ENVIRONMETRICS
Environmetrics 2010; **21**: 698–718
Published online 13 July 2010 in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1002/env.1060

Different ways to compute temperature return levels in the climate change context[‡]

Sylvie Parey[1*,†], Thi Thu Huong Hoang[1,2] and Didier Dacunha-Castelle[2]

[1] *EDF/R&D, 6 quai Watier, 78 401 Chatou Cedex, France*
[2] *Laboratoire de Mathématiques, Université Paris 11, Orsay, France*

SUMMARY

The climate change context has raised new problems in the computation of temperature return levels (RLs) in using the statistical extreme value theory. This arises since it is not yet possible to accept the hypothesis that the series of maxima or of high level values are stationary, without at least verifying the assumption. Thus, in this paper, different approaches are tested and compared to derive high order RLs in the nonstationary context. These RLs are computed by extrapolating identified trends, and a bootstrap method is used to estimate confidence intervals. The identification of trends can be made either in the parameters of the extreme value distributions or in the mean and variance of the whole series. Then, a methodology is proposed to test if the trends in extremes can be explained by the trends in mean and variance of the whole dataset. If this is the case, the future extremes can be derived from the stationary extremes of the centered and normalized variable and the changes in mean and variance of the whole dataset. The RL can then be estimated as nonstationary or as stationary for fixed future periods. The work is done for both extreme value methods: block maxima and peak over threshold, and will be illustrated with the example of a long observation time series for daily maximum temperature in France. Copyright © 2010 John Wiley & Sons, Ltd.

KEY WORDS: climate change; extreme value theory; trends; extrapolation; return level

1. INTRODUCTION

The statistical extreme value theory (EVT) is commonly used by engineers to evaluate the intensity of meteorological extreme events to which industrial installations or buildings must withstand. These events are evaluated as long period return levels (RLs), which correspond to very rare events.

This paper concerns the prediction of such very rare events in a nonstationary context. Climate change induces nonstationarity in the climatic series, at least for temperature. This means one must identify an evolution of some parameters with time and then extrapolate this evolution to predict the possible future RLs. Extrapolation involves the assumption that observed behavior remains valid in the future. This is a strong hypothesis, but difficult to avoid. It is thus important to determine such trends in an as intrinsic manner as possible, that is in avoiding high frequency variations or sample effects. For

*Correspondence to: S. Parey, EDF/R&D, 6 quai Watier, 78 401 Chatou Cedex, France.
[†]E-mail: Sylvie.parey@edf.fr
[‡]This article is published in *Environmetrics* as a special issue on TIES 2008: *Quantitative Methods for Environmental Sustainability*, edited by Sylvia R. Esterby, University of British Columbia Okanagan, Canada.

Received 30 October 2008
Accepted 7 May 2010

example, the trend used for extrapolation may be defined on a subperiod of the whole observation period or modified to rid it of sample effects. This will be illustrated by examples.

The EVT has existed since the mid 20th century (Gumbel, 1958; Leadbetter *et al*., 1983) and advances has been proposed in different fields: Smith (1989) gives basic examples for environmental applications, Coles (2001) and Katz *et al*. (2002) developed the EVT for the fields of oceanography and hydrology. The EVT relies on two general definitions of extreme events: extreme events can be considered as maxima of given blocks of time and then, they are described by the generalized extreme value (GEV) distribution, or they can be considered as exceedances over a defined high threshold, and then, when independent and in sufficient quantity, they follow a generalized Pareto distribution (GPD), whereas their dates of occurrence follow the trajectory of a Poisson process.

Probability theory assumes that the studied series are stationary (not necessarily independent), so they do not present any cycle or trend, these being easily identified forms of nonstationarity. However, when dealing with climatic data, this assumption has to be carefully considered. Since climatic data often show a seasonal cycle and include the possibility of recent trends, the evaluation of temperature extremes can now hardly be conducted without considering this issue of trends. The seasonal cycle issue can be sometimes overcome if the study is restricted to the season when the extremes of the studied variable mainly occur, but still needs some basic verification.

Extensions have been proposed to the EVT to deal with trends (Smith, 1989; Coles, 2001). Different ways of identifying trends in extremes have been tested in Zhang *et al*. (2004) and the study concluded that methods based on the modeling of trends in the parameters of the distribution of the extremes are the most powerful. Hundecha *et al*. (2008) used this approach to look for trends in extreme annual wind speed over the Gulf of St Lawrence in Canada. This same methodology of trend identification in the parameters of the distribution of extremes will be used in this paper. The probability theory is applied for each block separately and so gives an estimate of the parameters depending on the block. Then one can constraint these parameters to some predetermined analytical dependency such as polynomials or continuous piecewise linear (CPL) functions. The question arises as to how the identified trend can be used to compute a future RL. In the literature, this question is rarely addressed. Generally, RLs are computed from recent rather than historical parts of the observed time series or by using climate model simulations.

In Parey *et al*. (2007), a methodology is proposed to estimate high temperature RLs in a nonstationary context from observed temperature series in France, by extrapolating recent observed trends in the parameters of the distribution of extremes. Some insights are given on the choice of the trend and the impact of observed extremes like the 2003 heat wave, based on many observation series in France. Using this methodology, the present paper further studies the identification and extrapolation of trends in the estimation of temperature RLs. The issue of trend identification is considered first, including the questions of possible attribution to a background climatic evolution and the influence of sample effects, for example several consecutive very hot or very cold years at the end of the observation period. Once trend identification has been settled, estimation of the confidence interval for the RL in a nonstationary context is considered. These questions will be carefully investigated in this paper by applying methodologies to identify trends in both the extreme distribution parameters and the mean and variance of the whole series. A methodology is proposed to test the link between these two types of trends, for the two EVT methods, block maxima and peak over threshold (POT), and to compute future RLs.

The paper is organized as follows: first the mathematical tools are described in Section 2, then the main issues are illustrated on the example of the observed daily maximum temperature at the station of Déols in France for the 1901–2006 period (Klein Tank *et al*., 2002). In Section 3, the extrapolation of

DOI: 10.1002/env

trends in the extreme value distribution parameters is discussed whereas the verification and the use of the link between trends in extremes and trends in the mean and variance of the whole dataset are exposed in Section 4, before coming to the discussion in Section 5.

2. MATHEMATICAL TOOLS

2.1. General framework

The general framework of the study is the statistical EVT. According to the block maxima method, the block maxima M_j of block j are modeled as independent variables with a GEV distribution with three parameters: the location parameter μ, the scale parameter σ and the shape parameter ξ. With the POT method, the distribution of independent values selected over a fixed high threshold asymptotically converges to a GPD and their dates converge to a Poisson process. The GPD has two parameters: the scale parameter σ and the shape parameter ξ. In theory, the shape parameter of the GEV and the GPD distributions for the same series are identical. The intensity of the Poisson process is another parameter of the POT method.

In the application of the methods, choices have to be made in order to verify the assumptions. For the block maxima method, the block length must be sufficiently long to allow the asymptotical convergence of the maxima distributions and the independence between two consecutive maxima. For the POT method, the threshold has to be selected in such a way that the process of the dates of occurrence follows a Poisson process and that the property of threshold stability of the GPD is fulfilled (Smith, 1989). On the other hand, the values have to be independent, and a declustering technique has often to be used. Goodness of fit tests allow us to check the fit of the distribution of exceedances to the GPD and the set of dates to a Poisson process, and to validate the hypotheses of independence. The threshold selection can be based on the properties of the GPD, such as the constancy of the shape parameter or the linearity of the mean of the exceedances over an adequate threshold value.

2.2. Trends in the parameters of the extreme value distributions

2.2.1. Trend identification. Nonstationarity is translated for EVT methods by modeling the parameters as nonconstant functions of time. The scale and shape parameters of the GEV and GPD, the location parameter of GEV and the Poisson process are allowed to evolve with time or with other covariates. The shape parameter ξ is the most difficult to estimate, and it could be tricky to differentiate possible evolutions from estimation errors. In their study, Zhang *et al*. (2004) did not consider any trend in this parameter, as they assume that it is not likely to show a trend in the climate series. The tests on different periods of a long observation series have shown that this parameter does not significantly evolve with time (Parey *et al*., 2007), and more sophisticated nonparametric studies lead to the same conclusion (Hoang *et al*., 2009). Thus, in the following, the shape parameter ξ will be considered as constant for both GEV and GPD distributions.

As the ultimate goal is to extrapolate trends in parameters in order to compute RLs, parametric forms given by closed formulas have to be considered. Initially, nonparametric approaches may be used to suggest the general form of these trends, which can then be reasonably approximated by polynomials or other family like the family of CPL functions (Parey *et al*., 2007). These families of functions are often ordered (or partially ordered) by an integer d as the polynomial degree or the number of pieces of a CPL.

DOI: 10.1002/env

The choice of d is based on the likelihood ratio test or another procedure such as the Akaike criterion (Davison, 2008).

For the GEV distribution, the procedure is the following: for every pair of integers (dmu,dsigma) one computes the coefficients which maximize the log-likelihood for the block maxima M_j with a location parameter $\mu(j)$ and a scale parameter $\sigma(j)$ considered as polynomials of degree dmu for $\mu(j)$ and dsigma for $\sigma(j)$. To choose the optimal model in this polynomial class for $\mu(j)$ and $\sigma(j)$, one of the previously mentioned procedures can be used. The Akaike penalization is 2(dmu + dsigma). The same procedure can be applied for CPL functions. For the POT method, the same principle is applied for the intensity of the Poisson process $I(t)$ and the scale parameter of the GPD $\sigma(t)$.

The advantage of CPL models is that they capture more details and provide linear trends which seem to be more appropriate for the RL extrapolations. Still, statistical optimization is harder for CPL functions than for polynomials and the mathematical theory for model choice and estimation is more difficult (Dacunha-Castelle and Gassiat, 1999) due to nonidentifiability. The classical theory of likelihood is no longer valid and consistent estimators are not available. This can be avoided under the constraint of separation of angular nodes, *e.g.*, by 10 years. A misuse of CPL models could however be to systematically conduct the prediction by extrapolating the last linear piece. The slope of this linear fragment may be too dependant on the distribution of the last observed values in the last period. We thus define a specific statistical procedure in order to control these edge effects and to choose the extrapolated trend.

2.2.2. Return level in a nonstationary context. Estimation of the RL by extrapolating these trends in the parameters requires that the parameters are monotonous functions, and thus polynomials will be restricted to degree 2. In the stationary case, the RL z_a for a years is the level for which the probability of exceedance every year is equal to $1/a$. In the nonstationary case, knowing the identified trend, z_a will then be re-defined as the unique level such that the expectation of the number of exceedances over z_a in the next a years will be 1 (Parey *et al.*, 2007). We shall define the RL z_a for a years starting from the date t_0, so the RL is a function of the initial date and of the number of years taken into account to make the projection.

Let $D(t_0,a)$ be the set of all days that belong to $[t_0, t_0 + 365a]$. Then, for the block maxima method, RL z_a is such that:

$$\frac{1}{\text{nb}} \sum_{t=t_0}^{t_0+365a} \left\{ 1 - \exp\left[-\left(1 + \frac{\xi}{\sigma(t)} (z_a - \mu(t)) \right)^{-1/\xi} \right] \right\} = 1 \tag{1}$$

nb is the number of days in each block, as each day in a block shares the same probability of being the maximum.

For the POT method, z_a is such that:

$$\sum_{t \in D(t_0,a)} \left(1 + \frac{\xi}{\sigma(t)} (z_a - u) \right)^{-\frac{1}{\xi}} I(t) = 1 \tag{2}$$

$I(t)$ is the mathematical expectation of the number of exceedances at date t and the Pareto term $\left(1 + \frac{\xi}{\sigma(t)} (z_a - u)\right)^{-\frac{1}{\xi}}$ is the mathematical expectation that the exceedance is larger than the level z_a, u being the threshold.

The definition for the stationary case is a particular case of this one.

DOI: 10.1002/env

In each case, z_a is then the result of an optimization procedure and we will consider that the optimization procedure properly converged if the value of the function for the computed level z_a is lower than 10^{-5}.

2.2.3. Associated confidence intervals. In the stationary context, confidence intervals can be derived from the standard deviation of the estimated RLs, using the approximate normality of the maximum likelihood estimator or of any function of it. This so-called ''delta method'' will be used here to evaluate the confidence interval in the stationary case.

If a trend is identified, the confidence interval has to take both sampling errors and trend identification errors into account. The method is based on a bootstrap technique, which will be described first for the block maxima method.

Once the trend has been identified, the block maxima series can be expressed as:

$$M_j = \hat{\mu}(j) + \hat{\sigma}(j)e_j \tag{3}$$

where M_j are the block maxima, $\hat{\mu}(j)$ and $\hat{\sigma}(j)$ the estimated optimal functions for the location and scale parameters and e_j are residuals having a GEV distribution of location parameter 0 and scale parameter 1.

Then, the residuals are computed from the observed maxima and the estimated optimal trends. This residual series is bootstrapped and a new sample of block maxima is computed. From this new sample, the trend identification method allows one to compute the optimal trend of the same form as the original one and a new RL is estimated. By applying this procedure a high number of times (500 times here), a distribution of the estimated RL is produced, from which a confidence interval can be derived for a defined level, for example 90%.

For the POT method, a similar procedure can be applied, but with the help of simulation. The number of dates of threshold exceedances follows a Poisson distribution. Thus, first a new number of dates N^* will be sampled from a Poisson distribution of integers with parameter $\overline{I}$ corresponding to the mean value of $I(t)$ over the observation period. Then, a new series of N^* dates of exceedances is drawn (simulated) as a sample of the probability density function $f(t) = \frac{I(t)}{\overline{I}}$ giving sets of new dates $t_1 \ldots t_{N^*}$ and so new estimates $I^*(t)$ of the intensity.

At these new dates we shall affect new values V_u over the threshold u. Observed values can be expressed as:

$$V_u = u + \pi\hat{\sigma}(t) \tag{4}$$

where $\hat{\sigma}(t)$ is the estimated scale parameter and π a residual following a GPD of scale 1. A new set of values V_u^* can be produced by bootstrapping the iid π. Then, values sampled in this way are associated to the new set of dates of exceedances, while for the other dates of the series, a value is sampled from the original values under the threshold u. This procedure is of course applied on the ''declustered'' series, that is the series from which the dependent values over the threshold and their dates have been eliminated. Here again, by applying this sampling procedure a high number of times (for instance 500), it is possible to estimate a high number of new sets of parameters $I^*(t)$, $\hat{\xi}$, and $\hat{\sigma}^*(t)$ from which a distribution of the RL can be obtained with the previous formula, distribution used to derive the desired confidence interval.

2.3. Link between trend in extremes and trends in the whole series: the K hypothesis

2.3.1. Nonparametric tools to derive trends. In Parey *et al*. (2010), nonparametric methods have been introduced and used to derive trends in mean and variance of the summer daily maximum (and

DOI: 10.1002/env

winter daily minimum) temperature series. A link has been identified between the trend in mean and the trend in variance, the variance increasing when mean increases in summer for different European locations, and especially in France. The local polynomial estimators (LOESS) will be used here, with the same smoothing parameter, to compute the trends in mean $m(t)$ and in variance $s^2(t)$ of the observed summer daily maximum temperature $X(t)$.

2.3.2. Link with the trends in extremes. The question then is whether a link can be identified between these nonparametric trends in mean and variance of the whole dataset and the trends in extremes (Hoang *et al*., 2009). To study it, $Y(t)$ is defined as the centered and normalized variable computed from the observed series $X(t)$ as:

$$Y(t) = \frac{X(t) - m(t)}{s(t)} \tag{5}$$

For GEV, the parameters for $X(t)$ can be obtained from those of $Y(t)$ using the following relationships:

$$\begin{cases} \xi_X = \xi_Y \\ \sigma_X(t) = \sigma_Y(t) * s(t) \\ \mu_X(t) = m(t) + \mu_Y(t) * s(t) \end{cases} \tag{6}$$

If the extreme value model for $Y(t)$ is stationary, then the RL of $X(t)$ can be obtained from the stationary RL for $Y(t)$ and the trends in mean and variance of the whole dataset. Thus the trend in extremes is mainly explained by the trends in mean and variance of the whole series. Then, the trends are to be identified from the whole ensemble of data, which makes the identification more robust, having been obtained from a much larger dataset.

A methodology to test the so called ''K hypothesis'', which states that the extremes of $Y(t)$ can be considered as stationary, has been proposed and detailed in the GEV case in Hoang *et al*. (2009). Only the general principles are recalled here. The test consists in comparing the distance between the two functions of time corresponding to the stationary and nonstationary estimations of the parameters of the extreme distribution of $Y(t)$ to a table of such distances constructed by using a known stationary extreme value distribution with the same parameters. The steps are then the following:

(1) compute a nonparametric trend for the mean $m(t)$ of the observed values $X(t)$, using LOESS: $\hat{m}(t)$
(2) compute the variance var(t) as $(X(t) - \hat{m}(t))^2$ and its non parametric trend using the same LOESS method: $\hat{s}^2(t)$
(3) compute $Y(t) = \frac{X(t) - \hat{m}(t)}{\hat{s}(t)}$
(4) estimate $\hat{\mu}(j)$ and $\hat{\sigma}(j)$, j being the block index, the location and scale parameters of the GEV distribution:
    (a) as constant in time $\hat{\mu}_0$ and $\hat{\sigma}_0$
    (b) as time varying $\hat{\mu}(j)$ and $\hat{\sigma}(j)$
    (c) and their distances: $\Delta\mu = \left(\int_{t \in D} (\hat{\mu}(t) - \hat{\mu}_0) \mathrm{d}t \right)^{1/2}$ and $\Delta\sigma = \left(\int_{t \in D} (\hat{\sigma}(t) - \hat{\sigma}_0)^2 \mathrm{d}t \right)^{1/2}$, D being the number of days in the series
(5) compute 1000 samples of the same number of maxima with the constant parameters $\hat{\mu}_0$ and $\hat{\sigma}_0$ and the 1000 distances between the parameters estimated from these samples as constant or time varying

DOI: 10.1002/env

(6) situate $\Delta\mu$ and $\Delta\sigma$ in the distributions of distances obtained from the stationary distribution and accept or reject the hypothesis.

If the distance computed from the observed series lies inside the distributions, then, the K hypothesis can be accepted and the distance is only due to statistical errors.

The same procedure can be applied for the POT method by computing ΔI and $\Delta\sigma$ and comparing them to the distributions of such distances coming from a sample of a known stationary situation. The main difference lies in the fact that the nonparametric evolution of $I(t)$ is computed using a kernel method rather than LOESS.

2.3.3. The use of this link to compute return levels and their associated confidence intervals. Once the K hypothesis has been tested and can be accepted, one can imagine different ways to use it in the computation of future RLs. On one hand, the RLs and associated confidence intervals can be evaluated in the nonstationary context, but by extrapolating the trends in mean and variance of the whole dataset instead of the trends in the parameters of the extreme value distributions. The trends in mean and variance of the whole dataset will be computed here as a two-part CPL function, based on the identification of one break point in their evolutions. This will be done by using the so-called ''brut force'' procedure proposed by Mudelsee (2009). Thus, $m(t)$ and $s(t)$, respectively the time evolutions of the mean and standard deviation of the whole dataset, are considered as two linear parts, and the second part is extrapolated to compute the nonstationary RLs, as follows:

$$\frac{1}{\text{nb}}\sum_{t=t_0}^{t_0+365a}\left\{1-\exp\left[-\left(1+\frac{\xi}{\sigma_Y}\left(\frac{z_a-m(t)}{s(t)}-\mu_Y\right)\right)^{-1/\xi}\right]\right\}-1=0\,\text{for GEV} \tag{7}$$

and

$$\sum_{t\in D(t_0,a)}\left(1+\frac{\xi}{\sigma_Y}\left(\frac{z_a-m(t)}{s(t)}-u\right)\right)^{-\frac{1}{\xi}}I_Y-1=0\,\text{for POT} \tag{8}$$

Then, the confidence intervals are constructed by applying a bootstrap procedure:

- bootstrap Y(t) to obtain a high number of samples $Y^*(t)$ and compute the stationary parameters for the extreme value distribution of each $Y^*(t)$
- compute $X^*(t)$ from each $Y^*(t)$ and m(t) and s(t)
- from each $X^*(t)$, compute linear trends for m and s from the previously identified break point, to obtain $m^*(t)$ and $s^*(t)$
- compute each RL using (7) or (8): RL^*
- get the confidence interval as the desired inter-quintile interval of the distribution of the RL^* according to the chosen confidence level (5 and 95 percentiles for the 90% confidence interval).

Another way is to consider that the RLs can be computed in the stationary context for fixed periods of time, for example 30 years. Thus, the present day RLs are the stationary ones evaluated over the most recent period of the observed series. Then, using the K hypothesis, the future RLs can be computed from the stationary distribution of the extremes of $Y(t)$ and the values of mean m and standard deviation s over a future 30-year period. The confidence interval is computed by using the delta method, taking both the errors in the extreme value distribution parameters of $Y(t)$ and those of m and s into account,

DOI: 10.1002/env

these last ones being evaluated using a bootstrap technique. The values for m and s for a future period can be retrieved either from the trend extrapolations or from climate model simulations.

3. EXAMPLE: EXTRAPOLATION OF TRENDS IN EXTREME VALUE DISTRIBUTION PARAMETERS

The temperature series chosen to illustrate the previously described methodology to compute RLs in the nonstationary context is provided by the European Climate Assessment and Dataset (ECA&D) project (Klein Tank *et al*., 2002). It consists of daily maximum temperature observed at the station named Déols (in the center of France) over the 1901–2006 period (106 years). In order to limit the problem linked to the seasonal cycle in temperatures, the high temperature RLs will be computed from the sub-series corresponding to the summer days, when high temperatures are susceptible to occur. The summer period has been defined from empirical statistical analyses regarding the occurrence of very high temperatures as the 100 days between the 14th of June and the 21st of September (Parey *et al*., 2010). Thus a series of 106 times 100 days is extracted from the observation series to produce the series of daily maximum temperatures in summer in Déols between 1901 and 2006. The observed maximum occurred on the 2nd of August 1906 with a temperature of 40.5°C.

3.1. The block maxima method

The study has firstly been conducted using the block maxima method. The first step consisted in choosing the block length and verifying the identical distribution of the selected maxima between the blocks. To do so, the occurrence of block maxima and of the annual maxima within each block have been compared for two block sizes: 25 days (four blocks per summer) and 50 days (two blocks per summer). The results are summarized in Table 1. The maxima are reasonably well represented in each block. However, for the 25-day blocks, annual maxima seems to occur more rarely at the beginning (block 1) and at the end of the summer (block 4), revealing the remaining of a seasonal cycle, even though the summer season only has been retained. In order to avoid a fastidious de-seasonalization procedure, a 50-day block length will be chosen, for which the repartition in each block is more similar. This leads to 2 maxima per summer, thus 212 values over the period and is a good compromise between the length of the series of extremes and the length of the block needed to ensure independence and identical distribution of the selected maxima.

The trend identification procedure for polynomials leads to a quadratic trend for the location parameter $\mu(j)$, j being the block number, whereas the scale parameter σ can be considered as constant.

Table 1. Repartition of the 424 block maxima and the 106 summer maxima in the 4 blocks for 25-day blocks and of the 212 block maxima and 106 summer maxima in the 2 blocks for 50-day blocks

	25-day blocks in summer				50-day blocks in summer	
	14/06–08/07	09/07–02/08	03/08–27/08	28/08–21/09	14/06–02/08	03/08–21/09
Block maxima	92	115	105	112	102	110
Summer maxima	16	44	37	9	60	46

DOI: 10.1002/env

Figure 1 (left panel) shows the optimal identified trend superimposed on the block maxima evolution. The extrapolation of this trend gives the following results for the 30-, 50-, and 100-year RLs (RL30, RL50, and RL100, respectively), with their 90% confidence intervals:

RL30 = 42.0°C [40.0–43,5]; RL50 = *44.6°C [41.2–46.3]*; RL100: *54.0°C [42.0–58.6]*

The 50- and 100-year RLs are in italics as the convergence of the optimization is not robust. This example shows how polynomial trends may lead to very high, physically unrealistic temperature values. The quadratic form is due to the occurrence of high temperature values at the beginning and at the end of the sample.

Another proposed way of evaluating the RLs is to look for the optimal CPL trend, and use it to define the period over which the trend should better be considered, if one wants to avoid high frequency fluctuations contained in the sample. The optimal CPL trend for the daily maximum temperature in summer in Déols is found in two parts for the location parameter $\mu(j)$, while the scale parameter remains constant. The identified break point is in year 1974. Figure 1 (right panel) shows this optimal CPL trend, superimposed on the block maxima evolution.

The extrapolation of this last linear part, which could be judged as more representative of the recent warming trend, leads to the following RLs:

RL30 = 42.0°C [39.7–43.9]; RL50 = 44.1°C [40.6–45.7]; RL100 = *49.3°C [41.4–51.7]*

Here again, although the convergence is better achieved, the optimization is still difficult for the 100-year RL.

The two approaches, using the optimal polynomial trend or the last part of the optimal CPL trend, lead to very similar results for the 30- and 50-year RLs. The extrapolation to 100-year RLs remains however not realistic, probably because of the very high values encountered at the end of the sample, whereas when beginning after the break point, the sub-sample starts with the lowest block temperatures reached in the sample. It thus does not seem reasonable to assume an unchanged trend on such a long period.

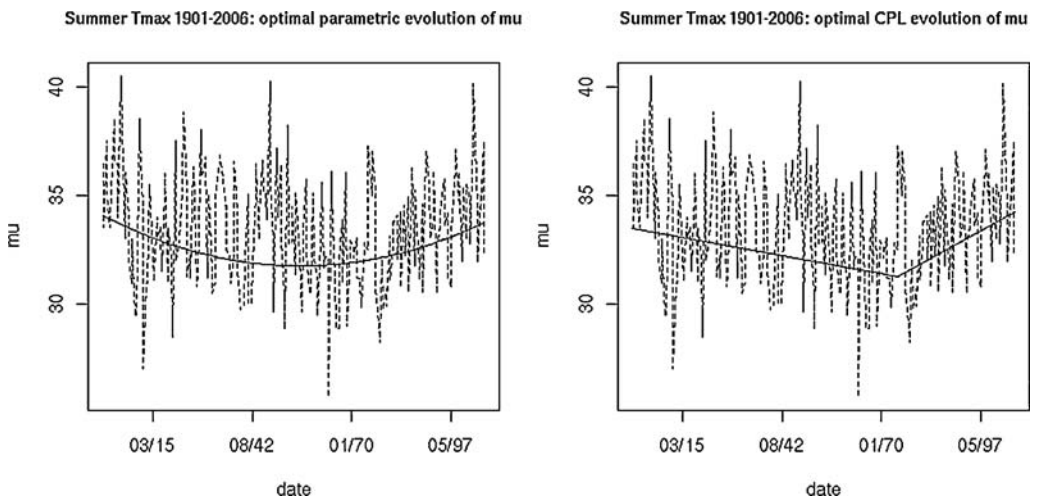

Figure 1. Optimal polynomial trend for the location parameter $\mu(j)$ (j is the block index) (left panel) and optimal CPL trend for the location parameter $\mu(j)$ (right panel), superimposed on the evolution of the block maxima, for the daily maximum temperature in summer in Déols between 1901 and 2006

DOI: 10.1002/env

3.2. The POT method

3.2.1. Trend extrapolation. The same analysis has then been conducted with the POT method.

The first task is here to choose the threshold and the declustering procedure. The declustering procedure used is a simple one, consisting in retaining only the maximum value when several exceedances occur on consecutive days. The threshold selection is then guided by the properties of the GPD, as stated in Section 2.1, to which we add the necessity for the identified trends in the Poisson process intensity and in the scale parameter of the GPD to remain constant. This leads to a threshold of 31.4°C, with 400 independent exceedances for the whole series length.

Then, the optimal polynomial trend is identified for both the Poisson process intensity $I(t)$ and the scale parameter of the GPD $\sigma(t)$: the result is a quadratic trend for $I(t)$ and a constant $\sigma(t)$. The trend in the Poisson process intensity is illustrated in Figure 2 (left panel), superimposed on the frequency of threshold exceedances each summer. The 30-, 50-, and 100-year RLs obtained by extrapolating this trend are the following:

RL30 = 39.6°C [38.8–40.9]; RL50 = 40.1°C [39.3–41.4]; RL100 = 40.7°C [39.9–42.5]

The obtained 30- and 50-year RLs are lower than those obtained previously with the block maxima method.

As previously, the optimal CPL trend has then been identified. Here, the results are different, since the optimal trend for the Poisson process intensity has three parts, with the last one strongly influenced by the edge effect of the recent hot summers in France, as can be seen in Figure 2 (right panel), with break points, respectively, in 1965 and 1999. The scale parameter is still constant. The best fitted two parts evolution, although not optimal, identifies the break point in 1997, which confirms the strong edge effect of the last hot summers in the series. In order to overcome this effect, the first idea could be to extrapolate the second part rather than the last one. This trend is however very low, and probably artificially low, as the effect of the recent high values is totally excluded. It has thus been decided to cut the sample at the date of the identified second break point, and to look for the same optimal trends over

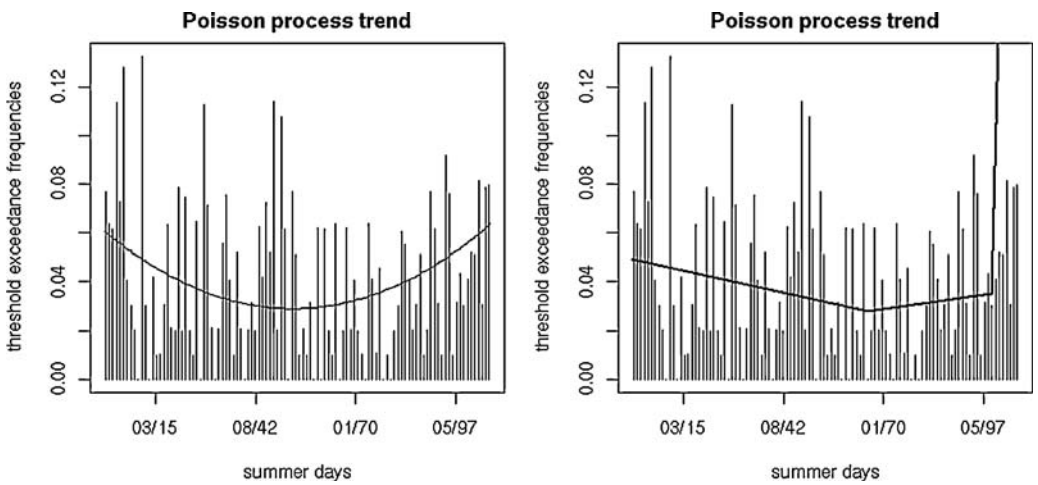

Figure 2. Optimal polynomial (left panel) and CPL (right panel) trend for the Poisson process intensity superimposed on the frequencies of threshold exceedances each summer, for the daily maximum temperature in summer in Déols between 1901 and 2006

DOI: 10.1002/env

the 1965–2006 period. The optimal trend is then linear for the intensity of the Poisson process, while the scale parameter of the GPD remains constant. The extrapolated RLs are the following:

RL30 = 39.3°C [37.5–41.3]; RL50 = 39.7°C [37.7–42.2]; RL100 = 40.4°C [37.8–44.5]

3.2.2. Effect of a change in the location parameter of the GEV or in the intensity of the Poisson process. It seems from the previous results that an evolution in the Poisson process intensity of the POT method has a lower impact on the RL than an evolution of the location parameter of the GEV distribution. In order to confirm this observation, the derivation of the RL for the two methods against these parameters are computed. The calculations here are conducted for the stationary RLs, as their analytical expressions are easier to derivate than the proposed expressions for the non stationary RLs (Equations (1) and (2) above). The only aim is to understand the effects of an increase in the location parameter of the GEV distribution or in the intensity of the Poisson process on the estimated RL.

For GEV, the T-year stationary RL Z_T writes:

$$Z_T = \mu - \frac{\sigma}{\xi}\left[1 - (-\log(1-T))^{-\xi}\right] \tag{9}$$

in the case where $\xi \neq 0$ like here, and then $\frac{\partial Z_T}{\partial \mu} = 1$.For POT, the same T-year stationary RL Z_T writes:

$$Z_T = u + \frac{\sigma}{\xi}\left[\left(Tn_yI\right)^{\xi} - 1\right] \tag{10}$$

where n_y is the number of exceedances each year and thus the mean number of declustered exceedances of the threshold per year. Then:

$$\frac{\partial Z_T}{\partial I} = \sigma\left(Tn_yI\right)^{\xi-1} \text{ and } \frac{\partial Z_T}{\partial \sigma} = \frac{1}{\xi}\left[\left(Tn_yI\right)^{\xi} - 1\right] \tag{11}$$

with the values of the parameters in the studied case, this leads to the results summarized in Table 2. One can see that the impact of an increase in I on the RL is much lower and decreases when the return period increases, whereas the impact of an increase of the scale parameter σ is much higher and increases with the return period. This explains the observed differences between the RL extrapolation according to the two extreme value methods.

3.3. Research of a break point in the evolution of the mean

Until now, the trends have been identified for the whole sample of extremes, that is over the entire period, and the CPL form of the trends lead us to use sub-periods. Another idea could be to look for a break in the evolution of the mean daily maximum temperatures in summer, and then to consider only the period beginning after this break. In order to do this, the method used consists in adjusting a

Table 2. Values of the derivation of the RL against the Poisson process intensity I and the scale parameter of the Pareto distribution σ for the 30-, 50-, and 100-year return periods

Return period	30-year	50-year	100-year
$\partial Z_T/\partial I$	0.0077	0.0041	0.0017
$\partial Z_T/\partial \sigma$	2.87	3.03	3.22

DOI: 10.1002/env

two-part linear trend to the series of daily maximum temperatures in summer and identifying the break point as the one for which the standard deviation of the residuals is the minimum. All dates of the series are tested as potential break points, except the 5 first and last summers in order to minimize the edge effects. This follows the method proposed by Mudelsee (2009). This procedure leads to a break in the trend of mean daily maximum temperature evolution in 1956, as illustrated in Figure 3, together with the smooth loess trend used in the following to compute the centered and normalized variable.

Starting from this result, the evolution of the parameters of the extreme value distributions for the GEV and POT methods has been computed over the period 1957–2006. For GEV, the optimal polynomial trends are of degree 1 (linear) for the location parameter μ and 0 (stationary) for the scale parameter σ. The extrapolated RLs and their 90% bootstrap confidence interval are then:

RL30 = 40.8°C [39.3–42.1]; RL50 = 42.2°C [40.2–43.5]; RL100: *45.6°C [40.8–46.5]*

These values are lower than those obtained previously for the trend identified over the total period length or over the 1975–2006 period after the break in the trend of the location parameter, as the 50-year RL now is comparable to the 30-year RL previously obtained.

When using the POT method, first the threshold has been selected in the same way as previously as 31.0°C, corresponding to 200 independent exceedances. The optimal trends are of degree 2 (quadratic)

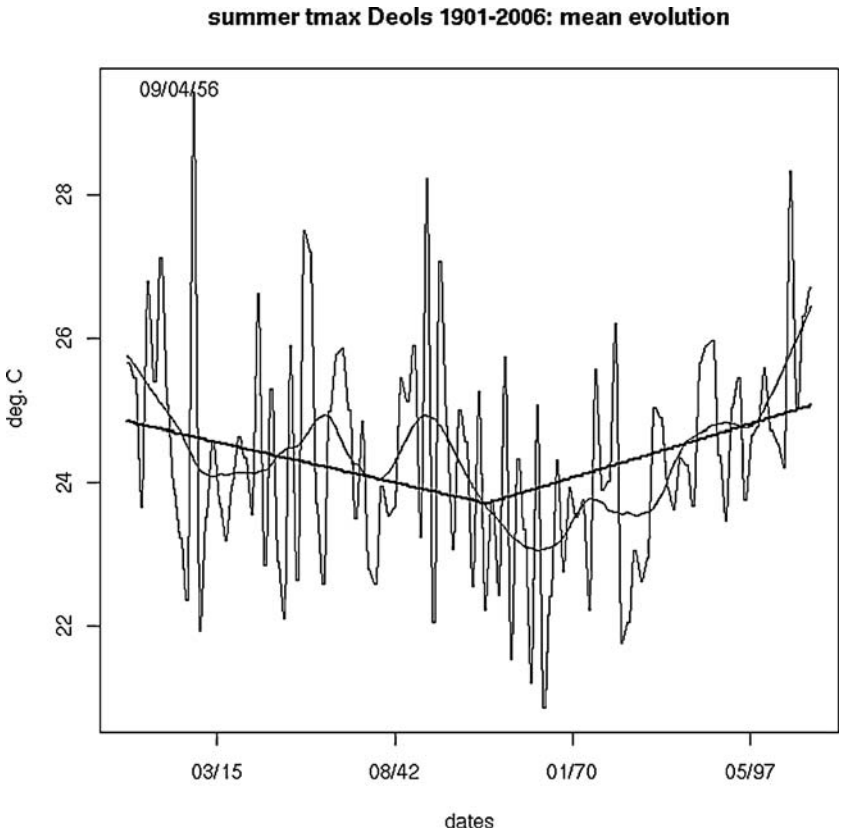

Figure 3. Nonparametric (LOESS, thin smooth line) and optimal two pieces linear trends (bold) for the summer daily maximum temperature in Déols between 1901 and 2006, superimposed on the evolution of the seasonal means. The date of the break point is written in the top left corner

DOI: 10.1002/env

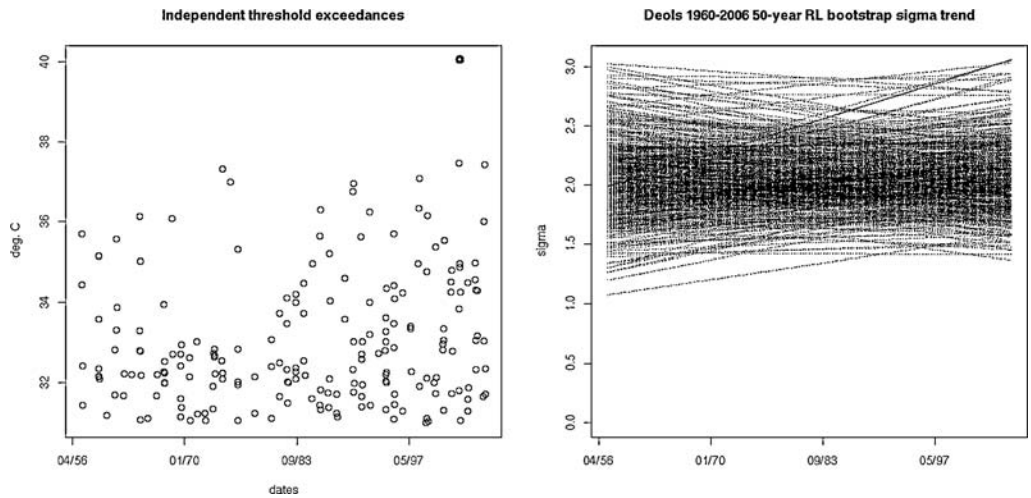

Figure 4. Threshold exceedances over the 1957–2006 period with the 2003 maximum in bold (left panel) and the optimal scale parameter trend (solid line) together with the 500 bootstrap scale parameter trends (dotted lines) (right panel)

for the intensity I of the Poisson process and 1 (linear) for the scale parameter of the Pareto distribution σ. Then the extrapolated RLs and their 90% bootstrap confidence intervals are the following:

RL30 = 41.3°C [35.9–38.6]; RL50 = 42.9°C [35.9–39.9]; RL100 = 46.9°C [36.0–42.1]

what appears first is the fact that the extrapolated RLs lie outside their bootstrap confidence intervals. This is due to the fact that the 2003 maximum is far above the other high levels and over this shorter period, it alone implies the obtained linear trend for the scale parameter σ. Figure 4 shows the selected independent values over the threshold on the left, and the identified optimal trend of the scale parameter together with all the different trends computed from the bootstrap samples in the evaluation of the confidence interval (dotted lines) on the right. Very few trends in the bootstrap samples are as steep as the optimal trend of the original sample (some are even decreasing, depending on the location of the 2003 event in the bootstrap sample). And, as seen before, the evolution of the scale parameter has the greatest influence on the computed RL in the POT method. This explains why the extrapolated RLs are outside the bootstrap confidence intervals. This shows then that this methodology of extrapolation of the trends in the parameters of the POT distributions is very sensitive to unusually high values encountered at the end of the period, and cannot be reliably applied in such cases. In this case, it is therefore recommended to deal with the longest possible series to avoid such a drawback.

4. EXAMPLE: USE OF THE K HYPOTHESIS

As discussed in Section 2.3, previous studies opened the possibility to use the link between the evolution of the mean and variance of the whole summer daily maximum temperature series and the trend in extremes to derive future extreme levels. The steps are first to check the K hypothesis for the Déols series and then to chose the way to evaluate mean and variance in the desired future period.

DOI: 10.1002/env

4.1. The K hypothesis

As stated before, the K hypothesis supposes that the extremes of the centered and normalized series $Y(t)$ are stationary. It has to be tested in comparing the distance between the evaluation of the parameters of the extreme value distribution as constant and as time varying to a distribution of such distances obtained from a similar distribution known as stationary, as described in Section 2.3. The results for the Déols series in summer in the case of block maxima and POT are illustrated in Figure 5 and show that the K hypothesis can be accepted.

4.2. Nonstationary return levels

As stated before, the K hypothesis can first be used to compute RLs in the nonstationary context, but by extrapolating the trends in mean and variance of the whole data series instead of those of the extreme value distribution parameters. In order to do so, it is first necessary to identify a parametric trend in the mean and variance of the observed summer temperatures.

In Section 3.3, a method has been used to identify a break point in the trend in mean daily maximum temperature in summer in Déols. The same method is applied to the evolution of the variance of the same temperatures, and leads to the identification of a break point in the same 1956 year. These two-part linear trends will then be used to compute the RLs by extrapolating the second parts (Figure 6).

The 30-, 50-, and 100-year RLs are then computed following (7) and (8) and their 90% confidence intervals using the bootstrap technique described in Section 2.3.3 above. The results are summarized in Table 3. Both methods lead to comparable values.

4.3. Stationary return levels

Another proposed approach consists in supposing that over fixed periods of time, the RLs can be estimated in the stationary context. These periods have however to be long enough to allow reliable RL estimations, but not too long so that trends cannot be ignored. The period length will be chosen here as

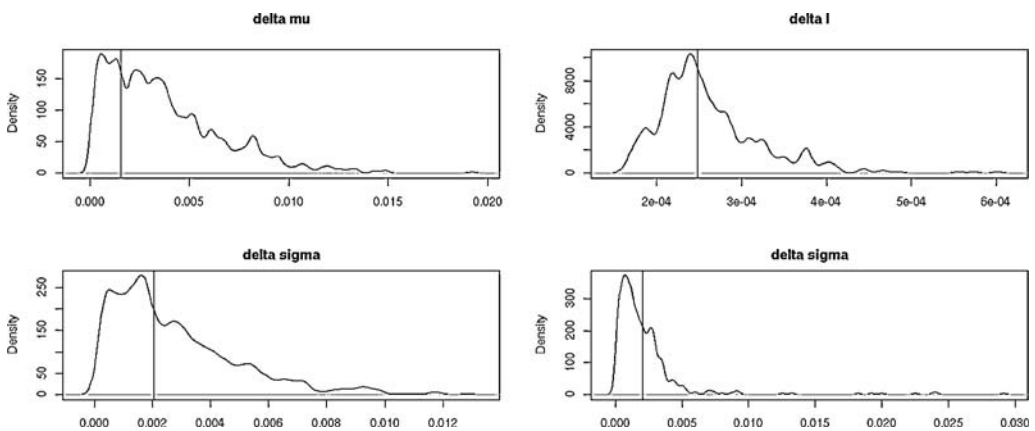

Figure 5. Distributions of the distances between constant and time varying parameters of the extreme value distributions in the case of constant distribution and the same distances for $Y(t)$ in summer in Déols (vertical lines), for GEV (location parameter μ and scale parameter σ; left panel) and for POT (intensity I of the Poisson process and scale parameter σ of the GPD; right panel)

DOI: 10.1002/env

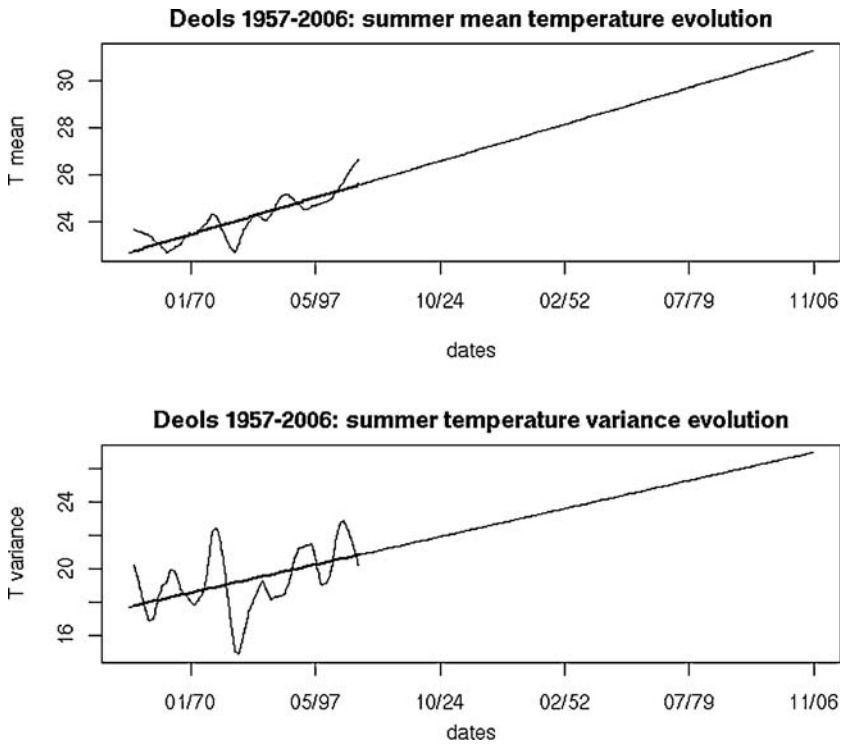

Figure 6. Extrapolation of the second part of the 2-part linear trends in mean and variance of the daily maximum temperatures in summer in Déols, beginning 1956

30 years, the length being recommended by the World Meteorological Organization to define climatology. Thus, the present day climate RLs will be computed for the last 30 years of the observation series, that is here for the 1977–2006 period. Then, two future 30-year fixed periods will be defined in accordance with the current outputs of the climate simulations: the periods 2021–2050 and 2046–2065. Over these future periods, the RLs will be computed from the stationary RLs of $Y(t)$, defined over the observed period, and the changes in mean and standard deviation between each future period and the present one. It must be recalled here that the definition of the RL in this stationary context differs a bit from the one used in the previously set nonstationary context: here, the A-year RL is the value reached or exceeded average once every A years over the studied period, whereas in the non stationary context, the same level corresponds to a value whose probability to be exceeded in the next A years equals 1, that is the value which is expected to be reached or exceeded only 1 day in the A future years.

Table 3. Nonstationary RLs using the K hypothesis for GEV and POT, with their bootstrap 90% confidence intervals

	30-year RL	50-year RL	100-year RL
GEV	40.3 [38.9–42.0]	41.2 [39.4–43.3]	43.3 [40.5–46.3]
POT	39.9 [38.4–41.0]	40.8 [38.9–42.2]	42.9 [40.0–45.4]

DOI: 10.1002/env

Two ways of computing the changes in mean and standard deviation of the summer daily maximum temperatures will be tested: the extrapolation of the previously identified linear trends, and the use of climate model simulations.

4.3.1. Mean and variance trends extrapolation. The previously identified trends in mean and variance of the observed summer series are extrapolated, and the mean and standard deviations over the periods 2021–2050 and 2046–2065, respectively, are calculated. The results are summarized in Table 4. Thus the future RLs are estimated using the stationary RLs for the centered and normalized variable $Y(t)$, z_Y, and the future values of the mean (m_F) and standard deviation (s_F) as: $z_X = z_Y s_F + m_F$ where z_X is the desired RL for the original series $X(t)$. The 90% confidence intervals are computed by using the delta-method, where the variances for the mean and variance are added to the variance of z in the computation of the covariance matrix. The results are shown in Table 5. Not surprisingly, the two extreme value methods give similar results in that case. However, the confidence intervals are much larger for the POT method than for GEV. This is due to a larger variance of the estimated RL with POT, probably meaning that the estimation is less robust.

4.3.2. Use of climate model results. In the framework of the European project ENSEMBLES, different Regional Climate Models (RCM), forced using different General Circulation Models (GCM) have been used to simulate future climate under different greenhouse gas evolution scenarios. We will use here six of these models, with a 25 km resolution and using the Intergovernmental Panel on Climate Change (IPCC) Special Report on Emission Scenarios (SRES) A1B scenario. The simulations cover the period 1950–2100 (1960–2100 for one of the models). In order to assess RLs for the two selected future periods, the mean and variance changes will be computed between the 2021–2050 and 1977–2006 periods, and the 2046–2065 and 1977–2006 periods, respectively, for the modeled series corresponding to the nearest grid point of the station Déols in France, and added to the observed

Table 4. Mean, standard deviation, and their respective increases for the periods 1977–2006, 2021–2050, and 2046–2065 computed from the observed series of summer daily maximum temperatures in Déols (1977–2006) and from the extrapolation of the second part of the identified 2-part linear trends for the future 2021–2050 and 2046–2065 periods

	1977–2006	2021–2050	Increase	2046–2065	Increase
Mean	24.56	25.87	1.31	26.55	1.99
Standard deviation	4.50	4.68	0.18	4.79	0.29

Table 5. 30-, 50-, and 100-year RLs and their 90% confidence intervals estimated in the stationary context according to the K hypothesis for the two future periods, by extrapolating the mean and standard deviation trends derived from the observation series, with the GEV and POT methods

		30-year RL	50-year RL	100-year RL
2021–2050	GEV	40,3 [39,5–41,1]	40,7 [39,8–41,6]	41,1 [40,1–42,1]
	POT	40,3 [33,3–47,4]	40,7 [33,6–47,9]	41,2 [33,9–48,4]
2046–2065	GEV	41,2 [40,3–42,0]	41,5 [40,6–42,3]	42,0 [41,0–43,0]
	POT	41,4 [34,1–48,6]	41,7 [34,4–49,1]	42,2 [34,8–49,7]

DOI: 10.1002/env

Table 6. List of the ENSEMBLE project model simulations used

Acronym	DMI-ARPEGE	DMI-ECHAM5	KNMI-RACMO	MPI-ECHAM5	SMHI-BCM	SMHI-ECHAM5
Institute	DMI	DMI	KNMI	MPI	SMHI	SMHI
Scenario	A1B	A1B	A1B	A1B	A1B	A1B
GCM	ARPEGE	ECHAM5	ECHAM5	ECHAM5	BCM	ECHAM5
RCM	HIRHAM	HIRHAM	RACMO	REMO	RCA	RCA
Resolution	25 km	25 km	25 km	25 km	25 km	25 km
Period	1950–2100	1950–2100	1950–2100	1950–2100	1960–2100	1950–2100

1977–2006 mean and standard deviation. Table 6 presents the simulations used and Table 7 summarizes the mean and standard deviation changes for the two periods. As expected, there is an important dispersion in the climate model results for both the change in mean and the change in standard deviation, and the values previously obtained by extrapolating the observed trends (Table 4) lie in the range of the model values. For the 2021–2050 period, the change in mean given by the models is rather lower than the extrapolated one, except for DMI-ARPEGE. For the 2046–2065 period however, the change in mean computed from the extrapolated trend is rather low compared to model values. A comparison between the series of observed annual temperature maxima and modeled ones for the nearest grid points (not shown) have shown that the DMI-ARPEGE model exhibits a warm bias, DMI-ECHAM5, SMHI-BCM, and SMHI-ECHAM5 a rather cold one and KNMI-RACMO and MPI-ECHAM5 give correct results. The use of these changes, together with the stationary RLs for $Y(t)$ as previously described, lead to the 30-, 50-, and 100-year RLs (with their 90% confidence intervals) summarized in Table 8 for both GEV and POT methods. Here again, the two extreme value methods lead to comparable results, although the POT method gives larger confidence intervals. Concerning climate models, besides DMI-ARPEGE which gives much higher levels (but has been seen to have a warm bias over the current period), they all give comparable orders of magnitude for the future RLs.

5. DISCUSSION AND PERSPECTIVES

In previous papers, especially Parey *et al*. (2007), a computation of temperature RLs in the nonstationary context had been proposed, based on time evolutions of the extreme value distribution parameters.

Table 7. Changes in mean and standard deviation between each future period and the current 1977–2006 period given by the climate simulations of each model for the nearest grid point of Déols

Model	2021–2050		2046–2065	
	Change in mean	Change in std	Change in mean	Change in std
DMI-ARPEGE	2.03	0.29	3.57	0.81
DMI-ECHAM5	0.47	0.01	1.47	−0.02
KNMI-RACMO	0.89	0.17	2.62	0.33
MPI-ECHAM5	0.70	0.09	2.20	1.16
SMHI-BCM	−0.03	−0.11	1.38	0.46
SMHI-ECHAM5	1.05	0.21	2.49	0.31

DOI: 10.1002/env

Table 8. 30-, 50-, and 100-year RLs and their 90% confidence intervals computed in the stationary context according to the K hypothesis by using climate simulation results to calculate the mean and standard deviation changes in the future, with both GEV and POT methods

Methods	2021–2050			2046–2065		
	30-year RL	50-year RL	100-year RL	30-year RL	50-year RL	100-year RL
			GEV			
DMI-ARPEGE	41.4 [40.6–42.3]	41.8 [40.9–42.7]	42.2 [41.3–43.2]	44.6 [43.7–45.5]	45.0 [44.1–46.0]	45.5 [44.4–46.5]
DMI-ECHAM5	39.0 [38.2–39.8]	39.4 [38.5–40.2]	39.8 [38.8–40.7]	39.9 [39.1–40.7]	40.3 [39.4–41.1]	40.7 [39.7–41.6]
KNMI-RACMO	39.9 [39.1–40.7]	40.3 [39.7–41.7]	40.7 [39.7–41.7]	42.2 [41.3–43.0]	42.5 [41.6–43.4]	43.0 [42.0–43.9]
MPI-ECHAM5	39.5 [32.5–46.4]	39.9 [32.9–46.8]	40.3 [33.2–47.4]	41.2 [40.4–42.0]	41.6 [40.7–42.5]	42.0 [41.0–43.0]
SMHI-BCM	38.1 [37.3–38.9]	38.5 [37.6–39.3]	38.9 [37.9–39.8]	41.3 [40.5–42.1]	41.7 [40.8–42.6]	42.1 [41.1–43.1]
SMHI-ECHAM5	40.2 [39.4–41.0]	40.6 [39.7–41.4]	41.0 [40.0–42.0]	42.0 [41.1–42.8]	42.3 [41.5–43.2]	42.8 [41.8–43.8]
			POT			
DMI-ARPEGE	41.4 [34.2–48.7]	41.8 [34.5–49.2]	42.3 [34.8–49.7]	44.6 [36.8–52.4]	45.0 [37.1–52.9]	45.6 [37.5–53.5]
DMI-ECHAM5	39.0 [32.2–45.8]	39.4 [32.5–46.3]	39.8 [32.8–46.8]	39.9 [32.9–46.9]	40.3 [33.2–47.3]	40.7 [33.5–47.8]
KNMI-RACMO	39.9 [32.9–46.9]	40.3 [33.2–47.4]	40.8 [33.6–47.9]	42.2 [34.7–49.5]	42.6 [35.1–50.0]	43.0 [35.4–50.5]
MPI-ECHAM5	39.5 [38.2–40.8]	39.9 [38.5–41.8]	40.3 [38.8–41.8]	41.2 [34.0–48.4]	41.6 [34.3–48.9]	42.1 [34.6–49.4]
SMHI-BCM	38.1 [31.4–44.8]	38.5 [31.7–45.2]	38.9 [32.0–45.7]	41.3 [34.0–48.5]	41.7 [34.4–49.0]	42.2 [34.7–49.6]
SMHI-ECHAM5	40.2 [33.1–47.2]	40.6 [33.4–47.7]	41.1 [33.8–48.2]	42.0 [34.6–49.3]	42.4 [34.9–49.8]	42.9 [35.3–50.3]

DOI: 10.1002/env

In this paper, we extend these results in both introducing the computation of bootstrap confidence intervals and discussing different approaches to determine the time evolution of the parameters, in the context of both the GEV and POT methods. Furthermore, the identified link between mean and variance evolution and the evolution of extremes (Hoang *et al*. 2009) is used to estimate future RLs from these mean and variance evolutions. The different approaches are illustrated on the example of the daily maximum summer temperature in Déols, center of France, over the 1901–2006 period. This example shows that the extrapolation of the optimal parametric trends as polynomials or CPL functions generally leads to similar results, but extrapolation sometimes gives unrealistic levels for long return periods, especially for quadratic evolutions. The use of the CPL functions is helpful to identify sub-periods over which the trend can be more reasonably extrapolated, but lets the sample begin with the lowest identified temperature levels.

Another way to select such sub-periods is to search for a break point in the linear evolutions of the mean and variance of the whole dataset and select the second part as the studied sub-period. In this non stationary context, the block maxima method generally leads to higher RLs than the POT one, as the influence of an increase in the location parameter for the GEV distribution is higher than that of an increase in the Poisson process intensity for the POT method. On the other hand, the POT method has been shown to be very sensitive to very high values located at the end of the sample.

Another proposed way of estimating RLs in the nonstationary context is through the use of the link between the evolution of extremes and the evolution of the mean and variance of the whole dataset. The application of the described test to the summer temperature series of Déols showed that the so-called K hypothesis, under which the extremes of the centered and normalized data can be considered as constant, is accepted for both GEV and POT approaches. Then, again two ways can be used to estimate the future RLs. First, in keeping the nonstationary RL definition, one can compute the value reached or exceeded once in a future period of time by extrapolating the trends in mean and standard deviation of the whole data series instead of those of the extreme value distribution parameters. On the other hand, the RLs for future fixed periods (2021–2050 and 2046–2065 in this paper) can be obtained from the stationary extremes of the centered and normalized quantity and the mean and standard deviation of the summer daily maximum temperature in the future period, in a stationary context. The RL is then the stationary one, that is the value reached or exceeded in average once over the future period.

Table 9. 30-, 50-, and 100-year non stationary RLs computed by extrapolating the identified trends in the parameters of the extreme value distributions or the identified trend in mean and variance of the whole sample according to the K hypothesis

	30-year RL	50-year RL	100-year RL
Trends in EVT parameters			
GEV 1901–2006	42.0°C [40.0–43.5]	*44.6°C [41.2–46.3]*	*54.0°C [42.0–58.6]*
1975–2006	42.0°C [39.7–43.9]	44.1°C [40.6–45.7]	*49.3°C [41.451.7]*
1957–2006	40.8°C [39.3–42.1]	42.2°C [40.2–43.5]	*45.6°C [40.8–46.5]*
POT 1901–2006	39.6°C [38.8–40.9]	40.1°C [39.3–41.4]	*40.7°C [39.8–42.0]*
1965–2006	39.3°C [37.5–41.3]	39.7°C [37.7–42.2]	40.4°C [37.8–44.5]
1957–2006	41.3°C [35.9–38.6]	42.9°C [35.9–39.9]	46.9°C [36.0–42.1]
Trends in total mean and variance			
GEV	40.3°C [38.9–42.0]	41.2°C [39.4–43.3]	43.3°C [40.5–46.3]
POT	39.9°C [38.4–41.0]	40.8°C [38.9–42.2]	42.9°C [40.0–45.4]

DOI: 10.1002/env

Table 10. 30-, 50- and 100-year stationary RLs computed for two future periods by evaluating the changes in mean and variance of summer daily maximum temperature by extrapolating the observed trends or by using climate model simulations, according to the K hypothesis

Methods	2021–2050			2046–2065		
	30-year RL	50-year RL	100-year RL	30-year RL	50-year RL	100-year RL
GEV						
Observed trends	40.3 [39.5–41.1]	40.7 [39.8–41.6]	41.1 [40.1_42.1]	41.2 [40.3–42.0]	41.5 [40.6–42.3]	42.0 [41.0–43.0]
DMI-ARPEGE	41.4 [40.6–42.3]	41.8 [40.9–42.7]	42.2 [41.3–43.2]	44.6 [43.7–45.5]	45.0 [44.1–46.0]	45.5 [44.4–46.5]
DMI-ECHAM5	39.0 [38.2–39.8]	39.4 [38.5–40.2]	39.8 [38.8–40.7]	39.9 [39.1–40.7]	40.3 [39.4–41.1]	40.7 [39.7–41.6]
KNMI-RACMO	39.9 [39.1–40.7]	40.3 [39.7–41.7]	40.7 [39.7–41.7]	42.2 [41.3–43.0]	42.5 [41.6–43.4]	43.0 [42.0–43.9]
MPI-ECHAM5	39.5 [38.7–40.3]	39.9 [39.0–40.7]	40.3 [39.3–41.2]	41.2 [40.4–42.0]	41.6 [40.7–42.5]	42.0 [41.0–43.0]
SMHI-BCM	38.1 [37.3–38.9]	38.5 [37.6–39.3]	38.9 [37.9–39.8]	41.3 [40.5–42.1]	41.7 [40.8–42.6]	42.1 [41.1–43.1]
SMHI-ECHAM5	40.2 [39.4–41.0]	40.6 [39.7–41.4]	41.0 [40.0–42.0]	42.0 [41.1–42.8]	42.3 [41.5–43.2]	42.8 [41.8–43.8]
POT						
Observed trends	40.3 [33.3–47.4]	40.7 [33.6–47.9]	41.2 [33.9–48.4]	41.4 [34.1–48.6]	41.7 [34.4–49.1]	42.2 [34.8–49.7]
DMI-ARPEGE	41.4 [34.2–48.7]	41.8 [34.5–49.2]	42.3 [34.8–49.7]	44.6 [36.8–52.4]	45.0 [37.1–52.9]	45.6 [37.5–53.5]
DMI-ECHAM5	39.0 [32.2–45.8]	39.4 [32.5–46.3]	39.8 [32.8–46.8]	39.9 [32.9–46.9]	40.3 [33.2–47.3]	40.7 [33.5–47.8]
KNMI-RACMO	39.9 [32.9–46.9]	40.3 [33.2–47.4]	40.8 [33.6–47.9]	42.2 [34.7–49.5]	42.6 [35.1–50.0]	43.0 [35.4–50.5]
MPI-ECHAM5	39.5 [38.2–40.8]	39.9 [38.5–41.8]	40.3 [38.8–41.8]	41.2 [34.0–48.4]	41.6 [34.3–48.9]	42.1 [34.6–49.4]
SMHI-BCM	38.1 [31.4–44.8]	38.5 [31.7–45.2]	38.9 [32.0–45.7]	41.3 [34.0–48.5]	41.7 [34.4–49.0]	42.2 [34.7–49.6]
SMHI-ECHAM5	40.2 [33.1–47.2]	40.6 [33.4–47.7]	41.1 [33.8–48.2]	42.0 [34.6–49.3]	42.4 [34.9–49.8]	42.9 [35.3–50.3]

DOI: 10.1002/env

The computation of future mean and standard deviation is proposed in two ways: first, by extrapolating the identified trends in mean and variance of the observed series, and secondly by using climate model simulations. This has been tested for a selection of RCM simulations conducted in the framework of the European ENSEMBLES project.

Table 9 summarizes the different RLs obtained with the nonstationary definition and Table 10 those obtained with the stationary one for fixed future periods. Globally, the extrapolation of trends in the parameters of the extreme value distributions have limitations when levels associated to high return periods are concerned. As a matter of fact, the assumption that the trend will remain the same becomes very strong for long future periods. On the other hand, the trends in EVT are derived from quite limited samples regarding the total number of values in the series, and are therefore certainly less robust than trends identified from the whole sample. In that sense, the use of the K hypothesis seems more promising. The best way could then be to derive the mean and standard deviation changes from an as wide as possible range of climate simulations, to get a distribution of possible RLs for different future periods. However, further studies should be devoted to the analysis of the climate model behavior regarding mean, and especially variance evolution at the local scale. This could then be used to weight the different simulations in the distribution of obtained RLs.

ACKNOWLEDGEMENTS

The authors thank the referees and the Editor for their helpful comments.

REFERENCES

Coles S. 2001. *An Introduction to Statistical Modeling of Extreme Values. Springer Series in Statistics XIV*, 228 p. 77 illus., ISBN 978-1-85233-459-8, Hardcover.

Dacunha-Castelle D, Gassiat E. 1999. Testing the order of a model using locally conic parametrization: population mixtures and stationary ARMA processes. *The Annals of Statistics* **27**(4): 1178–1209.

Davison AC. 2008. Statistical Models Cambridge Series in Statistical and Probabilistic Mathematics (No. 11).

Gumbel EJ. 1958. *Statistics of Extremes*. Columbia University Press: New York; 375.

Hoang TTH. 2010. Modélisation de series chronologiques non stationnaires, non linéaires: application à la definition des tendances sur la moyenne, la variabilité et les extrêmes de la température de l'air en Europe. PhD thesis work, chapter 6, pp 120-138.

Hundecha Y, St-Hilaire A, Ouarda TBMJ, El Adlouni S, Gachon P. 2008. A nonstationary extreme value analysis for the assessment of changes in extreme annual wind speed over the Gulf of St Lawrence, Canada. *Journal of Applied Meteorology and Climatology* **47**: 2745–2759.

Katz RW, Parlange MB, Naveau P. 2002. Statistics of extremes in hydrology. *Advances in Water Ressources* **25**: 1287–1304.

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

Leadbetter MR, Lindgren, G, Rootzen, H. 1983. *Extremes and Related Properties of Random Sequences and Series*. Springer Verlag: New York.

Mudelsee M. 2009. Break function regression. *The European Physical Journal Special Topics* **174**: 49–63.

Parey S, Malek F, Laurent C, Dacunha-Castelle D. 2007. Trends and climate evolutions: statistical approach for very high temperatures in France. *Climatic Change* **81**: 331–352.

Parey S, Dacunha-Castelle D, Hoang TTH. 2010. Mean and variance evolutions of the hot and cold temperatures in Europe. *Climate Dynamics* **34**: 345–369. DOI: 10.1007/s00382-009-0557-0.

Smith RL. 1989. Extreme value analysis of environmental time series: an application to trend detection in ground-level ozone. *Statistical Science* **4**: 367–393.

Zhang XF, Zwiers FW, Li G. 2004. Monte Carlo experiments on the detection of trends in extreme values. *Journal of Climate* **17**: 1945–1952.