Estimation of extreme wind speeds in Sweden

by
Olivier Perrin†, Holger Rootzén‡, and Roger Taesler§

†Université de Tolouse 1, ‡Chalmers University of Technology,
§Swedish Meteorological and Hydrological Institute

October 23, 2001

Abstract

Wind speeds in extra tropical latitudes are known to be approximately Weibull distributed. Hence a Weibull distribution fitted to all available data is often used to predict sizes of future extreme winds. The most extreme values then, however, have little influence on the estimated parent Weibull distribution, and the accuracy of the extreme value predictions obtained in this manner may be questioned. In the present paper such a " Weibull method" is compared to a method based on statistical extreme value theory, "the Annual Maxima method". The comparison is based on 30 years of 10 minute wind speed averages measured hourly at 12 meteorological stations located at airports in Sweden. Results include that the Weibull method gives incorrect estimates of the tails of the distributions of wind speeds and of the distribution of yearly maximum wind speeds, that it doesn't provide any confidence bounds for the estimates, and that serial dependence of individual measurements has to be taken into account. The Annual Maxima method avoids these problems. A further, "technical" results is that rounding to entire meters per second decreases the accuracy of estimates, and that if such rounding is done, it should be taken into account in the estimation procedure.

[0]

Keywords and phrases: annual maxima method, building codes, extreme value distribution, extreme winds, Weibull method, wind norms, wind standards.

Research supported by the European Union's research network "Statistical and Computational Methods for the Analysis of Spatial Data", grant ERB-FMRX-CT960095.

1 Introduction

Wind speeds in extra tropical latitudes are well known to conform closely to the Weibull distribution. Consequently, methods based on fitting a Weibull distribution to all available data are often applied to predict sizes of future extreme winds. This can be done by using the tail of the Weibull distribution directly – here this will be called "the Weibull method". However, since the most extreme values have very little influence on the parameters of the parent Weibull distribution, there is reason to question the accuracy of the extreme value predictions obtained in this manner.

An alternative that is sometimes used is to assume that extreme winds follow a Gumbel distribution with parameters derived from the parent Weibull distribution (cf. e.g. Davenport 1967, Alexandersson 1979, Bergström 1992). Since the Gumbel distribution is obtained as the limit of the distribution of maxima of Weibull variables, this is expected to give similar results to "the Weibull method". However, still this includes one more, perhaps unnecessary, approximation, and we hence will not study it further in the body of this paper. Still, a brief discussion is given in an Appendix.

In the present paper the "Weibull method" is compared to a method based on statistical extreme value theory, "the Annual Maxima method". The methods are briefly introduced in this section, and are discussed more in detail in Section 3 below. The comparison is aimed at developing methods for construction of Swedish wind standards.

The Weibull method is briefly alluded to in Lundtang Petersen et al. (1981), with substantial reservations, and Weibull estimation is surveyed in Conradsen et al. (1984). There are many accounts of the Annual Maxima method, see e.g. Embrechts et al. (1997, Section 6.3). Teugels (2000) contains an extensive collection of references on the interface between wind speeds and statistics, and Paultikof et al. (1999) make a very useful survey of methods to calculate extreme wind speeds.

In discussions of statistical methods for extremes it is convenient to use the concept of an *extremal index*. The idea, which is due to Leadbetter, is as follows: often extreme values come in small clusters. For instance, a single wind storm often leads to several high measured wind speeds. Roughly, the extremal index, θ, is one divided by the average number of values in such a cluster of large values. Under rather general conditions, the maximum of n dependent observations have approximately the same distribution as the maximum of θn independent observations. For formal discussion and proofs see Embrechts et al. (1997, Section 8.1) and Leadbetter et al. (1983, Section 3.7).

Similar ideas existed also earlier, e.g. Davenport (1967) defines an effective sampling frequency, $n_{eff} = \nu_T$, where T is time in hours and n_eff is the "number of independent samples". Davenport e.g. found $\nu = 0.1$ to be representative for 1 minute wind speed averages.

A main aim is to estimate 50-year or 100-year winds or, in different terminology, the 50- or 100-year return levels. By definition these are the yearly maximum windspeeds which on the average are exceeded once every 50 or 100 years. Equivalently, the 50- and 100-year winds are the $1 - 1/50$ or $1 - 1/100$ quantiles of the distribution of the annual maximum wind speed. (The 50- or 100-year winds can also be thought of as close approximations to the mean value of the distribution of the maximum windspeed during a 50 or 100 year period. Since the distributions of maxima are skewed, the mean values are a bit different from the modal values for these distributions.)

In the Weibull Method a two parameter Weibull distribution is fitted to all available wind speed recordings. The 50- and 100-year winds are then computed from the fitted Weibull distribution by assuming that the recorded wind speeds are serially independent.

Weibull distributions have often been observed to give a good fit to the bulk of wind speed data. However, the Weibull method relies on two further assumptions,

(1) a suitable independence of the recorded wind speeds, and
(2) that the fit of the Weibull distribution is good also in the extreme right tail of the distribution of wind speeds.

For (1) it should be noted that independence is not necessary for correct estimation of the parameters of the Weibull distribution. It is enough that dependence decreases suitably fast with increasing time separation. (Exact conditions for this does not seem to be available. However, it is known that they would be of very general applicability.)

However, independence is also used in the step where 50- or 100-year winds are computed from the Weibull distribution. Again, full independence is not needed for the calculation to be valid, it is enough that the extreme part of the data is approximately independent. In technical terms it is sufficient that the extremal index equals one, so that there is no clustering of extremes.

Thus Assumption (1) of the Weibull Method is not as restrictive as it would appear at first. Nevertheless, if dependence extends into the extreme right tail of the distribution of the wind speed recordings, then the Weibull Method can give wrong results.

As for (2) there is no physical or theoretical reason that it should be satisfied for wind speeds recordings. If it is not then the Weibull Method, again, can give wrong results.

In the Annual Maxima method, an Extreme Value (EV) distribution is fitted to the annual maximum wind speeds, assuming that maxima from different years are serially independent. The argument for using the EV distribution comes from statistical extreme value theory (see e.g. Coles (1999), Embrechts et al. (1997), and Rootzén and Tajvidi (1997) for a discussion of this).

An additional advantage of the Annual Maxima method is that maxima taken over different years are likely to be much more independent than hourly measurements. Thus, this method circumvents the problematic Assumption (1) of by pooling the potentially locally dependent recordings into a single value for each year. Assumption (2) is circumvented, since the statistical fitting only uses extreme values.

There of course is a cost associated with using the Annual Maxima method. This is the very drastic reduction of the number of observations used in the statistical analysis. As a consequence estimates obtained from the Annual Maxima method are much more variable than estimates provided by the Weibull Method. However, if the Weibull Method estimates the wrong quantities, it is not much help if this is done with high precision.

In a second semiparametric method based on extreme value theory, the Peaks over Thresholds (PoT) method, a Generalized Pareto distribution is fitted to the declustered exceedances of a suitably chosen high level or "threshold". Similarly as for the Annual Maxima method this avoids using Assumption (2) since only large values are used in the analysis. Further, Assumption (1) is handled by the declustering procedure.

The cost compared to the Weibull Method is the same as for the Annual Maxima method. As compared to the Annual Maxima method, PoT includes more of the data and hence potentially could give more precise estimates. Instead it involves the choice of two somewhat arbitrary tuning parameters, the threshold and the "length of declustering". We will not use PoT in this paper.

Our comparison of the Weibull and Annual Maxima methods is based on a data set provided by the Swedish Meteorological and Hydrological Institute (SMHI). The data set contains thirty years of hourly windspeed recordings (10 minute averages) for 12 stations in Sweden. We applied both the Weibull and the Annual Maxima method to estimate the distribution of the annual maximum windspeed and compared the results to the observed annual maxima. We also compared the 5-year, 50-year and 100-year winds obtained from the two methods.

The data base is described in Section 2. The Weibull and Annual Maxima methods and the statistical methods are discussed in Section 3. The next section, Section 4, summarises the results of the analysis, and a final discussion and conclusions are given in Section 5.

2 The wind data base

The wind database used in this paper was provided by the SMHI. It consists of 10 minute wind speed averages recorded, with some exceptions, at the start of each hour during 1961 – 1990 at 12 synoptic meteorological stations in Sweden (Table 1), and thus contains roughly $30 \times 365 \times 24 \approx 2.6 \times 10^5$ recordings per station. The stations are all located at military airports except for Arlanda, Bromma, Karlstad and Visby which are civil aviation airports.

WMO regulations for siting of different types of meteorological stations and procedures of observations are given in WMO (1983). Regulations for wind measurements are given in chapter 6. The standard exposure for instruments over land is 10 m.a.g over level, open terrain. Open terrain is defined as an area where the distance between the instrument and any obstruction is at last 10 times the height of the obstruction. In practice, it is not always possible to follow these regulations completely. Therefore, wind speeds in particular should always be regarded as influenced to some degree by local effects. Airports generally present open fetch over the nearest area surrounding the instrument. However, at greater distances (say > 1 km) there may be considerable inhomogeneities also around airports.

Airports have the advantage of using instruments to measure wind. During past decades at most other synoptic stations wind speed has only been estimated by human observers. This has now changed completely and the present meteorological network is almost completely automated with high quality wind measuring instruments. Unfortunately, we will have to wait many years to have a basis for studies of extreme values from these stations.

All stations had missing observations, ranging from 0.06% for Arlanda to 17% for Karlstad. In addition the entire first year was not included in the database for Arlanda.

Major reasons for "missing observations" was that some airports, in particular minor civil aviation airports, only make observations during hours of operation. Others, notably military airports, make observations every hour during hours of operation but only every 3 hours (the regular synoptic hours of observation) during the remaining part of the day.

Further, a number of data points seemed dubious or plainly wrong. The reasons were recordings at not existing time points, forgotten conversions from knots to m/s, and measurements which were very out of line with adjacent measurements, such as one very high wind speed recorded when all the measurements around it were very low. We deleted or corrected these data points manually. Only recordings with wind speeds in excess of 14 m/s were corrected, since errors in lower wind speeds were deemed to have a negligible influence on the results.

All synoptic observations (from civil as well as military stations) are made in knots. The data is then converted to m/s and rounded off to the nearest m/s by the SMHI when the data are stored in the historical database.

Table 1: The recording sites

site	missing	deleted	corrected
Arlanda	145	3	0
Barkåkra	1,999	0	4
Bredåkra	1,996	6	8
Bromma	20,339	1	3
Karlstad	44,916	5	1
Ljungbyhed	2,014	2	1
Norrköping-Sörby	2,153	1	7
Säve	3,400	4	2
Söderhamn	2,093	6	4
Uppsala	2,166	7	5
Västerås-Hässlö	12,782	3	6
Visby	8,090	4	0

All the stations in Table 1 (except perhaps Bredåkra) are situated in landscape with rather gentle topography. Their elevation above sea level is rather similar. Changes in local topography are rarely documented or documented in a way that does not permit assessments of resulting changes in wind speed. Changes in instruments and equipment are documented but it is not possible to determine the effects of such changes on the measured data.

It is likely that over a 30-year period the landscape around the recording instruments have changed enough to affect the measurements. A particular problem arises at near urban sites, where urban-industrial growth causes gradual changes in fetch conditions. Arlanda airport is an example of this sort. This civil aviation airport has grown extremely over the past decades and today presents a much more urbanised landscape with significantly higher surface roughness than during the 1960-ies. Most likely this has caused a general reduction in wind speed. However, to our knowledge, no proper analysis of this has yet been made. It is also likely that such an effect is more pronounced for some wind directions than for others. In particular topographical changes could lead to spurious time trends or correlations. We have not had the possibility to correct for this possibility.

Further, to actually establish wind standards, the raw wind speed would have to be calibrated to a standard height and standard surface roughness. Since this only means a change of scale, it does not affect the present comparison of methods. Hence we have not made such standardisation.

3 The Weibull and the Annual Maxima methods

The analysis was performed separately and in the same way for each of the recording sites. To describe it, denote the (rounded) hourly wind speed recordings at the site by $\{x_i\}$, and let $\{m_i\}$ be the yearly maxima of these (rounded) recordings. Further, let F be the distribution of the yearly maximum wind speed. Then the T-year wind u_T is obtained as a solution of

$$T(1 - F(u_T)) = 1. \tag{1}$$

(For a discussion of slightly different versions of this, see Rasmusen and Rosbjerg (1991).)

In the Weibull method the x_i's at a site are treated as having the same two-parameter Weibull distribution function $W(x) = 1 - \exp\{-(\frac{x}{\sigma})^k\}$ with density function

$$w(x) = \frac{k}{\sigma}\left(\frac{x}{\sigma}\right)^{k-1} \exp\{-(\frac{x}{\sigma})^k\}.$$

The distribution function F needed for the computation of return levels in Equation (1) is then computed as if the observations were independent, that is as $F(x) = W(x)^{365\times 24}$, so that

$$u_T = \sigma(-\ln(1 - (1 - 1/T)^{1/(365\times 24)}))^{1/k}. \tag{2}$$

Finally, the parameters have to be estimated from the data. This is here done by the so called maximum likeness method, for grouped samples, which is to use the values of σ and k which maximize

$$\sum_i n_i \ln(W(i + 1/2) - W(i - 1/2)),$$

for n_i = the number of recordings with wind speed i m/s. (In the standard Weibull method the rounding often is not taken into account. This, however, would change the estimates only marginally.) The estimates of course are the same as would be obtained by ordinary Maximum Likelihood estimation if the hourly observations were assumed to be independent (for studies of estimation for binned data, see Kulldorf (1961), Sen and Singer (1993)). The Weibull method does not provide any confidence intervals for the estimated return values.

In the Annual Maxima method, the yearly maxima, m_i, are assumed to be independent and to have the Extreme Value (EV) distribution function $G(x) = \exp\{-(1+\gamma\frac{x-\mu}{\sigma})_+^{-1/\gamma}\}$, where μ is a location parameter, σ is a scale parameter, γ is a shape parameter and $+$ signifies "positive part". That is, for x-values which make the expression in the inner parenthesis negative, the expression should be replaced by zero. The density of the EV distribution is

$$g(x) = \frac{1}{\sigma}(1+\gamma\frac{x-\mu}{\sigma})_+^{-\frac{1}{\gamma}-1}\exp\{-(1+\gamma\frac{x-\mu}{\sigma})_+^{-1/\gamma}\}.$$

The case $\gamma = 0$ is defined by continuity, that is for $\gamma = 0$ the EV distribution has the Gumbel distribution function $G_0(x) = \exp(-\exp(-\frac{x-\mu}{\sigma}))$ with density

$$g_0(x) = \frac{1}{\sigma}\exp(-\frac{x-\mu}{\sigma})\exp(-\exp(-\frac{x-\mu}{\sigma})).$$

The distribution G_0 occupies a special place in extreme value theory, both for theoretical reasons and because it has been found to fit many data sets well.

Finally, the distribution function F in Equation (1) of course is replaced by G (or G_0) in the Annual Maxima method, and hence

$$u_T = \mu + \frac{\sigma}{\gamma}\{(-\ln(1-\frac{1}{T}))^{-\gamma} - 1\} \tag{3}$$

or, for $\gamma = 0$,

$$u_T = \mu - \sigma\ln(-\ln(1-\frac{1}{T})). \tag{4}$$

The maximum likelihood estimates if the rounding of the data is disregarded are the values of the parameters which maximise the log likelihood function

$$\sum_j \ln(g(m_j)) \tag{5}$$

where the summation is over the number of data points. If the rounding (or *"binning"*) is taken into account the log likelihood function instead becomes

$$\sum_i n_i \ln\{G(m_i+1/2) - G(m_i-1/2)\}, \tag{6}$$

where summation again is over groups (or "bins"), and then the maximum likelihood estimates are obtained by maximising (6).

For likelihood computations without taking rounding into account we used a set of S+ routines provided by Coles (http://www.stats.bris.ac.uk/ masgc/) and for the binned likelihood computations we wrote new S+ routines. Unlike Coles we computed the information matrix analytically and used expected information instead of observed.

We studied two different methods to compute confidence intervals. The first one used the inverse of the expected information together with the delta method. The second one was to compute profile likelihood confidence intervals, see Coles (2001). The calculations with expected information are simpler, but profile likelihood intervals are expected to be superior when the distribution of the estimates are skewed. The two methods yield similar results if the distribution of the estimate is reasonably symmetric.

The calculations described above were repeated with G replaced by the Gumbel distribution function G_0. The hypothesis that $\gamma = 0$ was tested using a likelihood ratio test (Coles (2001)) at the 5% significance level.

The goodness-of-fit of the estimated distributions of the annual maxima were checked by Chi-Square tests and graphically. For the Chi-Square test we used 3 - 5 groups and computed the degrees of freedom for the Weibull method as *the number of groups - 1* and for the Annual maxima method as *the number of groups - the number of estimated parameters - 1*. The reason we did not subtract the number of estimated parameters in the Weibull method was that for this method the parameters were not estimated from the yearly maxima but from the (very many) hourly observations.

To check the fit we used standard empirical distribution function and quantile-quantile plots, and plots of the return levels. The empirical distribution function plot for G consists of a "staircase" through the points $(m_{(i)}, N_i/(n+1) : i = 1, \ldots, K)$ and the quantile-quantile plot shows the points $(m_{(i)}, G^{-1}(N_i/(n+1)) : i = 1, \ldots, K)$ where $m_{(1)} < m_{(2)} < \ldots < m_{(K)}$ are the ordered yearly maxima, (that is the ordered centres of bins which contain maxima), N_i is the number of yearly maxima less than or equal to $m_{(i)}$, K is the number of bins and $n = 30$ (for Arlanda 29) is the number of years of observation. The empirical distribution function plots in addition show the estimated distribution function and the quantile-quantile plots includes an $y = x$ line. Plots for G_0 were made similarly.

If G (or G_0) fits the data, the empirical and estimated distribution functions should be close and the quantile plot should consist of points closed to the $y = x$ line.

Return level plots show the empirical return levels $m_{(i)}$ plotted against $(-1/\ln(N_i/(n+1))$ and the T-year return levels estimated from the model plotted against $-1/\ln(1-1/T)$. (It is customary to use $-1/\ln(1-1/T)$ instead of T. For large values of T this does not make any difference. For small values, the return period can be obtained from $T = 1/(1-\exp\{1/(1/ln(1-1/T))\})$ If the model fits the data, the estimated and empirical return levels should be close together. Further the x-axis is on a log scale, which produces a straight line for the Gumbel distribution.

For further discussion of these plots, see e.g. Coles (2001).

The extremal index was estimated using a number of different thresholds. To find the estimate for a threshold of say 14 m/s, the number of “clusters” was divided by the number of recorded wind speeds strictly larger than 14 m/s. A cluster of exceedances of 14 m/s was defined as follows: the recordings were divided into blocks of size $p \times 24$ (we tried with $p = 3, 6, 9$ and 12); if a block contained at least one record strictly larger than 14 m/s then this block is one cluster.

4 Results

"Technical" results: In the Annual Maxima method the differences between the estimates obtained from the binned likelihood function (6) and the likelihood function (5) were quite small for the location parameter, less (usually much less) than 33% of the estimated standard error for the scale parameter and less than 19% of the estimated standard error for the shape parameter, for all sites. The binned estimated standard errors for the location parameter were very similar to the not binned ones while the binned estimated standard errors for the scale parameter were larger (4% to 62%) than the not binned ones, and even larger for the shape parameter (10% to 53%).

Confidence intervals for return levels based on expected information and on profile likelihood were rather similar. For instance, for the 50-year winds, with one exception, the profile likelihood intervals typically were very slightly shorter, and shifted roughly 0.5m/s towards larger values. The exception was Säve where the expected information confidence interval was $(23.1, 27.0)$ and the profile likelihood interval was $(23.9, 28.3)$.

The Weibull Method: The Weibull distribution often underestimated the tail of the distribution of the 10 minute wind speeds by about 1 m/s, see Figure 1.

Estimation of the extremal index is well known to be sensitive to the choice of level. This was the case also for the present data. However, nevertheless, the extremal index seemed clearly less than one for all sites, see Table 2.

The distribution for yearly maxima obtained by the Weibull Method were very strongly rejected by the Chi square test. The p-value was 0.02 for Norrköping-Sörby, 0.007 for Söderhamn and less than 10^{-5} for the other sites. That the Weibull distribution for yearly maxima did not fit was also very evident from plots of estimated distribution functions (Figures 3, 4).

The *Annual Maxima method:* Plots of autocorrelation's (Figure 2) indicated little serial dependence of the annual maxima. Formal portmanteau tests (Brockwell and Davis (1996)) showed a small deviation from independence for the station Bredåkra, but not for the other stations. Likelihood ratio tests at the 5% level rejected the hypothesis that the shape parameter γ in the EV distribution was zero for three sites, Bredåkra, Säve and Visby (Figure 4). The p-values were 0.04, 0.02, and 0.02, respectively. The estimates of the shape parameters for these three sites were negative, corresponding to somewhat smaller extremes than for $\gamma = 0$. The Chi square goodness of fit tests did not reject the EV distribution assumption for any of the sites, and did not reject the Double Exponential distribution for any of the 9 sites where $\gamma = 0$ was not rejected by the likelihood ratio test.

Comparison of results: The quality of the fit for the two methods and return levels for other periods than 5, 50 and 100 years are illustrated in Figures 3 and 4. The sites shown in the figure were chosen as follows: Barkåkra is a typical site, Arlanda is the site where the Gumbel distribution has the visually poorest fit although it nevertheless was not rejected by the likelihood ratio test, and Bredåkra, Säve and Visby are the three sites were the hypothesis that $\gamma = 0$ was rejected. The 50 year return levels estimated by the Weibull Method and by the Annual Maxima method were the same for 4 sites, 1 or 2 m/s larger with the Weibull Method for 4 sites, 4 m/s larger with the Weibull Method for Säve, 5 m/s larger with the Weibull Method for Ljungbyhed and 2 m/s smaller for Norrköping-Sörby and Söderhamn (Table 2). For the 2 sites Ljungbyhed and Säve, the 50 and 100 year Weibull estimates were outside of the Annual Maxima profile likelihood confidence intervals (Table 3).

5 Discussion and conclusions

The goal of this paper was to compare the Weibull Method to compute return levels with methods based on extreme value theory. The comparison was based on 30 years of hourly wind speed recordings from 12 sites in Sweden.

The main results were that the Weibull Method

- does not provide any confidence bounds for the estimates,

Table 2: Return levels (in m/s) and extremal index. W = Weibull, EV = Annual Maxima with EV distribution and binned likelihood, G = Annual Maxima with Gumbel distribution and binned likelihood, * = the Double Exponential distribution was rejected at the 5 % level, and EI = the extremal index estimated with the characteristic value, i.e. the 37% quantile of the yearly maxima, as level and with block size 9 days.

period (ys)	5			50			100			
	W	EV	G	W	EV	G	W	EV	G	EI
Arlanda	18	16	16	20	20	19	21	21	20	0.39
Barkåkra	24	21	21	28	24	28	29	25	29	0.29
Bredåkra*	18	16	16	20	18	20	21	19	21	0.43
Bromma	17	16	16	19	19	19	20	19	20	0.38
Karlstad	19	17	17	21	20	20	22	21	21	0.48
Ljungbyhed	24	18	18	28	22	23	29	22	24	0.42
Norrköping	21	20	20	24	25	26	25	27	28	0.36
Säve*	25	22	22	29	25	29	30	26	31	0.40
Söderhamn	17	17	17	19	22	21	20	24	22	0.44
Uppsala	19	18	18	22	23	22	22	25	23	0.38
Västerås	19	18	18	22	21	22	23	22	23	0.33
Visby*	22	21	21	25	23	26	26	23	27	0.35

Table 3: Profile likelihood 95% confidence intervals for return levels. The binned likelihood function with the Gumbel distribution for annual maxima was used except where marked by *, where the EV distribution was used.

site	5 year	50 year	100 year
Arlanda	(15.2, 16.9)	(17.5, 20.9)	(18.2, 22.2)
Barkåkra	(20.1, 23.4)	(24.9, 31.4)	(26.2, 33.7)
Bredåkra*	(15.3, 16.9)	(17.3, 20.6)	(17.7, 21.7)
Bromma	(14.8, 16.6)	(17.4, 21.0)	(18.1, 22.2)
Karlstad	(15.7, 17.7)	(18.5, 22.6)	(19.3, 24.0)
Ljungbyhed	(17.4, 19.8)	(20.8, 25.4)	(21.8, 27.1)
Norrköping-Sörby	(18.9, 22.2)	(23.6, 30.1)	(24.9, 32.4)
Säve*	(20.9, 23.3)	(23.9, 28.3)	(24.5, 29.7)
Söderhamn	(15.7, 18.0)	(18.9, 23,4)	(19.8, 25.0)
Uppsala	(17.0, 19.2)	(20.0, 24.5)	(20.9, 26.0)
Västerås-Hässlö	(16.6, 18.9)	(20.0, 24.7)	(21.0, 26.5)
Visby*	(20.3, 21.8)	(22.1, 25.0)	(22.4, 25.9)

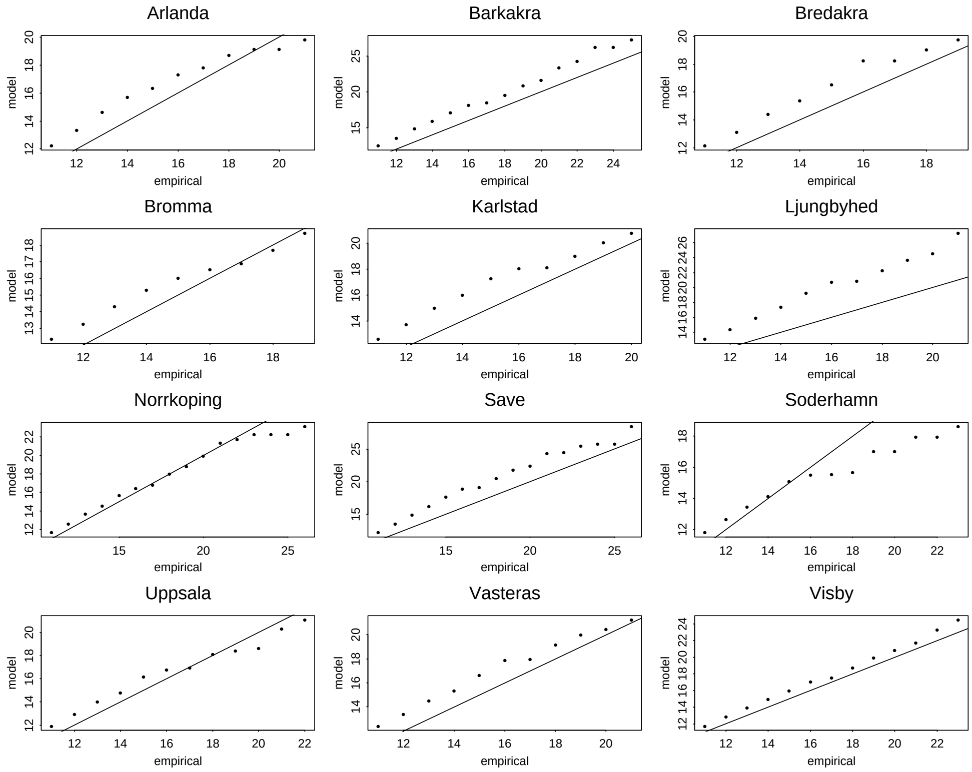

Figure 1: Tail of qq-plot of the empirical quantiles of the hourly 10 minute windspeed measurements against quantiles of fitted Weibull distribution ($\cdots$) and the x=y line (——)

- gives incorrect estimates of the distribution of yearly maximum wind speeds, and
- does not take into account that the serial dependence of hourly wind

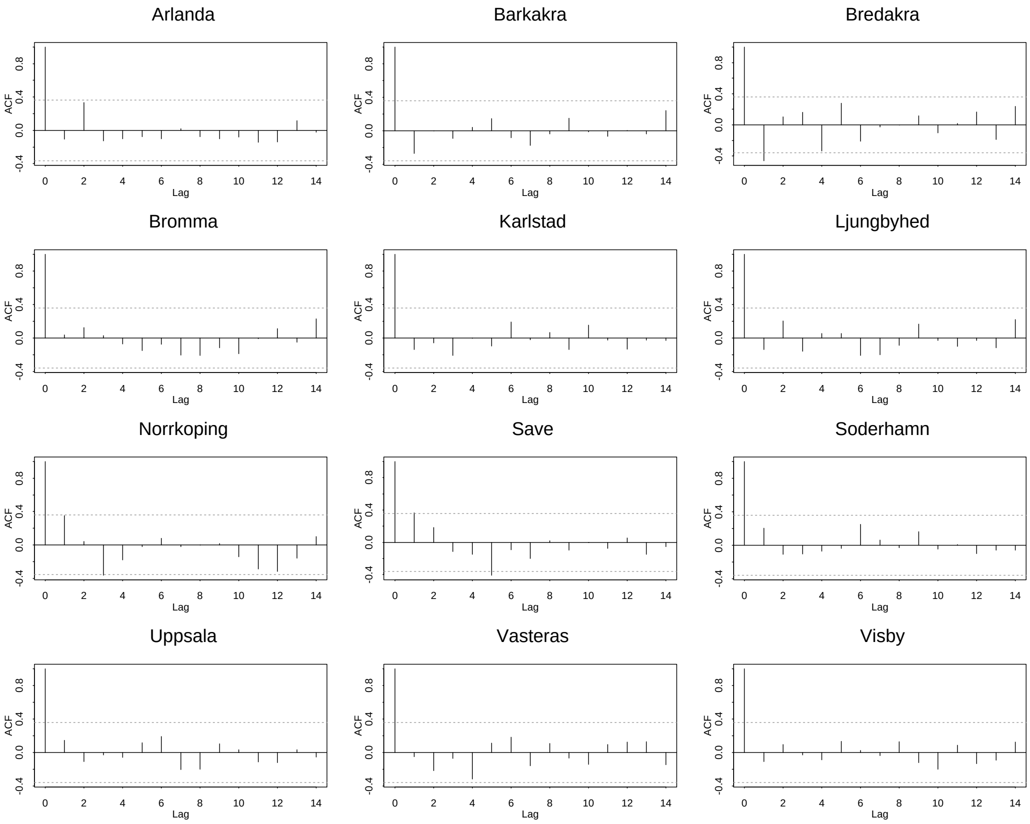

Figure 2: Autocorrelation functions for annual maximum wind speeds

measurements extends into the tail of the distribution of extreme wind speeds, and uses an incorrect form of the marginal distribution of wind speeds.

Here the second item is based on Chi-square goodness of fit tests and on graphical analysis (Section 3) and the third on estimates of the extremal index of the wind measurements (Table 2) and on comparison of fitted and empirical tail functions (Figure 1).

The Annual Maxima method with EV distributions gives credible confidence intervals, a good fit to the tail of the distribution of extreme wind

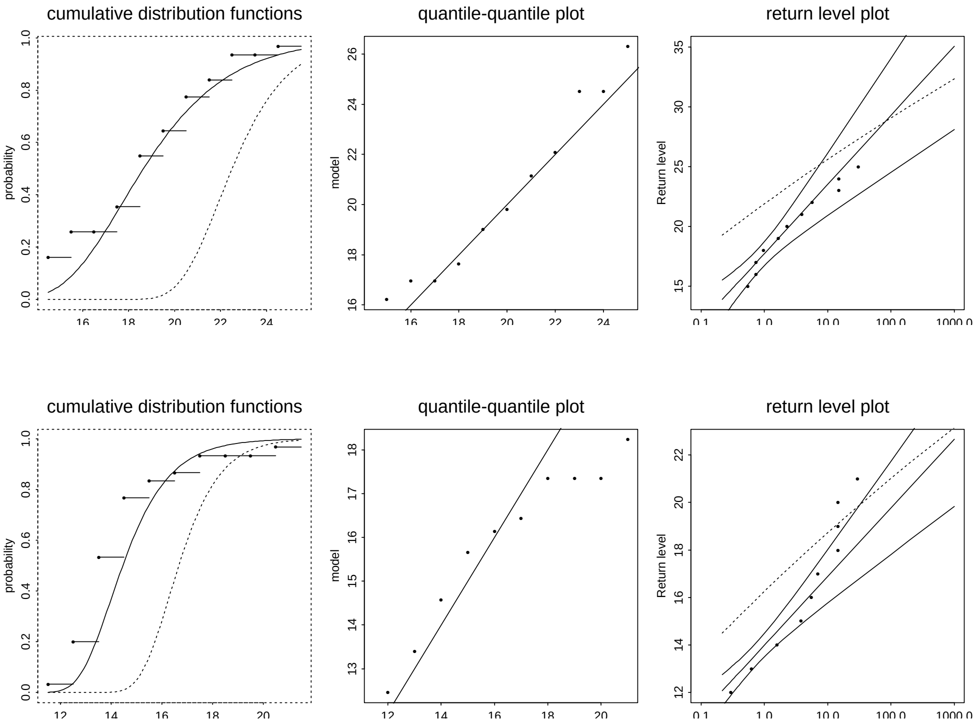

Figure 3: Top Barkåkra (a typical station), bottom Arlanda (worst fit of stations with $\gamma = 0$ not rejected). Solid lines: estimated values and 95% confidence bounds for Annual Maxima estimates. Dotted lines: Weibull estimates

speeds and seems to avoid the problem of serial dependence connected with hourly measurements (Section 3). The different versions of the Annual Maxima method gave rather similar results, and in particular likelihood ratio tests of the assumption that the shape parameter of the EV distribution was zero, or equivalently that the annual maxima followed a Gumbel distribution, was only rejected for three sites, and then only with a small margin. Estimates and confidence intervals obtained assuming an EV distribution and assuming a Gumbel distribution were quite similar. We had expected that the point estimates would not change, but had expected a clear shortening of the confidence intervals when the Gumbel distribution was used, and hence the latter finding was a surprise.

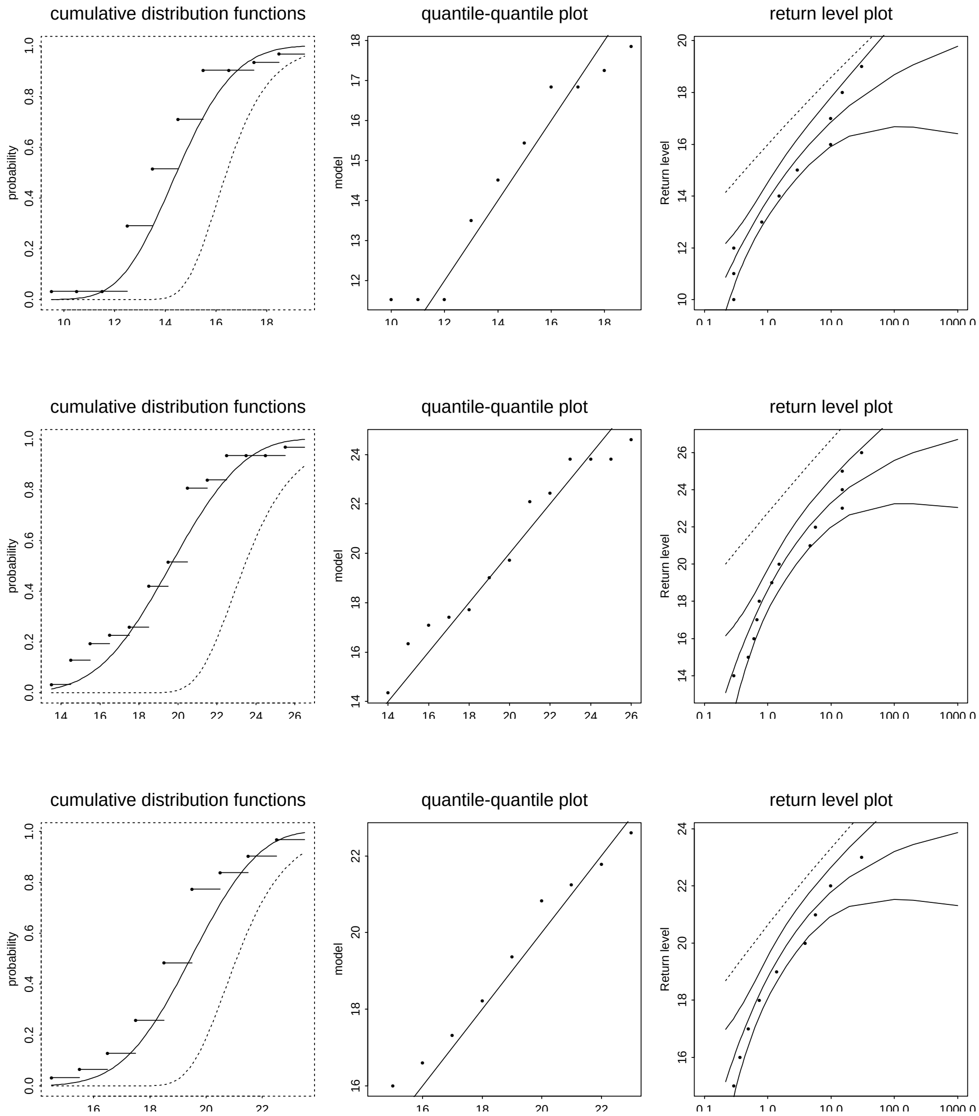

Figure 4: Stations with $\gamma = 0$ rejected. Top Bredåkra, middle Säve, bottom Visby. Solid lines: estimated values and 95% confidence bounds for Annual Maxima estimates. Dotted lines: Weibull estimates

In this paper we have used calendar years. It may be somewhat preferable to have different storm seasons in separate years, that is to let years start, say, at July 1 instead. Cook (1982, 1985) finds that dynamic pressure leads to a better fit of extreme winds to the Gumbel distribution. We have not investigated if this is the case also for the present data set.

For the Weibull Method the 50 year return level estimates were close to those given by the Annual Maxima method for some of the sites, but still rather different for others. For other return periods the Weibull results differed from those obtain by the Annual Maxima method.

On a more technical level, taking into account that the observations were rounded to m/s without decimals changed the parameter estimates rather little in the Annual Maxima analysis. However, still

- the rounding to m/s caused a loss of precision in estimates of extreme wind speeds.

In fact, as discussed at the beginning of Section 4, rounding made the standard errors of the estimates of the scale and shape parameters substantially larger, and also made confidence intervals for the return levels markedly wider. Both results were as expected.

Confidence intervals obtained using profile likelihood and expected information were rather similar. Since the shape parameters of the fitted EV distribution were close to zero this was as expected. The sample sizes are rather small, and it would be interesting to check the accuracy of the confidence intervals by simulation.

In conclusion, it is our opinion that the Weibull Method should not be used for estimation of extreme winds in Sweden.

A simple alternative is the Annual Maxima method assuming a Gumbel distribution, disregarding the binning of the observations and using maximum likelihood estimation. This may give slightly conservative estimates, but should still provide reasonably reliable results. However, if confidence intervals are computed, they will show a too optimistic view of the accuracy of the estimates.

The preferred alternative, however, is the Annual Maxima method with the binned likelihood functions and the EV distribution and profile likelihood confidence intervals. If not rejected by a likelihood ratio test the Gumbel distribution could alternatively be used. The quality of the fit should always be checked graphically.

The Peaks over Thresholds method is expected to produce similar results to the Annual Maxima method but has a potential to be slightly more accurate, at the expense of the arbitrariness involved in the choice of storm events, or more or less equivalently, in the choice of level and in the declustering. We have not investigated this for the present data set.

A further problem is the influence of inhomogeneous site conditions (topography and land use) at meteorological stations and the effects of temporal changes in site conditions or instruments at meteorological stations. This might in principle be approached by careful adjustments for inhomogeneities and temporal changes.

An attractive alternative way to avoid these problems is to use so called surface geostrophic wind speeds, i.e. winds derived from surface pressure fields and adjusted to account for surface friction, instead of observed winds. In sections 4.4 - 4.9 of the Danish Wind Atlas (Lundtang Petersen et al. (1981)) this so called "geostrophic wind" approach is described in some detail. The method is applied - with some modifications - by the SMHI to obtain primary data on surface geostrophic wind speeds for a somewhat nonstandard PoT analysis of extreme winds. Among the results from these analyses are the current Swedish reference wind speed data included in "Snö och Vindlast" (Nero and Åkerlund 1997). However, this is outside of the scope of the present paper.

It is an interesting future task to investigate the spatial dependence of measurements taken at different recording stations, and how it can be used to improve estimation of extreme winds.

Acknowledgment: We are grateful to Stuart Coles for providing the S+ functions for the non-binned likelihood method and for inspiration on how to adapt these functions to binned data.

References

[1] Alexandersson, H. (1979) A statistical analysis of wind, wind profiles and gust ratios at Gränby, Uppsala. *Report No 55, Dep. of Meteorology, Univ. of Uppsala.*

[2] Bergström, H. (1992) Distribution of extreme wind speed. *Wind energy report WE 92:2, Dep. of Meteorology, Univ. of Uppsala.*

[3] Brockwell, P. J. and Davis, R. A. (1996) *Introduction to time series and forecasting.* Springer, New York.

[4] Conradsen, K., Nielsen, L.B. and Prahm, L.P. (1984) Review of Weibull statistics for estimation of Wind speed distributions. *J. Climate & Applied Meteorology* **23**, 1173 - 1183.

[5] Coles, S. (2001) *An Introduction to Statistical Modelling of Extreme Values.* Springer, New York.

[6] Cook, N.J. (1982) Towards better estimation of extreme winds. *J. Wind Engineering and Industrial Aerodynamics* **9**, 295-323.

[7] Cook, N.J. (1985) *The Designers guide to wind loading of building structures* BRE dep. of the environm., Garston, & Butterworths, London.

[8] Davenport, A.G. (1967) The dependence of wind loads on meteorological parameters. *Proc. Int. Res. Seminar, Wind Effects on Buildings and Structures*, Ottawa, Univ. of Toronto Press, 19 - 82.

[9] Embrechts, P., Klüppelberg, C. and Mikosch, T. (1997) *Modelling extremal events for insurance and finance.* Springer, Berlin.

[10] Kulldorf, G. (1961). *Contribution to the theory of estimation from grouped and partially grouped samples* Almqvist &Wiksell, Stockholm.

[11] Leadbetter, M.R., Lindgren, G. and Rootzén, H. (1983) *Extremes and related properties of random sequences and processes.* Springer, New York.

[12] Nero, K and Åkerlund, S. (eds.) (1997) *Snö och vindlast.* BSV97, *utgåva 2.* Publikationsservice, Karlskrona.

[13] Lundtang Petersen, E., Troen, I., Frandsen, S. and Hedegaard, K. (1981) Danish Wind Atlas, a rational method of wind energy siting. *Report Riso-R-428*, Riso National Laboratories, Denmark.

[14] Palutikof, J.P., Brabson, B.B., Lister, D.H. and Adcock, S.T. (1999) A review of methods to calculate extreme wind speeds. *Meteorol. Appl.* **6**, 119-132.

[15] Rasmussen, P. F. and Rosbjerg, D. (1991) Evaluation of risk concepts in partial duration series. *Stochastic Hydrology and Hydraulics.*

[16] Rootzén, H. and Tajvidi, N. (1997) Extreme value statistics and wind storm losses: a case study. *Scand. Actuarial J.* No. 1, 70 - 94.

[17] Teugels, J.L. (2000) Wind and statistics, a guide to the literature. *Report, Dept. of Statistics, Katholike Universiteit Leuven.*

[18] Sen, K.S. and Singer J.M. (1993) *Large samples methods in statistics.* Chapman & Hall, New York.

[19] World Meteorological Organization (1983) *Guide to Meteorological Instruments and Methods of Observation*, WMO No 8, Geneva.

Appendix: a comparison with the Weibull-Gumbel method

In connection with producing wind norms for Sweden, the SMHI made a comparison of different approaches to extreme wind estimation. In this appendix we compare results from the present paper with a part of the SMHI results on "the Weibull-Gumbel approach". In this approach, a Weibull distribution is fitted to all wind speed data, and maxima of wind speeds are then assumed to follow a Gumbel distribution with parameters derived from the parent Weibull distribution.

Table 4: 50-year return levels. W = Weibull, EV = Annual Maxima with EV distribution and binned likelihood, G = Annual Maxima with Gumbel distribution and binned likelihood, * = the Gumbel distribution was rejected at the 5 % level, and WG = the Weibull-Gumbel method.

	W	EV	G	WG
Arlanda	20	20	19	19
Bromma	19	19	19	18
Karlstad	21	20	20	20
Ljungbyhed	28	22	23	24
Norrköping	24	25	26	22
Säve*	29	25	29	26
Söderhamn	19	22	21	19
Uppsala	22	23	22	21
Västerås	22	21	22	20

Specifically, if wind measurements were made every hour for thirty years, then the T-year wind is assumed to have the Gumbel distribution $G_0(u) = \exp(-\exp(-\frac{u-\mu_T}{\sigma_T}))$ where μ_T and σ_T are called the modal value and dispersion factor, respectively, and are obtained from the Weibull parameters μ, σ as

$$\begin{aligned} \mu_T &= \sigma(\log(N))^{1/k} \\ \sigma_T &= \frac{\sigma}{k}(log(N))^{1/k-1} \end{aligned}$$

for $N = 24 \times 365 \times 30 \times T$. Variants of this method include adjustment for dependence in the data. The SMHI results for the Weibull-Gumbel method are given in Table 4. The consistently lower values for the Weibull-Gumbel method as compared to the Weibull Method is due to adjustments for dependence in the version of the former method which is used by the SMHI.