



# Modeling and forecasting extreme hot events in the central Ebro valley, a continental-Mediterranean area

J. Abaurrea\*, J. Asín, A.C. Cebrián, A. Centelles

*Statistical Methods Department, Zaragoza University, Spain*

## Abstract

This work has three objectives, first, to analyze the observed change in the summer maximum daily temperature during the period 1951–2004, in the centre of the Ebro river basin, a region situated in the NE of the Iberian Peninsula. Secondly, to characterize the extreme hot event behaviour by means of a statistical model consisting of a non-homogeneous Poisson process, to represent the occurrence, and three regression models, each with an adequate non-Normal error distribution, to model its severity. The model parameters are allowed to depend on temperature covariates, to take into account the influence of global warming in hot event generating process. Finally, using the fitted model and different outputs from a GCM, we obtain a medium term projection, up to 2050, of the expected behaviour of these extreme events.

© 2006 Elsevier B.V. All rights reserved.

*Keywords:* Ebro valley; heat waves; marked Poisson model; future climate projections

## 1. Introduction

The global warming undergone by mean surface air temperature values during the 20th century and, particularly, in its last decade, will probably continue, associated to the climate change induced by the increasing concentration of greenhouse gases in the composition of the atmosphere (Houghton et al., 2001). This likely scenario, together with the increasing occurrence of extremely hot events (EHE) and, particularly, the exceptionally long heat wave observed in Europe in August 2003, led to the publication of many studies dealing with climatic analyses of these events, such as Beniston and Stephenson (2004) and Schär et al. (2004).

According to the IPCC TAR (Houghton et al., 2001), the observed warming during the 20th century in land-

surface air temperature signals mainly occurred during 1910–45 and 1976–2000, whereas the intermediate period, 1946–75, did not show any significant increase. In Europe, spring and winter mean temperatures followed this pattern, with winters between 1973 and 2002 being, very likely, the warmest of the last half-millennium (Luterbacher et al., 2004). Regarding the summer season, the averaged European temperature anomaly series for the period 1977–2003, shows a linear trend whose slope is 0.7 °C/decade and an exceptionally strong, unprecedented, warming in the 1994–2003 decade. Autumn is the only season where a decrease of temperature is observed in extensive areas of Europe during the 1976–99 period (Klein Tank et al., 2005).

For the European territory, the European Climate Assessment (ECA) project is carrying out a study about the evolution of many signals related to mean values, variability and extreme indices of precipitation and temperature, particularly, for 1946–75 and 1976–2004 periods (Klein Tank et al., 2002; Klein Tank and

\* Corresponding author.

*E-mail address:* [abaurrea@unizar.es](mailto:abaurrea@unizar.es) (J. Abaurrea).

Können, 2003). In the ECA approach, the results for a region are obtained from point results of selected observatories situated in the corresponding area or close to it, when they show coherent evolutions. This can pose serious difficulties: the number of observatories in the Iberian Peninsula for the period 1946–75 is small, 5, and they are concentrated in the West, in Portugal or close to it.

According to the ECA results regarding the Ebro basin, the mean value of maximum daily temperature during 1976–2004 was stable in autumn and increased in the other seasons; this increment was statistically significant in spring and summer and more intense in spring. In a previous work, Abaurrea et al. (2001) analyzed the evolution of maximum and minimum daily temperatures, denoted  $T_x$  and  $T_n$  respectively, at the centre of the Ebro basin during the period 1938–97, obtaining similar conclusions to the ECA project but more precise results. We emphasize that: (i) the  $T_x$  and  $T_n$  evolution is non-homogeneous in the different seasons; during the period 1976–97, for example, the  $T_x$  mean value increases 1.43 °C/decade in spring, 0.96 °C/dec. in summer and 0.51 °C/dec. in winter, being stable in autumn; (ii) for the same time period, a generalized decrease of interannual variability, except in autumn, is observed, which is statistically significant for winter and spring  $T_x$  series and only in spring for  $T_n$ . Our results agree with those

obtained by Brunet et al. (2001) for the Catalanian region, also situated in the NE of Spain.

The aim of this work is to study the extreme hot events observed during the summer season at the centre of the Ebro valley, an area represented by Zaragoza and Huesca observatories. In spite of the small size of the territory, the study is interesting because of the climatic conditions of this region, which undergoes a dry and extreme climate, influenced by the Mediterranean sea, situated to the East, and the continental conditions of the Iberian central plateau to the WSW, and also due to the shortage of studies for this type of climate. We aim to characterize the EHE generating process by fitting a statistical model relating the EHE behaviour and the temperature evolution. We also show the use of this model for providing possible trajectories of the EHE process in a medium term horizon, 2050, as an alternative to regional climate models (RCM), Beniston and Diaz (2004), Schär et al. (2004), Sánchez et al. (2004).

In next section we analyze the observed change in the summer maximum daily temperature during the period 1951–2004 and describe the occurrence rate and severity of the observed hot events. In Section 3 a statistical model able to characterize the EHE generating process is developed. In Section 4 we use the fitted model to obtain a scenario for the EHE process in the interval 2005–2050.



Fig. 1. Map locating Zaragoza, Huesca and HadCM3 grid points (in black).

The most important conclusions resulting from this work are summarized in the last section.

## 2. Descriptive analysis

### 2.1. Daily temperature series

#### 2.1.1. Data

The data to be analyzed are the daily temperature series from the Zaragoza Airport and Huesca Monflorte observatories, provided by the Spanish National Meteorological Institute. Both locations are situated in the central Ebro valley, see Fig. 1, at 240 m and 540 m above sea level, respectively. This basin, with an area of 85,362 km<sup>2</sup>, lies South of the Pyrenees and the Cantabrian mountains and borders to the Southwest on the Iberian chain and the East on the Coastal-Catalonian chain and the Mediterranean sea. Due to this orography, in the central part of the valley there is a NW wind predominance and a dry climate affected by climate extremes. Maximum and minimum daily temperatures are available for both observatories from 1951 to 2004, although some months, 8/1978, 6–8/1999, 6/2000 and also 6 days in 7/2002 are missing in the Huesca record.

#### 2.1.2. Analyzing temperature evolution

To show the evolution of temperature mean values during the period 1951–2004, we plot the lowest smoothed curves (using a 0.30 window parameter) of the maximum and minimum temperature series, for the summer months, JJA, in Zaragoza and Huesca, see Fig. 2. Both signals show a similar evolution in both locations: after a stable period 1951–75, both variables increase until, approximately, 1990; later on, a short

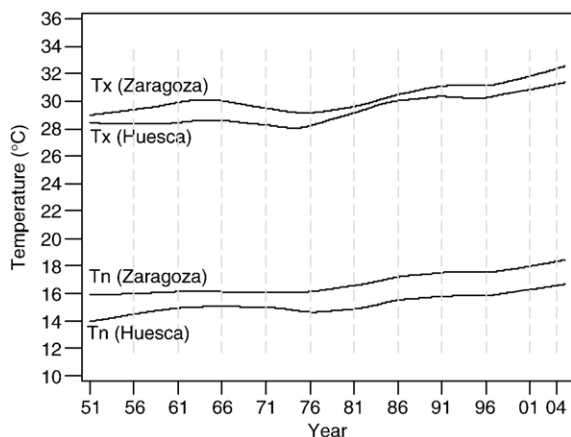


Fig. 2. Evolution (lowest 30%) of summer (JJA) Tx and Tn variables, Zaragoza and Huesca.

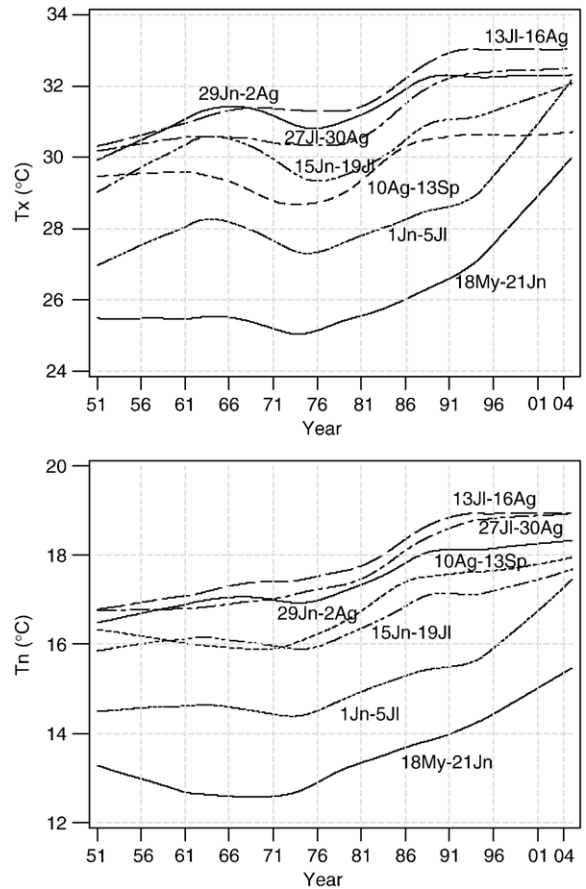


Fig. 3. Evolution (lowest 40%) of Tx (top) and Tn (bottom) in 7 overlapping 5-week periods during the summer, Zaragoza.

stable period, from 1991 to 1996, precedes a final increasing phase.

To analyze the temperature evolution during the summer we plot, see Fig. 3, the smoothed Tx and Tn series from Zaragoza, corresponding to 7 overlapping 5-week length periods, covering the whole summer: 18My–21Jn, 1Jn–5Jl, 15Jn–19Jl, 29Jn–2Ag, 13Jl–16Ag, 27Jl–30Ag and 10Ag–13Sp. As we can see, the 1975–2004 evolution shows two different profiles: one corresponds to the first two curves and the other to the four last, being the third period, 15Jn–19Jl, a transitional one. From the analysis of these plots we notice that:

- (i) During 1975–2004, the greatest Tx increase, nearly 5 °C measured on the smoothed values, occurs at the end of the spring and the beginning of the summer. In these periods, the Tx slope is always positive and becomes steeper from 1994 onwards. The increment of Tx in the middle and

the end of the summer is lower, about 2 °C, and occurs over a shorter interval, 1976–92. The smoothed curves corresponding to four last periods show stability during 1992–2004.

- (ii) The Tn evolution is similar but it shows a smaller increase, 3 °C, at the beginning of the summer and about 1.5 °C in July and August.

As a consequence of this evolution, the summer season in this region has become hotter and more homogeneous; the summer is now more uniformly hot than 30 yr ago. In fact, the temperature difference between the coolest and the hottest summer periods, those centered in June 4th and July 30th respectively, has undergone an important decrease from 1975: in Tx, this difference has changed from 6.25 °C to 3 °C, and, in Tn, from 4.7 °C to 3.5 °C. The observed evolution in Huesca is similar with greater differences between the hottest and the coolest summer periods: 3.5 °C and 1.4 °C in Tx and Tn, respectively.

Fig. 4 shows more simply the situation described above; we see the evolution corresponding to June, July and August temperatures in Zaragoza. For the Tx values we observe the convergence, by the end of the period, of the three monthly signals. We also notice that the increase observed in August from 1975 on, is greater than in July, so that Tx smoothed signals reach the same level and, for Tn, August exceeds July by 0.5 °C.

The seasonal pattern of Tx and Tn is not homogeneous in time. Fig. 5 shows Zaragoza seasonal profiles of Tx and Tn for periods 1951–75, 1976–90, 1991–95 and 1996–2004; the curves are calculated by smoothing the daily data from March 1st to November 30th, after subtracting the corresponding annual JJA mean. It can be observed that: (i) the patterns from periods 1951–75

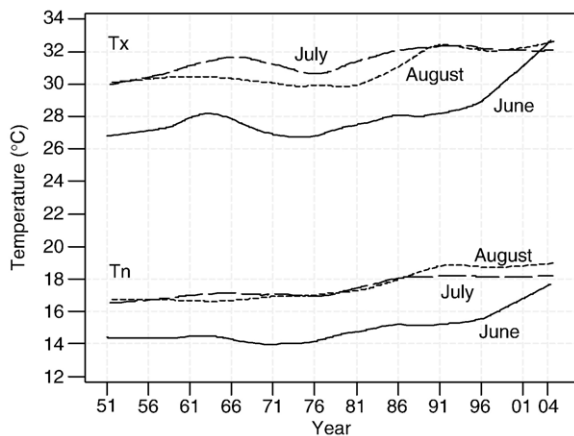


Fig. 4. Evolution (lowest 40%) of Tx and Tn data, per month, Zaragoza.

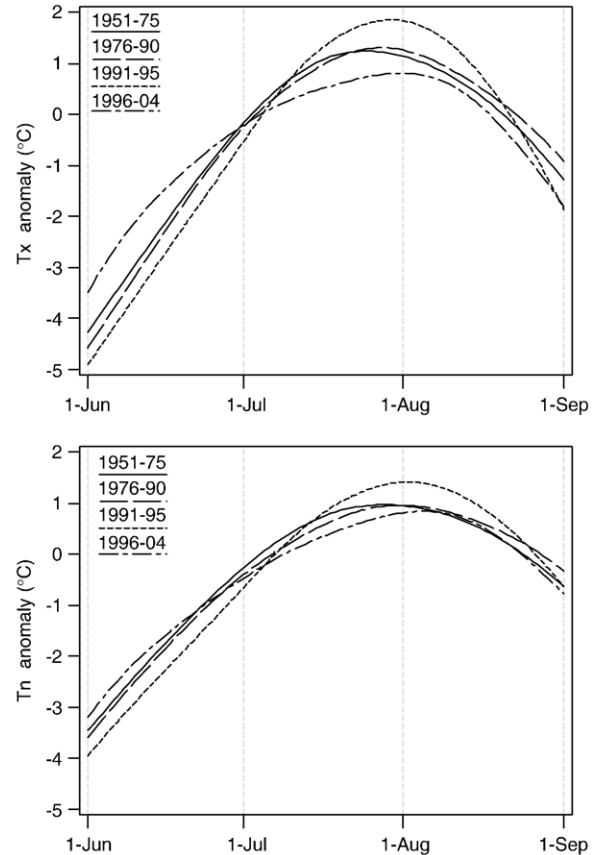


Fig. 5. Seasonal profiles of Tx (top) and Tn (bottom) anomaly series in 4 different periods, Zaragoza.

and 1976–90 are almost identical; (ii) the increase of the maximum in the first half of the 90s is a consequence of the warming observed in the midsummer by 1991, and (iii) in 1996–04, the curve reflects the temperature increase in May and June and the greater uniformity of the season in this period.

To complete this exploratory study, we show for Zaragoza and Huesca some statistics (mean value, standard deviation, variation coefficient in %, 90th and 99th percentiles) describing Tx and Tn distributions, see Table 1. Four separate periods, 1951–75, 1976–90, 1991–96 and 1997–2004 are considered and the statistics are calculated for the whole summer period (JJA) and per month. The following results, observed in both observatories, complement the graphical information:

- (i) A decrease of the global JJA variability is observed for the period 1997–2004.
- (ii) June data show the highest dispersion; the temperature increment in this month is observed in both the mean and extreme values.

Table 1

Statistics of summer (JJA) and monthly Tx and Tn distributions in four different periods, Zaragoza and Huesca

Zaragoza	June					July					August					JJA				
	Mean	Stdev	CV	p90	p99	Mean	Stdev	CV	p90	p99	Mean	Stdev	CV	p90	p99	Mean	Stdev	CV	p90	p99
1951–75	27.3	4.7	17	33.6	37.6	31.0	3.9	12	36.1	38.6	30.3	3.9	13	35.3	38.5	29.5	4.4	15	35.3	38.5
1976–90	27.7	4.5	16	33.8	36.6	31.4	4.3	14	36.4	40.3	30.6	3.9	13	35.4	38.7	29.9	4.5	15	35.6	38.7
1991–96	28.4	4.9	17	34.6	38.0	32.6	4.3	13	37.6	40.9	32.2	4.1	13	37.4	39.3	31.1	4.8	15	37.0	39.5
1997–04	30.4	5.0	16	36.4	40.2	31.6	4.1	13	37.1	39.0	32.3	3.7	11	37.4	39.3	31.5	4.3	14	36.9	39.3

Zaragoza	June					July					August					JJA				
	Mean	Stdev	CV	p90	p99	Mean	Stdev	CV	p90	p99	Mean	Stdev	CV	p90	p99	Mean	Stdev	CV	p90	p99
1951–75	14.1	2.7	19	17.4	20.2	16.9	2.2	13	19.6	21.4	16.7	2.3	14	19.6	21.5	15.9	2.7	17	19.2	21.3
1976–90	14.9	2.9	19	18.8	21.3	17.6	2.5	14	21.0	22.7	17.6	2.4	14	20.6	23.0	16.7	2.9	17	20.4	22.4
1991–96	15.2	2.9	19	19.3	21.6	18.3	2.6	14	21.5	23.0	18.7	2.4	13	21.6	23.2	17.4	3.1	18	21.2	23.0
1997–04	16.4	2.8	17	19.7	22.8	17.8	2.3	13	21.0	22.6	18.7	2.2	12	21.3	22.5	17.6	2.6	15	21.0	22.5

Huesca	June					July					August					JJA				
	Mean	Stdev	CV	p90	p99	Mean	Stdev	CV	p90	p99	Mean	Stdev	CV	p90	p99	Mean	Stdev	CV	p90	p99
1951–75	25.7	4.2	16	31.0	34.6	29.9	3.3	11	34.0	37.1	29.0	3.5	12	33.4	36.1	28.2	4.1	14	33.2	36.5
1976–90	26.8	4.5	17	32.8	37.3	31.0	3.9	13	35.6	39.2	30.1	3.8	13	34.9	38.5	29.3	4.5	15	34.6	38.2
1991–96	27.3	4.4	16	32.6	36.4	32.0	3.8	12	36.6	38.9	31.2	3.9	13	36.1	38.0	30.2	4.5	15	35.6	38.3
1997–04	29.2	4.9	17	35.1	37.9	30.5	3.7	12	35.0	37.5	31.3	3.4	11	36.4	38.0	30.4	4.1	13	35.4	37.8

Huesca	June					July					August					JJA				
	Mean	Stdev	CV	p90	p99	Mean	Stdev	CV	p90	p99	Mean	Stdev	CV	p90	p99	Mean	Stdev	CV	p90	p99
1951–75	12.8	3.3	26	16.8	20.0	15.6	3.0	19	19.4	22.0	15.6	3.0	19	19.4	21.4	14.7	3.4	23	18.8	21.4
1976–90	13.0	3.4	26	17.4	20.1	16.1	3.3	20	20.2	23.4	16.0	3.1	19	19.6	23.4	15.0	3.6	24	19.6	22.7
1991–96	13.1	3.7	28	17.6	21.0	16.7	3.4	21	20.6	23.2	17.0	3.1	18	20.8	22.6	15.6	3.8	25	20.2	22.6
1997–04	14.9	3.8	25	19.4	23.1	15.7	3.1	20	19.4	22.0	17.2	3.2	18	21.3	24.0	16.0	3.5	22	20.3	23.0

(iii) In August, the Tx variability decreases in the last period; the increase observed in Tx mean and moderate extreme values, p90, is greater than the corresponding to most extreme, p99.

(iv) The maximum values of the location statistics for July occur by the period 1991–96; afterwards, the mean, the extreme values and also the variability of Tx and Tn decrease.

## 2.2. Analysis of extreme hot events

### 2.2.1. Defining extreme hot events

After analyzing the temperature evolution, our aim is the study of extreme hot events. Using an “excess over threshold” approach, an EHE is defined as a spell, of arbitrary length, of consecutive days with their Tx values over an extreme threshold; we select as threshold the 95th percentile of the JJA Tx daily data from 1971–2000. This percentile is 37 °C in Zaragoza and 35.8 °C in Huesca, values which are very close to the heat wave alarm thresholds, 37.3 °C and 35.8 °C, respectively, established by the Spanish public health authorities from the studies

by Díaz et al. (2005). Due to the climate characteristics of the observatories, EHEs appear only in the warm season and, consequently, only observations from the months of June, July and August are considered. This period seems to be enough, since no EHE is observed in May during the analysed period, while 3 events are found in Zaragoza and 4 in Huesca, in September.

In order to fit a model based on a point process, an occurrence point must be associated to each event; we choose the day of the spell where the maximum Tx value is observed. To summarize the severity of each EHE we use three variables: the *length*,  $L$ , the number of days of the event, the *maximum intensity*,  $I_x$ , the maximum value of the Tx excess over the threshold during the spell, and the *mean intensity*,  $I_m$ , the ratio between the total excess, defined as the accumulated exceedances of Tx over the threshold, and the length of the event.

### 2.2.2. Analyzing EHEs

Applying the previous definition we found 146 EHEs in Zaragoza and 110 in Huesca. To describe their characteristics and evolution, some basic statistics are

Table 2

Summer (JJA) mean and maximum number of EHEs and mean, 50th and 90th percentiles of severity variable distributions, per decade, Zaragoza and Huesca

Zaragoza	JJA mean frequency	JJA max. frequency	L			Ix			Im		
			Mean	p50	p90	Mean	p50	p90	Mean	p50	p90
51–60	1.2	3	1.4	1.0	2.7	1.3	0.9	3.0	1.06	0.90	2.21
61–70	2.8	6	1.6	1.0	3.1	1.2	0.6	3.1	0.96	0.60	2.25
71–80	2.1	3	1.4	1.0	2.8	0.8	0.4	1.6	0.57	0.40	1.60
81–90	2.5	5	1.8	1.0	3.8	1.3	1.0	4.0	0.92	0.80	2.84
91–00	3.9	7	1.9	2.0	3.0	1.3	1.0	2.5	0.99	0.90	1.83
01–04	5.3	7	2.5	2.0	4.8	1.3	1.0	2.8	0.99	0.90	1.63

Huesca	JJA mean frequency	JJA max. frequency	L			Ix			Im		
			Mean	p50	p90	Mean	p50	p90	Mean	p50	p90
51–60	1.2	2	1.5	1.0	2.7	1.1	1.0	2.2	0.87	0.73	1.78
61–70	1.0	3	1.7	1.0	4.7	1.1	0.9	3.6	0.76	0.80	1.86
71–80	1.0	3	1.9	2.0	3.0	1.4	1.3	3.0	0.99	0.70	2.17
81–90	2.5	5	2.3	2.0	4.4	1.6	1.2	4.7	1.13	0.87	3.18
91–00	3.8	8	1.9	1.0	4.2	1.3	1.0	2.7	0.95	1.00	1.80
01–04	5.1	7	2.3	1.5	3.9	1.1	1.0	2.2	0.87	0.72	1.89

shown in Table 2. For each decade we show the mean and maximum number of EHEs observed in the summer and the mean, 50th and 90th percentiles of the three severity variables. From the analysis of this table, we observe that: (i) The EHE occurrence mean rate and the mean length of events are increasing whereas the mean values of the intensity measures remain stable. (ii) Regarding the extreme values, the maximum number of events and the 90th percentile of the length in Zaragoza show an increasing trend while the intensity measures show again a stable evolution.

In order to analyze the EHE seasonal profile, we calculate, per month and decade, the mean annual occurrence rate and the median of the severity variables, see Table 3. The occurrence rate shows a seasonal pattern during the period 1951–2000: July is the month most prone to have EHEs, followed by August and, finally, June. This pattern disappears in the last decade due to the already mentioned increase of the temperature in June. Concerning the severity variables, no clear seasonal pattern is observed.

As a conclusion of the results in this section, we can assert that:

- (i) The evolution of Tx distribution is not homogeneous inside the summer period. From 1996, an important increase of the temperature mean level is observed in June, while in July and August it remains stable.
- (ii) The occurrence rate of EHEs has a seasonal behaviour inside the considered period, JJA, as

could be expected, but, as a consequence of the Tx evolution, this pattern could be changing.

These facts indicate that the EHE occurrence can be only partially explained by the long-term evolution and the seasonal cycle of Tx and that it will be necessary to consider temperature signals able to reproduce the different temperature evolution inside the summer.

Table 3

Mean number of EHEs and 50th percentile of severity variable distributions, per month and decade, Zaragoza and Huesca

Zaragoza	Mean frequency			p50			L			Ix			Im		
	Jn	Jl	Ag	Jn	Jl	Ag	Jn	Jl	Ag	Jn	Jl	Ag	Jn	Jl	Ag
51–60	0.1	0.9	0.2	1.0	1.0	1.5	1.9	0.7	1.8	1.9	0.7	1.3			
61–70	0.5	1.2	1.1	1.0	1.0	1.0	1.6	0.8	0.4	1.4	0.6	0.4			
71–80	0.0	1.3	0.8	–	1.0	1.0	–	0.5	0.3	–	0.5	0.3			
81–90	0.2	1.5	0.8	1.5	2.0	1.0	0.7	1.2	0.6	0.5	0.8	0.6			
91–00	0.2	2.3	1.4	1.5	1.0	2.0	0.6	1.2	1.1	0.5	1.2	0.9			
01–04	2.0	2.0	1.3	1.5	2.0	2.0	1.8	0.8	0.9	1.4	0.5	0.7			

Huesca	Mean frequency			p50			L			Ix			Im		
	Jn	Jl	Ag	Jn	Jl	Ag	Jn	Jl	Ag	Jn	Jl	Ag	Jn	Jl	Ag
51–60	0.1	0.9	0.2	1.0	1.0	2.5	1.5	0.7	2.2	1.5	0.7	1.5			
61–70	0.3	0.3	0.4	2.0	1.0	1.0	1.8	1.0	0.5	0.9	1.0	0.5			
71–80	0.0	0.6	0.4	–	1.5	2.0	–	1.3	1.3	–	1.0	0.6			
81–90	0.4	1.5	0.6	2.0	2.0	2.0	1.8	0.8	1.0	1.6	0.8	0.9			
91–00	0.1	2.1	1.4	1.0	1.0	2.0	0.2	1.0	1.4	0.2	1.0	1.0			
01–04	2.0	1.6	1.5	1.0	2.0	1.5	1.3	0.7	1.0	0.9	0.7	0.7			

### 3. Modeling extreme hot events

In order to model the EHE behaviour, we consider first the occurrence process and then, the three severity variables: length, maximum and mean intensities.

#### 3.1. Occurrence model

For modeling the EHE occurrence we choose a non-homogeneous Poisson process, NHPP. There are some reasons that lead us to this model: the homogeneous Poisson process (HPP) has been widely used to model the occurrence of extreme events, since a well known Extreme Value theory result, Embrechts et al. (1997) Section 5.3, states that the occurrence of excesses over increasing thresholds converges to a Poisson process. Madsen and Rosbjerg (1997) and Abaurrea and Cebrián (2002) are examples of the use of HPP to model hydrological and meteorological droughts respectively. In this case, however, the process is clearly non-homogeneous since, in the previous exploratory analysis, an upward trend and a seasonal behaviour are observed in the occurrence rate. The frame of point processes, and particularly the non-homogeneous Poisson process, still enables a simple formulation of non-stationarity, see Smith (1984) for an application.

Thus, it seems reasonable to model the EHE occurrence by a NHPP, where points occur randomly

in time, at a variable rate which depends on influential covariates. Recall that a NHPP in  $R$ , the real number set, with rate  $\lambda(t)$ , is a point process with independent increments and such that the number of points (hot events) occurring in a set  $A$ , a certain period of time, follows a Poisson distribution with its mean equal to the integral of  $\lambda(t)$  on  $A$ .

We model the intensity  $\lambda(t)$ , where  $t$  denotes time measured on a daily scale, as a deterministic function and consider two different types of covariates:

1. Temperature terms. Two temperature related variables are considered:
  - A long-term signal, TTx, defined using a lowess smoother with a 30% window parameter, which represents the decadal evolution.
  - A short-term signal,  $Tx_{m31}$ , defined as a 31 days moving average centered in each day, which provide information on the local “state” of the temperature.
2. Seasonal terms. The seasonality of the process could be partly taken into account by the short-term temperature signal but we still allow the process to have specific seasonal terms; more precisely, we consider as covariates, the restriction to three summer months of first order harmonic functions describing the annual cycle,  $\cos(2\pi i)$  and  $\sin(2\pi i)$ , where  $i=152/365, \dots, 243/365$ , is an index of the day position inside the year.

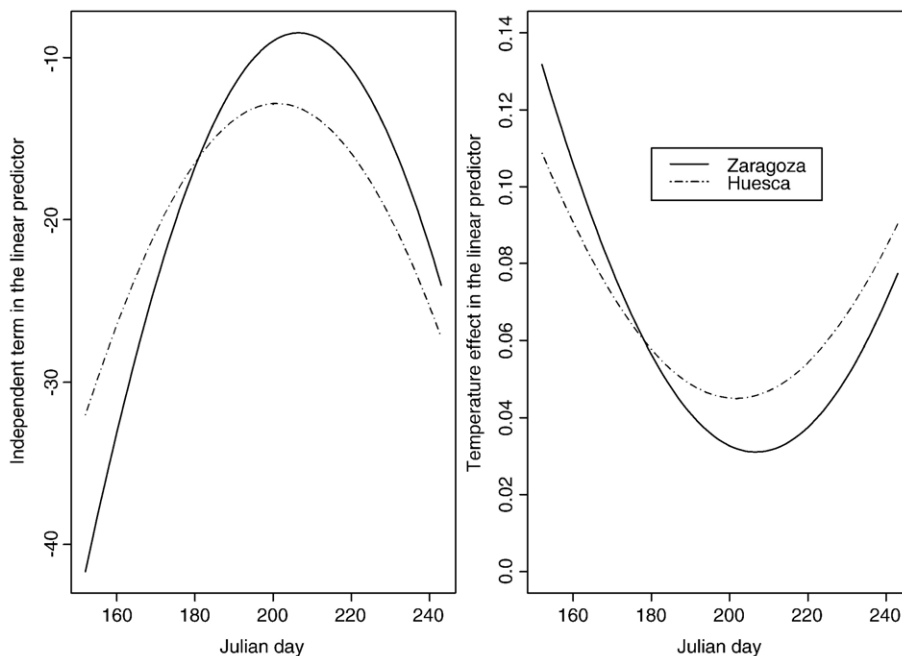


Fig. 6. Partial effect of the independent and temperature terms in the linear predictor of EHE occurrence model, Zaragoza and Huesca.

A quadratic temperature term and the interaction terms between the temperature and the seasonal components are also considered as possible covariates. Since the occurrence rate must be a positive value, we use a logarithmic function to link the rate with the linear predictor. In short, we allow the intensity function to have this structure:

$$\begin{aligned} \log(\lambda(t)) = & \beta_0 + \beta_1 \cos(2\pi i) + \beta_2 \sin(2\pi i) \\ & + f_1(\text{temperature variables}; \beta_j) \\ & + f_2(\cos(2\pi i) \text{ temperature variables}; \beta_k) \\ & + f_3(\sin(2\pi i) \text{ temperature variables}; \beta_l) \end{aligned}$$

### 3.1.1. Model estimation

The parameter estimation is performed by maximum likelihood and the selection of the variables to be included in the linear predictor is based on the likelihood ratio test, which allows us to compare nested models.

*Results.* The final model for Zaragoza is:

$$\begin{aligned} \log(\lambda(t)) = & -89.8 - 74.5 \cos(2\pi i) - 32.6 \sin(2\pi i) \\ & + 0.277 \text{Tx}_{m31} + 0.224 \text{Tx}_{m31} \cos(2\pi i) \\ & + 0.099 \text{Tx}_{m31} \sin(2\pi i) \end{aligned}$$

The likelihood ratio test  $p$ -values associated to the different terms are: 0.001 for  $\text{Tx}_{m31}$ , 0.006 for the seasonal effect given by the harmonic terms,  $\cos(2\pi i)$  and  $\sin(2\pi i)$ , and 0.009 for the interaction effect between  $\text{Tx}_{m31}$  and the harmonic terms.

From this model it can be concluded that the EHE occurrence shows a seasonal behaviour which varies with the temperature signal. The effect of harmonic terms is to match up the highest value of the independent term with the end of July, the hottest period of the summer, see left side of Fig. 6. The interaction terms

reflect that seasonality affects the temperature effect; the  $\text{Tx}_{m31}$  influence, always positive, changes during the summer period: for an identical  $\text{Tx}_{m31}$  input, the slope is lower in the mid summer than at the beginning, see right side of Fig. 6. For example, if  $\text{Tx}_{m31}$  increased 1 °C, the occurrence rate for July 15th would increase by a factor 1.43, while for June 15th this factor would be 2.45.

Although  $\text{TTx}$  is an influential covariate, it is not significant when the short-term temperature is included in the linear predictor and, consequently, it is removed from the model; the  $\text{Tx}_{m31}$  term is enough to represent both the local effect of the temperature and the influence of its long-term evolution.

In Fig. 7 we can observe the general evolution of the fitted rate throughout the observed period, showing its increasing trend.

### 3.1.2. Checking the model

*3.1.2.1. NHPP checking.* To validate the model, we followed a two step procedure:

1. A NHPP can be made homogeneous of rate 1 by transforming the time scale (Cox and Isham, 1980). If we transform the occurrence points of the original process,  $t_i$ , in this way,

$$t_i^* = E[(0, t_i)] = \int_{(0, t_i)} \hat{\lambda}(t) dt$$

the new point process will be a homogeneous Poisson process.

2. The Poisson character of the transformed process can be checked by controlling the exponential distribution of inter-event distances,  $d_i$ , or, alternatively, the

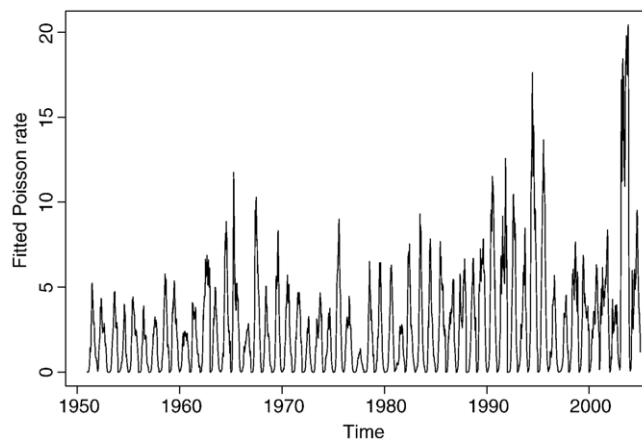


Fig. 7. Fitted NHPP occurrence rate of the EHE process, Zaragoza.



standard uniform behaviour of the transformed distances,  $\exp(-d_i)$ , using, for example, a Kolmogorov–Smirnov test and a qq-plot with a confidence band.

**3.1.2.2. Residuals.** To validate the fitted linear predictor, some type of residuals has to be used. In this case, since there are no observed occurrence rates, the usual raw residuals (observed minus fitted values) cannot be determined; so we decided to use alternative residuals defined as the difference between empirical and fitted rates. The empirical rates are based on the observed number of events in a period of time of certain length,  $l$ , centered on the corresponding day  $t$ , while the fitted ones are based on the sum of the fitted daily rates,  $\lambda(t)$ , over the same period. After trying different lengths, 1, 3 and 6 months, we chose the value corresponding to the summer length,  $l=3$ .

**Results.** The validation process was found to be satisfactory for the model fitted to Zaragoza data. Concerning the NHPP checking, the uniform qq-plot, not shown, shows a linear behaviour inside the confidence band and the p-value of the corresponding Kolmogorov–Smirnov test is 0.38.

In Fig. 8, the residuals are plotted versus four relevant variables: time, TTx, Julian day and  $\text{Tx}_{m31}$ . It

seems that the short-term temperature signal is enough to represent the time evolution, since no systematic pattern is observed in the plot of residuals versus time. The seasonality also seems to be properly modeled, since a random cloud of points is observed in the plot of residuals versus the Julian day.

### 3.2. Modeling severity

To complete the description of the EHE process we model the severity variables:  $L$ ,  $I_x$ , and  $I_m$ . We describe these variables by adequate probability distributions whose parameters are allowed to depend on influential covariates. The same covariates used for modeling the occurrence process are considered.

#### 3.2.1. Length

Taking into account that length can only take integer values equal or greater than 1 and usually not very high (after eliminating a highly anomalous observation, 14 day long, in 2003, the maximum of the sample is 6 days), a shifted Poisson distribution was firstly considered. Since the resulting model showed overdispersion, we chose a shifted negative binomial distribution. We fitted by maximum likelihood a model with constant size parameter,  $N$ , and the mean

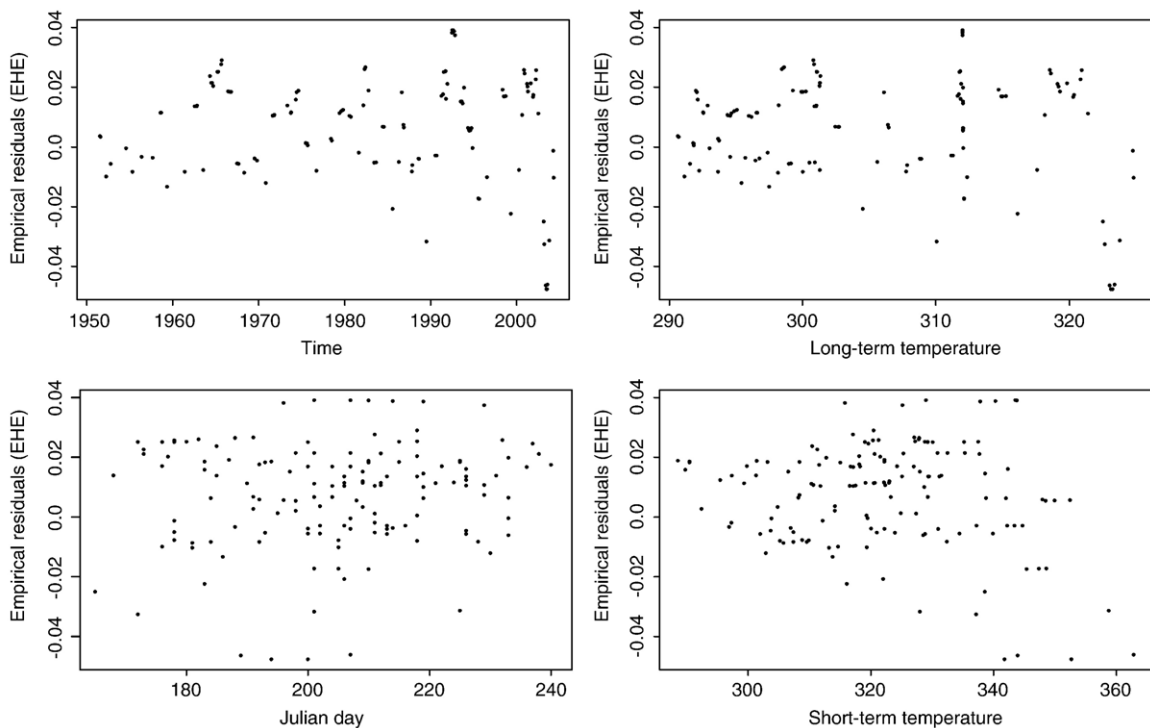


Fig. 8. Residual versus potential covariate plots for the EHE occurrence model, Zaragoza.

Table 4  
Fitted models to severity variables, Zaragoza

Variable	Distribution	Parameters
Length	Negative binomial	$\mu(t) = \exp(-8.5 + 0.025 T_{x_{m31}})$ $N=2$
Mean intensity	Generalized Pareto	Shape $\xi = -0.280$ Scale $\sigma = 12.357$
Maximum intensity	Generalized Pareto	Shape $\xi = -0.127$ Scale $\sigma = 13.981$

parameter,  $\mu$ , was allowed to depend on the considered covariates; we also included a factor to distinguish the different months and the corresponding interaction terms. To assure a positive mean, a logarithmic link function was used.

**Results.** In Table 4 we summarize the model resulting from the selection process. It can be observed that the mean parameter only depends on the short-term temperature and that no additional seasonal term is needed.

Given the discrete character of the length, the usual goodness of fit graphics cannot be applied; alternatively, we use plots, see Fig. 9, where we compare the observed

frequencies and the fitted probabilities of EHEs observed in similar temperature conditions. More precisely, we divide the observed events into 6 groups, according to their  $T_{x_{m31}}$  values, using as cutpoints the 17th, 33rd, 50th, 67th and 83rd percentiles. For each group, whose size is about 25 observations, we calculate the fitted probabilities using a negative binomial distribution, whose  $\mu$  parameter has input the  $T_{x_{m31}}$  mean value of the group. Even if the probability distributions used in these plots are not exactly the fitted ones to each observation, the model seems to reproduce properly the observed behaviour.

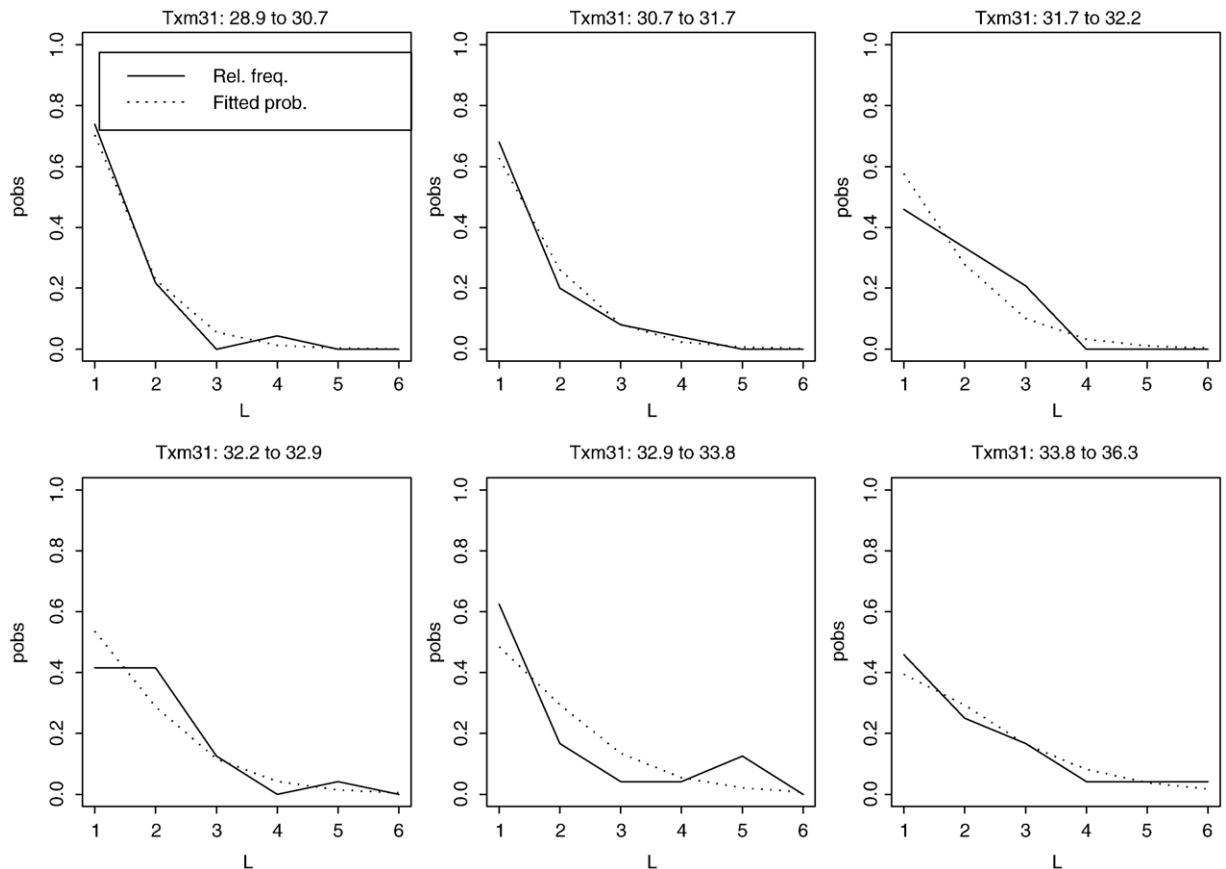


Fig. 9. Relative frequency and negative binomial probability distribution for EHE lengths, grouped by similar  $T_{x_{m31}}$  value, Zaragoza.

### 3.2.2. Maximum and mean intensity

To fit the maximum intensity we tried different continuous positive distributions such as gamma, Gumbel, Weibull, exponential, lognormal, loglogistic and generalized Pareto. The last one provided the best fitting in the exploratory analysis and is also justified by an Extreme Value theory asymptotic result, see [Embrechts et al. \(1997\)](#), Section 3.4. We allow the scale and shape distribution parameters,  $\sigma$  and  $\xi$ , to vary monthly and the scale parameter, also, to depend on temperature signals; we fit the model by the maximum likelihood method. Applying the same exploratory analysis, the generalized Pareto distribution was also selected for modeling Im.

*Results.* The resulting model for Ix is shown in [Table 4](#) and its both parameters are constant, not depending on any covariate. The parameters of the selected model for Im are also constant, see [Table 4](#).

The generalized Pareto qq-plots corresponding to both samples, not shown, are linear; the Kolmogorov–Smirnov test  $p$ -values are 0.21 and 0.15, respectively.

To summarize the results concerning the severity variables, we can assert that only the EHE length is being affected by temperature evolution, through the short-term temperature covariate, while intensity variables remain stable during the considered period.

### 3.3. Final modeling remarks

As in the exploratory analysis, in order to confirm the validity of our results, the Huesca series is also modeled. The similarity of the independent term and the global slope of  $T_{x_{m31}}$  term in the occurrence model is shown in [Fig. 6](#). The main difference between both models appears in the seasonal behaviour, which is less significant than in Zaragoza: the  $p$ -value associated with seasonal terms is 0.16. Concerning the length, a similar model to the one fitted for Zaragoza,  $\log(\mu) = -8.33 + 0.026 T_{x_{m31}}$ , is obtained.

Thus, the fitted model allows us to conclude that the mean temperature increase observed from 1975 onwards yields an important increase in the frequency and length of EHEs but not in their intensity, which shows a stationary behaviour. This fact agrees with the hypothesis of a variability reduction of the temperature values within the year, stated by the IPCC TAR ([Houghton et al., 2001](#)): “..., there is now little evidence to suggest that the interannual variability of global temperatures has increased over the past few decades, but there is some evidence to suggest that the variability of intra-annual temperatures has actually quite widely decreased. Several analyses find a decrease in spatial and temporal variability of temperatures on these shorter time-scales.”

## 4. Medium-term projection in a climate change scenario

The final part of this study shows the use of the fitted model to obtain projections for the EHE evolution, in a climate change scenario, in the central Ebro valley. The hypothesis of an increment in the number of EHEs associated to global warming process has led to an increasing interest in this kind of events; thus, several recent studies have tried to quantify, at a regional scale, this evolution. [Beniston and Diaz \(2004\)](#), using a HIRHAM4-RCM output for the scenario A2 at the grid point closest to Basel (Switzerland), suggest that the 2003 heat wave may be considered as a “shape of things to come” in the latter part of the 21st century. They find that, in the forced climate, the number of days where Tx would exceed the 30 °C threshold, the 90th percentile of Tx in Basel for the period 1961–1990, would increase from the current 8 days, to more than 40 days in 2071–2100.

[Sánchez et al. \(2004\)](#) use their model PROMES-RCM to obtain two simulated trajectories over the Mediterranean basin, one for the period 2071–2100 in an A2 scenario and another corresponding to the current climate (1961–90). For the centre of the Iberian Peninsula they project an increase of summer Tx mean value greater than 6 °C and a slightly lower one for the 90th percentile. According to this scenario, the JJA Tx mean would reach in Zaragoza 36 °C at the end of the 21st century; assuming for the future observations a Gaussian distribution with a standard deviation equal to the observed one in 1961–90, the 95th percentile of the 1971–2000 data, 37 °C, would become the 59th in the 2071–2100 distribution.

To obtain possible EHE trajectories we feed the fitted model with the covariate values obtained from the temperature series generated by a general circulation model (GCM), in the climate change scenario. Two preliminary steps are required: (a) the comparison of the observed temperatures with the series generated by the GCM, in order to estimate the possible bias with regard to Zaragoza data; and (b) to check that the “scaled” GCM series and the observed data follow a similar distribution. Only the EHE occurrence and length are projected since they are the only properties that we find to be influenced by the temperature evolution.

### 4.1. GCM data

Among the GCMs available, we selected the HadCM3 developed by the Hadley Centre in 1998, [Gordon et al. \(2000\)](#), [Pope et al. \(2000\)](#). From this model, the monthly mean series of Tx values were obtained but not the

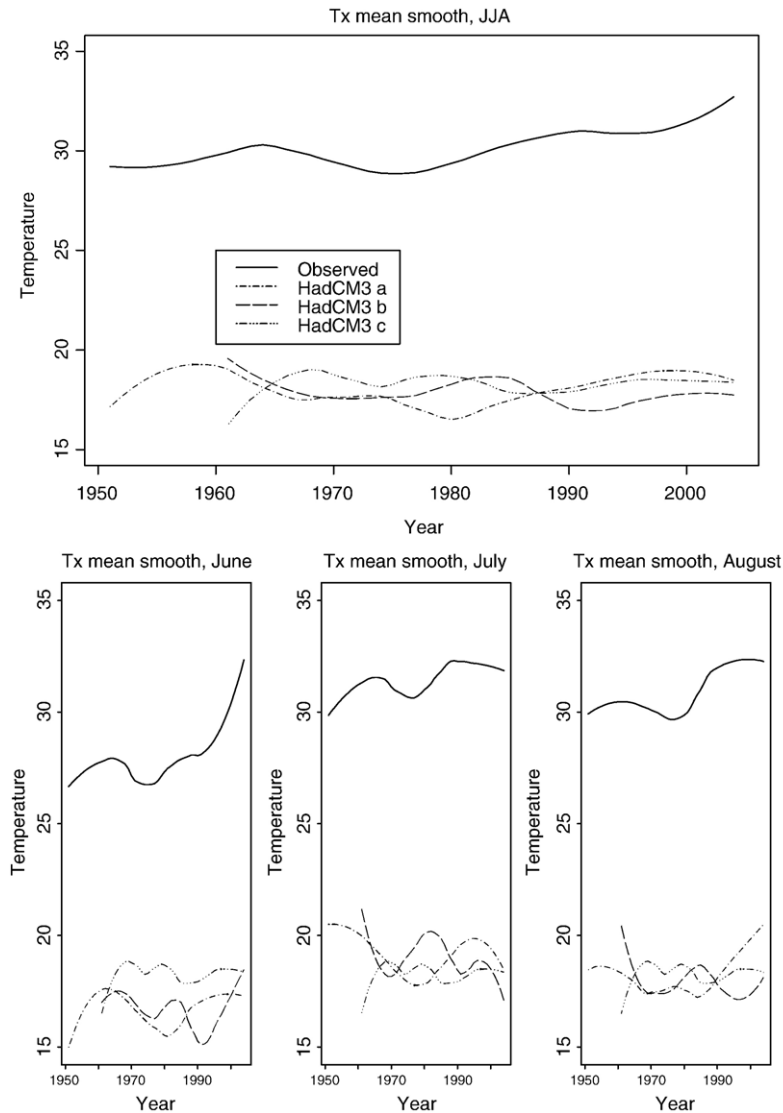


Fig. 10. Smoothed curves of the observed and (HadCM3) simulated monthly Tx mean series, 1951–2004, for summer, JJA, (top) and per month (bottom), Zaragoza.

corresponding daily trajectories, which were not available. The HadCM3 has a spatial resolution of  $3.75^\circ$  longitude by  $2.5^\circ$  latitude and we get the data corresponding to the grid point closest to Zaragoza ( $41.6^\circ\text{N}$ ,  $1^\circ\text{W}$ ), whose coordinates are ( $42.5^\circ\text{N}$ ,  $0^\circ\text{W}$ ), see Fig. 1, for the period 1951–2050. Among the several scenarios available, we select the SRES-A2, which represents a world with continuously increasing population and regionally oriented economic growth. The three HadCM3 trajectories have been generated under observed conditions until 1990 and under the A2 scenario from 1991 onwards. Data from this and other GCMs can be downloaded from the IPCC data distribution centre webpage: <http://ipcc-ddc.cru.uea.ac.uk/>.

The GCM generated data for the grid point average the conditions over the surrounding area, in this case, a surface of almost  $90,000 \text{ km}^2$  including heights varying from 100 m in the Ebro valley, to more than 2000 m in the Pyrenees. Consequently, even though the temperature evolution could be homogeneous in this area, the mean levels of separate points can be very different. In Fig. 10 we show the smoothed signals of summer Tx mean series observed in Zaragoza and of three analogous series obtained from HadCM3 at the aforementioned grid point, for the same period. As expected, the simulated series evolves at a lower level, since Zaragoza height is only 200 m. The different level is also apparent when we compare the evolution by month, see Fig. 10, bottom.

Table 5

Mean, standard deviation, 25th and 75th percentiles of the observed and three HadCM3 simulated Tx monthly mean series, 1961–1990, Zaragoza

	June				July				August			
	Mean	Stdev	p25	p75	Mean	Stdev	p25	p75	Mean	Stdev	p25	p75
Observed	27.6	1.67	26.4	28.7	31.4	1.54	30.5	32.6	30.4	1.51	29.4	31.4
HadCM3a	16.5	1.97	15.4	17.9	18.8	2.19	17.2	20.2	17.7	2.48	16.3	18.4
HadCM3b	16.8	2.00	15.2	18.3	19.3	2.07	17.9	20.0	18.2	1.96	17.1	19.5
HadCM3c	17.4	2.09	15.9	18.8	19.3	1.96	17.7	20.5	18.0	2.32	16.7	19.4

In Table 5 some location and variability statistics are calculated for observed and simulated (1961–1990) Tx trajectories. Concerning variability, the results of the  $F$

test for checking the variance equality between the observed and each one of three GCM generated series, depend on the month: while in June the lowest of the

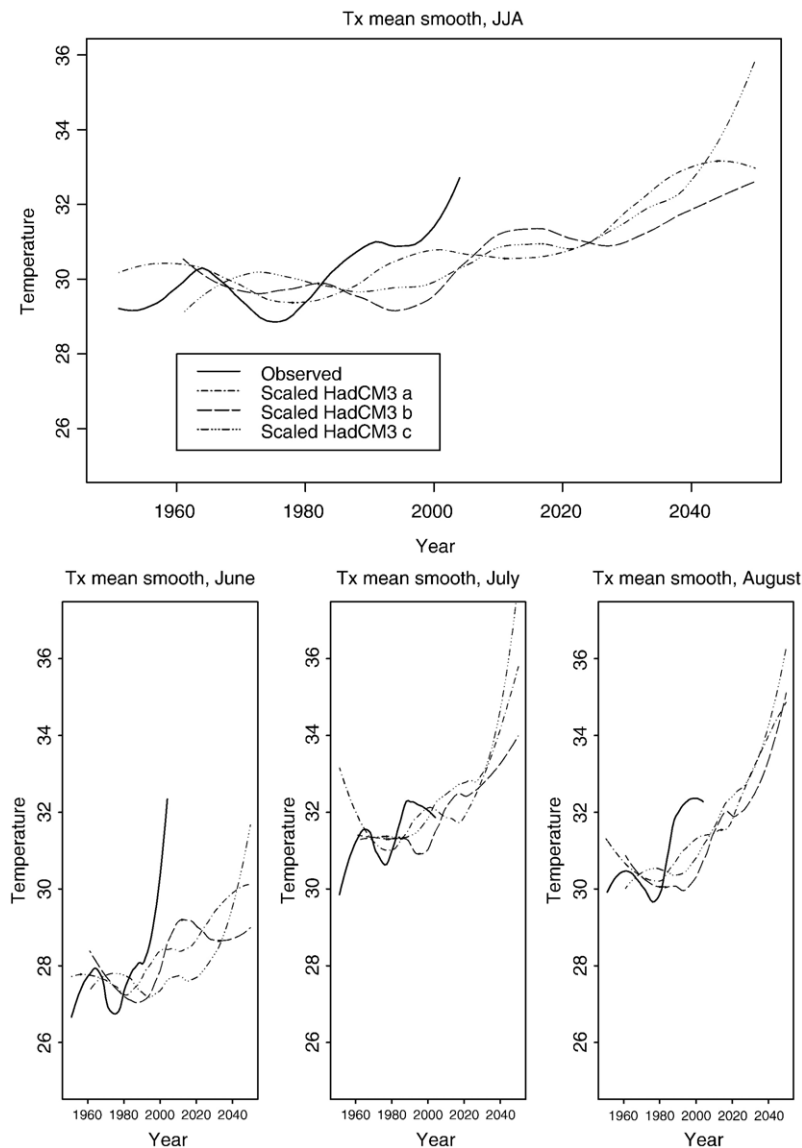


Fig. 11. Smoothed curves of the observed and scaled (HadCM3) simulated monthly Tx mean series, 1951–2050, for summer, JJA, (top) and per month (bottom), Zaragoza.

Table 6

EHE mean observed frequencies and fitted occurrence rates, using the model and HadCM3 temperature output, for summer and per month, Zaragoza

	Mean observed frequency				Fitted occurrence rates											
					A2a				A2b				A2c			
	Jn	Jl	Ag	JJA	Jn	Jl	Ag	JJA	Jn	Jl	Ag	JJA	Jn	Jl	Ag	JJA
1951–80	0.2	1.1	0.7	2.0	0.1	1.4	0.9	2.5								
1961–90	0.2	1.3	0.9	2.5	0.1	1.2	0.7	2.0	0.1	1.2	0.6	2.0	0.1	1.2	0.7	2.0
1971–00	0.1	1.7	1.0	2.8	0.1	1.3	0.8	2.2	0.1	1.1	0.5	1.8	0.1	1.2	0.6	1.9
1981–10					0.2	1.4	0.9	2.4	0.2	1.2	0.7	2.1	0.1	1.3	0.7	2.1
1991–20					0.2	1.5	1.1	2.7	0.4	2.6	1.8	4.7	0.3	2.9	1.9	5.0
2001–30					0.2	1.5	1.3	2.9	0.4	2.9	2.4	5.7	0.3	3.5	2.3	6.1
2011–40					0.5	3.0	2.9	6.4	0.4	3.7	4.2	8.3	0.4	4.0	3.0	7.5
2021–50					0.9	4.9	3.8	9.6	0.4	3.4	13.6	17.3	1.1	5.4	4.2	10.6

three  $p$ -values is 0.236, much lower values, 0.062 and 0.009, are obtained in July and August. Given these results, we decided to scale the HadCM3 series in order to match, by month, the Zaragoza mean level and variability observed in the 1961–1990 interval.

The Kolmogorov–Smirnov test which compares the distribution of monthly Tx mean values in the observed and the (1961–1990) scaled simulated series, generates  $p$ -values over 0.83 for the three trajectories, which support the use of them as input to the model. In Fig. 11 we show the evolution of scaled simulated series and the observed one. As we can see, the range of values in the four trajectories until 1990 is similar; afterwards, the observed one, influenced by the temperature increase from the 90s, exceeds the simulated under the A2 scenario.

#### 4.2. Projection of the EHE process

To explore the EHE behaviour in the climate change scenario we use three HadCM3 SRES A2 Tx trajectories, up to 2050. The temperature input,  $T_{x_{m31}}$ , is a moving average of daily Tx series; as HadCM3 daily data were not available,  $T_{x_{m31}}$  is built from the available series by assigning to each day the corresponding monthly mean.

##### 4.2.1. Analyzing the EHE simulated trajectories in 1951–90

In Table 6 we compare the fitted and observed mean annual EHE rates, per month and for the summer period, using 30 yr long moving periods which are updated every 10 yr. There is a good agreement in the 1951–80 and 1961–90 values and in the seasonal profile observed inside the summer. We noticed that the simulated temperatures under A2 scenario were lower than the ones observed during the 90s in the midsummer and, consequently, the EHE rates obtained from these

trajectories are also lower than the observed ones for July and August, 1971–2000.

The observed and generated EHE lengths are compared in Table 7. As the daily GCM trajectories are not available, it is not possible to identify the EHEs occurred in them. Thus, we estimate the fitted lengths of EHEs occurring in a certain month as the value obtained from the length model using as input the largest  $T_{x_{m31}}$  in this month. So, each value in Table 7 is a weighted average of these lengths, with weights equal to the frequencies fitted by the occurrence model, rounded to the next integer value. If the product of the fitted length by the fitted frequency overcomes the month length, we consider that only one EHE, of length the whole month, occurs. The results for periods 1961–90 and 1971–2000 show a good agreement with observed data.

##### 4.2.2. Projection of the EHE process up to 2050

The smoothed signals corresponding to observed Tx means (JJA, 1951–2004) and to three HadCM3 (JJA, 1951–2050) scaled series are plotted in Fig. 11. The projected Tx evolution shows different cycles with increasing maxima; the increase of the seasonal signal between 1980 and 2040 would be over 2 °C, and, regarding the monthly signals, we emphasize the rise, of nearly 4 °C, projected in August. The results for 2050, the end of the period analysed, are more extreme but they can be influenced by the edge effect of the smoother.

Table 6 shows monthly and seasonal occurrence mean rates calculated for 30 yr long periods. Globally, the number of EHEs would increase, at least, by a factor 3 by 2040 and by a factor greater than 4.5 by 2050. Regarding the seasonal shape, the trend towards a more uniform summer, observed in the exploratory analysis of Tx and Tn during 1951–2003, would go on in the future.

The projection obtained for the EHE lengths in two trajectories shows that they would be doubled in 50 yr

Table 7  
EHE mean observed and fitted lengths, 1951–2050, Zaragoza

	1951–1980	1961–1990	1971–2000	1981–2010	1991–2020	2001–2030	2011–2040	2021–2050
Observed	1.5	1.6	1.8					
HadCM3 A2a	2.1	1.7	1.9	1.9	2.0	2.2	3.6	3.9
HadCM3 A2b		1.7	1.7	1.9	2.4	2.6	2.7	3.4
HadCM3 A2c		1.9	1.6	1.8	2.7	3.3	4.1	5.4

and that they would increase by 230% according to the third one. We notice that combining both projections, for the month of July in 2021–50, the average number of days in the condition identified as EHE, would range from 11.6 days in HadCM3b to 29.2 days in HadCM3c.

## 5. Conclusions

The results of this work can be summarized in the following conclusions:

1. Concerning the temperature evolution in the central Ebro valley, we remark that the observed warming from 1975 is not homogeneous throughout the summer period, particularly from the 90s onwards. Summer is now more uniformly hotter than 30 yr ago and the period prone to EHE occurrence is longer.
2. A statistical model explaining the extreme hot event process is developed. It allows us to explain the occurrence and severity of extreme events using daily-time dependent covariates. Our approach is more flexible than the one proposed by Prieto et al. (2004) for the occurrence of extreme cold days, which does not provide rates in a daily scale. Usually, extreme event analysis is performed seasonally in order to get homogeneous samples, Möberg and Jones (2005) and Hundecha and Bardossy (2005). However, our model proves that there are different evolutions inside the summer season and that it is able to deal with the changing Tx evolution. Regarding the conclusions about the EHE behaviour in the central Ebro valley, obtained from the fitted models for Zaragoza and Huesca, we remark the non-stationarity of their generating processes. The occurrence rates are highly dependent on temperature and its evolution is determined by a short-term signal, which reflects also the long-term trend. The severity characteristics, however, show a mixed pattern: the length evolves influenced by the temperature change while the intensity measures remain stable.
3. With regard to the projected EHE evolution under scenario SRES A2, we point out that the occurrence

mean rate would increase, from 2000 to 2050, by a factor greater than 4 and that the mean length would increase by 103%, according to two trajectories, and by 230% according to the third one.

## Acknowledgements

The authors thank the Spanish National Meteorological Institute for providing the temperature data and the Hadley Centre for providing the HadCM3 output data. This work has been sponsored by the Spanish Ministry of Education and Science with the grant REN2002–00009/CLI.

## References

- Aburrea, J., Asin, J., Erdozain, O., Fernández, E., 2001. Climate variability analysis of temperature series in the medium Ebro river basin. In: Brunet, M., López, D. (Eds.), *Detecting and Modelling Regional Climate Change*. Springer-Verlag, pp. 109–118.
- Aburrea, J., Cebrián, A.C., 2002. Drought analysis based on a cluster Poisson model: distribution of the most severe drought. *Clim. Res.* 22, 227–235.
- Beniston, M., Diaz, H.F., 2004. The 2003 heat wave as an example of summers in a greenhouse climate? Observations and climate model simulations for Basel, Switzerland. *Glob. Planet. Change* 44, 73–81.
- Beniston, M., Stephenson, D.B., 2004. Extreme climatic events and their evolution under changing climatic conditions. *Glob. Planet. Change* 44, 1–9.
- Brunet, M., Aguilar, E., Saladié, O., Sigró, J., López, D., 2001. The variations and trends of the surface air temperature in the Northeastern Spain from middle nineteenth century onwards. In: Brunet, M., López, D. (Eds.), *Detecting and Modelling Regional Climate Change*. Springer-Verlag, pp. 81–94.
- Cox, D.R., Isham, V., 1980. *Point Processes*. Chapman and Hall. 188 pp.
- Embrechts, P., Kluppelberg, C., Mikosch, T., 1997. *Modelling Extremal Events*. Springer-Verlag. 645 pp.
- Diaz, J., Linares, C., García, R., 2005. Impacto de las temperaturas extremas en la salud pública: futuras actuaciones. *Rev. Esp. Salud Pública* 79, 145–157.
- Gordon, C., Cooper, C., Senior, C.A., Banks, H., Gregory, J.M., Johns, T.C., Mitchell, J.F.B., Wood, R.A., 2000. The simulation of SST, sea ice extents and ocean heat transports in a version of the Hadley Centre coupled model without flux adjustments. *Clim. Dyn.* 16, 147–168.
- Houghton, J.T., Ding, Y., Noguer, M., 2001. *Climate Change 2001: The Scientific Basis*. Cambridge University Press. 881 pp.
- Hundecha, Y., Bardossy, A., 2005. Trends in daily precipitation and temperature extremes across western Germany in the second half of the 20th century. *Int. J. Climatol.* 25, 1189–1202.

- Klein Tank, A.M., Winjgaard, J., Van Engelen, A., 2002. Climate of Europe. Assessment of Observed Daily Temperature and Precipitation Extremes. KNMI, De Bilt, the Netherlands. 36 pp.
- Klein Tank, A.M., Können, G.P., 2003. Trends in indices of daily temperature and precipitation extremes in Europe 1946–99. *J. Climate* 16, 3665–3672.
- Klein Tank, A.M., Können, G.P., Selten, F.M., 2005. Signals of anthropogenic influence on European warming as seen in the trend patterns of daily temperature variance. *Int. J. Climatol.* 25, 1–16.
- Luterbacher, J., Dietrich, D., Xoplaki, E., Grosjean, M., Wanner, H., 2004. European seasonal and annual temperature variability, trends, and extremes since 1500. *Science* 303, 1499–1503.
- Madsen, H., Rosbjerg, D., 1997. The partial duration series method in regional index-flood modeling. *Water Res. Res.* 33, 737–746.
- Möberg, A., Jones, P.D., 2005. Trends in indices for extremes in daily temperature and precipitation in central and western Europe, 1901–99. *Int. J. Climatol.* 25, 1149–1171.
- Pope, V.D., Gallani, M.L., Rowntree, P.R., Stratton, R.A., 2000. The impact of new physical parametrizations in the Hadley Centre climate model — HadAM3. *Clim. Dyn.* 16, 123–146.
- Prieto, L., Garcia-Herrera, R., Díaz, J., Hernández, E., del Teso, T., 2004. Minimum extreme temperatures over Peninsular Spain. *Glob. Planet. Change* 44, 59–71.
- Sánchez, E., Gallardo, C., Gaertner, M.A., Arribas, A., Castro, M., 2004. Future climate extreme events in the Mediterranean simulated by a regional climate model: a first approach. *Glo. Planet. Change* 44, 163–180.
- Schär, C., Vidale, P.L., Lüthi, D., Frei, C., Häberli, C., Liniger, M.A., Appenzeller, C., 2004. The role of increasing temperature variability in European summer heatwaves. *Nature* 427, 332–336.
- Smith, R.L., 1984. Threshold methods for sample extremes. In: Tiago de Oliveira, J. (Ed.), *Statistical Extremes and Applications*. NATO Advanced Study Institute. Reidel, pp. 621–638.