Accessing extremes of mid-latitudinal wave activity: methodology and application

Article

Published Version

It is advisable to refer to the publisher's version if you intend to cite from the work. See Guidance on citing.

www.reading.ac.uk/centaur

CentAUR

TELLUS

Accessing extremes of mid-latitudinal wave activity: methodology and application

By R. VITOLO[1]*, P. M. RUTI[2], A. DELL'AQUILA[2], M. FELICI[1,3], V. LUCARINI[4]
and A. SPERANZA[1],   [1]*PASEF – Physics and Applied Statistics of Earth Fluids, Dipartimento di Matematica ed Informatica, Università di Camerino, Via Madonna delle Carceri – 62032 Camerino (MC), Italy;* [2]*ENEA, via Anguillarese 301 – 00060 S. Maria di Galeria, Roma, SP 091 Italy;* [3]*Dipartimento di Matematica U. Dini, Università di Firenze, viale Morgagni 67/A – 50134 Firenze, Italy;* [4]*Dipartimento di Fisica, Università di Bologna, viale Berti Pichat 6/2 – 40126 Bologna, Italy*

ABSTRACT

A statistical methodology is proposed and tested for the analysis of extreme values of atmospheric wave activity at mid-latitudes. The adopted methods are the classical block-maximum and peak over threshold, respectively based on the generalized extreme value (GEV) distribution and the generalized Pareto distribution (GPD). Time-series of the 'Wave Activity Index' (WAI) and the 'Baroclinic Activity Index' (BAI) are computed from simulations of the General Circulation Model ECHAM4.6, which is run under perpetual January conditions. Both the GEV and the GPD analyses indicate that the extremes of WAI and BAI are Weibull distributed, this corresponds to distributions with an upper bound. However, a remarkably large variability is found in the tails of such distributions; distinct simulations carried out under the same experimental setup provide sensibly different estimates of the 200-yr WAI return level. The consequences of this phenomenon in applications of the methodology to climate change studies are discussed. The atmospheric configurations characteristic of the maxima and minima of WAI and BAI are also examined.

1. Introduction

Recently, different aspects of the general circulation of the atmosphere have been investigated within the framework of either very simplified models or state-of-the-art GCMs, namely:

(1) the statistical properties, specifically extreme value statistics, of time-series of total energy generated by a baroclinic model of intermediate complexity for the atmospheric mid-latitudes (Felici et al., 2007a,b; Lucarini et al., 2007a,b);

(2) the description of wave propagation in the mid-latitude atmosphere as simulated by a state-of-the-art coupled GCMs (simulations for the IPCC report) and associated products such as reanalyses (Dell'Aquila et al., 2005; Ruti et al., 2006; Lucarini et al., 2007c).

While in the papers cited in (1) the procedures concerning extreme statistics and their parametric inference have been defined,

in the papers quoted in (2), dynamically oriented climate metrics aimed at capturing the basic statistical properties of the fundamental features of the mid-latitude atmospheric variability – for example, synoptic baroclinic waves and planetary waves – have been introduced with the purpose of auditing GCMs and reanalyses. In this paper our purpose is to join the two research streams described above. Namely, we aim at developing a sound methodology to attack the meteo-climatic problem of determining the sensitivity of the (extreme) statistics of mid-latitudinal disturbances to CO_2 increase. The change in the statistics of mid-latitude disturbances is considered a complex issue, as it is not obvious to determine which mechanisms are involved in the change of the meridional temperature gradient at the various pressure levels. The method of investigation we propose is based on performing rigorous statistical analysis of extreme values of 'dynamically oriented metrics'.

The Northern Hemisphere mid-latitude variability is mainly driven by large-scale (specifically: planetary scale) processes which affect, at different spatial and temporal scales, the variability of surface parameters (i.e. precipitation, wind, temperature). The latter parameters are those typically observed in the meteorological networks and stored in the records. Nevertheless, time-series of sufficient length, suitable for reliable statistical analysis of extreme values, are not always available. This has

---

*Corresponding author.
Current address: School of Engineering, Computing and Mathematics, University of Exeter, Harrison building, North Park Road, Exeter, EX4 4QF, UK.
e-mail: r.vitolo@exeter.ac.uk

sometimes led the researchers to adopt weaker criteria for the selection of events to be used for analysis of extreme value statistics. In parametric inference this may introduce a substantial bias (Felici et al., 2007a); in practice, different studies based on the same data may obtain different quantitative results.

With respect to observations, the usage of model-generated data has the advantage that it makes it possible to satisfy the necessary requirements of length and quality of the record, discussed in, for example, Felici et al. (2007a). Indeed, statistical analysis of extreme events has been successfully applied to time-series of surface parameters produced by climate models. Beniston et al. (2007) consider heat waves, heavy precipitation, drought, wind storms and storm surges changes between present and future climate on the basis of regional climate model simulations produced by the PRUDENCE project. Potential future changes in temperature and precipitation extremes have been examined in an ensemble of global coupled climate models used for the preparation of the Fourth IPCC Assessment Report (Kharin et al., 2007). Temperature and precipitation extremes were already considered in Zwiers and Kharin (1998), Kharin and Zwiers (2000). For evident reasons, most of the research on extremes in climate models focuses on surface parameters such as temperature, precipitation and wind. Here, we wish to explore a different aspect: extremes of planetary and baroclinic atmospheric waves. Obviously, surface fields are strongly influenced by planetary-scale driving processes, but the characterization on sound theoretical grounds of a well-defined relation between the two is an open (and very hard) problem. One way of tackling this is through statistical downscaling, see, for example, Friederichs and Hense (2007).

Among the dominant physical processes featured in the midlatitude atmospheric dynamics, the synoptic waves and the interaction between ultra-long waves having zonal wavenumber smaller than five (Bordi et al., 2004) and topography are main ingredients. The synoptic travelling perturbations can be represented as high-frequency high-wavenumber eastward propagating spectral features, characterized by periods of order 2–7 d and by spatial scales of a few thousand kilometres. These perturbations can be associated with the release of available energy driven by conventional baroclinic conversion (Blackmon, 1976; Speranza, 1983; Wallace et al., 1988), so that they are often referred to as baroclinic waves. On the other hand, so-called planetary waves, which interact with orography (Charney and DeVore, 1979; Charney and Straus, 1980; Buzzi et al., 1984; Benzi et al., 1986) and are catalysed by the subtropical jet (Benzi and Speranza, 1989; Ruti et al., 2006), play a dominant role in the low-frequency low-wavenumber spectral region of so-called stationary waves, whose characteristic time and space scales belong to the interval 10–45 d and 7000–15 000 km, respectively (Hansen and Sutera, 1986).

The methodology of the present work consists in analysing the statistics of extreme values (in the parametric setting) for indexes large-scale atmospheric activity. A state-of-the-art GCM is run in perpetual winter conditions for present-day atmospheric CO_2 concentrations. We generate time-series of the Wave Activity Index (WAI) and of the Baroclinic Activity Index (BAI), which may be considered as proxies of the activity of the planetary and synoptic waves, respectively (Benzi et al., 1986; Hansen and Sutera, 1986; Dell'Aquila et al., 2007). Extreme value analysis is then performed by fitting generalized extreme value (GEV) distributions and generalized Pareto distributions (GPD) (Coles, 2001; Felici et al., 2007a,b) on subsequences of 'extreme values' extracted from the generated time-series. For the GEV, the extreme values are defined as maxima (or minima) computed over blocks of sufficiently large size, whereas for the GPD the extremes are values exceeding a sufficiently high threshold (for this reason, the latter is also called the peaks over threshold method – POT for shortness). The aim of studying the extremes of the WAI and BAI is to identify dynamically meaningful features of the general atmospheric circulation whose response under CO_2 increase can be accurately assessed. The latter issue is subject of current research by some of the authors. We emphasize that the above methods provide a common ground which can simplify the comparison of results from different studies, compare with, for example, Beniston et al. (2007).

We now give an outline of the paper. A short description of the model, metrics and statistical methodology can be found in Section 2. The general statistical properties of the time-series are analysed in Section 3, whereas the extreme value analysis is reported in Section 4. Conclusions and lines of future research are summarized in Section 6.

2. Data and methods

2.1. Description of the model and experimental setup

The atmospheric model used in this study is ECHAM4.6, an evolution of the model used by Roeckner and Arpe (1995), belonging to the fourth generation of GCM developed at the Max Planck Institute for Meteorology in Hamburg. It is an evolution of the spectral weather prediction model of the European Centre for Medium Range Weather Forecasts (Simmons et al., 1989). ECHAM4 uses the spectral transform method for 'dry dynamics' while water vapour, cloud water and trace constituents are advected by using a shape-preserving semi-Lagrangian scheme (Williamson and Rasch, 1994). The model atmosphere is vertically resolved in 19 layers, from the surface up to 10 hPa. The model contains a set of parametrization for unresolved or not explicitly represented dynamical and physical processes, including radiation (Fouquart and Bonnel, 1980; Morcrette, 1991), cumulus convection (Tiedke, 1989; Nordeng, 1994) stratiform clouds (Roeckner, 1995), gravity wave drag (Miller et al., 1989), vertical diffusion and surface fluxes, land surface processes and horizontal diffusion. A summary of the design and performance of ECHAM4 can be found in Roeckner and Arpe (1995).

The simulations analysed in this paper have been performed at T30 spectral horizontal resolution, corresponding approximately to a grid of $3.75° \times 3.75°$. The GCM has been run for 600 model months (30 d each), under perpetual January conditions, with present-day value of CO_2 concentration (360 ppmv). The sea surface temperature (SST) and sea ice cover are kept constant in time and fixed to the climatological mean of the month of January. The choice of using simulations without seasonal cycle has a two-fold motivation, first of all, mid-latitude atmospheric processes which might be affected by the climate change, specifically the baroclinic activities, 'take place under winter conditions' (Blackmon, 1976). Moreover, stationarity implies that there is no need to detrend the data; for inference of extreme values this is an important advantage, since it allows to reduce the number of parameters which are necessary in the statistical model. This, in turn, allows to more clearly understand the nature of variability in the time-series, see the discussion in Felici et al. (2007b).

2.2. Wave indexes used for the computation of time-series

Our study focuses on the Northern Hemisphere mid-latitude atmospheric winter variability, therefore, we consider the latitudinal belt 30°N–60°N, where both the baroclinic and the low-frequency planetary wave activities are present in the ECHAM4.6 model. The 500 hPa geopotential height field is latitudinally averaged over this belt, to define a one-dimensional longitudinal field representative of the atmospheric variability at mid-latitudes. We have verified that the results presented below are not sensitive with respect to small variations in the latitudinal band, this is compatible with the fact that they mainly refer to planetary-scale coherent atmospheric features. We choose the 500 hPa geopotential height because it is one of the typical variables used for the study of the planetary and synoptic scale atmospheric circulation (Blackmon, 1976) and for the comparison of different atmospheric data sets of climatological relevance. The indexes of dynamical activity for large-scale features of the mid-latitude troposphere are computed according to the following procedure:

(1) the 500 hPa geopotential height field $Z(\lambda, \phi)$ is averaged with respect to latitude ϕ over the latitudinal band bounded between 30°N and 60°N;

(2) for each day, the 500 hPa geopotential height is Fourier decomposed in the longitudinal direction λ;

(3) an index is finally computed from the variance associated to the Fourier coefficients Z_k for k ranging from k_1 up to k_n:

$$Z_{k_1, k_n}(t) = \left(\sum_{k=k_1}^{k_n} 2|Z_k(t)|^2 \right)^{\frac{1}{2}}. \tag{1}$$

The Wave Activity Index (Benzi et al., 1986; Hansen and Sutera, 1986), or WAI, is then computed as the root mean square

of the zonal wavenumbers 2–4 of the winter 500 hPa geopotential height variance over the channel 30°N–60°N, that is, formula (1) with $n = 3$, $k_1 = 2$ and $k_3 = 4$. Furthermore, an index of synoptic disturbances has been computed using $n = 3$, $k_1 = 6$ and $k_3 = 8$; we refer to this as the BAI. The physical meaning of the above introduced WAI and BAI is further discussed in Section 3.

2.3. Statistical inference of extreme values

We assume that the time-series generated by the ECHAM model (see Sections 2.1 and 2.2) are realizations of an unknown stochastic process and we use these time-series to perform two kinds of 'parametric inferences': point and interval estimations. Two parametric models are chosen for the statistical analysis of extremes: the GEV distribution and the GPD. Different inferential procedures are associated to the GEV and GPD: for the former one typically adopts the 'annual maximum' method, whereas the latter is used in conjunction with the POT method (Coles, 2001; Smith, 2004). The GEV distribution function is

$$G(x) = \exp\left[-\left(1 + \xi \frac{x - \mu}{\sigma} \right)_+^{-1/\xi} \right], \tag{2}$$

where the notation $F(x)_+$ means the following: $F(x)_+ = F(x)$ for all x such that $F(x) > 0$ and $F(x)_+ = 0$ otherwise. For $\xi = 0$, the GEV reduces to the Gumbel distribution (see Coles, 2001). Given a time-series, the distributional parameters (μ, σ, ξ) are inferred by maximum likelihood procedures from sequences of block maxima extracted from the time-series.[1] The methodology has been described in detail in Felici et al. (2007a,b), also see Coles (2001), Smith (2004) for basic theory and more examples. It should be emphasized that the GEV model is valid only 'asymptotically'. More precisely, one has to assume that the observed time-series is the realization of a stationary stochastic process X_i belonging to the so-called maximum domain of attraction (Resnick, 1987; Malevergne et al., 2006); denoting the maximum of the first n observations by M_n^*, one must assume that there exist constants a_n and b_n such that the normalized maximum $M_n = (M_n^* - b_n)/a_n$ converges to a non-degenerate distribution G. Under these assumptions, G is of the form (2). In the applications, the parent process X_i is a priori unknown and it is, therefore, not possible to determine whether the constants a_n and b_n exist or not; one tacitly assumes that they do exist. At any rate, the constants a_n and b_n can be 'absorbed' in the point estimates of the parameters μ and σ, assuming that one is in the asymptotic range of the validity of the GEV, that is,

---

[1] In environmental studies, a typical length for the blocks is 1 yr. For example, this overcomes the problem of seasonal trend in the data. From this derives the name of 'annual maximum' method.

$\Pr(M_n \le x) \approx G(x)$, then

$$\Pr(M_n^* \le x) = \Pr(M_n \le (x - b_n)/a_n) =$$
$$= \exp\left\{-\left[1 + \xi\left(\frac{x - \mu^*}{\sigma^*}\right)\right]^{-1/\xi^*}\right\}, \qquad (3)$$

where $\mu^* = a_n\mu + b_n$, $\sigma^* = a_n\sigma$ and $\xi^* = \xi$. This implies that the point estimates of GEV parameters μ^* and σ^*, as inferred by the block-maximum method, depend on the block length, whereas ξ^* is independent (asymptotically). Therefore, the assessment of the stability of the point estimates of ξ^* as block length is varied gives an indication of goodness-of-fit.

The distribution function of the GPD is

$$H(x) = 1 - \left(1 + \frac{\xi y}{\tilde{\sigma}}\right)_+^{-1/\xi}. \qquad (4)$$

Under rather general conditions, the GPD is the limit of the distribution of a random variable X conditionally on exceeding a (high) threshold u. More precisely, by defining, for $y > 0$,

$$F_u(y) = P[X > u + y | X > u], \qquad (5)$$

then for large u the conditional distribution F_u tends to (4). It turns out that the GEV and the GPD are very closely related; Pickands' theorem (Pickands, 1975) states that for any sequence $X_1, X_2, \ldots$ with a given common term X, a GPD arises from the large threshold limit of F_u in (5) if and only if there exist normalizing constants a_n, b_n and a non-degenerate distribution function G such that

$$P[(M_n - b_n)/a_n \le x] \to G(x). \qquad (6)$$

In this case, writing $G(x)$ as in (2), the shape parameter ξ is the same as in the GPD (4), whereas the scale parameters are related by

$$\tilde{\sigma} = \sigma + \xi(u - \mu). \qquad (7)$$

Given a sequence of independent and identically distributed data $x_1, \ldots, x_n$, extreme value modelling with the GPD is achieved by first selecting all exceedances over a sufficiently high threshold u. Suppose that there are n_{exc} ordered exceedances $x_{(1)} \ge x_{(2)} \ge \ldots \ge x_{(n_{\text{exc}})} > u$ and denote the threshold excesses by $y_j = x_{(j)} - u$ for $j = 1, \ldots, n_{\text{exc}}$. Then, the y_j are considered as independent realizations of a random variable whose distribution can be approximated by a member of the GPD family. A suitable choice for the threshold u can be obtained by examining the 'mean excess' plot (also known as mean residual life plot). This is based on the following idea: suppose that the exceedances above a threshold u_0 are 'exactly' GPD-distributed. Then the same holds for all thresholds u larger than u_0 and, by (7), the mean of the excesses above threshold u is given by

$$E[X - u | X > u] = \frac{\sigma_u}{1 - \xi} = \frac{\sigma_{u_0} + \xi(u - u_0)}{1 - \xi}. \qquad (8)$$

An empirical estimate for the mean excess is provided by the sample mean of the threshold excesses; by (8), the latter is ex-pected to vary linearly with u, at least for sufficiently high values of u, for which the GPD approximation is valid. Therefore, suitable threshold values are identified by looking for linearity in the plot of points

$$\left(u, \frac{1}{n_{\text{exc}}} \sum_{i=1}^{n_{\text{exc}}} (x_{(i)} - u)\right), \quad u \in [u_{\min}, u_{\max}]. \qquad (9)$$

Once a value for u has been selected, parameter estimation is carried out by the maximum likelihood method, which also provides confidence intervals. Model checking is performed by a variety of methods: among them, probability, quantile, density and return level plots. Also, the stability of the estimated model can be assessed by examining several fits obtained by varying the threshold. We refer to Coles (2001), Smith (2004) for details. A standard chi-squared test is used to assess goodness-of-fit (Perrin et al., 2006). The number of degrees of freedom of the chi-squared distribution is computed here as the number of classes minus four for the GEV (three for the GPD), to account for the parameters which are estimated from the data by the maximum likelihood. The number of classes may vary between a maximum of 20 and a minimum of 9, to have at least five residual degrees of freedom.

An important feature of the time-series analysed in the present paper is that they possess a non-negligible amount of autocorrelation. In the GEV modelling framework, this typically does not cause serious problems; taking blocks of sufficiently large size yields uncorrelated sequences of block maxima. Therefore, one can apply the GEV method as though the maxima were independent. However, this is not possible in the GPD approach: for time-series with autocorrelation, threshold exceedances tend to occur in 'clusters' and there is no general theory which specifies the joint distribution of neighbouring excesses (and, therefore, the likelihood function). A simple and widely used technique to counter this problem is the usage of so-called 'declustering' and is based on the idea that clusters of exceedances are asymptotically independent (Hsing, 1987). 'Runs declustering' (Leadbetter et al., 1989) consists in marking exceedances as belonging to the same cluster if they are separated by fewer than a fixed number r (the 'run length') of consecutive values below the threshold. Once clusters are identified, the GPD is fitted to the cluster maxima. We emphasize that the selection of the declustering parameter r is largely arbitrary. A more sophisticated declustering scheme, based on the estimation of the extremal index, is proposed by Ferro and Segers (2003). However, for simplicity runs declustering is used in the present paper.

A convenient way to summarize the statistical properties of extreme values is the 'return level'. Given a number p with $0 < p < 1$, the return level associated with the 'return period' $1/p$ is defined as the value z_p that has a probability p to be exceeded by the block maxima of the time-series. In the GEV modelling framework, a maximum likelihood estimator for z_p is obtained by plugging the estimates for (μ, σ, ξ) into the formulae for the

quantiles of $G(x)$, obtained by inverting (2):

$$z_p^* = \begin{cases} \mu^* - \frac{\sigma^*}{\xi^*}\{1 - [-\log(1-p)]^{-\xi^*}\} & \text{for} \xi^* \neq 0, \\ \mu^* - \sigma^* \log[-\log(1-p)] & \text{for} \xi^* = 0. \end{cases} \quad (10)$$

Confidence intervals for z_p may be obtained from those of (μ^*, σ^*, ξ^*) by the 'delta method' (Coles, 2001; Felici et al., 2007a). Notice that the delta method makes explicit use of the form of the likelihood. When referring to the estimates of the GEV parameters or of the return levels, for simplicity we will omit the star, taking for granted the notion that the inferred values may depend on the selected block length.

Similarly, in the GPD modelling approach, the m-observation return level is defined as the level x_m which is exceeded on average every m observations (in this case, $p = 1/m$). Again, the estimated values of (σ, ξ) are plugged into the quantile formula for the GPD, yielding the maximum likelihood estimator for x_m:

$$\hat{x}_m = u + \frac{\hat{\sigma}}{\hat{\xi}}\left[(m\hat{\zeta}_u)^{\hat{\xi}} - 1\right], \quad \hat{\zeta}_u = \frac{n_{\text{exc}}}{n}, \quad (11)$$

where n_{exc} is the number of exceedances and n is the total number of observations. If declustering is applied, however, one needs to use the rate at which clusters occur, rather than the rate of all exceedances: this leads to the formula

$$\hat{x}_m = u + \frac{\hat{\sigma}}{\hat{\xi}}\left[(m\hat{\zeta}_u\hat{\theta})^{\hat{\xi}} - 1\right], \quad \hat{\zeta}_u = \frac{n_{\text{exc}}}{n}, \hat{\theta} = \frac{n_{\text{clu}}}{n_{\text{exc}}}, \quad (12)$$

where $\hat{\theta}$ is an estimate for the extremal index and n_{clu} is the number of observed clusters (Coles, 2001). It is the latter formula that we will use for computation of return levels in the GPD framework.

All computations in this paper have been carried out with the statistical software **R** (**R** Development Core Team, 2008), available at `www.r-project.org` under the GPL license. The libraries `ismev`, which is an **R**-port of the routines written by Stuart Coles as a complement to Coles (2001), and `fExtremes` (both downloadable at `www.cran.r-project.org`) have been used, as well as self-programmed routines.

3. Statistics of wave activity: a general overview

In this section, we first discuss the physical meaning of the WAI and BAI introduced in Section 2.2, by giving a brief historical account. After this, the statistical properties of the time-series generated by the ECHAM model are discussed.

One of the classical problems in understanding the general atmospheric circulation is to characterize the recurrent atmospheric patterns of flow which are observed at mid-latitudes in the Northern Hemisphere winters (Rex, 1950; Baur, 1951; Dole, 1983). The practical need motivating this research effort is twofold: the feasibility of extended range weather forecasts and the detection of climate change (Corti et al., 1999). As a matter of fact, it still is a subject of debate whether the large-scale atmospheric circulation undergoes fluctuations around a sin-

gle equilibrium (Nitsche et al., 1994; Stephenson et al., 2004) or multiple equilibria (Charney and DeVore, 1979; Benzi and Speranza, 1989; Hansen and Sutera, 1995a; Mo and Ghil, 1988; Ruti et al., 2006). Hansen and Sutera (1986) find bimodality in the statistical distribution of time-series of the WAI computed from observed data and orographic resonance theories provide theoretical support to the hypothesis of a multimodal distribution for the activity of planetary waves. Specifically, Charney and DeVore (1979) propose that the interaction between zonal flow and wavefield (via form-drag) explains the occurrence of multiple equilibria for the amplitude of planetary waves. However, in their barotropic theory, the transitions between the two quasi-stable equilibria involve variations of the mean westerlies which are much too large ($\Delta u \approx 40\,\text{ms}^{-1}$) with respect to observations (Malguzzi and Speranza, 1981; Benzi et al., 1986). In Ruti et al. (2006), it has been shown that the bimodality of statistics of the planetary waves is modulated by the intensity of the subtropical jet, in accordance with the theoretical findings of Benzi et al. (1986).

Taking into account these features of the general circulation of the atmosphere, the model simulations gathered by the IPCC possess a reasonable skill in reproducing the high-frequency component of the spectral decomposition, less so for the low-frequency part (Calmanti et al., 2007; Lucarini et al., 2007c). In fact, considering the planetary wave index (WAI), few models reproduce bimodality as in the two reanalyses data sets (Ruti et al., 2006; Calmanti et al., 2007). This highlights a still existing problem for the coupled simulations to reasonably reproduce the low-frequency component of the atmospheric flow. In the case of ECHAM, the AMIP-like experiments show a similar deficiency. The ECHAM4.6 model we are using in this work shows a unimodal probability density function of the planetary wave index (WAI). The results of the IPCC simulation intercomparison (Calmanti et al., 2007; Lucarini et al., 2007c) suggest that the low resolution adopted in the present setup is not the main factor affecting the lack of bimodality.

Figure 1a displays the empirical cumulative distribution functions of the WAI and BAI time-series generated by the ECHAM model in the present setup (see Section 2.1). The corresponding probability density functions, shown in the panel b, are estimated using the kernel technique of Silverman (1986), where the smoothing parameter has been chosen as a Gaussian best-fit for each index. Regarding the first statistical moments of the time-series (see Table 1), we first note that both the mean value and the standard deviation of WAI are much larger than those of BAI. This confirms that a large portion of the planetary-scale activity is concentrated on the spatial scales pertaining to the planetary waves (Dell'Aquila et al., 2005; Lucarini et al., 2007c). Confidence intervals have been computed using a block-bootstrap method which takes care of the time autocorrelation of the time-series (Davison and Hinkley, 1997). Basically we have that the variance of the mean varies as $s/(L/\tau)$, where s is the sample variance, L is the length of the time-series and
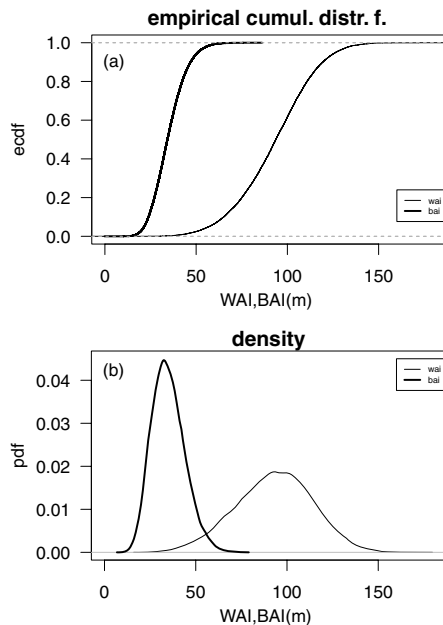
Fig. 1. Empirical cumulative distribution function (a) and probability density function (b) of the time-series of the WAI (thin lines) and BAI (thick lines) indexes. The PDF has been estimated using the kernel technique of Silverman (1986).

Table 1. 95% confidence interval of the mean and standard deviation (both expressed in m) of the WAI and BAI time-series

	Mean (95% confidence interval)	*SD*
WAI, $1CO_2$	93.0 ± 0.9	21.0
BAI, $1CO_2$	35.0 ± 0.3	9.0

τ is its decorrelation time (the latter is estimated by requiring that the linear autocorrelations drop below $1/e$). We have that in both cases $\tau \approx 6$ for the BAI and $\tau \approx 12$ for the WAI time-series.

Our interest in characterizing the extreme value statistics of the WAI and BAI falls in the context mentioned above; specifically we aim at understanding which properties may be relevant to assess climate change and the physical processes involved in it at the level of general atmospheric circulation. We emphasize that the bimodality (or lack thereof) of the time-series does not affect the applicability of the GEV/GPD methods: extreme value analysis aims at characterizing the upper or lower tails of a stochastic process, whereas bimodality takes place in the central part of the probability distribution function. More specifically, one can design a stochastic process having a pre-specified tail behaviour, with any other kind of behaviour (unimodal, bimodal, multimodal, discontinuous) in the central part, so the two are unrelated.

4. Extreme value analysis

In this section, the block-maximum and threshold-exceedance methods described in Section 2.3 are applied to the time-series of WAI and BAI indicators generated from the ECHAM model.

4.1. GEV analysis

In the block-maximum approach, the time-series are subdivided into B data blocks, each containing D daily values, where $B = L/D$ and L is the total length (number of daily observations) of the time-series. Maximum values over each data block are computed, producing the sequences of values from which the GEV parameters (μ, σ, ξ) are estimated. The same procedure is applied to sequences of block minima of the time-series. Since the GEV is a limit distribution, obtained in the limit of L and D going to infinity (see Section 2.3), to avoid bias in the estimates it is not sufficient to take a block length, that is, slightly larger than the decorrelation time scale τ (see Section 3). This implies that $D = 100$ d is likely to be the minimum acceptable block length. To assess the sensitivity and stability of the estimates, we have tried a few values for D: all of them are such that L is exactly divisible by D, to avoid bias due to a different length of the last data block. In Fig. 2, we display the estimated values of ξ (with the related uncertainty) against the value of D used for inference. The estimates of ξ are rather stable after sampling uncertainty is taken into account. Of course, uncertainties increase with D,
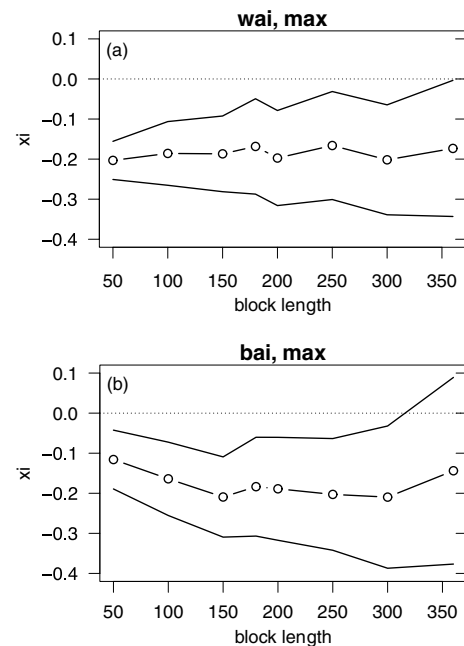

Fig. 2. Estimated values of the shape parameter ξ as a function of the block length D over which maxima are computed, for the WAI (a) and BAI (b) time-series. 95% confidence intervals (average plus and minus two standard deviations) are added.
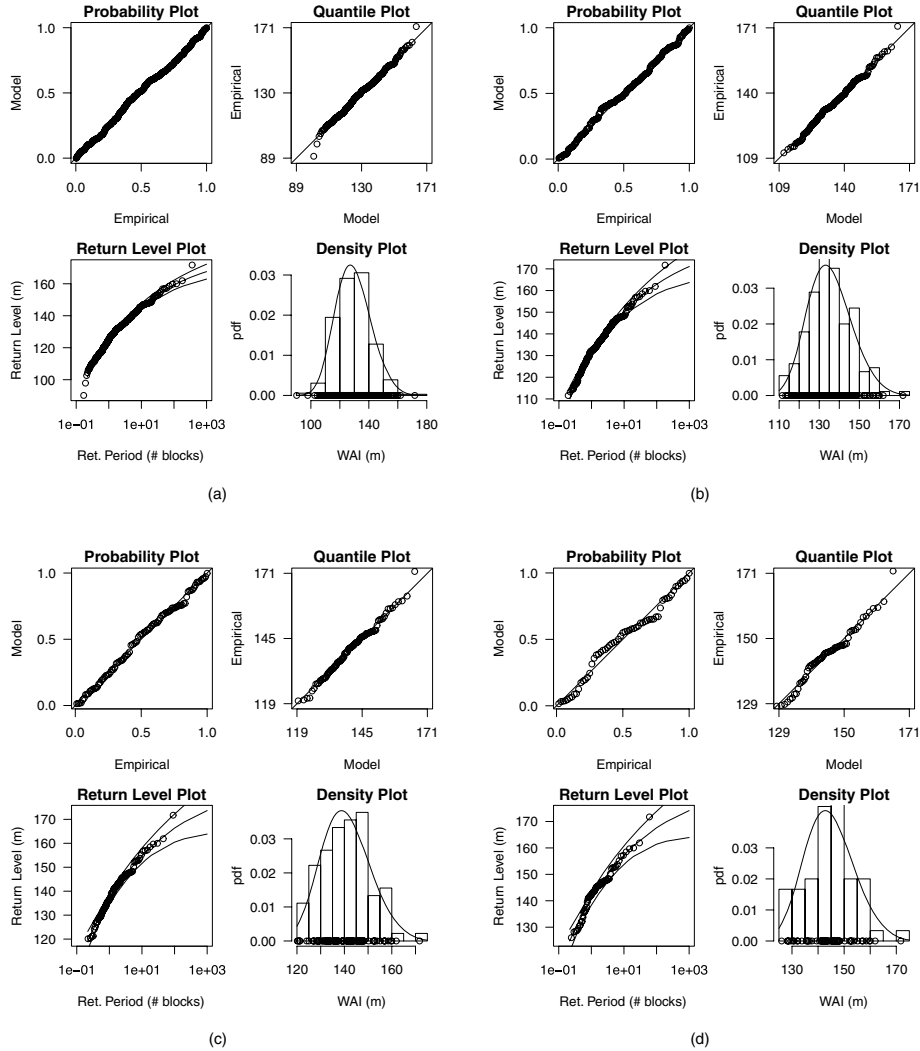
Fig. 3. Diagnostic plots for the GEV fits of the WAI computed with sequences of maxima with block length $D = 50$ (a), 100 (b) 200 (c) and 300 (d).

since one is using less values for the inference. Accordingly, we choose D in such a way that the inferences are reasonably stable and that the associated uncertainties are not too large. There is a good deal of subjectivity in this choice: as usual in GEV-based analysis, one has to adopt a reasonable compromise between long and short data blocks, which boils down to a trade-off between variance and bias (Coles, 2001; Felici et al., 2007a,b).

For the time-series of WAI and BAI, block lengths of $D = 100$ or $D = 200$ seem reasonable: they are large enough to ensure decorrelation of the extreme values and the point estimates of the GEV parameters remain almost constant for $D \geq 100$. The diagnostic plots in Fig. 3 confirm that there is nothing wrong with the inferences obtained for $D = 100, 200, 300$: for example, the displacements of points from the diagonals are relatively small in the probability and quantile plots. More graphical diagnostics (not shown), analogous to those in Figs. 2 and 3, suggest that the

same conclusions also hold for the minima. The choice $D = 50$ is not appropriate for various reasons: first of all, a non-trivial amount of linear autocorrelation persists in the relative sequence of block maxima (figure not shown). Second, the quantile plot reveals lack-of-fit at both ends of the empirical distribution and the confidence intervals of the return level plot look too narrow, indicating that the balance leans towards bias in the bias-versus-variance trade-off discussed above.

Plots of the return levels z_p as functions of the return period (expressed in days) are given in Fig. 4 for the maxima and in Fig. 5 for the minima, for $D = 100$ and 300. For the minima, estimation is performed by first multiplying the time-series by -1, then performing the inference on the maxima, then multiplying again by -1. We observe a very good agreement of the return levels for sufficiently large return periods: this confirms the stability of the inferences with respect to block length.
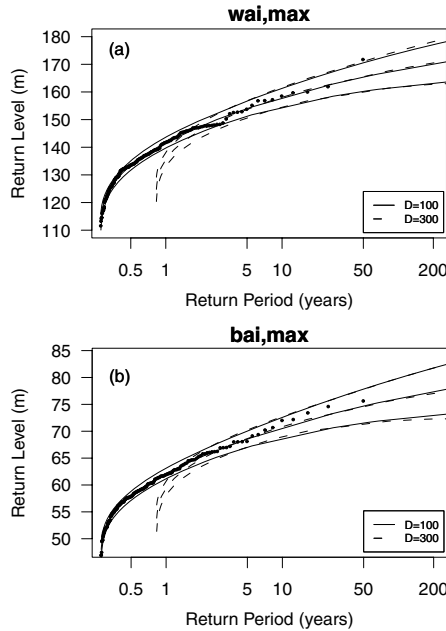
Fig. 4. Return level plots for maxima of the WAI (a) and BAI (b) time-series, using blocks of length $D = 100$ d (solid lines) and $D = 300$ (dashed lines). Time is expressed in years (365 model days) and WAI/BAI in metres.
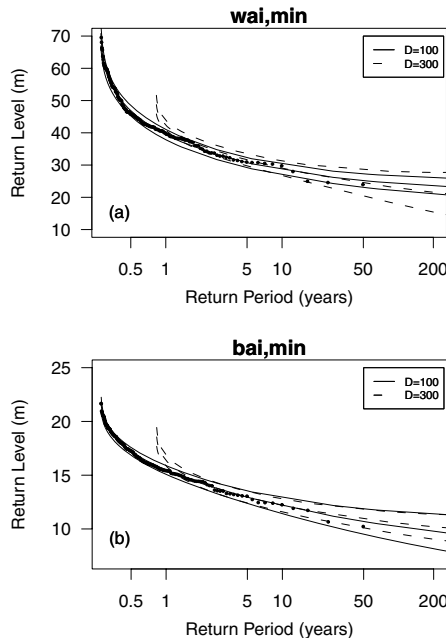

Fig. 5. Same as Fig. 4 for the minima of the same time-series.

However, the estimated uncertainties are in general fairly large: from a 'physical' point of view, a word of caution must be spent here, because a small (in relative terms) variation of absolute value of WAI corresponds to a large variation in the corresponding planetary-scale fields. For example, the upper and lower bounds of the 95% confidence interval for the 200-yr return

level of the WAI maxima differ by roughly 15.5 m. The WAI difference between amplified wave patterns and zonal flow of about 20 m corresponds to local anomalies up to 200–300 m (Hansen and Sutera, 1995b).

4.2. GPD analysis

We now compare the results of the GEV analysis reported in the previous section to those obtained by the GPD approach. Throughout this section, a run-declustering scheme is applied to the data, with $r = 10$; we have verified that for larger values of r (say, up to $r = 20$), the number of identified clusters does not change much, whereas there are larger differences between the cases $r = 0$ and $r = 1, 2$.

The first step is to identify a suitable threshold u for the selection of exceedances. Mean excess plots for the WAI and BAI series are given in Fig. 6 (Fig. 7) for exceedances beyond high (below low) levels. This suggests the following intervals of linearity: [135, 143] and [60, 65] for high levels of WAI and BAI, respectively; [35, 60] and [15, 20] for low levels of WAI and BAI, respectively. For each case, the GPD is fitted to the cluster maxima of the exceedances for a range of thresholds within the above ranges. The resulting estimates of ξ, displayed below the respective mean excess plots, are consistent across the selected range. We have also plotted the estimates obtained with run lengths $r = 3$ and $r = 20$ (dashed and dotted lines, respectively). This shows that there is little sensitivity with respect to run length, particularly so for larger thresholds.

We fix thresholds u of 136 and 60 for high levels of WAI and BAI, respectively. Return level curves computed by (12) are shown in Fig. 8 and reveal an excellent agreement with the GEV estimates obtained for $D = 200$, also see Tables 2 and 3. The thresholds have been chosen within the ranges discussed above, in such a way that the number of identified clusters is not too different from the number of block maxima used for the GEV inference. For $u = 140$, wider confidence intervals are obtained for WAI (shown with dotted lines in Fig. 8a), but the chi-squared test with 20 classes rejects the fit at the 95% level of confidence (the test with 19 classes is passed, with a t-value of 0.07 Table 2). All other tests do not indicate lack-of-fit and using a threshold of 131 yield a plot (not shown) which is very similar to that for $u = 136$.

We also emphasize that the estimated value of the parameter ξ is always negative, corresponding to a Weibull distribution (Coles, 2001), except for a few cases (not shown) where the number of values used for inference is too low and the sampling uncertainty is extremely large (indicating lack-of-fit). The support of Weibull probability density functions is bounded from above, there exists a value $z_\infty = \mu - \sigma/\xi$ which may be considered as a return level with unbounded return period, since values larger than z_∞ form a set having zero probability (the Weibull probability density function is identically zero for those values). The fact that ξ is negative with overwhelming confidence is
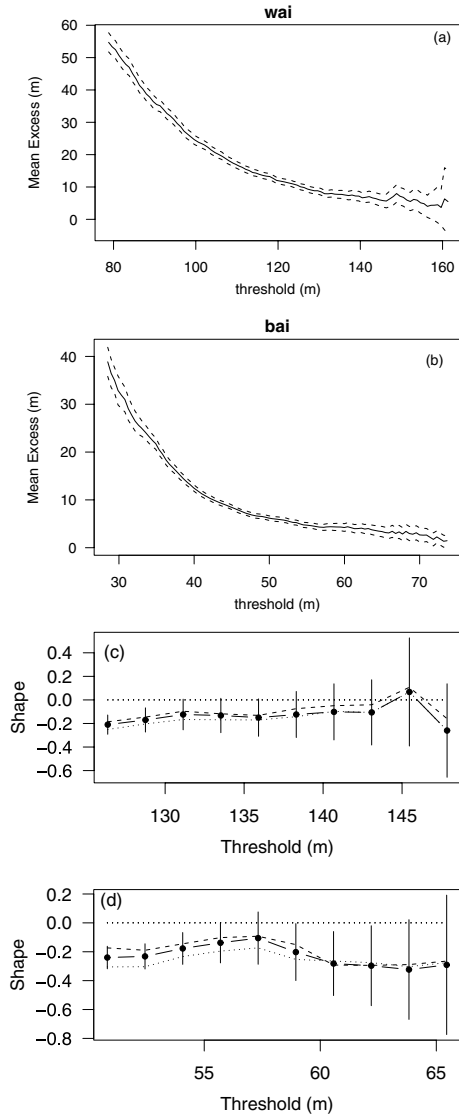
Fig. 6. (a and b) Mean excess plots for high level exceedances of the WAI (a) and BAI (b) time-series. (c and d) Shape parameter ξ of the GPD distribution for the WAI (c) and BAI (d) time-series, estimated from maxima of cluster exceedances for a range of values of the threshold u. A run-declustering algorithm is used with $r = 10$ (solid line and confidence bars), $r = 3$ (dashed lines) and $r = 20$ (dotted lines).

Fig. 7. Same as Fig. 6 for the exceedances below low levels.

compatible with the existence of upper and lower bounds for the considered indexes. In a much simpler model (Felici et al., 2007a), it has been shown that such bounds do exist, providing 'physical constraints' of the system. For the ECHAM model, we interpret this result as follows: in a system with a finite energy input, energy of the perturbations cannot grow indefinitely. Therefore, the probability density function of the (extreme) perturbations is bounded and the shape parameter is negative. This theory is fully compatible with all of the estimates which we have obtained from the simulation data. For more common physical
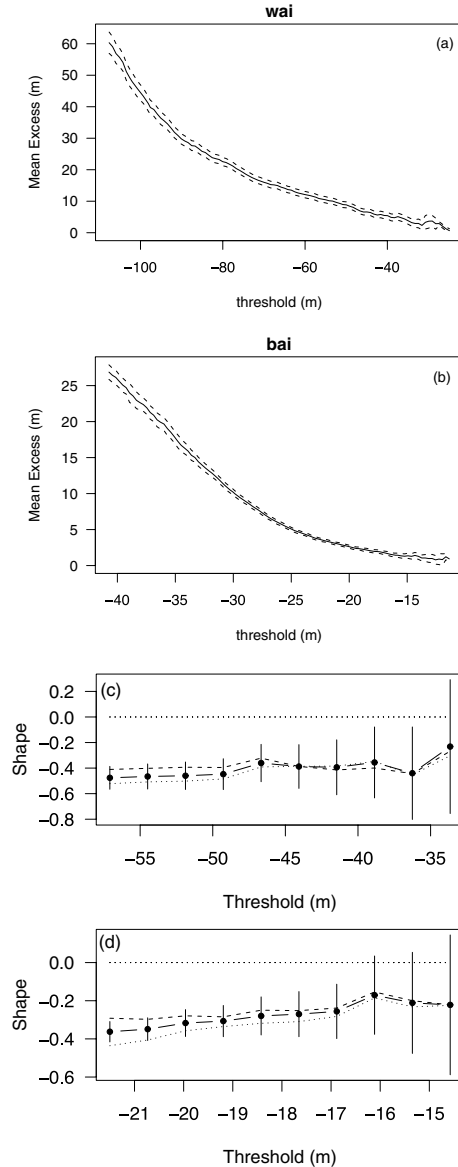
variables, wind speeds in both extratropical latitudes (Perrin et al., 2006) and in tropical storms and hurricanes (Neumann, 1987; Elsner et al., 2008) are often observed to be Weibull distributed, whereas hydrological variables such as precipitation often display heavy tails (with a positive value of the shape parameter), see Coles (2001), Felici et al. (2007a), Smith (2004) for more references.

4.3. On the possibility of slow convergence of GEV estimators

In this section, we analyse the sensitivity of the extreme value statistical properties with respect to changes in model
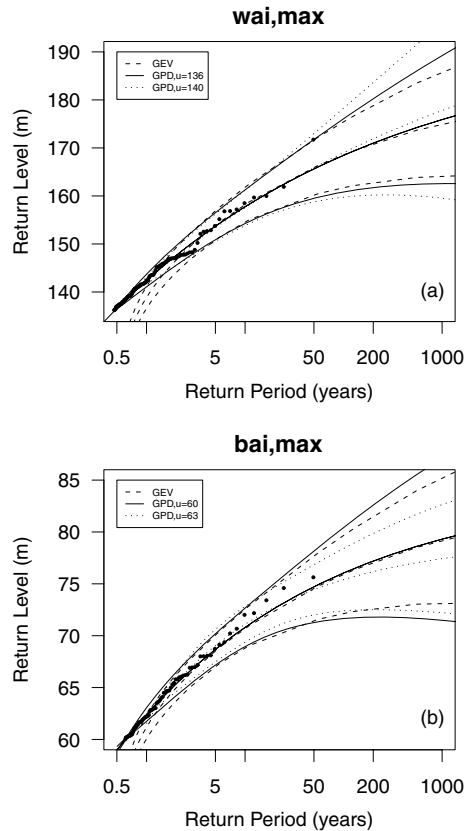
wai,max

bai,max

Fig. 8. Joint GEV–GPD return level plots for the maxima-high level exceedances of WAI (a) and BAI (b). The GEV maximum likelihood estimates and 95% confidence intervals are plotted with dashed lines, whereas solid and dotted lines are used for GPD fits obtained with two different threshold values. Empirical estimates (points) are added for the GPD fit lower threshold. The estimated parameter values and other information concerning the inferences are reported in Table 2 for WAI and in Table 3 for BAI.

Table 2. Summary of the GEV (top row) and GPD (bottom row) fits corresponding to the inference in Fig. 8a (the GPD fits with $r = 3$ and that with $u = 131$ are not shown in the plot). Standard errors for the parameters are reported inside parentheses, whereas nd is the number of degrees of freedom for the chi-squared goodness-of-fit test

μ	σ	ξ	B	D	t-value, nd
136.7 (1.1)	9.8 (0.8)	−0.20 (0.06)	90	200	0.75 (15)

u	$\tilde{\sigma}$	ξ	n_{clu}	r	
136	8.3 (1.0)	−0.14 (0.07)	121	3	0.15 (16)
136	9.0 (1.1)	−0.16 (0.08)	105	10	0.18 (16)
140	7.7 (1.3)	−0.12 (0.11)	68	10	0.1 (15)
131	9.1 (0.9)	−0.13 (0.07)	187	10	0.39 (16)

Table 3. Same as Table 2 for the inference Fig. 8b. The GPD fit with $r = 3$ is not shown in the plot

μ	σ	ξ	B	D	t-value (nd)
60.1 (0.5)	4.8 (0.4)	−0.19 (0.07)	90	200	0.33 (15)

u	$\tilde{\sigma}$	ξ	n_{clu}	r	
60	5.1 (0.8)	−0.22 (0.11)	82	3	0.84 (16)
60	5.1 (0.8)	−0.21 (0.11)	81	10	0.85 (16)
63	5.3 (1.1)	−0.33 (0.16)	41	10	0.47 (16)

configuration. Two runs of the ECHAM model, with CO_2 concentration equal to five times the present-day value (1800 ppmv) are carried out. For simplicity, the SSTs are kept the same as in the $1CO_2$ run examined in the rest of this paper. We emphasize that this rules out any interpretation of the results of the present section in terms of climatic change, based on the comparison with the $1CO_2$ run. For this reason, such a comparison is not even attempted. The present experiments aim exclusively at getting more information on the sensitivity of the statistical estimates based on data generated by the ECHAM model, to establish the experimental requirements for studies of climatic change based on the present methodology.

The two $5CO_2$ runs only differ from each other in the initial condition, the remaining parameters and settings being identical to those described in Section 2.1. For this reason, time-series extracted from the two runs should be considered as different realizations of the same stochastic process and should pos-

sess the same statistical properties. GEV analysis is carried out on time-series of WAI and BAI for these $5CO_2$ runs (to distinguish them we refer to the second one as $5CO_2$Bis). For the maxima of the WAI, one obtains huge differences in the point estimates of the return levels and even much larger variations if confidence intervals are taken into account, see Fig. 9a. The point estimates of the 1000-yr return levels of the WAI differ by roughly 15 m (an idea of the physically relevance of such a value is given in a remark at the end of Section 4.1). A similar discrepancy is also found using the GPD method, although a smaller uncertainty is obtained (Fig. 9b). This suggests the following scenario: even under identical experimental setup, except for the initial condition from which the model is started, one may obtain quantitatively different point estimates for the return values. Now in the experimental setting which we have considered, the wide confidence interval helps us reject the hypothesis that the two populations are different. However, if shorter data blocks are used, then narrower confidence intervals might be obtained. In this case, the different experiments would obtain different results at the same level of confidence.

These discrepancies are largely accounted for by a few exceptionally large values of the WAI in the $5CO_2$Bis series: for example, by taking away the two largest block maxima of $5CO_2$Bis, the difference in the return level plots is much smaller (Fig. 9c). Moreover, the remaining block maxima of $5CO_2$Bis are smaller then the maximum of $5CO_2$. Similar considerations hold for the GPD method, see Fig. 9d. This suggests that there exist states of
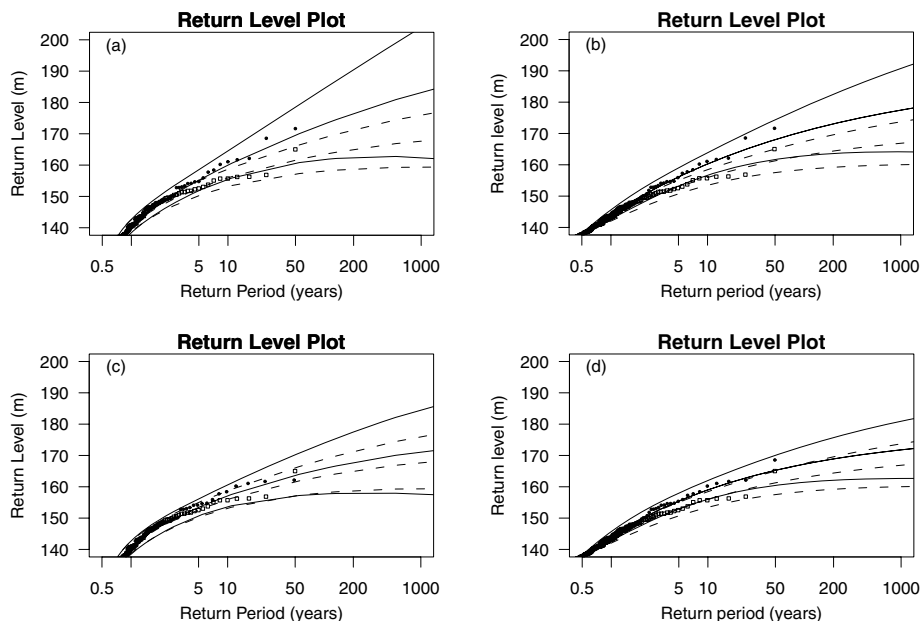
Fig. 9. (a) Return level plot obtained by the GEV method from the maxima of two distinct $5CO_2$ runs of the ECHAM model, see text for the details. (c) Same as (a), where the two largest block maxima of the $5CO_2Bis$ series have been removed. (b and d) Same as (a and c), respectively, for a GPD fit of the same time-series; threshold values for (b and d) are equal to those used for Figs. 8 (a) and (b), respectively.

remarkable WAI amplitude which are visited very rarely by the ECHAM model; it appears that to include such states in the analysis, time-series of great length are necessary. This requirement is likely to be even more stringent when simulations by fully coupled AOGCMs with seasonal cycle are examined. This phenomenon might bear relation to the slowness of convergence of the GEV estimators which has been found in a quasi-geostrophic model (Vannitsem, 2007) but also in a stochastic volatility process with long-memory modelling stock returns (Malevergne et al., 2006). We believe that assessing the speed of convergence of the extreme value distribution estimators and its possible relation to long-memory processes (Maraun et al., 2004; Eichner et al., 2006) is of remarkable scientific and, in view of the envisaged applications, societal importance.

5. Mapping the extremes

In this section, we examine the ensembles of 500 hPa geopotential height maps constructed by choosing the fields corresponding to the dates of the maxima or minima extracted from the time-series of the WAI and BAI. For each index, maxima and minima are computed over blocks of length $D = 100$ d (see Section 4). For each of the four cases, we analyse the ensemble mean of the eddy field, obtained by removing the zonal average in each map, and the map of the standard deviation, computed gridpoint by gridpoint.

Figure 10 shows the ensemble eddy mean and standard deviation of the maxima and of the minima of WAI and BAI. The composite average 500 hPa map constructed by selecting

the dates of extreme (block maxima) values of WAI (panel a) shows, as expected, an enhanced eddy field, with stronger positive centre over the Rockies and deepening of the negative centre over the Labrador with respect to the mean eddy field obtained by averaging over the whole time-series (not shown). Such a wavy pattern corresponds to a ridge over the Rockies and diffluent flow over Europe. Hansen et al. (1993) observed that this amplified planetary-wave regime corresponds to Pacific Blocking episodes. When considering the standard deviation of the 500 hPa map, we observe two main peaks over the north-central Atlantic and over the north-western Pacific, which correspond to the final part of the two oceanic storm-tracks, related to the jet exit region. For the same model, Carillo et al. (2000) found strong interaction between high frequency and low frequency over these regions.

As opposed to the previous case, the mean eddy pattern obtained by selecting the dates of maxima of BAI Fig. 10b shows an eddy pattern relatively similar to the mean eddy pattern (figure not shown), where an eastward shift over the Rockies and a northward shift of the Euro-Atlantic positive anomaly are detected. When considering the standard deviation map, large values are observed in the storm-tracks regions (Dell'Aquila et al., 2005).

The ensemble mean of the eddy field obtained by considering the days of extremely low value of WAI (panel c) features the prominence of wavenumber 1, which is connected to a relatively widespread zonal flow. When considering the standard deviation of the field, centres are present also in this case in the exit region of the Pacific jet and over the northern Atlantic, even if the
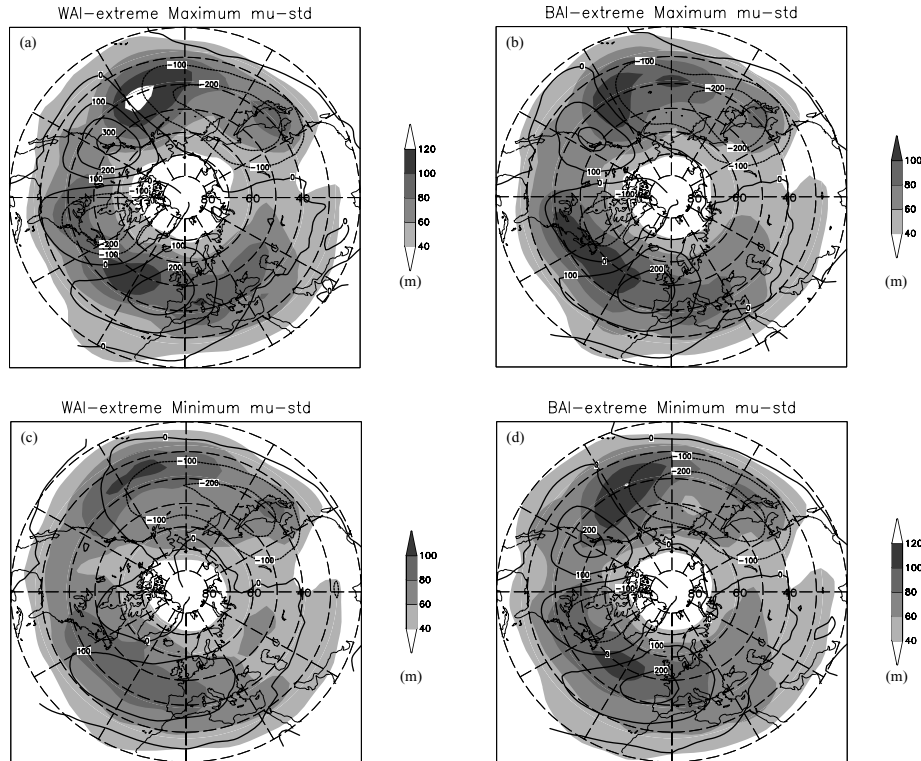
Fig. 10. The eddy ensemble mean (contour) and the standard deviation (shaded) of the 100-d maxima of WAI (a) and BAI (b) and of the 100-d minima of WAI (c) and BAI (d).

variability is much lower than what depicted in Fig. 10. As for the maxima, the ensemble mean of the eddy field obtained by selecting the extremely low BAI events Fig. 10d is similar to the mean eddy pattern (figure not shown), while the two standard deviation maps feature larger variance in correspondence with the Pacific and Atlantic jet exit regions, but with reduced variability with respect to what shown in Fig. 10.

6. Summary and conclusions

In this paper, we have proposed a methodology for the analysis of extreme atmospheric wave activity at mid-latitudes. The methodology has been applied to various simulations obtained by the General Circulation Model ECHAM4.6. The wave indexes we have examined can be considered as proxies of planetary-scale activity at different spatial and temporal scales. The mid-latitude dynamics features upper planetary and synoptic waves as typical ingredients. Since synoptic and planetary waves provide a relevant contribution to the meridional transport of energy and momentum in the mid-latitudes, the physical processes driving such phenomena are sensible ingredients for the characterization of the climate system (Speranza, 1983). The statistical behaviour of extreme values of the various time-series is studied by the block-maximum method (fitting GEV distributions) and by the

POT method (fitting GPDs). The statistical significance of these results has been thoroughly assessed by a variety of diagnostic tools. The analysis of extreme values is complemented by an examination of composite maps of the extremes of the maxima, to highlight the spatial patterns.

Rather large uncertainties are obtained here, both for the GEV parameters and for the return levels. It might be questioned whether there are problems of convergence for the GEV estimator. In this sense, our results bear some resemblance with those by Vannitsem (2007), who finds that extremely long time-series are necessary to reach satisfactory convergence of the statistical estimators in a quasi-geostrophic model. The fact that such a large amount of uncertainty is obtained for the extreme value statistics even in the present, idealized simulations (no seasonal cycle) suggests the possibility that time-series of almost prohibitive length might be necessary for accurate estimations in a more realistic setup, such as an atmospheric GCM coupled with a slab ocean or even a fully coupled AOGCM with seasonal cycle. Of course, this problem is likely to be even more serious when dealing with observed data or reanalyses, as sensitivity studies with respect to length of the time-series cannot be carried out. As expected, usage of the GPD-based approach may yield a somewhat reduced uncertainty. However, in our study of two stationary runs which only differ in the intial condition

(and, therefore, may be considered as two realizations of the same stochastic process), a marked difference in the two point estimates is obtained.

The analysis of the composite maps allow us to feature the spatial patterns associated with the extremes in low and high wavy regimes. The maxima in the WAI shows enhanced eddy field, with stronger positive centre over the Rockies and deepening of the negative centre over the Labrador. The corresponding standard deviation features high amplitudes, where strong interaction between high frequency and low frequency is observed. Considering the maxima of the BAI index, the standard deviation pattern is quite similar to what observed for the WAI, but stronger on the American east coast where the Atlantic storm-tracks origin. The eddy pattern is not enhanced as in the WAI maxima: this could suggest that the link between planetary and baroclinic waves is not linear. The analysis of the relationship between high and low frequencies will be further developed in a study with AMIP-like simulations forced by SST from a global scenario coupled simulation.

We conclude by remarking that, for the general strategy of setting up model comparisons, the statistical nature of the selected diagnostic indicators might be of essential importance. We believe that the indicators of planetary-scale activity considered in the present paper and, particularly, the extreme statistics of such indicators, can be used for model intercomparison as a complement to the surface parameters (such as temperature, precipitation, wind) typically examined in research on extremes in climate models.

7. Acknowledgments

References

Baur, F. 1951. Extended range weather forecasting. In: *Compendium of Meteorology*. Amer. Meteor. Soc., Boston, Massachussets, 814–833.

Beniston, M., Stephenson, D. B., Christensen, O. B., Ferro, C. A. T., Frei, C. and co-authors. 2007. Future extreme events in European climate: an exploration of regional climate model projections. *Clim. Change* **81**(Suppl. 1), 71–95.

Benzi, R. and Speranza, A. 1989. Statistical properties of low frequency variability in the Northern Hemisphere. *J. Clim.* **2**, 367–379.

Benzi, R., Malguzzi, P., Speranza, A. and Sutera, A. 1986. The statistical properties of general atmospheric circulation. Observational evidence and a minimal theory of bimodality. *Quart. J. Roy. Meteorol. Soc.* **112**, 661–674.

Blackmon, M. L. 1976. A climatological spectral study of the 500 mb geopotential height of the Northern Hemisphere. *J. Atmos. Sci.* **33**, 1607–1623.

Bordi, I., Dell'Aquila, A., Speranza, A. and Sutera, A. 2004. On the mid-latitude tropopause height and the orographic-baroclinic adjustment theory. *Tellus* **56A**, 278–286, doi:10.1111/j.1600-0870.2004.00065.x.

Buzzi, A., Trevisan, A. and Speranza, A. 1984. Instabilities of a baroclinic flow related to topographic forcing. *J. Atmos. Sci.* **41**, 637–650.

Calmanti, S., Dell'Aquila, A., Lucarini, V., Ruti, P. M. and Speranza, A. 2007. Statistical properties of mid-latitude atmospheric variability. In: *Nonlinear Dynamics in Geosciences* (eds. A. Tsonis and J. Elsner). Springer, (New York, USA), ISBN 978-0-387-34917-6.

Carillo, A., Ruti, P. M. and Navarra, A. 2000. Storm tracks and mean flow variability: a comparison between observed and simulated data. *Clim. Dyn.* **16**, 218–229.

Charney, J. G. and DeVore, J. G. 1979. Multiple flow equilibria in the atmosphere and blocking. *J. Atmos. Sci.* **36**, 1205–1216.

Charney, J. G. and Straus, D. M. 1980. Form-drag instability, multiple equilibria and propagating planetary waves in baroclinic, orographically forced, planetary wave systems. *J. Atmos. Sci.* **37**, 1157–1176.

Coles, S. 2001. *An Introduction to Statistical Modelling of Extremes Values*. Springer-Verlag, Berlin and Heidelberg, 208 pp.

Corti, S., Molteni, F. and Palmer, T. N. 1999. Signature of recent climate change in frequencies of natural atmospheric circulation regimes. *Nature* **398**, 799–802.

Davison, A. C. and Hinkley, D. V. 1997. *Bootstrap Methods and Their Application*. Cambridge University Press, Cambridge.

Dell'Aquila, A., Lucarini, V., Ruti, P. M. and Calmanti, S. 2005. Hayashi spectra of the Northern Hemisphere mid-latitude atmospheric variability in the NCEP-NCAR and ECMWF reanalyses. *Clim. Dyn.*, **25**, 639–652.

Dell'Aquila, A., Ruti, P. M. and Sutera, A. 2007. Effects of the baroclinic adjustment on the tropopause in the NCEP-NCAR reanalysis. *Clim. Dyn.* **28**, doi:10.1007/s00382-006-0199-4.

Dole, R. M. 1983. Persistent anomalies of the extratropical Northern Hemisphere wintertime circulation. In: *Large-Scale Dynamical Processes in the Atmosphere*, (eds. B. J. Hoskins and R. P. Pearce). Elsevier, New York, 95–109.

Eichner, J. F., Kantelhardt, J. W., Bunde, A. and Havlin, S. 2006. Extreme value statistics in records with long-term persistence. *Phys. Rev. E* **73**, 016130.

Elsner, J. B., Jagger, T. H. and Liu, K. b. 2008. Comparison of Hurricane return levels using historical and geological records. *J. Appl. Meteorol. Climatol.* **47**, 368–374.

Felici, M., Lucarini, V., Speranza, A. and Vitolo, R. 2007a. Extreme value statistics of the total energy in an intermediate complexity model of the mid-latitude atmospheric jet. Part I: stationary case. *J. Atmos. Sci.* **64**, 2137–2158.

Felici, M., Lucarini, V., Speranza, A. and Vitolo, R. 2007b. Extreme value statistics of the total energy in an intermediate complexity model of the mid-latitude atmospheric jet. Part II: trend detection and assessment. *J. Atmos. Sci.* **64**, 2159–2175.

Ferro, C. A. T. and Segers, J. 2003. Inference for clusters of extreme values. *J. R. Stat. Soc.* **65B**, 545–556.

Fouquart, Y. and Bonnel, B. 1980. Computations of solar heating of the Earth's atmosphere. A new parametrization. *Contrib. Phys. Atmos.* **53**, 35–62.

Friederichs, P. and Hense, A. 2007. Statistical downscaling of extreme precipitation events using censored quantile regression. *Mon. Wea. Rev.* **135**, 2365–2378.

Hansen, A. R. and Sutera, A. 1986. On the probability density distribution of planetary-scale atmospheric wave amplitude. *J. Atmos. Sci.* **43**, 3250–3265.

Hansen, A. R. and Sutera, A. 1995a. The probability density distribution of planetary-scale atmospheric wave amplitude revisited. *J. Atmos. Sci.* **52**, 2463–2472.

Hansen, A. R. and Sutera, A. 1995b. Large amplitude flow anomalies in Northern Hemisphere mid-latitudes. *J. Atmos. Sci.* **52**, 2133–2151.

Hansen, A., Pandolfo, J. and Sutera, A. 1993. Midtropospheric flow regimes and persistent wintertime anomalies of surface layer pressure and temperature. *J. Clim.* **6**, 2136–2143.

Hsing, T. 1987. On the characterization of certain point processes. *Stochast. Process. Appl.* **26**, 297–316.

Kharin, V. V. and Zwiers, F. W. 2000. Changes in the extremes in an ensemble of transient climate simulations with a coupled atmosphere-ocean GCM. *J. Clim.* **13**, 3760–3788.

Kharin, V. V., Zwiers, F. W., Zhang, X. and Hegerl, G. C. 2007. Changes in temperature and precipitation extremes in the IPCC ensemble of global coupled model simulations. *J. Clim.* **20**, 1419–1444.

Leadbetter, M. R., Weissman, I., de Haan, L. and Rootzén, H. 1989. On clustering of high values in statistically stationary series. *Technical Report 253*. Center for Stochastic Processes, University of North Carolina, Chapel Hill.

Lucarini, V., Speranza, A. and Vitolo, R. 2007a. Parametric smoothness and self-scaling of the statistical properties of a minimal climate model: what beyond the mean field theories? *Physica D* **234**, 105–123.

Lucarini, V., Speranza, A. and Vitolo, R. 2007b. Self-scaling of the statistical properties of a minimal model of the atmospheric circulation. In: *Nonlinear Dynamics in Geosciences* (eds. A. Tsonis and J. Elsner). Springer, New York, USA, ISBN 978-0-387-34917-6.

Lucarini, V., Calmanti, S., Dell'Aquila, A., Ruti, P. M. and Speranza, A. 2007c. Intercomparison of the northern hemisphere winter mid-latitude atmospheric variability of the IPCC models. *Clim. Dyn.* **28**, 829–849, doi:10.1007/s00382-006-0213-x.

Malevergne, Y., Pisarenko, V. and Sornette, D. 2006. On the power of generalized extreme value (GEV) and generalized Pareto distribution (GPD) estimators for empirical distributions of stock returns. *Appl. Financ. Econ.* **16**, 271–289.

Malguzzi, P. and Speranza, A. 1981. Local multiple equilibria and regional atmospheric blocking. *J. Atmos. Sci.* **38**, 1939–1948.

Maraun, D., Rust, H. W. and Timmer, J. 2004. Tempting long-memory—on the interpretation of DFA results. *Nonlin. Processes Geophys.* **11**, 495–503.

Miller, M. J., Palmer, T. N. and Swinbank, R. 1989. Parameterization and influence of sub-grid scale orography in general circulation and numerical weather prediction models. *Meteorol. Atmos. Phys.* **40**, 84–109.

Mo, K. and Ghil, M. 1988. Cluster analysis of multiple planetary flow regimes. *J. Geophys. Res.* **93**, 10 927–10 952.

Morcrette, J. J. 1991. Radiation and cloud radiative properties in the European Center for Medium Range Forecasts forecasting system. *J. Geophys. Res.* **96**, 9121–9132.

Morcrette, J. J. and Jakob, C. 2000. The response of the ECMWF Model to changes in the cloud overlap assumption. *Mon. Wea. Rev.* **128**, 1707–1732.

Guilyardi, E., Delecluse, P., Gualdi, S. and Navarra, A. 2002. Mechanisms for ENSO phase change in a coupled GCM. *J. Clim.* **16**, 1141–1158.

Neumann, C. J. 1987. The national hurricane center risk analysis program (HURISK). NOAA/National Hurricane Center *Technical Memo. NWS NHC 38*, 59 pp.

Nitsche, G., Wallace, J. M. and Kooperberg, C. 1994. Is there evidence of multiple equilibria in planetary wave amplitude statistics? *J. Atmos. Sci.* **51**, 314–322.

Nordeng, T. E. 1994. Extended versions of the convective parameterization scheme at ECMWF and their impact on the mean and transient activity of the model in the tropics. ECMWF, *Technical Memo n. 206*, Reading, England, 41 pp.

Perrin, O., Rootzén, H. and Taesler, R. 2006. A discussion of statistical methods used to estimate extreme wind speeds. *Theor. Appl. Climatol.* **85**, 203–215.

Pickands, J. 1975. Statistical inference using extreme order statistics. *Ann. Statist.* **3**, 119–131.

R Development Core Team 2008. ***R**: A Language and Environment for Statistical Computing*, **R** Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0, available at http://www.R-project.org.

Resnick, S. I. 1987. *Extreme Values, Regular Variation, and Point Processes*, Springer Series in Operations Research and Financial Engineering, 2nd printing, 2008, 320 pp, ISBN 978-0-387-75952-4.

Rex, D. F. 1950. Blocking action in the middle troposphere and its effect upon regional climate. Part 2: the climatology of blocking action. *Tellus* **2**, 275–301.

Roeckner, E. 1995. Parameterization of cloud radiative properties in the ECHAM4 model. In: *Proceedings of the WCRP Workshop: cloud microphysics parameterizations in global circulation models*, *WCRP Report n. 93, WMO/TD-no 713*, 105–116.

Roeckner, E. and Arpe, K. 1995. AMIP experiments with the new Max Planck Institute for Meteorology Model ECHAM4. In: *Proceedings of the "AMIP Scientific Conference", May 15-19, 1995*, Monterey, USA, WCRP-Report No. 92, 307-312, WMP/TD-No. 732.

Ruti, P. M., Lucarini, V., Dell'Aquila, A., Calmanti, S. and Speranza, A. 2006. Does the subtropical jet catalyze the midlatitude atmospheric regimes? *Geophys. Res. Lett.* **33**, L06814, doi:10.1029/2005GL024620.

Silverman, B. W. 1986. *Density Estimation for Statistics and Data Analysis*. Chapman & Hall, London.

Simmons, A. J., Burridge, D. M., Jarraud, M., Girard, C. and Wergen, W. 1989. The ECMWF medium-range prediction models. Development of the numerical formulations and the impact of increased resolution. *Meteorol. Atmos. Phys.* **40**, 28–60.

Smith, R. L. 1985. Maximum likelihood estimation in a class of non-regular cases. *Biometrika* **72**, 67–92.

Smith, R. L. 2004. Statistics of extremes, with applications in environment, insurance and finance. In: *Extreme Values in Finance, Telecommunications and the Environment* (eds. B. Finkenstadt and H. Rootzén). Chapman and Hall/CRC Press, London, 1–78.

Speranza, A. 1983. Deterministic and statistical properties of the Westerlies. *Paleogeophysics* **121**, 512–562.

Stephenson, D. B., Hannachi, A. and O'Neil, A. 2004. On the existence of multiple climate regimes. *Q. J. R. Meteorol. Soc.* **130**, 583–605.

Tiedke, M. 1989. A comprehensive mass flux scheme for cumulus parameterization in large-scale models. *Mon. Wea. Rev.* **117**, 1779–1800.

Vannitsem, S. 2007. Statistical properties of the temperature maxima in an intermediate order Quasi-Geostrophic model. *Tellus* **59A**, 80–95. doi:10.1111/j.1600-0870.2006.00206.x

Wallace, J. M., Lim, G. and Blackmon, M. L. 1988. Relationship between cyclone tracks, anticyclone tracks and baroclinic waveguides. *J. Atmos. Sci.* **45**, 439–462.

Williamson, D. L. and Rasch, P. J. 1994. Water vapour transport in the NCAR CCM2. *Tellus* **46A**, 34–51.

Zwiers, F. W. and Kharin, V. V. 1998. Changes in the extremes of the climate simulated by CCC GCM2 under CO_2 doubling. *J. Clim.* **11**, 2200–2222.