Uncertainties in extreme surge level estimates from observational records

By H.W. van den Brink, G.P. Können & J.D. Opsteegh

Royal Netherlands Meteorological Institute, P.O. Box 201, 3730 AE De Bilt, The Netherlands

Ensemble simulations with a total length of 7540 years are generated with a climate model, and coupled to a simple surge model to transform the wind field over the North Sea to the skew surge level at Delfzijl (NL). The 65 constructed surge records, each with a record length of 116 years, are analyzed with the Generalized Extreme Value (GEV) and the Generalized Pareto distribution (GPD) to study both the model and sample uncertainty in surge level estimates with a return period of 10^4 years, as derived from 116-year records.

The optimal choice of the threshold, needed for an unbiased GPD-estimate from Peak-Over-Threshold (POT) values, cannot be determined objectively from a hundred-year data set. This fact, in combination with the sensitivity of the GPD-estimate to the threshold, and its tendency towards too low estimates, leaves the application of the GEV distribution to storm-season maxima as the best approach. If the GPD-analysis is applied, the exceedance rate λ should be chosen not larger than 4.

The climate model hints at the existence of a second population of very intense storms. As the existence of such a second population can never be excluded from a hundred-year record, the estimated 10^4-year wind speed from such records has always to be interpreted as a lower limit.

Keywords: Extreme value statistics; surge; uncertainty; climate change

1. Introduction

In The Netherlands, a probability of 10^{-4} per year for flooding from the sea is used as baseline for dike design (Deltacommissie 1960). Several problems arise when translating this 'accepted risk' into the sea level being exceeded (on average) only once in 10^4 years. First, as the observational records of tidal stations are only 10^2 years in length, the surge level with an average return period of 10^4 years requires an extrapolation of two orders of magnitude. It is unclear how reliable the estimate from such an extrapolation is. Second, various probability functions can be fitted to the observational records of extreme surges, leading to different results in the 10^4-year return levels (Dillingh *et al.* 1993, de Haan 1990). Third, extrapolation from observational records does not contain information about surges in a greenhouse gas induced changing climate. Fourth, a second population of rare but intense storms, originating from a different kind of meteorological system, would result in higher return values than estimated from standard extreme value analysis of the available short records.

These problems are explored by analyzing two very long surge records for the Dutch coastal station Delfzijl, which were generated by a climate model. One series refers to the present-day climate; the second to the future (doubled greenhouse gas concentration) conditions. The length of these series (order 10^4 years) allow for exploring the far tail of the distribution, as well as for uncertainty estimates of the return values if calculated from much shorter (order 10^2 years) subsets.

2. Model descriptions

Wind data are generated by the General Circulation Model (GCM) ECBilt-Clio, consisting of an ocean model Clio (Goose & Fichefet 1999) and an atmospheric model ECBilt (Opsteegh *et al.* 1998, 2001). ECBilt is a spectral T21 global 3-level quasi-geostrophic model, with a time step of 4 hours. The T21-resolution corresponds (for the latitudes of interest) with a grid point distance of approximately 500 km.

The surge model we used is a simplified version of the Timmerman model (Timmerman 1977). It is described and validated in van den Brink *et al.* (2003). We calculated a surge level every 12 hours.

3. Methodology

(*a*) *Extreme Value distributions*

There are two commonly applied approaches in extreme value statistics: In the first approach, 'block maxima' are considered, to which the Generalized Extreme Value (GEV) distribution is applied. The GEV distribution function is given by:

$$\text{GEV} = P(Y \le y) = \exp\left(-\left[1 - \frac{\theta}{\alpha}(y - \mu)\right]^{1/\theta}\right) \tag{3.1}$$

with μ the location parameter, α the scale parameter, θ the shape parameter, and y the block maximum of the considered variable (de Haan 1976).

In the second approach, all values above a certain threshold u are considered. To these 'Peak over Threshold' (POT) values, the Generalized Pareto Distribution (GPD) is applied. The GPD distribution function is given by:

$$\text{GPD}_\lambda = P(Y - u \le y | Y > u) = 1 - \left[1 - \frac{\theta}{\alpha}(y - \mu)\right]^{1/\theta} \tag{3.2}$$

with μ the location parameter, α the scale parameter, θ the shape parameter, and y the variable above a chosen threshold u. We follow the common approach to choose μ equal to the threshold u (Palutikof *et al.* 1999). The exceedance rate λ, which depends on the threshold u, is estimated as the average number of exceedances over the threshold u per 'block'. From both approaches, the level belonging to a given probability of exceedance can be estimated by inverting equations 3.1 and 3.2.

The shape parameters θ of the GEV and the GPD distributions are equal if the threshold is large enough (Katz *et al.* 2002).

In order to come to an optimal estimate, it is desirable that the estimate is both unbiased (i.e., with the right expected value) and efficient (i.e., with a small uncertainty). The uncertainty depends mainly on the number of samples that are

considered, whereas a systematic bias will be introduced if a wrong distribution is used to describe the data. As both the GPD and the GEV distribution describe only the 'tail' of the parent distribution, a bias will be introduced if samples that do not belong to this 'tail' are also considered. Which samples belong to the tail, in the sense that they can be described with the same parameters as the more extreme events, depends on the convergence rate of the parent distribution to the asymptotic extreme value distribution. For the 'block maxima' approach, this convergence is assumed beforehand, whereas for the POT-approach, this question is commonly answered empirically, making use of the fact that for the 'tail', the estimated shape parameter θ should be independent of the threshold u, and thus of the exceedance rate λ. This can be explored by plotting the estimated shape parameter θ as a function of the threshold u or the exceedance rate λ. The chosen λ is then the largest one for which θ is stable (de Haan 1990, Coles 2001). If there are strong fluctuations or trends in the estimated θ, quantile estimates are difficult to obtain (see e.g. Brabson and Palutikof 2000).

So, the larger sample set that is considered in the POT-approach (if $\lambda > 1$) makes this method more efficient than the 'block maxima' approach. On the other hand, the POT-approach is more likely to be biased, as samples less far in the tail of the distribution are also used.

For a further overview of the advantages and disadvantages of the POT-approach and the 'block maxima' approach, we refer to Palutikof *et al.* (1999).

(*b*) *Set-up of the numerical experiment*

With ECBilt-Clio, 260 runs of 30 years each were generated, with a CO_2 concentration according to the period 1960-1989 (320 ppm on average). In addition to the control run, we also generated 233 ensemble runs of 30 years with estimated CO_2 concentrations according to the period 2050-2079 (following the SRES A1B CO_2 emission scenario (Houghton *et al.* 2001)). This emission scenario results in approximately doubled CO_2 concentration (620 ppm on average) in 2050-2079 with respect to the control run.

As every 30-year run contains 29 storm season periods, we have $260 \times 29 = 7540$ 'block maxima' for the control run, and 6902 for the greenhouse run.

(*c*) *Data handling*

To remove dependent events from the POT-selection, we require a minimal time separation between selected events of 96 hours, as in de Haan (1990).

We concentrate on storm season events (October till March) to improve homogeneity of the data set (de Haan 1990). We applied the surge model to the ECBilt-Clio grid point (6E,47N). This grid point best represents the North Sea winds (van den Brink *et al.* 2003).

We calculated the parameters of the GEV and GPD distributions via the method of Maximum Likelihood (Coles 2001). The 95%-confidence levels were estimated from the profile likelihood (Coles 2001).

Figure 1 shows the number of exceedances λ as a function of the threshold u for the observations and the ECBilt-Clio data. ECBilt-Clio has somewhat less exceedances over a given threshold than the observations. In this study, we compare
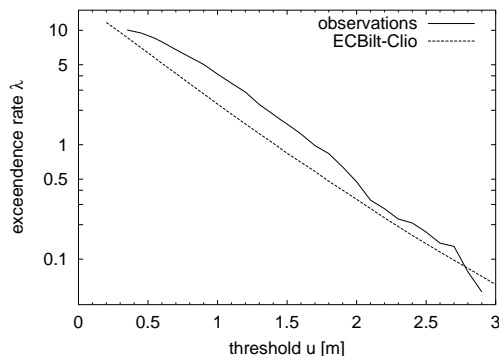
Figure 1. Number of exceedances λ per storm season (October to March) as a function of the threshold u for the observations and the ECBilt-Clio data. The vertical scale is logarithmic.
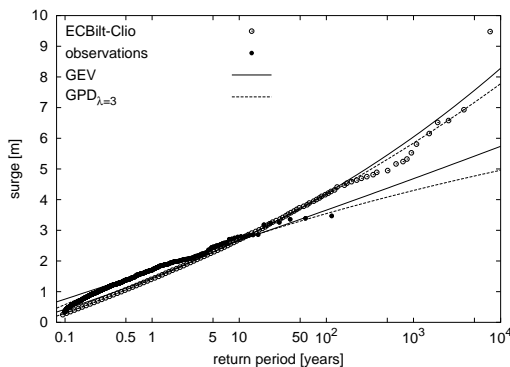

Figure 2. Return level plot for the observational set and the ECBilt-Clio set. The lines are GEV- and GPD-fits with exceedance rate $\lambda = 3$ to both sets. The horizontal scale is logarithmic.

both record for situations with equal exceedance rate λ, which means that for the observational record a higher threshold is chosen than for the ECBilt-Clio record.

Figure 2 shows a return level plot for the observational set and the ECBilt-Clio set. The extremes of ECBilt-Clio are in reasonable agreement for return periods up to 50 years. For return periods larger than 50 years, the extremes in ECBilt-Clio are higher than in the observational set.

4. Results

(a) Dependence of 10^4-year estimate on model choice

With the control run, we tested the uncertainty in the extrapolation of the extreme surges for Delfzijl. We applied the GPD distribution to the 116-year observational surge record of Delfzijl (1883-1999), to 65 subsets of 116 years each of ECBilt-Clio, and to the total set of 7540 years, all for several choices of λ. We
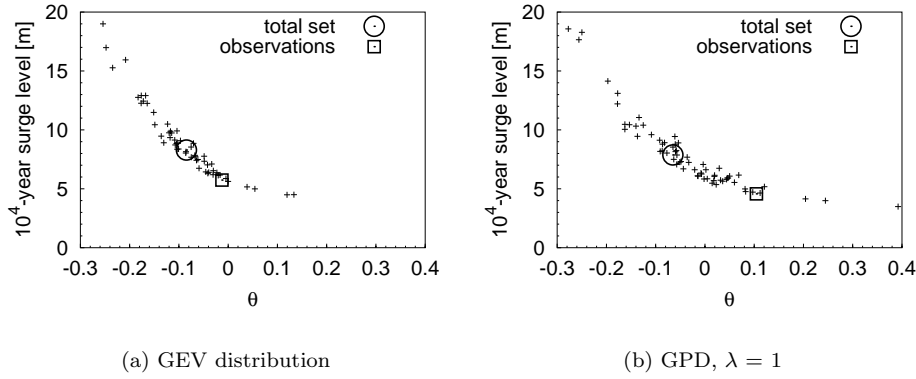
(a) GEV distribution   (b) GPD, $\lambda = 1$

Figure 3. The estimated 10^4-year return level for the surge as a function of the corresponding shape parameter θ for 65 subsets (each of 116 years length) from the ECBilt-Clio control run. Shown are the GEV-estimates (a) and the GPD-estimates for exceedance rate $\lambda = 1$ (b). Also shown are the estimates from the total ECBilt-Clio control run of 7540 storm seasons (circle), and the estimate from the 1883-1999 observational set of Delfzijl (square).

also applied the GEV distribution to the storm season block-maxima of all these sets. The 10^4-year estimates are shown in Figure 3 as a function of the estimated shape parameter θ from the GEV distribution (panel a) and from the $\text{GPD}_{\lambda=1}$ distribution (panel b). Figure 3 shows the following features: First, both panels resemble the strong correlation between the estimated shape parameter θ and the estimated 10^4-year return level. Second, the 10^4-year estimate from the total 7540-year ECBilt-Clio set are similar for the GEV (8.29 m (7.21,10.9)) and the $\text{GPD}_{\lambda=1}$ estimate (7.87 m (7.28,8.62)) (the values between brackets are the lower- and upper 95%-confidence levels). Third, the lower $\text{GPD}_{\lambda=1}$ estimate (4.66 m (3.70,8.95)) than the GEV estimate (5.85 m (4.17,11.5)) indicates that the two approaches can result in considerably different 10^4-year estimates (although they do not differ significantly in this case). Fourth, the 10^4-year estimates of the 116-year ECBilt-Clio subsets vary considerably, between 4 and 20 m, both for the GEV- and the $\text{GPD}_{\lambda=1}$ estimates.

(b) *Dependence of GPD 10^4-year estimate on exceedance rate*

We now want to explore if $\lambda > 1$ makes application of the GPD distribution more efficient than the GEV distribution. Figure 4 shows the estimated GPD shape parameters and 10^4-year return levels as a function of λ. Figure 4(a) shows the estimates from the total 7540-year set, two arbitrarily chosen 116-year ECBilt-Clio sets and the observational set. Figure 4(b) gives the estimated 10^4-year surge levels for the same sets.

Figure 4 gives the following information: First, if $\lambda < 1$, the estimate of θ (and thus the 10^4-year estimate) from a 116-year subset is very sensitive to λ, which is undesired. Second, also for $\lambda > 1$, considerable fluctuations in the estimated shape parameter θ remain in 116-year sets, as both the two ECBilt-Clio sets and
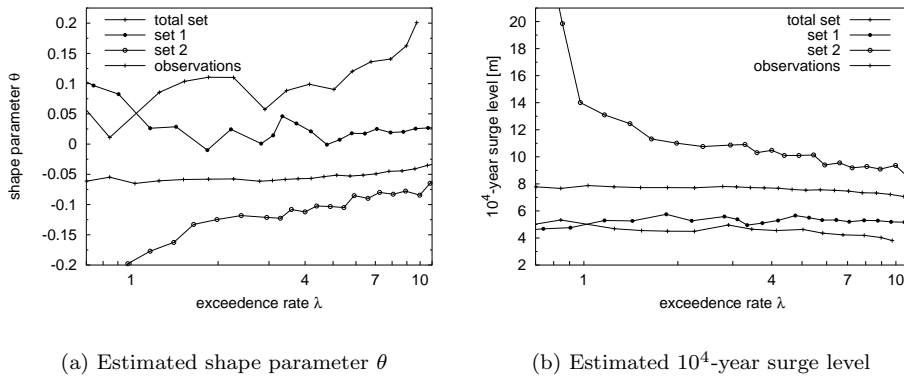
(a) Estimated shape parameter θ (b) Estimated 10^4-year surge level

Figure 4. (a) Estimated shape parameters θ of the GPD-distribution for the surge in Delfzijl according to the total 7540-year ECBilt-Clio set as a function of the exceedance rate λ. Also shown are estimates from two arbitrarily chosen 116-year ECBilt-Clio subsets (set 1, set 2), and the observational set. (b) The corresponding 10^4-year surge levels. The horizontal axes are logarithmic.

the observational set show. This fact, together with the different 'stable' regions for θ of the two ECBilt-Clio 116-year subsets, make it difficult or even impossible to choose an optimal value of λ from a 116-year record. Third, the two 116-year ECBilt-Clio subsets remain for all λ's either below, or above the estimate of the total 7540-year ECBilt-Clio set. Fourth, even the estimates of the total 7540-year ECBilt-Clio set are not threshold-independent. This suggests that the upward slope of θ (for $\lambda > 4$) in Figure 4(a) (and the corresponding decreasing 10^4-year estimate in Figure 4(b)) is a bias, caused by samples that do not belong to the tail of the parent distribution. Fifth, the fact that the estimates from the observational set are within the ECBilt-Clio range for the GEV and the GPD distribution for $\lambda \lesssim 3$, but outside that range for $\lambda \gtrsim 3$, might indicate that the observational set is even more biased for large values of λ than the ECBilt-Clio set.

We conclude that λ should be in the range between 1 and 4 to have a more or less unbiased 10^4-year surge estimate. However, such a range can only be determined from an extremely long data set. Data sets of order hundred years are too short to determine a maximal choice of λ (and thus of the minimal threshold) that results in an unbiased estimate. The strong dependence of the GPD-estimates on the choice of λ makes it difficult, or even impossible, to obtain reliable unbiased GPD-estimated 10^4-year surge levels from 10^2-year records.

(c) Optimal choice for threshold

In order to investigate the bias in more detail, we determine the fraction of the 65 subsets for which the actual value lies outside the 95%-confidence interval. These percentages are shown in Figure 5(a) for a return period of 100 years, split out to the lower- and upper 95%-confidence level. The actual 100-year value is 4.20 m (according to Figure 2). The open circles in Figure 5(a) show that the fraction of upper 95%-confidence levels is larger than 2.5% for all exceedance rates,

except between 2 and 4. For $\lambda \geq 5$, the number of subsets for which the upper 95%-confidence levels is below the actual value is large, whereas for none of the 65 subsets the actual value is below the the lower 95%-confidence levels. This indicates the existence of bias towards too low values.

Fitting a GEV-distribution to the 65 subsets gives one subset for which the 100-year upper 95%-confidence level is lower than the actual value, and one subset for which the lower 95%-confidence level is higher than the actual 100-year value. The corresponding percentages are plotted at the right side of Figure 5(a).

To highlight the effects of Figure 5(a), the calculations are repeated, but now based on 1000 116-years subsets, obtained by randomly sampling from the total ECBilt-Clio set, in order to decrease the noise. The results for a return period of 100 years are shown in Figure 5(b). This confirms the findings of Figure 5(a) that there is a bias in the GPD$_\lambda$-estimates. Whereas Figure 5(a) did not indicate a bias for $\lambda = 3$, Figure 5(b) shows that there is a small bias for $\lambda < 4$, which strongly increases for $\lambda > 4$. Figure 5(b) indicates no bias in the GEV-estimates (about 2.5% of the upper 95%-confidence intervals is lower than the actual value, and about 2.5% of the lower 95%-confidence intervals is above the actual value).

For the 10^4-year return periods, the percentages exceeding the upper- and lower 95%-confidence intervals are depicted in Figure 5(c–f). In this case, i.e., for 10^4-year return periods, the 'real' value for the 10^4-year return value has to be chosen, as it can not be determined directly from Figure 2. As possible 'real' values we considered both the GEV- and the GPD$_{\lambda=3}$-estimates, as obtained from the total 7540-year set. The reason for considering $\lambda = 3$ in the GPD-estimate is that this value of the exceedance rate λ turns out to be the best, according to Figure 5(a). Another reason is that the GPD$_{\lambda=3}$-estimate is correct for a 100-year return period (Figure 2).

Figures 5(c,e) show the results for the 65 subsets, respectively taking the 10^4-year GEV-estimate (8.29 m) as 'real' value, and the 10^4-year GPD$_{\lambda=3}$-estimate (7.78 m). Figures 5(d,f) show the results for the 1000 sampled subsets.

We see an even stronger bias towards too low values for the 10^4-year return periods than for the 100-year return periods for the GPD-estimates if $\lambda > 4$ (note the different vertical range of Figures 5(c–f) with respect to Figures 5(a,b)). No bias is detected for the GEV-estimates in the situation that the 10^4-year GEV-estimate is taken as 'real' value (Figures 5(d)), and for the GPD-estimates if $1 < \lambda < 4$ in the situation that the 10^4-year GPD$_{\lambda=3}$-estimate is taken as 'real' value (Figures 5(f)), as expected from consistency.

We conclude from Figure 5 that the GPD-estimates are more sensible for bias than the GEV-estimates, especially if the exceedance rate $\lambda > 4$. The GEV-estimates are unbiased. This leaves the GEV analysis as the preferred method.

(*d*) *Greenhouse effect on wind*

The effect of the greenhouse doubling on the extreme wind speed in ECBilt-Clio for the North Sea grid point is shown in Figure 6. Up to return periods of ≈ 100 years, no effect is apparent. However, for wind speeds with return periods of more than 250 years, the greenhouse run deviates systematically from the fitted GEV distribution. This suggests the existence of a second population in the extreme wind distribution. Fitting the GEV distribution to the deviating extremes only, results
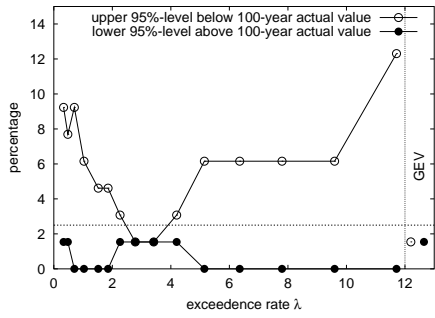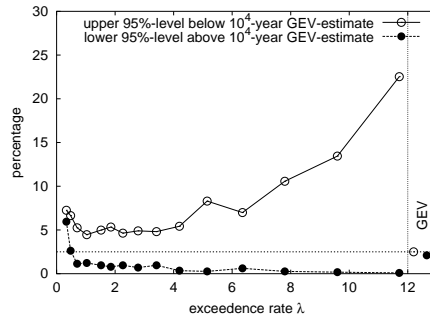
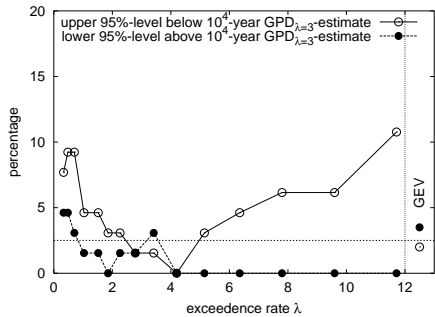(a) 100-year GPD- and GEV-estimates compared with the actual 100-year value (4.20 m), for 65 116-year subsets.
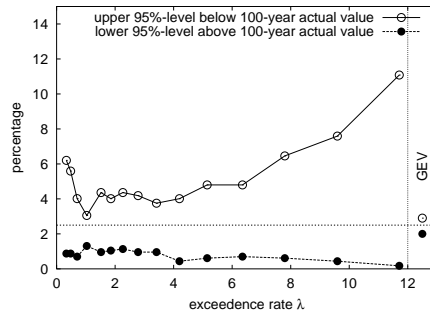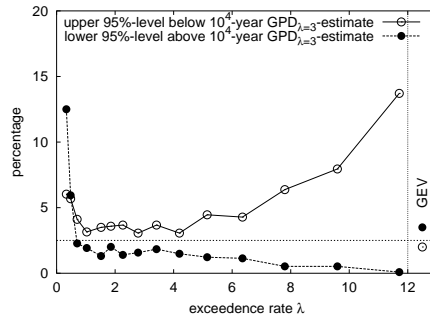
(b) idem as (a), for 1000 sampled sets.
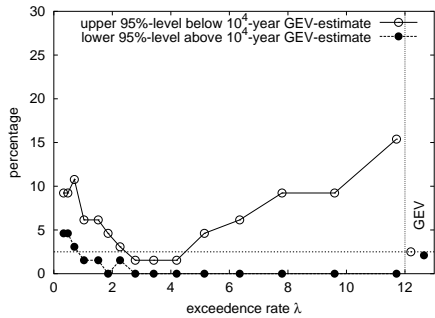
(c) 10^4-year GPD- and GEV-estimates compared with the 10^4-year GEV-estimate of the 7540-year set (8.29 m)

(d) idem as (c), for 1000 sampled sets.

(e) idem as (c), compared with 10^4-year the $GPD_{\lambda=3}$-estimate of the 7540-year set (7,78 m).

(f) idem as (e), for 1000 sampled sets.

Figure 5. Percentages of the number of sets that do not contain the 'real' value within its 95%-confidence intervals, estimated with the GPD_λ-distribution for different exceedance rates λ, and with the GEV-distribution.The 'real' value is chosen to be the actual 100-year value of the 7540-year set in (a,b), the 10^4-year GEV-estimate of the 7540-year set in (c,d), and the 10^4-year $GPD_{\lambda=3}$-estimate of the 7540-year set in (e,f). Left: 65 116-year subsets; Right: 1000 116-year sets, randomly sampled from the total ECBilt-Clio set.
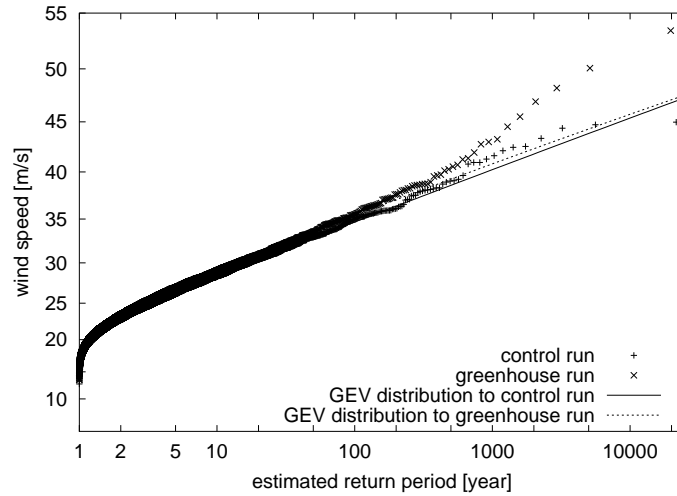
Article submitted to Royal Society

Figure 6. Return level plot of the observed and GEV-estimated 12-hourly averaged wind speeds for the control- and greenhouse runs in ECBilt-Clio for the North Sea representing grid point (6E,47N). The kink at a return period of 250 years in the greenhouse run suggests the presence of a double population in the extreme wind distribution.

in a considerably higher 10^4-year return value for the wind speed than fitting to the total set. For a more comprehensive description, we refer to van den Brink *et al.* (accepted by Journal of Climate).

5. Discussion and conclusions

The variance in the GEV-estimates from 65 records of 116 years indicates that only a crude estimate of the 10^4-year surge level can be made from a single record with a length of the order of a hundred years. The GPD-estimates give lower 95%-confidence intervals for exceedance rates $\lambda > 1$ than the GEV-estimates, but the total 7540-year ECBilt-Clio set shows that these GPD-estimates are biased toward lower 10^4-year values.

For the ECBilt-Clio data, the percentage of the 95%-confidence levels containing the actual value can be determined for a return period of 100 years, and estimated for a return period of 10^4 years. This analysis points out that application of the GPD-analysis to the ECBilt-Clio data leads to estimates that are biased towards too low values. We emphasize that this analysis can only be done for extremely long sets, and thus not for the short observational sets.

The unknown optimal value of the exceedance rate λ for the observational set, combined with the sensitivity of the GPD-estimate to the choice of λ, and the tendency towards too low estimates, leaves in our opinion the GEV analysis as the preferred method to apply to the observational data, despite of its large uncertainty. If the GPD-analysis is applied, λ should be chosen not larger than 4.

In the future, output of more advanced climate- and surge models will be used to calculate the 10^4-year surge level and its uncertainty. Another possibility may be to

apply the optimal λ, as obtained from the climate model, to the observational data and still estimate the 10^4-year surge level and its uncertainty from the observations.

ECBilt-Clio hints at the excitation of extratropical 'superstorms', defined as storms with more extreme winds than expected from extrapolation of less extreme events. The fact that this second population is only apparent in the greenhouse run for this grid point indicates that regions where second populations exist, can be shifted, enhanced or generated by climate change. The reality of this model-induced second population has still to be shown (van den Brink *et al.* , accepted by Journal of Climate). Due to their extreme rarity, they are not detectable from records of only hundred years in length. Reversing this argument implies that extrapolations from 100-year records to 10^4-year return levels are only valid under the condition that the extreme value distribution is single populated. As this condition can never be proved from 100-year records, the GEV- or GPD- (or any other distribution) estimated 10^4-year wind speed from 100-year records has always to be interpreted as a lower limit.

References

Brabson B. B., & J. P. Palutikof 2000 Tests of the Generalized Pareto Distribution for Predicting Extreme Wind Speeds. *Meteorological Applications*, **39** 1627–1640

Coles S. 2001 *An Introduction to Statistical Modeling of Extreme Values* Springer-Verlag London

de Haan L. 1976 Sample extremes: an elementary introduction. *Statistica Neerlandica*, **30**, 161-172.

de Haan L. 1990 Fighting the arch-enemy with mathematics. *Statistica Neerlandica*, **44**, 45–68.

Deltacommissie 1960 *Rapport Deltacommissie*. SDU uitgevers, Den Haag (in Dutch).

Dillingh D., L. de Haan, R. Helmers, G.P. Können, & J. van Malde 1993 De basispeilen langs de Nederlandse kust; statistisch onderzoek. Technical Report DGW-93.023, Ministerie van Verkeer en Waterstaat, Directoraat-Generaal Rijkswaterstaat (in Dutch).

Goose H & T. Fichefet 1999 Importance of the ice-ocean interactions for the global ocean circulation: a model study, *Journal of Geophysical Research*, **104**, 23 337–23 355.

Houghton J.T., Y. Ding, D. J. Griggs, M. Noguer, P.J. van der Linden, X. Dai, K. Maskell, & C.A. Johnson 2001 *IPCC Working Group I Third assessment report*. Cambridge University Press.

Katz, R.W., M.B. Parlange & P. Naveau 2002 Statistics of extremes in hydrology *Advances in Water Resources* **25** 1287-1304

Opsteegh J.D., R.J. Haarsma, & F.M. Selten 1998 ECBilt: A dynamic alternative to mixed boundary conditions in ocean models. *Tellus*, **50A**, 348–367.

Opsteegh J.D., F.M. Selten, & R.J. Haarsma 2001 Climate variability on decadal timescales. Technical Report 410 200 060, Dutch National Research Programme on Global Air Pollution and Climate Change.

Palutikof J.P., B. B. Brabson, D. H. Lister S. T. Adcock 1999 A review of methods to calculate extreme wind speeds, *Meteorological Applications*, **6**, 119-132

Timmerman H. 1977 *Meteorological effects on tidal heights in the North Sea*. Staatsdrukkerij, 's Gravenhage.

van den Brink H.W., G.P. Können & J. D. Opsteegh 2003 The reliability of extreme surge levels, estimated from observational records of order hundred years, *Journal of Coastal Research*, **19**, 376-388

van den Brink H.W., G.P. Können & J. D. Opsteegh 2004 Statistics of extreme synoptic-scale wind speeds in ensemble simulations of current and future climate *(accepted by Journal of Climate)*