Extreme Value Statistics of the Total Energy in an Intermediate-Complexity Model of the Midlatitude Atmospheric Jet. Part II: Trend Detection and Assessment

MARA FELICI

PASEF, Dipartimento di Matematica ed Informatica, Università di Camerino, Camerino, and Dipartimento di Matematica U. Dini, Università di Firenze, Florence, Italy

VALERIO LUCARINI, ANTONIO SPERANZA, AND RENATO VITOLO

PASEF, Dipartimento di Matematica ed Informatica, Università di Camerino, Camerino, Italy

(Manuscript received 1 December 2006, in final form 27 February 2007)

ABSTRACT

A baroclinic model for the atmospheric jet at middle latitudes is used as a stochastic generator of nonstationary time series of the total energy of the system. A linear time trend is imposed on the parameter T_E, descriptive of the forced equator-to-pole temperature gradient and responsible for setting the average baroclinicity in the model. The focus lies on establishing a theoretically sound framework for the detection and assessment of trend at extreme values of the generated time series. This problem is dealt with by fitting time-dependent generalized extreme value (GEV) models to sequences of yearly maxima of the total energy. A family of GEV models is used in which the location μ and scale parameters σ depend quadratically and linearly on time, respectively, while the shape parameter ξ is kept constant. From this family, a GEV model is selected with Akaike's information criterion, complemented by the likelihood ratio test and by assessment through standard graphical diagnostics. The inferred location and scale parameters are found to depend in a rather smooth way on time and, therefore, on T_E. In particular, power-law dependences of μ and σ on T_E are obtained, in analogy with the results of a previous work where the same baroclinic model was run with fixed values of T_E spanning the same range as in this case. It is emphasized under which conditions the adopted approach is valid.

1. Introduction

In the context of climate change, an intensely debated question is whether the statistics of extreme meteo-climatic events is changing (and/or will change) and, in case, how fast it is changing (and/or will change). For example, the role of time dependence in the statistics of extreme weather events has been at the heart of discussions about climate change since the work by Katz and Brown (1992). In particular, the detection of trends in the frequency of intense precipitation has been the object of much research, particularly at regional level [see, e.g., Karl et al. (1996), Karl and Knight (1998), for the United States and Brunetti et al. (2002, 2004) for the Mediterranean area]. The general relevance of the problem has been highlighted in the 2002 release of a specific Intergovernmental Panel on Climate Change (IPCC) workshop report on changes in extreme weather and climate events (available at http://www.ipcc.ch/pub/extremes.pdf). In fact, the emphasis laid on the subject by the IPCC report in many countries spread the question "Is the probability of major impact weather increasing?" This question reached the general public almost everywhere and innumerable studies of trends in series of "extremes" were undertaken. These studies mainly deal with variables of local character, typically precipitation and temperature at specific stations. Moreover, most studies are regional: see, for example, the proceedings of the Italy–United States meeting held in Bologna in 2004 (Diaz and Nanni 2006) for the relevance of the extreme events in the Mediterranean Climates and the INTERREG IIIB Central, Adriatic, Danubian, and Southeastern Euro-

Corresponding author address: Dr. Renato Vitolo, Department of Mathematics and Informatics, University of Camerino, via Madonna delle Carceri, 62032 Camerino (MC), Italy.
E-mail: renato.vitolo@unicam.it

DOI: 10.1175/JAS4043.1

pean Space (CADSES) hydrological cycle of the CADSES regions (HYDROCARE; http://www.hydrocare-cadses.net) for impacts of extreme events in the hydrological cycle of the central eastern Europe.

In a companion paper (Felici et al. 2007, hereafter Part I) we have addressed the problem of extreme value statistical inference on statistically stationary time series produced by a dynamical system providing a minimal model for the dynamics of the midlatitudes baroclinic jet. There reported is, from mathematical literature, a suitable, rigorous, universal setting for the analysis of the extreme events in stationary time series. This is based on Gnedenko's theorem (Gnedenko 1943) according to which the distribution of the block maxima of a sample of independent identically distributed variables converges, under fairly mild assumptions, to a member of a three–real parameter family of distributions, the so-called generalized extreme value (GEV) distribution (Coles 2001). The GEV approach to the analysis of extremes requires that three basic conditions are met, namely the independence of the selected extreme values, the consideration of a sufficiently large number of extremes, the selection of values that are genuinely extreme. These conditions could be achieved relatively easily for the case at hand.

Part I was originally motivated by the interest in weather having major impact (on human life and property) in the Mediterranean area, in particular, intense precipitation and heat waves over Italy. [See, e.g., Brunetti et al. (2002, 2004), Lucarini et al. (2004, 2006), Speranza et al. (2006), Speranza and Tartaglione (2006), Tartaglione et al. (2006), and the Mediterranean Experiment (MEDEX) phase 1 report (available at http://medex.inm.uib.es/) for related results and activities.] The study reported in Part I revealed, among other things, that diagnostics of extreme statistics can highlight interesting dynamical properties of the analyzed system. Properties that, thanks to the universality of the GEV, can be investigated in a low dimensionality space of parametric probability density functions, although at the expenses of the total number of events available in order to capture a sufficient number of independent extremes. A key role (that is presently being explored elsewhere, in the context of general atmospheric circulation theory) was played in Part I by the smoothness of variation of the extreme statistics parameters (average, variance, shape factor) upon the external (forcing) parameters of the system. In this paper, again, we devote attention to exploring the statistics of extremes as a dynamical indicator, this time in the framework of the (typically meteorological) statistical inference problem of detecting trends in observations.

The definition of a rigorous approach to the study of extremes is much harder when the property of stationarity does not hold. One basic reason is that there exists no universal theory of extreme values (such as, e.g., a generalization of Gnedenko's theorem) for nonstationary stochastic processes. Moreover, in the analysis of observed as well as synthetically generated sequences of data of finite length, practical issues, such as the possibility of unambiguously choosing the time scales that defines the statistical properties and their changes, become of critical importance. Nevertheless, GEV-based statistical modeling offers a practical unified framework also for the study of extremes in nonstationary time series. In the applications, the three parameters of the GEV distribution are taken as time-dependent and time is introduced as a covariate in the statistical inference procedure (Coles 2001). The practical meaning of this assumption is that the probability of occurrence (chance) of the considered extreme events evolves in time pretty much as we are inclined to think in our everyday life. However, giving a scientific meaning to such an assumption is possible only in an intuitive, heuristic fashion: in an adiabatic limit of infinitely slow trends (but rigorously not even in such a limit). We adopt this point of view not only because it is in line with the common practice and view of extremes, but also because interesting dynamical properties can be inferred from extremes, in analogy with the findings in Part I.

In the present paper we perform and assess time-dependent GEV inference on nonstationary time series $E(t)$ of the total energy obtained by the same simplified quasigeostrophic model that was used in Part I. The model undergoes baroclinic forcing toward a given latitudinal temperature profile controlled by the forced equator-to-pole temperature difference T_E; see Lucarini et al. (2005, 2007b) for a thorough description. We analyze how the parameters of the GEV change with time when a linear trend is imposed on the large-scale macroscopic forcing T_E; that is, when T_E is taken as a (linear) function of time. Since this functional relation is invertible, we derive a parameterization relating the changes in the GEV to variations in T_E (instead of time). One major goal here is to present a methodological framework to be adopted with more complex models and with data coming from observations, as well as an assessment of the performance of the GEV approach for the analysis of trends in extremes in the somewhat gray area of nonstationary time series. Methodologically, our setup is somewhat similar to that of Zhang et al. (2004) regarding the procedures of statistical inference. However, in this case we face two additional problems:

1) As in Part I, the statistical properties of the time series $E(t)$ cannot be selected a priori: in the stationary case (T_E constant in time) and much less in the nonstationary case there is no explicit formula for the probability distribution of the observable $E(t)$.
2) In the nonstationary case we even lack a definition (in fact, even a mere candidate) of what might be the probability distribution of $E(t)$: certainly not a frequency limit for $t \rightarrow \infty$; and not by construction, as opposed to Zhang et al. (2004), who use stochastic generators with a known distribution.

This also means that we have no hypothesis concerning the functional form of the trend in the statistics of extremes of $E(t)$, resulting from the trend imposed on the control parameter T_E. The lack of a general GEV theorem for nonstationary sequences implies that the choice of the time-dependent GEV as a statistical model is, in principle, arbitrary: other models might be equally (or better) suitable. Here the adiabaticity hypothesis mentioned above comes into play, which leads to the central statement of this paper: if the trend is sufficiently slow and the statistical behavior of the atmospheric model has a sufficiently regular response with respect to variations in the external parameters, the GEV remains a suitable model for inference of trend in extremes.

The structure of the paper follows. In section 2, we describe the general problem of the characterization of statistical trends in deterministic models, with both its conceptual and practical implications. Then in section 3 we describe how the GEV modeling can be applied to nonstationary time series and how the quality check of the fits is performed. In section 4 we present the time series considered in this work and the setup of the numerical experiments performed with the atmospheric model. The inferences for various values of the trend in the forcing parameter T_E are presented in section 5 and a sensitivity analysis is carried out in section 6. A comparison with the stationary case analyzed in Part I is given in section 7. In section 8, we summarize the main findings of this work, highlighting the future research developments.

2. Statistical trends: The theoretical problem

The stochastic generator used in this paper to produce the time series is a deterministic model (an ordinary differential equation), whose dynamics, for the considered range of values of T_E, is chaotic in the sense that it takes place on a strange attractor Λ in phase space (Eckmann and Ruelle 1985). See Lucarini et al. (2005, 2007b) for a study of the properties of this attractor, including sensitivity with respect to initial conditions. The statistical behavior of this type of time series is determined by the Sinai–Ruelle–Bowen (SRB) probability measure μ (Eckmann and Ruelle 1985): this is a Borel probability measure in phase space that is invariant under the flow f^t of the differential equation, is ergodic, is singular with respect to the Lebesgue measure in phase space and its conditional measures along unstable manifolds are absolutely continuous [see Young (2002) and references therein]. Moreover, the SRB measure is also a physical measure: there is a neighborhood U of Λ such that for every continuous observable $\phi: U \rightarrow \mathbb{R}$, we have the frequency-limit characterization

$$\lim_{t \to \infty} \frac{1}{t} \int_0^t \phi[f^t(x)]\, dt = \int \phi\, d\mu, \tag{1}$$

for time evolutions $f^t(x)$ starting from almost all $x \in U$ with respect to Lebesgue measure. We are adopting Ruelle's point of view: physical observability corresponds to positive Lebesgue measure in phase space (Eckmann and Ruelle 1985). Notice that the above definition of SRB measure allows for the existence of time evolutions that are pathological (e.g., they do not have an associated statistics μ), but requires that these occupy a set which is physically negligible, since it has Lebesgue measure zero in phase space. Conversely, time evolutions that behave according to a well-defined statistics, identified as the SRB measure μ, occur with probability one: in practice, an experimenter or numerical modeler would only see the latter.

Existence and uniqueness of an SRB measure μ, have been proved only for very special classes of flows f^t [in particular, for flows that possess an Axiom-A attractor; see Young (2002)]. However, existence and uniqueness of μ are necessary to define a stationary stochastic process associated to an observable ϕ. In sum, this allows one to consider a given time series of the form $[\phi[f^t(x)]: t > 0]$ as a realization of the stationary process, justifying statistical inference on a solid theoretical basis. In Part I, we conjectured existence the uniqueness of an SRB measure for the atmospheric model, providing the theoretical foundation to the application of GEV models in the inference of extreme values.

In certain cases of nonautonomous ordinary differential equations (e.g., if the dependence on time is periodic), it still is possible to define, at least conceptually, what an SRB measure is. However, in the present case, due to the form of time dependence adopted for the parameter T_E, the atmospheric model admits no invariant measure. This means that there is no (known) way to associate a stationary stochastic process to the time series of the total energy. In other words, it is even in doubt what we mean by statistical properties of the time series, since it is impossible to define a probability distribution associated with time averages. This conceptual problem has a very serious practical consequence: the operational definition of probability as a frequency limit [as in (1)] is not valid in the nonstationary case, since the time evolution is not a sampling of a unique probability distribution. Even if one assumes the existence of a sequence of distinct probability distributions, one for each observation, one realization (the time series) does not contain sufficient statistical information, since each distribution is very undersampled (with only one observation).

Despite all these problems, the results in Part I suggest a framework that is, for the moment, formulated in a heuristic way. Suppose you evolve an initial condition x in phase space by the flow f^t of the autonomous atmospheric model, that is the system in which T_E is kept fixed to some value T_E^0. After an initial time span (transient), say for t larger than some $t_0 > 0$, the evolution $f^t(x)$ may be thought to take place on the attractor Λ and time averages of the form

$$\frac{1}{t - t_0} \int_{t_0}^{t} \phi[f^t(x)]\, dt \tag{2}$$

may be considered as approximations of

$$\int \phi\, d\mu_0, \tag{3}$$

which is the average of ϕ by the SRB measure μ_0 existing at the value $T_E = T_E^0$ (the attractor average at T_E^0). Now suppose that at some $t_1 > t_0$ the parameter T_E is abruptly changed to some value $T_E^1 > T_E^0$: there will be some transient interval, call it $[t_1, t_2]$, during which the evolution $f^t(x)$ approaches the new attractor of the atmospheric model, that is the attractor existing for T_E fixed at T_E^1. After that time span, the evolution may be considered to take place on the new attractor.

In our case, though T_E varies continuously (linearly) with time, if the trend magnitude is low, then T_E may be considered constant (with good approximation) during time spans that are sufficiently long in order to have both convergence to the new attractor and good sampling of the new SRB measure, in the sense sketched above. Though admittedly heuristic, this scenario allows one to clarify under which condition it is still reasonable to speak of statistical properties of a time series generated by a nonautonomous system: namely, the closeness to a stationary situation for time spans that are sufficiently long. This is the adiabatic hypothesis that we mentioned in the introduction. An essential ingredient for this to hold is that the statistical properties of the autonomous model do not sensibly depend on the external parameter T_E, in the sense that no abrupt transitions (bifurcations) should take place as T_E is varied. This was indeed checked for the system at hand in Part I. Notice that the validity of the adiabatic hypothesis also has a useful practical consequence: one can use the statistics of the stationary system as a reference against which the results for the nonstationary case can be assessed. Having this scenario in mind, we proceed to the description of the time-dependent GEV approach in the next section.

3. GEV modeling for nonstationary time series

The GEV approach for sequences of independent, identically distributed (IID) random variables is by now rather standard (Castillo 1988; Embrechts et al. 1997; Falk et al. 1994; Galambos 1978; Gumbel 1958; Jenkinson 1955; Leadbetter 1974, 1983; Lindgren et al. 1983; Reiss and Thomas 2001; Tiago de Oliveira 1984). We refer the reader to Part I for a concise description of the GEV method in the stationary case. We describe here our approach in the nonstationary setting, referring to Coles (2001) for more details.

If stationarity of the time series does not hold, then the limiting distribution function is no longer bound to be the GEV (or any other prescribed family). Some exact results are known only in certain very specialized types of nonstationarity (Hüsler 1986; Lindgren et al. 1983), but it is very unlikely that a general theory can be established. However, a pragmatic approach is adopted for nonstationary statistical modeling of extreme values: the GEV distribution is used as a template in which time-dependent parameters $\mu(t)$ and $\sigma(t)$ are used, yielding a model of the form

$$G[x; \mu(t), \sigma(t), \xi] = \exp\left\{-\left[1 + \xi\left(\frac{x - \mu(t)}{\sigma(t)}\right)\right]^{-1/\xi}\right\}. \tag{4}$$

Usually ξ is kept time-independent in order to avoid numerical problems, since it is the most delicate parameter to estimate. Different kinds of time dependence can be imposed for $[\mu(t), \sigma(t)]$. In this paper, we adopt a simple polynomial family of models:

$$\mu(t) = \mu_0 + \mu_1 t + \mu_2 t^2, \quad \sigma(t) = \sigma_0 + \sigma_1 t, \tag{5}$$

with $\mu_{0,1,2}$ and $\sigma_{0,1} \in \mathbb{R}$. Polynomially time-dependent families, as in (5), have been already used in the geophysical literature for analysis of nonstationarity at extreme levels [see, e.g., Nogaj et al. (2006)]. Usually, to ensure positivity of the inferred values of $\sigma(t)$, the logarithm of $\sigma(t)$ is used instead of $\sigma(t)$ self. However we chose to directly use $\sigma(t)$ since this yielded a better numerical convergence. In our case, positivity of $\sigma(t)$ is preserved anyway, since the inferred values of the coefficients σ_0 and σ_1 turn out to be always positive.

Members of the family (5) are denoted by $G_{p,q}$ according to the degree of $\mu(t)$ and $\sigma(t)$ as functions of t: for example, $G_{1,0}$ is a model for which $\mu_2 = \sigma_1 = 0$, so that $\mu(t) = \mu_0 + \mu_1 t$ is a linear function of time and $\sigma(t) = \sigma_0$ is constant. Note that the ordinary GEV distribution is obtained from the time-dependent model (4) by setting $\mu_2 = \mu_1 = \sigma_t = 0$ in (5). For the above family, statistical inference amounts to estimating the parameter vector

$$\boldsymbol{\beta} = [\mu_0, \mu_1, \mu_2, \sigma_0, \sigma_1, \xi] \tag{6}$$

from the given time series. The time t is included as a covariate in the inference, which is performed as follows: the time series is divided into m consecutive blocks, each containing n observations, equally spaced in time. Denote by $z_1, \ldots, z_m$ the sequence of the maxima taken over each data block. A log-likelihood function is defined as

$$l(\beta) = -\sum_{t=1}^{m} \left\{ \log\sigma(t) + (1 + 1/\xi) \log\left[1 + \xi\left(\frac{z_t - \mu(t)}{\sigma(t)}\right)\right] + \left[1 + \xi\left(\frac{z_t - \mu(t)}{\sigma(t)}\right)\right]^{-1/\xi} \right\} \tag{7}$$

for the case $\xi \neq 0$, provided that

$$1 + \xi\left(\frac{z_t - \mu(t)}{\sigma(t)}\right) > 0, \quad t = 1, \ldots, m. \tag{8}$$

If $\xi = 0$, an alternative log-likelihood function, derived from the Gumbel distribution, must be adopted. Numerical procedures are used to estimate the value of the parameter vector $\boldsymbol{\beta}$, which maximizes the log-likelihood function in (7). Confidence intervals for $\boldsymbol{\beta}$ may be computed by means of the expected or observed information matrix, using the asymptotic multinormality of the maximum likelihood estimator (Coles 2001).

Although the complexity of our family of time-dependent GEV models, as defined by (5), might be easily increased by choosing polynomials of larger degrees for $\mu(t)$ and $\sigma(t)$, parsimony is recommended (Coles 2001): too many coefficients in the parameter vector $\boldsymbol{\beta}$ would result in unacceptably large uncertainties, especially if few data are available, as it is often the case with real observations. Therefore, in the search for the best estimate model in family (5), the right values for p and q should be determined, being as low as possible compatibly with satisfactory goodness-of-fit. Model selection is achieved through Akaike's information criterion (Akaike 1973, 1974): all models from the family $G_{p,q}$ are first inferred, searching for the one such that Akaike's error

$$\mathrm{AIC}(G_{p,q}) = -2l(G_{p,q}) + 2k \tag{9}$$

is minimal. Here, $k = p + q + 3$ is the number of parameters of the model $G_{p,q}$ and $l(G_{p,q})$ is the maximized log-likelihood corresponding to model $G_{p,q}$. Actually, following the suggestions of Burnham and Anderson (2002, 2004), we have considered the quantity

$$\Delta(G_{p,q}) = \mathrm{AIC}(G_{p,q}) - \min_{p,q} \mathrm{AIC}(G_{p,q}): \tag{10}$$

although the best model, in principle, is that for which $\Delta(G_{p,q}) = 0$, models with $\Delta \leq 2$ should also be considered as reasonable candidates. Therefore, in the presence of a pair of models with $\Delta < 2$, we have used the likelihood ratio test to select the best model (see section 5 for the description of our procedure). This is possible since our family $G_{p,q}$ is nested: simpler models are obtained from more complex ones by setting to zero some of the parameters. In other words, the choice of a model $G_{p1,q1}$ is rejected in favor of a more complex model $G_{p2,q2}$ (where more complex means that $p_2 \geq p_1$ and $q_2 \geq q_1$) if the maximized likelihood of $G_{p2,q2}$ is much larger than that of $G_{p1,q1}$. We refer to Coles (2001) for details.

Probability and quantile plots have been used as graphical diagnostics, where reduction to Gumbel scale is applied to deal with nonstationarity. Let z_t, $t = 1, \ldots, m$ be a sequence of block maxima extracted from a nonstationary time series, from which the time-dependent GEV model $G[\hat{\mu}(t), \hat{\sigma}(t), \hat{\xi}]$ has been inferred. The sequence of maxima is transformed according to

$$\tilde{z}_t = \log\left\{\left[1 + \hat{\xi}\left(\frac{z_t - \hat{\mu}(t)}{\hat{\sigma}(t)}\right)\right]^{-1/\hat{\xi}}\right\}, \quad t = 1, \ldots, m. \tag{11}$$

If Z_t are random variables with distribution $G[\hat{\mu}(t), \hat{\sigma}(t), \hat{\xi}]$, then the transformation (11) produces variables $\tilde{Z}_t$ that have the standard Gumbel distribution. In other words, transformation (11) attempts to remove the time dependence from the sequence of maxima bringing it as close as possible to a common scale. This way, the distribution function and the quantiles of the transformed sequence of maxima $\tilde{z}_t$ can be compared with the empirical ones of the standard Gumbel distribution. The probability and quantile plots are then constructed as in the stationarity case (see Part I), except for using the transformed variables $\tilde{z}_t$ instead of the z_t. For both kinds of plots, displacement of points from the diagonal indicates low quality of the inference.

We conclude this section by emphasizing that all of the above procedure is performed in the spirit of pure inference, that is, determining the likelihood of the adopted parametric hypothesis and not its truth or conceptual validity: these latter, in the absence of a supporting theorem, remain unknown. It should be kept in mind that several structurally different models might fit the observations with similar reliability (likelihood): in this case, as no universal model is suggested or enforced (as opposed to the stationary case), there is no reason to prefer the one above the other.

4. The time series: Total energy of the atmospheric jet model with a trend in average baroclinicity

We consider here the same baroclinic jet model used in Part I, where the spectral order JT is set to 32. The model temperature is relaxed toward a given equator-to-pole profile, which acts as baroclinic forcing. The statistical properties of the model radically change when the parameter T_E, determining the forced equator-to-pole temperature gradient, is varied. A physical and dynamical description of the model is given in Speranza and Malguzzi (1988), Malguzzi et al. (1990), and Lucarini et al. (2005, 2007b).

In Part I we performed an extreme value analysis of the system's response with respect to variations in T_E. Several stationary time series of the total energy $E(t)$ were used as a basis for GEV inference. Each time series was generated with T_E fixed at one value within a uniform grid on the interval [10, 50], with spacing of 2 units. We recall that, given the nondimensionalization of the system, $T_E = 1$ corresponds to 3.5 K, 1 unit of total energy corresponds to roughly 5×10^{17} J, and $t = 0.864$ is one day, see Lucarini et al. (2005, 2007b). In that case, all parameters of the system being kept fixed, after discarding an initial transient each time series of the total energy could be considered as a realization of a stationary stochastic process having weak long-range dependence. Therefore, the classical framework for GEV modeling was applied (see Part I).

In the present setting, a specific linear trend is imposed on T_E: starting at time $t = 0$, the model is run with a time-dependent forcing parameter

$$T_E(t) = (T_E^0 - 1) + t\Delta T_E, \quad t \in [0, t_0], \tag{12}$$

with $T_E^0 = 10$. Three values are chosen for the trend intensity ΔT_E: 2 units every $L_B = 1000$, 300, and 100 yr, yielding three time series for the total energy $E(t)$. The range swept by $T_E(t)$ during integration is kept fixed in all three cases to the interval [9, 51], so that the total length of the time series depends on ΔT_E. Each time series is split into 21 data blocks B^i, $i = 1, \ldots, 21$. The length L_B of each block corresponds to a time interval I^i such that, as t varies within I^i, the baroclinicity parameter $T_E(t)$ by (12) spans the interval

$$[T_E^i - 1, T_E^i + 1], \tag{13}$$

which is 2 units wide and centered around one of the values T_E^i considered in Part I:

$$(T_E^0, T_E^1, \ldots, T_E^{21}) = (10, 12, 14, \ldots, 50). \tag{14}$$

Therefore, the total length L of the time series depends on the trend intensity, so that we have $L = 21 \times 2/\Delta T_E = 21L_B$. Moreover, since the time span over which the maxima are computed is kept fixed to one year, the number of maxima in each data block B^i also depends on ΔT_E: in fact, it is equal to L_B, see Table 1.

Such a selection of the intervals as in (13) allows for a direct comparison of the present results with those obtained for stationary time series in Part I. Moreover, our choices regarding block length and other factors are based on the indications provided in Part I, where the goodness-of-fit assessments performed by a variety of means showed that:

- the adopted block length of 1 yr ensures that the extremes are uncorrelated and genuinely extreme;
- the minimum length (100 data) used for the sequences of maxima yields robust inferences.

TABLE 1. The length L of each of the three time series and the length L_B of each of the 21 the data blocks B_i (both are expressed in years), as a function of the intensity ΔT_E of the trend (12) imposed on the parameter T_E of the baroclinic model.

ΔT_E	2/(1000 yr)	2/(300 yr)	2/(100 yr)
L	21 000	6300	2100
L_B	1000	300	100

5. Time-dependent GEV analysis of the total energy

For each data block B^i, $i = 1, \ldots, 21$, we first extract a sequence of yearly maxima z_t^i, with $t = 1, \ldots, L_B$ For compactness, each sequence is denoted in vector form as $z^i = (z_1^i, z_2^i, \ldots, z_{L_B}^i)$. One GEV model of the form $G_{p,q}$ [see (5)] is fitted from each of the sequences z^i. In the following, by simpler model we mean a model having less nonzero parameters in the vector $\boldsymbol{\beta}$ as in (6). For each $i = 1, \ldots, 21$, the analysis follows three main steps:

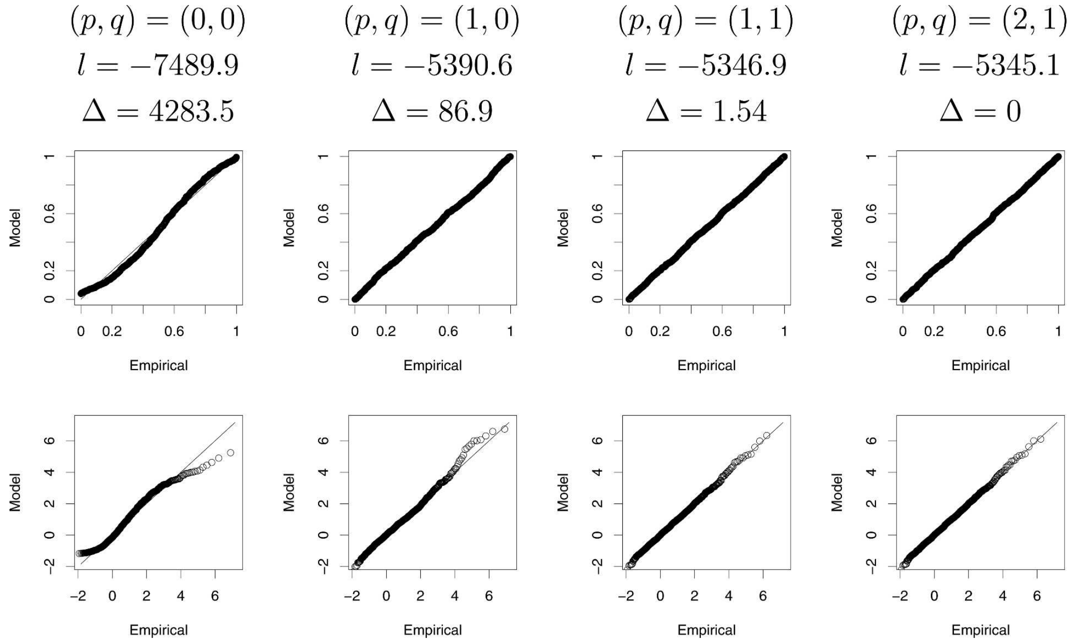

FIG. 1. Diagnostic plots of GEV inferences with model (4) and parameters as in (5), for block B_i with $i = 2$ [corresponding to $T_E^i = 12$ through (14)] and $\Delta T_E = 2/(1000 \text{ yr})$. (top) Probability plots; (bottom) quantile plots; (left to right) plots for models $G_{p,q}$ [see (5)], with (p, q) indicated on top, together with the corresponding log-likelihood l [Eq. (7)] and value of Δ [Eq. (10)].

1) nested models $G_{p,q}$, for $0 \le p \le 2$ and $0 \le q \le 1$, are fitted on the ith sequence of maxima z^i; the quantity $\Delta(G_{p,q})$ is computed for all p, q;
2) models for which $\Delta > 2$ are discarded; the model with $\Delta = 0$ is checked against each of the models with $\Delta < 2$, starting from the simplest and each time increasing the complexity: according to the likelihood ratio test, if a simpler model cannot be rejected in favor of that with $\Delta = 0$, then the simpler model is chosen as the best fit; conversely, if the model with $\Delta = 0$ must be rejected in favor of a more complex model, then the latter is chosen as the best fit;
3) lastly, the best-fit model is graphically checked by examining the probability and quantile plots, and it is possibly rejected in favor of a simpler model.

Following the above procedure, for each time interval I^i, $i = 1, \ldots, 21$, time-dependent GEV models $G_{\hat{p}^i,\hat{q}^i}(z^i)$ with parameters $[\hat{\mu}^i(t), \hat{\sigma}^i(t), \hat{\xi}^i]$ are inferred from the data block z^i. Model $G_{\hat{p}^i,\hat{q}^i}(z^i)$ (denoted for shortness $G_{\hat{p}^i,\hat{q}^i}$ in the rest of this section) is the best estimate for the ith data block, relative to the family of models $G_{p,\,q}$ and to the selection criteria listed above. Choosing a model with different orders (p, q) would either give poor results in the graphical checks, or fail to pass Akaike's information criterion, or the likelihood ratio test.

An example is given in Fig. 1, for the data block $i = 2$ in the time series with $\Delta T_E = 2/(1000 \text{ yr})$. According to Akaike's criterion, models $G_{0,0}$ and $G_{1,0}$ are immediately discarded and one should choose model with $(p, q) = (2, 1)$. However, by the likelihood ratio test, model $(p, q) = (1, 1)$ cannot be rejected at the 5% level of confidence in favor of model $(p, q) = (2, 1)$, since the deviance statistics satisfies $\mathcal{D} = 2\{l_{2,1} - l_{1,1}\} = 3.6 < 3.84$ (the latter is the 0.95-quantile of the χ_1^2 distribution). Moreover, the graphical diagnostics reveal that both models fit the data quite well. Therefore, the simpler model $(\hat{p}^i, \hat{q}^i) = (1, 1)$ is selected as best fit.

Another example of out procedure is given in Fig. 2, for the data block $i = 18$ in the time series with $\Delta T_E = 2/(100 \text{ yr})$ According to Akaike's criterion, models $G_{0,0}$ is discarded, whereas $G_{1,0}$, $G_{1,1}$, and $G_{2,1}$ are all plausible. By the likelihood ratio test model $(p, q) = (1, 0)$ cannot be rejected at the 5% level of confidence in favor of model (p, q) (2, 1), since the deviance statistics satisfies $\mathcal{D} = 2\{l_{2,1} - l_{1,0}\} = 4.3 < 5.99$, which is the 0.95-quantile of the χ_2^2 distribution. The simplest nonstationary model $(\hat{p}^i, \hat{q}^i) = (1, 0)$ is therefore selected as best fit (the graphical diagnostics does not indicate lack of fit). Model $(p, q) = (1, 1)$ is not taken into account, since we have found a simpler model with a satisfactory fit.

Plots of the best estimate parameters $[\hat{\mu}^i(t), \hat{\sigma}^i(t), \hat{\xi}^i]$ as functions of time are proposed in Fig. 3. Confidence

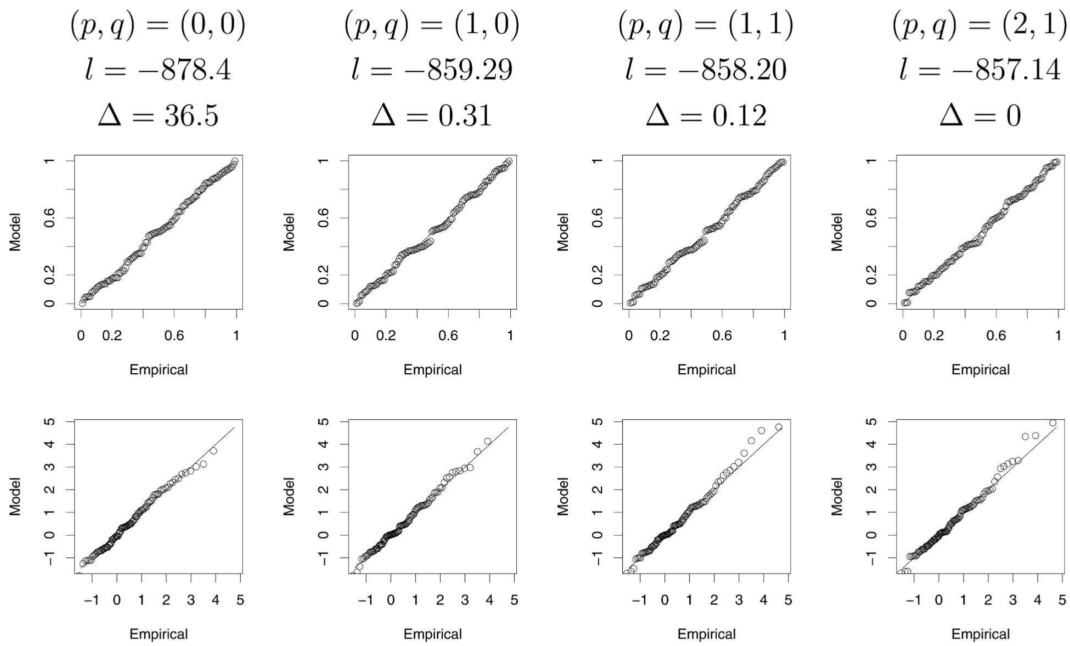

FIG. 2. Same as Fig. 1 but for block $i = 18$ [corresponding to $T_E^i = 44$ through (14)] and $\Delta T_E = 2/(100 \text{ yr})$.

intervals are computed by assuming exact multinormality of the maximum likelihood estimator, which only holds asymptotically (Coles 2001): to fix ideas, suppose that for the ith data block a best-estimate model is obtained with $\hat{p}^i = 1$, that is, $\hat{\mu}^i(t) = \mu_0^i + \mu_1^i t$. Then the uncertainty of $\hat{\mu}^i(t)$ is computed as though μ_0^i and μ_1^i were normal variables with variances and covariance given by $\sigma_{\mu_0^i}$, $\sigma_{\mu_1^i}$ and $\text{cov}(\mu_0^i, \mu_1^i)$, respectively, where the latter are provided by the observed information matrix (see Part I):

$$\sigma_{\hat{\mu}^i(t)} = \sqrt{\sigma^2_{\mu_1^i} t^2 + 2t\,\text{cov}(\mu_0^i, \mu_1^i) + \sigma^2_{\mu_0^i}}. \quad (15)$$

The corresponding confidence interval at time t is computed as

$$[\hat{\mu}^i(t) - 2\sigma_{\hat{\mu}^i(t)}, \hat{\mu}^i(t) + 2\sigma_{\hat{\mu}^i(t)}]. \quad (16)$$

These confidence intervals have been systematically compared with those obtained by a standard bootstrap procedure, which uses the same covariance matrix: no significant difference has been detected.

For most of the time intervals I^i, the best-estimate model is such that $\hat{\mu}^i(t)$ and $\hat{\sigma}^i(t)$ are respectively linear and constant in time; that is, $(\hat{p}^i, \hat{q}^i) = (1, 0)$. (See Tables 2, 3, and 4.) The best fit includes a quadratic term in $\mu(t)$ for only three cases, whereas models of the form $(\hat{p}^i, \hat{q}^i) = (1, 1)$ are typically found for small values of i. The fact that a statistical model of enhanced complexity (more parameters) is needed to achieve goodness-of-fit may be interpreted as follows: although the hypothesis of smoothness, described in section 2, may be considered valid, the rate of variation of the SRB measure with respect to variations in T_E is comparatively larger for small values of T_E, particularly when viewed through the extremes of the energy observable.

In concluding this section, we emphasize that the convergence of the numerical procedure used in the maximization of the likelihood function is here considerably more problematic than in the stationary case studied in Part I. Indeed, in the present case it is often necessary to accurately choose a suitable starting point for the maximization procedure, in order to achieve convergence. In particular, ad hoc starting points had to be searched for all cases with $i = 1$.

6. Trend assessment

When dealing with nonstationary data, the problem of assessing the sensitivity of trend inferences is particularly delicate. Beyond the serious conceptual problems explained in section 2, one is confronted with several practical issues. Most of the sensitivity tests in Part I were based on examining a shorter portion of same time series or on calculating the maxima on data blocks of different lengths. In the present nonstationary context, both operations would result in an alteration of the statistical properties of the sample (exactly because of the nonstationarity) and this makes comparisons somewhat ambiguous. An example is provided in Fig. 4 where we compute the best-estimate GEV fits using

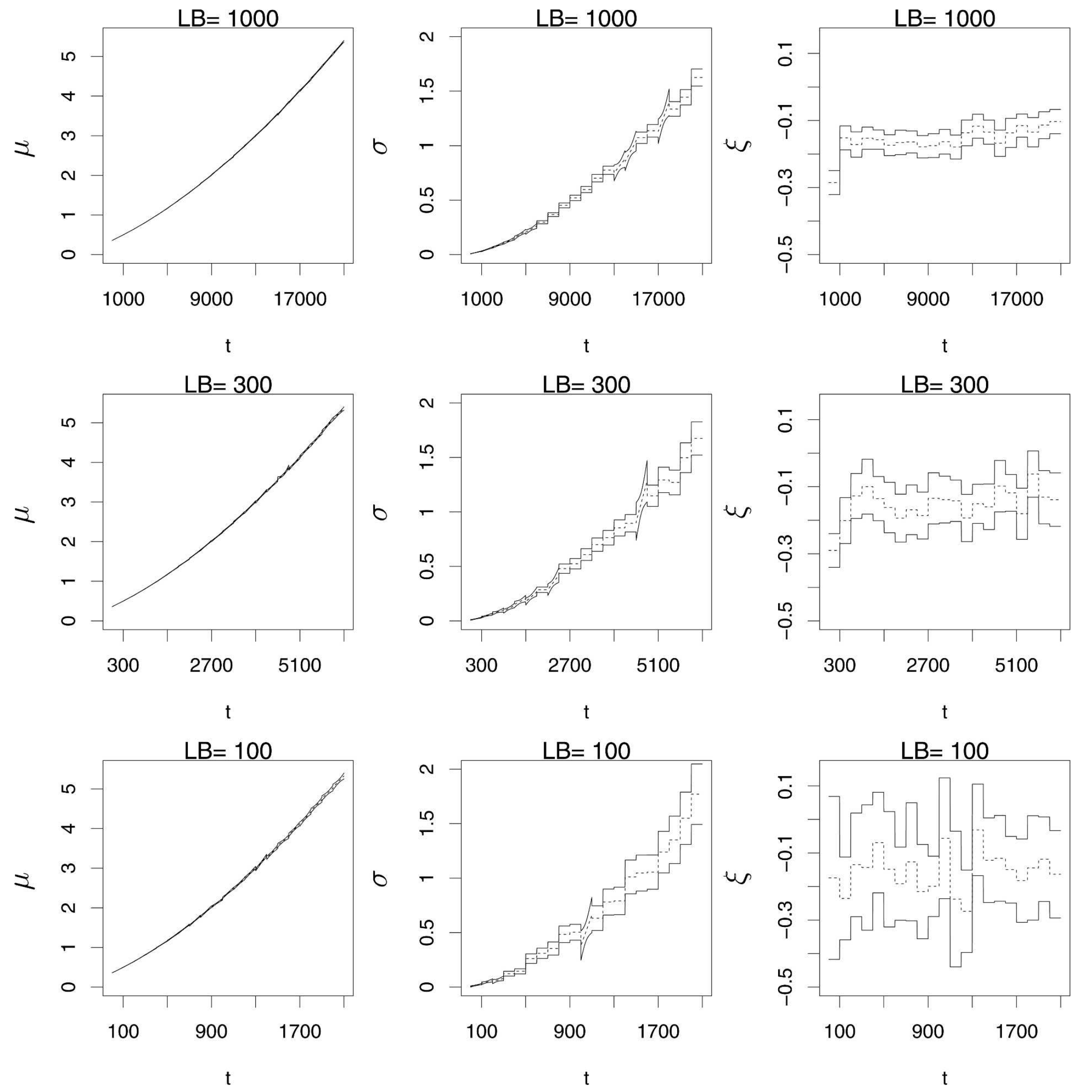

FIG. 3. GEV parameters as functions of time, for the three considered values of trend intensity ΔT_E: (top) ΔT_E = 2/(1000 yr), (middle) 2/(300 yr), and (bottom) 2/(100 yr). For each trend intensity, the inferred time-dependent parameters—(left) $\mu(t)$, (center) $\sigma(t)$, and (right) ξ—of the best estimate model $G_{\hat{p}^i,\hat{q}^i}(z^i)$ are plotted together with confidence intervals computed according to (15).

sequences of yearly maxima having different lengths—but starting at the same instant (year 15 000)—extracted from the time series with the slowest trend intensity ΔT_E = 2/(1000 yr). Notice that the best fit obtained by taking 100 yearly maxima is stationary. The corresponding extrapolations in time are, of course, completely wrong. By using 500 and 1000 maxima, the best estimates obtained (not shown) fall inside the confidence band of the 2000-yr-based estimate for most values of time.

The above example illustrates the trend dilemma: on the one hand, in order to be detected, a statistical trend has to be sufficiently fast with respect to the length of the record of observations; on the other hand, if the trend is too fast then the adiabatic hypothesis discussed in section 2 is no longer valid: one is left with no reference statistics against which the inferred models can be compared.

Moreover, when considering large time spans a further practical complication arises: because of the nonlinear dependence of the statistical properties with respect to the external parameter T_E, a functional relation between the GEV parameters and time might require many parameters to achieve goodness of fit. An indication of this phenomenon is the fact that for small values of i typically one needs more parameters, as dis-

TABLE 2. Best estimate GEV fits $G_{\hat{p}^i,\hat{q}^i}(z^i)$ with parameter vector as in (6) for the nonstationary time series with trend speed $\Delta T_E = 2/(1000\text{ yr})$; see text for details.

i	$\hat{\mu}_0$	$\hat{\mu}_1$	$\hat{\mu}_2$	$\hat{\sigma}_0$	$\hat{\sigma}_1$	$\hat{\xi}$
1	3585.34	1.41	0	7.79	0.02	−0.29
2	4990.60	1.50	0	30.71	0.04	−0.15
3	6498.00	1.61	0	66.34	0.04	−0.17
4	8106.62	1.77	0	105.56	0.04	−0.15
5	9851.78	1.84	0	155.78	0.06	−0.16
6	11 709.87	1.96	0	206.43	0.06	−0.17
7	13 699.46	2.00	0	296.81	0	−0.17
8	15 718.89	2.14	0	367.36	0	−0.16
9	17 936.64	2.15	0	451.82	0	−0.18
10	20 141.28	2.35	0	521.44	0	−0.17
11	22 527.79	2.25	0	597.35	0	−0.16
12	24 934.78	2.48	0	702.10	0	−0.18
13	27 452.54	2.58	0	774.77	0	−0.14
14	30 061.01	2.61	0	749.40	0.13	−0.12
15	32 686.08	2.88	0	855.56	0.19	−0.13
16	35 348.30	3.16	0	1072.35	0	−0.17
17	38 384.21	3.06	0	1136.50	0	−0.14
18	41 335.49	2.88	0	1131.06	0.27	−0.12
19	44 351.82	3.11	0	1336.62	0	−0.13
20	47 501.96	3.15	0	1443.89	0	−0.11
21	50 675.26	3.20	0	1625.06	0	−0.10

cussed in section 5. In more problematic cases, one faces the problem of large uncertainties in the parameter estimates or even lack of convergence. This has indeed been observed for the present time series: if we consider a long record, such that the change in T_E is large, the model family $G_{p,q}$ with parameters as in (5) becomes inadequate to catch the time dependence of the statistics of extremes. As a further example, we have examined a data block of length 5000 starting at year 14 500 in the time series with $\Delta T_E = 2/(1000\text{ yr})$. Inspection of graphical diagnostics (probability and quantile plots) reveals that no model in the family $G_{p,q}$ produces an acceptable inference. It should be emphasized that goodness-of-fit is achieved for the same time series using blocks of length 1000; that is, performing inferences that are more localized in time. Thus we infer that in this case the problem is not the failure of the adiabatic or smoothness hypothesis, but the nonlinear dependence of the attractor on the parameter T_E, which manifests itself on sufficiently large time intervals.

TABLE 3. As in Table 2 but for trend speed $\Delta T_E = 2/(300\text{ yr})$.

i	$\hat{\mu}_0$	$\hat{\mu}_1$	$\hat{\mu}_2$	$\hat{\sigma}_0$	$\hat{\sigma}_1$	$\hat{\xi}$
1	3583.57	4.70	0	8.29	0.08	−0.29
2	4988.11	5.00	0	35.09	0.09	−0.20
3	6491.71	5.41	0	68.78	0.10	−0.13
4	8115.16	5.82	0	88.98	0.18	−0.10
5	9843.28	6.29	0	152.46	0.17	−0.14
6	11 735.94	6.45	0	176.65	0.30	−0.16
7	13 750.43	6.20	0	285.48	0	−0.19
8	15 763.80	6.99	0	286.16	0.44	−0.17
9	17 844.04	7.86	0	479.55	0	−0.19
10	20 107.74	7.57	0	524.39	0	−0.13
11	22 485.55	7.75	0	608.08	0	−0.14
12	24 954.35	8.26	0	699.97	0	−0.14
13	27 314.52	9.46	0	763.00	0	−0.19
14	29 928.19	10.07	0	853.05	0	−0.15
15	32 723.97	8.88	0	896.10	0	−0.16
16	36 056.60	1.29	0.03	912.06	1.24	−0.10
17	38 284.71	9.73	0	1147.70	0	−0.12
18	41 437.32	9.95	0	1293.54	0	−0.18
19	44 386.77	10.05	0	1270.29	0	−0.06
20	47 704.61	10.77	0	1496.65	0	−0.13
21	50 890.62	9.09	0	1673.99	0	−0.14

TABLE 4. As in Table 2 but for trend speed $\Delta T_E = 2/(100\text{ yr})$.

i	$\hat{\mu}_0$	$\hat{\mu}_1$	$\hat{\mu}_2$	$\hat{\sigma}_0$	$\hat{\sigma}_1$	$\hat{\xi}$
1	3586.92	13.53	0.01	5.94	0.22	−0.17
2	4989.65	14.84	0	34.92	0.26	−0.24
3	6494.95	15.26	0.01	50.34	0.30	−0.14
4	8118.51	17.11	0	123.09	0	−0.14
5	9905.01	17.29	0	144.91	0	−0.07
6	11 679.56	19.67	0	261.31	0	−0.15
7	13 628.09	21.49	0	309.89	0	−0.19
8	15 542.18	24.39	0	354.45	0	−0.13
9	17 757.54	25.41	0	484.45	0	−0.22
10	20 245.62	19.27	0	503.43	0	−0.20
11	22 528.24	24.12	0	378.94	2.83	−0.06
12	24 848.74	25.25	0	632.31	0	−0.24
13	27 442.05	27.17	0	780.44	0	−0.27
14	29 638.34	35.68	0	790.87	0	−0.03
15	32 618.26	28.05	0	1010.55	0	−0.12
16	35 812.81	21.48	0	1046.24	0	−0.12
17	38 422.25	28.22	0	1055.08	0	−0.15
18	41 120.50	31.06	0	1238.43	0	−0.18
19	44 510.73	27.28	0	1350.89	0	−0.14
20	47 641.12	25.50	0	1549.09	0	−0.12
21	50 452.92	28.15	0	1770.02	0	−0.16

7. Smooth dependence on the forcing

The setup of the present analysis (see section 4) has been chosen to allow comparison of the nonstationary GEV inferences with the results of Part I, obtained from statistically stationary time series. To perform the comparison, for each $i = 1, \ldots, 21$ the best estimate parameters $[\hat{\mu}^i(t), \hat{\sigma}^i(t), \hat{\xi}^i]$ inferred from data block B^i are first expressed as functions of T_E inside the interval (13). This is achieved by inverting the trend Eq. (12), (writing time as a function T_E):

$$t(T_E) = \frac{T_E - T_E^0 + 1}{\Delta T_E}, \quad T_E \in [9, 51], \tag{17}$$

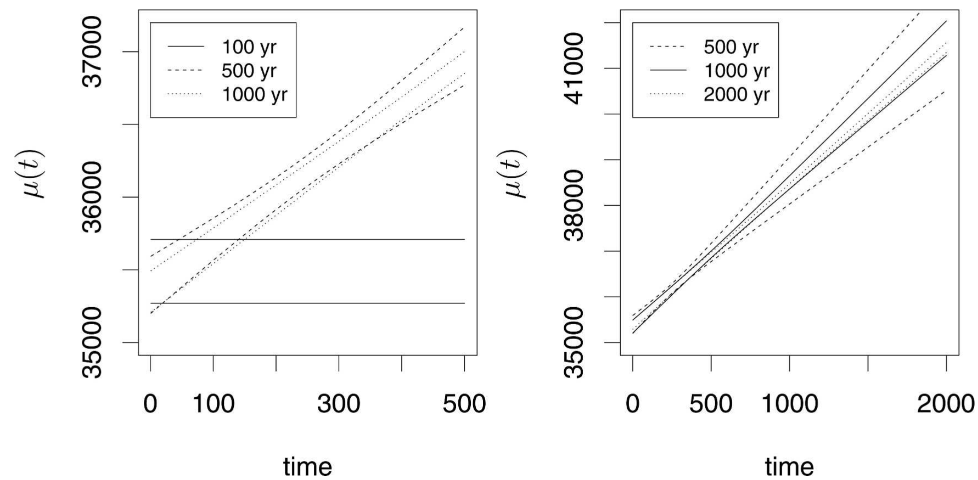

FIG. 4. Parameter $\mu(t)$ of the best estimate GEV inferences as a function of time. The time series with slowest trend intensity $\Delta T_E = 2/(1000 \text{ yr})$ has been used, taking yearly maxima over a data block starting at year 15 000. For legibility, only the confidence intervals have been plotted. (left) Inferences obtained with 100-, 500-, and 1000-yearly maxima. The best estimate fit based on 100 data is stationary ($\mu_1 = 0$) and it has been extrapolated to 500 yr. (right) Inferences obtained with 500-, 1000-, and 2000-yearly maxima. For 500 and 1000, the best estimate GEV model is of the form $G_{1,0}$; that is, $\sigma(t) = \sigma_0$ is constant and $\mu(t) = \mu_0 + \mu_1 t$ is a linear function of time. For 2000, also $\sigma(t) = \sigma_0 + \sigma_1 t$ is linear in time.

and inserting this into the expression of $[\hat{\mu}^i(t), \hat{\sigma}^i(t), \hat{\xi}^i]$. This yields functions that are denoted as $[\hat{\mu}^i(T_E), \hat{\sigma}^i(T_E), \hat{\xi}^i]$. These are evaluated at the central point T_E^i of the interval of definition and plotted in Fig. 5. Confidence intervals are given by the same estimate as in (16), used for Fig. 3, again evaluated at the value of t corresponding to the central point T_E^i via (12). A rather smooth dependence on T_E is observed, especially for the GEV parameters μ and σ. The location parameter μ turns out to be not very sensitive to changes in the trend intensity, being much more sensibly dependent on variations in T_E. Moreover its confidence intervals are always very small (relatively to the size of μ).

Denote by w^i the sequence of yearly maxima of the total energy, computed for T_E fixed at T_E^i, which were used in Part I to infer time-independent GEV models. These stationary GEV models are here denoted by $G_{0,0}(w^i)$ and the inferred values of their parameters by $\tilde{\mu}^i$, $\tilde{\sigma}^i$, and $\tilde{\xi}^i$. Since the graphs of the parameters $\tilde{\mu}^i$, $\tilde{\sigma}^i$, and $\tilde{\xi}^i$ versus T_E^i very closely match those in Fig. 5, comparison with the stationary data is presented under the form of relative differences (Fig. 6). To be precise, on the left column the absolute values of the ratios

$$\frac{\hat{\mu}^i(T_E^i) - \tilde{\mu}^i}{\hat{\mu}^i(T_E^i) + \tilde{\mu}^i} \tag{18}$$

are plotted against T_E^i (similarly for the GEV parameters σ and ξ). Remarkable agreement is obtained for the parameter μ: the relative differences less than 10% and drop below 5% for large T_E and for all considered trend intensities. Excellent agreement is also obtained for σ (particularly for large T_E) and for ξ except for the fastest trend intensity $\Delta T_E = 2/(100 \text{ yr})$. In the latter case, indeed, the sample uncertainty is as large as (or even larger than) the estimates self.

We emphasize that inferring time-independent models $G_{0,0}(z^i)$ from the nonstationary data z^i might induce very large errors, particularly in the scale and shape parameters. This is also reflected in the diagnostic plots, compare Fig. 1 (leftmost column). This lack of fit obtained with the time-independent model is particularly bad for the lowest values of T_E^i. A much better (even surprising) agreement between the stationary and nonstationary estimates is obtained with the procedure described in the previous section: first fitting the time-dependent model $G_{\hat{p}^i,\hat{q}^i}(z^i)$ and then evaluating its parameters at the central point T_E^i. There is agreement even in the estimates of the parameter ξ, which is usually the most difficult one to infer. In the case of stationary time series, since the attractor is bounded and since the energy observable $E(t)$ is a continuous function of the phase space variables, the total energy is bounded on any orbit lying on (or converging to) the attractor. Therefore, the total energy extremes are necessarily Weibull-distributed (ξ is negative); see Part I. Although this property is not bound to hold for nonstationary forcing, it is still verified; see Table 2.

Two distinct power-law regimes are identified for the

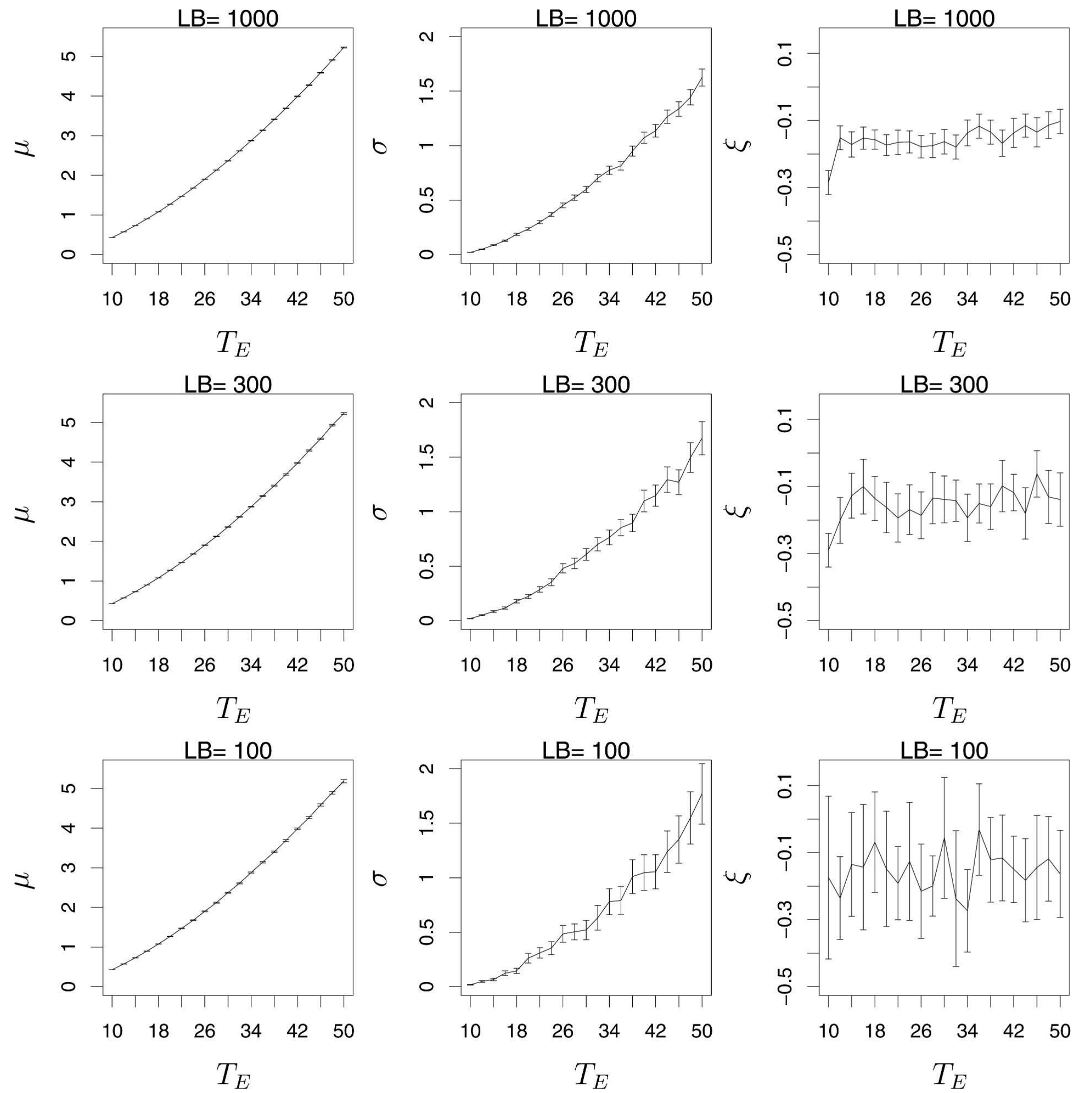

FIG. 5. (left to right) Parameters $\hat{\mu}^i(T_E^i)$, $\hat{\sigma}^i(T_E^i)$, $\hat{\xi}^i$ of the best estimate GEV model $G_{\hat{\rho}^i,\hat{q}^i}(z^i)$ evaluated at the central point T_E^i of each of the 21 intervals (13). The trend intensity ΔT_E is equal to (top) 2/(1000 yr), (middle) 2/(300 yr), and (bottom) 2/(100 yr).

GEV parameters ($\hat{\mu}^i$, $\hat{\sigma}^i$, $\hat{\xi}^i$) as functions of T_E^i, having the form

$$\hat{\mu}^i(T_E^i) = \alpha_\mu (T_E^i)^{\gamma_\mu} \quad \text{and} \quad \hat{\sigma}^i(T_E^i) = \alpha_\sigma (T_E^i)^{\gamma_\sigma}; \tag{19}$$

see Figs. 7 and 8. The values of the fitted exponents γ_μ and γ_σ in each scaling regime are reported in Tables 5 and 6, respectively. A similar power law dependence of the GEV parameters on T_E was already observed in Part I for the stationary datasets w^i: indeed, the exponents obtained there are very similar to those in Tables 5 and 6, particularly for large T_E. The lack of a power-law scaling regime for the parameter σ for small T_E explains both the more pronounced differences between the stationary and nonstationary estimates (Fig. 6) and the necessity of including a linear term for σ in the statistical model to get acceptable inferences. This highlights the strongly nonlinear behavior of the baroclinic model, whose response to changes of T_E has different features depending on the considered range of variation.

Two factors explain the qualitative analogies and the quantitative agreements between the time-dependent models discussed here and the stationary results of Part I. First of all, the trend intensity imposed on T_E in all cases is sufficiently slow with respect to the time of relaxation of the baroclinic model to the statistics of

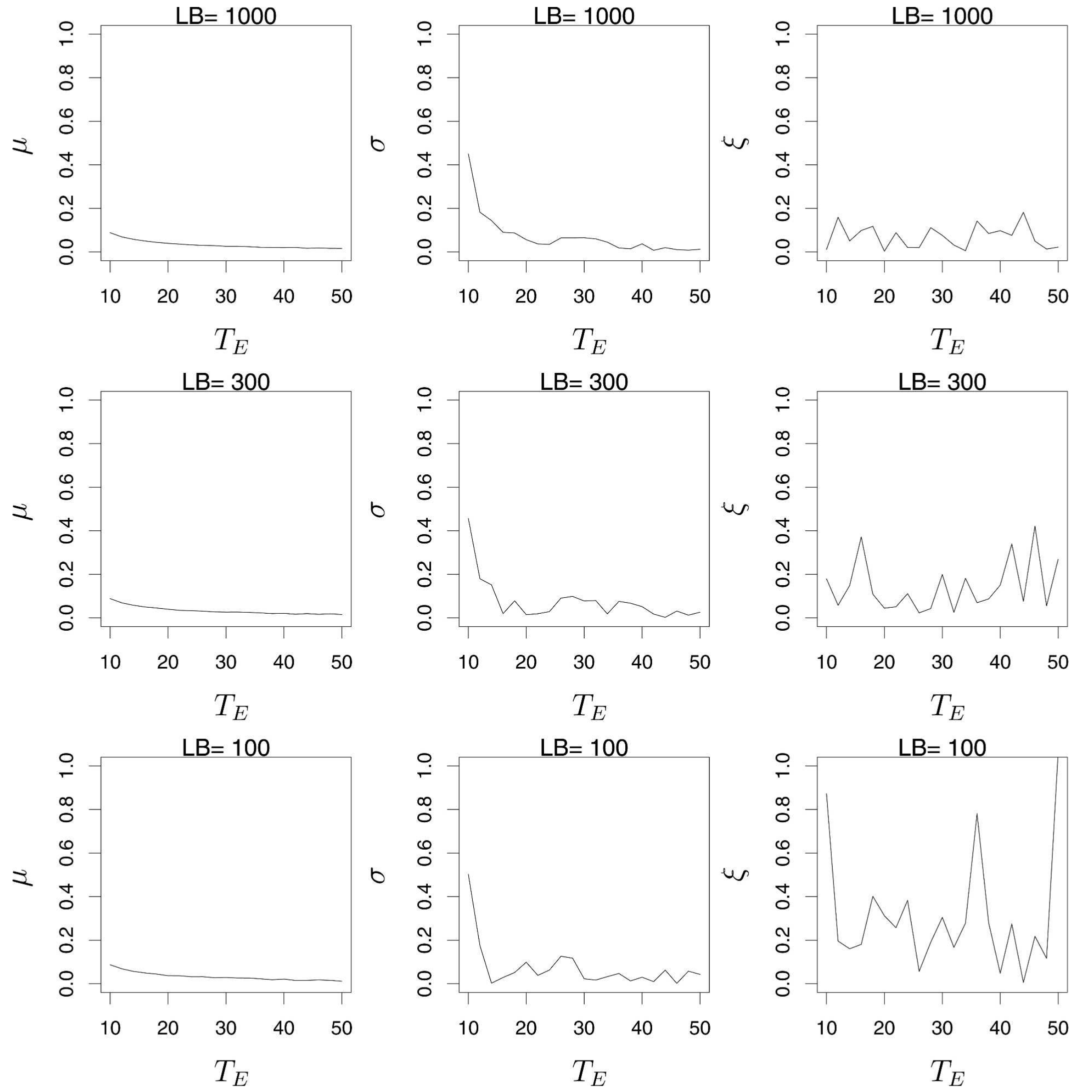

FIG. 6. Relative differences of the values of the GEV parameters $\tilde{\mu}^i$, $\tilde{\sigma}^i$, and $\tilde{\xi}^i$, inferred in Part I from stationary data for $T_E = T_E^i$ fixed, and the best estimate GEV parameters [$\hat{\mu}^i(T_E^i)$, $\hat{\sigma}^i(T_E^i)$, $\hat{\xi}^i$] obtained from the nonstationary simulations in the present setting (see the explanation in section 7 for details).

extreme values of the total energy. For clarity, we emphasize that the latter time scale is that used in section 2 to define the adiabatic hypothesis: it is the time necessary to obtain a good sampling of the SRB measure on the attractor, provided that one may consider the system as frozen (with constant T_E) for sufficiently long time spans. We do not know yet (the problem is under examination) whether this time scale bears any physical relation to other time scales, such as those of baroclinic instability or low-frequency variability [both have been described in Speranza and Malguzzi (1988) for the present model]. A second factor is that the system's statistical behavior responds rather smoothly to the imposed time-dependent variation of the parameter T_E. This smooth dependence on T_E of the statistical properties of the baroclinic model was analyzed in detail in Lucarini et al. (2005, 2007b) by considering not only global physical quantities such as total energy and average wind profiles, but also finer dynamical indicators, such as the Lyapunov exponents and dimension. Both properties of smoothness and adiabaticity are of crucial importance in order to justify the usage of nonstationary GEV models that are (locally) smooth functions of time, such as our polynomial family $G_{p,q}$.

8. Summary and conclusions

In this paper, we have investigated a general, although not universal, framework for the analysis of

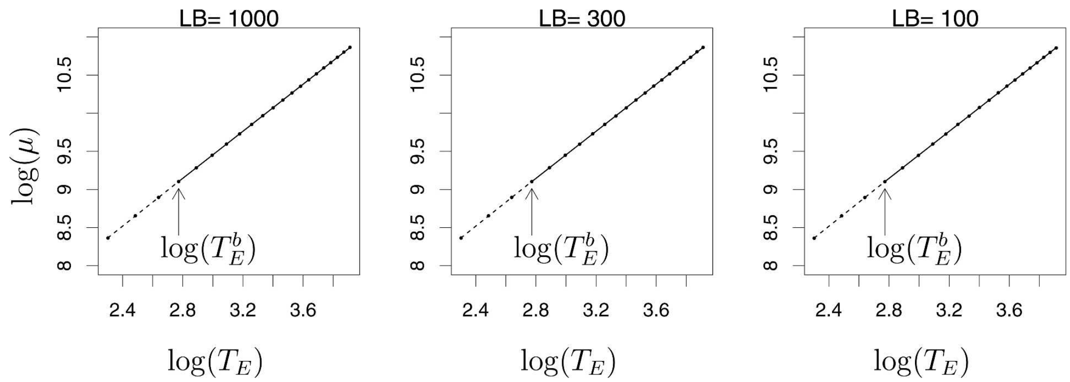

FIG. 7. Power-law fits of the inferred values of $\hat{\mu}^i(T_E^i)$ as a function of T_E^i [see (14)]. Trend intensities of (left) 2/(1000 yr), (middle) 2/(300 yr), and (right) 2/(100 yr) have been used. In each case, there are two intervals of T_E (separated by the common endpoint T_E^b, see Table 5) characterized by two distinct scaling laws.

trends in extremes of climatic time series. When all the shortcomings that are present in datasets and observations have to be considered, a rigorous definition of extremes and a neat, clean, and legible approach to the evaluation of trends is necessary in order to get useful and reliable information (Zhang et al. 2005). The time-dependent approach allows the expression the inferred GEV distributional parameters as functions of time. As expected, it is found that trend in the statistics of extreme values is detectable in a reliable way, provided that the record of observations is sufficiently long, depending on the time scale of the trend itself. Trend inference and assessment is much more problematic than in the statistically stationary inference. First, one is faced with a serious conceptual problem: there is no operational definition of probability, since, to say it in loose words, the time series is not a sampling of a unique probability distribution, as it is in the stationary case. Even if one assumes that the time series is a realization of a sequence of random variables (with different distributions), the statistical properties of the sample are altered by any operation such as resampling or taking shorter subsamples, which makes sensitivity studies somewhat ambiguous. One must assume that the distributions of the random variables vary slowly and smoothly with time, so that the time series contains sufficient sampling information on the local (in time) statistical behavior.

In the present context, we have adopted GEV models whose parameters are polynomial in time: the location parameter μ is at most quadratic with respect to time and the scale parameter σ is at most linear in time. Since the relation between the macroscopic forcing T_E and time is invertible, the time dependence of the inferred GEV models can be expressed as a relation between the GEV parameters and T_E, showing rather interesting properties. The location and scale parameters feature power-law dependence with respect to T_E, while the shape parameter has in all cases a negative value. As expected, both results are in agreement with what obtained in the companion paper (Part I) for stationary data. Since the parameter T_E increases mono-

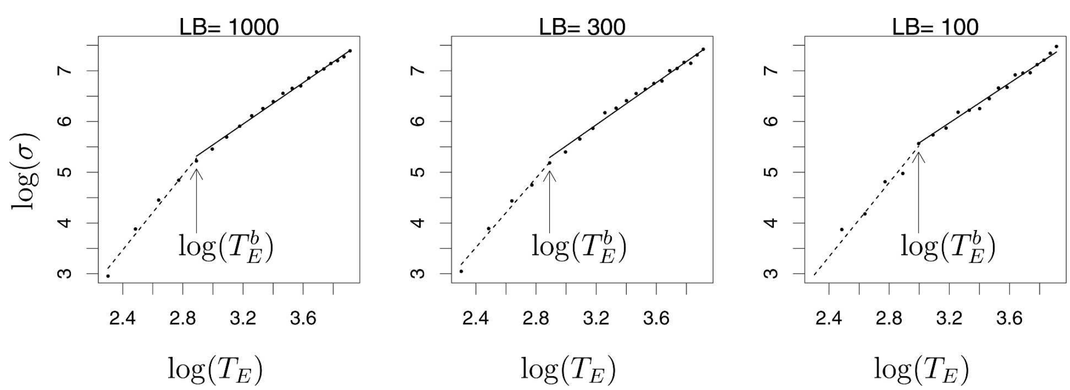

FIG. 8. As in Fig. 7 but for the inferred values $\hat{\sigma}^i(T_E^i)$.

TABLE 5. Power-law fits of the inferred location parameter $\hat{\mu}^i(T_E^i)$ as a function of T_E^i [see (14)] of the form $\hat{\mu}^i(T_E^i) = \alpha_\mu(T_E^i)^{\gamma_\mu}$. Two distinct scaling regimes (with distinct exponents $\gamma_{\mu,1}$ and $\gamma_{\mu,2}$) are identified and the corresponding adjacent intervals in the T_E axis are separated by the common endpoint T_E^b reported in the third column from left.

L_B	$\gamma_{\mu,1}$	T_E^b	$\gamma_{\mu,2}$
1000	1.573 ± 0.004	16	1.54 ± 0.01
300	1.572 ± 0.002	16	1.545 ± 0.002
100	1.572 ± 0.001	16	1.542 ± 0.001

TABLE 6. As in Table 5 but for the inferred scale parameter $\hat{\sigma}^i(T_E^i)$.

L_B	$\gamma_{\sigma,1}$	T_E^b	$\gamma_{\sigma,2}$
1000	3.69 ± 0.06	18	2.04 ± 0.04
300	3.4 ± 0.1	18	2.08 ± 0.07
100	3.7 ± 0.2	20	2.0 ± 0.1

tonically in the simulations with the baroclinic model, the system certainly does not possess any invariant measure. However, the results suggest that, as T_E increases, the system explores statistical states that vary smoothly with T_E and whose properties are locally quite similar to those obtained in the stationary setting. This is even captured for the relatively fast trend intensity $\Delta T_E = 2/(100 \text{ yr})$. The proposed explanation is that:

1) the system's statistical properties depend rather smoothly on T_E (cf. Lucarini et al. 2005, 2007b);
2) the adopted time scales of variation of T_E (i.e., the trend intensity ΔT_E) are sufficiently slow compared to the relaxation time to the statistics of extreme values.

The second condition, which was explained in more detail in section 2, amounts to the heuristic statement that for sufficiently short time spans the system's statistical properties can be considered frozen to those holding for a corresponding value of T_E. The possibility of using GEV models that are locally smooth (polynomial functions of time) depends essentially on these two conditions. For example, for a system having several bifurcations as the control parameter is changed the time-dependent GEV modeling would be much more complicated. However, even if the above two conditions do hold, the inference of time-dependent GEV models is valid locally in time, that is, if the sequences of maxima used for the inference span not too large time periods. For large time spans, indeed, the nonlinear response of the baroclinic model to variations in T_E becomes dominant and polynomial GEV models are no longer suitable. On the other hand, if the sequences of maxima used for the inference are too short (depending on the trend intensity), the wrong trend estimates may be obtained.

There are several ways in which we plan to extend the present study. First, it would be very useful to give quantitative criteria for the definition of the slowness of a statistical trend. This problem is very hard when dealing with observed data, where one does not dispose of a stationary model, justified on theoretical and physical grounds, which can act as a reference statistic, to be used as comparison as well as test (as we have done in Fig. 6). However, in the theoretical context of dynamical systems having well-behaved statistical properties (e.g., hyperbolic attractors or Anosov systems), one might try to formalize the heuristic notion of adiabaticity presented in section 2. Another aspect of this investigation would be to examine what might be the impact of trend in systems that are not smooth with respect to external parameters, for example, in systems having bifurcations. A second point is that we have considered an indicator of global character, the total energy of the system. Other choices might be to analyze the wave kinetic energy, the available energy or also the maximum vorticity on the domain of the model, which might behave in different ways as T_E is changed. Moreover, there are delicate issues connected with reducing the scale from a global indicator to a local one, such as the value of the wind on a grid point. This brings into play all complications due to the multifractality and the spatial dependence of the process. A more fundamental problem is the characterization of SRB measures and observables such that the associated stochastic process falls into the domain of attraction of the GEV family. First rigorous results in this sense were obtained by Haiman (2003) and J. M. Freitas and A. C. Moreira-Freitas (2007, personal communication), who examined the one-dimensional dynamical systems given by the tent map and the quadratic map, respectively. In particular, in both cases it was proved that the extremes are Weibull distributed (under mild conditions on the observable and for suitable values of the parameters), contradicting previous speculations (Balakrishnan et al. 1995). Lastly, a further development of the present work is the usage of extreme statistics as a dynamical indicator, in the sense of process-oriented metrics (Lucarini et al. 2007a). All these issues are currently under investigation.

We conclude by observing that the present and the companion paper (Part I) are devoted not merely to the statistical inference of extremes and their trends but also to explore the possibility of using extreme statistics in diagnosing the dynamical state of a geophysical fluid. Our analysis of the problem reveals, in fact, that diagnostics, which are based on universal (GEV theorem), robust (smoothness properties), simple (power-law scaling), controllable (low-dimensional parametric) statistical models, can be very helpful in setting up well-targeted models of the general circulation (see Lucarini et al. 2005, 2007b).

Acknowledgments. The authors are indebted to Nazario Tartaglione for useful conversations and wish to thank two anonymous referees for carefully reading the manuscript. This work has been supported by MIUR PRIN Grant "Gli estremi meteo-climatici nell'area mediterranea: Proprietà statistiche e dinamiche," Italy, 2003.

REFERENCES

Akaike, H., 1973: Information theory and the maximum likelihood principle. *Second International Symposium on Information Theory,* B. N. Petrov and F. Csaki, Eds., Academical Kiado, 267–281.

——, 1974: A new look at the statistical model identification. *IEEE Trans. Automatic Control,* **19,** 716–723.

Balakrishnan, V., C. Nicolis, and G. Nicolis, 1995: Extreme value distributions in chaotic dynamics. *J. Stat. Phys.,* **80,** 307–336.

Brunetti, M., M. Maugeri, T. Nanni, and A. Navarra, 2002: Droughts and extreme events in regional daily Italian precipitation series. *Int. J. Climatol.,* **22,** 543–558.

——, L. Buffoni, F. Mangianti, M. Maugeri, and T. Nanni, 2004: Temperature, precipitation and extreme events during the last century in Italy. *Global Planet. Change,* **40,** 141–149.

Burnham, K. P., and D. R. Anderson, 2002: *Model Selection and Inference: A Practical Information–Theoretic Approach.* Springer-Verlag, 488 pp.

——, and ——, 2004: Multimodel inference: Understanding AIC and BIC in model selection. *Soc. Meth. Res.,* **33,** 261–304.

Castillo, E., 1988: *Extreme Value Theory in Engineering.* Academic Press, 389 pp.

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

Diaz, H., and T. Nanni, 2006: Historical reconstruction, climate variability and change in Mediterranean regions. *Nuovo Cimento,* **29C,** 1–2.

Eckmann, J.-P., and D. Ruelle, 1985: Ergodic theory of chaos and strange attractors. *Rev. Mod. Phys.,* **57,** 617–655.

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

Falk, M., J. Hüsler, and R. Reiss, 1994: *Laws of Small Numbers: Extremes and Rare Events.* Birkhauser, 389 pp.

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

Galambos, J., 1978: *The Asymptotic Theory of Extreme Order Statistics.* Wiley, 352 pp.

Gnedenko, B. V., 1943: Sur la distribution limite du terme maximum d'une série aléatorie. *Ann. Math.,* **44,** 423–453.

Gumbel, E. J., 1958: *Statistics of Extremes.* Columbia University Press, 375 pp.

Haiman, G., 2003: Extreme values of the tent map process. *Stat. Probab. Lett.,* **65,** 451–456.

Hüsler, J., 1986: Extreme values of nonstationary random sequences. *J. Appl. Probab.,* **23,** 937–950.

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.

Karl, T. R., and R. W. Knight, 1998: Secular trend of precipitation amount, frequency, and intensity in the United States. *Bull. Amer. Meteor. Soc.,* **79,** 231–242.

——, ——, D. R. Easterling, and R. G. Quayle, 1996: Indices of climate change for the United States. *Bull. Amer. Meteor. Soc.,* **77,** 279–292.

Katz, R. W., and B. G. Brown, 1992: Extreme events in a changing climate: Variability is more important than averages. *Climatic Change,* **21,** 289–302.

Leadbetter, M. R., 1974: On extreme values in stationary sequences. *Z. Wahrsch. Verw. Gebiete,* **28,** 289–303.

——, 1983: Extremes and local dependence in stationary sequences. *Z. Wahrsch. Verw. Gebiete,* **65,** 291–306.

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

Lucarini, V., T. Nanni, and A. Speranza, 2004: Statistics of the seasonal cycle of the 1951–2000 surface temperature records in Italy. *Nuovo Cimento,* **27C,** 285–298.

——, A. Speranza, and R. Vitolo, cited 2005: Physical and mathematical properties of a quasi-geostrophic model of intermediate complexity of the mid-latitudes atmospheric circulation. [Available online at http://arxiv.org/abs/physics/0511208.]

——, ——, and T. Nanni, 2006: Statistics of the seasonal cycle of the 1951–2000 surface temperature records in Italy and in the Mediterranean area. *Nuovo Cimento,* **29C,** 21–31.

——, S. Calmanti, A. dell'Aquila, P. M. Ruti, and A. Speranza, 2007a: Intercomparison of the Northern Hemisphere winter mid-latitude atmospheric variability of the IPCC models. *Climate Dyn.,* **28,** 429–448.

——, A. Speranza, and R. Vitolo, 2007b: Self-scaling of the statistical properties of a minimal model of the atmospheric circulation. *20 Years of Nonlinear Dynamics in Geosciences,* J. Elsner and A. Tsonis, Eds., Springer, in press.

Malguzzi, P., A. Trevisan, and A. Speranza, 1990: Statistic and predictability for an intermediate dimensionality model of the baroclinic jet. *Ann. Geophys.,* **8,** 29–36.

Nogaj, M., P. Yiou, S. Parey, F. Malek, and P. Naveau, 2006: Amplitude and frequency of temperature extremes over the North Atlantic region. *Geophys. Res. Lett.,* **33,** L10801, doi:10.1029/2005GL024251.

Reiss, R.-D., and M. Thomas, 2001: *Statistical Analysis of Extreme Values: With Applications to Insurance, Finance, Hydrology and Other Fields.* 2d ed. Birkhauser, 443 pp.

Speranza, A., and P. Malguzzi, 1988: The statistical properties of a zonal jet in a baroclinic atmosphere: A semilinear approach. Part I: Two-layer model atmosphere. *J. Atmos. Sci.,* **45,** 3046–3061.

——, and N. Tartaglione, 2006: Extreme events in the Mediterranean area: A mixed deterministic–statistical approach. *Nuovo Cimento,* **29C,** 81–88.

——, and Coauthors, 2006: Estremi nelle scienze ambientali. *L'Acqua,* **3/2006,** 19–30.

Tartaglione, N., A. Speranza, F. Dalan, T. Nanni, M. Brunetti, and M. Maugeri, 2006: The mobility of Atlantic baric depressions leading to intense precipitation over Italy: A preliminary statistical analysis. *Nat. Hazards Earth Syst. Sci.,* **6,** 451–458.

Tiago de Oliveira, J., 1984: *Statistical Extremes and Applications.* Kluwer Academic, 708 pp.

Young, L. S., 2002: What are SRB measures, and which dynamical systems have them? *J. Stat. Phys.,* **108,** 733–754.

Zhang, X., F. W. Zwiers, and G. Li, 2004: Monte Carlo experiment on the detection of trends in extreme values. *J. Climate,* **17,** 1945–1952.

——, G. Hegerl, F. W. Zwiers, and J. Kenyon, 2005: Avoiding inhomogeneity in percentile-based indices of temperature extremes. *J. Climate,* **18,** 1641–1651.