship given by harmonic terms; harmonic terms are very flexible and satisfy circular boundary conditions. Once the model is selected and parameters are estimated, we consider for a fixed direction two quantities: return-levels (or quantiles) providing information about the most extreme value to be expected within a given time-period and probabilities of exceeding certain given design values within one year. In the railway application return-levels can be employed to judge which wind-speeds on a longer time horizon have to be expected, exceedance probabilities are useful once knowledge about a certain wind speed likely to cause derailment is determined by engineers. For both quantities we consider confidence intervals based on the delta method and profile likelihood method. The former is easier to apply but results in symmetric intervals; the latter is computationally more expensive but allows for asymmetric intervals. Results of a simulation study confirm that allowing for asymmetric confidence intervals is a more appropriate choice with this model.

The extreme value model employing the k largest order statistics was fitted to raw data and component data of Würzburg. We estimated return-levels and exceedance probabilities of design values for both data cases as well as related confidence intervals. Though the two types of data have a different definition of the problem, the model based on component data appeared to agree better with the data than the one based on raw data. This may be partly attributed to the masking problem, which can be regarded as a shortcoming of the common recording mechanism in extracting the largest daily observation only but leaving observations in all other directions unrecorded. To get further insight we compare for each direction annual maxima of ten-minutes data with annual maxima of the daily data and resolved data. The comparison clearly shows the presence of masking in the case of daily raw data and an alleviation of this problem when using component data.

In Chapter 3 we suggest a model, which we refer to as subinterval model, to account for the masking problem and compare it with a classical modelling approach similar to the one in Chapter 2, which we refer to as daily maxima model. Instead of using order statistics we employ a threshold approach for both models. For the subinterval model we require additionally to daily maxima the information of directional occurrence frequency. As for many weather stations a data set of smaller time intervals is available at least for a short observation period, the directional occurrence distribution can be estimated empirically. The approach is to include additionally to the daily maximum the information that occurrences in all other directions are no larger than the daily maximum. Parameter estimation is carried out using a likelihood which possibly wrongly assumes independence between sub-intervals. In most extreme value applications interest is in return-levels or high quantiles; as for any fixed direction the number of occurrences within sub-intervals vary from interval to interval we use a version of return-levels taking this variation into account. The performance of the new approach is then investigated by using a simulation study and comparing the mean square errors of high quantiles of the subinterval model with those of the daily maxima model.

In the first study, data are simulated from a chosen subinterval model assuming independence between subintervals. The subinterval model was then re-estimated with the same number of harmonic terms as was used in the simulation and compared with a daily maxima model found by a forward selection procedure. The subinterval model turns out to be considerably better than the daily maxima model in terms of mean square error of high quantiles for all directions and choices of quantiles. A surprising result is that the bias in the case of the daily model is mostly small, and the size of the mean square error is rather due to the variance contribution; a larger bias for the daily maxima model is mainly found in directions with a low occurrence frequency.

For data sets without sub-interval information it is tempting to substitute this information from a neighbouring one assuming differences to be little. This requires the subinterval model to be robust against not correctly specified directional occurrence distributions. A simulation study was carried out analysing different degrees

of deviation of the directional occurrence distribution from the true one which the data are simulated from. The subinterval model turns out to be robust against departures from a correctly specified occurrence distribution.

The strong assumption of independence between sub-intervals in the likelihood function may not hold in most real data applications but environmental time series often exhibit serial correlation instead. A simulation study using different degrees of dependence of successive sub-interval values was carried out. Although the performance of the subinterval model is slightly worse than in the independence study there is still clear superiority over the daily maxima model.

Finally the new model was applied to two sets of real data consisting of daily maxima and their directions; for one of these data-sets an additional data-set with ten-minutes recordings was available for a shorter observation period. For the data set of Würzburg with additional sub-interval information the new model shows slight improvement. In the case of the Hannover data set without any sub-interval data the directional occurrence distribution was substituted by using the ten-minutes data of Würzburg. The estimated subinterval model of Hannover reflects much better the structure of the data than the daily maxima model, and return-level estimates of the largest observation within the recording period are much more in agreement with meteorological judgement.

In Chapters 4 and 5 we apply the multivariate conditional extreme value model. Chapter 4 gives a short overview over existing multivariate extreme value models and discusses their applicational shortcomings which the conditional model overcomes. One difficulty with the conditional multivariate model is to quantify exactly the degree of dependence between variables as this is given by parameters of two functions and a residual distribution. Based on the conditional model we study the behaviour of the well known scalar dependence measure $\bar{\chi}$ and different failure regions via simulation. Data are simulated from a bivariate normal distribution with a range of different correlations, which is a distributional choice where the extremal behaviour is well understood. The $\bar{\chi}$ measure shows good results for strong positive correlation, but is not reliable for non-positive dependence; a non-parametric equivalent of $\bar{\chi}$ is much more reliable in the presence of negative correlation. An

alternative to the scalar dependence measure is to consider visual summaries of dependence based on failure regions. Via simulation we studied the behaviour of four different failure regions and their ability to distinguish between different degrees of dependence. For positive dependence, summary measures based on all failure regions discriminate well between different degrees of correlation. However, just two of them are capable of separation in the case of non-positive correlation. While one of this remaining two summary measures fails clear discrimination, especially between strongly negatively correlated data, the other one provides good and clear results for all degrees of dependence and is therefore our preferred choice.

For the parametrisation of the conditional multivariate model we followed the suggestion of Heffernan and Tawn (2004). They use a combination of two different parametrisations accounting for the different decay of the upper and lower tail of the Gumbel distribution. The estimation procedure first calculates parameters for one type of parametrisation and, given the values of these parameters, either stops or continues to calculate the parameters of the other type of parametrisation; the breakpoint between this two parametrisations is within the range of near extremal independence. Our simulation study shows that for a range of negative correlations the application of this procedure leads to estimates resulting from a local but not a global maximum of the likelihood. We therefore applied an information criteria to avoid this shortcoming.

In Chapter 5 we consider dependence at extreme levels between component data of two weather stations. We are interested in whether extreme wind events along a railway track occur rather localized or exhibit joint dependence. A simplified version of the track consists of two points only; in the present application these are the endpoints of a railway track for high speed trains. Using this simplified version, one necessary component to calculate the probability of at least one station facing an extreme storm event is a conditional distribution comprising the dependence structure; the conditional model introduced in Chapter 4 is thus a natural choice. As we consider all directions we require a model capable of accounting for a broad range of different forms of dependence; the conditional model provides exactly this flexibility. We, however, need a model which allows smooth transition over different types of dependence, so the breakpoint within the parametrisation employed in Chapter 4

is not convenient. Therefore all margins are transformed to follow a double exponential distribution, which is symmetric and has the same upper tail as a Gumbel distribution, thus one parametrisation can be used for all types of dependence.

To enable directional variation of parameters we employ a functional relationship depending on the directions of both stations; it consists of two stages in each of which harmonic terms or adaptions of them are used. This allows a considerable reduction in parameters compared with a sector by sector analysis and the pooling of information to improve parameter estimation. Point estimation was carried out using an objective function derived from common likelihood methods. As the objective function is not a real likelihood, standard methods for model selection cannot be employed. We suggest a selection method based on a bootstrap procedure and consider the variation of bootstrap estimates of the location parameter; the bootstrap procedure is adapted to component data and the nature of extreme values.

Estimation of return-levels is based on simulation, which in turn requires simulation from the distribution of the residuals. Usually, the empirical distribution function is taken as an estimate of the residual distribution but we consider an extension which allows neighbouring information to be incorporated. Return-levels are used to assess the fit of the model, by comparing quantities based on data with the corresponding ones from the estimated model. For the analysis presented there was good agreement between return-levels derived from the model and the corresponding ones based on the data.

Appendix A

Gradient for Chapter 5

To find estimates using objective function (5.8), the optimization procedure requires the gradient function of parameters of the dependence structure. The gradient for parameter ϑ is

$$\frac{\partial l}{\partial \vartheta} = -\sum_{\phi_1}\sum_{\phi_2}[+]$$

with $[+]$ given by

$$[+] = \begin{cases} \frac{\partial \eta_\zeta}{\partial \vartheta}\sigma^{-1}_{\phi_1|\phi_2}\sum_{i=1}^{n_{\phi_1}}(*) & \text{if} \quad \zeta = \mu \\ \frac{\partial \eta_\zeta}{\partial \vartheta}\exp(\eta_\zeta)\sigma^{-1}_{\phi_1|\phi_2}\{n_{\phi_1} - \sum_{i=1}^{n_{\phi_1}}(*)^2\} & \text{if} \quad \zeta = \sigma \\ \frac{\partial \eta_\zeta}{\partial \vartheta}(-2)[\exp(\eta_\zeta)+1]^{-2}\exp(\eta_\zeta)\sigma^{-1}_{\phi_1|\phi_2}\sum_{i=1}^{n_{\phi_1}}(*)y_{\phi_1,i}^{1-b_{\phi_1|\phi_2}} & \text{if} \quad \zeta = a \\ \frac{\partial \eta_\zeta}{\partial \vartheta}\exp(-\eta_\zeta)\sum_{i=1}^{n_{\phi_1}}\log(y_{\phi_1,i})\left\{1 - \sigma^{-1}_{\phi_1|\phi_2}(*)[y_{\phi_1|\phi_2} - a_{\phi_1|\phi_2}y_{\phi_1,i}]y_{\phi_1,i}^{-b_{\phi_1|\phi_2}}\right\} & \text{if} \quad \zeta = b \end{cases}$$

where $(*) = \sigma^{-1}_{\phi_1|\phi_2}\left[y_{\phi_1,i}^{-b_{\phi_1|\phi_2}}(y_{\phi_1|\phi_2} - a_{\phi_1|\phi_2}y_{\phi_1,i}) - \mu_{\phi_1|\phi_2}\right]$ and $\partial\eta_\zeta/\partial\vartheta$ is given by

$$\frac{\partial \eta_\zeta}{\partial \vartheta} = \begin{cases} 1 & \text{if} \quad \vartheta = \lambda_{\psi_{\gamma_\zeta}} \\ \cos(l\phi_1 - \nu_{\psi_{\gamma_\zeta l}}) & \text{if} \quad \vartheta = \kappa_{\psi_{\gamma_\zeta l}} \\ \sin(l\phi_1 - \nu_{\psi_{\gamma_\zeta l}})\kappa_{\psi_{\gamma_\zeta l}} & \text{if} \quad \vartheta = \nu_{\psi_{\gamma_\zeta l}} \\ \\ \cos(j\phi_2 - \psi_{\omega_\zeta j}(\phi_1) - \phi_1) & \text{if} \quad \vartheta = \lambda_{\psi_{\beta_\zeta j}} \\ \cos(j\phi_2 - \psi_{\omega_\zeta j}(\phi_1) - \phi_1)\cos(l\phi_1 - \nu_{\psi_{\beta_\zeta jl}}) & \text{if} \quad \vartheta = \kappa_{\psi_{\beta_\zeta jl}} \\ \cos(j\phi_2 - \psi_{\omega_\zeta j}(\phi_1) - \phi_1)\sin(l\phi_1 - \nu_{\psi_{\beta_\zeta jl}})\kappa_{\psi_{\beta_\zeta jl}} & \text{if} \quad \vartheta = \nu_{\psi_{\beta_\zeta jl}} \\ \\ \psi_{\beta_\zeta j}(\phi_1)\sin(j\phi_2 - \psi_{\omega_\zeta j}(\phi_1) - \phi_1) & \text{if} \quad \vartheta = \lambda_{\psi_{\omega_\zeta j}} \\ \psi_{\beta_\zeta j}(\phi_1)\sin(j\phi_2 - \psi_{\omega_\zeta j}(\phi_1) - \phi_1)\cos(l\phi_1 - \nu_{\psi_{\omega_\zeta jl}}) & \text{if} \quad \vartheta = \kappa_{\psi_{\omega_\zeta jl}} \\ \psi_{\beta_\zeta j}(\phi_1)\sin(j\phi_2 - \psi_{\omega_\zeta j}(\phi_1) - \phi_1)\sin(l\phi_1 - \nu_{\psi_{\omega_\zeta jl}})\kappa_{\psi_{\omega_\zeta jl}} & \text{if} \quad \vartheta = \nu_{\psi_{\omega_\zeta jl}} \\ \\ 0 & \text{otherwise} \end{cases}$$

Bibliography

Barnett, V. (1976). The ordering of multivariate data (with discussion), *J. Royal Statist. Soc. A* **139**: 318–355.

Beirlant, J., Geogebeur, Y., Segers, J. and Teugels, J. (2004). *Statistics of extremes: theory and application*, Wiley, Chichester.

Casson, E. and Coles, S. G. (1998). Extreme hurricane wind speeds: Estimation, extrapolation and spatial smoothing, *J. Wind Eng. Ind. Aerodyn.* **74–76**: 131–140.

Coles, S. G. (2001). *An Introduction to Statistical Modeling of Extreme Values*, Springer–Verlag, London.

Coles, S. G. and Dixon, M. J. (1999). Likelihood-based inference for extreme value models, *Extremes* **2**: 5–23.

Coles, S. G., Heffernan, J. E. and Tawn, J. A. (1999). Dependence measures for extreme value analyses, *Extremes* **2**: 339–365.

Coles, S. G. and Pericchi, L. (2003). Anticipating catastrophes through extreme value modelling, *Appl. Stat.* **25**: 405–416.

Coles, S. G. and Powell, E. A. (1996). Bayesian methods in extreme value modelling: a review and new developments, *Int. Stat. Rev.* **64**: 119–136.

Coles, S. G. and Tawn, J. A. (1991). Modelling extreme multivariate events, *J. Royal Statist. Soc. Series B* **53**: 377–392.

Coles, S. G. and Tawn, J. A. (1994). Statistical methods for multivariate extremes: an application to structural design (with discussion), *Appl. Stat.* **43**: 1–48.

Coles, S. G. and Tawn, J. A. (1996). A Bayesian analysis of extreme rainfall data, *Appl. Stat.* **45**: 463–478.

Coles, S. G. and Walshaw, D. (1994). Directional modelling of extreme wind speeds, *Appl. Stat.* **43**: 139–157.

Cook, N. J., Harris, R. I. and Whiting, R. (2003). Extreme wind speeds in mixed climates revisited, *J. Wind Eng. Ind. Aerodyn.* **91**: 403–422.

Davison, A. C. and Hinkley, D. V. (1997). *Bootstrap methods and their application*, Cambridge University Press, Cambridge.

Davison, A. C. and Smith, R. L. (1990). Models for exceedances over high thresholds (with discussion), *J. Royal Statist. Soc. Series B* **52**: 393–442.

de Haan, L. (1985). Extremes in higher dimensions: the model and some statistics, *In: Proceedings of the 45th Session*, International Statistical Institute.

de Haan, L. and Ferreira, A. (2006). *Extreme Value Theory*, Springer–Verlag, New York.

Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997). *Modelling Extremal Events for Insurance and Finance*, Springer, Berlin.

Fisher, N. I. (1993). *Statistical analysis of circular data*, Cambridge University Press, Cambridge.

Fisher, R. A. and Tippet, L. H. C. (1928). Limiting forms of the frequency distribution of the largest or smallest member of a sample, *Proc. Cambridge Philos. Soc.* **24**: 180–190.

Gnedenko, B. V. (1943). Sur la distribution limité du terme d'une série aléatoire, *Ann. Math.* **44**: 423–453.

Heffernan, J. E. and Tawn, J. A. (2004). A conditional approach to multivariate extreme values (with discussion), *J. Royal Statist. Soc. Series B* **66**: 1–34.

Hosking, J. R. M. (1985). Maximum–likelihood estimtation of the parameter of the generalised extreme-value distribution, *Appl. Stat.* **44**: 301–310.

Jenkinson, A. F. (1955). The frequency distribution of the annual maximum (or minimum) of meteorological elements, *Quart. J.R. Met. Soc* **81**: 158 – 171.

Joe, H. (1997). *Multivariate models and dependence concepts*, Chapman & Hall, London.

Kasperski, M. (2002). A new wind zone map of Germany, *J. Wind Eng. Ind. Aerodyn.* **90**: 1271–1287.

Küchenhoff, H. and Thamerus, M. (1996). Extreme value analysis of munich air pollution data, *Environmental and Ecological Statistics* **3**: 127–141.

Leadbetter, M. R., Lindgren, G. and Rootzén, H. (1983). *Extremes and Related Properties of Random Sequences and Processes*, Springer, New York.

Leford, A. W. and Tawn, J. A. (1996). Statistics for near independence in multivariate extreme values, *Biometrika* **83**: 169–187.

Leford, A. W. and Tawn, J. A. (1997). Modelling dependence within joint tail regions, *J. Royal Statist. Soc. Series B* **59**: 475–499.

Mardia, K. V. and Jupp, P. E. (2000). *Directional Statistics*, Wiley, Chichester.

Moriarty, W. W. and Templeton, J. I. (1983). On the estimation of extreme wind gusts by direction sector, *J. Wind Eng. Ind. Aerodyn.* **13**: 127–138.

Nelsen, R. B. (1999). *An introduction to copulas*, Springer, New York.

O'Brien, G. L. (1987). Extreme values for stationary and Markov sequences, *Ann. Probability* **15**: 281–291.

Pandey, M. D. (2002). An adaptive exponential model for extreme wind speed estimation, *J. Wind Eng. Ind. Aerodyn.* **90**: 839–866.

Pandey, M. D., Van Gelder, P. H. A. J. M. and Vrijling, J. K. (2001). The estimation of extreme quantiles of wind velocity using L-moments in the peaks-over-threhold approach, *Struct. Safety* **23**: 179–192.

Payer, T. and Küchenhoff, H. (2004). Modelling extreme wind speeds at a German weather station as basic input for a subsequent risk analysis for high-speed trains, *J. Wind Eng. Ind. Aerodyn.* **92**: 241–261.

Pickands, J. (1975). Statistical inference using extreme order statistics, *Ann. Statistics* **3**: 119–131.

Prescott, P. and Walden, A. T. (1980). Maximum-likelihood estimation of the parameters of the three-parameter extreme-value distribution, *Biometrika* **67**: 723–724.

Resnick, S. I. (1987). *Extreme values, regular variation and point processes*, Springer–Verlag, New York.

Robinson, M. E. and Tawn, J. A. (1995). Statistics for exceptional athletics records, *Appl. Stat.* **44**: 499–511.

Robinson, M. E. and Tawn, J. A. (1997). Statistics for extreme sea currents, *Appl. Stat.* **46**: 183–205.

Smith, R. L. (1985). Maximum likelihood estimation in a class of non-regular cases, *Biometrika* **72**: 67–90.

Smith, R. L. (1986). Extreme value theory based on the r largest annual events, *J. Hydrology* **86**: 27–43.

Smith, R. L. (1990). Extreme value theory, *In: W. Ledermann, Handbook of applicable mathematics, Supplement, pp. 437–472*, Wiley, Chichester.

Stephenson, A. and Tawn, J. A. (2004). Baysian inference for extremes: Accounting for the three extremal types, *Extremes* **7**: 291–307.

Tawn, J. A. (1988). An extreme value theory model for dependent observations, *J. Hydrology* **101**: 227–250.

Van Gelder, P. H. A. J. M. (1996). A Bayesian analysis of extreme water levels along the Dutch coast using flood historical data, *Stochastic Hydraulics* **7**: 243–251.

von Mises, R. (1954). La distribution de la plus grande n valeurs, *In: Selected Papers, Volume II, 271 – 294*, American Mathematical Society, Providence, Rhode Island.

Walshaw, D. (1991). *Statistical analysis of extreme wind speeds*, PhD thesis, University of Sheffield, Sheffield.

Walshaw, D. (2000). Modelling extreme wind speeds in regions prone to hurricanes, *Appl. Stat.* **49**: 51–62.

Walshaw, D. and Anderson, C. W. (2000). A model for extreme wind gusts, *Appl. Stat.* **49**: 499–508.