Clim Dyn (2008) 30:99–112
DOI 10.1007/s00382-007-0275-4

Extremely high temperatures in France at the end of the century

S. Parey

Received: 9 January 2007 / Accepted: 10 May 2007 / Published online: 29 June 2007

Abstract Power plant construction requires anticipation to achieve a liable dimensioning on the long functioning time of the installation. In the present climate change context, dimensioning towards extremely high temperature for installations intended to run until the 2070s or later implies an evaluation of plausible extreme values at this time scale. This study is devoted to such an estimation for France, using both observation series and climate model simulation results. The climate model results are taken from the European PRUDENCE (Prediction of Regional scenarios and Uncertainties for Defining European Climate change risks and Effects) project database of regional climate change scenarios for Europe. Comparison of high summer temperature distributions given by observations and climate models under current climate conditions, conducted using Generalized Extreme Value distribution, reveals that only a few models are able to correctly reproduce it. For these models, climate change under IPCC A2 and B2 scenarios leads to differences in the variability of high values, whose proportion has an important impact on future 100-year return levels.

Keywords Climate change · High temperature extremes · Regional climate models

This study was first presented at the EGU General Assembly in Vienna, 2–7 April 2006.

S. Parey (✉)
EDF/R&D, 6, quai Watier, 78401 Chatou Cedex, France
e-mail: sylvie.parey@edf.fr

1 Introduction

Renewal of currently running electricity production facilities has been undertaken in France. The design of such long term running facilities involves the necessity to anticipate future climate evolution. As a matter of fact, dimensioning adaptation measures to a warmer climate taken from the initial stage are less costly than adaptation measures taken during the life cycle of the installation. Therefore, a probable extremely high air temperature for the end of the century needs to be estimated. Such extreme values are usually computed by applying statistical Extreme Value Theory (EVT) to sufficiently long observation series. French climate at the end of the century will probably differ from current observed climate and values estimated in this way will no longer be suitable. Therefore, the only way to obtain such extreme values for the end of the century is to use climate change simulations under different greenhouse gas concentration scenarios.

Impact of climate change on extreme events remains a prevalent enquiry. The task is still difficult because of the lack of very long homogeneous series and the difficulty of climate models to correctly represent rare events (see IPCC report 2001 or Moberg and Jones 2004). Nevertheless, much work has been done or is in progress on this subject, especially within the framework of the European projects PRUDENCE (Prediction of Regional scenarios and Uncertainties for Defining European Climate change risks and Effects, climatic change special issue 2007), STARDEX (Statistical and Regional Dynamical downscaling of Extremes for European regions), MICE (Modelling the Impacts of Climate Extremes) and ENSEMBLES (ENSEMBLE-based Predictions of Climate Changes and their Impacts). Work on extremes for simulated climate are

also conducted in Semenov and Bengtsson (2002) and Huntingford et al. (2003). After comparing present simulated climate to observations regarding extreme events, evaluations are made with EVT in a stationary context for fixed periods in the future according to available simulations.

Some evidence on the evolution of the distribution of extreme and very high values can be found in papers of ''Understanding changes in weather and climate extremes'' (2000 in Bulletin of the American Society: Easterling et al. 2000; Meehl et al. 2000). These papers also include a large bibliography on physical aspects of the problem and on data and studies for various areas in the world.

Many recent works focus on regions of different size, from parts of Europe to the entire world. They are concerned with the evolution of universally accepted indices proposed in order to represent high values of climate variables, such as temperature or precipitation. Based on particular standards (see standards of ECA for instance), these indices do not examine very high, rare events, but enable fair evaluations for very large scales (Yan et al. 2002). Kharin and Zwiers (2000) have compared 20-year return levels obtained from transient simulations to NCEP reanalysis and then studied the changes on a large scale basis.

This paper presents a systematic comparison of high levels of summer temperature in France between observation series and model simulation series at the nearest grid point, for a wide range of regional climate models (RCM). The more recent database of regional climate model simulations for Europe is the one provided by European PRUDENCE project (www.prudence.dmi.dk). Simulations for current climate (1960–1990 period) and future climate (2070–2100 period), using the same sea surface temperature conditions (observed for current climate, issued from coupled Hadley Centre (HC) HadAM3H model under IPCC-A2 or B2 scenarios for future climate), conducted with ten different RCM, have been used. Observation series for France are daily maximum temperature series provided by Météo-France. Only series beginning at least in 1960 are considered here.

In a first part, the climate model simulations and observation series will be presented. Then, the statistical EVT is recalled, and the hypotheses for its application are verified using observation series. The sensitivity of 100-year return levels towards each of the statistical law parameters is discussed. The fourth part is dedicated to comparison of observations and climate simulations under current climate conditions, regarding the characteristics of the GEV distribution for high summer temperature, before coming to future climate. Finally, conclusion and discussion are given.

2 Regional climate model experiments and observational series

2.1 Regional climate model experiments

In the framework of European PRUDENCE project, simulations using both global and RCM have been conducted. Precise description of simulation conditions and models can be found in Déqué et al. (2005). Simulations used in this study are briefly described hereafter.

Regional climate simulations come from one variable resolution Atmospheric General Circulation Model (AGCM): ARPEGE3, used by the Centre National de Recherches Météorologiques (CNRM) and nine RCM: HIRHAM used by the Danish Meteorological Institute (DMI), HadRM3H used by the HC, CHRM used by Eidgenössische Technische Hochschule Zürich (ETH), CLM used by Geesthacht Institute for Coastal Research (GKSS), REMO used by Max Planck Institut für Meteorologie (MPI), RCAO used by Swedish Meteorological and Hydrological Institute (SMHI), PROMES used by Universidad Complutense de Madrid (UCM), RegCM used by International Centre for Theoretical Physics (ICTP) and RACMO2 used by Koninklijk Nederlands Meteorologisch Instituut (KNMI). All regional models are forced with lateral conditions given by simulations of HadAM3H General Circulation Model (GCM) of the HC. Concerning sea surface temperature, both the variable resolution AGCM and the RCMs are forced using observed conditions between 1960 and 1990 for current climate and sea surface temperature anomaly from HC HadAM3H GCM under IPCC-A2 (or B2) scenario added to observed conditions for future climate (2070–2100 period). All model simulations studied here have a resolution of ~50 km over France and provide 30-year simulations for current (1960–1990) and future (2070–2100) climate. CNRM, DMI and HC provide three simulations for each period. SMHI performed one current climate and one future climate simulation using MPI ECHAM4/OPYC sea surface temperature anomaly under IPCC-A2 and B2 conditions. HC, SMHI, UCM, ICTP simulations under IPCC-B2 conditions can also be considered.

All considered model simulations are summarized in Table 1.

2.2 Observational data series

The observational data series used for the study are the daily maximum 2 m temperature series provided by Météo-France (air temperature measured at the 2 m height). These series have been carefully scrutinized by Météo-France in order to provide as homogeneous as possible series for climate evolution studies and are

Table 1 Summary of the model simulations considered

Institute	Model	SST forcing	Simulations
CNRM	ARPEGE	Hadley Centre, A2	3 Current climate + 3 future climate
DMI	HIRHAM	Hadley Centre, A2	3 Current climate + 3 future climate
Hadley Centre (HC)	HadRM3H	Hadley Centre, A2	3 Current climate + 3 future climate
		Hadley Centre, B2	1 Future climate
ETH	CHRM	Hadley Centre, A2	1 Current climate + 1 future climate
GKSS	CLM	Hadley Centre, A2	1 Current climate + 1 future climate
MPI	REMO	Hadley Centre, A2	1 Current climate + 1 future climate
SMHI	RCAO	Hadley Centre, A2	1 Current climate + 1 future climate
		Hadley Centre, B2	1 Future climate
		MPI, A2	1 Current climate + 1 future climate
		MPI, B2	1 Future climate
UCM	PROMES	Hadley Centre, A2	1 Current climate + 1 future climate
		Hadley Centre, B2	1 Future climate
ICTP	RegCM	Hadley Centre, A2	1 Current climate + 1 future climate
		Hadley Centre, B2	1 Future climate
KNMI	RACMO	Hadley Centre, A2	1 Current climate + 1 future climate

referred to as Daily Reference Series (SQR for Séries Quotidiennes de Référence in French). Only series beginning at least in 1960 are considered here, in order to compare with current climate simulations: 36 series over the French territory are selected. Their locations are shown in Fig. 1. All series are considered on the 1960–1990 period, in order to compare with current climate model simulations.

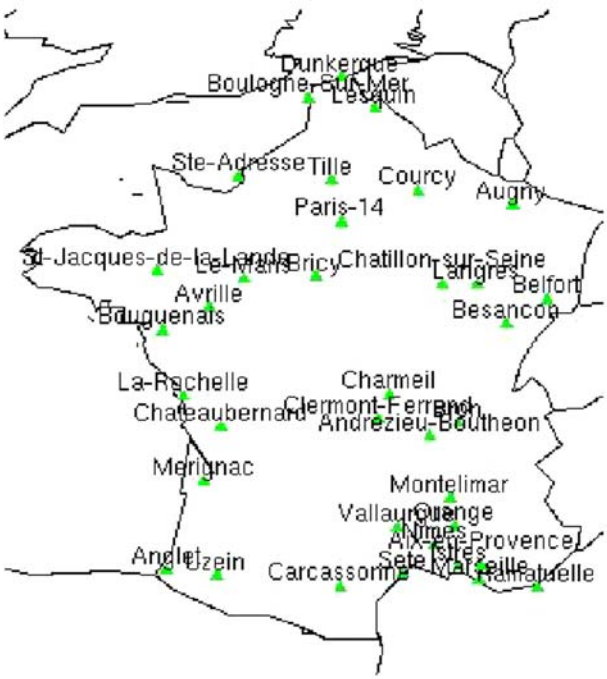

Fig. 1 Locations of the observational data series

2.3 Grid point selection

The first step in comparing models with observations consists in extracting time-series from model simulations, which can correspond to each observational series. The choice has been made here to consider the series of the nearest grid box. As the aim is to study the behaviour of extreme values, this has been preferred to any averaging or combination of grid points, which could smooth or alter extreme values. The selection of the nearest grid point for each model experiment has been done automatically, in comparing grid point distances from observational series location and then choosing the nearest grid point over land. Thirty-six time series are extracted from each model results for each simulation. Then, each temperature series from model simulations is corrected to fit the same altitude as the nearest observation series. This has been made in adding the standard atmosphere gradient of 6.5 K/km to the model temperature values, according to the altitude difference between the model grid point and the observational station. Note that for each series, daily maxima are considered, which correspond to computed daily temperature maxima for model simulations and observed daily maxima for observation series, all reported to the altitude of the observation station. Both observations and simulations provide air temperature at the 2 m height level.

2.4 Summer season simple statistics

Before addressing the study of extremely high summer temperature, mean and standard deviation of the whole analysed sample of temperature values are compared

between observations and models for current climate. For extreme value analyses, a season where extremes currently occur has been defined in previous studies as the 100 days between 14 June and 21 September (Parey et al. 2007). Thus, for each observation and corresponding grid point series, seasonal mean and standard deviation over the 31 100-day seasons have been compared. These comparisons are illustrated in Fig. 2 for each of the 36 observation stations and their corresponding grid point in climate simulations. Generally, models better reproduce seasonal mean temperature than standard deviation. HC model simulates too high temperature levels and a too important variability. On the contrary, UCM model tends to under-estimate summer temperatures, but to a lesser degree. DMI model shows correct results. These comparisons show that the chosen nearest grid point with altitude correction is a legitimate model representation of the observation series.

3 Extreme value theory

3.1 Classical models

The foundation of the EVT relies on two general definitions of extreme events. Those can be considered as maxima of given blocks of time (e.g. a year or a month), described by the Generalized Extreme Value (GEV) distribution. Another specification consists of Peaks Over Threshold (POT), where extremes are retained values over a properly chosen high threshold. These exceedances, when independent and in sufficient quantity, follow a Generalized Pareto Distribution (GPD). The series of dates of occurrences of these events is the trajectory of a Poisson process. The EVT has been primarily developed for the fields of oceanography and hydrology (Coles 2001; Katz et al. 2002).

The GEV distribution has been preferred in this study, because of difficulty in the threshold choice for the POT, which could make the comparison between models and observations more delicate.

Then, under stationary conditions, if M_j are the maxima of a series X_t over defined time blocks, then the distribution of M_j asymptotically converges towards a so called GEV Distribution, for which:

$$\Pr(M_j \leq x) = \begin{cases} \exp\left\{-\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-1/\xi}\right\} & \text{if} \quad \xi \neq 0, \\ \exp\left\{-\exp\left[-\frac{x-\mu}{\sigma}\right]\right\} & \text{if} \quad \xi = 0 \end{cases}$$

and for some defined probability p (for example ones in 100 years, $p = 1/100$), the return level z_p is computed as:

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

The location parameter μ and the scale parameter σ are, respectively, proportional to the mean and the variance of extremes, while the shape parameter ξ describes the shape of the tail ($\xi < 0$ implies that it admits a right bound $z_b = \mu - \sigma/\xi$, $\xi > 0$ describes a left bounded tail, with the same bound, and $\xi = 0$ gilts for an unbounded tail, the well known Gumbel distribution).

3.2 Hypotheses

3.2.1 *Homogeneity and stationarity*

The hypothesis which needs to be examined before applying EVT concerns the homogeneity and stationarity of the series.

Observation series used in this study are the best daily series available regarding homogeneity, as they have been checked for homogeneity by Météo-France (Moisselin 2004). This does not mean that they are fully homogeneous, but that non-homogeneities are minimum.

Stationary means no cycle, nor trends. Temperature data have an obvious seasonal cycle. In order to get rid of it, only days of the hot season, where temperature maxima are more likely to occur, are selected. This season has been defined in a previous study (Parey et al. 2007) as the 100 days between 14 June and 21 September, for France. The same study suggests a method to deal with climate change induced trends. Climate simulations used in this study have been conducted under evolving sea surface temperatures and greenhouse gas concentrations. Some trend identifications conducted on the results of the CNRM ARPEGE-Climate model showed that the simulated trends are not realistic, mainly because sea-surface temperature evolutions are not homogeneous over time and space. As the aim here is to estimate the order of magnitude of future extremes on a fixed period of future time, and as the studied periods are rather short, it has been chosen not to deal with trends. Moreover, this could bring a supplementary source of uncertainty in comparing the behaviour of the different models and make it more complicated.

3.2.2 *Block size*

Maxima on fixed blocks of time have to be extracted from each daily maximum temperature series over the 31 years multiplied by 100 days per season, that is 3,100 values, for the observation and model series between 1960 and 1990. An equilibrium has to be found here between a long block

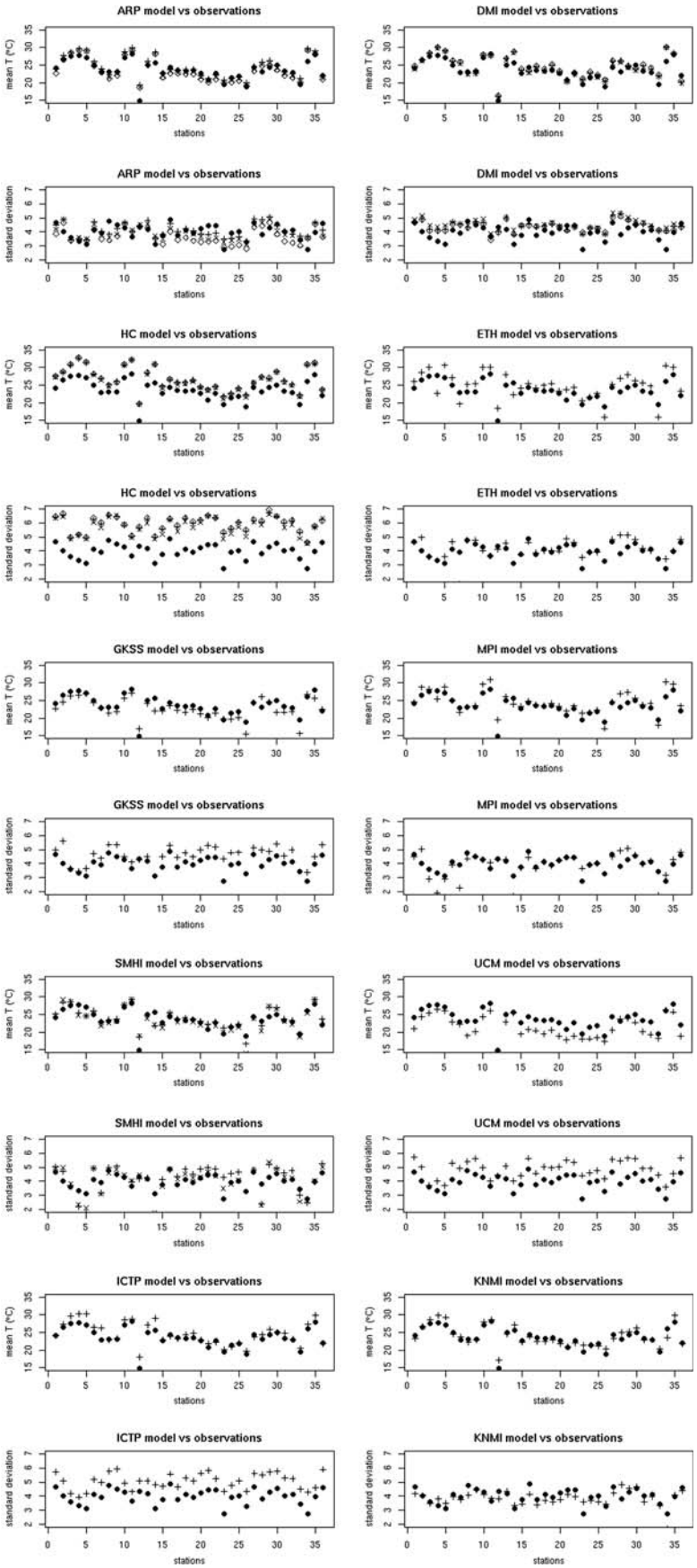

Fig. 2 Seasonal mean (*top*) and standard deviation (*bottom*) at each of the 36 stations for observations (*black dots*) and model simulations (*crosses* and *open signs* for the different simulations)

size, compatible with the asymptotic assumption, and a sufficient number of high level values for a liable statistical approximation. Ideally, the block size should be the season, whose number of days is more compatible with the asymptotic convergence of the theory. But in our case, with only 31 years, this leads to 31 element samples, which is very short for a liable distribution adjustment. Thus, in addition to using the 100-day blocks, we also use 25-day blocks (leading to 124 maxima instead of 31), and compare the results. These results are summarized in Table 2 for 12 series representative of the different geographical locations. When smaller blocks are considered, the location parameter μ logically decreases as the scale parameter σ increases. The shape parameter ξ, which is the most delicate to estimate, is however more precisely evaluated with more blocks, as the standard error is almost divided by 2 using 25-day blocks instead of 100-day blocks. As 100-year return levels are of the same order, with smaller confidence intervals, this 25 days block size has then been used throughout the study. The adjustment is made using maximum likelihood with *R* statistical library.

3.2.3 Independence

To apply EVT, selected high values have to be independent. The study by Parey et al. (2007) showed that high air temperature clusters are generally not very long, with a mean of around 2 days. But a coherent and convergent result of climate projections for the end of the century is towards more frequent and longer heat waves, thus it could be necessary to deal with higher cluster lengths for future climate. This issue is handled in the EVT by introducing the so called ''extremal index'' θ, which can be estimated as the inverse of the mean cluster length. Then, according to Leadbetter et al. (1983), the GEV distribution becomes:

$$\Pr(M_j \le x) = \begin{cases} \exp\left\{-\theta\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-1/\xi}\right\} & \text{if} \quad \xi \neq 0, \\ \exp\left\{-\theta\exp\left[-\frac{x-\mu}{\sigma}\right]\right\} & \text{if} \quad \xi = 0 \end{cases}$$

and the return level for a low probability p is obtained by:

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

Parameter θ, called the extremal index and estimated as the inverse mean cluster length over a very high threshold, varies between 0 and 1. For $\theta = 1$, the traditional formulation of GEV distribution is found.

Mean cluster length has been evaluated for values higher than the 92nd percentile. This percentile value has been chosen because it leads to a similar number of independent values above the threshold as retained using 25-day blocks, that is around 124 (namely between 120 and 130). Figure 3 shows the computed extremal index θ (i.e. inverse of the mean cluster length) for each of the 36 observation series. The extremal index lies between 0.5 and 0.7, which corresponds to a mean cluster length for such extremely high temperatures of less than 2 days.

3.3 Confidence intervals for return levels

The computation of confidence intervals for the return levels can be done using different approaches. Two of

Table 2 GEV parameters and 100-year return levels for 12 selected observation stations with 100-day blocks (left) and 25-day blocks (right), with standard error for ξ and 95% confidence intervals length for return levels in brackets

Station	100-Day blocks				25-Day blocks			
	μ	σ	ξ (SE)	NR100 (IC length)	μ	σ	ξ (SE)	NR100 (IC length)
Charmeil	33.9	2.21	–0.185 (0.092)	40.8 (4.5)	30.7	3.13	–0.253 (0.037)	40.3 (2.5)
Carcassonne	35.0	1.70	–0.194 (0.114)	40.2 (3.8)	31.7	3.01	–0.318 (0.038)	39.8 (1.8)
Chateaubernard	34.3	2.09	–0.393 (0.140)	38.7 (2.4)	31.1	3.23	–0.384 (0.047)	38.6 (1.7)
Montélimar	35.3	1.65	–0.261 (0.113)	39.8 (2.9)	32.1	2.91	–0.329 (0.051)	39.7 (2.2)
St. Jacques de la Lande	31.5	1.98	–0.128 (0.145)	38.4 (6.8)	28.0	3.06	–0.225 (0.057)	38.0 (4.2)
Bricy	32.1	1.84	–0.186 (0.142)	37.7 (5.0)	29.0	2.87	–0.269 (0.053)	37.5 (3.0)
Langres	29.7	1.76	–0.199 (0.123)	35.0 (4.1)	26.8	2.83	–0.299 (0.051)	34.7 (2.4)
Augny	32.1	1.70	–0.204 (0.141)	37.1 (4.3)	28.7	2.89	–0.283 (0.060)	37.0 (3.1)
Tille	30.6	2.09	–0.211 (0.125)	36.8 (4.7)	27.4	2.93	–0.232 (0.051)	36.9 (3.9)
Boulogne	27.6	2.82	–0.252 (0.153)	35.3 (6.1)	24.2	3.47	–0.245 (0.055)	35.1 (4.2)
Uzein	34.3	1.96	–0.316 (0.108)	39.1 (2.5)	31.1	2.78	–0.297 (0.044)	38.9 (2.1)
Ramatuelle	32.8	1.39	–0.335 (0.096)	36.0 (1.5)	30.0	2.57	–0.380 (0.049)	36.1 (1.3)

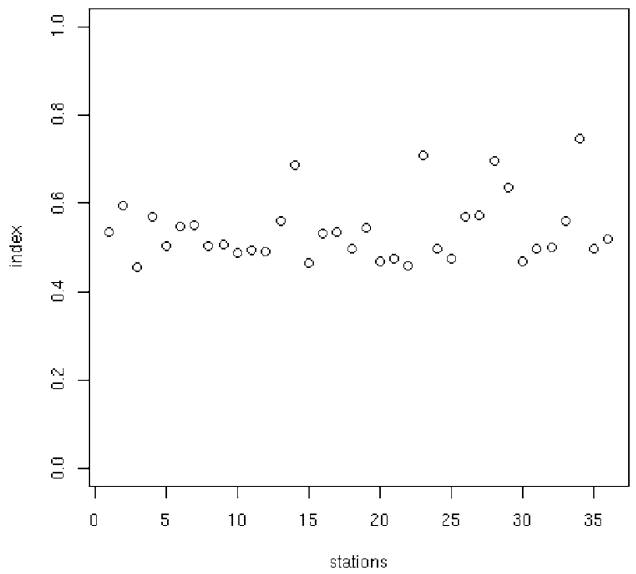

Fig. 3 Extremal index for each of the 36 observation series, computed for 92nd percentile as threshold

them will be exposed and compared here, but the evaluation can be conducted using bootstrap or Bayesian techniques too.

Using the approximate normality of the maximum likelihood estimator or of any function of it, confidence intervals for the return levels can be derived from the standard deviation of the estimated return levels. This is called the ''delta method''.

Another way of estimating confidence intervals is the so called ''profile likelihood method''. This method is based on the following theorem:

Let $x_1,\ldots, x_n$ be independent realizations from a distribution within a parametric family F, and let θ_0 denote the maximum likelihood estimator of the d-dimensional model parameter $\theta_0 = (\theta_i,\ \theta^{(2)})$, where θ_i is a one dimensional subset of θ_0. Then, under suitable regularity conditions, for large n:

$$D_p(\theta_i) = 2\left\{\ell(\hat{\theta}_0) - \ell_p(\theta_i)\right\} \approx \chi^2,$$

where $\ell_p(\theta_i) = \max \ell(\theta_i,\ \theta_{-i})$ is the profile log-likelihood for θ_i, that is, for each value of θ_i, the maximized log likelihood with respect to all other components of θ.

This requires a reparameterization of the GEV model so that the return level z_p is one of the model parameters. This is straightforward as $\mu = z_p + \frac{\sigma}{\xi}\left\{1 - [-\log(1-p)]^{-\xi}\right\}$. Then the deviance D_p is evaluated for different values of z_p and the confidence interval is derived using the $(1 - \alpha)$ quantile of the chi-square distribution, α being the desired confidence level.

3.4 Return level sensitivity to GEV parameters

If the more general formulation is considered for the return level evaluation, z_p, the return level associated to a low probability p, depends on four parameters: the extremal index θ and the three parameters of the GEV distribution: the location parameter μ, the scale parameter σ and the shape parameter ξ. In order to get an idea of the respective influence of each parameter on the 100-year return level, z_p expression has been successively derived against each parameter and a numerical evaluation of each contribution is made for $p = 0.01$ using the observation series.

From $z_p = \mu - \frac{\sigma}{\xi}\left[1 - \left\{\frac{-\log(1-p)}{\theta}\right\}^{-\xi}\right]$, we obtain:

$$\frac{\partial z_p}{\partial \theta} = \sigma(-\log(1-p))^{-\xi}\theta^{\xi-1},$$

$$\frac{\partial z_p}{\partial \mu} = 1,$$

$$\frac{\partial z_p}{\partial \sigma} = -\frac{1}{\xi}\left\{1 - \left[\frac{-\log(1-p)}{\theta}\right]^{-\xi}\right\},$$

$$\frac{\partial z_p}{\partial \xi} = \frac{\sigma}{\xi}\left\{\frac{1}{\xi} - \frac{1}{\xi}\left[\frac{-\log(1-p)}{\theta}\right]^{-\xi} + \left[\frac{-\log(1-p)}{\theta}\right]^{-\xi} \times \log\left(\frac{-\log(1-p)}{\theta}\right)\right\}.$$

Thus, return level z_p linearly depends on μ and σ, while the dependence itself is based on the parameter for θ and ξ. From the numerical application for $p = 0.01$ using observation series, it can be seen that $\frac{\partial z_p}{\partial \sigma}$ is around 2, so an increase of the scale parameter σ has a higher (doubled) influence on 100-year return level than the same increase of the location parameter μ.

The shape of the influence of extremal index θ and shape parameter ξ on 100-year return level, for observed range of values of each parameter, is shown in Fig. 4. θ Lies between 0 and 1, and its influence on the return level is very large for low values. For values higher than around 0.2, the impact of θ on the return level can be neglected. The shape parameter ξ typically lies between –0.125 and –0.44 for the observation series, and it can be seen that the influence of ξ on the return level is higher for lower ξ values (in absolute value).

4 Comparison between observation and simulations for current climate

4.1 Cluster length

The first step in comparing observations and simulations is to verify if climate simulations reproduce a mean cluster

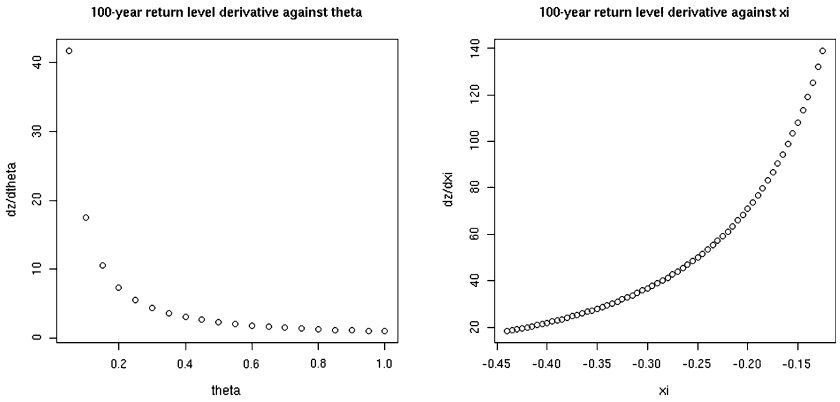

Fig. 4 numerical evaluation of the influence of the extremal index θ (*left*) and the shape parameter ξ (*right*) on the 100-year return level, for the respective ranges of θ and ξ and the estimated parameters for station Charmeil

length for high level temperatures of the same order as observations. The difference in mean between the two 36-value samples of extremal index have been tested using a two sided *t*-test assuming different variances at a 95% confidence level. The results show that all models tend to significantly under-estimate extremal index (or over-estimate cluster length), with values around 0.4, which leads to a mean cluster length of around 2.5 days rather than 0.54 (less than 2 days) for observations. Thus, all models tend to over-estimate persistence over a high level. Extremal index averages for observations and each model simulations are summarized in Table 3. Nevertheless, extremal index remains larger than 0.2 and the influence of θ on return level is weak. Return level can then be evaluated with $\theta = 1$ as for the observations, in order to better analyse the influence of the GEV parameters.

4.2 GEV parameters μ, σ and ξ, and 100-year return levels

In order to objectively compare the results, it was chosen to count the number of stations where the simulated values for parameters μ, σ and ξ of the fitted GEV distribution, as well as 100-year return levels, fall inside the 95% confidence interval estimated from the corresponding observation series. GEV parameter 95% confidence interval bounds are constructed using the delta method, i.e. considering the distribution of parameters as Gaussian. For the return level, the comparison has been made using the profile likelihood method for confidence intervals. Results are shown in Fig. 5 for the GEV parameters and in Fig. 6 for the return levels. For each model and each simulation, the percentage of grid points falling inside the 95% confidence interval of the corresponding observation series is plotted in a bar plot and the 50% line is added. The figures show that models generally have difficulties to correctly reproduce the parameters of the GEV distribution, and particularly the location parameter μ. This parameter is either 1 or 2 degrees lower or around 1–3 degrees higher than the observed one, without any clear behaviour from the models. Except for the HC RCM, all models better reproduce scale parameter σ and shape parameter ξ. Thus, as the location parameter has a smaller influence on the return level than the scale parameter for instance, when the error on the location parameter is not too high, the return level may still be correct, and thus the percentage of return levels falling inside the 95% confidence interval of observations is always higher than the one for the location parameter. The way of computing the confidence interval has a weak influence on the percentage of grid points. Profile likelihood allows asymmetrical intervals towards warm levels, which generally slightly enhances the number of values falling inside the interval, except for models with cold bias as GKSS or UCM models.

As stated in Moberg and Jones (2004), high temperatures in France simulated with HC RCM are much larger than observed, coherently with the previous comparison of mean values. The shape parameter is however rather well reproduced. The tail of the distribution is too high and with a too large variability but its shape is generally correct. On the contrary, GKSS, UCM and KNMI models tend to simulate too cold extremes. This is true for mean values

Table 3 Mean of extremal index over the 36 stations (or grid points) for observations and models current climate simulations (one line per simulation)

Observations	CNRM	DMI	HC	ETH	GKSS	MPI	SMHI	UCM	ICTP	KNMI
0.54	0.36	0.44	0.38	0.43	0.40	0.46	0.42	0.46	0.36	0.47
	0.37	0.44	0.41				0.41			
	0.40	0.45	0.37							

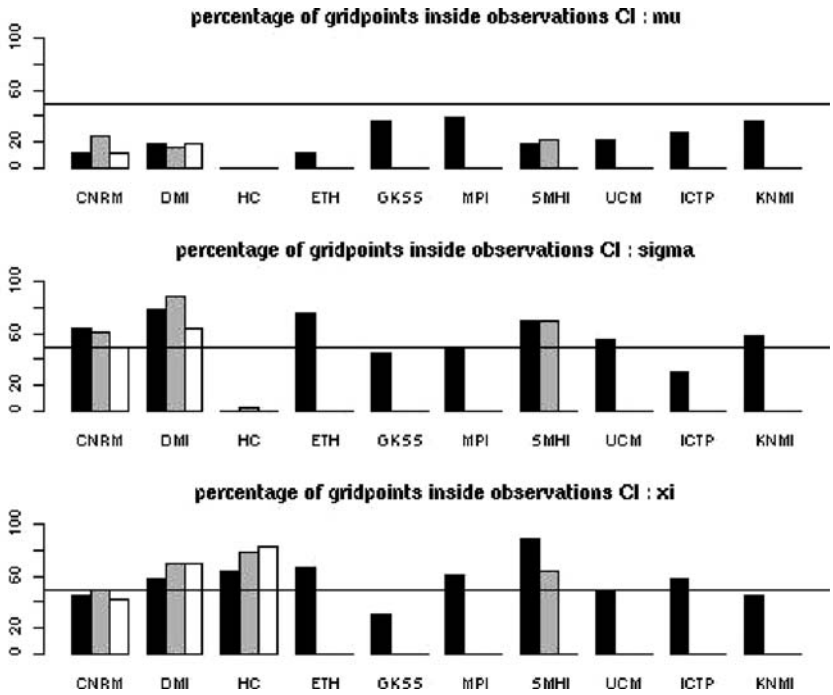

Fig. 5 Percentage of grid points for each model simulation (*black*: simulation 1, *grey*: simulation 2, *white*: simulation 3) falling inside the 95% confidence interval of observations for the parameters of GEV distribution, *top*: the location parameter, *middle*: the scale parameter, *bottom*: the shape parameter

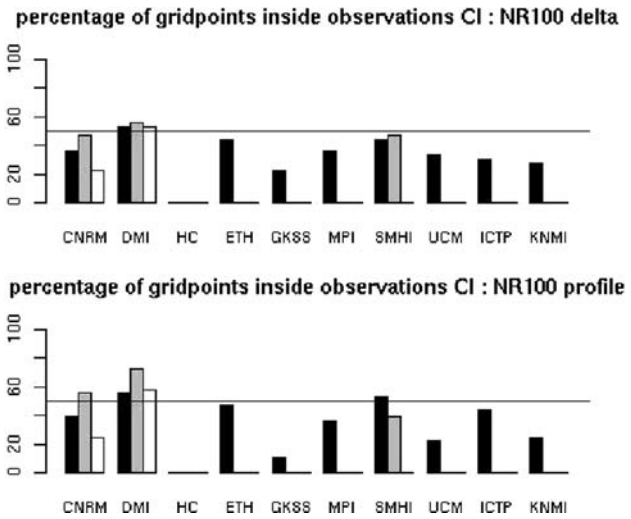

Fig. 6 Percentage of grid points for each model simulation (*black*: simulation 1, *grey*: simulation 2, *white*: simulation 3) falling inside the 95% confidence interval of observations for the 100-year return levels. *Top*: confidence interval calculated with the delta method, *bottom*: confidence intervals calculated with the profile likelihood method

too, as stated in §2.4 above. Therefore, models show generally a similar behaviour in reproducing extreme values as mean values, but the correct reproduction of the whole tail of the distribution seems much more difficult.

If we wished to select only the models showing an acceptable representation of high temperature distribution, where more than half of the grid points fall inside the 95% confidence interval of observation series for at least two parameters, including the return level, then only DMI model, SMHI model, ETH model and simulation 2 of CNRM model can be retained. As an illustration of a model with such a correct reproduction of high levels, 100-year return levels estimated from observations and DMI model simulation 2 are shown in Fig. 7, where 100-year return levels falling inside the profile likelihood 95% confidence interval of the observations are written in larger print.

5 Estimations for future climate

The previous section aimed at verifying model representation of extremes. Even if a correct representation for current climate is not a warranty for a correct representation of the future, it seems reasonable to base the evaluation of future 100-year return levels on models which lead to the better representations of present day levels. Thus, the analysis for future climate will be made for DMI, SMHI and ETH models and simulation 2 of CNRM model. Between those, only SMHI performed simulations under both A2 and B2 greenhouse gas emission scenarios.

5.1 Extremal index

In a first step, in order to verify if the classical formulation of GEV can still be applied, the mean cluster length for these simulations are computed in the same way, that is with a threshold taken as the 92nd percentile. Simulations for the end of the century present a significant decrease in

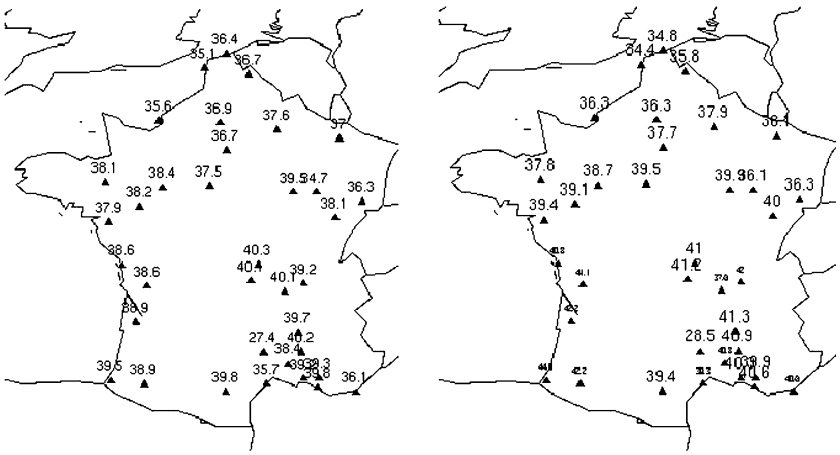

Fig. 7 100-Year return levels computed from observation series (*left*) and current climate DMI simulation 2 (*right*), where grid points where simulated 100-year return levels are correctly reproduced (confidence interval computed with the profile likelihood method) are written in larger print

the extremal index in comparison with simulations of present climate (significance has again been tested using a two-sided *t*-test assuming different variances with a 95% confidence level). The values lie between 0.3 and 0.4 (0.35 for CNRM; 0.39, 0.42 for DMI simulations 1 and 2; 0.39 for ETH; 0.35, 0.38 for SMHI), which corresponds to a mean cluster length of 2.5–3 days. Only DMI simulation 3 (0.46) does not show such a difference. On the other hand, SMHI results (Fig. 8 for simulation 1) show that the mean cluster length increases (that is the extremal index decreases) from current to future climate under B2 scenario and then future climate under A2 scenario. Differences between future climate under B2 scenario and current climate are significant for both simulations (mean extremal index: 0.42, 0.41 for current climate; 0.38, 0.39 for future climate under B2 scenario). On the other hand, significant differences between future climate under B2 scenario and future climate under A2 scenario are found for simulation 1 only (mean extremal index: 0.38 for future climate B2 scenario and 0.35 for future climate A2 scenario, whereas for simulation 2, mean extremal index is 0.39 for B2 scenario and 0.38 for A2 scenario). This may be due to the differences in the applied sea surface temperature forcing. Mean cluster length for high temperature levels thus tends to increase with the severity of induced climate change, from around 2.4 days under current conditions (models over-estimate the observed length of less than 2 days) to around 2.8 days at the end of the century under A2 scenario, with a value around 2.6 days at the end of the century under B2 scenario, for SMHI simulation 1. This result is however not robust because SMHI simulation 2 computes similar extremal index for future climate under A2 and B2 scenario, still significantly different from current index. Even though model simulated length is larger than the observed one and the computed lengths are not precise, the result seems to show a possible increase in hot period durations, which is coherent with a convergent signal of lengthening of heat waves. There is thus globally an increase in the duration of high level temperature episodes, but not enough to justify that the extremal index must be taken into account in the return level evaluation.

5.2 Future 100-year return levels

The aim of the study is to estimate possible future 100-year return levels considering that greenhouse gas emission is

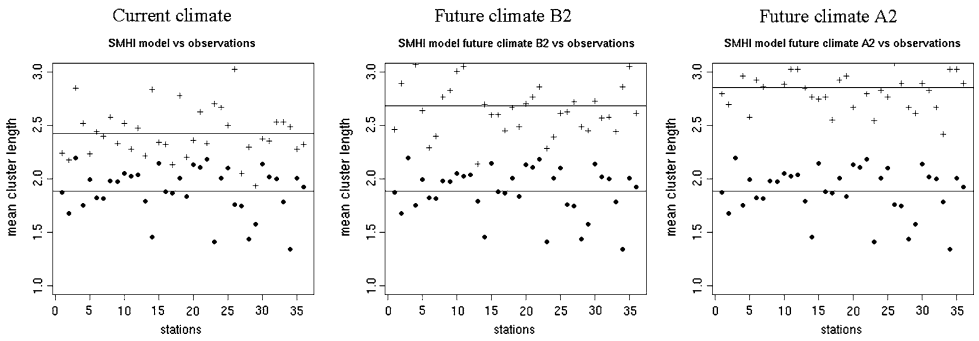

Fig. 8 Mean cluster length (days) for each of the 36 observation series, computed for 92nd percentile as threshold, for observations (*black dots*) and for SMHI model simulation 1 (*crosses*) for current climate, future climate under B2 scenario and future climate under A2 scenario. *Straight lines* represent the mean cluster length over all 36 stations for observations and model results, respectively

going to evolve. Thus, the next step, after having verified the consistency of the extremal index, is to compare 100-year return levels obtained with the selected model simulations. Here again, results for the grid points where current 100-year return levels are correctly reproduced will be emphasized in the analysis.

5.2.1 100-year return level results

Observed 100-year return levels for the 36 considered stations range between 27.4 and 40.3°C. All models give significantly higher 100-year return levels for the end of the century ranging between (considering all 36 grid points in order to avoid geographical bias):

CNRM simulation 2: 35.7 and 50.6°C,
DMI simulation 1: 40.5 and 54.1°C,
DMI simulation 2: 40.9 and 55.8°C,
DMI simulation 3: 40.6 and 52.8°C,
ETH: 24.7 and 48.9°C,
SMHI simulation 1: 25.9 and 52.2°C; B2: 24.6 and 50.1°C,
SMHI simulation 2: 24.2 and 53.3°C; B2: 23.6 and 50.5°C.

These results show that 100-year return levels are of the same order of magnitude for greenhouse gas scenario A2 with ETH model and greenhouse gas scenario B2 with SMHI model, whereas all models besides ETH give higher levels for A2 scenario. An illustration is given in Fig. 9 for DMI simulation 2, ETH and SMHI simulation 1 with B2 scenario, compared to current observed levels. ETH model behaviour is then different from other models, and this may be due to different behaviours in the evolution of the parameters of the GEV distribution.

5.2.2 Differences in shape and scale parameters

In order to understand why the ETH model shows a different behaviour from the other models, differences in shape and scale parameters of the GEV distribution between future and current climate have been investigated. Location parameter has a lower impact on return levels and with higher temperatures, it will necessarily increase, thus the explanation cannot be found through this parameter. Results are summarized in Fig. 10 for each model simulation, that is all three DMI simulations, both SMHI simulations under A2 and B2 scenario, CNRM simulation 2 and ETH simulation. The bars correspond to the percentage of grid points where the parameter falls inside the 95% confidence interval of current climate simulation, i.e. points where parameters are similar in the future and currently. The results show that the shape parameter ξ is, for most of the grid points, inside the confidence interval of current climate series, unlike scale parameter σ, which is not, except for ETH simulation. Thus, models generally produce a significantly different scale parameter for the tail of summer temperature distributions of the future from the one of the current climate, except ETH, whereas the shape parameter remains globally of the same order of magnitude. The comparison of scale parameters between present and future climate, for grid points where differences are significant, shows that the scale parameter is higher in future climate simulations, for all model simulations. Table 4 summarizes these comparisons for SMHI simulation 1 under both A2 and B2 scenarios. The increase in scale parameter is slightly lower for B2 scenario than for A2, but there is a clear increase in the future. The same is true for the other models. Thus, the difference in behaviour of ETH model seems to be clearly linked to this difference in change of the scale parameter: when all other considered models show an increase in the scale parameter in the future, ETH model does not and rather tends to simulate a scale parameter of the same order of magnitude. We have seen in Sect. 3 that 100-year return levels depend linearly on the shape parameter with a factor of around 2, so this can be an explanation for the higher return levels simulated by models predicting an increase in the shape parameter in the future.

6 Conclusion and perspectives

The aim of the study was to estimate a possible extremely high temperature level for the end of the century in France. To do this, ten climatic models results have been analysed: one variable resolution AGCM and nine RCM, over Europe, and compared to nearest observation series. Simulations concern current climate (1960–1990 period) and future climate (2070–2100 period) under IPCC-A2 and B2 scenarios, using mostly HC coupled model forcing for future sea surface temperature conditions, and have been produced in the framework of European PRUDENCE project. Comparisons between current climate observed and modelled mean and standard deviation over the whole sample of summer temperatures show that the models generally correctly reproduce the main features, except for the HC model which is much too warm, with a too high variance. Generally, the average is better simulated than the variance.

First, the conditions to apply EVT, and particularly block maxima, have been verified. The cluster length of very high temperature is small enough for the data to be considered as independent. A compromise between asymptotic convergence and statistical approximation leads to retain 25-day blocks over the 3,100 days of the hot seasons, taken between 14 June and 21 September each year. A sensitivity study of return levels towards each EVT

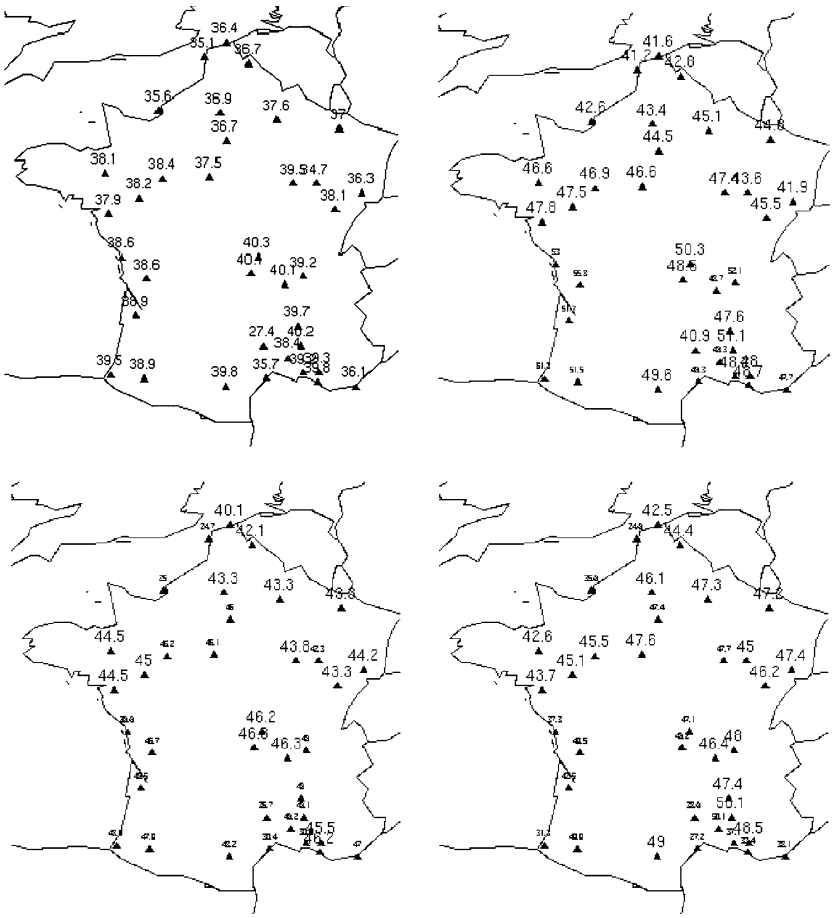

Fig. 9 100-Year return levels for the end of the century compared to currently observed ones (*top left*) as computed for DMI simulation 2 under A2 scenario (*top right*), ETH simulation under A2 scenario (*bottom left*) and SMHI simulation 1 under B2 scenario (*bottom right*), with grid points where current climate levels are correct (according to profile likelihood confidence intervals) written in larger print

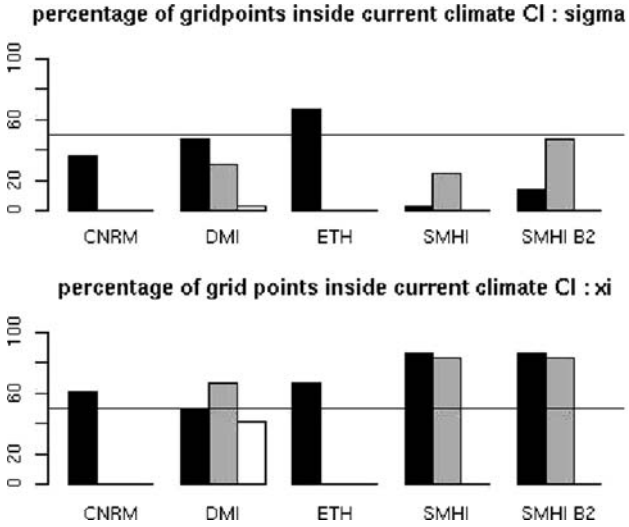

Fig. 10 Percentage of grid points for each future climate model simulation (*black*: simulation 1, *grey*: simulation 2, *white*: simulation 3) falling inside the 95% confidence interval of current climate simulation for the parameters of GEV distribution, *top*: the scale parameter, *bottom*: the shape parameter

parameter has shown that the scale parameter has a greater influence than the location parameter. The influence of the extremal index is weak as long as this index remains lower than ~0.2, which is the case in our applications. The shape parameter has a higher influence when its absolute value is small, corresponding to less bounded tails.

Then, reproduction of high summer temperature distributions by the different models, fitting GEV distribution, has been compared to observations on the same 1960–1990 period. The first step has been to compare the so called extremal index, which corresponds to the inverse of the mean cluster length over a high threshold. It has been shown that models tend to over-estimate cluster length (or under-estimate the extremal index) compared to observations, leading to around 2.5 days rather than less than 2 for the observations. This means that models tend to simulate a higher persistence than that found in observations.

The comparison of GEV parameters has shown that all models fail in reproducing a correct level for the location parameter. Regarding the other two parameters and

Table 4 Scale parameter σ for current and future climate, for grid points where they significantly differ, for SMHI model simulation 1 under A2 and B2 greenhouse gas emission scenarios

SMHI simulation 1					
Grid point	Current climate	Future climate A2	Grid point	Current climate	Future climate B2
1	2.822974	3.696963	1	2.822974	3.993877
2	2.959558	3.566646	2	2.959558	3.921388
3	2.52289	3.155155	3	2.52289	3.536717
4	1.584072	1.949148	4	1.584072	1.97172
5	1.268759	1.639007	5	1.268759	1.556168
6	2.926473	4.06866	6	2.926473	3.896918
7	2.01453	2.654727	7	2.01453	2.431664
8	2.960686	4.127374	8	2.960686	3.873434
9	2.783691	3.981223	9	2.783691	3.76786
10	2.66498	3.457086	10	2.66498	3.72882
11	2.880008	3.52405	11	2.880008	3.553151
12	2.637765	3.541674	12	2.637765	3.967978
13	2.524574	2.98561	13	2.524574	3.155046
14	0.9904021	1.14157	15	3.257985	3.741271
15	3.257985	4.258016	16	2.994726	3.913747
16	2.994726	3.73497	17	2.964155	3.669509
17	2.964155	4.054219	18	3.083787	3.841554
18	3.083787	4.134179	19	3.188855	3.891762
19	3.188855	4.119651	20	3.129922	3.75909
20	3.129922	4.386538	21	2.786116	3.805349
21	2.786116	3.989547	22	2.954091	3.869512
22	2.954091	4.175814	24	3.155425	3.693365
23	3.314985	4.455214	25	3.193285	3.777688
24	3.155425	4.511124	27	2.723795	3.817253
25	3.193285	4.34547	29	2.703549	3.720798
26	1.28188	1.665232	30	2.949653	3.73442
27	2.723795	3.456022	31	3.19993	3.915745
29	2.703549	3.431978	32	3.17893	3.851154
30	2.949653	3.636723	34	1.838623	2.26755
31	3.19993	4.050481	35	2.879387	3.552137
32	3.17893	4.317181	36	2.898331	3.738558
33	2.322299	3.024954			
34	1.838623	2.180379			
35	2.879387	3.523843			
36	2.898331	3.778232			

100-year return levels, only a few models are able to correctly reproduce at least two of them for more than half of the considered grid points. Generally, model behaviour is similar for mean and extreme value reproduction, but a good mean and standard deviation simulation does not necessarily induce a good reproduction of all features of high temperature distribution.

The models that succeed relatively well in reproducing high summer temperature distribution have future forecasts showing:

1. lower extremal index, that is higher high temperature cluster lengths, increasing from current climate to future climate A2 scenario, with intermediate values for future climate under B2 scenario, or similar values for future climate under A2 and B2 scenario;
2. a significantly higher scale parameter for the GEV distribution, while the shape parameter seems to remain rather unchanged, except for ETH model for which the scale parameter remains quite unchanged too.

The behaviour regarding scale parameter modifications has a non-negligible impact on the 100-year return levels, as shown by the sensitivity study. One hundred-year return levels for the end of the century range roughly between 40.1 and 54.1°C over the grid points with correct current 100-year return levels reproduction, and between 40.1 and 46.3°C for ETH model for which scale parameter does not change much in the future. Thus, the scale parameter increase seems to have a greater influence on 100-year return levels in the future than the greenhouse gas emission scenario. As a matter of fact, future levels obtained with ETH model under A2 scenario are of the same order as those obtained with SMHI model under B2 scenario.

These results have to be regarded as preliminary in the sense that they rely on a few models for which the GEV description of high summer temperature distributions can be considered as quite correct, although 10 models had been studied. Besides, correct reproduction of current feature is not a warranty for correct behaviour under future climate conditions.

Climate models are however designed to correctly reproduce current observed climate in its mean features, and looking for correct reproductions of the tail of distribution on the grid point level is quite a severe criteria regarding this. A recent study (Nogaj et al. 2007) has shown that trends in high temperature distributions in France are strongly linked to the trends in mean and variance of the whole distribution. This result could be used to go further in such comparisons, in focusing on the whole distribution and studying implications for extremes, as models seem to better reproduce the centre of the distribution than the tail. This could also allow us to deal with fully coupled simulations over longer periods, in comparing observed and modelled trends for current periods and exploring climate model results for the future in the light of their behaviour regarding central values.

Acknowledgements This work uses results from European Project PRUDENCE, supported by the European Commission Programme Energy, Environment and Sustainable Development under contract EVK2-2001-00156, and from French GICC-IMFREX project supported by the Department of Environment (MEDD). The author is grateful to Dr. O. B. Christensen (DMI) for preparing the database with regional scenarios and to Dr. M. Déqué for managing project IMFREX and preparing the database. Author's thanks go to Professor Didier Dacunha-Castelle too, for his kind advises regarding statistical aspects of the study. Lastly, thanks to the reviewers whose comments have contributed to greatly improve the paper.

References

Coles S (2001) An introduction to statistical modeling of extreme values. Springer Series in Statistics. Springer, Berlin

Déqué M, Jones RG, Wild M, Giorgi F, Christensen JH, Hassel DC, Vidale PL, Rockel B, Jacob D, Kjellström E, de Castro M, Kucharski F, van den Hurk B (2005) Global high resolution versus Limited Area Model climate change projections over Europe: quantifying confidence level from PRUDENCE results. Clim Dyn 25(6):653–670

Easterling DR, Evans JL, Groisman PYa, Karl TR, Kundel KE, Ambernje P (2000) Observed variability and trends in extreme climate events: a brief review. Bull AMS 81(3):417–425

Huntingford C, Jones RG, Prudhomme C, Lamb R, Gash HHC, Jones DA (2003) Regional climate-model predictions in extreme rainfall for a changing climate. Q J R Meteorol Soc 129:1607–1621

IPCC Third Assessment Report, 2001

Kharin V, Zwiers F (2000) Changes in the extremes in an ensemble of transient climate simulations with a coupled Atmosphere-Ocean GCM. J Clim 12:3760–3788

Katz RW, Parlange MB, Naveau P (2002) Statistics of extremes in hydrology. Adv Water Ressour 25:1287–1304

Leadbetter MR, Lindgren G, Rootzen H (1983) Extremes and related properties of random sequences and series. Springer, New York

Meehl GA, Karl T, Easterling DR, Changnon S, Pielke R, Changnon D, Evans J, Groisman PY, Knutson TT, Kunkel KE, Mearns LO, Parmesan C, Pulwarty R, Root T, Sylves RT, Whetton P, Zwiers F (2000) An introduction to trends in extreme weather and climate projections: observations, socioeconomic impacts and model projections. Bull AMS 81(3):413–416

Moberg A, Jones PD (2004) Regional climate model simulations of daily maximum and minimum near-surface temperature across Europe compared with observed station data 1961–1990. Clim Dyn 23:695–715

Moisselin J-M (2004) Long term reference series of Météo-France. In: Proceedings of the EMS/ECAC conference, Nice

Nogaj M, Parey S, Dacunha-Castelle D (2007) Non-stationary extreme models and a climatic application. Nonlinear process in Geophys (in press)

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

Prediction of Regional Scenarios and Uncertainties for Defining European Climate Change Risks and Effects: the PRUDENCE project. Climatic Change, volume 81, Supplement 1, May 2007

Semenov VA, Bengtsson L (2002) Secular trend in daily precipitation characteristics/greenhouse gas simulation with a coupled AOGCM. Clim Dyn 19:123–140

Yan Z, Jones PD, Davies TD, Moberg A, Bergström H, Camuffo D, Cocheo C, Maugeri M, Demarée GR, Verhoeve T, Thoen E, Barriendos M, Rodríguez R, Martín-Vide J, Yang C (2002) Trends in extreme temperatures in Europe and China based on daily observations. Clim Change 52:355–392