Statistics of Extreme Synoptic-Scale Wind Speeds in Ensemble Simulations of Current and Future Climate

H. W. VAN DEN BRINK, G. P. KÖNNEN, AND J. D. OPSTEEGH

Royal Netherlands Meteorological Institute, De Bilt, Netherlands

(Manuscript received 18 July 2003, in final form 7 June 2004)

ABSTRACT

Statistical analysis of the wind speeds, generated by a climate model of intermediate complexity, indicates the existence of areas where the extreme value distribution of extratropical winds is double populated, the second population becoming dominant for return periods of order 10^3 yr. Meteorological analysis of the second population shows that it is caused when extratropical cyclones merge in an extremely strong westerly jet stream such that conditions are generated that are favorable for occurrence of strong diabatic feedbacks. Doubling of the greenhouse gas concentrations changes the areas of second population and increases its frequency. If these model results apply to the real world, then in the exit areas of the jet stream the extreme wind speed with centennial-to-millennial return periods is considerably larger than extreme value analysis of observational records implies.

1. Introduction

Motivated by safety and dike design demands, much statistical research with observational data has been done on the estimation of extreme wind speeds and storm surges (Cook 1982; Simiu et al. 2001; de Haan 1990) for return periods up to 10^4 yr. Although a wide variety of methods has been developed (see, e.g., Palutikof et al. 1999 for an overview), all practical applications are hampered from the restricted length of the observational series (order 100 yr), on which the statistical extrapolations are based.

Several assumptions underly the statistical estimate of the wind speed with a return period of 10^4 yr. The most important one that all extratropical extremes (up to return periods of 10^4 yr) belong to the same population, is hard to verify from the available short observational sets.

We evaluated this problem within the context of a climate model of intermediate complexity. We have generated ensemble runs (consisting of 3509 yr in total) with a climate model, both for the current climate (~1975) and for a climate at doubled CO_2 concentrations (~2065). With these long records we searched for double populations in the extreme value distributions of annual wind extremes for return periods up to 10^4 yr. In addition, we explored the effect of increased greenhouse gas concentrations on the mean annual wind and on the double populations. Finally, we analyzed the meteorological conditions of the small but violent second population of extreme winds.

The paper is structured as follows: Section 2 describes the theoretical statistical basis, and section 3 the data handling of the model output and the detection of double populations from extreme value distributions. Section 4 describes the climate model used, and section 5 the statistical results. Section 6 analyzes the second population in a meteorological sense, and section 7 gives the discussion and conclusions.

Corresponding author address: H. W. van den Brink, Royal Netherlands Meteorological Institute, P.O. Box 201, 3730 AE De Bilt, Netherlands.
E-mail: brinkvdh@knmi.nl

2. Extreme value analysis

a. General arguments

Let M_m be the maximum of m independent observations $\xi_1, \xi_2, \ldots, \xi_m$ from distribution $F(x)$:

$$M_m = \max(\xi_1, \xi_2, \ldots, \xi_m), \tag{1}$$

then the distribution of M_m is given by

$$P(M_m \le x) = P(\xi_1 \le x)P(\xi_2 \le x) \cdots P(\xi_m \le x) = F^m(x). \tag{2}$$

Extreme value theory states that the distribution of maxima of many probability distributions $F(x)$ (properly normalized with α_m and μ_m) approaches asymptotically a specific class of functions $G(x)$:

$$\lim_{m\to\infty} P\left(\frac{M_m - \mu_m}{\alpha_m} \le x\right) = \lim_{m\to\infty} F^m(\alpha_m x + \mu_m) = G(x), \tag{3}$$

where $G(x)$ is given by

$$G(x) = e^{-(1-\theta x)^{1/\theta}}, \tag{4}$$

with the parameter $\theta \in \mathbb{R}$ determining the nature of the function [see e.g., de Haan (1990) and Kotz and Nadarajah (2000)]. A special case of $G(x)$ is the Gumbel distribution, which is $G(x)$ with $\theta = 0$. Then, interpreting $\theta = 0$ as the limit $\theta \to 0$, (4) reduces to

$$G(x) = e^{-e^{-x}}. \tag{5}$$

There are many examples of parent distributions $F(x)$ whose normalized extremes converge to the Gumbel distribution, for example, the normal distribution, the Poisson distribution, the exponential distribution and the Weibull distribution (Embrechts et al. 1997).

b. Generalized extreme value distribution of wind data

We rewrite m in (1) as rp, with r the effective number of independent daily averaged wind speeds in a year, and p the sample period in years from which the maxima M_p are extracted. If we suppose that the distribution of the normalized p-yr wind maxima follows $G(x)$ for $p \geq 1$, we can write (4) in the form of the so-called generalized extreme value (GEV) distribution $G_p(u)$ (Jenkinson 1955):

$$G_p(u) \equiv P(M_p \leq u) = e^{-e^{-x_p}}, \tag{6}$$

with M_p the p-yr wind maxima, $G_p(u)$ the GEV distribution resulting from p-yr sampling, and x_p a substitute for

$$x_p = \ln\left(1 - \theta\frac{u - \mu_p}{\alpha_p}\right)^{-1/\theta}, \tag{7}$$

in which μ_p is the location parameter, α_p the scale parameter, θ the shape parameter, and u the wind speed. The location parameter μ_p can be interpreted as the wind speed that is exceeded on average once per time interval p in the original record. Hence in the total record an exceedance of the level with value μ_p recurs with an average period p (Buishand and Velds 1980; see also Langbein 1949). For $\theta > 0$, u is bounded by an upper limit of value $\mu_p + \alpha_p/\theta$; for $\theta \leq 0$, u can approach infinity.

The Gumbel distribution ($\theta = 0$) is interpreted as the limit of (7) as $\theta \to 0$, leading to

$$x_p = \frac{u - \mu_p}{\alpha_p}. \tag{8}$$

Extreme value distributions are often plotted as a so-called Gumbel plot, where the variable u is on the ordinate, and the abscissa is transformed into the Gumbel variate:

$$\text{Gumbel variate} = -\ln\{-\ln[F^m(u)]\}. \tag{9}$$

On a Gumbel plot, a Gumbel distribution is represented by a straight line, whereas a GEV distribution ($\theta \neq 0$) is curved, downwardly for $\theta > 0$ and upwardly for $\theta < 0$.

Using that $[G_1(u)]^p = G_p(u)$ (Leadbetter et al. 1983, p. 8), from (6) follows:

$$x_p = x_1 - \ln p. \tag{10}$$

Substituting (10) into (7) gives

$$\mu_p = \mu_1 + \alpha_1\frac{1 - p^{-\theta}}{\theta}, \qquad \alpha_p = \alpha_1 p^{-\theta}. \tag{11}$$

In the special case of the Gumbel distribution, the right-hand sides of (11) are $\mu_1 + \alpha_1 \ln p$ and α_1, respectively.

In extreme value studies, the probability of exceedance of a certain value u is usually expressed in terms of the *return period* T. The return period T is the average number of years between two succeeding exceedances of the corresponding *return value* u:

$$T(u) \equiv \frac{1}{1 - G_1(u)} \approx pe^{x_p} \quad \text{for } T \gg p. \tag{12}$$

c. Two-component extreme value distribution

The local wind can be caused by two meteorological systems a and b of different physical nature, each of them generating its own distribution $F_a(u)$ and $F_b(u)$. Then, the parent distribution $F_{a,b}(u)$ is said to be mixed, and can be decomposed into

$$F_{a,b}(u) = (1 - \epsilon)F_a(u) + \epsilon F_b(u), \tag{13}$$

with $0 < \epsilon < 1$. An interpretation of (13) is that of every m samples, $(1 - \epsilon)m$ originates from mechanism a, and ϵm from b. Especially interesting is the case in which $\epsilon \ll 1$, and where the far tail of $F_b(u)$ is heavier than that of $F_a(u)$. Then $F_{a,b}(u) \approx F_a(u)$, and $F_b(u)$ can not easily be detected from the parent distribution. However, the extremely large events will originate from system b, which existence may be detected from the observed distribution of the extremes.

If $F_a^{(1-\epsilon)m}(u) \to G_a(u)$ and $F_b^{\epsilon m}(u) \to G_b(u)$, then the distribution of the extremes $G_{a,b}(u)$ of the mixed distribution is given by (Cook et al. 2003):

$$G_{a,b}(u) \equiv F_a^{(1-\epsilon)m}(u)F_b^{\epsilon m}(u) \to G_a(u)G_b(u), \tag{14}$$

where the subscripts a, b of $G(u)$ refer to the populations of systems a and b.

The simplest case of $G_{a,b}(u)$ represents the multiplication of two Gumbel distributions $G_a(u)$ and $G_b(u)$, which Rossi et al. (1984) calls the two-component extreme value (TCEV) distribution:

$$G_{a,b}(u) = \exp[-e^{-(u-\mu_a)/\alpha_a} - e^{-(u-\mu_b)/\alpha_b}]. \tag{15}$$

If transformed into the Gumbel variate, (9) becomes for the TCEV distribution

$$\text{Gumbel variate} = -\ln[e^{-(u-\mu_a)/\alpha_a} + e^{-(u-\mu_b)/\alpha_b}], \tag{16}$$

which shows that a Gumbel plot of extremes from a

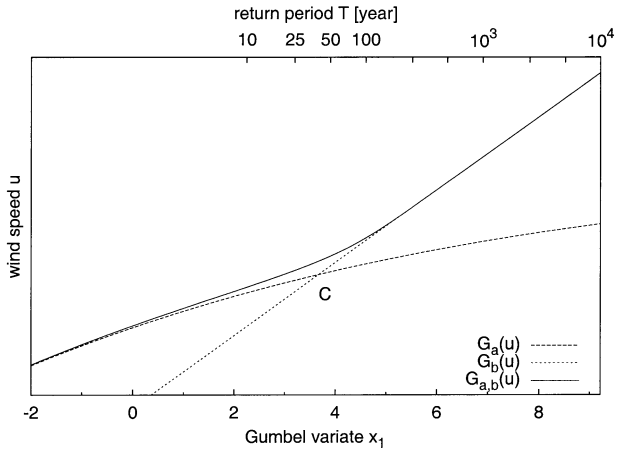


FIG. 1. A GEV distribution $G_a(u)$ and a Gumbel distribution $G_b(u)$, with the corresponding GTCEV distribution $G_{a,b}(u)$. The intersection point C of $G_a(u)$ and $G_b(u)$ is here at $x_1 = 3.67$, corresponding with a return period T_C of 40 yr. The distributions are shown on a Gumbel plot.

mixed distribution $G_{a,b}(u)$ results in a smooth transition from the asymptote of the distribution of the extremes $G_a(u)$ originating from the dominant population *a*, to the distribution of the violent extremes $G_b(u)$, originating from the rare population *b*. The intersection point C of $G_a(u)$ and $G_b(u)$ marks the sampling period T_C where the probability for sampling an extreme from population *a* or *b* is the same.

Figure 1 illustrates this behavior for the second-simplest case of a TCEV distribution, which is the case that $G_a(u)$ in (14) is allowed to generalize to a GEV distribution, but $G_b(u)$ remains a Gumbel distribution. We denote this type of two-component distribution, which we shall concentrate on in this paper, by the generalized two-component extreme value (GTCEV) distribution, of which the TCEV distribution is a special case. The reason for analyzing the second-simplest case is that the simplest case is not appropriate, as population *a* cannot be described by a Gumbel distribution. The combination of two GEV distributions is not suitable, as population *b* is too small to estimate the shape parameter θ.

For a GTCEV distribution to become apparent in the data, three conditions have to be fulfilled. First, the series length in years Y should amply exceed the return period T_C of the crossing point C. Second, the sampling period p should be sufficiently large to achieve convergence for the extremes of both populations *a* and *b* to their respective limits $G_a(u)$ and $G_b(u)$. Third, the sampling period p should be much smaller than T_C, since in the opposite case $G_a(u) \to 1$ and hence the GTCEV approaches $G_b(u)$, which is the ultimate extreme value limit of both $G_b(u)$ and $G_{a,b}(u)$. Note however, that the detection of the presence of a GTCEV distribution is easiest for $\theta_a > 0$, as in the opposite case $G_a(u)$ curves upwardly so that the GTCEV distribution becomes more difficult to distinguish from a single-component GEV distribution with $\theta_a < 0$. Figure 1 may help to illustrate these points.

3. Data handling

In empirical studies, the parameters of the GEV distribution $G_p(u)$ are obtained from a series with a finite length of Y yr. Taking the maxima of every p yr, n values remain to fit, with

$$n = \frac{Y}{p}. \tag{17}$$

Traditionally, the GEV distribution is applied to the annual maxima, so with sample period $p = 1$, giving $n = Y$ values to fit. This practice implicitly assumes that convergence to the GEV limit [Eq. (3)] is achieved for $m = r$, with $r \approx 50$, the effective number of independent daily averaged wind speeds in a year (Coles 2001, p. 98). This number r is less than 365 because exceedances of a high threshold tend to occur in clusters. It depends on the mathematical form of the parent distribution $F(u)$ whether the annual maxima have indeed converged to the asymptotic distribution $G_{p=1}(u)$ for $m \approx 50$. If the convergence is too incomplete for $p = 1$ to achieve a meaningful GEV analysis, one remedy is to increase the sampling period p. However, this leads to a proportionally decreased number of points n on which the GEV fit is based, and hence to increased sampling noise and standard errors in the parameter estimation of the GEV distribution. In our analysis, we apply an alternative method, namely to improve convergence by transforming the data in such a way that they become distributed according to a faster converging parent distribution. The method makes use of the fact that, for extratropical wind speeds, the Weibull distribution is well established as the parent distribution:

$$F(u) = 1 - e^{-(u/a)^k}, \tag{18}$$

with u the wind speed, a the Weibull scale parameter, and k the Weibull shape parameter. Theory shows that the maxima of observations from a Weibull distribution converge asymptotically (for any $k > 0$) to the Gumbel distribution, with the convergence rate depending on k, being largest for the exponential distribution ($k = 1$) (Embrechts et al. 1997; Cook and Harris 2001). Hence, improved convergence to the Gumbel distribution can be obtained if u^k instead of u is the fitted parameter, as this transforms the Weibull distribution into an exponential distribution. We made use of this property, and determined k in the tail of the parent distribution, after which the parameter u^k was fitted to a GEV distribution [Eqs. (6),(7)]. The underlying conjecture is that, even if deviations from the Weibull distribution in the far tail would lead to convergence to the GEV distribution instead of to the Gumbel distribution, the convergence rate to that GEV distribution is still faster for u^k than for u.

In the analysis of wind maxima, we restrict ourself

to the storm season (October–March) instead of to the annual maxima, assuming better homogeneity of the parent distribution within the storm season (Dillingh et al. 1993).

a. GEV parameter estimation

The parameters of the GEV distribution were estimated by the method of probability-weighted moments (Hosking et al. 1985). We also used his estimate of the plot positions:

$$x^i = -\ln\left[-\ln\left(\frac{i - 0.35}{n}\right)\right], \tag{19}$$

with x^i the plot position of the ith maximum in the set of n ordered maxima. Equation 19 can be regarded as a discrete version of (9). The estimated return value of the wind speed $u_{p,T}$ for a given x [which is determined by the sample period p and return period T via (12)] follows from inverting (7) and back-transforming u^k to u:

$$u_{p,T} = \left[\mu_p + \frac{\alpha_p}{\theta}(1 - e^{-\theta x_p})\right]^{1/k}, \tag{20}$$

with k the Weibull shape parameter. Neglecting the sampling error in the Weibull shape parameter k, the standard error $\sigma_{u,p,T}$ in the estimated return value $u_{p,T}$ is calculated by the so-called delta method (see, e.g., Coles 2001, p. 33):

$$\sigma^2_{u,p,T} = \nabla u^T \mathbf{V} \nabla u \tag{21}$$

with

$$\nabla u = \left(\frac{\partial u}{\partial \mu}, \frac{\partial u}{\partial \alpha}, \frac{\partial u}{\partial \theta}\right)^{\mathrm{T}}$$

and $\mathbf{V}$ the variance–covariance matrix

$$\mathbf{V} = \begin{pmatrix} \sigma^2_\mu & \sigma_{\mu\alpha} & \sigma_{\mu\theta} \\ \sigma_{\mu\alpha} & \sigma^2_\alpha & \sigma_{\alpha\theta} \\ \sigma_{\mu\theta} & \sigma_{\alpha\theta} & \sigma^2_\theta \end{pmatrix},$$

with σ_μ the standard error of μ etc. The values of $\mathbf{V}$ are given by Hosking et al. (1985).

b. Detection of two-component extreme value distributions

The statistical analysis was performed in four steps: First, the Weibull parameters a and k were determined for each grid point on the Northern Hemisphere. These parameters a and k were obtained by least mean-square fitting all daily wind speeds with $u > a$ (i.e., the upper e^{-1} part of the distribution, to exclude the influence of the lower wind speeds) in the winter season in a 150-yr record, hence 9932 daily values per grid point. Second, the GEV distribution was fitted to the set of 3509 annual maxima u^k for each grid point, so with $p = 1$ and $n = Y = 3509$. The third step was to identify possible GTCEV distributions, that is, locations where the extreme value distribution originates from populations a and b. In that procedure, we assume that the data are GTCEV-distributed if the 10^4-yr return value as obtained from the GEV fit to p-yearly sampled maxima $u_{p,10^4}$ exceeds the 10^4-yr return value as obtained from the GEV fit to annual maxima $u_{1,10^4}$ by more than two standard deviations. Expressed in a signal-to-noise ratio SN

$$\mathrm{SN} \equiv \frac{u_{p,10^4} - u_{1,10^4}}{\sqrt{\sigma^2_{u,p,10^4} + \sigma^2_{u,1,10^4}}}, \tag{22}$$

the criterion reads

$$\mathrm{SN} > 2, \tag{23}$$

where $\sigma_{u,1,10^4}$ is the sampling uncertainty in $u_{1,10^4}$, and $\sigma_{u,p,10^4}$ the standard error in $u_{p,10^4}$ when fitting a GEV distribution to p-yearly sampled maxima originating from $G_{p=1}(u^k)$ [using (11) to calculate μ_p and α_p]. If the wind were single-populated, then the probability is less than 5% that criterion (23) would result in erroneously concluding that wind is double populated. (This probability is, due to the skew distribution of rare return levels, somewhat larger than the 2.5% corresponding with a normal distribution.) This process is visualized in Fig. 2, which shows 3509 annual maxima with the fitted GEV distributions to the annual ($p = 1$) and centennial maxima ($p = 100$). We will apply (22) to annual

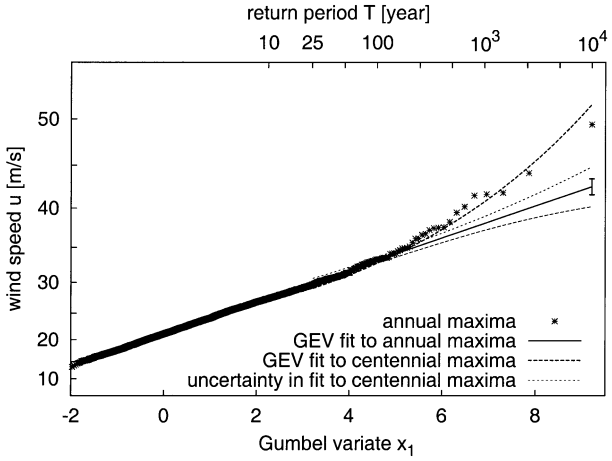


FIG. 2. Visualization of the procedure to detect GTCEV distributions. The 10^4-yr return value $u_{100,10^4}$, estimated by fitting the GEV distribution to the centennial maxima, is compared to the 10^4-yr return value $u_{1,10^4}$, estimated by fitting the GEV fit to the annual maxima. $G_a(u^k)$ is the estimate of the maxima of population $F_a(u)$ and $G_b(u^k)$ is the estimate of population b. For this case $u_{1,10^4} = 42.5$ m s^{-1} and $u_{100,10^4} = 51.6$ m s^{-1}, and $\sigma_{u,1,10^4} = 0.9$ m s^{-1} (indicated by the bar). The uncertainty in the fit to centennial maxima $\sigma_{u,100,10^4} = 2.1$ m s^{-1}, giving SN = 3.9. According to (23), this implies detection of a GTCEV distribution, and hence of a double population in the extreme winds. Shown is a set of 3509 daily averaged annual maxima from the greenhouse run at (47°N, 6°E). The vertical axis is linear in u^k, with $k = 1.74$.

and centennial maxima, although the outcome is rather robust for other choices of p.

The fourth step was to estimate the GTCEV distribution $G_{a,b}(u)$ for specific grid points where SN > 2. Here, we assumed that $F_b(u)$ is also Weibull distributed with the same shape parameter k as $F_a(u)$. $G_b(u)$ was estimated from the maxima which deviate considerably from the GEV fit to annual maxima, and $G_a(u^k)$ by adjusting μ_a, α_a, and θ_a iteratively in such a way that fitting a GEV distribution to the given distribution $G_{a,b}(u^k)$ [assuming that $G_b(u^k)$ is correctly estimated from the deviating maxima] results in the same GEV parameters as the original dataset. From $G_a(u^k)$ and $G_b(u^k)$, the intersection point C and the corresponding return period T_C were derived.

4. ECBilt-Clio Model description

The climate model used in this study is a coupled atmosphere–ocean–sea ice model of intermediate complexity, called ECBilt-Clio [referred to as ECBilt in van den Brink et al. (2003)]. The atmospheric component ''ECBilt'' is a spectral T21 global three-level quasi/geostrophic model. The atmospheric time step is 4 h. It is coupled to a dynamic ocean model ''Clio,'' which has a dynamic sea ice component and a relatively sophisticated parameterization of vertical mixing (Goosse and Fichefet 1999). For a more detailed description of the model, we refer to Opsteegh et al. (1998), Goosse and Fichefet (1999), and van den Brink et al. (2004).

a. Experimental setup

A transient run was generated for the period 1860–2080, using historical greenhouse forcings for 1860–2000, and the Special Report on Emissions Scenarios (SRES) A1B CO_2 emission scenario (Nakicenovic 2000) for 2000–80. This emission scenario results in approximately doubled CO_2 concentrations in 2050 (550 ppm) with respect to the emission in 1860 (290 ppm).

An ensemble of 121 runs of 30 yr each was generated, starting from the situation in 1960 of the transient run. The set of all 121 runs for the period 1960–89 is called the *control experiment.* For each grid point, vector-averaged daily mean extreme wind speeds were sampled from each October–March period in the set, giving 29 extremes per ensemble member and grid point, and 3509 extremes per grid point for the entire control experiment. Note that we often refer to annual extremes, whereas only storm-season extremes are sampled. For five ensemble runs, also all 27 000 daily averaged wind speeds in each grid point in the storm season were archived.

We also generated 121 ensembles runs of 30-yr starting from the situation in 2050 of the transient run. This set for the period 2050–79 is called the *greenhouse experiment.* As before, the series from which the extremes are sampled has a total length Y of 3509 yr, and the subseries for which all daily values were archived was 150 yr.

b. Validation of extreme statistics

For validation of the extreme wind distribution in ECBilt-Clio, we used the reanalysis dataset of the National Centers for Environmental Prediction (NCEP; Kalnay et al. 1996). This dataset provides the wind on a global $2.5° \times 2.5°$ grid every 6 h. We used the July 1965–June 2002 NCEP data. Lacking a 1000-hPa layer in ECBilt-Clio, we sampled wind speeds at 800 hPa ($\sim$2 km height) instead, being the lowest wind level in ECBilt-Clio. Comparison of the wind distributions at 850 and 1000 hPa for ocean grid points within the NCEP data shows similarity between the extreme value distributions (van den Brink et al. 2003). So, we assume that the ECBilt-Clio 800-hPa extreme winds have the same behavior as the extreme surface winds.

A suitable parameter to illustrate the ability of generating extremes is the GEV location parameter $\mu_{p=1}$, as it represents the wind speed that is exceeded on average once a year. Figure 3 shows $\mu_{p=1}$ (back-transformed from u^k to the wind u) as estimated from the 3509 annual wind extremes in the control experiment of ECBilt-Clio at 800 hPa and from the 36 annual wind extremes in the NCEP dataset at 850 hPa, respectively. Figure 3 shows an overall agreement in the patterns over sea, although the position of the storm tracks in ECBilt-Clio and NCEP are slightly different, with the Pacific storm track in ECBilt-Clio being too strong. Enhanced land–sea gradients and underestimations over land are apparent in ECBilt-Clio, probably caused by the simplified parameterizations of the boundary layer over land and the extremely low vertical resolution.

The wind speeds with return periods longer than 10^2 yr are considerably larger in ECBilt-Clio than in NCEP. This is illustrated in Fig. 4 for grid point (47°N, 65°W), for which ECBilt-Clio and NCEP have comparable estimates of the annual wind $\mu_{p=1}$, but differ considerably for larger return periods. Apparently, the variability in extremes is much larger in ECBilt-Clio than in the NCEP data.

In conclusion, it is clear that in the verification of ECBilt-Clio, considerable differences emerge. This is to be expected from models of intermediate complexity like ECBilt-Clio. Despite these shortcomings, the ECBilt-Clio values seem close enough to reality to justify studies like the present one, that is, exploration of the statistical nature of extreme winds, like the potential existence of a double population. However, it should be emphasized that the question of the reality of specific features generated in ECBilt-Clio can only be answered with results of models of higher complexity.

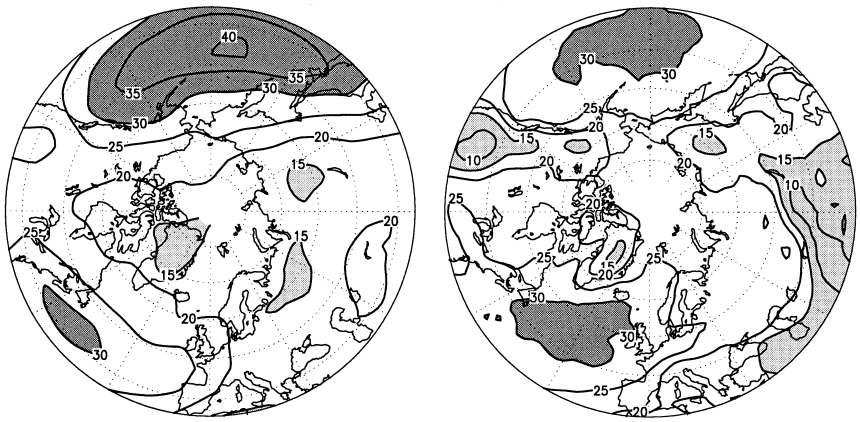


FIG. 3. Location parameter $\mu_{p=1}$ of the annual wind speeds, estimated from fitting a GEV distribution to (left) the ECBilt-Clio control experiment (3509 yr) and to (right) NCEP dataset (36 yr). The location parameter $\mu_{p=1}$ represents the wind speed, which is exceeded on average once a year. Values larger than 30 m s^{-1} and smaller than 15 m s^{-1} are shaded.

5. Results

a. Spatial distribution of GEV parameters in the control experiment

The estimated GEV scale parameter $\alpha_{p=1}$ and shape parameter $\theta_{p=1}$ of the daily averaged annual maxima of the wind speed for the EcBilt-Clio control experiment at 800 hPa are shown in Fig. 5. All three GEV parameters are largest over the oceanic storm tracks. In these regions, $\theta_{p=1}$ is lightly positive, that is, downwardly curved on a Gumbel plot. A possible reason might be that here there is a physical upper limit to the wind speed (although this limit is far beyond the 10^4-yr wind). The small range of absolute values of θ in Fig. 5b indicates that the annual maxima of u^k in EcBilt-Clio do not strongly deviate from the Gumbel distribution.

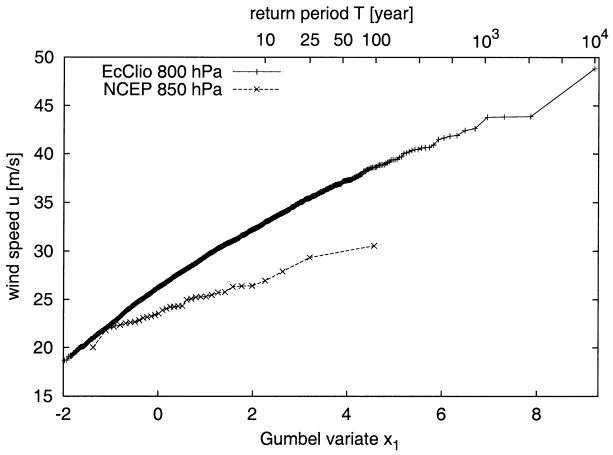


FIG. 4. Distribution of the annual maxima at 47°N, 65°W in ECBilt-Clio and NCEP. The variability in extremes in ECBilt-Clio is considerably larger than in the NCEP data.

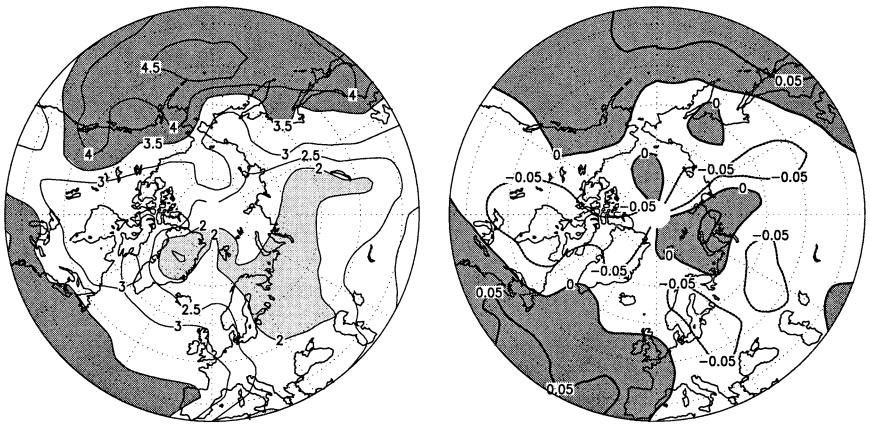
b. Two-component extreme value distributions in the control experiment

The spatial distribution of SN is shown in Fig. 6 for $|\mathrm{SN}| > 1.5$. Only patterns of large *positive* signal-to-ratios SN are detected. This indicates the reality of the patterns, as only situations with positive SN can be attributed to second populations. For 9.4% of the area shown in Fig. 6, SN is larger than 2, which is a factor 2 more than the expected 5% from Monte Carlo simulations. Figure 6 shows patterns of SN that fulfill our criterion [Eq. (23)] over the Atlantic, the east Pacific and Siberia, which indicates that in the control run, double populations in the extreme wind speeds are apparent at the end of both storm tracks.

c. Greenhouse effect on wind extremes

The change in the annual extreme wind due to the greenhouse effect is represented by the change in the GEV location parameter $\mu_{p=1}$, shown in Fig. 7. It shows a significant increase of the once-a-year exceeded wind speed over the northern Atlantic and Europe, as well as over the Pacific and North America. Maximum increase (5%) is found over Scandinavia. Comparison with Fig. 3 shows a zonally more elongated storm track, which is consistent with the positive NAO-like response to enhanced greenhouse gas forcing in EcBilt-Clio (Fig. 11 in Schaeffer et al. 2003, manuscript submitted to *Climate Dyn.*). Apparently, the change in the annual wind maxima $\mu_{p=1}$ behaves similarly to the change in the mean wind in winter.

The regions where TCEV distributions are detected in the greenhouse experiment are shown in Fig. 8. It shows the same patterns as in the control run, with the Atlantic region shifted to the east and elongated from Spain to Finland. For 10.6% of the area shown in Fig.



(a) scale parameter $\alpha_{p=1}$. Values larger than 3.5 and smaller than 2.0 are shaded

(b) shape parameter $\theta_{p=1}$. Positive values are shaded

FIG. 5. Estimated (left) GEV scale parameter $\alpha_{p=1}$ (back-transformed from u^k to u) and (right) shape parameter $\theta_{p=1}$ of the daily averaged wind speed for the EcBilt-Clio control experiment at 800 hPa as derived from 3509 annual maxima of u^k. In the left panel values larger than 3.5 and smaller than 2.0 are shaded; in the right panel positive values are shaded. The standard error σ_θ according to Hosking et al. (1985) is between 0.012 and 0.015.

8, SN is larger than 2, which is a slight increase with respect to the control run.

6. Meteorology of the second population

Close inspection of the grid points with large positive SN reveals that the deviating extremes of neighboring grid points all originate from a restricted number of storms. Apparently, the storms are so intense, that they influence the highest extremes over a larger area during their track to the east.

To find the meteorological circumstances responsible for this second population, we concentrate on the events which clearly belong to that second population. We will consider grid point (42°N, 17°W) in the control run, for which the maximum signal-to-noise ratio SN of 3.8 occurs. Figure 9 shows that the three most severe events substantially deviate from the fit, and thus can safely be assumed to originate from the second population. This is also apparent from the fact that SN decreases from 3.8 to $0.1 \ll 2$ if these three points are omitted from the dataset, which means that there is no significant detection of a second distribution possible without the three largest events. The relative vorticity for the most extreme event is shown in Fig. 10. In the first 2 days, it displays a mature cyclone which is no longer devel-

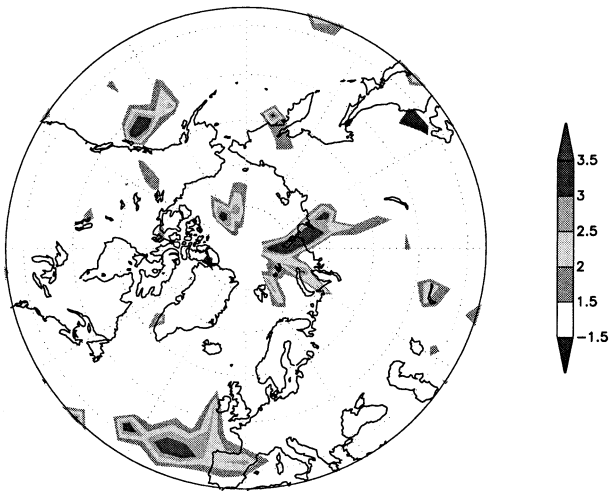


FIG. 6. Signal-to-noise ratio SN for $|\mathrm{SN}| > 1.5$ in the control experiment. According to our criterion (23), $\mathrm{SN} > 2$ indicates the existence of a double population.

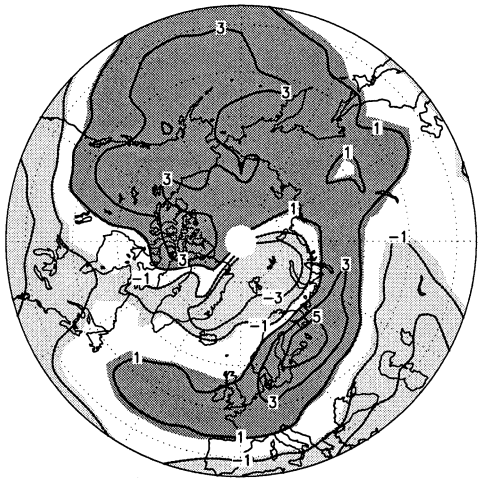


FIG. 7. Relative changes (%) in the GEV location parameter $\mu_{p=1}$ due to the greenhouse effect. The shaded areas are significant at 5% level.

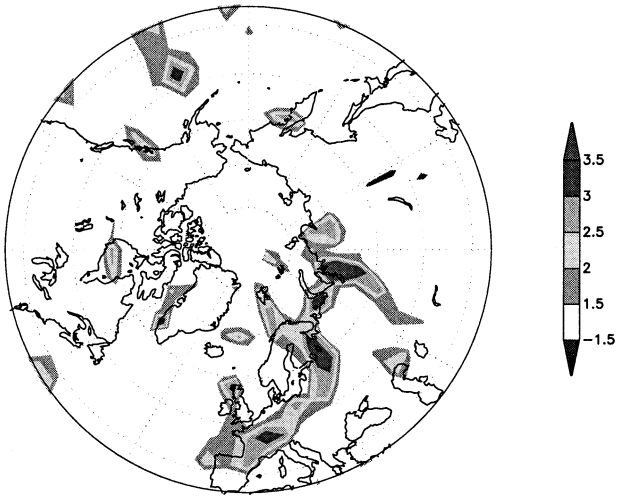


FIG. 8. Signal-to-noise ratio SN for $|\mathrm{SN}| > 1.5$ in the greenhouse experiment. According to our criterion (23), $\mathrm{SN} > 2$ indicates the existence of a double population.

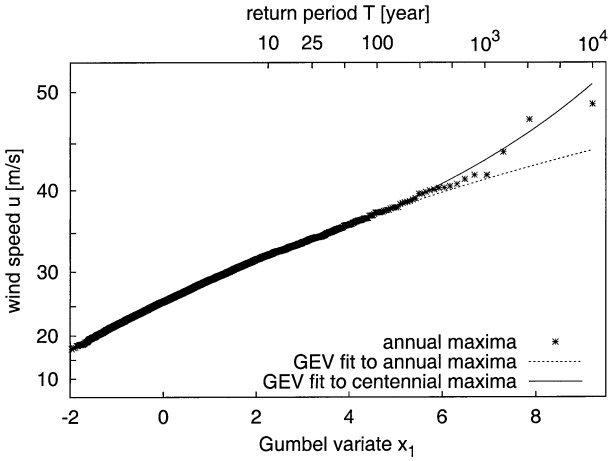


FIG. 9. Gumbel plot for grid point (42°N, 17°W) in the control run. The signal-to-noise ratio SN according to (22) is 3.8. The three most severe events especially deviate from the fit and are assumed to originate from the second population. The vertical scale is linear in u^k, with $k = 1.71$.

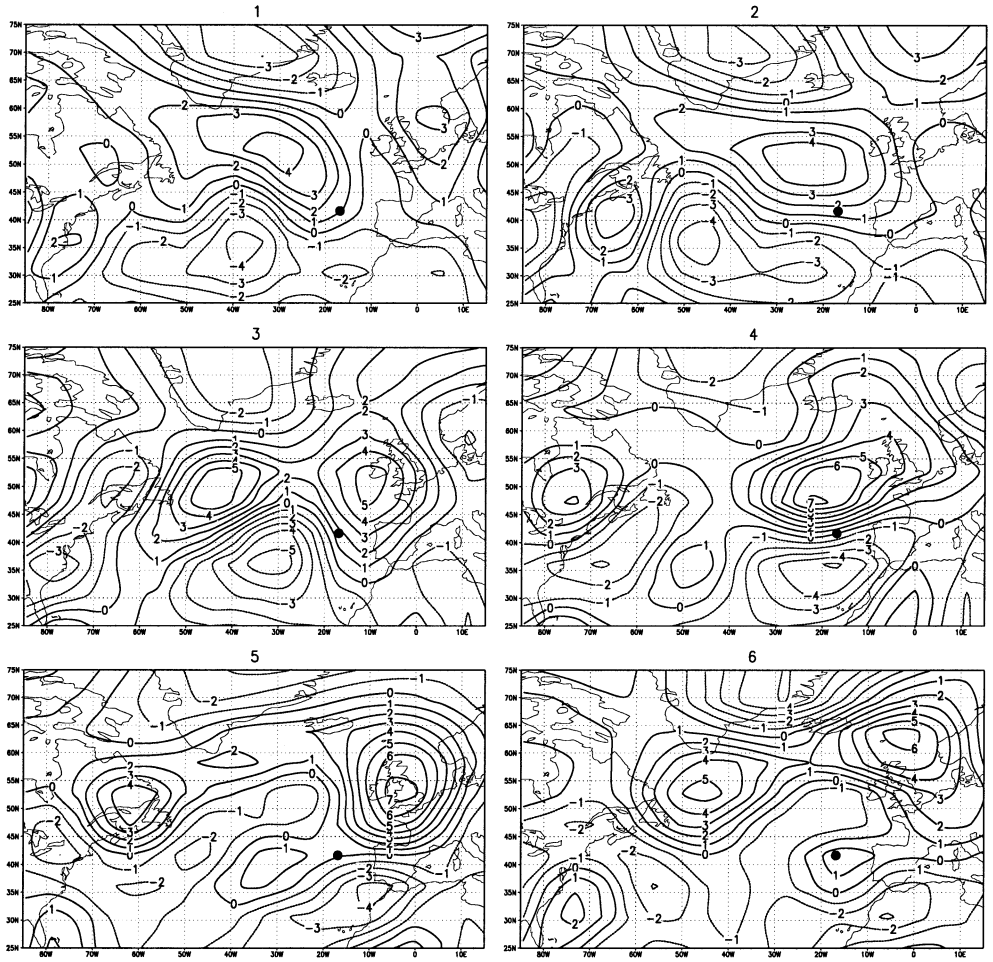


FIG. 10. Daily averaged 800-hPa relative vorticity (10^{-5} s^{-1}) during merging. At day 4, the most extreme wind in 3509 yr is reached at the location indicated by a black dot (42°N, 17°W).

oping. However, at day 3, another cyclone is starting to merge with the original cyclone, resulting in explosive cyclogenesis and extreme wind speeds at the location of interest until day 5. After day 5, the decay phase of the eddy sets in. The importance of wave merging for the process of explosive cyclogenesis is stressed in several observational studies of cyclogenesis, for example, Hakim et al. (1995, 1996) and Gaza and Bosart (1990). In our results, merging occurs in the 1st, 2d, 3d, and 5th most extreme events, but not in the situations of the other 10 largest extremes. We hypothesize that merging is a crucial condition for a second population of extreme wind speeds to occur.

Analysis of normal annual extremes shows that, at this location, wave merging is not exceptional. So, although important for the cyclogenesis process, it is clearly not a sufficient condition for a second population to occur. To distinguish between normal annual extremes and the merging events of the second population, we examined the anomalous time mean 500-hPa streamfunction pattern. The anomaly pattern was computed by first averaging the 500-hPa streamfunction $\boldsymbol{\Psi}$ over a period of 7 days preceding the day for which the maximum wind occurred in the grid point (42°N, 17°W). This was done for 600 cases belonging to the first population *a*, and the four cases belonging to the second population *b*. The anomalous pattern $\boldsymbol{\Psi}_{\text{an}}$ is defined as the mean of the mentioned cases $\boldsymbol{\Psi}_b$ minus the mean of the 600 cases $\boldsymbol{\Psi}_a$ in the area between 20° and 65°N and between 80° and 10°W:

$$\boldsymbol{\Psi}_{\text{an}} = \boldsymbol{\Psi}_b - \boldsymbol{\Psi}_a,$$

$$\boldsymbol{\Psi}_a = \frac{1}{600}\sum_{i=1}^{600} \boldsymbol{\Psi}_{a,i}, \qquad \boldsymbol{\Psi}_b = \frac{1}{4}\sum_{i=1}^{4} \boldsymbol{\Psi}_{b,i}, \quad (24)$$

where $\boldsymbol{\Psi}_{a,i}$ is the 7-day averaged 500-hPa streamfunction pattern of case *i* in the first population *a*, and $\boldsymbol{\Psi}_{b,i}$ is the same for the second population *b*. Figure 11 displays $\boldsymbol{\Psi}_a$ and $\boldsymbol{\Psi}_{\text{an}}$, and Fig. 12 the corresponding zonal wind pattern.

Figure 11 and 12 show that annual extremes develop in a mean circulation which is in a strong westerly phase, with maximum time mean zonal winds of 24 m s^{-1}. The anomaly pattern of the second population has a large positive amplitude in the model's version of the North Atlantic Oscillation pattern, and leads to a much stronger jet than normal annual extremes (up to 31 m s^{-1}). The pattern has an extension in the easterly direction. We have computed the probability that the anomaly pattern of a member in the first population projects just as strong on the $\boldsymbol{\Psi}_{\text{an}}$ pattern as the four members of the second population. For the projection we used the squared norm:

$$p_i = \frac{\langle \boldsymbol{\Psi}_{a,i} - \boldsymbol{\Psi}_a, \boldsymbol{\Psi}_{\text{an}} \rangle}{\langle \boldsymbol{\Psi}_{\text{an}}, \boldsymbol{\Psi}_{\text{an}} \rangle}. \quad (25)$$

The events of population *a* will have an average projection of zero, and population *b* of one.

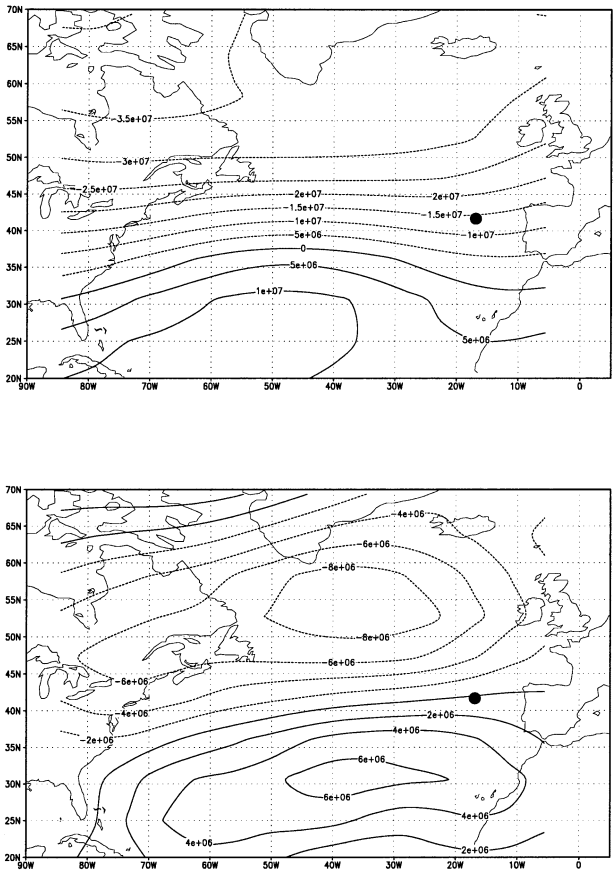


FIG. 11. (top) Mean streamfunction $\boldsymbol{\Psi}_a$ of the first population *a* and (bottom) the difference $\boldsymbol{\Psi}_{\text{an}}$ between the second population *b* and the first population *a*.

There are only six events of the 600 with a projection larger than unity, which means that the location and intensity of the jet stream of the second population has a frequency of order once in 10^2 yr. Another striking feature of the second population is that the events are accompanied by extreme precipitation, where large-scale precipitation and convective precipitation contribute equally. All four events of the second population have daily precipitation rates which have return periods of order 10^3 yr.

We tentatively conclude that an extremely strong jet stream in which wave merging occurs can generate conditions which are favorable for the occurrence of strong diabatic feedbacks. This leads to anomalously strong cyclogenesis and the generation of a second population of wind extremes.

We checked this hypothesis with the data from the greenhouse experiment. We consider grid point (47°N, 5°E), which has a maximum in the signal-to-noise ratio SN of 3.9. Here, the 1st to 7th and 10th largest extremes belong to cyclones that originated after merging. Skipping the seven largest events from the dataset reduces SN from 3.9 to $0.35 \ll 2$. So, again, the detection is

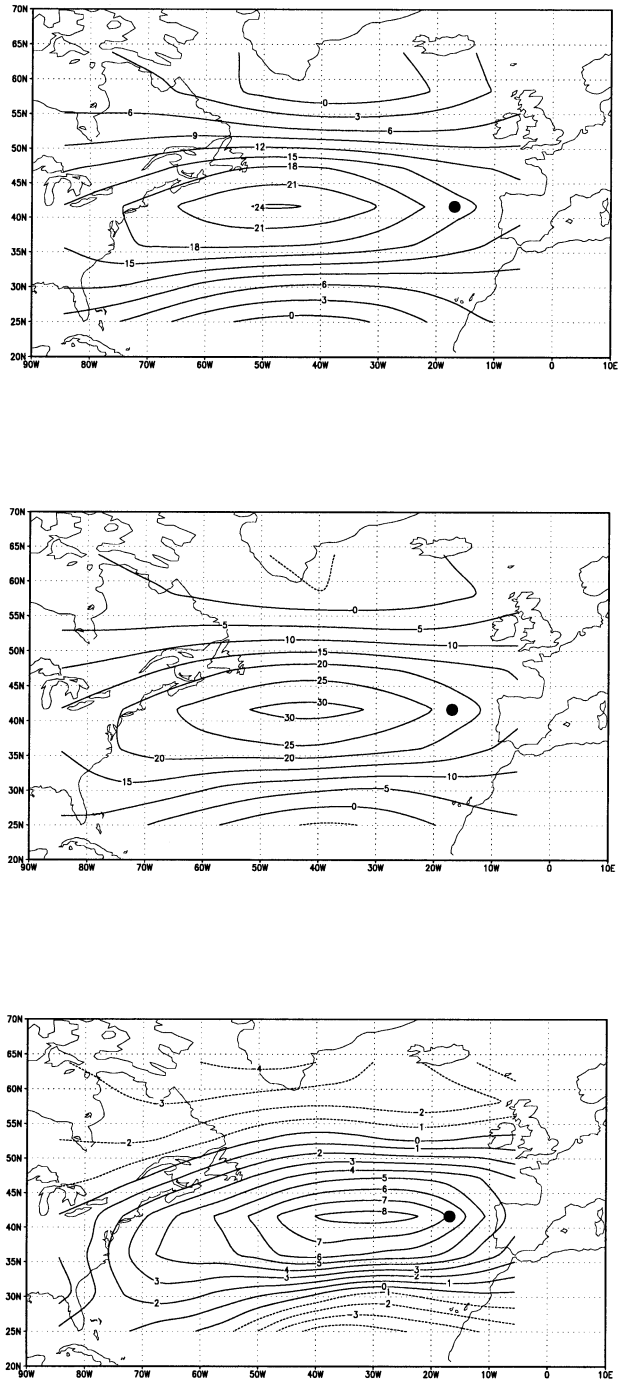


FIG. 12. Time mean zonal wind at 500 hPa of (top) the first population, (middle) the second population, and (bottom) their difference.

only significant when the merging events are incorporated.

Projection on the area between 80°W and 10°E and between 20° and 65°N shows an increase of the maximum mean zonal wind from 22 to 31 m s^{-1}. Once in 56 yr, the projection is larger than unity. Also the precipitation rates are extraordinary. For six of the seven events, the return periods of the precipitation rates are of order 10^3 yr. So, our hypothesis based on the control experiment is confirmed by the results of the greenhouse experiment.

We conclude that the extreme wind speed belongs to a second population if the following three conditions are fulfilled: First, there is an intense jet stream, corresponding with a positive NAO. Second, two cyclones merge to a single intense cyclone. Third, the cyclone is accompanied by extreme precipitation.

7. Discussion and conclusions

The climate model ECBilt-Clio shows preferred regions for the extratropics in which the annual wind extremes with return periods of order 10^3 yr belong to another population than the more frequent annual winds. One consequence of this result is that in such regions, the 10^4-yr wind speed can not be estimated from annual extremes in observed series with time lengths of order 100 years. Only a lower limit of the 10^4-yr wind can be estimated from such a series, as the existence of a second population in the extremes always increases the 10^4-yr wind. Another, closely related, consequence is that this low frequency of the second population prevents detection from single-station observational records.

We found that the second population in EcBilt-Clio exists of merging cyclones embedded in a strong jet stream, and that they are accompanied with extreme precipitation. The robustness of these results has to be confirmed by analyzing the results of more advanced models.

Doubling of the greenhouse gas concentrations has two important effects on the second population in the wind speed. The first is that the regions change for which second populations appear. This implies that regions, which are single populated in the current climate, may be double populated in a doubled CO_2 climate and vice versa. The second effect of CO_2 doubling in EcBilt-Clio is that, in double-populated areas, the frequency of cyclones from the second population increases. Whereas the second population is dominant over the first population for return periods of 600 yr and larger for the control run, this turning point lays at a return period of 40 yr for the greenhouse run. This implies that not only the 10^4-yr winds are influenced by the second population, but also the 10^2-yr winds.

We attribute the eastward shift of the Atlantic area with double populations in the greenhouse experiment to the response in the climatological winter mean, which resembles a positive NAO pattern with largest westerly wind increase over Scandinavia (see Fig. 11 in Schaeffer et al. 2003, manuscript submitted to *Climate Dyn.*). This response causes the eastward elongation of the storm track as shown in Fig. 7, and consequently of the area in which a second population occurs.

REFERENCES

Buishand, T. A., and C. A. Velds, 1980: *Neerslag en Verdamping.* Staatsdrukkerij, 206 pp.

Coles, S., 2001: *An Introduction to Statistical Modelling of Extreme Values.* Springer-Verlag, 208 pp.

Cook, N. J., 1982: Towards better estimation of extreme winds. *J. Wind Eng. Ind. Aerodyn.,* **9,** 295–323.

——, and R. I. Harris, 2001: Discussion on application of the generalized Pareto distribution to extreme value analysis in wind engineering by J. D. Holmes, W. W. Moriarty. *J. Wind Eng. Ind. Aerodyn.,* **89,** 215–224.

——, ——, and R. Whiting, 2003: Extreme wind speeds in mixed climates revisited. *J. Wind Eng. Ind. Aerodyn.,* **91,** 403–422.

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

Dillingh, D., L. de Haan, R. Helmers, G. P. Können, and J. van Malde, 1993: De basispeilen langs de Nederlandse kust; statistisch onderzoek. Tech. Rep. DGW-93.023, Ministerie van Verkeer en Waterstaat, Directoraat-Generaal Rijkswaterstaat, 132 pp.

Embrechts, P., C. Klüppelberg, and T. Mikosch, 1997: *Modelling Extremal Events for Insurance and Finance.* Springer-Verlag, 645 pp.

Gaza, R. S., and L. F. Bosart, 1990: Trough-merger characteristics over North America. *Wea. Forecasting,* **5,** 314–331.

Goosse, H., and T. Fichefet, 1999: Importance of ice–ocean interactions for the global ocean circulation: A model study. *J. Geophys. Res.,* **104,** 23 337–23 355.

Hakim, G. J., L. F. Bosart, and D. Keyser, 1995: The Ohio Valley wave-merger cyclogenesis event of 25–26 January 1978. Part I: Multiscale case study. *Mon. Wea. Rev.,* **123,** 2663–2692.

——, D. Keyser, and L. F. Bosart, 1996: The Ohio Valley wave-merger cyclogenesis event of 25–26 January 1978. Part II: Diagnosis using quasigeostrophic potential vorticity inversion. *Mon. Wea. Rev.,* **124,** 2176–2205.

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. *Quart. J. Roy. Meteor. Soc.,* **81,** 158–171.

Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.,* **77,** 437–471.

Kotz, S., and S. Nadarajah, 2000: *Extreme Value Distributions: Theory and Applications.* Imperial College Press, 185 pp.

Langbein, W. B., 1949: Annual floods and the partial-duration flood series. *Eos, Trans. Amer. Geophys. Union,* **30,** 879–881.

Leadbetter, M. R., G. Lindgren, and H. Rootzén, 1983: *Extremes and Related Properties of Random Sequences and Processes.* Springer-Verlag, 336 pp.

Nakicenovic, N., Ed., 2000: *Special Report on Emission Scenarios.* Cambridge University Press, 595 pp.

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

Palutikof, J. P., B. B. Brabson, D. H. Lister, and S. T. Adcock, 1999: A review of methods to calculate extreme wind speeds. *Meteor. Appl.,* **6,** 119–132.

Rossi, F., M. Fiorentino, and P. Versace, 1984: Two-component extreme value distribution for flood frequency analysis. *Water. Resour. Res.,* **20,** 847–856.

Simiu, E., N. A. Heckert, J. J. Filliben, and S. K. Johnson, 2001: Extreme wind local estimates based on the Gumbel distribution of dynamic pressures: An assessment. *Struct. Saf.,* **23,** 221–229.

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

——, ——, ——, G. J. van Oldenborgh, and G. Burgers, 2004: Improving 10^4-year surge level estimates using data of the ECMWF seasonal prediction system. *Geophys. Res. Lett.,* **31,** L17210, doi:10.1029/2004GL020610.