

# Comparison of extremes in GCM-simulated, downscaled and observed central-European temperature series

Jan Kyselý\*

Institute of Atmospheric Physics AS CR, Boční II/1401, 14131 Prague 4, Czech Republic

**ABSTRACT:** This study concentrates on the comparison of 20 and 50 yr return values of annual maximum and minimum temperatures in (1) observations, (2) general circulation model (GCM) simulated control climates, (3) statistical downscaling from observations, and (4) statistical downscaling from GCMs. Temperature series at 4 sites in central Europe and in the nearest GCM gridpoints corresponding to the stations are analysed. Extreme value analysis was performed by fitting the generalized extreme value (GEV) distribution using  $L$  moment and maximum likelihood estimators. The comparison does not appear to be sensitive to which statistical method of the estimation of parameters of the GEV distribution is used, although individual return values are influenced by the choice of method. The skill of both GCMs in reproducing extreme high and low temperatures is limited, and statistical downscaling from GCMs tends to improve their performance, although it generally yields extremes that are too moderate compared to observed values (this holds for downscaling both from observations and GCMs). White noise addition in downscaling from observations leads to more realistic return values of annual maximum and minimum temperatures than variance inflation.

**KEY WORDS:** General circulation model · Statistical downscaling · Validation · Surface temperature · Central Europe · Extreme value analysis · GEV distribution · Maximum likelihood method ·  $L$  moments

*Resale or republication not permitted without written consent of the publisher*

## 1. INTRODUCTION

Extreme climatic phenomena are a subject of investigation because of both their current impacts on ecosystems and society and the threat of their possible increases in severity and duration in the climate perturbed by enhanced concentrations of greenhouse gases in the atmosphere. Impacts of climate change would 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 mean and variance of climate variables can induce considerable changes in the severity of extreme events (Katz

& Brown 1992, Hennessy & Pittock 1995, Colombo et al. 1999).

General circulation models (GCMs) are the most frequently used tool in climate modelling. They are able to qualitatively reproduce many of the 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. Studies of climate change impacts usually require daily time series of climate variables for a future climate state at a specific site. There are several ways to obtain site-specific daily time series, which are to a different extent based on GCM outputs, one of them being statistical downscaling. Statistical downscaling takes advantage of the fact that GCMs simulate large-scale upper-air fields more

\*E-mail: honza@ufa.cas.cz

accurately than the surface local variables (Kim et al. 1984, Huth 1999). It consists in identifying in the observed data the statistical 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 phenomena has recently been paid in GCM studies, in both validating the simulated present-day climate and analyzing the possible future climate (Meehl et al. 2000). These studies have mostly concentrated on precipitation characteristics such as frequencies of extreme rainfall events and/or dry spells (e.g. Joubert et al. 1996, Gregory et al. 1997, Huth et al. 2000, Kharin & Zwiers 2000). In regard to extremes of surface temperature, relatively little work has been done in GCM and downscaling studies. Recently, Zwiers & Kharin (1998) and Kharin & Zwiers (2000) analysed return periods of annual maxima and minima of surface temperature (and other climatic variables) in 2 GCMs of the Canadian Centre for Climate Modelling and Analysis; Trigo & 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 5 GCMs with a coarse spatial resolution; and Huth et al. (2001) dealt with heat and cold waves in 2 GCMs and in downscaled and stochastically generated temperature series. Generally, the skill of GCMs in reproducing extreme temperature events is limited, although it is better than the simulation of extreme precipitation events, where sub-grid scale processes play a more important role in producing observed extremes.

Here we concentrate on the comparison of 20 and 50 yr return values of annual maximum and minimum temperatures in the observed, GCM-simulated and downscaled series in central Europe. The paper is organized as follows: Data and methods used are described in Section 2, including the extreme value analysis. The performance of the models is evaluated in Section 3, and main conclusions are drawn in Section 4.

## 2. DATA AND METHODS

### 2.1. Data

#### 2.1.1. GCMs

Simulations of the present climate of 2 GCMs were used in this study. The ECHAM GCM originates from the ECMWF (European Centre for Medium range

Weather Forecast) model, modified at the Max Planck Institute for Meteorology in Hamburg. A detailed description of its version 3, used here, is given in DKRZ (1993). It has a T42 resolution, corresponding approximately to a  $2.8^\circ$  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. The Canadian Climate Centre Model (CCCM) of the second generation is described in McFarlane et al. (1992), where also its basic validation is presented. CCCM has a T32 resolution, corresponding to a  $3.75^\circ \times 3.75^\circ$  grid. It was run with a mixed-layer ocean model and a thermodynamic sea ice model; 20 yr of its control integration were available. Both these models were used in Huth et al. (2001) to examine the time structure of the temperature series and heat and cold wave characteristics. The 10, 20 and 50 yr return values of annual maximum and minimum temperatures in CCCM2 were analysed by Zwiers & Kharin (1998). Since CCCM2 simulates winter temperatures in central Europe unrealistically (Zwiers & Kharin 1998, Kalvová et al. 2000), due to deficiencies in parameterization schemes for soil moisture (Palutikof et al. 1997, Laprise et al. 1998), annual minimum temperatures were not examined in this model here. Recent analyses indicate that the same problem persists in the new coupled version of the model (Kharin & Zwiers 2000).

#### 2.1.2. Downscaling

The downscaled temperatures were calculated by the multiple linear regression with stepwise screening from gridded 500 hPa heights and 1000/500 hPa thickness over the region bounded by the  $16.9^\circ$ W and  $28.1^\circ$ E meridians and the  $32.1$  and  $65.6^\circ$ N parallels with a gridstep of  $5.6^\circ$  in both latitude and longitude (the area covers a large portion of Europe and the adjacent Atlantic Ocean; for a detailed description of the procedure see Huth 1999, Huth et al. 2001). The stepwise regression of gridded values was selected because in the intercomparison study by Huth (1999) it turned out to perform best among several linear methods, including regression of principal components and canonical correlation analysis. The relationships between large-scale fields and local daily maximum and minimum temperatures were identified in observations for 2 seasons (May–September and November–March) separately and then applied to both observations and control GCM outputs. The observed 1000 and 500 hPa height fields were taken from the NCEP reanalyses (Kalnay et al. 1996), interpolated using bicubic splines from the original  $5^\circ \times 5^\circ$  grid onto the ECHAM's grid with double spacing ( $5.6^\circ \times 5.6^\circ$ ). The geopotential data from the CCCM were interpolated

onto the same grid. 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 & von Storch 1999), were applied in downscaling from observations and are compared here; as regards the downscaling from GCMs, inflation of variance was used as a standard procedure.

Since downscaled temperatures reproduce the observed means and variances, for a fair comparison between direct GCM outputs and downscaling, the distributions of GCM-produced temperatures were resized to have the observed mean and standard deviation, and both these versions of GCMs (the resized one and non-resized one) were analysed and compared.

### 2.1.3. Observations

The models were evaluated against observations (covering the period 1961–1990) at 4 sites in central Europe (Fig. 1): Neuchâtel (Switzerland), Hamburg, Würzburg (both Germany) and Prague (Czech Republic). All datasets spanned 30 yr, except for CCCM and downscaling from CCCM, which covered 20 yr.

## 2.2. 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.

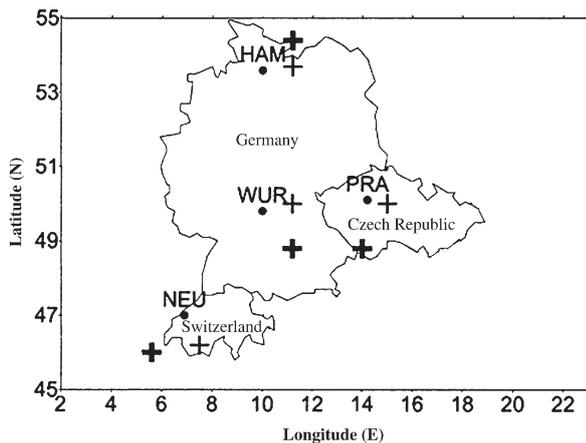


Fig. 1. Location of stations (NEU: Neuchâtel; WUR: Würzburg; HAM: Hamburg; PRA: Prague) and the closest GCM grid points (bold crosses for ECHAM, thin ones for CCCM)

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

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

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

where  $F(x)$  is the distribution function of random variable  $X$  (annual extreme in our case) and  $\xi$ ,  $\beta$  and  $k$  are the location, scale and shape parameters of the distribution, respectively. (Above are shown expressions for the distribution of maximum values; to deal with minima, the easiest way is to transform the random variable  $X$  to  $-X$  and use the same formulae.) The 2-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 method for the estimation of the parameters of extreme value distributions has emerged that is based on  $L$  moments (Hosking 1990; see Appendix 1). The method of  $L$  moments is closely related to the method of probability-weighted moments, described, for example, in Hosking et al. (1985), Buishand (1986) and Faragó & Katz (1990).

Recent climatological studies mostly employed either the maximum likelihood method (Brown & Katz 1995) or the method of  $L$  moments (Zwiers & Kharin 1998, Kharin & Zwiers 2000) to estimate parameters of the GEV distribution; however, only rarely were results obtained using  $L$  moments compared to those of maximum likelihood estimators or of other methods (as e.g. by Angel & 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; 3 tests (the median test, Gumbel 1965; the maximum likelihood test, Otten & Van Montfort 1980; and the probability-weighted moments test, Hosking et al. 1985; see also Faragó & Katz 1990 for their brief description) were performed. The results of the

Table 1. Percentage of datasets where the null hypothesis,  $H_0: k = 0$ , was rejected at  $\alpha = 0.05$  and  $\alpha = 0.10$  for the samples of annual maximum and minimum temperatures

	Maximum temperature		Minimum temperature	
	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.10$
Maximum likelihood test	30.6	44.4	3.6	7.1
Median test	25.0	38.9	7.1	14.3
Probability-weighted moments test	50.0	66.7	10.7	17.9

hypothesis testing were different for annual maximum and minimum temperatures (Table 1). Whereas for maximum temperatures, the null hypothesis was rejected in about 50 % of the datasets ( $\alpha = 0.10$ ), 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 was rejected in only 10 to 15 % of the cases ( $\alpha = 0.10$ ). Unfortunately, the null-hypothesis was rejected at 2 of the 4 stations in the observed data, which prevented the Gumbel distribution from being used in the comparison for annual minima as well.

The standard Kolmogorov-Smirnov goodness-of-fit test was also applied for both the GEV and Gumbel distributions 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 & Zwiers 1999). In such cases, a bootstrap procedure yields more appropriate estimates of the critical values (Efron 1982, Kharin & Zwiers 2000). 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,

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

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. However, only rather exceptionally does this test lead to the rejection of the null hypothesis that the annual extremes are drawn from the distribution  $F(x)$  for the Gumbel distribution, and in no cases for the GEV distribution. This indicates that both these distributions can be applied for annual maximum and minimum temperatures; particularly for annual maxima, the GEV distribution fits the data (as measured by MAX) considerably better than the Gumbel distribution, whereas for annual minima it is only slightly superior. The value of MAX does not exceed 0.15 (which is far from the critical value at 0.10 level) in any dataset

for both the annual maximum and minimum temperatures when the GEV distribution is considered.

### 3. EVALUATION OF PERFORMANCE OF THE MODELS

#### 3.1. Reproduction of the shape parameter of the GEV distribution

When evaluating the performance of the models as regards annual extremes, one may look first at whether the models reproduce the shape parameter ( $k$ ) of the GEV distribution (Table 2). (Since the adjustment procedure applied to GCMs has no effect on  $k$ , no difference is made between the resized and non-resized GCM outputs in this subsection.) For annual maximum temperatures,  $k$  is positive at all the 4 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 CCCM in Würzburg, where  $k$  is slightly negative ( $-0.11$ ). Also the value of the shape parameter averaged over the stations ( $k \sim 0.3$ ) is captured very well by all the models ( $k \sim 0.29$  to  $0.37$ ) except for downscaling from CCCM ( $k = 0.13$ ). 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.

#### 3.2. Reproduction of 20 and 50 yr return values of annual maximum and minimum temperatures

The 20 and 50 yr return values of annual maximum temperature (Fig. 2) were obtained by fitting the GEV distribution using the method of maximum likelihood (Fig. 2a) and  $L$  moments (Fig. 2b). The 4 bars for obser-

vations and for each of the 8 models correspond to the 4 examined stations (Neuchâtel, Würzburg, Hamburg and Prague).

The  $L$  moment and maximum likelihood estimators provide almost an identical picture; although some individual values are slightly different, there are no differences in the overall results. The standard deviations of the 20 and 50 yr return values of annual maximum temperature obtained by the maximum likelihood method are mostly about 0.5 and 0.6°C, respectively, which is much less than a typical difference among the various datasets (Figs. 2 & 3). CCCM strongly underestimates the 20 and 50 yr return values, and the improvement of the resized version (where means and variances are the same as in observations) is not satisfactory (the return values tend to be overestimated). On the other hand, ECHAM is relatively successful, and the adjustment to the observed means and variances further improves its performance. Of the 2 downscaling methods from observations, the white noise addition yields better results than the variance inflation, for which the return values are too low; concerning downscaling from the GCMs, downscaling from ECHAM worsens the results compared to direct GCM output, whereas downscaling from CCCM improves them.

The performance of the models is evaluated in Fig. 3 using the mean error (defined as the average of the dif-

ferences between the simulated and observed values) and the mean absolute error (defined as the average of the absolute values of the differences between the simulated and observed values) of the 20 and 50 yr return values of annual maximum temperature calculated from the 4 stations. The mean error (Fig. 3a) 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. Conspicuous are the improvement of downscaling from CCCM compared to the CCCM 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; Fig. 3b). The differences between the  $L$  moment estimators (left pair of bars) and maximum likelihood estimators (right pair) are very small.

For annual minimum temperatures, there are only 6 models, since CCCM outputs were not analysed. The differences among stations (as regards observed minima) are somewhat larger than in the case of maximum temperatures, so much better insight can again be obtained using the mean error and the mean absolute error (Fig. 4). The standard deviations of the 20 and 50 yr return values of annual minimum temperatures (about 1.2 and 1.5°C, respectively) are larger than for annual maxima, which indicates that some of the differences among the return values derived from the

Table 2. Maximum likelihood estimates of the shape parameter,  $k$ , of the GEV distribution and the standard deviations of the estimates [denoted  $\text{std}(k)$ ] in observed, GCM-simulated and downscaled datasets of annual (a) maximum and (b) minimum temperatures. The average over the stations is shown in the last row. OBS: observations; ECHAM and CCCM: the 2 GCMs; DWI (DWW): downscaling from observations with variance inflation (white noise addition); DWE (DWC): downscaling from ECHAM (CCCM)

Station		OBS	ECHAM	CCCM	DWI	DWW	DWE	DWC
<b>(a) Annual maximum temperatures</b>								
Neuchâtel	$k$	0.30	0.31	0.30	0.20	0.32	0.47	0.30
	$\text{Std}(k)$	0.11	0.12	0.14	0.12	0.10	0.12	0.22
Würzburg	$k$	0.14	0.25	0.40	0.28	0.25	0.40	-0.11
	$\text{Std}(k)$	0.22	0.10	0.18	0.10	0.12	0.10	0.25
Hamburg	$k$	0.48	0.37	0.21	0.60	0.17	0.33	0.05
	$\text{Std}(k)$	0.10	0.23	0.17	0.13	0.17	0.09	0.18
Prague	$k$	0.23	0.32	0.23	0.39	0.47	0.27	0.28
	$\text{Std}(k)$	0.12	0.09	0.20	0.15	0.13	0.09	0.14
Mean	$k$	0.29	0.31	0.29	0.37	0.30	0.37	0.13
<b>(b) Annual minimum temperatures</b>								
Neuchâtel	$k$	0.06	0.33		0.14	0.40	0.06	-0.11
	$\text{Std}(k)$	0.16	0.16		0.09	0.11	0.14	0.21
Würzburg	$k$	0.56	0.06		-0.02	0.16	0.19	0.26
	$\text{Std}(k)$	0.17	0.25		0.14	0.14	0.11	0.16
Hamburg	$k$	0.49	-0.11		0.03	-0.17	-0.05	-0.19
	$\text{Std}(k)$	0.14	0.16		0.13	0.17	0.15	0.25
Prague	$k$	0.26	-0.02		0.03	-0.15	0.08	-0.50
	$\text{Std}(k)$	0.16	0.17		0.19	0.17	0.13	0.50
Mean	$k$	0.34	0.07		0.05	0.06	0.07	-0.14

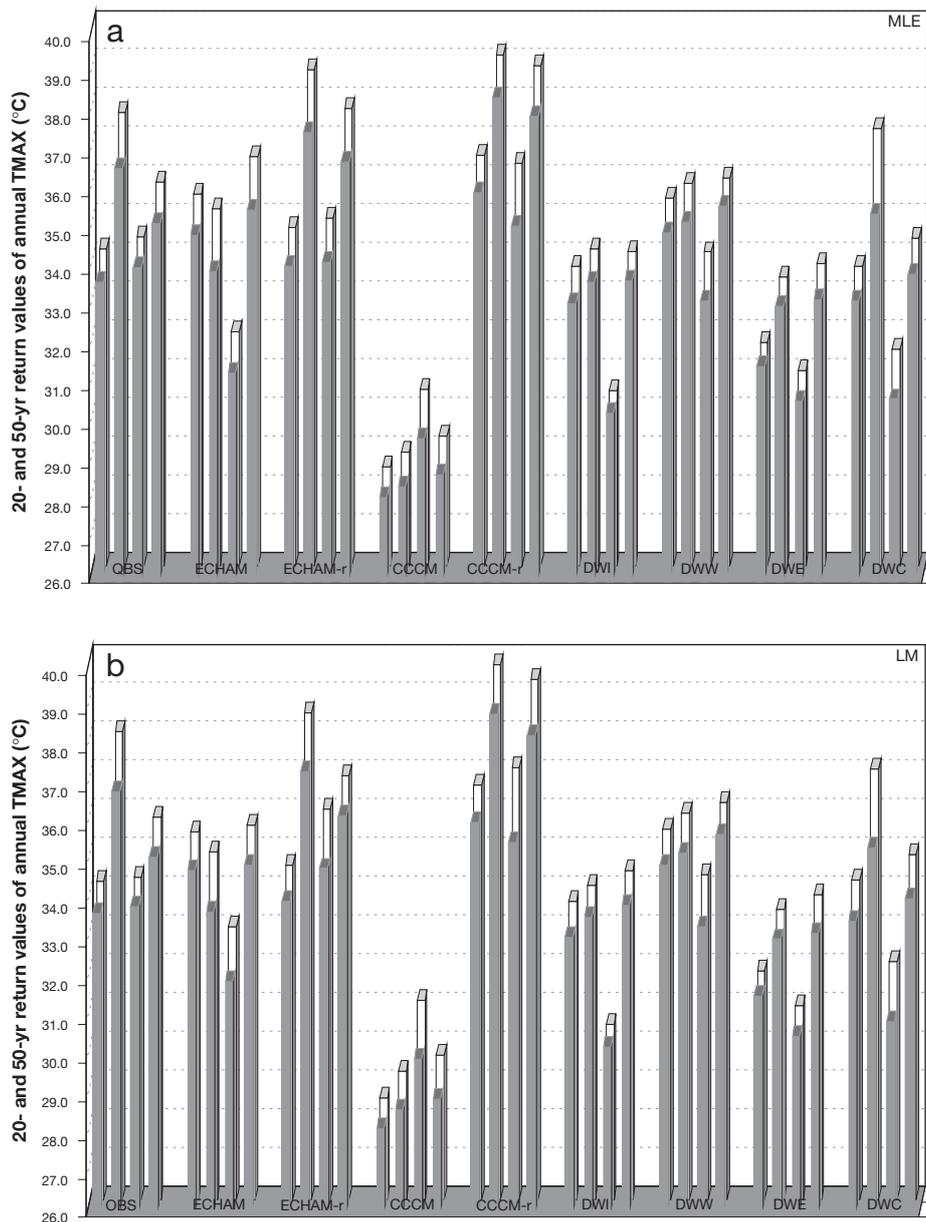


Fig. 2. The 20 and 50 yr return values of annual maximum temperature in the observed and modelled series. White tips of bars show the 50 yr values, grey parts depict 20 yr values. The 4 bars within each cluster correspond to the stations Neuchâtel, Würzburg, Hamburg and Prague (from left to right). The values were obtained by fitting the GEV distribution using the method of (a) maximum likelihood and (b) *L* moments. OBS: observations; ECHAM-r and CCCM-r: resized versions of the 2 GCMs; DWI (DWW): downscaling from observations with variance inflation (white noise addition); DWE (DWC): downscaling from ECHAM (CCCM)

analyzed datasets are not significant with respect to the errors. The mean error (Fig. 4a) 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. 4b).

An interesting feature in Fig. 4b is that for 20 yr return values, the non-resized and resized ECHAM

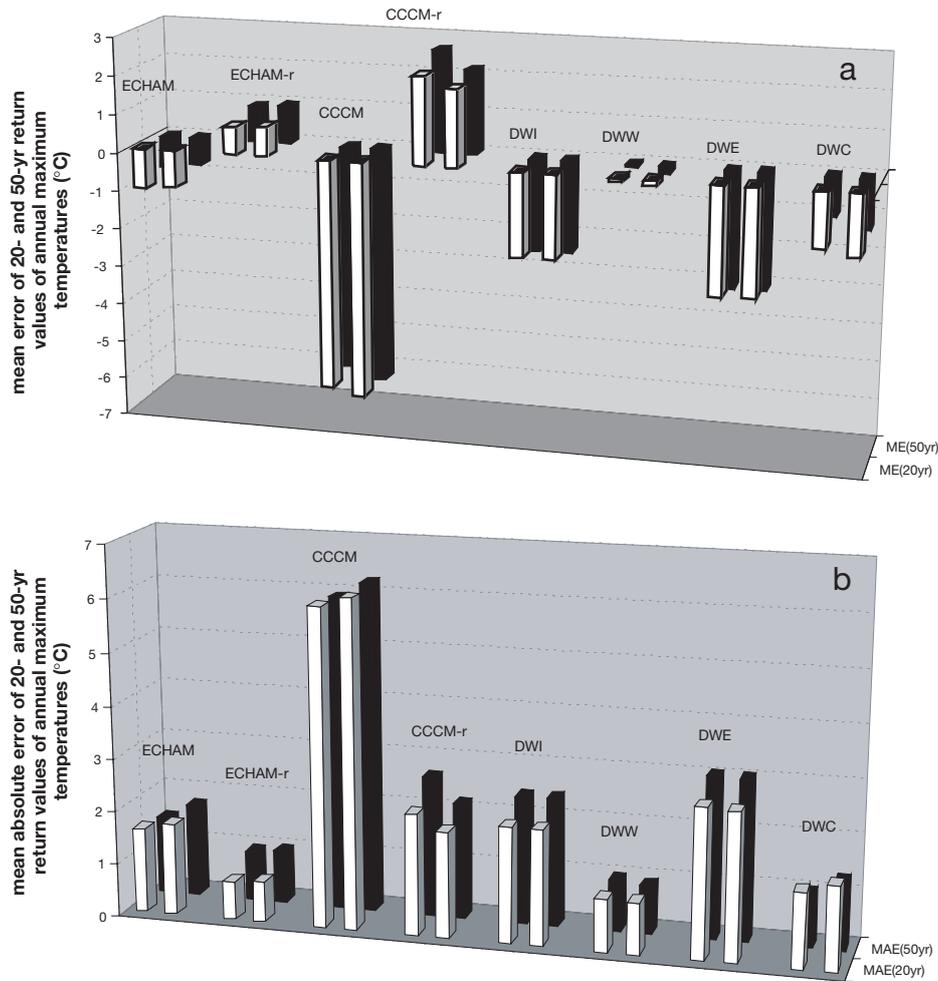


Fig. 3. (a) Mean error and (b) mean absolute error of the 20 and 50 yr return values of annual maximum temperature in the modelled series. The light (dark) pair of bars is for 20 yr (50 yr) return values, the left (right) pair for values derived using the  $L$  moment (maximum likelihood) estimators. For the dataset notation see Fig. 2

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 they were for annual maxima, but the conclusions are again unaffected by the choice of the method. Worth noting is that downscaling from CCCM 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 parameterizations of land surface processes (Palutikof et al. 1997, Laprise et al. 1998), whereas the large-scale upper-air circulation and tem-

perature fields and their temporal variability are reproduced reasonably well at least over large parts of Europe and the adjacent Atlantic Ocean (Huth & Pokorná 2001).

### 3.3. Reproduction of mean annual extremes

The models' performance with regard to the 20 and 50 yr return values may be compared to the ability of models to reproduce mean annual extremes (Fig. 5). Generally, the mean annual extremes are simulated much better than the 20 and 50 yr return values by downscaling methods and the resized GCM outputs, particularly for minimum temperatures. 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 con-

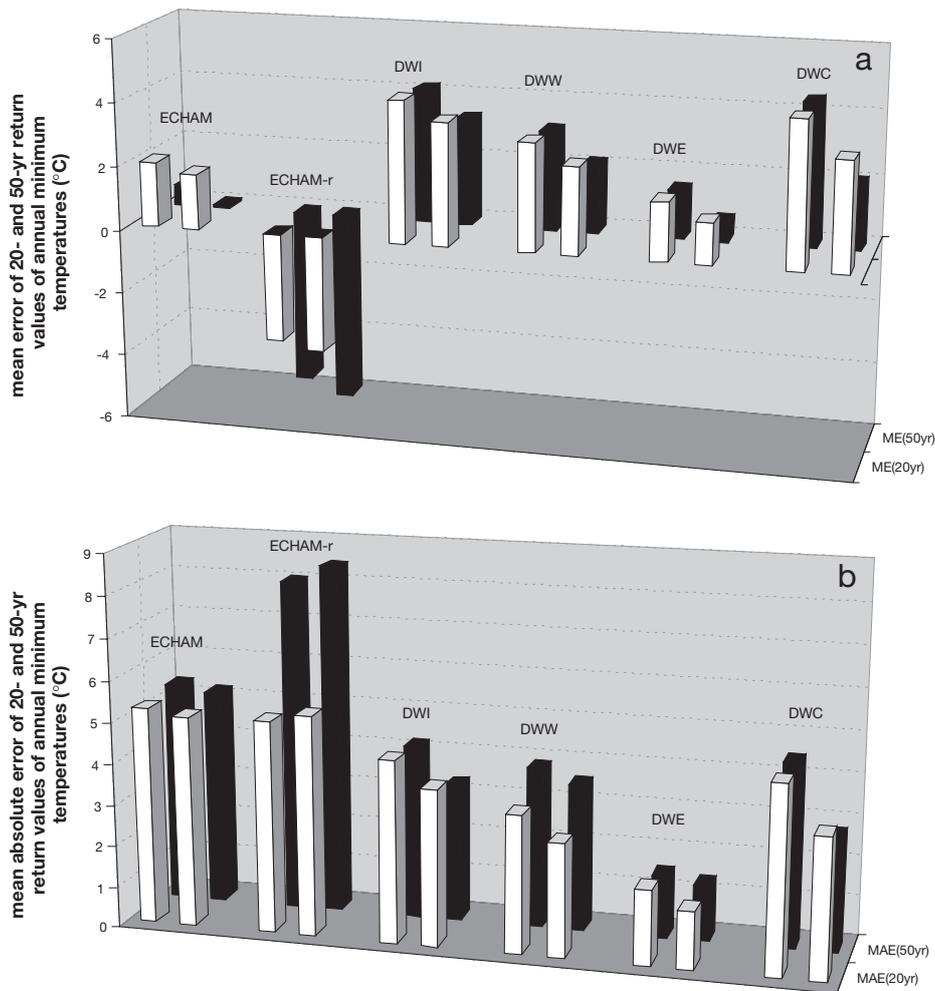


Fig. 4. (a) Mean error and (b) mean absolute error of the 20 and 50 yr return values of annual minimum temperature in the modelled series. Details as in Fig. 3

siderably 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 (CCC) 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.

#### 4. CONCLUSIONS

This study tries to fill a gap in validation of climate models in terms of extremes of surface temperature,

since only occasionally were various methods of construction daily data compared with regard to extreme events (Trigo & Palutikof 1999, Huth et al. 2001). 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 at least partly better for annual maxima but is even worse for ECHAM for annual minima, particularly when 50 yr return values are compared. This implies the need for a more appropriate, likely non-parametric adjustment procedure (used, for example, by Trewin & Trewitt 1996, and proposed for validations of climate models by Huth & Kysely 2001) to de-bias a control GCM output since the standard resizing procedure which adjusts the first 2 statistical moments may deteriorate the model's performance in tails of the distribution, and this can happen even if the

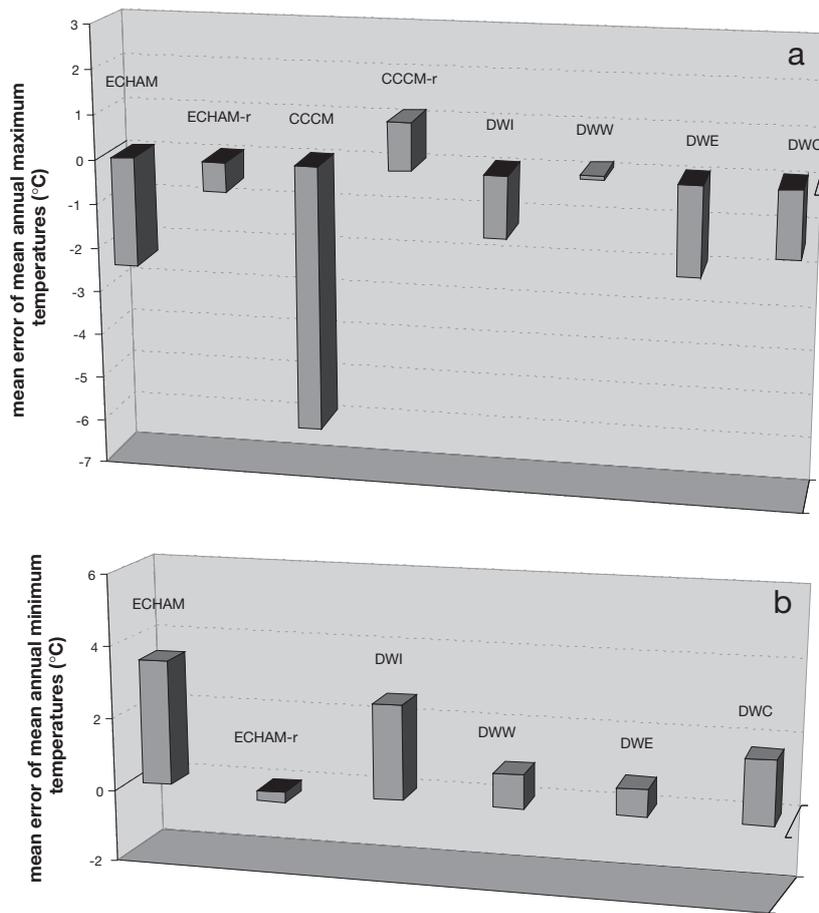


Fig. 5. Mean error of mean annual (a) maximum and (b) minimum temperatures in modelled series. For the notation of datasets see Fig. 2

original (non-resized) distribution is not severely distorted (which is the case of ECHAM).

Up to now, only rarely were various methods of treating the missing variance in downscaling compared with each other (cf. von Storch 1999). In regard to 20 and 50 yr return values of surface temperature, the white noise addition definitely yields better results than variance inflation. This may be explained by a much larger day-to-day variability (which further leads to more extreme annual maxima and minima) in downscaling with white noise compared to variance inflation. Downscaling from GCMs tends to improve results obtained from GCMs directly; however, this does not hold for annual maximum temperatures in ECHAM. Generally, downscaling 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. 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) in downscaling from GCM may considerably improve the performance of downscaling in regard to the time structure of the temperature series and extreme events.

An important finding is that the comparison 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. The insensitivity of the main results to the statistical method used is related to the fact that typical differences in the return values obtained from various datasets are much larger than differences resulting from the application of the 2 methods. A similar comparison between 2 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 one of the methods (Zwiers & Kharin 1998, Kharin & Zwiers 2000).

Worth noting is that the performance of the models is generally worse for annual minimum than maximum temperatures. A possible explanation is 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 a skewness close to zero (e.g. Huth et al. 2002). This 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, can be expected to be limited. This is in agreement with the bad simulation of the shape of the GEV distribution for annual minimum compared to maximum temperatures.

The presented analysis provides only one special view of the datasets and the ability of the models even in regard to extreme temperatures. A relatively successful simulation of 20 and 50 yr return values (as, for example, by downscaling with white noise) does not indicate that the model is able to reproduce, for exam-

ple, properties of heat and cold waves that are strongly related to the time structure of temperature series, particularly to the distribution of day-to-day temperature changes and to lag-1 correlations (Huth et al. 2001).

The ability of the analyzed climate models to capture characteristics of extreme temperature events is limited, but with advances in climate modelling and in the methods of the adjustment of model outputs (namely the methods of dynamical and statistical downscaling), an improvement in this domain is expected.

*Acknowledgements.* This research was supported by the Grant Agency of the Czech Republic and Grant Agency of the Academy of Sciences of the Czech Republic under Grant Projects No. 205/99/1561 and A3042903, and by the Ministry of Environment of the Czech Republic within the National Climate Programme. Thanks are due to R. Huth, Institute of Atmospheric Physics, Prague, for helpful comments on the manuscript and providing the datasets of downscaled temperatures; and to the 3 anonymous reviewers for suggestions which helped to improve the text. Maximum likelihood estimators of parameters of the GEV distribution were calculated using an updated version of 'Extreme value analysis' computer code described in Faragó & Katz (1990).

#### Appendix 1. Use of $L$ moments in estimating parameters of the GEV distribution

Sample  $L$  moments are based on order statistics which are obtained simply by sorting the sample  $\{X_1, X_2, \dots, X_n\}$  of  $n$  independent realizations of variable  $X$  in ascending order  $\{X_{1:n}, X_{2:n}, \dots, X_{n:n}\}$ ; the subscript  $k:n$  denotes the  $k$ th smallest number in the sample of length  $n$ .  $L$  moments are defined as expectations of linear combinations of these order statistics,

$$\lambda_1 = E(X_{1:n})$$

$$\lambda_2 = \frac{1}{2}E(X_{2:n} - X_{1:n})$$

$$\lambda_3 = \frac{1}{3}E(X_{3:n} - 2X_{2:n} + X_{1:n})$$

and generally for the  $k$ th  $L$  moment

$$\lambda_k = \frac{1}{k} \sum_{j=0}^{k-1} (-1)^j \binom{k-1}{j} E(X_{k-j:n})$$

where  $E$  denotes expectation operator (see, for example, von Storch & Zwiers 1999). The first  $L$  moment is the conventional first moment; the second  $L$  moment (an analogue to the conventional second moment) measures the variance of the sample; and the third and fourth  $L$  moments are shape parameters. Standardized  $L$  moments are the  $L$ -coefficient of variation  $\frac{\lambda_2}{\lambda_1}$ , the  $L$ -skewness  $\frac{\lambda_3}{\lambda_2}$  and the  $L$ -kurtosis  $\frac{\lambda_4}{\lambda_2}$ ; they take values between  $-1$  and  $+1$  (except for some special cases of small samples).

For details concerning the estimation of  $L$  moments see, for example, Hosking (1990) and von Storch & Zwiers (1999); Hosking (1990) showed that the  $k$ th  $L$  moment  $\lambda_k$  ( $k \leq n$ ) can be estimated as

$$I_k = \sum_{l=0}^{k-1} (-1)^{k-l-1} \binom{k-1}{l} \binom{k+l-1}{l} b_l$$

where

$$b_l = \frac{1}{n} \sum_{i=1}^n \frac{(i-1)(i-2)\dots(i-l)}{(n-1)(n-2)\dots(n-l)} X_{i:n}, \quad l \geq 1,$$

and

$$b_0 = \frac{1}{n} \sum_{i=1}^n X_{i:n}$$

For the first 3  $L$  moments, estimators can be expressed in a simpler form as

$$I_1 = \frac{\sum X_i}{n}$$

$$I_2 = \frac{\sum_{i>j} (X_{i:n} - X_{j:n})}{2 \binom{n}{2}} \quad \text{and}$$

$$I_3 = \frac{\sum_{i>j>k} (X_{i:n} - 2X_{j:n} + X_{k:n})}{3 \binom{n}{3}}$$

## Appendix 1 (continued)

Details concerning  $L$  moments for probability distributions can be found in Hosking (1990). If  $X$  has the GEV distribution ( $k \neq 0$ ), the first 3  $L$  moments  $\lambda_1, \lambda_2, \lambda_3$  are given by (note that in Zwiers & Kharin 1998 there is a misprint in the expression for  $\lambda_3$ )

$$\lambda_1 = \xi + \beta \frac{1 - \Gamma(1+k)}{k}$$

$$\lambda_2 = \beta \frac{(1 - 2^{-k})\Gamma(1+k)}{k} \quad \text{and}$$

$$\lambda_3 = \beta \frac{\Gamma(1+k)(-1 + 3 \cdot 2^{-k} - 2 \cdot 3^{-k})}{k}$$

where  $\Gamma$  stands for the gamma function and  $(\xi, \beta, k)$  are parameters of the GEV distribution (e.g. Kharin & Zwiers 2000). The resulting  $L$  moment estimators of parameters  $\xi, \beta$  and  $k$  of the GEV distribution are given by (in Zwiers & Kharin 1998 there is a misprint in the expression for scale parameter  $\beta$ )

$$k = 7.8590z + 2.9554z^2$$

$$\beta = \frac{l_2 k}{(1 - 2^{-k})\Gamma(1+k)} \quad \text{and}$$

$$\xi = l_1 + \beta \frac{\Gamma(1+k) - 1}{k}$$

where  $z = \left( \frac{2}{3 + \frac{l_3}{l_2}} - \frac{\ln 2}{\ln 3} \right)$  and  $l_1, l_2$  and  $l_3$  are estimates of

the first 3  $L$  moments. For the 2-parameter Gumbel distribution the  $L$  moments estimators are  $\beta = \frac{l_2}{\ln 2}$  and  $\xi = l_1 - \frac{\gamma l_2}{\ln 2}$ , where  $\gamma$  is the Euler constant.

The main advantage of  $L$  moments over conventional moments is that the higher  $L$  moments can be estimated more reliably and they are less sensitive to outlying data values. This is because ordinary moments (unlike  $L$  moments) require involution of the data which causes greater weight to be given to the outlying values. Robust estimators of higher moments are needed to identify and fit distributions used in extreme value analysis.

## LITERATURE CITED

- Angel JR, Huff FA (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, p 255–260
- Brown BG, Katz RW (1995) Regional analysis of temperature extremes: spatial analog for climate change? *J Clim* 8: 108–119
- Buishand A (1986) Extreme value analysis of climatological data. In: Proc 3rd International Conference on Statistical Climatology. Vienna, p 145–158
- Colombo AF, Etkin D, Karney BW (1999) Climate variability and the frequency of extreme temperature events for nine sites across Canada: implications for power usage. *J Clim* 12:2490–2502
- DKRZ (1993) The ECHAM3 Atmospheric General Circulation Model. Report No. 6, Deutsches Klimarechenzentrum, Hamburg
- Efron B (1982) The Jackknife, the Bootstrap and other resampling plans. JW Arrowsmith, Bristol
- Faragó T, Katz RW (1990) Extremes and design values in climatology. WMO/TD-No. 386, WMO, Geneva
- Gregory JM, Mitchell JFB, Brady AJ (1997) Summer drought in northern midlatitudes in a time-dependent CO<sub>2</sub> climate experiment. *J Clim* 10:662–686
- Gumbel EJ (1965) A quick estimation of the parameters in Fréchet's distribution. *Rev Int Stat Inst* 33:349–363
- Hennessy KJ, Pittock AB (1995) Greenhouse warming and threshold temperature events in Victoria, Australia. *Int J Climatol* 15:591–612
- Hosking JRM (1990) L-moments: analysis and estimation of distributions using linear combinations of order statistics. *J R Stat Soc* 52B:105–124
- Hosking JRM, Wallis JR, Wood EF (1985) Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. *Technometrics* 27: 251–261
- Huth R (1999) Statistical downscaling in central Europe: evaluation of methods and potential predictors. *Clim Res* 13: 91–101
- 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. GKSS Forschungszentrum, Geesthacht
- 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, 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, Dubrovsky M (2001) Time structure of observed, GCM-simulated, downscaled, and stochastically generated daily temperature series. *J Climate* 14: 4047–4061
- Huth R, Kyselý J, Dubrovsky M (2002) Simulation of surface air temperature by GCMs, statistical downscaling and weather generator: higher-order statistical moments. *Studia Geoph Geod* (in press)
- Jenkinson AF (1969) Statistics of extremes. In: Estimation of maximum floods, WMO Tech Note No. 98, WMO, Geneva, p 183–257
- Joubert AM, Mason SJ, Galpin JS (1996) Droughts over southern Africa in a doubled CO<sub>2</sub> climate. *Int J Climatol* 16:1149–1156
- Kalnay E, and 21 others (1996) The NCEP/NCAR 40-year reanalysis project. *Bull Am Meteorol Soc* 77:437–471
- Kalvová J, Raidl A, Trojáčková 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 TR, Wang WC, Schlesinger ME, Knight RW, Portman D (1990) A method of relating general circulation model sim-

- ulated climate to the observed local climate. Part I: Seasonal statistics. *J Clim* 3:1053–1079
- Katz RW, Brown BG (1992) Extreme events in a changing climate: variability is more important than averages. *Clim Change* 21:289–302
- Kharin VV, Zwiers FW (2000) Changes in the extremes in an ensemble of transient climate simulations with a coupled atmosphere-ocean GCM. *J Clim* 13:3760–3788
- Kim JW, Chang JT, Baker NL, Wilks DS, Gates WL (1984) The statistical problem of climate inversion: determination of the relationship between local and large-scale climate. *Mon Weather Rev* 112:2069–2077
- Laprise R, Caya D, Giguère M, Bergeron G, Côté H, Blanchet JP, Boer GJ, McFarlane NA (1998) Climate and climate change in western Canada as simulated by the Canadian Regional Climate Model. *Atmos Ocean* 36:119–167
- McFarlane NA, Boer GJ, Blanchet JP, Lazare M (1992) The Canadian Climate Centre second-generation general circulation model and its equilibrium climate. *J Clim* 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 GA, Zwiers FW, 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 Am Meteorol Soc* 81: 427–436
- Otten A, Van Montfort MAJ (1980) Maximum likelihood estimation of the general extreme value distribution parameters. *J Hydrol* 47:187–192
- Palutikof JP, Winkler JA, Goodess CM, Andresen JA (1997) The simulation of daily temperature time series from GCM output. Part I: Comparison of model data with observations. *J Clim* 10:2497–2513
- Parmesan C, Root TL, Willig MR (2000) Impacts of extreme weather and climate on terrestrial biota. *Bull Am Meteorol Soc* 81:443–450
- Trewin BC, Trevitt ACF (1996) The development of composite temperature records. *Int J Climatol* 16:1227–1242
- Trigo RM, Palutikof JP (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 Clim* 12:3505–3506
- von Storch H, Zwiers F (1999) *Statistical analysis in climate research*. Cambridge University Press, Cambridge
- Watson RT, Zinyowera MC, Moss RH, Dokken DJ (eds) (1996) *Climate change 1995: impacts, adaptations, and mitigation of climate change: scientific-technical analyses*. Published for the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge
- Wilby RL, Hay LE, Leavesley GH (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 Clim* 12:2472–2489
- Zwiers FW, Kharin VV (1998) Changes in the extremes of the climate simulated by CCC GCM2 under CO<sub>2</sub> doubling. *J Clim* 11:2200–2222

*Editorial responsibility: Hans von Storch, Geesthacht, Germany*

*Submitted: April 22, 2001; Accepted: November 18, 2001  
Proofs received from author(s): February 13, 2002*