

The statistical distribution of meteorological outliers

H. W. van den Brink¹ and G. P. Können^{1,2}

Received 18 September 2008; revised 21 October 2008; accepted 30 October 2008; published 5 December 2008.

[1] A general expression for the statistical distribution of the probability of the highest event occurring in a record is presented. This result can be empirically applied to situations where records are available for multiple geographical locations. The empirical estimation of the probability of the highest events provides a means to assess whether the assumed (extreme value) distribution is appropriate for extrapolation or not. The approach allows for combining the highest events from different records and to validate estimated return periods in the order of the length of the combined records. The method is illustrated with an analysis of the annual extreme wind speeds over the North Atlantic area according to the ERA40 dataset, showing that the Gumbel distribution is in favor of the GEV distribution to estimate the (appropriately transformed) extreme wind speeds up to return periods of 10^4 years. **Citation:** van den Brink, H. W., and G. P. Können (2008), The statistical distribution of meteorological outliers, *Geophys. Res. Lett.*, 35, L23702, doi:10.1029/2008GL035967.

1. Introduction

[2] In the perspective of safety protection, much scientific attention is given to the estimation of hydrological and climatological extreme events [Brabson and Palutikof, 2000; van den Brink et al., 2004b].

[3] From a theoretical point of view, several distributions are possible [e.g., Coles, 2001]. However, although these distributions often fulfill the standard goodness of fit criteria, they may differ considerably in their estimates for large return periods. This makes it hard to verify which distribution correctly describes the extremes for return periods exceeding the length of the observational record.

[4] A frequent situation is that multiple time series of a certain meteorological variable are available (e.g., for different stations) for which one can assume the same distribution, albeit with different parameters. A common approach is then to combine the (normalized) data of multiple stations [e.g., Buishand, 1991; Hosking and Wallis, 1997], but this requires spatial homogeneity over the whole area.

[5] Here we present a robust diagnostic tool able to detect whether the most extreme events of multiple records are well described by the fitted distributions or not. The tool detects how strongly the highest extremes in records (the “outliers”) are over- or underestimated. The only conditions to get the detection tool to work are that enough

records are available, and that the most extreme events of different records are distinct.

2. Methodology

[6] We consider the probability integral transform theorem [Folland and Anderson, 2002]:

$$\Pr(F(y) \leq x) = x \quad 0 \leq x \leq 1 \quad (1)$$

with F the cumulative distribution function of a (meteorological) variable y , from which follows:

$$\Pr(F(y) \leq G(x)) = G(x) \quad 0 \leq G(x) \leq 1 \quad (2)$$

$$\Pr(G^{-1}(F(y)) \leq x) = G(x) \quad (3)$$

Equation (1) states that $F(y)$ is uniformly distributed between 0 and 1 for every cumulative function F , whereas equation (3) states that the inverse function of G applied to $F(y)$ is distributed according to G .

[7] Let y_n be the maximum of n independent observations $y_1 \leq y_2 \leq \dots \leq y_n$ from distribution $F(y)$, then the distribution of y_n is given by:

$$\begin{aligned} F(y_n) &= \Pr(y_n \leq y) = \Pr(y_1 \leq y) \cdots \Pr(y_n \leq y) \\ &= F^n(y) \end{aligned} \quad (4)$$

Using that G^{-1} is monotonically increasing, it follows:

$$\Pr(G^{-1}(F(y_n)) \leq x) = [\Pr(G^{-1}(F(y)) \leq x)]^n = [G(x)]^n \quad (5)$$

We now choose $G(x)$ to be a distribution that satisfies the max-stable property [e.g., de Haan and Ferreira, 2006, p. 9]:

$$[G(x)]^n = G(\alpha_n x + \beta_n). \quad (6)$$

with α_n and β_n appropriate normalizing constants, and select the simplest among them, which is the normalized Gumbel distribution:

$$G(x) = e^{-e^{-x}}. \quad (7)$$

With this choice, equation (6) reduces to:

$$[G(x)]^n = G(x - \ln(n)). \quad (8)$$

Incorporation in equation (5) gives:

$$\Pr(-\ln(-\ln(F(y_n))) \leq x) = G(x - \ln(n)) \quad (9)$$

¹Royal Netherlands Meteorological Institute, De Bilt, Netherlands.

²Now at Soest, Netherlands.

If we define ΔX_n as:

$$\Delta X_n = -\ln(-\ln(F(y_n))) - \ln(n) \quad (10)$$

then we have for the distribution of ΔX_n :

$$\Pr(\Delta X_n \leq x) = G(x) \quad (11)$$

[8] Note that equation (11) does neither depend on the underlying distribution F nor on the parameters of F , nor on the sample size n . This implies that the distribution of ΔX_n can be constructed by combining (different or equal) meteorological variables from different or equal parent distributions, with correlated or uncorrelated parameters.

[9] Equation (11) can be visualized on a so-called Gumbel plot, where the abscissa is transformed into the Gumbel variate:

$$x_r = -\ln(-\ln(\hat{F}(y_r))) = -\ln\left(-\ln\left(\frac{r}{n+1}\right)\right) \quad (12)$$

where $\hat{F}(y_r)$ is the expected value of $F(y_r)$ [Benard and Bos-Levenbach, 1953]. From equation (12) follows for $r = n$:

$$x_n = -\ln\left(-\ln\left(\frac{n}{n+1}\right)\right) \simeq \ln(n) \quad (13)$$

The approximation is useful for the graphical illustration of the concept, as the highest event y_n is plotted at a plotting position $x_n \simeq \ln(n)$. Hence the horizontal ‘distance’ ΔX_n between x_n and the distribution (with abscissa $-\ln(-\ln(F(y_n)))$) is Gumbel distributed with location parameter 0 and scale parameter 1. The concept is schematically illustrated in Figure 1.

[10] If m records with independent outliers are used of (not necessarily equal) lengths T_i , $1 \leq i \leq m$, the distribution of the outliers can be tested up to a maximum return period T_{\max} :

$$T_{\max} \simeq \sum_{i=1}^m T_i \quad (14)$$

which implies that T_{\max} can be orders larger than the individual record lengths.

3. Verification

[11] We applied equation (11) to 10^4 computer-generated independent records of random length, from a randomly chosen distribution (Gumbel, Weibull, Gaussian or GEV) with randomly chosen parameters. The result is shown in Figure 2, in which $\Delta \hat{X}_n$ is the estimated value of ΔX_n . It shows an excellent agreement with theory, independent of the choice of the distribution, parameters or record length. Note that Figure 2 clearly shows that the transformation according to equation (11) fully determines the distribution of ΔX_n , with no degrees of freedoms left.

[12] Equation (11) holds for an exactly known distribution F . However, in practical applications, an *a priori* type of distribution is assumed, and then the parameters are to be estimated from the data. This means that F is replaced by $\tilde{F}_{\hat{\theta}}$, where \tilde{F} stands for the assumed distribution, and $\hat{\theta}$ for the vector of the parameters fitted to the data (using Maximum Likelihood estimates).

[13] A common choice for F to describe meteorological annual maxima is the Generalized Extreme Value (GEV) distribution [Jenkinson, 1955]:

$$F(y) = e^{-e^{-x}} \quad (15)$$

with the Gumbel variate x a substitute for:

$$x = \begin{cases} \ln(1 - \xi \frac{y - \mu}{\alpha})^{-1/\xi} & \xi \neq 0 \\ \frac{y - \mu}{\alpha} & \xi = 0 \end{cases} \quad (16)$$

in which μ is the location parameter, α the scale parameter, ξ the shape parameter, and y the annual maxima of the considered variable. The Gumbel distribution is the special case with $\xi = 0$.

[14] A well-known problem with goodness of fit tests is the influence of finite n and of the unknown parameters on the test criteria [Laio, 2004]. In our case, both the sampling effect and the effect of y_n on $\hat{\theta}$ result in a standard deviation of $\Delta \hat{X}_n$ that is too small. The effect of the sample length n on the distribution of $\Delta \hat{X}_n$ is illustrated in Figure 3 for several values of n . Shown is the case that the original distribution F is a Gumbel distribution and the fitted distribution \tilde{F} is either a Gumbel- or a GEV-distribution. It shows that if the fitted distribution \tilde{F} is a Gumbel distribution, the distribution of $\Delta \hat{X}_n$ is near the theoretical line if the record lengths $n \geq 40$.

[15] However, the presence of an extra parameter (ξ) in the GEV distribution leads to a severe underestimation of the larger values of ΔX_n even if the record length $n = 300$. So, although the extra parameter of the GEV distribution may better fit to the observations than the Gumbel distribution, this improvement holds only for return periods within the record length—for return periods exceeding the record length the results of the GEV fit are worse. This makes the GEV distribution inappropriate for estimates of return periods that exceed the record length. The same conclusion can be drawn from situations in which the parent distribution F is a GEV distribution, or that L-moments estimates [Hosking et al., 1985] are used [Landwehr et al., 1979; Kochanek et al., 2008] (see auxiliary material).¹

[16] A possible solution may be to make assumptions about the shape parameter, e.g., to be constant over a certain area [Buishand, 1991].

[17] A single extreme meteorological event may determine the value of y_n (and thus of $\Delta \hat{X}_n$) for multiple records. To remove this spatial and temporal correlation, only those values of $\Delta \hat{X}_n$ should be considered that belong to distinct meteorological events. For a given extreme event, that record should be considered for which the event is most exceptional, i.e., for which $\Delta \hat{X}_n$ is maximal. Note that this condition is much weaker than those required for Regional Frequency Analysis [e.g., Reed et al., 1999].

4. Interpretation of the Empirical $\Delta \hat{X}_n$ Distribution

[18] The following situations may occur:

[19] 1. If all m (temporally independent) points in a $\Delta \hat{X}_n$ -plot follow the theoretical line, the chosen distribution \tilde{F}

¹Auxiliary materials are available in the HTML. doi:10.1029/2008GL035967.

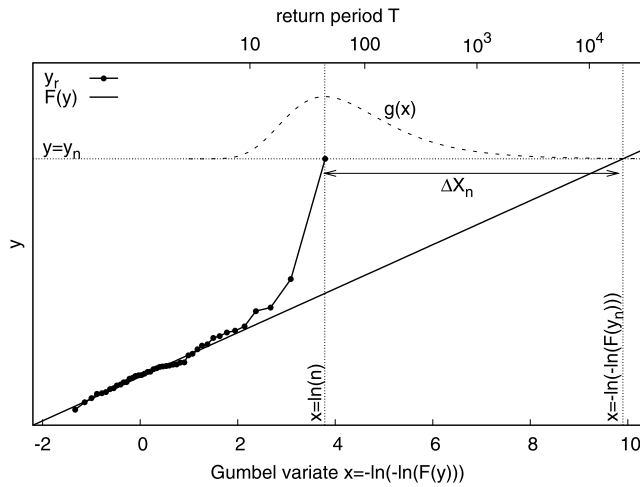


Figure 1. Visualization of the concept of equation (11) on a Gumbel plot. The probability of the most extreme event in the record (the outlier) is Gumbel-distributed, which is depicted by its probability density function g . In this example the number of extremes $n = 44$.

describes the extremes properly up to a return period corresponding to the length of the combined records (equation (14)).

[20] 2. If all m points in a $\Delta\hat{X}_n$ -plot are below the theoretical line, the outliers are lower than \hat{F} implies. This means that \hat{F} overestimates the probability of the high extremes. In case of F being an extreme value distribution, this points to incomplete convergence of the normalized annual maxima to the extreme value distribution.

[21] 3. If all m points in a $\Delta\hat{X}_n$ -plot are above the theoretical line, the outliers are higher than \hat{F} implies. This means that \hat{F} underestimates the probability of the high extremes. In case of F being an extreme value distribution, this points to incomplete convergence, or to the presence of

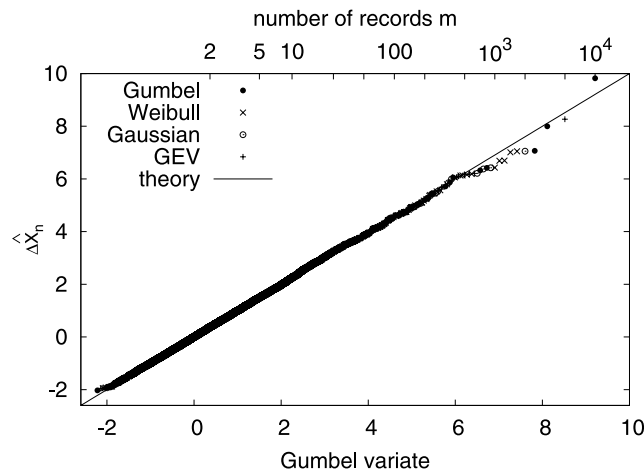


Figure 2. Gumbel plot for $\Delta\hat{X}_n$ for 10^4 records of random length, from a randomly chosen distribution with randomly chosen parameters. The theoretical distribution of ΔX_n is the normalized Gumbel distribution (solid line). The abscissa represents the Gumbel-transformed number of records m , as shown on the upper axis.

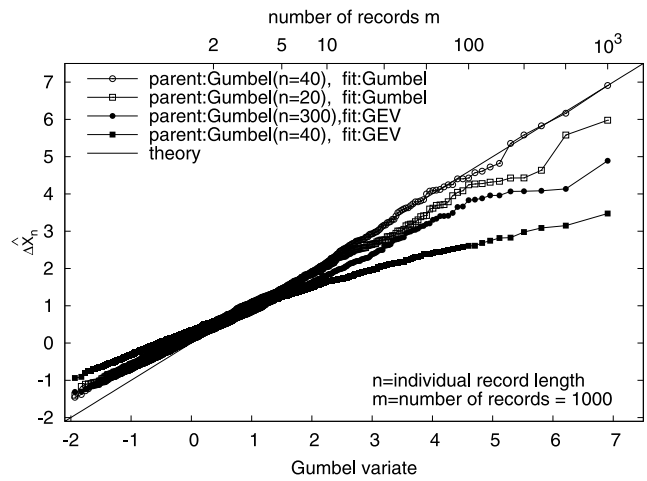


Figure 3. Gumbel plots of the distribution of $\Delta\hat{X}_n$ for fitting either a Gumbel- or a GEV distribution to $m = 1000$ records (each of length n) sampled from a Gumbel distribution. The distribution of ΔX_n is strongly underestimated if a GEV is fitted (even to records of 300 years each), but is well represented if a Gumbel distribution is fitted (even to records of only 40 years each).

a second population in the far tail of F [van den Brink et al., 2004a].

[22] 4. If the m points in a $\Delta\hat{X}_n$ -plot follow the theoretical line up to a certain Gumbel variate, and then start to be higher, then this means that \hat{F} describes only the more frequently occurring extremes well, but underestimates the probability of the less frequent extremes. This points to the presence of a second population in the far tail of F .

[23] 5. If all lower points in a $\Delta\hat{X}_n$ -plot are above the theoretical line, and all higher points are below that line, F has too many free parameters, and is inappropriate to describe the extremes for return periods exceeding the individual record length.

5. Example: Extreme Wind Speed in ERA40

[24] As an example of the method, we tested Cook's [1982] hypothesis that the normalized annual maxima of the wind speed u follow a Gumbel distribution, if not u itself, but u^k is the fitted variable, with k the shape parameter of the Weibull distribution:

$$F(u) = 1 - \exp\left\{-\left(u/a\right)^k\right\} \quad (17)$$

with a the scale parameter. He argues that the normalized block-maxima of the Weibull distribution converge to the Gumbel distribution for every value of k , but that the convergence speed strongly depends on k , being optimal for $k = 1$, which is equivalent to fitting u^k .

[25] We tested Cook's [1982] hypothesis by determining the distribution of $\Delta\hat{X}_n$ for the North-Atlantic region (59W-10E and 41N-70N) by fitting a GEV- and a Gumbel distribution to the annual maxima of both u and u^k . For u , we used the 44 annual maxima of the 10 m wind speed from the ERA40-dataset [Uppala et al., 2005] for the period

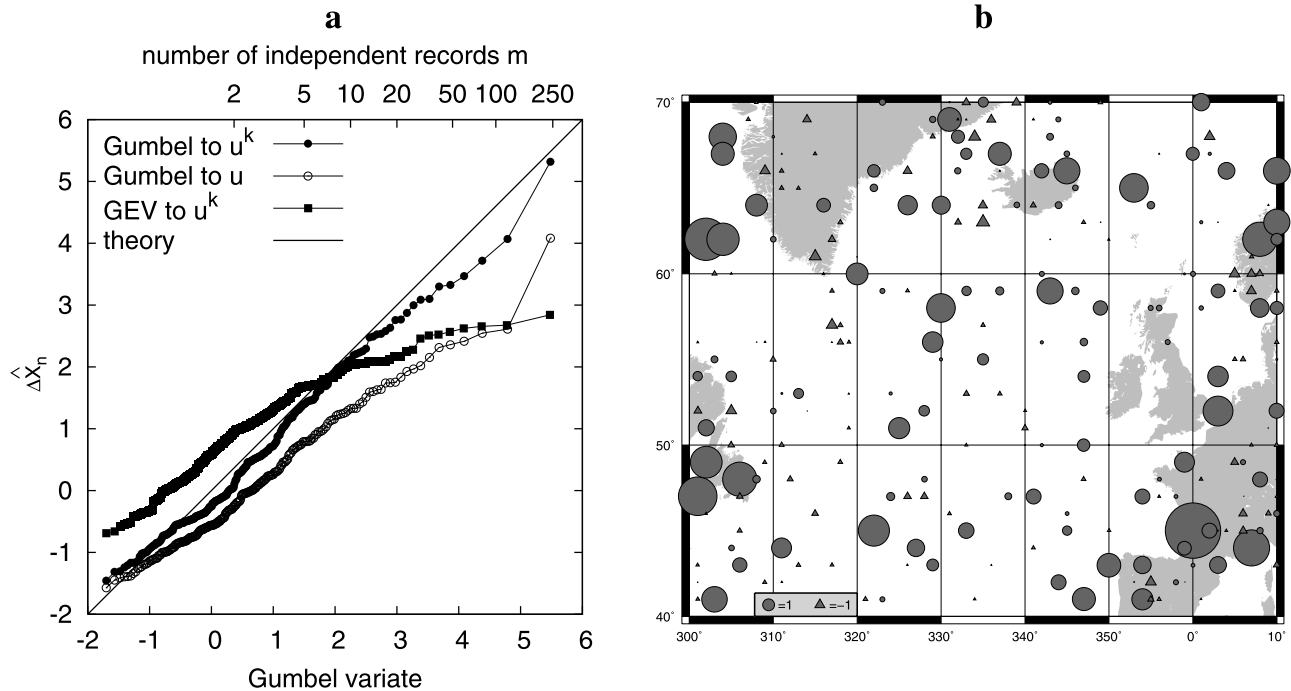


Figure 4. (a) Gumbel plots of the distribution of $\Delta\hat{X}_n$ for the annual maximum wind speed u for the North-Atlantic region. The fitted distribution \tilde{F} is either a GEV distribution to u^k , with k the Weibull shape parameter (squares), or a Gumbel distribution to u (open circles) or u^k (closed circles). The solid black line represents the theoretical normalized Gumbel distribution of $\Delta\hat{X}_n$. (b) Geographical locations of the 240 independent values of $\Delta\hat{X}_n$ when fitting a Gumbel distribution to u^k . The size of the symbols represent the value of $\Delta\hat{X}_n$, where triangles indicate negative values of $\Delta\hat{X}_n$.

1958–2001, interpolated to a spatial resolution of 1° , resulting in 2100 values for y_n .

[26] The Weibull shape parameter k was determined for every grid point from all 6-hourly wind speeds with $u > a$, with a the Weibull scale parameter (see auxiliary material).

[27] If an extratropical cyclone determines $\Delta\hat{X}_n$ for multiple grid points, we only use the maximum value of $\Delta\hat{X}_n$, i.e., we consider the cyclone only on its most exceptional moment. We assume that cyclones are distinct if the dates corresponding to $\Delta\hat{X}_n$ differ by more than 3 days.

[28] This selection procedure leads in this example to 240 independent values of $\Delta\hat{X}_n$. Equation (14) implies that it is likely that somewhere in the North-Atlantic area a 10^4 -year (240×44) event happened during the 1958–2001 period.

[29] Fitting a GEV distribution leads to a severely biased distribution of $\Delta\hat{X}_n$, both for u (not shown) and u^k (squares in Figure 4a). The maximum value of $\Delta\hat{X}_n$ of 2.84 corresponds to a return period of about 750 years (equation (10)), i.e., an underestimation by a factor 14. Fitting a Gumbel distribution to the annual maxima of u (open circles in Figure 4a) results in an underestimation of $\Delta\hat{X}_n$, indicating incomplete convergence to the Gumbel distribution. However, fitting a Gumbel distribution to the annual maxima of u^k (closed circles in Figure 4a) gives a satisfactory agreement with the theory, which confirms that the North-Atlantic annual extremes of u^k can be described by a Gumbel distribution up to return periods of 10^4 years (equation (14)). The same distribution for $\Delta\hat{X}_n$ is obtained for the winds on a $2.5^\circ \times 2.5^\circ$ resolution. We stress that those return periods are only valid for the realization of the

climate represented by the 1958–2001 period of the ERA40 dataset. Analyses for detrended series (as a first estimate of low-frequent variability) and of an apparently more homogeneous subset (1958–1990) lead to the same conclusion. However, inhomogeneities may be an issue in other parts of the world, especially on the Southern Hemisphere [Wang *et al.*, 2006].

[30] The 240 geographical locations of $\Delta\hat{X}_n$ for the Gumbel-fit to u^k are shown in Figure 4b. The event with the highest $\Delta\hat{X}_n$ of 5.32 (return period $9 \cdot 10^3$ years) is ‘Martin’, one of the Christmas-storms in France [Ulbrich *et al.*, 2001] at (0E,45N) on 27-12-1999.

[31] A second result of Figure 4a is that in this area, no second population of extreme wind speeds is detected.

6. Discussion and Conclusions

[32] We derived an expression for the distribution of the probability of the most extreme events in (meteorological) records. By combining the results for multiple records, the empirical distribution of the probability of record-extremes can be compared with the theoretical distribution of the outliers.

[33] In climatology, the Gumbel and GEV distribution are often applied to annual maxima. Here we showed that the fitted GEV distribution on average severely overestimates the probability of the record-extremes for most observational record lengths (<100 year). This is caused by the presence of an extra parameter in the GEV distribution, which makes the fit too sensitive for sampling effects, and thus inappropriate for extrapolation. Possible practical solutions are to

assume the GEV shape parameter to be constant over the area, or to fit a Gumbel distribution either to the variable y itself or to an appropriate power y^k .

[34] Whereas the standard goodness of fit tests evaluate the residuals *within* the observed time-domain, this tool tests if the fit is appropriate *outside* the observed time-domain. This is especially of importance for extreme-value statistics as it is often used to make estimates for return periods that amply exceed the record length.

References

- Benard, A., and E. C. Bos-Levenbach (1953), The plotting of observations on probability paper, *Stat. Neer.*, 7, 163–173.
- Brabson, B. B., and J. P. Palutikof (2000), Tests of the Generalized Pareto Distribution for predicting extreme wind speeds, *Meteorol. Appl.*, 39, 1627–1641.
- Buishand, T. A. (1991), Extreme rainfall estimation by combining data from several sites, *Hydrol. Sci. J.*, 36, 345–365.
- Coles, S. (2001), *An Introduction to Statistical Modelling of Extreme Values*, 208 pp., Springer, London.
- Cook, N. J. (1982), Towards better estimation of extreme winds, *J. Wind Eng. Ind. Aerodyn.*, 9, 295–323.
- de Haan, L., and A. Ferreira (2006), *Extreme Value Theory: An Introduction*, 417 pp., Springer, New York.
- Folland, C., and C. W. Anderson (2002), Estimating changing extremes using empirical ranking methods, *J. Clim.*, 15, 2954–2960.
- Hosking, J. R. M., and J. R. Wallis (1997), *Regional Frequency Analysis: An Approach Based on L-Moments*, Cambridge Univ. Press, Cambridge, U. K.
- Hosking, J. R. M., J. R. Wallis, and E. F. Wood (1985), Estimation of the Generalized Extreme-Value distribution by the method of probability-weighted moments, *Technometrics*, 27, 251–261.
- Jenkinson, A. F. (1955), The frequency distribution of the annual maximum (or minimum) values of meteorological elements, *Q. J. R. Meteorol. Soc.*, 81, 158–171.
- Kochanek, K., W. G. Strupczewski, V. P. Singh, and S. Weglarczyk (2008), The PWM large quantile estimates of heavy tailed distributions from samples deprived of their largest element, *Hydrol. Sci. J.*, 53, 367–386.
- Laio, F. (2004), Cramer-von Mises and Anderson-Darling goodness of fit tests for extreme value distributions with unknown parameters, *Water Resour. Res.*, 40, W09308, doi:10.1029/2004WR003204.
- Landwehr, J. M., N. C. Matalas, and J. R. Wallis (1979), Probability weighted moments compared with some traditional techniques in estimating gumbel parameters and quantiles, *Water Resour. Res.*, 15, 1055–1064.
- Reed, D. W., D. S. Faulkner, and E. J. Stewart (1999), The FORGEX method of rainfall growth estimation II: Description, *Hydrol. Earth Syst. Sci.*, 3, 197–203.
- Ulbrich, U., A. H. Fink, M. Klawka, and J. G. Pinto (2001), Three extreme storms over Europe in December 1999, *Weather*, 56, 70–80.
- Uppala, S. M., et al. (2005), The ERA-40 re-analysis, *Q. J. R. Meteorol. Soc.*, 131, 2961–3012.
- van den Brink, H. W., G. P. Können, and J. D. Opsteegh (2004a), Statistics of extreme synoptic-scale wind speeds in ensemble simulations of current and future climate, *J. Clim.*, 17, 4564–4574.
- van den Brink, H. W., G. P. Können, J. D. Opsteegh, G. J. van Oldenborgh, and G. Burgers (2004b), Improving 10^4 -year surge level estimates using data of the ECMWF seasonal prediction system, *Geophys. Res. Lett.*, 31, L17210, doi:10.1029/2004GL020610.
- Wang, X. L., V. R. Swail, and F. W. Zwiers (2006), Climatology and changes of extratropical cyclone activity: Comparison of ERA-40 with NCEP-NCAR reanalysis for 1958–2001, *J. Clim.*, 19, 3145–3166.

G. P. Können, Sophialaan 4, 3761 DK, Soest, Netherlands.

H. W. van den Brink, Royal Netherlands Meteorological Institute, P.O. Box 201, NL-3780 AE De Bilt, Netherlands. (brinkvdh@knmi.nl)