EXTREME TEMPERATURE EVENTS IN CENTRAL EUROPE: ARE CLIMATE MODELS ABLE TO REPRODUCE THEM?

Jan Kyselý, Radan Huth
Institute of Atmospheric Physics AS CR, Boční II/1401, 141 31 Prague, Czech Republic:
honza@ufa.cas.cz

Abstract: Extreme weather and climate events (including extreme temperatures) severely influence ecosystems and human society. Impacts of climate change are likely to result rather from changes in climate variability and extremes than from an increase in the mean temperature, which underlines the need for validation of extreme events in present climate simulations. This study compares 20- and 50-year return values of annual maximum and minimum temperatures and heat/cold waves (periods of high/low temperature) in (i) observation, (ii) GCM simulated control climates, (iii) statistical downscaling from observation, and (iv) statistical downscaling from GCMs, in central Europe.
The comparison of the 20- and 50-year return values does not appear to be sensitive to what statistical method of the estimation of parameters of extreme value distribution is used. The skill of GCMs in reproducing extreme high and low temperatures is limited; the statistical downscaling from GCMs tends to improve the results although it generally yields extremes that are too moderate. On the other hand, heat and cold waves are better reproduced by GCMs than by statistical downscaling. Different results hold also when various methods of treating the missing variance in downscaling are compared: while the 20- and 50-yr return values are better captured by downscaling with noise additions, heat and cold waves are much better reproduced if variance inflation is performed.

1. Introduction

Extreme weather and climate phenomena are a subject of investigation because of both their current impacts on ecosystems and society and the threat of their possible increases in frequency, duration and severity in the climate perturbed by enhanced concentrations of greenhouse gases in the atmosphere. Impacts of climate change are expected to result rather from changes in climate variability and extreme event occurrence than from an increase in mean temperature (*Watson et al. 1996, Parmesan et al. 2000*) and even relatively small shifts in the means and variances of climate variables can induce considerable changes in the severity of extreme events (*Katz and Brown 1992, Colombo et al. 1999*).

General circulation models (GCMs) are currently the most frequently used tool in climate modelling. They are able to reproduce many features of the observed climate system not only in terms of means but also naturally occurring variability; however, they were not designed for simulating local climates and their reliability decreases with increasing spatial and temporal resolution required. There are several ways of obtaining site-specific daily time series, which are to a different extent based on GCM outputs, one of them being the statistical downscaling. Statistical downscaling takes advantage of the fact that GCMs simulate large-scale upper-air fields more accurately than the surface local variables (*Huth 1999*). It consists in identifying in the observed data the relationships between the upper-air variables and the local surface ones, and applying them to control and/or perturbed GCM runs. The downscaled time series are fitted to a specific site and, if applied to the present climate, can be adjusted to reproduce the original mean and variance.

Growing attention to extreme events has recently been paid in GCM studies, both in validating the simulated present-day climate and analyzing the possible future climate (*Meehl et al. 2000*). As regards extremes of surface temperature, relatively little work has been done in GCM and downscaling studies. Recently, *Zwiers and Kharin (1998)* and *Kharin and Zwiers (2000)* analysed return periods of annual maxima and minima of surface air temperature in two GCMs of the Canadian Centre for Climate Modelling and Analysis; *Trigo and Palutikof (1999)* examined heat and cold waves in the HadCM2 GCM and in the downscaling from the same GCM; *McGuffie et al. (1999)* compared return periods for extreme temperatures in five GCMs with a coarse spatial resolution; and *Huth et al. (2001a)* dealt with heat and cold waves in two GCMs and in downscaled and stochastically generated temperature series. Generally, GCMs reproduce extreme temperature events better than extreme precipitation events (where sub-grid scale processes play a more important role in producing observed extremes) but their skill remains limited.

Here we focus on the comparison of extreme high/low one-day maximum/minimum air temperatures (examining the distribution and 20- and 50-yr return values of annual maximum/minimum temperature) and periods of extreme high/low temperatures lasting at least 3 days, but typically about a week or longer (heat/cold waves). The analysis was performed at 4 sites in central Europe and the comparison involved the observed, GCM-simulated (3 models) and downscaled (4 models) temperature series.

2. Data

a. GCMs

Simulations of present climate of three GCMs were used in this study.

ECHAM3. The ECHAM GCM originates from the European Centre for Medium Range Weather Forecast model, modified in Max-Planck-Institute for Meteorology in Hamburg. Its version 3 is described in *DKRZ (1993)*. It has a resolution corresponding to a 2.8° gridstep both in longitude and latitude; years 11 to 40 of the control run, in which climatological SSTs and sea ice extent were employed, are examined.

AGCM2 (denoted CCCM2 in *Huth et al. 2001a*, *Kyselý 2002a* and *Kyselý et al. 2001*). The Canadian Centre for Climate Modelling and Analysis model of the second generation is described in *McFarlane et al. (1992)*. Its resolution corresponds to a 3.75° x 3.75° grid and the atmospheric model was run with a mixed layer ocean model and a thermodynamic sea ice model; 20 years of its control integration were available.

CGCM1. The first version of the Canadian Global Coupled Model is described in *Flato et al. (2000)*. The atmospheric component of the model is essentially AGCM2 which was coupled to the ocean dynamic model. Daily data are available for one of three transient climate change simulations for the period 1900-2100, which uses an effective greenhouse gas forcing change corresponding to that observed from 1850 to the present, and a forcing change corresponding to an increase in CO_2 at a rate of 1% per year thereafter until 2100. The period 1961-1990 was used as a control one.

The first two GCMs discussed here were used in several previous validation studies dealing with central European temperature series (*Nemešová and Kalvová 1997; Kalvová and Nemešová 1998; Kyselý and Kalvová 1998; Huth et al. 2000; Huth et al. 2001a; Kyselý 2002a* - ECHAM3; *Kalvová et al. 2000; Huth et al. 2001a; Kyselý 2002a* - AGCM2). The coupled model CGCM1 has not yet been examined from this point of view.

Since AGCM2 simulates winter temperatures in central Europe unrealistically (*Zwiers and Kharin 1998, Kalvová et al. 2000*) due to deficiencies in parametrization schemes for soil moisture (*Palutikof et al. 1997, Laprise et al. 1998*), annual minimum temperatures are not examined in this model. Recent analyses indicate that the same problem persists in the coupled version of the model, CGCM1 (*Kharin and Zwiers 2000*), whose winter temperatures are omitted here as well.

Since the downscaled temperatures reproduce the observed means and variances (see below), for a fair comparison between direct GCM outputs and downscaling, distributions of GCM-produced temperatures were resized to have the observed mean and standard deviation (for periods May-September and November-March separately), and both these versions of GCMs (the resized one and non-resized one) were analysed and compared. The standard de-biasing procedure consisted in subtracting the mean of the simulated series, multiplying the anomalies by the ratio std_{obs} / std_{mod} where std_{obs} is the observed and std_{mod} the simulated standard deviation, and adding the observed mean.

b. Downscaling

The downscaled temperatures were calculated by the linear regression with stepwise screening from gridded 500 hPa heights and 1000/500 hPa thickness over the region which covers large portion of Europe and the adjacent Atlantic Ocean (for a detailed description of the procedure see *Huth 1999, Huth et al. 2001a*). The relationships between large-scale fields and local daily maximum and minimum temperatures were identified in observations for two seasons (May-September, November-March) separately and then applied both to observations and control GCM outputs. Two possible ways of retaining the variance of the downscaled series, namely the variance inflation (*Karl et al. 1990*) and the addition of a white noise process (cf. *Wilby et al. 1999, Zorita and von Storch 1999*) were applied in downscaling from observations and are compared here (the models are denoted DWI and DWW, respectively). As to the downscaling from GCMs, inflation of variance was used as a standard procedure; the downscaling was applied for ECHAM3 (denoted DWE) and AGCM2 (DWC).

c. Observations

The models have been evaluated against observations (covering the period 1961-1990) at four sites in central Europe (Fig. 1): Neuchâtel (Switzerland), Hamburg, Würzburg (both Germany) and Prague (the Czech Republic). All datasets span 30 years except for AGCM2 and downscaling from AGCM2, which cover 20 years.

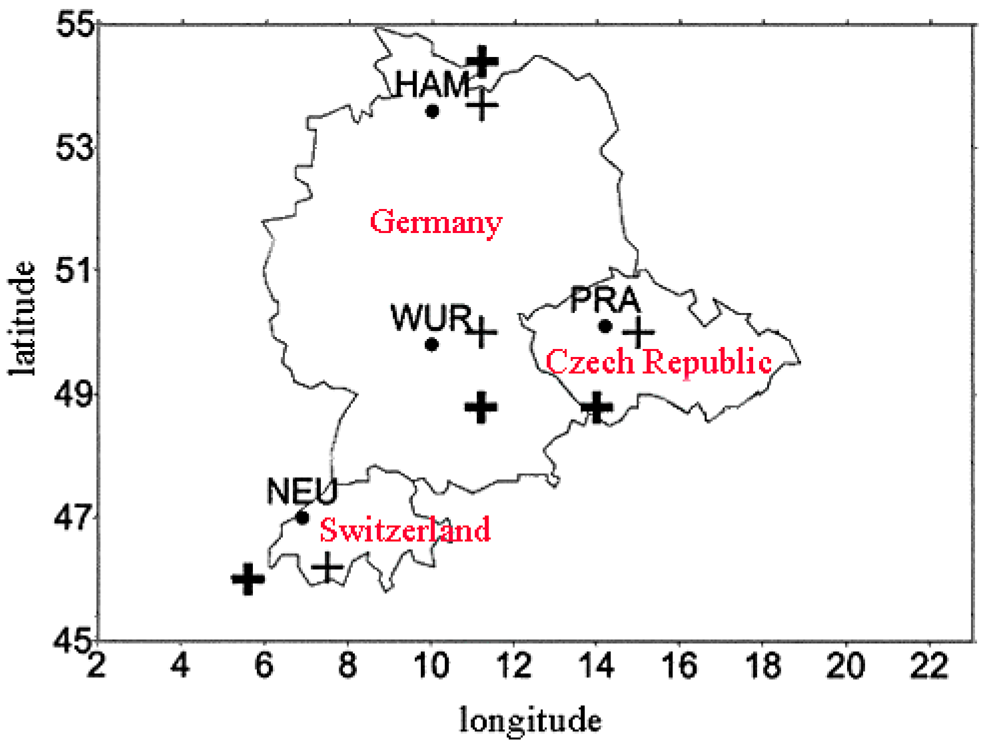

Fig. 1. Location of stations (NEU = Neuchâtel; WUR = Würzburg; HAM = Hamburg; PRA = Prague) and the closest GCM gridpoints (bold crosses for ECHAM, thin ones for AGCM2 and CGCM1).

3. Methods

a. Extreme value analysis

The extreme value analysis was performed by fitting the generalized extreme value (GEV) distribution (sometimes called the Fisher-Tippett distribution) to the sample of annual extremes of surface air temperature. The GEV distribution is expressed as

$$F(x) = \exp\left\{-\left(1-k\frac{x-\xi}{\beta}\right)^{\frac{1}{k}}\right\}, \quad x < \xi + \frac{\beta}{k}, \quad k > 0$$

$$F(x) = \exp\left\{-\exp\left(-\frac{x-\xi}{\beta}\right)\right\}, \quad k = 0$$

$$F(x) = \exp\left\{-\left(1-k\frac{x-\xi}{\beta}\right)^{\frac{1}{k}}\right\}, \quad x > \xi + \frac{\beta}{k}, \quad k < 0$$

where *F(x)* is the distribution function of random variable X and ξ, β and k are the location, scale and shape parameters of the distribution, respectively. The two-parameter Gumbel distribution is a special case (k=0) of the GEV distribution. The introduction of the shape parameter k in the GEV distribution improves the fit to the upper tail if the extremes are not Gumbel distributed; for $k < 0$ ($k > 0$) the probability density function of the GEV distribution converges more slowly (more rapidly) to zero compared to $k = 0$.

Various methods are used to estimate the parameters of the Gumbel / GEV distribution. The maximum likelihood estimators (*Jenkinson 1969*) are asymptotically optimal but they are not necessarily the best ones for finite sample sizes. Recently a new and in some aspects superior method of the estimation of the parameters of extreme value distributions has emerged that is based on L moments (*Hosking 1990*); their brief description is given e.g. in *Kharin and Zwiers (2000)* and *Kyselý (2002a, 2002b)*. Recent studies mostly employed either of the two methods (*Brown and Katz 1995, Zwiers and Kharin 1998, Kharin and Zwiers 2000*) to estimate parameters of the GEV distribution; however, results obtained by different methods were only rarely compared to each other (as e.g. by *Angel and Huff 1992*). Here, both the L moment and maximum likelihood estimators are calculated and compared.

The decision whether to use the Gumbel or the GEV distribution was based on results of the hypothesis testing, where the null-hypothesis was that the shape parameter of the GEV distribution k equals zero (*Kyselý 2002a*); three tests (the median test, *Gumbel 1965*; the maximum likelihood test, *Otten and Van Montfort 1980*; and the probability-weighted moments test, *Hosking et al. 1985*; see also *Faragó and Katz 1990* for their brief description) were performed. The results of the hypothesis testing were different for annual maximum and minimum temperatures: Whereas for maximum temperatures, the null-hypothesis was rejected in about 50% datasets (at $\alpha = 0.10$), which means that the introduction of the third parameter k improves the fit and the Gumbel distributions cannot be applied, for minimum temperatures, the null-hypothesis is rejected only in 10-15% cases. However, the fact that the null-hypothesis was rejected at two of the four stations in the observed data prevented the Gumbel distribution from being used in the comparison for annual minima as well.

Also the standard Kolmogorov-Smirnov goodness-of-fit test was applied for both the GEV and Gumbel distribution in all datasets. Since parameters of the distributions were estimated from the same sample which is compared in the test, commonly used critical values from statistical tables should not be employed (*von Storch and Zwiers 1999*). In such cases, a parametric bootstrap procedure (*Efron 1982*) yields more appropriate estimates of the critical values (*Kharin and Zwiers 2000, Kyselý 2002a*). From each fitted GEV and Gumbel distribution (for all datasets) 300 samples of the same size as observed or modelled series of annual maxima and minima were generated, and the 90% quantile of the distribution of statistic MAX used in the Kolmogorov-Smirnov test

$$\mathrm{MAX} = \max_{x \in R} \left| F(x) - G(x) \right|$$

where *F(x)* denotes the fitted distribution function and *G(x)* the empirical distribution function estimated from the sample, was taken as the critical value of the test at the significance level $\alpha = 0.10$. (These critical values are smaller in most datasets compared to the values given in tables.) The null-hypothesis that the annual extremes are drawn from the distribution *F(x)* is not rejected in any data for the GEV distribution, and in a few cases only for the Gumbel distribution. This indicates that both these distributions may be appropriate for annual maximum and minimum temperatures, but particularly for annual maxima the GEV distribution fits the data (as measured by MAX) considerably better than the Gumbel distribution.

b. Heat and cold waves

Heat and cold waves are defined as in *Huth et al. (2000, 2001a)* and *Kyselý et al. (2001)*. The definition consists of three requirements imposed on a period to be treated as a heat / cold wave; they are

(i) TMAX (daily maximum air temperature) ≥ T1 in at least 3 days; (ii) mean TMAX over the whole period ≥ T1; and (iii) TMAX ≥ T2 in each day - for heat waves, and

(i) TMIN (daily minimum air temperature) ≤ T1 in at least 3 days; (ii) mean TMIN over the whole period ≤ T1; and (iii) TMIN ≤ T2 in each day - for cold waves.

The threshold values in the heat wave definition were set to T1=30.0 °C, T2=25.0 °C, in accordance with a practice commonly applied in the Czech Republic which refers to the days with TMAX reaching or exceeding 30.0 °C and 25.0 °C as tropical and summer days, respectively. For cold waves, the threshold values were set to T1=-12.0 °C, T2=-5.0 °C so that the probabilities of TMIN lower than T1 (T2) would be comparable with frequencies of TMAX higher than T1 (T2) at the stations under study; see *Huth et al. (2001a)*.

4. Simulation of extreme one-day temperatures

a. Reproduction of the shape parameter of the GEV distribution

In evaluating the performance of the models as regards annual extremes, one may look first whether the models reproduce the shape parameter (k) of the GEV distribution (Table 1). For annual maximum temperatures, k is positive at all the four stations in the observed data, which means that the probability density function of the GEV distribution converges more rapidly to zero compared to that of the Gumbel distribution. This is reproduced by all the models at all stations except for downscaling from AGCM2 in Würzburg where k is slightly negative (-0.05). Also the value of the shape parameter averaged over the stations ($k \sim 0.3$) is captured well by all the models ($k \sim 0.20$-0.37) except for downscaling from AGCM2 (k = 0.13). For annual minimum temperatures, k is positive again at all the stations (but close to zero in Neuchâtel); the downscaling methods produce shape parameters that are mostly close to zero (with some exceptions), and so does the ECHAM GCM (except for Neuchâtel where the value is 0.33). It is obvious that the shape of the distribution of extremes is captured in the models much better for annual maximum than minimum temperatures.

Table 1. Maximum likelihood estimates of the shape parameter k of the GEV distribution in observed, GCM-simulated and downscaled datasets of annual maximum (a) and minimum (b) temperatures. The average over the stations is shown in the last row. OBS stands for observations, ECHAM, AGCM2 and CGCM1 are the three GCMs; DWI (DWW) denotes downscaling from observations with variance inflation (white noise addition) and DWE (DWC) downscaling from ECHAM (AGCM2). *na* denotes that the model was not analysed because of an unrealistic simulation.

a. annual maximum temperatures

station	OBS	ECHAM	AGCM2	CGCM1	DWI	DWW	DWE	DWC
Neuchâtel	0.30	0.31	0.30	0.17	0.20	0.32	0.47	0.30
Würzburg	0.14	0.25	0.40	0.33	0.28	0.25	0.40	-0.11
Hamburg	0.48	0.37	0.21	0.12	0.60	0.17	0.33	0.05
Prague	0.23	0.32	0.23	0.16	0.39	0.47	0.27	0.28
mean	0.29	0.31	0.29	0.20	0.37	0.30	0.37	0.13

b. annual minimum temperatures

station	OBS	ECHAM	AGCM2	CGCM1	DWI	DWW	DWE	DWC
Neuchâtel	0.06	0.33	*na*	*na*	0.14	0.40	0.06	-0.11
Würzburg	0.56	0.06	*na*	*na*	-0.02	0.16	0.19	0.26
Hamburg	0.49	-0.11	*na*	*na*	0.03	-0.17	-0.05	-0.19
Prague	0.26	-0.02	*na*	*na*	0.03	-0.15	0.08	-0.50
mean	0.34	0.07			0.05	0.06	0.07	-0.14

b. Reproduction of 20- and 50-yr return values of annual maximum and minimum temperatures

The performance of the models is evaluated in Fig. 2 using the mean error and the mean absolute error of the 20- and 50-yr return values of annual maximum temperature calculated for the four stations. The mean error (Fig. 2a) measures the model bias; among the downscaling methods, only downscaling from observations with white noise addition (and only for annual maxima) does not yield extremes that are too moderate. Both AGCM2 and CGCM1 strongly underestimate the 20- and 50-yr return values; the de-biasing procedure (after which means and variances are the same as in observations) leads to a slight but not satisfactory improvement for AGCM2 whereas the performance of CGCM1 is considerably improved. This may indicate that the newer coupled version of the Canadian climate model simulates the shape of the temperature distribution somewhat better although it still yields (if unadjusted) too low summer temperatures in central Europe. Conspicuous is the improvement of downscaling from AGCM2 compared to the AGCM2 direct output, and a relatively good performance of both the non-resized and resized versions of ECHAM (particularly the mean absolute error is reduced in the latter; see Fig. 2b). The differences between the L moment estimators (left pair of bars) and maximum likelihood estimators (right pair) are small.

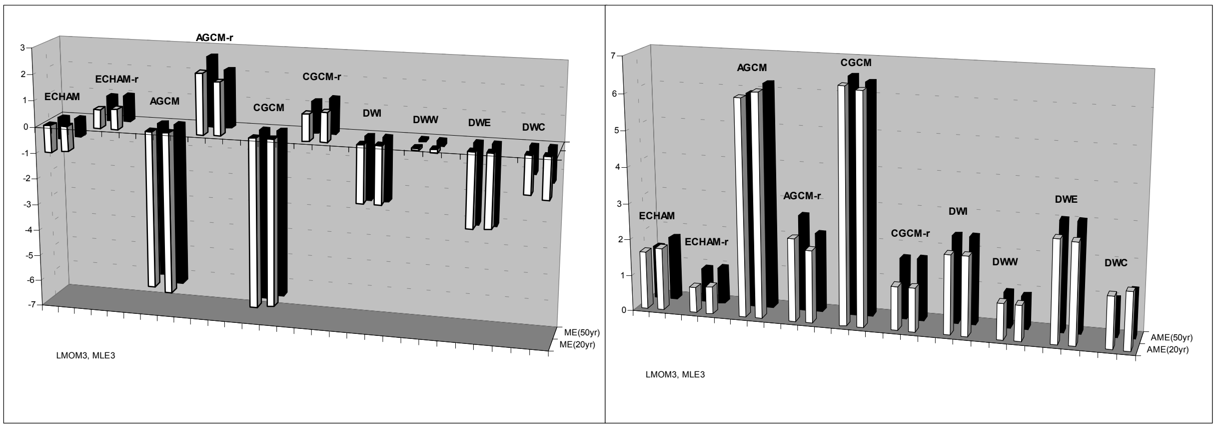

Fig. 2. Mean error (left) and mean absolute error (right) of the 20- and 50-yr return values of annual maximum temperature in the modelled series. The light (dark) pair of bars is for 20-yr (50-yr) return values, the left (right) pair for values derived using the L moment (maximum likelihood) estimators. OBS stands for observations, ECHAM-r, AGCM-r and CGCM-r are the resized versions of the ECHAM3, AGCM2 and CGCM1 GCMs; DWI (DWW) denotes downscaling from observations with variance inflation (white noise addition) and DWE (DWC) downscaling from ECHAM (AGCM2).

For annual minimum temperatures, there are only 6 models since AGCM2 and CGCM1 outputs were not analysed. The mean error (Fig. 3a) reveals that all the downscaling methods tend to provide extremes that are too moderate, and that the resized version of ECHAM does not yield better results compared to the non-resized one. Downscaling with the white noise addition attains again at least slightly better results compared to downscaling with variance inflation, and the downscaling from ECHAM performs comparably to ECHAM direct output. The latter holds, however, only when the mean error is considered; downscaling from ECHAM yields much better results when measured by the mean absolute error than ECHAM directly (Fig. 3b).

An interesting feature in Fig. 3b is that for 20-yr return values, the non-resized and resized ECHAM versions perform similarly, whereas for 50-yr values the performance of the resized version is much worse. This indicates that the application of the standard resizing procedure based on shifts in the mean and standard deviation may considerably worsen the model performance in tails of the temperature distribution. The differences between values derived using L moment and maximum likelihood estimators are for annual minima somewhat larger than those for annual maxima, but the conclusions are again unaffected by the choice of the method. Worth noting is that downscaling from AGCM2 leads to relatively good results also for annual minimum temperatures, which supports the idea that the completely unrealistic simulation of winter temperatures in this model in central Europe is caused by deficiencies in parametrizations of land surface processes (*Palutikof et al. 1997, Laprise et al. 1998*), whereas the large-scale upper-air circulation and temperature fields and their temporal variability are reproduced reasonably well at least over large parts of Europe and the adjacent Atlantic Ocean (*Huth and Pokorná 2001*).

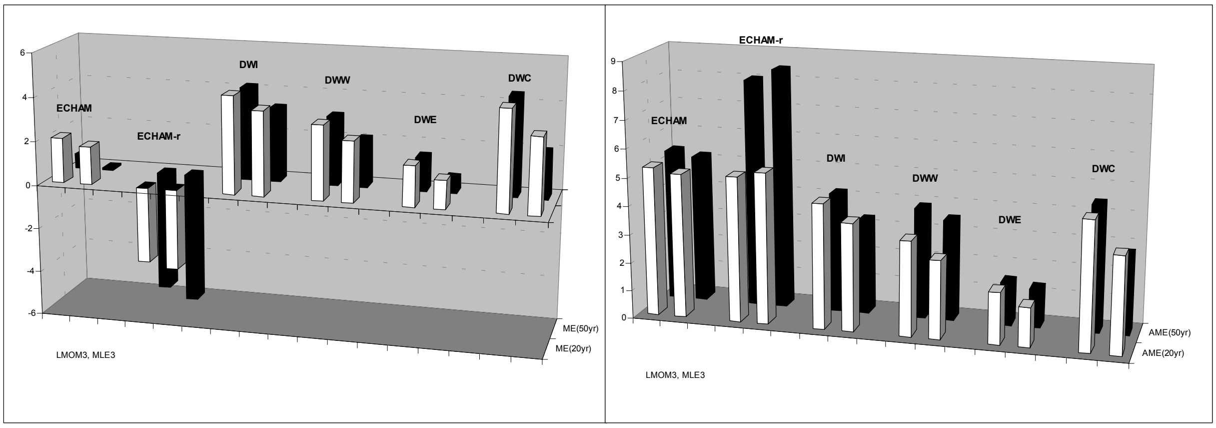

Fig. 3. The same as in Fig. 2 except for annual minimum temperature.

c. Reproduction of mean annual extremes

Generally, mean annual extremes (shown in *Kyselý 2002a*) are simulated much better than the 20- and 50-yr return values by downscaling methods and the resized GCM outputs, particularly for minimum temperatures (Table 3). The white noise addition in downscaling captures also the mean annual extremes better than the variance inflation, and the resized versions of GCMs perform in all cases considerably better compared to the non-resized ones. For direct GCM outputs, the simulation of the mean annual extremes is comparable to that of the 20- and 50-yr return values (AGCM2) or even worse (ECHAM). This is because ECHAM tends to underestimate the mean annual maxima, and the right tail of the distribution of daily temperatures in summer is flatter in ECHAM than in observations (*Huth et al. 2000*) which leads to a smaller underestimation of the 20- and 50-yr return values. A similar explanation holds for annual minima which are overestimated in ECHAM, but the left tail of daily winter temperatures is again flatter compared to observations.

5. Simulation of heat and cold waves

Generally, GCMs (provided that they are adjusted to reproduce the observed mean and variance) yield better results than statistical downscaling as regards the simulation of heat and cold waves (Table 2 and 3, Fig. 4; see *Huth et al. 2001a* and *Kyselý et al. 2001* for more details). The ECHAM3 GCM is the best among the models in simulation of cold waves (although the unadjusted temperatures are too high); of the models analyzed in *Huth et al. (2001a)*, including two versions of the four-variable stochastic weather generator, the ECHAM3 GCM is the only one that does not underestimate the extremity of cold waves. All the three GCMs are relatively successful in reproducing frequencies (Table 2) and some other properties of heat waves, e.g. the temporal evolution with the highest temperature typically reached in the second half of their duration (*Kyselý et al. 2001*). Due to the overestimated persistence of TMAX (*Kalvová and Nemešová 1998*), heat waves are too long in ECHAM (Fig. 4a), peak at too high temperatures and the inclusion of tropical days into prolonged periods is overestimated. As to heat wave characteristics, AGCM2 and CGCM1 yield better results than ECHAM3. A deformation of the temperature annual cycle with a maximum shifted towards August leads to an unrealistic position of a typical heat wave in a year in all the three GCMs analyzed. Since the errors in temperature persistence and annual cycles over continents appear to be common to many GCMs, at least some of the heat wave characteristics are likely to be misreproduced by other GCMs as well.

Since physical processes are explicitly included in GCMs only, some properties of time series that are related to them (mainly to the radiation balance and atmospheric fronts) cannot be reproduced by statistical models. Too low frequency of both heat and cold waves in the downscaled times series is influenced by the unrealistic symmetry of the day-to-day temperature change distribution and (if variance is retained by adding white noise) by a too high interdiurnal variability. The underestimation is more pronounced in winter when it is strongly influenced by a lower number of extremely cold days due to an unrealistic symmetry of the TMIN distribution. The white noise addition leads to temperature series that are too variable, and is therefore unsuitable if one is concerned with the time structure and heat and cold waves. The statistical downscaling from GCMs does not improve heat and cold wave characteristics derived from direct (adjusted) GCM outputs.

Table 2. Mean annual frequencies of heat and cold waves. Model abbreviations are given in Tab. 1, *-r* denotes re-sized versions of GCM outputs with means and variances the same as observed ones.

a. heat waves

station	OBS	ECHAM -r	AGCM2 -r	CGCM1 -r	DWI	DWW	DWE	DWC
Neuchâtel	0.60	0.90	0.75	0.53	0.30	0.33	0.17	0.45
Würzburg	0.77	0.87	0.80	0.53	0.43	0.20	0.33	0.50
Hamburg	0.20	0.40	0.20	0.07	0	0.07	0	0
Prague	0.50	0.77	0.75	0.53	0.33	0.13	0.20	0.35

b. cold waves

station	OBS	ECHAM -r	DWI	DWW	DWE	DWC
Neuchâtel	0.17	0.03	0.03	0	0.03	0
Würzburg	0.62	0.52	0.17	0.07	0.34	0.11
Hamburg	0.59	0.48	0.14	0.03	0.34	0.05
Prague	1.38	0.97	0.72	0.38	0.90	0.63

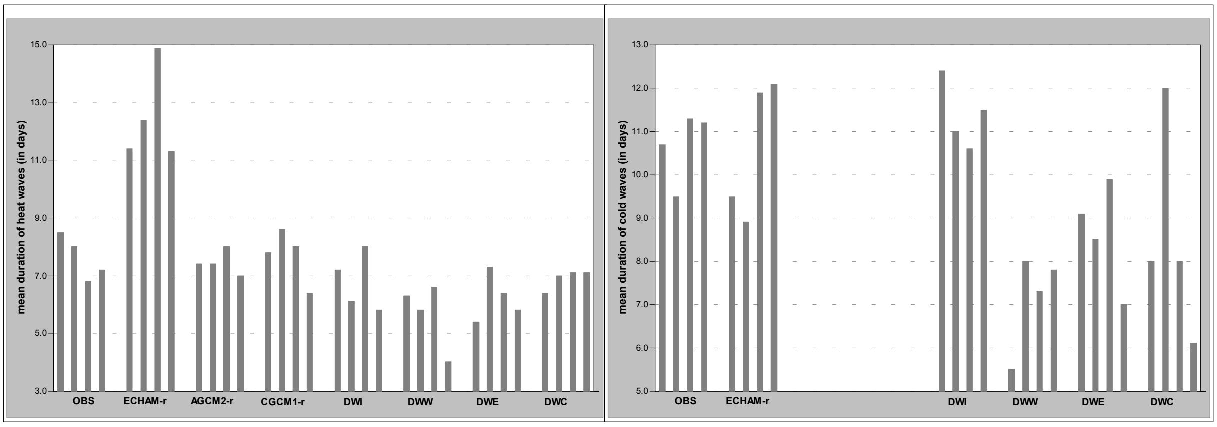

Fig. 4. Mean duration of heat waves (left) and cold waves (right). Models' abbreviations are the same as in Fig. 2, four bars for each model represent four stations. Because of a very low frequency of heat waves in Hamburg and cold waves in Neuchâtel, Strážnice (the Czech Republic) was used as the fourth station in both cases (see *Kyselý et al. 2001*).

6. Conclusions

This contribution tries to compare a set of climate models employing two views of the simulation of surface air temperature extremes: namely that of one-day extremes and of multi-day extremes. Although there is a growing number of studies dealing with extremes in climate models (mainly in GCMs), various methods of construction daily data were only occasionally compared as regards extreme events (*Trigo and Palutikof 1999, Kyselý 2000, Huth et al. 2001a*) and no study discussing and comparing results of two or more views of what is considered an 'extreme event' in temperature series in the context of climate modelling is known to the authors.

The skill of GCMs in reproducing extremely low and high temperatures is limited; of the GCMs analysed here, only ECHAM in summer is partly able to simulate extremely high temperatures. When GCM outputs are resized to preserve the observed mean and variance, their performance is much better for heat and cold waves, but not generally for one-day temperature extremes; it is partly better only for annual maxima but even worse in ECHAM for annual minima, particularly if 50-yr return values are compared. This implies the need for a more appropriate, likely non-parametric adjustment procedure (used e.g. by *Trewin and Trewitt 1996*, and proposed for validations of climate models by *Huth and Kyselý 2001*) to de-bias a control GCM output since the standard resizing procedure which adjusts the first two statistical moments may deteriorate the model's performance in tails of the distribution. This can happen even if the original (non-resized) distribution is not severely distorted (which is the case of ECHAM).

Various methods of treating the missing variance in downscaling were only rarely compared with each other as well (cf. *von Storch 1999*). Here the results are rather ambiguous since the one-day temperature extremes are much better reproduced by the white noise addition whereas heat and cold waves are better captured by downscaling with variance inflation. The relatively good simulation of one-day extremes in downscaling with white noise may be explained by much larger (compared to variance inflation) day-to-day temperature variability that further leads to more extreme annual maxima and minima; however, the same (enhanced) temperature variability prevents longer periods of hot / cold weather from being simulated and leads to a very bad reproduction of heat / cold waves.

Table 3. Models' ability to reproduce selected properties of extreme temperature events (summary). OK (**<u>OK</u>**) denotes a relatively good (very good) simulation, - (--) underestimation (strong underestimation), + (++) overestimation (strong overestimation) in the involved characteristics. *na* indicates that the characteristics was not analyzed in the model. Models' abbreviations are the same as in Tab. 1, *-r* denotes re-sized versions of GCM outputs with means and variances the same as observed ones. Inclusion of T1 days means the ratio of days with TMAX ≥ 30.0 °C (TMIN ≤ -12.0 °C) occurring within heat (cold) days to their total counts; shape parameter is the parameter *k* of the GEV distribution.

a. annual maximum and minimum temperatures

model	maximum temperatures			minimum temperatures		
	shape parameter	20-/50-yr values	mean annual extremes	shape parameter	20-/50-yr values	mean annual extremes
ECHAM3 -r	**<u>OK</u>**	OK	OK	-	--	OK
AGCM2 -r	**<u>OK</u>**	+	OK	*na*	*na*	*na*
CGCM1 -r	OK	OK	OK	*na*	*na*	*na*
ECHAM3	**<u>OK</u>**	-	-	-	+	++
AGCM2	**<u>OK</u>**	--	--	*na*	*na*	*na*
CGCM1	OK	--	--	*na*	*na*	*na*
DWI	OK	-	OK	-	++	+
DWW	**<u>OK</u>**	OK	OK	-	+	OK
DWE	OK	--	-	-	OK	OK
DWC	-	-	OK	--	++	+

b. heat and cold waves

model	heat waves			cold waves		
	frequency	duration	inclusion of T1 days	frequency	duration	inclusion of T1 days
ECHAM3 -r	+	++	+	OK	OK	**<u>OK</u>**
AGCM2 -r	OK	**<u>OK</u>**	OK	*na*	*na*	*na*
CGCM1 -r	OK	OK	-	*na*	*na*	*na*
ECHAM3	+	++	+	--	*na*	*na*
AGCM2	--	*na*	*na*	--	*na*	*na*
CGCM1	--	*na*	*na*	--	*na*	*na*
DWI	-	-	OK	--	+	-
DWW	--	--	--	--	--	--
DWE	--	-	-	-	-	-
DWC	-	-	OK	--	-	--

Also the question whether downscaling from GCMs tends to improve results obtained from GCMs directly cannot be answered for the models examined here in general; among others it depends on whether unadjusted or adjusted GCM outputs are considered. The adjustment of a GCM output to preserve the observed mean and variance appears to yield multi-day extremes that are in a better consent with observations than extremes produced by the downscaling method applied here; for one-day extremes it depends on the model and variable examined. But what is a general feature of the statistical downscaling is that it leads to extremes that are too moderate compared to observed values. This is likely strongly connected with the assumption of linearity of the downscaling method used, and an application of a more sophisticated non-linear model and a more realistic treatment of the missing variance (by adding a noise which is temporally correlated) may considerably improve performance of downscaling as to the time structure of the temperature series and extreme events.

Another important finding is that the comparison of the return values of annual maxima and minima does not appear to be sensitive to whether the L moment or maximum likelihood method is applied to estimate parameters of the GEV distribution, although individual return values are influenced by the choice of the method. A similar comparison between two statistical procedures used in extreme value analysis has only rarely been performed in relevant studies and gives a greater confidence to findings of other studies where return periods were examined using only a single method (*Zwiers and Kharin 1998, Kharin and Zwiers 2000*).

Worth noting is that the performance of the models is generally worse for annual minimum than maximum temperatures. Possible explanation takes account of the fact that in central Europe, distributions of daily minimum temperatures in winter are negatively skewed (with heavy left tail) while those of daily maximum temperatures in summer have skewness close to zero (e.g., *Huth et al. 2001b*). It means that extremely low temperatures are relatively more distant from the centre of the distribution of daily temperatures in winter compared to extremely high temperatures in summer, and the skill of climate models (particularly of the statistical models) to reproduce these low extremes, strongly influenced by radiation balance and local climatic settings of the stations, may be expected to be limited. This is in consensus with the bad simulation of the shape of the GEV distribution for annual minimum compared to maximum temperatures, too.

Acknowledgements: This research was supported by the Grant Agency of the Czech Republic (205/99/1561), Grant Agency of the Academy of Sciences of the Czech Republic (A3042903) and by the Ministry of Environment of the Czech Republic within the National Climate Programme.

References:

Angel J.R., Huff F.A., 1992: Comparing three methods for fitting extreme rainfall distributions: L-moments, maximum likelihood, and graphical fit. In: Proc. 12th Conference on Probability and Statistics in the Atmospheric Sciences, Toronto, Canada, 255-260.

Brown B.G., Katz R.W., 1995: Regional analysis of temperature extremes: Spatial analog for climate change? J. Climate 8, 108-119.

Colombo A.F., Etkin D., Karney B.W., 1999: Climate variability and the frequency of extreme temperature events for nine sites across Canada: Implications for power usage. J. Climate 12, 2490-2502.

DKRZ, 1993: The ECHAM3 Atmospheric General Circulation Model. Report No. 6, Deutsches Klimarechenzentrum, Hamburg, 184 pp.

Efron B., 1982: The Jackknife, the Bootstrap and Other Resampling Plans. SIAM, 92 pp.

Faragó T., Katz R.W., 1990: Extremes and Design Values in Climatology. WMO/TD-No 386, 43 pp.

Flato G.M., Boer G.J., Lee W.G., McFarlane N.A., Ramsden D., Reader M.C., Weaver A.J., 2000: The Canadian Centre for Climate Modelling and Analysis Global Coupled Model and its climate. Clim. Dyn., 16, 451-467.

Gumbel E.J., 1965: A quick estimation of the parameters in Fréchet's distribution. Rev. Int. Stat. Inst. 33, 349-363.

Hosking J.R.M., Wallis J.R., Wood E.F., 1985: Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. Technometrics 27, 251-261.

Hosking J.R.M., 1990: L-moments: Analysis and estimation of distributions using linear combinations of order statistics. J. Roy. Stat. Soc. 52B, 105-124.

Huth R., 1999: Statistical downscaling in central Europe: Evaluation of methods and potential predictors. Clim. Res. 13, 91-101.

Huth R., Kyselý J., Pokorná L., 2000: A GCM simulation of heat waves, dry spells, and their relationships to circulation. Clim. Change 46, 29-60.

Huth R., Kyselý J., 2001: Removing errors in simulated distributions of climate elements using a non-parametric adjustment procedure. In: Proc. 8th Int. Meeting on Statistical Climatology, Lüneburg, Germany.

Huth R., Pokorná L., 2001: Simulation of selected climatic elements by the model of the Canadian Centre for Climate Modelling. Meteorol. Zpr., 54, 129-138. [in Czech]

Huth R., Kyselý J., Dubrovský M., 2001a: Time structure of observed, GCM-simulated, downscaled, and stochastically generated daily temperature series. J. Climate, 14, 4047-4061.

Huth R., Kyselý J., Dubrovský M., 2001b: Simulation of surface air temperature by GCMs, statistical downscaling and weather generator: higher-order statistical moments. Stud. Geoph. et Geod. (accepted)

Jenkinson A.F., 1969: Statistics of extremes. In: Estimation of maximum floods, WMO Tech. Note No. 98, p. 183-257.

Kalvová J., Nemešová I., 1998: Estimating autocorrelations of daily extreme temperatures in observed and simulated climates. Theor. Appl. Climatol., 59, 151-164.

Kalvová J., Raidl A., Trojáková A., Žák M., Nemešová I., 2000: The Canadian climate model - air temperature in Europe and in the Czech Republic. Meteorol. Zpr. 53, 137-145 [in Czech]

Karl T.R., Wang W.C., Schlesinger M.E., Knight R.W., Portman D., 1990: A method of relating general circulation model simulated climate to the observed local climate. Part I: Seasonal statistics. J. Climate 3, 1053-1079.

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

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

Kyselý J., Kalvová J., 1998: Comparison of heat wave occurrence in South Moravia with outputs of the ECHAM climate model. Meteorol. Zpr., 51, 136-141. [in Czech]

Kyselý J., 2000: Changes in the occurrence of extreme temperature events. PhD thesis, Faculty of Mathematics and Physics, Charles University, Prague, 97 pp. [in Czech, with 23pp summary in English]

Kyselý J., Huth R., Dubrovský M., 2001: Simulation of extreme temperature events using general circulation models, statistical downscaling and stochastic generator. Meteorol. Zpr., 54, 73-82. [in Czech]

Kyselý J., 2002a: Comparison of extremes in GCM-simulated, downscaled and observed central-European temperature series. Clim. Res. 20, 211-222.

Kyselý J., 2002b: Probability estimates of extreme temperature events: stochastic modelling approach vs. extreme value distributions. Studia geoph. et geod., 46, 93-112.

Laprise R., Caya D., Gigučre M., Bergeron G., Côté H., Blanchet J.-P., Boer G.J., McFarlane N.A., 1998: Climate and climate change in western Canada as simulated by the Canadian Regional Climate Model. Atmos. Ocean. 36, 119-167.

McFarlane N.A., Boer G.J., Blanchet J.-P., Lazare M., 1992: The Canadian Climate Centre second-generation general circulation model and its equilibrium climate. J. Climate 5, 1013-1044.

McGuffie K., Henderson-Sellers A., Holbrook N., Kothavala Z., Balachova O., Hoekstra J., 1999: Assessing simulations of daily temperature and precipitation variability with Global Climate Models for present and enhanced greenhouse climates. Int. J. Climatol. 19, 1-26.

Meehl G.A., Zwiers F.W., Evans J., Knutson T., Mearns L., Whetton P., 2000: Trends in extreme weather and climate events: Issues related to modeling extremes in projections of future climate change. Bull. Amer. Meteorol. Soc. 81, 427-436.

Nemešová I., Kalvová J., 1997: On the validity of ECHAM-simulated daily extreme temperatures. Studia geoph. et geod., 41, 396-406.

Otten A., Van Montfort M.A.J., 1980: Maximum likelihood estimation of the general extreme value distribution parameters. J. Hydrol. 47, 187-192.

Palutikof J.P., Winkler J.A., Goodess C.M., Andresen J.A., 1997: The simulation of daily temperature time series from GCM output. Part I: Comparison of model data with observations. J. Climate 10, 2497-2513.

Parmesan C., Root T.L., Willig M.R., 2000: Impacts of extreme weather and climate on terrestrial biota. Bull. Amer. Meteorol. Soc. 81, 443-450.

Trewin B.C., Trevitt A.C.F., 1996: The development of composite temperature records. Int. J. Climatol. 16, 1227-1242.

Trigo R.M., Palutikof J.P., 1999: Simulation of daily temperatures for climate change scenarios over Portugal: a neural network model approach. Clim. Res. 13, 45-59.

von Storch H., 1999: On the use of 'inflation' in statistical downscaling. J. Climate 12, 3505-3506.

von Storch H., Zwiers F., 1999: Statistical Analysis in Climate Research. Cambridge University Press, 484 pp.

Watson R.T., Zinyowera M.C., Moss R.H., Dokken D.J. (eds), 1996: Climate Change 1995: Impacts, Adaptations, and Mitigation of Climate Change: Scientific-Technical Analyses. Published for Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, 878 pp.

Wilby R.L., Hay L.E., Leavesley G.H., 1999: A comparison of downscaled and raw GCM output: implications for climate change scenarios in the San Juan River basin, Colorado. J. Hydrol. 225, 67-91.

Zorita E., von Storch H., 1999: The analog method as a simple downscaling technique: Comparison with more complicated methods. J. Climate 12, 2472-2489.

Zwiers F.W., Kharin V.V., 1998: Changes in the extremes of the climate simulated by CCC GCM2 under CO_2 doubling. J. Climate 11, 2200-2222.