INTERNATIONAL JOURNAL OF CLIMATOLOGY
Int. J. Climatol. (2009)
Published online in Wiley InterScience
(www.interscience.wiley.com) DOI: 10.1002/joc.2047

Estimating 10000-year return values from short time series

H. W. van den Brink* † and G. P. Können‡
Royal Netherlands Meteorological Institute, PO Box 201, 3730 AE De Bilt, The Netherlands

ABSTRACT: The distribution of outliers is used as a tool for finding the extreme value distribution of meteorological parameters and to provide return values for large return periods from short records. Its potential is demonstrated for five cases. For extreme winds in the Northern Hemisphere (NH) the method shows that appropriately transformed annual maximum wind speeds can be described by a Gumbel distribution; for extreme waves it rejects the proposed adoption of an exponential distribution and points to a Gumbel distribution; for extreme daily European precipitation R it confirms the theoretically predicted value $k = 2/3$ in its Weibull distribution and it also justifies the application of the Gumbel distribution to $R^{2/3}$ up to return periods of about 50 000 years; for seasonal precipitation in the Netherlands it highlights enhanced extreme precipitation in the coastal area in December-January-February (DJF) and failure of the $k = 2/3$ hypothesis outside June-July-August (JJA); for sea levels in the Southern North Sea it points to the Gumbel distribution and provides improved estimates for the 10^4-return value of the sea level at coastal stations, which is elaborated for the Dutch tidal station Scheveningen. Copyright © 2009 Royal Meteorological Society

KEY WORDS meteorological extremes; surges; waves; precipitation; risk analysis; statistical uncertainties; extreme value statistics; outliers

Received 20 March 2009; Revised 17 September 2009; Accepted 5 October 2009

1. Introduction

Much scientific research has been done on finding the best estimates of weather (related) extremes, like precipitation, wind speed, sea surges (e.g. Cook 1982; Koutsoyiannis 2004a; Van den Brink *et al.* 2004b), as well as other geophysical hazards, like earthquakes or tsunamis. An important issue in any of these studies is the choice for the distribution to be made to fit the observed extremes. Mostly, the generalized pareto distribution (GPD) is applied to the (independent) exceedances over a high threshold, or the generalized extreme value (GEV) distribution to annual maxima. However, estimates of high return values depend strongly on the value of the shape parameter of these distributions, which is hard to estimate with sufficient accuracy from the observational records. A method to bypass the problem is to try to lengthen the observational series by means of combining. This approach is applied by Van den Brink *et al.* (2005) by treating the sequence of operational seasonal ECMWF (European Centre for Medium-Range Weather Forecasts, Reading UK) forecasts as observational series of the current climate. It also forms the basis of the regional frequency analysis, where spatially homogeneous records are combined to reduce the statistical uncertainty (Buishand 1991; Hosking and Wallis 1997).

In an earlier article (Van den Brink and Können 2008) we developed an alternate approach to the problem. It is based on the empirical distribution of the highest value in an observational series (the 'outlier'). The method provides a diagnostic whether extrapolation to high return values is justified, and is able to yield an estimate of that return value with a higher precision than the traditional methods. By applying the method on the ERA40 wind data, we illustrated (Van den Brink and Können 2008) the potentials of that approach showing that the transformed ERA40 extreme wind speeds over the North Atlantic area are well described by a Gumbel distribution up to return periods of 10^4 years.

As an elaboration of Van den Brink and Können (2008), we here apply the same method to five quantities related to severe weather events. First, to extreme ERA40 wind speeds over the Northern Hemisphere. Second, to extreme significant ERA40 wave heights over the Northern part of the Northern Hemisphere (NH). Third, to extreme daily precipitation in Europe, using the ECA-D dataset (Klein Tank *et al.* 2002). Fourth, to extreme daily precipitation in the Netherlands. Fifth, to extreme surges in the Southern North Sea, by merging observations with data generated by the WAQUA surge model (de Vries 2000) driven by the ECHAM5-MPI climate model (Jungclaus *et al.* 2006).

This paper is structured as follows: Section 2 describes the theoretical framework; Section 3 gives some remarks about the interpretation of the theory; Section 4 describes the extreme value analysis of the applications mentioned and Section 5, the discussion and conclusions.

† Current address: MeteoGroup, PO Box 617, 6700 AP Wageningen, The Netherlands.
‡ Now: retired.
* Correspondence to: H. W. van den Brink, Royal Netherlands Meteorological Institute, PO Box 201, 3730 AE De Bilt, The Netherlands. E-mail: Henk.vandenBrink@meteogroup.com

2. Theory

In this section we present a short outline of the theory. For a more general formalism, reference is made to Van den Brink and Können (2008), who gives a slightly different derivation. The Appendix proves the equivalence of the two derivations.

The probability of an extreme event can be expressed in its return period T:

$$T = \frac{1}{1-F} \tag{1}$$

with F the cumulative distribution function of the annual maxima of the variable y. The probability that a T-year event y_T happens to occur in a certain year is given by:

$$\Pr(y > y_T) = 1 - F(y_T) \equiv 1 - \left(1 - \frac{1}{T}\right) = \frac{1}{T} \tag{2}$$

The value of y_T is called the return level or the return value. The probability that a T-year event happens to occur at least once in a n-year period is (assuming independence):

$$\Pr(y_n > y_T) = 1 - F(y_T)^n \equiv 1 - \left(1 - \frac{1}{T}\right)^n. \tag{3}$$

in which y_n is the highest value (the outlier) in a n-year period. It then follows:

$$\begin{aligned}
&\Pr(y_n \le y_T) = \left(1 - \frac{1}{T}\right)^n \\
&-\ln(\Pr(y_n \le y_T)) = -n \ln\left(1 - \frac{1}{T}\right) \\
&-\ln(-\ln(\Pr(y_n \le y_T))) = -\ln\left(-\ln\left(1 - \frac{1}{T}\right)\right) - \ln(n)
\end{aligned} \tag{4}$$

If we define:

$$\Delta X_T = -\ln\left(-\ln\left(1 - \frac{1}{T}\right)\right) - \ln(n) \tag{5}$$

then it follows:

$$\Pr(y_n \le y_T) = G(\Delta X_T) \tag{6a}$$

in which G is the standardized Gumbel distribution:

$$G(x) = \mathrm{e}^{-e^{-x}} \tag{6b}$$

For all practical situations, Equation (5) can be approximated by:

$$\Delta X_T \cong \ln(T) - \ln(n) \tag{7}$$

Van den Brink and Können (2008) gives a graphic visualization of Equation (6), showing how the horizontal 'distance' on a Gumbel plot between the plotting position of the highest event in a record and its theoretical distribution is Gumbel distributed.

3. Interpretation

Here, we summarize some of the practical aspects of Equation (6).

1. Equation (6) is exact if and only if $F(y)$ is known. In practical situations, F has to be empirically determined. Then, F is assumed to be of a certain distribution-type $\tilde{F}$, and its parameters $\hat{\boldsymbol{\theta}}$ are estimated from the data. Given that situation, Equation (6) can serve as a goodness-of-fit measure for F, as a wrong choice for $\tilde{F}_{\hat{\theta}}$ results in a wrong estimate of ΔX_n (denoted as $\Delta\hat{X}_n$) and hence in $\Delta\hat{X}_n$ being not Gumbel distributed. Note that no free parameters are left in Equation (6). We refer to Van den Brink and Können (2008) for the influence of sampling effects on the distribution of $\Delta\hat{X}_n$.
2. Equation (6) becomes of practical relevance in the case that multiple records are available, as every record ends up with a single value of $\Delta\hat{X}_n$. Combination of all values of $\Delta\hat{X}_n$ enables to test whether they are Gumbel distributed, and thus whether the fit to the data is appropriate for extrapolation. See Van den Brink and Können (2008) for a summary of the interpretation of the statistical distribution of $\Delta\hat{X}_n$.
3. Equation (6) focuses on the highest event in a record, and is thus especially of interest for the analysis of extremes, and for extrapolation purposes. Although not necessary, it is often convenient to use one of the extreme value distributions for F. In that case, Equation (6) represents the application of extreme value theory twice: first by bringing F to one of the extreme value distributions, and then by bringing $\Delta\hat{X}_n$ to the standardized Gumbel distribution (Equation (6b)).
4. While the standard goodness-of-fit tests are focused on testing how well the fit behaves in interpolation, this method focuses on extrapolation, which is much more appropriate in extreme value analysis.
5. For convenience, we assumed in Equation (1) F to be the distribution of the annual maxima. However, the derivation by Van den Brink and Können (2008) shows that Equation (6) holds for any arbitrary parent distribution.
6. In case of applying the method to time series of a given element observed on different locations, the first step is usually in the time domain and the second step in the spatial domain. In case of application to one single time series, the second step can be achieved by splitting the record into multiple (independent) shorter records and then by bringing the set of $\Delta\hat{X}_n$ of the sub-series to the standardized Gumbel distribution.
7. Equation (6) is valid for every $F(y)$ and n, as it analyses the probability of occurrence of the

DOI: 10.1002/joc

observed highest event (its 'exceptionality') and not its value. This implies that the records on which the method is applied, neither need to run simultaneously in time (like in gridded data) nor need to be of equal length. Also, records of different distributions (e.g. for land and sea or tropics and extratropics) could be combined.

8. The method can thus be considered as a pooling technique on the frequency of the variable (which is more general than pooling on its amplitude).
9. Special attention has to be paid to obtain independent values of $\Delta\hat{X}_n$, as required by Equation (3). The method has no restrictions due to statistical (in)dependence of the elements in the underlying. records, as it only requires that the events that determine the values of $\Delta\hat{X}_n$, i.e. the most exceptional events in every record are independent. In case of meteorological-related events this is guaranteed when the most exceptional events originate from distinct meteorological systems. A simple criterion is to require a minimum temporal interval between the dates that these events occurred. If an extreme event determines $\Delta\hat{X}_n$ for multiple records, then the event should be considered only at its most exceptional moment, i.e. only that record where the event has the highest (estimated) return period. In the case of spatially distributed time series (like gridded data) where the area of interest is very large, one can require that either the time interval exceeds a certain threshold, or the spatial distance is large enough. These thresholds depend on the variable of interest: extreme hourly precipitation will have a much smaller spatial and temporal correlation than daily temperature extremes. In the first situation, a time interval of 1 day or a spatial distance of 500 km will satisfy, whereas for temperatures, a time interval of 14 days or a distance of 2000 km will be more appropriate. Too large values for these thresholds only marginally influence the empirical distribution of $\Delta\hat{X}_n$, as the probability that multiple independent record extremes happen to occur within the time interval is very small.
10. The largest value of $\Delta\hat{X}_n$ in a set of records does not necessarily imply that it also corresponds to the highest event in that set but that, in the perspective of the location-specific climatologies, the event is most exceptional to occur in the given n-year period.
11. The analysis of the distribution of $\Delta\hat{X}_n$ can be performed by plotting the ordered values of $\Delta\hat{X}_n$, which is expressed in Gumbel-variate units, for m independent records on a Gumbel plot. In such a representation, $\Delta\hat{X}_n$ should be at the ordinate and the Gumbel-variate-transformed m at the abscissa. Note that m in this case represents the number of records, and not the number of years in a record. The distribution of $\Delta\hat{X}_n$ can thus easily be compared with the theoretical distribution (Equation (6b)), which is represented by the diagonal on such a Gumbel-Gumbel plot.
12. If the m values are Gumbel distributed according to Equation (6b), it can be concluded that the assumption about the type of $\tilde{F}$ is confirmed, and that the extrapolated fit of $\tilde{F}_{\hat{\theta}}$ is unbiased for all m records, up to return periods of $\sum_i^m n_i$ years, i.e. the total number of years in the m records.
13. While a Gumbel plot of $\Delta\hat{X}_n$ gives information about its statistical properties, a spatial representation of the values of $\Delta\hat{X}_n$ informs about the areas where the assumptions about F may fail. If low (or high) values of $\Delta\hat{X}_n$ are clustered in certain areas, this indicates that the assumptions about F fail for its extremes. (e.g. over sea, over mountains or latitude-bounded)
14. The method indicates whether the fitted distributions lead to systematic biases in the extrapolation or not, but do not provide information about the statistical uncertainty in the estimate of extreme return values.

4. Five Applications

4.1. Extreme NH wind speeds

In our work Van den Brink and Können (2008), we showed that the ERA40 annual extreme wind speed u over the North Atlantic area is Gumbel distributed if u^k is fitted instead of u, with k the locally determined Weibull shape parameter – a hypothesis originally put forward by Cook (1982). Here we extend this analysis to the entire NH (latitude $>10\,^\circ$N).

We use the 44 annual maxima for every grid point of the 10 m wind speed from the ERA40-dataset (Uppala *et al.* 2005) for the period 1958-2001. The T159 resolution is interpolated to a spatial resolution of 1° for the whole NH. Note that inhomogeneities in the wind may exist (especially on the Southern Hemisphere, Wang *et al.* 2006).

We calculated the Weibull shape parameter k from the upper 36% of all 6-hourly wind speeds for every grid point on the NH, fitted a Gumbel distribution (using maximum likelihood estimation, MLE) to the annual maxima of u^k, and calculated $\Delta\hat{X}_n$ according to Equation (5). Then the NH was subdivided into 24 boxes, each of size 20° in latitude and 60° in longitude (Figure 1).

We required a minimum interval of 3 days between the 1200 outliers in each box in order to ensure mutual independence. This yielded 11 up to 256 independent values of $\Delta\hat{X}_n$, depending on the box. The Gumbel plots of the distribution of $\Delta\hat{X}_n$ for every box are included in Figure 1, together with the positions of the largest outliers. The size of the circles corresponds to the value of $\Delta\hat{X}_n$.

Figure 1 shows that outside the tropics (latitude $>30\,^\circ$N), the assumption that the extremes of u^k can be described by a Gumbel distribution is confirmed, but that in the tropics the assumption fails. The kinks in the Gumbel plots of $\Delta\hat{X}_n$ in that region can be attributed to the occurrence of tropical cyclones, which generate a second population in the distribution of extreme winds (Van den Brink *et al.* 2004a).

DOI: 10.1002/joc

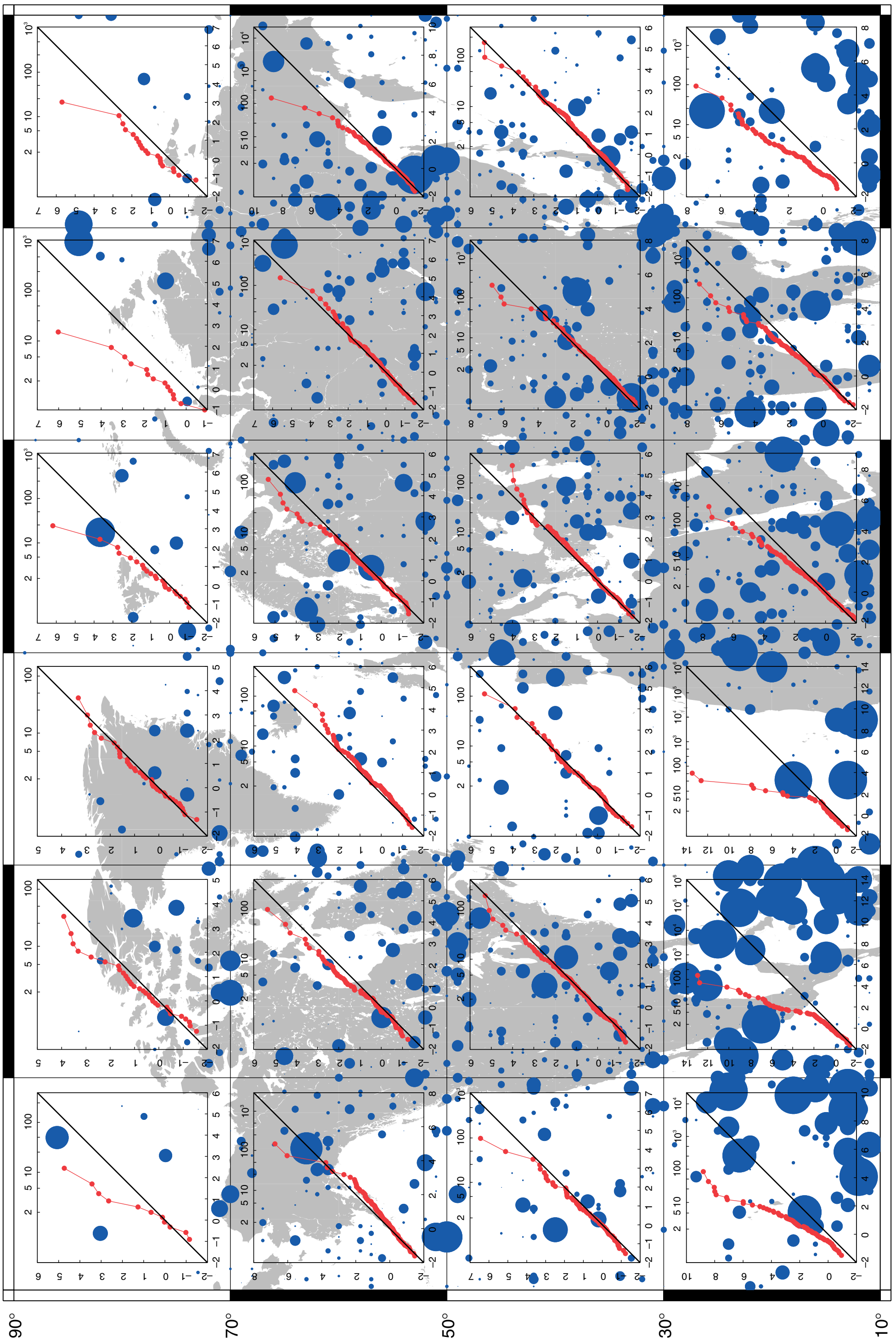

Figure 1. Extreme wind speed according to the 1958-2001 ERA40 data. The NH is divided into 24 boxes. The blue circles in them give the grid points where the biggest outliers in wind speed took place. The Gumbel plot in each box shows the distribution of the outliers $\Delta\hat{X}_n$ in that box.

4.2. Extreme waves

Caires and Sterl (2005) present global estimates of 100-year return values of significant wave heights, based on the ERA-40 reanalysis data. Their calculation of return values is based on the peaks-over-threshold (POT) method, with a threshold on the 93% level of all 6-hourly data. They conclude that 'the large amount of data used in this study provides evidence that the distribution of significant wave height belongs to the domain of attraction of the exponential'. Here, we test this conclusion by calculating the distribution of $\Delta \hat{X}_n$ for the waves in the 1958-2000 period between 30 and 70 °N. To ensure independence in the set of calculated values of $\Delta \hat{X}_n$ we required each pair of outliers with a mutual distance less than 7500 km to be separated by more than 4 days (>96 h). This selection procedure results in 192 independent values for $\Delta \hat{X}_n$, representative for a total of $192 \times 43 = 8256$ years.

Figure 2 shows the Gumbel plot of the 192 independent values of $\Delta \hat{X}_n$, for three choices of the distribution $\tilde{F}$: an exponential distribution (using L-moments, like in Caires and Sterl 2005), a GEV and a Gumbel distribution (using MLE). It shows that the exponential fit clearly underestimates the distribution of $\Delta \hat{X}_n$, and thus overestimates the extremes. According to Equation (7), the average 'distance' $\Delta \hat{X}_n$ between the highest event in a 43-year record and the 100-year return value should be $\ln(100) - \ln(43) = 0.84$. Instead, the estimated value $\Delta \hat{X}_n = 0.84$ corresponds in Figure 2 to a theoretical value $\Delta X_n = 2.04$, i.e. a return period of 330 years (Equation (7)). Fitting a GEV distribution leads to underestimation of high values of ΔX_n, a general feature of the GEV distribution, see Van den Brink and Können (2008).

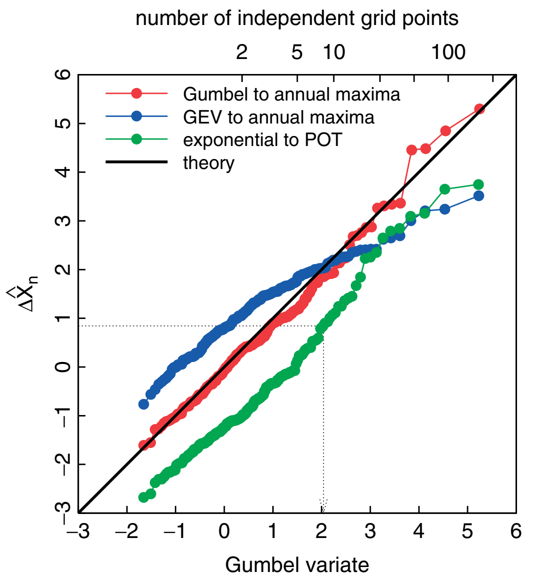

Figure 2. Gumbel plot of the statistical distribution of $\Delta \hat{X}_n$ for the significant wave heights in the 1958-2000 period between 30 and 70 °N for three assumed extreme value distributions of the wave heights (Gumbel, GEV and exponential). Assuming them Gumbel distributed performs best. The exponential to POT is the distribution applied by Caires and Sterl (2005). POT stands for peak-over-threshold.

However, fitting a Gumbel distribution to the annual maxima gives good results for the distribution of $\Delta \hat{X}_n$. This result is in accordance with Bouws *et al*. (1998) who state that "The Fisher-Tippett Type I (i.e. Gumbel) distribution often seems to give a good fit to 3-hourly data from the North Atlantic and North Sea" (p. 106).

We conclude that the proposal of Caires and Sterl (2005) to use an exponential distribution for the distribution of extreme wave heights must be rejected. The consequence of using a Gumbel distribution instead is that the 100-year return values become on average 15% lower than the values of Caires and Sterl (2005).

4.3. Extreme precipitation rates in Europe

Wilson and Toumi (2005) argue on theoretical grounds that the extreme precipitation R is Weibull distributed:

$$\Pr(R < r) = 1 - \exp\left[-\left(\frac{r}{R_0}\right)^k\right] \quad (8)$$

with the shape parameter k equal to 2/3. Analogously to the argument that justifies the transformation of the wind speed in Section 4.1, this implies that $R^{2/3}$ is exponentially distributed, with a fast convergence of its normalized maxima to the Gumbel distribution.

We extracted daily precipitation of 2482 European stations from the ECA-D dataset Klein Tank *et al*. 2002 of different lengths running in the period 1951-2008. We required that the records are more than 20 years in length. This leads to 2147 records with a total of 88 470 annual maxima. We fitted a Gumbel distribution to the annual maxima of $R^{2/3}$ for the 2147 records (using MLE), and calculated $\Delta \hat{X}_n$ according to Equation (5).

To ensure independence in the set of calculated values of $\Delta \hat{X}_n$ we required each pair of outliers with a mutual distance less than 1000 km to be separated by 2 or more days. This leads to 1379 independent values for $\Delta \hat{X}_n$, representing a total of 56 261 years. The Gumbel plot of the 1379 independent values of $\Delta \hat{X}_n$ is shown in Figure 3, together with the locations of the stations used. Only the 1379 mutually independent stations are shown. The colours correspond to the values of $\Delta \hat{X}_n$.

Figure 3 shows that the distribution of $\Delta \hat{X}_n$ corresponds well with theory, which confirms the assumption of Wilson and Toumi (2005) that extreme precipitation is Weibull distributed with shape parameter $k = 2/3$ up to return periods of about 50 000 years.

The highest $\Delta \hat{X}_n$ value of 10.02 is found in the 1974-2005 record of Meknes, Morocco (5.53 °W, 33.88 °N) on 24 October 1977, with 249.9 mm. Its estimated return period of 584 500 years has a probability of almost 10% to occur within 56 261 years. Note that the precipitation amount of this event is only the 25th highest in the whole record, but nevertheless the most exceptional event given its climatology (i.e. its distribution $\tilde{F}_{\hat{\theta}}$) and its length n of 26 years. The even higher value for $\Delta \hat{X}_n$ that is present in the set (10.89 for the 204 mm event in Lien i Selbu, Norway on 17 March 2003, with an estimated

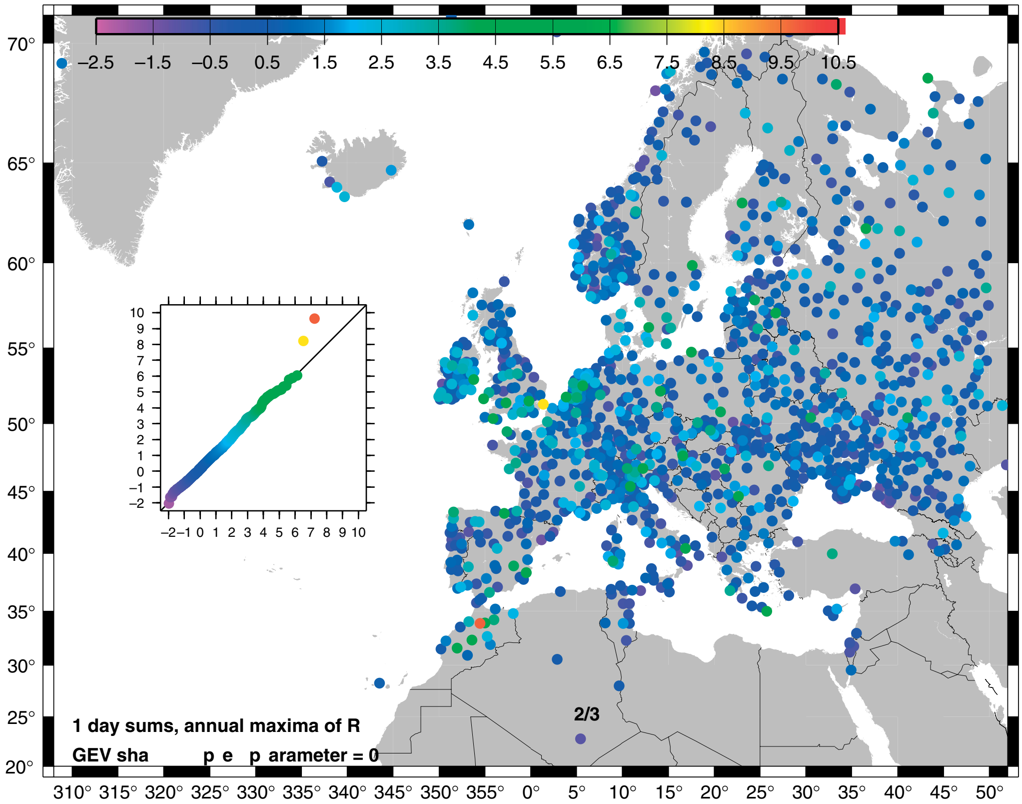

Figure 3. Gumbel plot of the statistical distribution of $\Delta\hat{X}_n$ and the spatial distribution of $\Delta\hat{X}_n$ for the annual maximum 1-day precipitation sums R, obtained by fitting a Gumbel distribution to $R^{2/3}$ to the 2147 records of the ECA-D dataset (1951-2008). The colour coding in plot and map indicate the magnitude of $\Delta\hat{X}_n$ and relate the points on the graph with their positions on the map. Only the 1379 independent values of $\Delta\hat{X}_n$ (out of 2147) are shown.

return period of 2.8×10^6 years) showed after detailed inspection to be erroneous.

Our conclusion, above, that the (transformed) precipitation is Gumbel distributed, seems to be contradictory to many papers that state that precipitation is heavy tailed (e.g. Koutsoyiannis 2004b, and references therein), i.e. is distributed according the GEV distribution with its shape parameter $\theta > 0$:

$$\mathrm{GEV}(y) = \exp\left\{-\left[1+\theta\left(\frac{y-\mu}{\sigma}\right)\right]^{-1/\theta}\right\} \qquad (9)$$

The answer can be found in the work of Furrer and Katz (2008), who give an expression for the GEV shape parameter if the GEV distribution is used as a pen-ultimate distribution for normalized maxima from the Weibull distribution:

$$\theta \approx \frac{1-k}{k\ln(i)} \qquad (10)$$

with k the Weibull shape parameter, and i the number of independent events in a year or season.

To compare both approaches, we also fitted a GEV distribution to R itself, assuming the GEV shape parameter to be constant over Europe, and estimated its value by iteratively fitting a GEV distribution to all 88 470 annual maxima (normalized by the local estimates of the location and scale GEV parameters, see Buishand 1991, Appendix A for details). This yields a value for θ of 0.1008, in good agreement with literature (e.g. Gellens 2002) and consistent with the value according to Equation (10) for $k = 2/3$ and $i = 150$.

Figure 4 shows that the distribution of $\Delta\hat{X}_n$ obtained by fitting a GEV distribution to R with a constant shape parameter of $\theta = 0.1008$ to the 2147 records, is as good as in the case that a Gumbel distribution is fitted to $R^{2/3}$ (Figure 3). This empirically proves that the GEV distribution with Equation (10) can indeed (Furrer and Katz 2008) serve as a very good pen-ultimate distribution for the Gumbel distribution.

Figure 5 shows the Gumbel plot for the 1951-1998 record of Mont-Aigoua, France (44.1 °N, 3.583 °E), which is the record with the highest absolute precipitation amount in the ECA-D dataset, namely 520 mm on 24 February 1964.

The fits are the Gumbel distribution to $R^{2/3}$, the GEV distribution to R with the shape parameter fixed to $\theta = 0.1008$ and the GEV distribution to R. The values for $\Delta\hat{X}_n$ of 1.05, 1.94 and 2.15, respectively, indicate a return period of the largest event of 137, 335 and 413 years, depending on which of the three respective fits is adopted. The vertical axis is linear in $R^{2/3}$, which transforms the GEV fit with shape parameter 0.1008 almost into a straight line. The GEV fit with a fixed shape parameter and the Gumbel fit are almost similar,

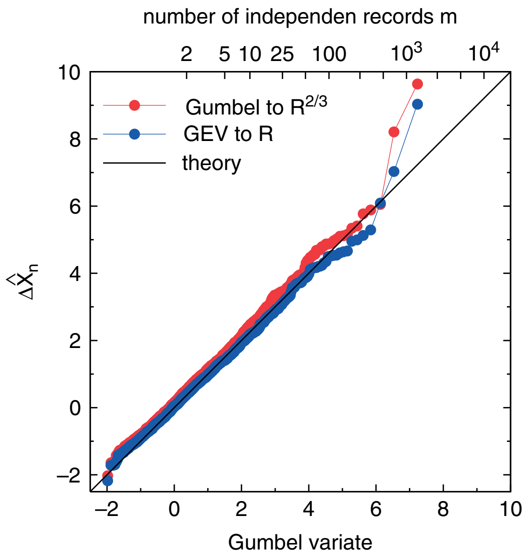

Figure 4. Comparison of two Gumbel plots of the statistical distribution of the 1379 independent values of $\Delta\hat{X}_n$ over Europe for the annual maximum 1-day precipitation sums R. Red: as obtained when a Gumbel distribution is fitted to $R^{2/3}$ (identical to the Gumbel plot in Figure 3). Blue: as obtained by fitting a GEV distribution with constant shape parameter $\theta = 0.1008$ to R.

both in the estimates and in the uncertainty ranges. They outperform the GEV fit with a free shape parameter, which is too sensitive to the outlier.

In most practical cases fitting a Gumbel distribution to $R^{2/3}$ will thus be preferred over fitting a GEV with fixed shape parameter to R, as the more symmetrical confidence intervals of the former leads to better estimations of the upper confidence interval of extrapolated return values.

4.4. Extreme precipitation in the Netherlands

We applied the GEV distribution with fixed shape parameter to 1-day precipitation maxima to the four seasons separately, using 294 station records in the Netherlands with a length of at least 20 years. The total sets contains 16 524 years and covers the period 1906-2007. The records can be downloaded from http://www.knmi.nl/klimatologie/monv/reeksen/. The GEV shape parameter is fixed and is like before empirically determined.

To ensure independence in the set of calculated values of $\Delta\hat{X}_n$, we required each pair of outliers to be separated by 2 or more days.

Figure 6 shows the following features:

- The JJA 1-day precipitation maxima are well described by a GEV distribution with a fixed shape parameter of 0.109. This value is almost the same as the value derived from the ECA-D dataset and is in accordance with the $k = 2/3$ Weibull shape parameter predicted by Wilson and Toumi (2005). The distribution of the 119 independent values of $\Delta\hat{X}_n$ representing a total of 6672 JJA maxima, closely coincides with the diagonal of the Gumbel-Gumbel plot (Equation (6b)).
- The fixed GEV shape parameter for DJF is almost zero. This deviates considerably from the value of 0.1 expected for a Weibull shape parameter $k = 2/3$.

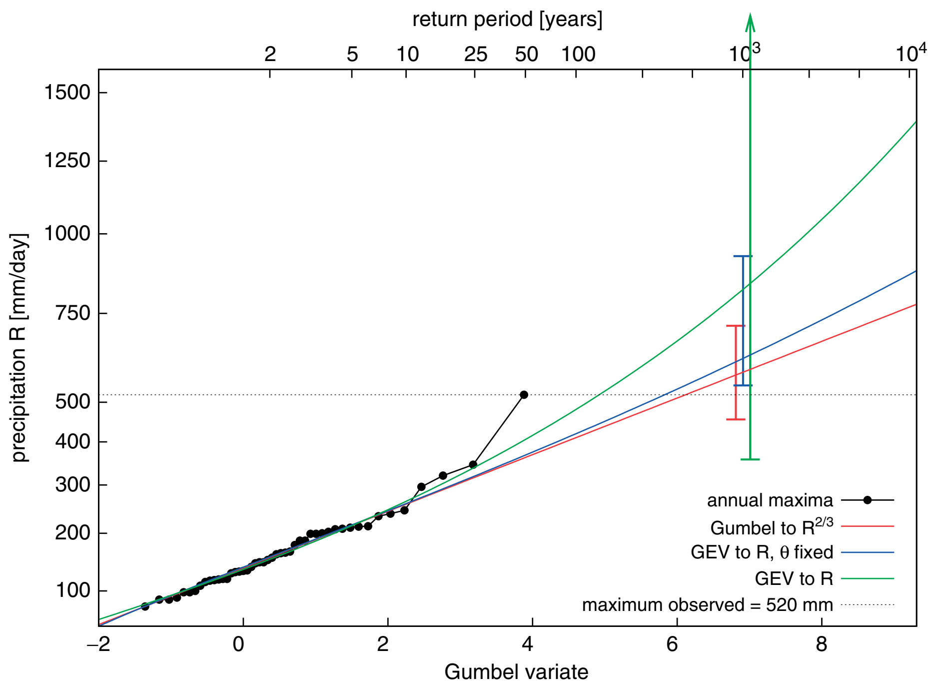

Figure 5. Gumbel plot for the annual maximum 1-day precipitation sums R in Mont-Aigoua, France (1951-1998). The fits are the Gumbel distribution to $R^{2/3}$ (red), the GEV distribution with $\theta = 0.1008$ to R (blue), and the GEV distribution with free shape parameter (green). The vertical axis is linear in $R^{2/3}$. The largest observed 1-day amount (520 mm) is indicated by a dashed horizontal line. The horizontal distances of the maximum observed value to the three fits (the magnitude of $\Delta\hat{X}_n$) of 1.05, 1.94 and 2.15 indicate return periods of this event of 137, 335 and 413 years, respectively. The vertical bars indicate the 95% confidence intervals. The green bar runs till R = 2361 mm/day.

DOI: 10.1002/joc

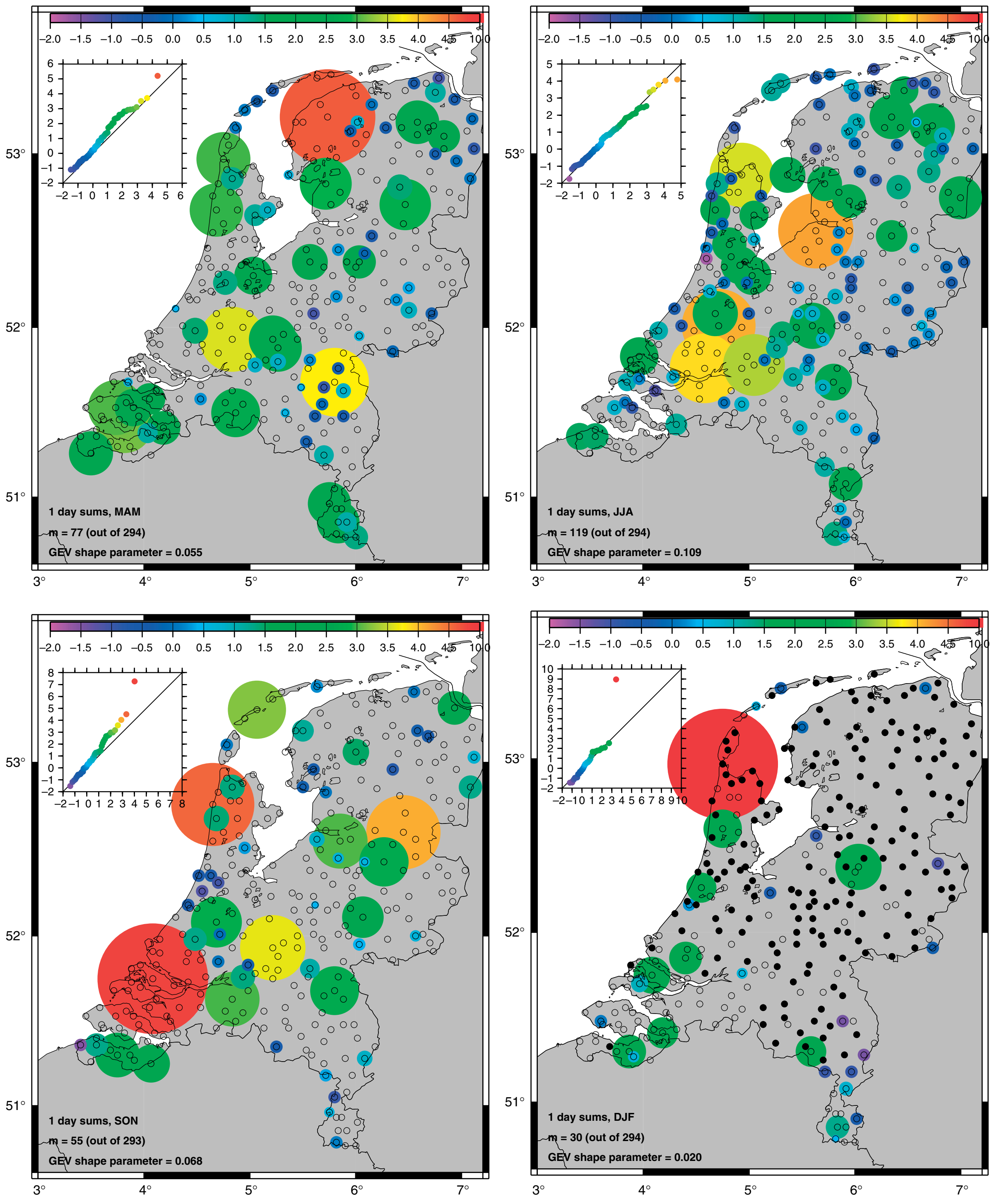

Figure 6. Gumbel plot of the statistical distribution of $\Delta\hat{X}_n$ and the spatial distribution of $\Delta\hat{X}_n$ for the 1-day precipitation sums for the Netherlands in MAM, JJA, SON and DJF (1906-2007). The colour coding in plot and map indicate the magnitude of $\Delta\hat{X}_n$ and relate the points on the graph with their positions on the map. The independent values of $\Delta\hat{X}_n$ are shown in colour; the dependent values are indicated by the open circles. The number of independent values m is indicated for each season. The total number of stations is 294; the closed circles in the DJF graph are the 178 stations where the event on 4 December 1960 caused the outlier.

This suggest that the conditions behind the Wilson and Toumi (2005) formula are violated. We speculate that this failure is related to the absence of deep convection in winter, combined with the lack of orographic forcing in this flat country.

- The DJF maxima are much more spatially correlated than the JJA maxima, with only 30 independent values of $\Delta\hat{X}_n$, representing 1562 years. This reduction can be attributed to the fact the DJF maxima are due to large-scale precipitation, with a larger spatial correlation than the convective precipitation that causes the JJA maxima. The precipitation on 4 December 1960 caused the record extreme for 178 stations (indicated

DOI: 10.1002/joc

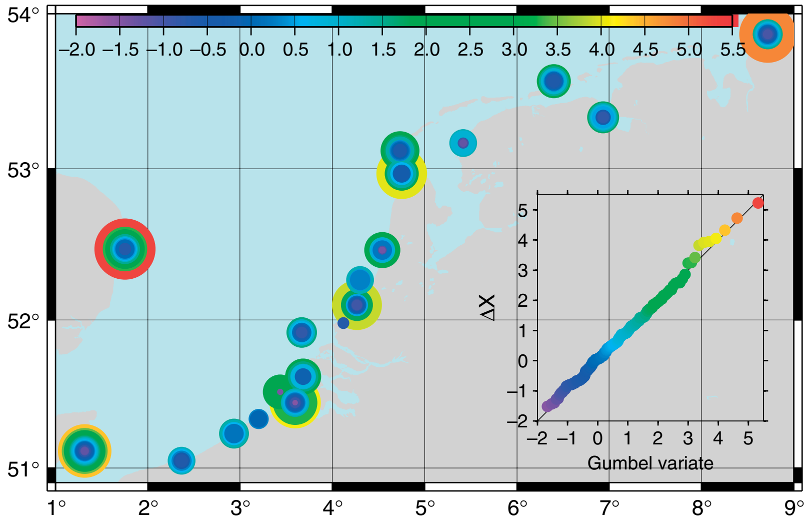

Figure 7. Gumbel plot of the statistical distribution of $\Delta\hat{X}_n$ and the spatial distribution of extreme sea levels in the southern North Sea, generated by the WAQUA surge model coupled to the ECHAM5/MPI-OM GCM. The colour coding in plot and map indicate the magnitude of $\Delta\hat{X}_n$, and relate the points on the graph to their positions on the map. The total number of stations is 969; only the 204 independent values of $\Delta\hat{X}_n$ are shown.

by the filled black circles in Figure 6), with a maximum amount of 83.9 mm in Joure (5.82 °E, 52.98 °N). This event also caused the highest $\Delta\hat{X}_n$ value of 8.95 in Den Helder (4.75 °E, 52.97 °N) with 83.3 mm. Its estimated return period of 5×10^5 years is very unlikely to happen in a 1562-year period, and hints on underestimation of the DJF extremes by a GEV distribution with fixed shape parameter and hence on rejection of the Wilson and Toumi (2005) hypothesis. Evaluation of the DJF 1-day precipitation extremes in the ECA-D dataset (1379 independent values, figure not shown) leads to the same conclusion; due to the larger dataset the deviation of the most extreme outliers from the diagonal in the Gumbel-Gumbel plot is more pronounced. The outlier in the plot in Figure 6 suggests that our method is able to detect this bias in a dataset consisting of as few as 30 independent values.

- The independent values of $\Delta\hat{X}_n$ in DJF are clustered in the coastal zone. The coastal clustering can be attributed to the relatively warm sea and the cold land, resulting in nocturnal convection over sea which occasionally affects the coastal zone.
- The MAM and SON seasons show a mix of summer- and winter features, which is apparent in values of the GEV shape parameter that are in between the summer and winter values. It is also be read from the Gumbel plots of $\Delta\hat{X}_n$ as the dots in them less perfectly follow the diagonal, particularly in SON.

4.5. Extreme sea levels

In the ESSENCE project (Sterl *et al.* 2008) the ECHAM5/MPI-OM climate model (Jungclaus *et al.* 2006) has been used to simulate the climate from 1950 to 2100, assuming future greenhouse gas concentrations to follow the SRES A1b scenario (Nakicenovic *et al.* 2000). Winds and sea level pressures from 17 integrations are used to drive the WAQUA/DCSM98 surge model (de Vries 2000). This results in 17 independent records of 150 years for 19 stations. For details we refer to Sterl *et al.* (2009). Here we test the assumption whether the GEV shape parameter can be taken to be constant for the 19 available stations. We found an optimal value of −0.005, i.e. practically equal to zero, which means that a Gumbel distribution performs best for extrapolation purposes. We divided the 17 records in subsets of 50 years, resulting in $3 \times 17 \times 19 = 969$ values of $\Delta\hat{X}_n$. We required a minimum time interval between independent events of 3 days, resulting in 204 independent values of $\Delta\hat{X}_n$, resembling a dataset of 10 200 years. Figure 7 shows the distribution of $\Delta\hat{X}_n$, as well as the locations of the outliers. It shows that fitting a Gumbel distribution results in very good agreement with theory, i.e. in unbiased estimates up to return periods of 10^4 years.

We checked this result with observations from 30 stations along the Dutch coast, with a total length of 1832 years, running within the period 1851-2007. In order to reduce sampling effects in the distribution of $\Delta\hat{X}_n$, long records were split into sub-series of at least 20 years, leading to 78 (sub)series. The distribution of $\Delta\hat{X}_n$ is given in Figure 8, which shows that the 19 independent values agree very well with theory, confirming the result of Figure 7 to return periods of at least 400 years.

The Gumbel plot of the 1896-2005 record of observed water levels at Scheveningen is shown in Figure 9. This record gives the highest value of $\Delta\hat{X}_n$. Although the GEV and Gumbel fits are almost similar, the application of the Gumbel distribution results in a considerable reduction of the uncertainty for high return periods. The Gumbel fit gives a 17 cm higher 10^4-year return level, but has a four times smaller 95% confidence interval than the GEV fit. The estimated return period of the 1953-disaster (the

DOI: 10.1002/joc

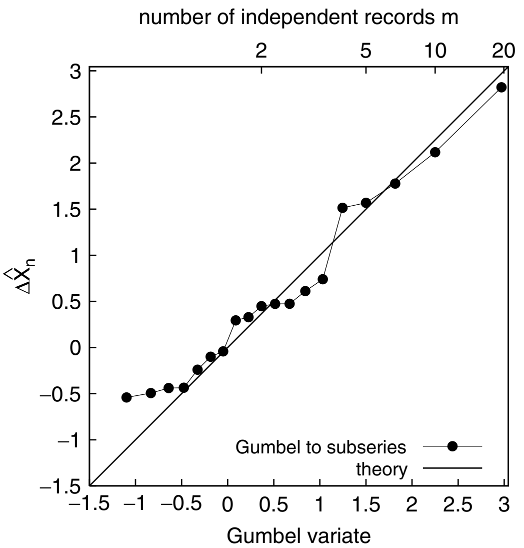

Figure 8. Gumbel plot of the statistical distribution of $\Delta\hat{X}_n$ of extreme sea levels along the Dutch coast. The period covered is 1851-2007; the 30 available records are split into 78 sub-series. Only the 19 independent values of $\Delta\hat{X}_n$ are plotted.

highest observation in the record) is 455 years ($\Delta\hat{X}_n = 1.47$, probability of 23% to occur in a 105-year period) according to the Gumbel fit, and 602 years ($\Delta\hat{X}_n = 1.75$, probability of 17%) according to the GEV fit.

5. Discussion and Conclusions

We have developed and explored a new tool for extreme value analysis, i.e. the evaluation of the distribution of the probability of outliers ΔX_n in observational series. It is proved that this distribution should obey the standardized Gumbel distribution. The absence of adjustable parameters in the ΔX_n distribution causes empirical analysis of this distribution to be a powerful instrument for verifying whether an adopted distribution to the (annual) extremes of a series is justified. The availability of this tool enables the calculation of large return values with a higher accuracy than hitherto seen.

We tested in five examples the potential of this tool on a variety of meteorological-related problems, which encompasses extremes of precipitation, wind speed, waves and surges. In all cases, it is found that the method is able to reject or accept the choices of the adopted (extreme value) distributions or to minimize the number of free parameters in them. Additionally to that, the method also turns out to be able to identify special features, like coastal effects or second populations in the distribution of extreme events.

The method can equally well be applied to results from more advanced statistical techniques (for instance to take dependence and/or non-stationarity into account) than the currently used classical extreme value theory, as the method only uses the 'distance' between the highest observation and its estimated return period for validation. The (more or less advanced) way how this estimate is obtained is not relevant in our method.

An interesting application might be to use several (advanced) techniques to minimize the statistical uncertainty in estimates, and then to use the current method to determine possible systematic biases. In this way, an optimal balance can be found between the systematic bias and the statistical uncertainty in the estimates of large return values from short observational time series.

The results of this paper lead us to believe that the method offers an effective and very robust instrument for extreme value analysis that is applicable to a much wider range of problems than yet explored. These may include risk analysis due to non-meteorological hazards such as earthquakes and maybe even to problems in fields outside geophysics. With this in mind, we believe that the present analysis represents only a first step in exploring the full

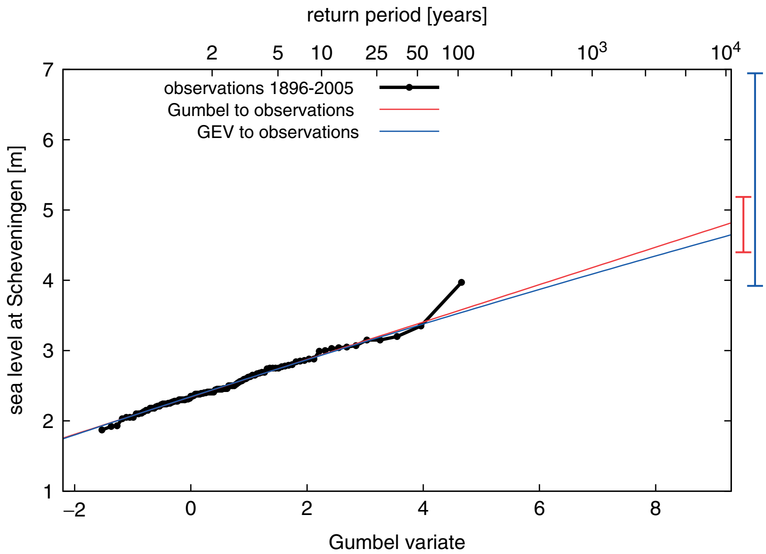

Figure 9. Gumbel plot of the annual maxima of the sea level at Scheveningen (1896-2005), the Netherlands. Shown are the fits according to a Gumbel (red) and a GEV distribution (blue). The vertical bars indicate the 95% confidence intervals.

DOI: 10.1002/joc

potential of this kind of method in extrapolating, selecting and empirically verifying the shape of the distributions of extreme events and for investigating the properties in its far tail.

Acknowledgements

The ESSENCE project, lead by Wilco Hazeleger (KNMI) and Henk Dijkstra (UU/IMAU), was carried out with support of DEISA, HLRS, SARA and NCF (through NCF projects NRG-2006.06, CAVE-06-023 and SG-06-267). We thank the DEISA Consortium (co-funded by the EU, FP6 projects 508830/031513) for support within the DEISA Extreme Computing Initiative (www.deisa.org). The authors thank Andreas Sterl (KNMI), Camiel Severijns (KNMI), and HLRS and SARA staff for technical support.

6. Appendix

The derivation in Section 2 uses the return period T as the basic variable, while Van den Brink and Können (2008) use the highest observed value y_n in a n-year record as starting point. The first evaluates ΔX_T, i.e. the probability that a T-year events happens in n years, whereas the second considers ΔX_n, i.e. the return period T for the highest event y_n in n years. Here we show the equivalence of both approaches.

In order to transform Equation (6) into Equation (11) of Van den Brink and Können (2008), we use that $F(y)$ and $-\ln(-\ln(F))$ are monotonic functions:

$$\Pr(y_n \leq y_T) = G(\Delta X_T) \tag{A.1}$$

$$\Pr(F(y_n) \leq F(y_T)) = G(\Delta X_T) \tag{A.2}$$

$$\Pr(-\ln(-\ln(F(y_n))) \leq -\ln(-\ln(F(y_n))) = G(\Delta X_T) \tag{A.3}$$

Subtracting $\ln(n)$ on both sides of the sign $\leq$ gives:

$$\Pr(-\ln(-\ln(F(y_n))) - \ln(n) \leq -\ln(-\ln(F(y_n))) - \ln(n)) = G(\Delta X_T) \tag{A.4}$$

Using the definitions of ΔX_n (Equation (11) of Van den Brink and Können 2008) and ΔX_T (Equation (6) of the present paper):

$$\Delta X_n = -\ln(-\ln(F(y_n))) - \ln(n) \tag{A.5}$$

$$\Delta X_T = -\ln(-\ln(F(y_T))) - \ln(n) \tag{A.6}$$

gives:

$$\Pr(\Delta X_n \leq \Delta X_T) = G(\Delta X_T) \tag{A.7}$$

Substituting x for ΔX_T gives:

$$\Pr(\Delta X_n \leq x) = G(x) \tag{A.8}$$

which shows that Equation (6) of the present paper and Equation (11) of Van den Brink and Können (2008) are equivalent.

The probability that a T-year events happens in n years (i.e. the ΔX_T formulation) is commonly used in risk analysis, whereas the return period of the outlier (i.e. the ΔX_n formulation) is easier to apply in practical situations.

References

Bouws E, Draper L, Shearman EDR, Laing AK, Feit D, Mass W, Eide LI, Francis P, Carter DJT, Battjes JA. 1998. *Guide to Wave Analysis and Forecasting*, 2nd edn, WMO - No.702. World Meteorological Organization: Geneva; Switzerland.

Buishand TA. 1991. Extreme rainfall estimation by combining data from several sites. *Hydrological Scienes Journal* **36**: 345–365.

Caires S, Sterl A. 2005. 100-year return value estimates for ocean wind speed and significant wave height from the era-40 data. *Journal of Climate* **18**: 1032–1048.

Cook NJ. 1982. Towards better estimation of extreme winds. *Journal of Wind Engineering and Industrial Aerodynamics* **9**: 295–323.

De Vries JW. 2000. 'A concise evaluation of water level forecasts from WAQUA/CSM-8', Memorandum WM–00–01 KNMI http://www.knmi.nl/jwdv/WAQUA/evalCSM8/index.html.

Furrer EM, Katz RW. 2008. 'Improving the simulation of extreme precipitation events by stochastic weather generators'. *Water Resources Research* **44**: W12439.

Gellens D. 2002. Combining regional approach and data extension procedure for assessing gev distribution of extreme precipitation in Belgium. *Journal of Hydrology* **268**: 113–126.

Hosking JRM, Wallis JR. 1997. *Regional Frequency Analysis: An Approach Based on L-moments*. Cambridge University Press: Cambridge; UK.

Jungclaus J, Keenlyside N, Botzet M, Haak H, Luo JJ, Latif M, Marotzke J, Mikolajewicz U, Roeckner E. 2006. Ocean circulation and tropical variability in the coupled model ECHAM5/MPI-OM. *Journal of Climate* **19**: 3952–3972.

Klein Tank AMG, Wijngaard JB, Können GP, Böhm R, Demarée G, Gocheva A, Mileta M, Pashiardis S, Hejkrlik L, Kern-Hansen C, Heino R, Bessemoulin P, Müller-Westermeier G, Tzanakou M, Szalai S, Pálsdóttir T, Fitzgerald D, Rubin S, Capaldo M, Maugeri M, Leitass A, Bukantis A, Aberfeld R, van Engelen AFV, Forland E, Mietus M, Coelho F, Mares C, Razuvaev V, Nieplova E, Cegnar T, López JA, Dahlström B, Moberg A, Kirchhofer W, Ceylan A, Pachaliuk O, Alexander LV, Petrovic P. 2002. Daily dataset of 20th-century surface air temperature and precipitation series for the European Climate Assessment. *International Journal of Climatology* **22**: 1441–1453.

Koutsoyiannis D. 2004a. Statistics of extremes and estimation of extreme rainfall: I. Theoretical investigation. *Hydrological Sciences* **49**: 575–590.

Koutsoyiannis D. 2004b. Statistics of extremes and estimation of extreme rainfall: II. Empirical investigation of long rainfall records. *Hydrological Sciences* **49**: 591–610.

Nakicenovic N, Alcamo J, Gruebler A, Riahi K, Roehrl RA, Rogner H-H, Victor N. 2000. Special Report on Emissions Scenarios (SRES), A Special Report of Working Group III of the Intergovernmental Panel on Climate Change. Cambridge University Press: Cambridge, UK.

Sterl A, Severijns C, Dijkstra H, Hazeleger W, van Oldenborgh GJ, van den Broeke M, Burgers G, van den Hurk B, van Leeuwen PJ, van Velthoven P. 2008. When can we expect extremely high surface temperatures? *Geophysical Research Letters* **35**: L14703.

Sterl A, van den Brink H, de Vries H, Haarsma R, van Meijgaard E. 2009. An ensemble study of extreme North Sea storm surges in a changing climate. *Ocean Science* **5**: 369–378.

Uppala SM, Kallberg PW, Simmons AJ, Andrae U, da Costa Bechtold V, Fiorino M, Gibson JK, Haseler J, Hernandez A, Kelly GA, Li X, Onogi K, Saarinen S, Sokka N, Allan RP, Andersson E, Arpe K, Balmaseda MA, Beljaars ACM, van de Berg L, Bidlot J, Bormann N, Caires S, Chevallier F, Dethof A, Dragosavac M, Fisher M, Fuentes M, Hagemann S, Holm E, Hoskins BJ, Isaksen L, Janssen PAEM, Jenne R, McNally AP, Mahfouf J-F, Morcrette J-J, Rayner NA, Saunders RW, Simon P,

DOI: 10.1002/joc

Sterl A, Trenberth KE, Untch A, Vasiljevic D, Viterbo P, Woollen J. 2005. The ERA-40 re-analysis. *Quarterly Journal of Royal Meteorological Society* **131**: 2961–3012.

Van den Brink HW, Können GP. 2008. The statistical distribution of meteorological outliers. *Geophysical Research Letters* **35**: L23702.

Van den Brink HW, Können GP, Opsteegh JD. 2004a. Statistics of extreme synoptic-scale wind speeds in ensemble simulations of current and future climate. *Journal of Climate* **17**: 4564–4574.

Van den Brink HW, Können GP, Opsteegh JD, van Oldenborgh GJ, Burgers G. 2004b. Improving 10^4-year surge level estimates using data of the ECMWF seasonal prediction system. *Geophysical Research Letters* **31**: L17201.

Van den Brink HW, Können GP, Opsteegh JD, van Oldenborgh GJ, Burgers G. 2005. Estimating return periods of extreme events from ecmwf seasonal forecast ensembles. *International Journal of Climatology* **25**: 1345–1354.

Wang XL, Swail VR, Zwiers FW. 2006. Climatology and changes of extratropical cyclone activity: comparison of ERA-40 with NCEP-NCAR reanalysis for 1958–2001. *Journal of Climate* **19**: 3145–3166.

Wilson PW, Toumi R. 2005. A fundamental probability distribution for heavy rainfall. *Geophysical Research Letters* **32**: L14812.

DOI: 10.1002/joc