Payer, Küchenhoff:

Modelling extreme wind speeds in the context of risk analysis for high speed trains

Sonderforschungsbereich 386, Paper 295 (2002)

Online unter: http://epub.ub.uni-muenchen.de/

Modelling extreme wind speeds in the context of risk analysis for high speed trains

T. Payer and H. Küchenhoff
Department of Statistics, Ludwig–Maximilians–Universität München,
D–80 799 München, Germany

November 21, 2002

Abstract

For high speed trains there is a potential risk of derailment when driving very fast and being hit by an extraordinary strong gust at the same time. The risk depends on both the wind speed and the angle between train and gust. Several techniques have been established to minimize this risk to acceptable values. To decide which of these techniques at a given site is most appropriate, precise knowledge of the wind process at extreme levels is necessary. Therefore methods adapted to the special requirements of the application are needed. We discuss directional modelling using an approach proposed by Coles and Walshaw [2]. We focus on estimating extreme quantiles and their confidence intervals. Different types of confidence intervals are compared and we show how these calculations can be used for risk analysis.

Keywords: Directional data; Extreme values; Risk assessment; Precision of estimates

1 Introduction

To shorten the travelling time, and therefore improve the attractiveness of using railway as choice of transport, modern high speed trains are built to go faster, have increased acceleration, and run most energy efficient. One strategy to achieve this goal is to use less heavy materials. So due to higher speed and lower weight, the stability of the train is reduced and the effect of extreme winds should be investigated closer. The problem we face therefore 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. For the simultaneous analysis of wind speed and its direction we apply a model proposed by Coles and Walshaw [2]. It uses the r 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. We apply two methods. The first simply employs the raw data as recorded. The second uses a special procedure processing data before model estimation; data of one direction contribute to neighboring directions by their value being multiplied by the cosine of the distance in angle. So an order statistic of any one direction might be each an actual observation or a reduced value of a neighboring one. It might therefore be regarded as the power of the wind which is being modelled rather than the wind speed itself. To distinguish the two approaches we refer to the second as component model and corresponding data as component data. We discuss both methods and their different interpretation and usefulness in the context of risk analysis.

After having estimated the model, there are two possibilities of risk assessment we will 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 want to know which wind speeds we must expect to face in order, for example, to think about measures like wind protection, while the second is favorable 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 will discuss both methods.

The whole risk assessment is based on the assumption that the applied model using harmonic terms is an appropriate choice. 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. Conversely, this simulation study at the same time serves as an analyze on different methods of constructing confidence intervals of return-levels.

The paper is organized as follows. In the subsequent section we give a short summary of the theory, which is followed by simulation results on different strategies for calculating confidence intervals. In Section 4 we study data from a German gauging station. We conclude with some discussion in Section 5.

2 Theoretical Background

2.1 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 $M_n = \max\{Z_1, \ldots, Z_n\}$ of independent and identically distributed random variables Z_i. For large n the distribution of the maximum M_n is usually approximated by the Generalized Extreme Value family, which, as $n \to \infty$, constitutes the entire class of non-degenerate limiting distributions of normalized maxima. The distribution function of the GEV is given by

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

whenever $\{x : 1 + \xi(x - \mu)/\sigma > 0\}$ with $\mu \in I\!R$ and $\sigma > 0$ being location and scale parameters, respectively. The shape parameter ξ determines whether or not the distribution has an upper bound. The former is true whenever $\xi < 0$, which corresponds to a Weibull distribution, while there is no upper limit for $\xi > 0$, which is of Frechet type and $\xi = 0$ being interpreted as $\xi \to 0$ yielding a Gumbel distribution. We analyze annual maxima of wind speed in direction ϕ, which may be regarded as the maximum of 365 daily maxima. As the asymptotic theory is still valid under mild dependence conditions (see for example [4]), the deviation of the independence assumption is not essential. We therefore assume the GEV to be an appropriate model for annual maxima; for a comprehensive account see [3].

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 r largest order statistics. This distribution is characterized by the same parameters as the GEV. Denote the order statistics for a given direction $\phi \in \Phi \subset (0, 2\pi]$ by $x_{\phi j}^{(1)} \geq x_{\phi j}^{(2)} \geq \cdots \geq x_{\phi j}^{(n)}$. Then the joint density of $\vec{x}_{\phi j} = (x_{\phi j}^{(1)}, \ldots, x_{\phi j}^{(r)})$ for $\{x_{\phi j} : 1 + \xi_\phi(x_{\phi j} - \mu_\phi)/\sigma_\phi > 0\}$ is

$$h_\phi^{(r)}(\vec{x}_{\phi j}) = \sigma_\phi^{-r} \exp\left\{ - \left[1 + \xi_\phi \left(\frac{x_{\phi j}^{(r)} - \mu_\phi}{\sigma_\phi} \right) \right]^{-1/\xi_\phi} - \right.$$
$$\left. - \left(1 + \frac{1}{\xi_\phi} \right) \sum_{k=1}^{r} \log \left[1 + \xi_\phi \left(\frac{x_{\phi j}^{(k)} - \mu_\phi}{\sigma_\phi} \right) \right] \right\}. \qquad (2)$$

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. The functional relationship is given by harmonic terms having the form

$$\theta_c(\phi) = a_c + \sum_{t=1}^{n_c} b_{ct} \cos(t\phi - w_{ct}) \qquad (3)$$

with θ_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. 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 $3 + 2\sum_{j=1}^{3} n_c$ parameters. Let N be the number of intervals, say years, and r denote the number of order statistics for a subset $\Phi \subset (0, 2\pi)$, then the logarithm of the likelihood is

$$l = \sum_{\phi \in \Phi} \sum_{j=1}^{N} \log h_\phi^{(r)}, \qquad (4)$$

with $h_\phi^{(r)}$ being the density given in (2). After substituting the parameters of the density by harmonic terms as given in (3) usual maximization procedures will supply parameter estimates of a_c, b_{ct}, and w_{ct}. Related standard errors are calculated from the observed Hessian $\mathbf{H}_O = -\nabla^2 l(\vec{\vartheta})$ evaluated at $\vec{\vartheta} = \hat{\vec{\vartheta}}$, where $\vec{\vartheta}$ denotes the vector of all parameters a_c, b_{ct}, and w_{ct}.

The alternative approach is by using component data, which implies a processing of data before analyzing them. For each direction α the data consist of all values $Y_\alpha = Y_\phi \cos(\alpha - \phi)$ whenever $|\alpha - \phi|(\text{modulo } \pi) < \pi/2$ holds; Y_ϕ represents a gust in direction ϕ. From these values the r largest ones of any direction contribute to the likelihood in the usual way. The dependence induced by the processing procedure does not alter the maximum likelihood estimate itself; it needs, however, to be accounted for when calculating the standard errors of estimated parameters. Let $l(\vec{\vartheta})$ denote the logarithm of the likelihood as given in (4) stressing the dependence on parameters. Then, by applying an approximation using Taylor series expansion, the covariance matrix of $\hat{\vec{\vartheta}}$ becomes

$$\mathrm{cov}(\hat{\vec{\vartheta}}) \approx \mathbf{H}^{-1}\mathbf{V}\mathbf{H}^{-1}, \tag{5}$$

where $\mathbf{H} = -E(\nabla^2 l(\vec{\vartheta}))$ and $\mathbf{V} = \mathrm{cov}(\nabla l(\vec{\vartheta}))$; ∇ 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 $h_{\phi j}^{(r)}$ denote the density of the r largest order statistics in year j; with the annual contributions $\vec{u}_j(\vec{\vartheta}) = \nabla \sum_{\phi\in\Phi} \ln h_{\phi j}^{(r)}(\vec{x})$ being independent and identically distributed random variables, the score vector can be restated as $\nabla l(\vec{\vartheta}) = \sum_{j=1}^{n} \vec{u}_j(\vec{\vartheta})$ and therefore its corresponding covariance matrix is given by

$$\mathbf{V} = \mathrm{cov}(\nabla l(\vec{\vartheta})) = n\mathbf{V}_{\vec{u}_j}$$

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

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

Substitution of $\mathbf{V}_{\vec{u}_j}$ by $\hat{\mathbf{V}}_{\vec{u}_j}$ 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 (5), the desired covariance matrix.

2.2 Risk assessment via quantiles

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

$$x_{\phi}^{(J)} = \mu_{\phi} - \frac{\sigma_{\phi}}{\xi_{\phi}}\{1 - [-\log(1 - 1/J)]^{-\xi_{\phi}}\} \tag{6}$$

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 [1]. 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,

$$V_{\eta_{\phi}^{J}} \approx \vec{d}'\mathbf{V}_{\vec{\vartheta}}\vec{d}. \tag{7}$$

In (7) we have $\vec{d} = \nabla\eta_{\phi}^{(J)}(\vec{\vartheta})$, with $\eta_{\phi}^{(J)}(\vec{\vartheta})$ denoting the return-level given in (6) stressing dependence on the vector of parameters $\vec{\vartheta}$.

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 $x_\phi^{(J)}$ and all remaining parameters. Using (6) and (3) this is

$$a_1 = x_\phi^{(J)} + \frac{\theta_{2\phi}}{\theta_{0\phi}} \left\{1 - [-\log(1 - 1/J)]^{-\theta_{0\phi}}\right\} - (\theta_{1\phi} - a_1) \tag{8}$$

where $\theta_{i\phi}$, $i = 0, 1, 2$, are the parameters according to (3) at the point ϕ. Maximization of the likelihood (4) after substitution of a_1, and maximizing over a reasonable range of return-level-candidates $x_\phi^{(J)}$ for every $\phi \in \{10, \ldots, 360\}$ yields (after comparison with the required quantiles of the χ^2-distribution) the desired confidence bands.

2.3 Risk assessment via probability of exceedance

While the preceding paragraph focuses on calculating extreme quantiles, risk assessment here results from determining 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) we will get the probability of v_{crit} being exceeded in any one year by $P(V > v_{crit}) = 1 - G_{(\xi,\mu,\sigma)}(v_{crit}) =: \nu(\vec{\vartheta})$. In practice, however, the parameters will be replaced by their estimates, which are subject to sampling error, and so, in turn, is the estimated probability of the risk. To get an impression of the precision of the estimated risk probability, confidence bounds or bands are desirable.

One way to calculate confidence intervals is via the delta method. Setting $\mathbf{d} = \nabla\nu(\vec{\vartheta})$ and replacing this in (7) yields the variance of the exceedance probability. Because of the approximate normality, again, it is straightforward to calculate confidence bounds. It is worth mentioning that this way of determining confidence intervals may result in a negative lower interval bound; it is clear that it needs to be set equal zero. When using component data the covariance matrix $\mathbf{V}_{\vec{\vartheta}}$ is replaced by (5).

It is also possible to apply the profile likelihood method to gain confidence intervals for the risk probability. Set $p = 1/J$ and replace $x_\phi^{(J)}$ by v_{crit} equation (8) can be stated as

$$a_1 = v_{crit} + \frac{\theta_{2\phi}}{\theta_{0\phi}} \left\{1 - [-\log(1 - p)]^{-\theta_{0\phi}}\right\} - (\theta_{1\phi} - a_1). \tag{9}$$

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

3 Simulation Study

To investigate the validity and performance of the model a simulation study was carried out assuming true parameters according to the estimated values described in Section 4, see Table 3. As the data there are discretized to ten degree, we will have $\phi \in \{10, \ldots, 360\}$. The simulation procedure then works as follows. For any direction the parameter values of ξ_ϕ, μ_ϕ, and σ_ϕ for direction ϕ are re-calculated from the given parameters a_c, b_{ct}, and w_{ct} using (3). As simulation of the r largest order statistics directly from (2) is complicated, we sample 365 daily maxima and take the five largest values. Employing the max-stability property of the GEV, the distribution F for the daily maxima is calculated by

$$F(x) = G^{1/n}(x).$$

Then by definition G is the distribution of the annual maxima. Note that F is again of extreme value type with a change in the parameters μ_ϕ and σ_ϕ while ξ_ϕ remains the same. Then for each direction $\phi \in \{10, \ldots, 360\}$ n values are simulated from the distribution F. The r largest values of each direction ϕ 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 0° 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 .

The simulation study uses the delta method yielding symmetric confidence intervals. In the following we will use a model having a constant for the parameter ξ_ϕ, one and four harmonic terms to describe variation in σ_ϕ and μ_ϕ, respectively; this model will be 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) are estimated. Due to high computational costs the simulation size being 200 is rather small. However, we can recognize basic features and get a rough impression of the model's performance. The results are shown in Table 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 neighboring directions to have more values outside the required interval as others. However, as repeated simulations with different seeds show the complete opposite phenomenon, we assume it to be random. A striking fact, in contrast, is that most points outside the interval are larger than the interval and only a very little amount being smaller than the required bounds. This might be justified by considering the general shape of confidence intervals based on profile likelihood, which demonstrate 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 produced by the delta method the number of points lying outside the interval must be higher for larger values.

4 Analysis of German wind data

The model described above is now applied to data of the DWD's (Deutscher Wetter Dienst) gauging station of Würzburg, a town in the southern part of Germany, which constitutes one end of the railway track under consideration. The data consist of 22 years (1976-1997) of daily maximum wind speeds and the time of day they occurred. The wind-direction of the maximum itself is not available, but the average of wind-direction within each hour being recorded with an accuracy of 10 degree. Analyses of data from shorter time intervals for one year have shown that the average of the wind direction constitutes a reliable measurement for the exact direction corresponding to the maximum. Having 144 missing values the data to analyse consist of 7892 obersvations. To get an impression of the data, Figure (1) shows boxplots of the wind speeds of daily maxima for all directions as well as a histogram reflecting corresponding frequencies of directions. The angle $\phi = 0 = (360)$ corresponds to the direction north and angles are recorded clockwise. There is a clear pattern supporting the choice of a model for wind speeds which varies smoothly over directions.

One common problem with wind data when considering direction is the masking problem, which is described in detail at [5]. This problem is easiest understood by an example: there is a very strong gust from, say east, and at the same day a slightly stronger one from west. If maxima are recorded daily, the data contains the one from west, but that one of east, which might rank among the greatest ones of this direction, got lost. In this case we say that the gust from east was masked by the one from west and this may apparently cause biased estimates. An immediate consequences 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 [5] 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 consequents of which is a larger estimate of the scale parameter. Considering equation (6), a larger scale parameter, in turn, results in a larger return-level. In our case the situation is not clear. Analysis of 10-minute data over an investigation period of one year, however, have shown, that within directions containing large wind speeds the amount of masking is little.

A convenient feature of the r 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 analyzing 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 analyzed at the same time contributing to the same likelihood function. In the present case, the number of order statistics varies with different directions, but not over years. We restrict the number of largest observations r to be at most five, so each direction within each year contributes by $r \leq 5$ values. Table 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 2 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 of directions indicated by NA are excluded. Note that the masking effect could be high in that directions and that the most extreme values of all directions do not occur in that direction. For directions being essential for the risk analysis the data basis we use is sufficient. The model is then estimated for 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 non 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 3.

As we are to judge the risk of 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.

Graph (2) 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 to be found in Figure 3. The alternative approach is to process the data receiving components as described in Section 2.1. In this case we have $r = 5$ for any year and direction. The estimation results are given in Table 3.

For comparison of the two methods we have calculated the 100 year return-levels and corresponding confidence bands, see Figure 4. 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 it's calculation see [2]. 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.

The alternative way to quantify the risk when knowing a critical value is to calculate its probability of being exceeded. In section 2 the two alternative approaches on how to calculate confidence intervals considered in this paper are described. The delta method is now applied to the unprocessed data shown in Figure 5 and to component data in Figure 6 for a critical wind speed of 38 m/s. Using just the non-processed data for one direction (260°) but two different critical values profile likelihood intervals are visualized in Figures 7 and 8. As with return-levels the confidence interval of the exceedance risk is getting the more asymmetric the more extreme the values considered are and the bigger is the departure from symmetric approximations based on the delta method. Unfortunately, this departure of the delta method is anti-conservative leaving its applicant possibly expecting himself in a safer position than he actually is.

5 Discussion

To model extreme wind behavior 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; for a recent publication see [6]. We have discussed a directional model for extreme value data and two risk measures derived therefrom. The first risk measure is the classical approach using quantiles, while the second is based on the exeedance probability. As the model parameters are subject to sampling variation, so are the risk measures. 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.

To estimate the model's parameters, two approaches were applied. In both cases, a fixed number of order statistics in each direction were extracted for parameter estimation. The first approach took data from different directions without being processed. 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 do not hold to justify application of the model used. As the extreme value family, however, is very flexible, this doesn't really pose difficulties. The other problem is possible dependencies, both between neighboring directions and successive data. Unlike many other analysis working with hourly data, the present one uses daily maxima - a much longer period - so the problem of dependence is not that critical.

By using components and calculating the variance adjusted for this situation, in contrast, 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: both, components have to be computet as well as the variance needs to be adjusted; additionally a reference distribution for model discrimination via a likelihood ratio test needs to be calculated, which implies a high amount of additional computing cost.

Considering the most extreme observation of the whole investigation period, which is 42.7 m/s having 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 casts serious doubts on the practical applicability of using a model with non-processed data. Though, a comparison of the largest observation would be more appropriate with a return-period of 22 years, a comparison of the 100-year return-level using the component model proves this model to supply far more plausible and reliable results.

The analysis can be used for risk 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; it's probability of being exceeded has then to be estimated. This can be done by our methods. Furthermore we can give confidence intervals for this probability. This local risk estimates serve as a sensible input for a risk measure of the whole track which adds up the risk over all points. So the methods give substantial improvement of the risk measurement compared to the usage of empirical quantiles of the wind speed distribution.

6 Acknowledgements

The authors' research was part of a joint project with the German Rail. T. Payer was supported by a (Dr. Hermann und Ellen) Klaproth-Stiftung grant. We are grateful to P. Deeg (German Rail) and J. Schorlemmer (Deutscher Wetterdienst) for many fruitful discussions. We also want to thank S. Krieger and H. Gnann for some helpful comments.

References

[1] S. Coles, *An Introduction to Statistical Modeling of Extreme Values,* Springer–Verlag, London, 2001.

[2] S. Coles and D. Walshaw, Directional Modelling of Extreme Wind Speeds, *Appl. Statist.* **43** (1994) 139–157

[3] P. Embrechts, C. Klüppelberg and T. Mikosch *Modelling Extremal Events for Insurance and Finance* Springer, Berlin, 1997

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

[5] W.W. Moriarty and J.I. Templeton, On The Estimation of Extreme Wind Gusts by Direction Sector, *Journal of Wind Engineering and Industrial Aerodynamics,* **13** (1983) 127–138

[6] M.D. Pandey, An adaptive exponential model for extreme wind speed estimation, *Journal of Wind Engineering and Industrial Aerodynamics,* **90** (2002) 839–866

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

ϕ	190°	200°	210°	220°	230°	240°	250°	260°	270°	280°	290°	300°	310°	320°	330°	340°	350°	360°
10y	3	1	0	0	0	1	2	3	3	2	2	4	2	3	5	7	6	6
	7	8	8	9	8	9	14	9	9	7	5	5	7	6	4	5	5	6
50y	1	1	1	0	0	0	0	1	2	2	3	4	2	3	3	4	4	3
	7	7	11	10	10	10	12	9	9	9	7	6	7	7	8	9	9	7
100y	1	2	1	0	0	0	0	1	2	2	3	3	2	3	3	3	3	2
	5	9	11	11	10	11	11	10	9	9	7	7	6	7	9	10	10	9
1000y	1	0	1	0	0	0	0	1	1	1	2	2	1	1	2	3	3	2
	7	7	9	10	9	11	9	9	10	8	7	8	7	7	10	9	10	10

Table 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

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: 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

		raw data (0,4,1)	components (1,2,1)
	$\hat{a}_0$	-0.197 (0.011)	-0.106 (0.023)
	$\hat{b}_{01}$	NA	0.061 (0.022)
$\hat{\xi}_\phi$	$\hat{w}_{01}$	NA	1.014 (0.517)
	$\hat{a}_1$	14.451 (0.150)	20.061 (0.451)
	$\hat{b}_{11}$	4.843(0.157)	6.668 (0.380)
$\hat{\mu}_\phi$	$\hat{w}_{11}$	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
	$\hat{a}_2$	3.733 (0.06)	2.705 (0.186)
	$\hat{b}_{21}$	0.890 (0.074)	0.793 (0.137)
$\hat{\sigma}_\phi$	$\hat{w}_{21}$	4.536 (0.116)	4.603 (0.141)

Table 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.

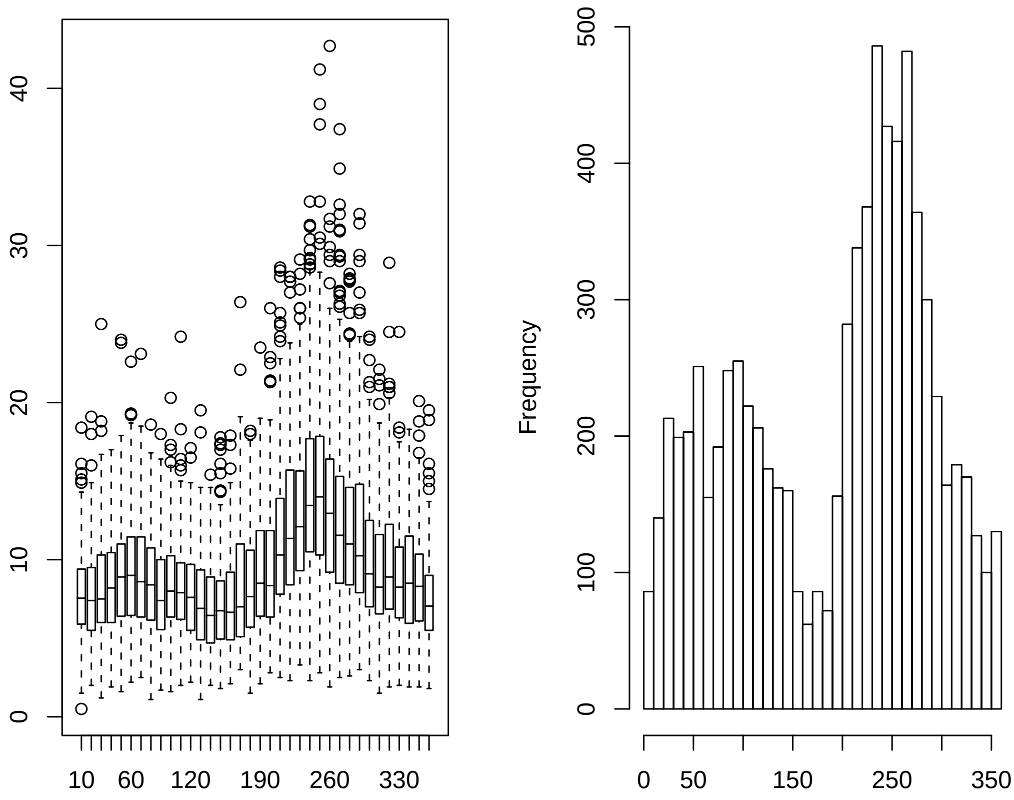

Figure 1: Boxplots of the wind speeds of all observations within the 22 years for different directions; histogram of directions

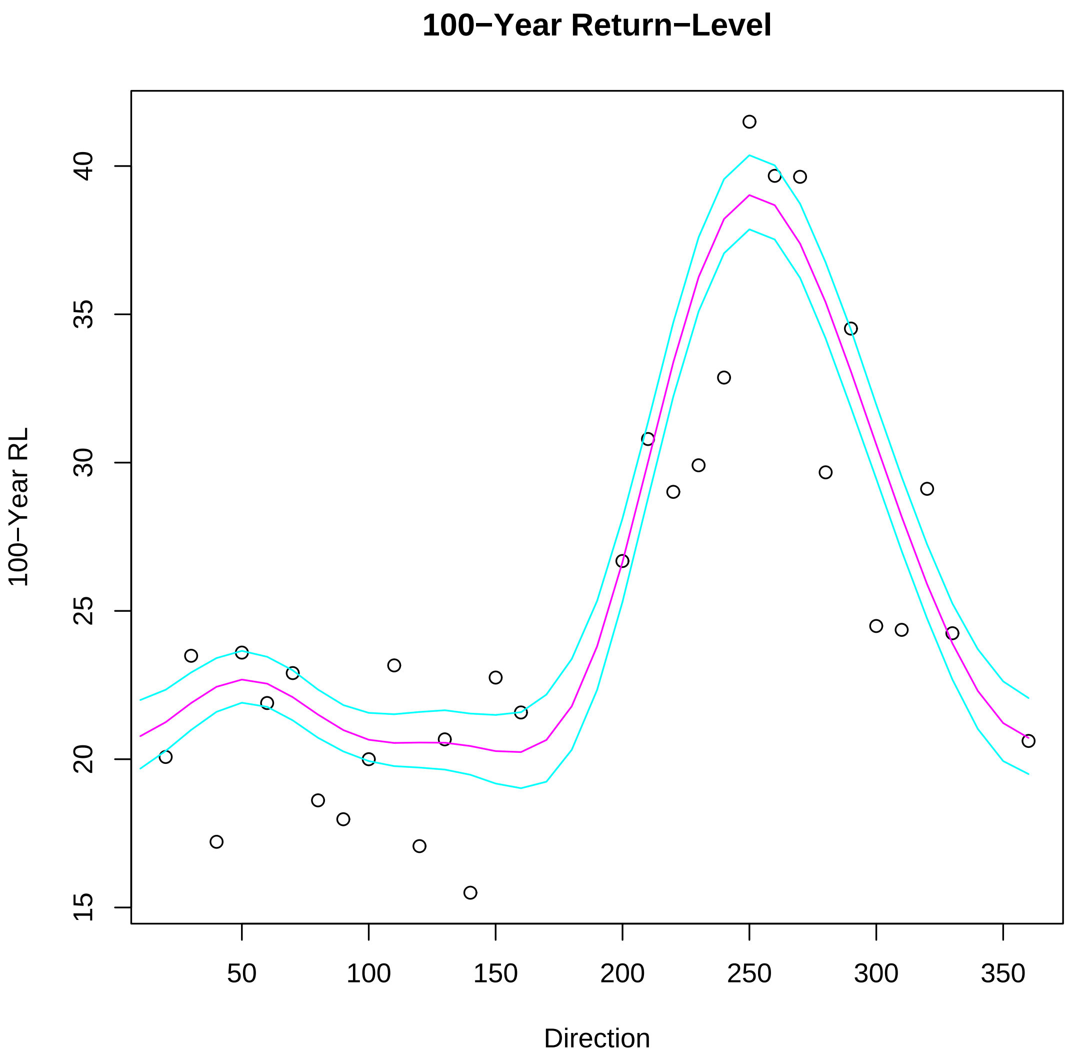

Figure 2: Plot of the (0,4,1)-model of Würzburg: maximum likelihood estimate and 95%-profile likelihood confidence bands; points are estimated return-levels based on data of that direction only.

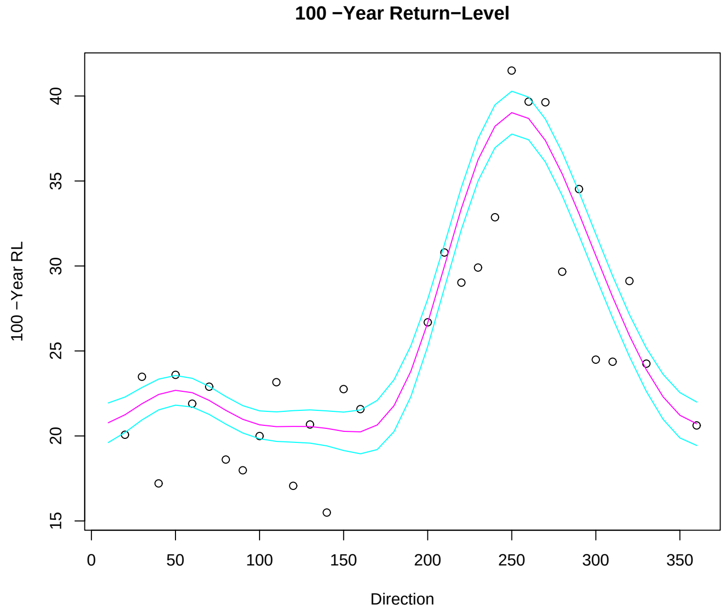

Figure 3: Plot of the (0,4,1)-model of Würzburg: maximum likelihood estimate and 95%- confidence bands by the delta method; points are estimated return-levels based on data of that direction only.

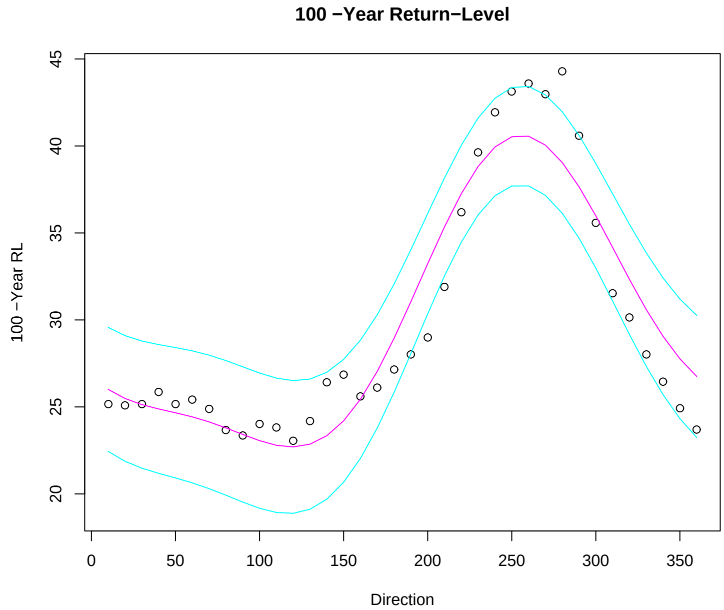

Figure 4: Plot of the (1,2,1) harmonic model of Würzburg: maximum likelihood estimate of the COMPONENT data and 95%- confidence bands by the delta method; points are estimated return-levels based on data of that direction only.

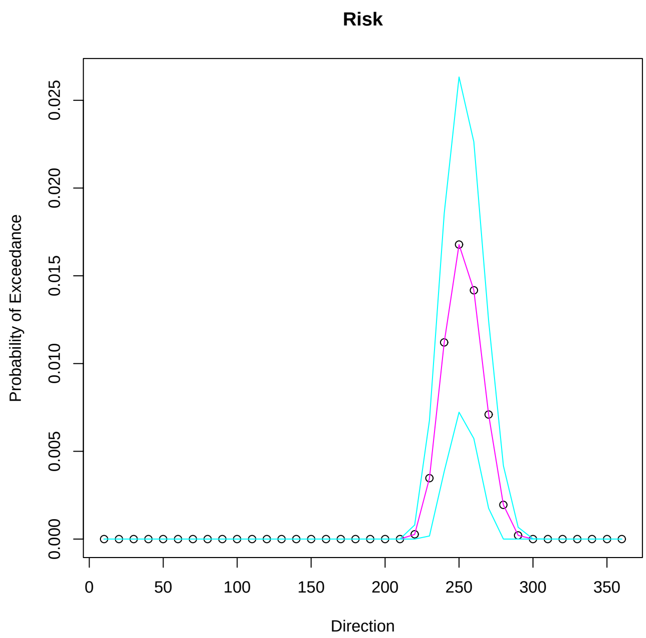

Figure 5: Plot of the exceedance probability of the critical wind speed 38m/s across directions using the (0,4,1)-model based on unprocessed data

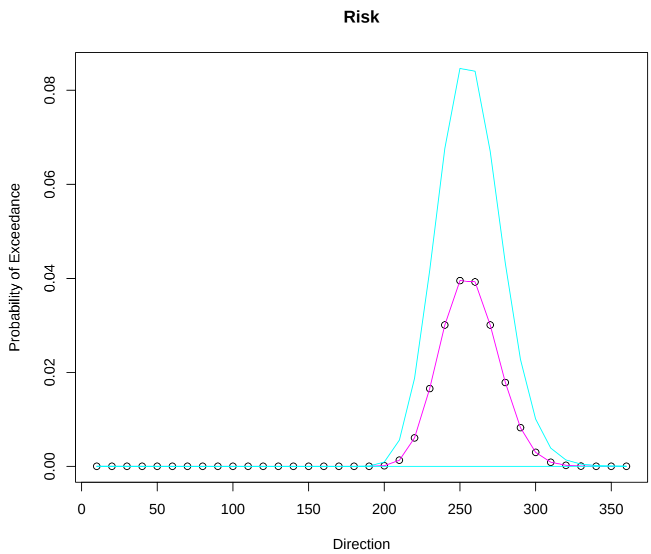

Figure 6: Plot of the exceedance probability of the critical wind speed 38m/s across directions using the (1,2,1)-model based on component data

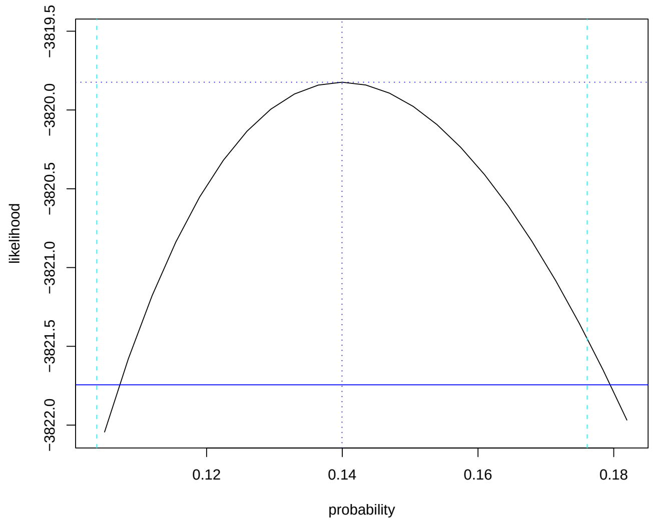

Figure 7: Plot of the exceedance probability of the critical wind speed 32m/s using the (0,4,1)-model; profile likelihood with horizontal line indicating the interval limits and dashed vertical lines indicate corresponding interval based on the delta method

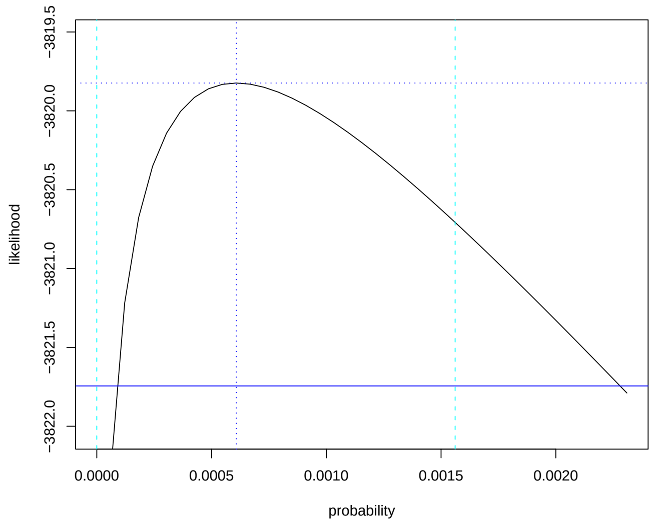

Figure 8: Plot of the exceedance probability of the critical wind speed 42.7 m/s using the (0,4,1)-model; profile likelihood with horizontal line indicating the interval limits and dashed vertical lines indicate corresponding interval based on the delta method