Time Structure of Observed, GCM-Simulated, Downscaled, and Stochastically Generated Daily Temperature Series

RADAN HUTH

Institute of Atmospheric Physics, Prague, Czech Republic

JAN KYSELÝ

Institute of Atmospheric Physics, and Department of Meteorology and Environment Protection, Charles University, Prague, Czech Republic

MARTIN DUBROVSKÝ

Institute of Atmospheric Physics, Prague, Czech Republic

(Manuscript received 7 September 2000, in final form 20 February 2001)

ABSTRACT

The time structure of simulated daily maximum and minimum temperature series, produced by several different methods, is compared with observations at six stations in central Europe. The methods are statistical downscaling, stochastic weather generator, and general circulation models (GCMs). Outputs from control runs of two GCMs are examined: ECHAM3 and CCCM2. Four time series are constructed by statistical downscaling using multiple linear regression of 500-hPa heights and 1000-/500-hPa thickness: (i) from observations with variance reproduced by the inflation technique, (ii) from observations with variance reproduced by adding a white noise process, and (iii) from the two GCMs. Two runs of the weather generator were performed, one considering and one neglecting the annual cycle of lag-0 and lag-1 correlations among daily weather characteristics. Standard deviation and skewness of day-to-day temperature changes and lag-1 autocorrelations are examined. For heat and cold waves, the occurrence frequency, mean duration, peak temperature, and mean position within the year are studied.

Possible causes of discrepancies between the simulated and observed time series are discussed and identified. They are shown to stem, among others, from (i) the absence of physics in downscaled and stochastically generated series, (ii) inadequacies of treatment of physical processes in GCMs, (iii) assumptions of linearity in downscaling equations, and (iv) properties of the underlying statistical model of the weather generator. In downscaling, variance inflation is preferable to the white noise addition in most aspects as the latter results in highly overestimated day-to-day variability. The inclusion of the annual cycle of correlations into the weather generator does not lead to an overall improvement of the temperature series produced. None of the methods appears to be able to reproduce all the characteristics of time structure correctly.

1. Introduction

Studies of climate change impacts frequently require daily time series of climate variables such as temperature and precipitation for a future climate state at a specific site. There are several ways of obtaining site-specific daily time series, which are to a different extent based on general circulation model (GCM) outputs (Giorgi and Mearns 1991; Kattenberg et al. 1996; Dubrovský 1997). Those frequently used include taking the GCM output itself, regional climate models (RCMs), statistical downscaling, and weather generators. In order to have confidence in simulations of a future climate, we must verify that the particular method is capable of simulating present climate conditions reliably. In this paper, we concentrate on the evaluation of simulations of present climate by several methods of producing site-specific daily temperature series; we do not deal with applications of the methods to future climate conditions. The latter subject is discussed, for example, in Hulme et al. (1994) for the use of a direct GCM output; in Hay et al. (2000) for a direct GCM output and downscaling; in Winkler et al. (1997) and Huth and Kyselý (2000) for downscaling; and in Semenov and Barrow (1997) and Dubrovský et al. (2000) for weather generators.

Many studies evaluate the performance of the above-mentioned approaches against observed data, but the vast majority concerns a single method only. A simultaneous use of several methods and a subsequent comparison would, nevertheless, aid in selecting the most

Corresponding author address: Dr. Radan Huth, Institute of Atmospheric Physics, Boční II 1401, 141 31 Prague, Czech Republic. E-mail: huth@ufa.cas.cz

appropriate technique for specific applications in future studies. There are only few studies that attempted to make such a comparison for daily temperature. Kidson and Thompson (1998) compared statistical downscaling with a limited-area model over New Zealand, arriving at the conclusion that both methods perform with a similar accuracy. Hay et al. (2000) found that downscaling from observations approximates daily temperatures in selected North American basins better than a direct GCM output and downscaling from GCM.

In most studies, simulated time series are verified against observations in terms of distance measures such as root-mean-square error and correlation coefficient, and in terms of the first two statistical moments, that is, the mean and standard deviation. These are, however, only a few of the possible criteria. In many applications, for example, in crop growth modeling, the reliability of the time structure of a series is of crucial importance. In spite of this, little has been done in examining the time structure of temperature series produced by various methods. Several studies investigated autocorrelations of temperature series in GCM and RCM outputs (Buishand and Beersma 1993; Mearns et al. 1995; Kalvová and Nemešová 1998) and the occurrence of prolonged extreme events, such as heat waves (Huth et al. 2000). No study has so far analyzed the time structure of downscaled temperature series, except for Trigo and Palutikof (1999) who examined the numbers of heat and cold waves. Hayhoe (2000) validated the weather generator in terms of the distribution of the length of frost-free periods.

The aim of this study is to analyze time series of daily extreme temperatures produced by two GCMs, a statistical downscaling procedure, and a weather generator, and verify them against observations at six central European sites. The analysis is focused on empirical distributions of day-to-day temperature changes and, primarily, on prolonged periods of high and low temperatures, that is, heat and cold waves. Although extreme events have recently begun to play a more important role in analyses of climate change simulations (Meehl et al. 2000), little attention has so far been paid to heat and cold waves. We also examine the effect on time series of (i) the way the variance is reproduced in statistical downscaling and (ii) whether the annual cycle of correlations and lagged correlations among simulated variables is implemented in the weather generator or not.

2. Datasets and methods of their construction

a. GCMs

The simulations of present climate (control runs) of two GCMs have been used in this study. The ECHAM GCM originates from the European Centre for Medium-Range Weather Forecasts model—hence EC, modified by the Max Planck Institute for Meteorology in Hamburg, Germany—hence HAM. A detailed description of its version 3, used here, is given in Deutsches Klimarechenzentrum (DKRZ 1993). It has a T42 resolution, corresponding approximately to a 2.8° grid spacing both in longitude and latitude. Here we examine years 11–40 of the control run, in which climatological SSTs and sea ice were employed. The validation of daily extreme temperatures produced by ECHAM was performed for selected areas of the Czech Republic by Nemešová and Kalvová (1997) and Nemešová et al. (1998).

The Canadian Climate Centre Model (CCCM) of the second generation is described in McFarlane et al. (1992) where also its basic validation is presented. The CCCM model has a T32 resolution, roughly corresponding to a $3.75° \times 3.75°$ grid. Its interactive lower boundary consists of a mixed layer ocean model and a thermodynamic ice model. Twenty years of its control integration have been available. The validation of its surface temperature characteristics over central Europe was performed by Kalvová et al. (2000).

There is a continuing debate as to whether GCM outputs should be treated as gridbox or gridpoint values (Skelly and Henderson-Sellers 1996), that is, if the gridded output of a model should be compared with individual station values or areal aggregates (e.g., spatial means). Huth et al. (2000) have shown that for daily maximum temperatures and heat waves, there is virtually no difference between the two approaches, and a grid point versus station comparison is thus justified. Therefore, we compare temperature characteristics at individual stations with GCM outputs at the grid points closest to them.

b. Downscaling

Downscaled temperatures were calculated by multiple linear regression with stepwise screening from gridded 500-hPa heights and 1000-/500-hPa thickness over the domain covering most of Europe and extending over the adjacent Atlantic Ocean (bounded by 32.1°, 65.6°N, 16.9°W, and 28.1°E). The stepwise regression of gridded values was selected because in the intercomparison study by Huth (1999), it turned out to perform best among other linear methods, including regression of principal components and canonical correlation analysis. In that study, a detailed description of the procedure can be found, including the selection of the most informative predictors.

The downscaling procedure is designed so that it retains the mean of the original time series. It is important to retain also variance; there are two possible ways of doing that, which are both applied and compared in this study. The commonly used way, variance inflation, consists in enhancing each day's anomaly by the same factor, defined by the share of variance explained by downscaling (Karl et al. 1990). However, the variance inflation implicitly assumes that all local variability originates in large-scale variability, which is not true (von

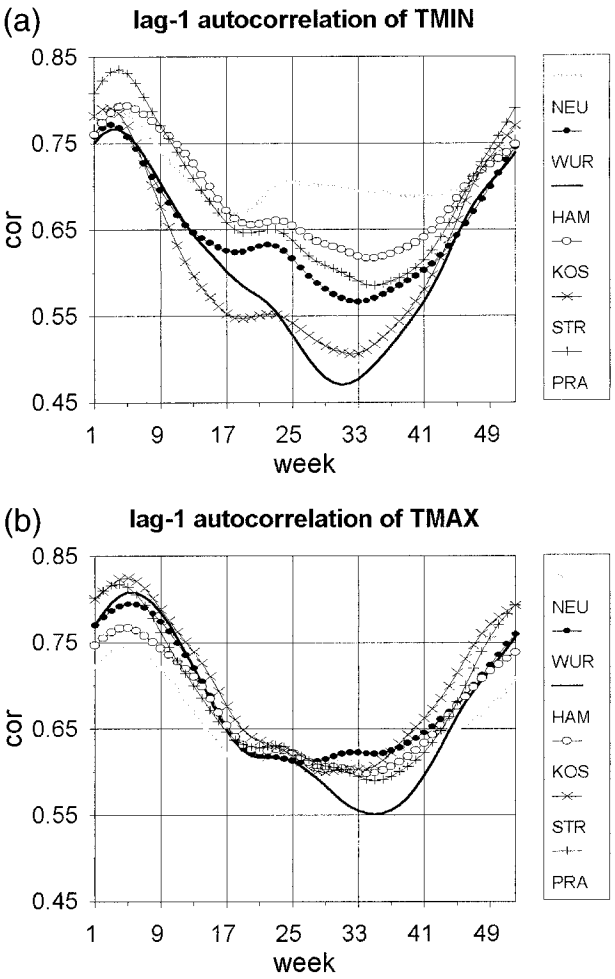


FIG. 1. Annual cycles of lag-1 autocorrelation at six stations of (a) daily minimum temperatures and (b) daily maximum temperatures. For the explanation of abbreviations of stations see the caption of Fig. 2.

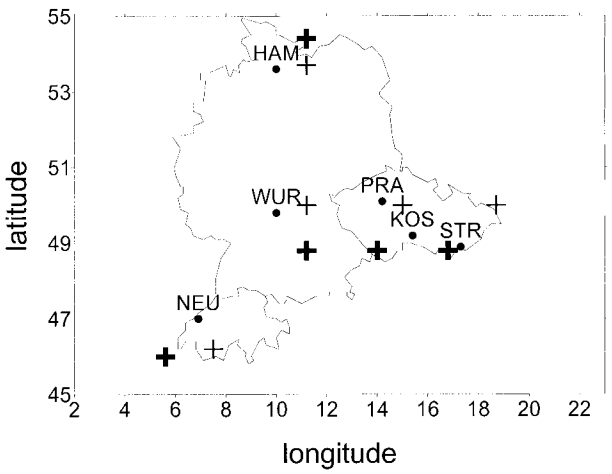


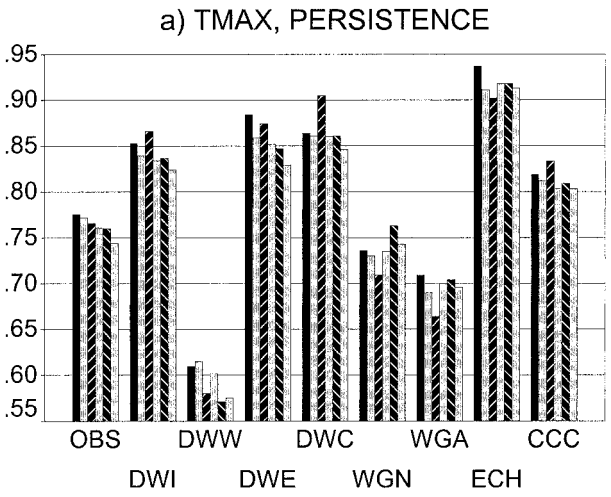
FIG. 2. Location of stations (dots; NEU = Neuchâtel; WUR = Würzburg; HAM = Hamburg; KOS = Kostelní Myslová; STR = Strážnice; PRA = Prague) and the closest GCM grid points (bold crosses for ECHAM, light crosses for CCCM).

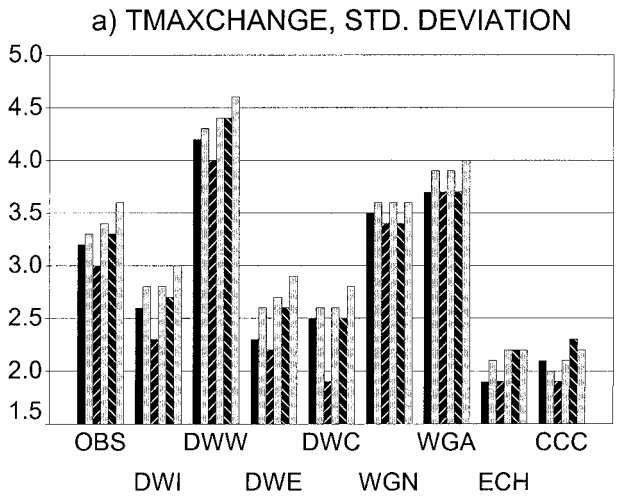
Storch 1999). Instead, it is possible to enhance the variability of a downscaled series by adding noise, which is assumed to represent the processes unresolved by the large-scale predictor. Here we enhance the variance by a white noise process, similarly to Wilby et al. (1999) and Zorita and von Storch (1999).

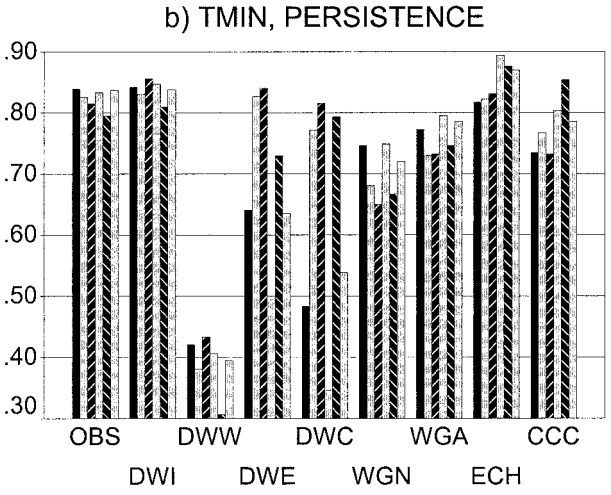
The relationships between large-scale fields and local temperature were first identified in observations, and then applied to GCM outputs. The observed 1000- and 500-hPa height fields were taken from the National Centers for Environmental Prediction reanalyses (Kalnay et al. 1996), interpolated using bicubic splines from the original $5° \times 5°$ grid onto ECHAM's grid with double spacing, that is, $5.6° \times 5.6°$. The geopotential data from CCCM were interpolated onto the same grid. In observations, the regression is performed between normalized anomalies (i.e., gridpoint values free of gridpoint long-term means, divided by temporal standard deviation at that grid point) of large-scale fields and normalized anomalies of local temperatures, for both seasons (Nov–Mar and May–Sep, see section 2d) separately. The normalization is employed because it results in a better fit of downscaled values with observations than if raw data are used (Winkler et al. 1997; Huth and Kyselý 2000). The variance of the downscaled series (which is less than 1) is first enhanced to equal 1 either by inflation or by adding white noise, and the series is then denormalized by the observed mean and standard deviation. In the application of downscaling to GCM outputs, the simulated large-scale fields are first normalized by their own mean and standard deviation. Then they enter the regression equations developed on observed data. The variance of the downscaled output is enhanced to equal 1 by inflation only, and then the time series are multiplied by the observed standard deviation and added to the observed mean similarly to observations. This procedure eliminates a GCM's bias and allows the observed mean and standard deviation to be reproduced in the GCM-downscaled time series.

TABLE 1. List of datasets and their abbreviations.

Dataset	Abbreviation
Observed	OBS
Downscaled from reanalyses; variance retained by inflation	DWI
Downsclaed from reanalyses; variance retained by adding white noise	DWW
Downscaled from the ECHAM GCM	DWE
Downscaled from the CCCM GCM	DWC
Weather generator with lag-0 and lag-1 correlations constant throughout the year	WGN
Weather generator with annual cycle of lag-0 and lag-1 correlations included	WGA
Direct output from the ECHAM GCM	ECH
Direct output from the CCCM GCM	CCC







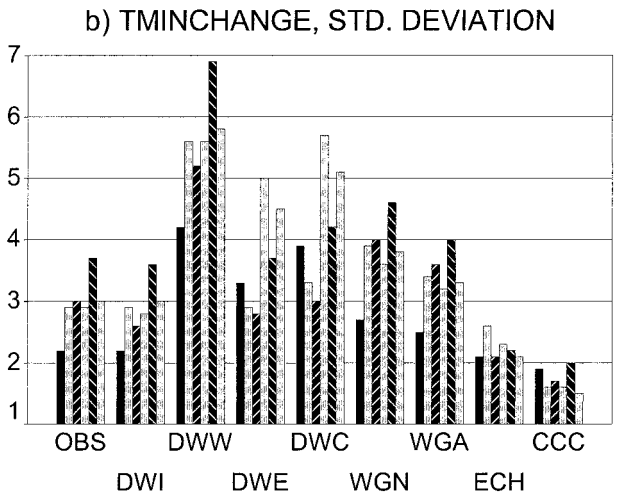


FIG. 3. Lag-1 autocorrelations of temperature for the nine datasets (clusters of bars) at six stations (from left to right within each cluster: Neuchâtel, Würzburg, Hamburg, Kostelní Myslová, Strážnice, and Prague).

FIG. 4. Standard deviation (in °C) of day-to-day temperature change distributions: (a) maximum temperature in summer and (b) minimum temperature in winter. Legends as in Fig. 3.

c. Weather generator

The stochastic weather generator Met&Roll used in this study is designed to produce synthetic weather series required, for example, in crop growth modeling (Dubrovský 1997; Dubrovský et al. 2000). Its model originates from Wilks (1992). It deals with four daily weather characteristics: maximum temperature (TMAX), minimum temperature (TMIN), sum of global solar radiation (SRAD), and the precipitation amount. Precipitation occurrence is modeled by the first-order Markov chain, precipitation amount on a wet day is fitted by the Gamma distribution. Normalized anomalies of TMAX, TMIN, and SRAD are modeled by the first-order autoregressive model; their means and standard deviations are conditioned by a precipitation occurrence and the day of the year. The annual cycles are smoothed by robust locally weighted regression (Solow 1988). The parameters of the precipitation model are determined separately for each month. The statistical structure of synthetic series produced by the generator was validated in detail by Dubrovský (1997).

The previous version of Met&Roll assumed lag-0 and lag-1 correlations among SRAD, TMAX, and TMIN to be constant. However, in reality, they vary considerably throughout the year, as can be seen in Fig. 1 for autocorrelations of TMAX and TMIN at six stations used in this study. Therefore, the annual cycle of both lag-0 and lag-1 correlations has been implemented into the model. To reveal the effect of implementing the correlations, the generator was run with and without their annual cycle. In the latter case, the correlations calculated from the whole year data were used in the autoregressive model.

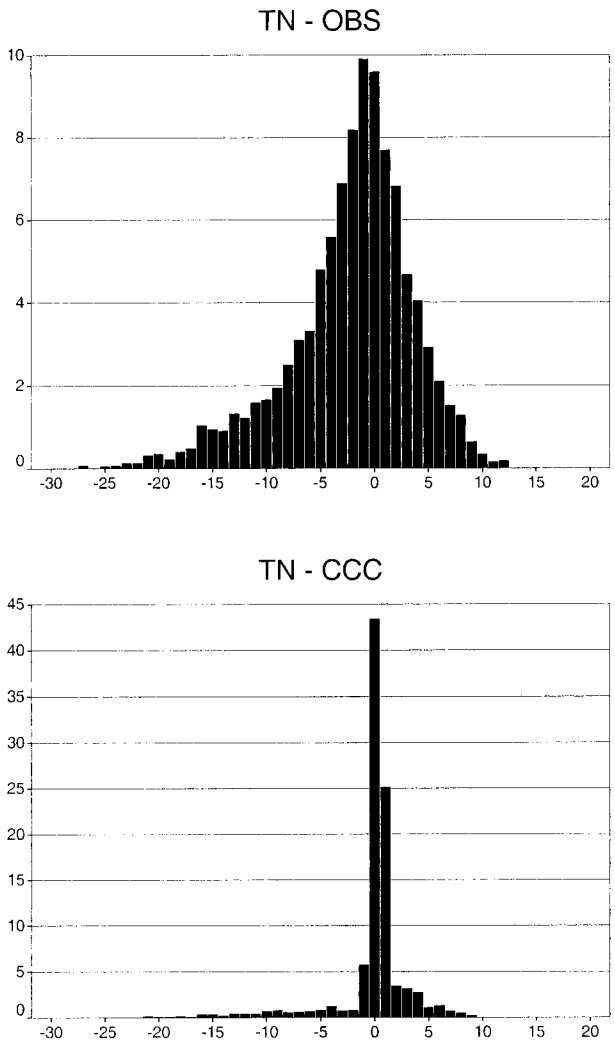


FIG. 5. Histograms of daily minimum temperature in winter (binned into 1°C intervals) at Strážnice for (top) observations and (bottom) the CCCM model.

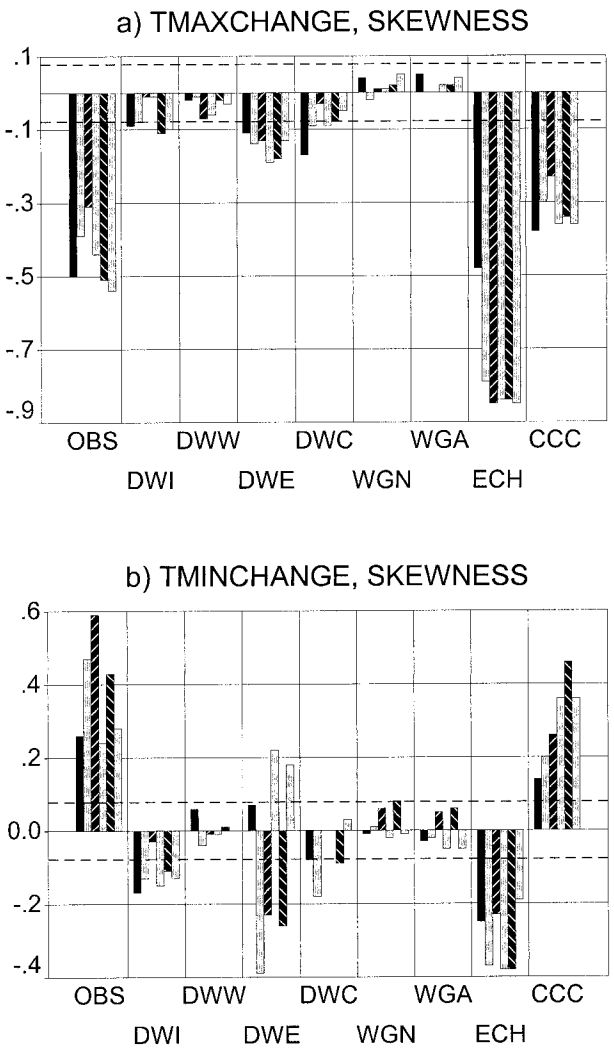


FIG. 6. Same as in Fig. 4 except for skewness. The dashed lines indicate the critical value for skewness to be different from zero at the 95% significance level, assuming autocorrelation of 0.1 (typical observed value in both seasons).

d. Datasets

Daily maximum temperature in the summer period (May–Sep) and daily minimum temperature in the winter period (Nov–Mar) are analyzed at six stations in central Europe. The stations are located in the Czech Republic, Germany, and Switzerland (see Fig. 2) in various climatic settings. The observations span the period 1961–90, and so do the downscaled time series derived from observed large-scale fields. Stochastically generated series are 30 yr long as well, but of course cannot be attributed to a specific historic period. The parameters of the generator were derived from observations in the period 1961–90. The outputs from the ECHAM and CCCM models are 30 and 20 yr long, respectively; so are the corresponding GCM-downscaled time series. The GCM grid points closest to the stations that were used for comparison are shown in Fig. 2.

3. Analysis of time series

In this section, we compare statistical properties of temperature time series produced by different methods. Namely, lag-1 autocorrelations and the second and third moments of day-to-day temperature change distributions are examined. Altogether nine time series are available at each station for either season. They are listed in Table 1 together with the abbreviations used throughout the text.

The day-to-day temperature variability is first expressed in terms of lag-1 autocorrelations (persistence; Fig. 3). The most striking feature is a large underestimation of persistence for the white noise downscaling in both seasons. The downscaling with inflation reproduces persistence well in winter but overestimates it considerably in summer; as a result, persistence values

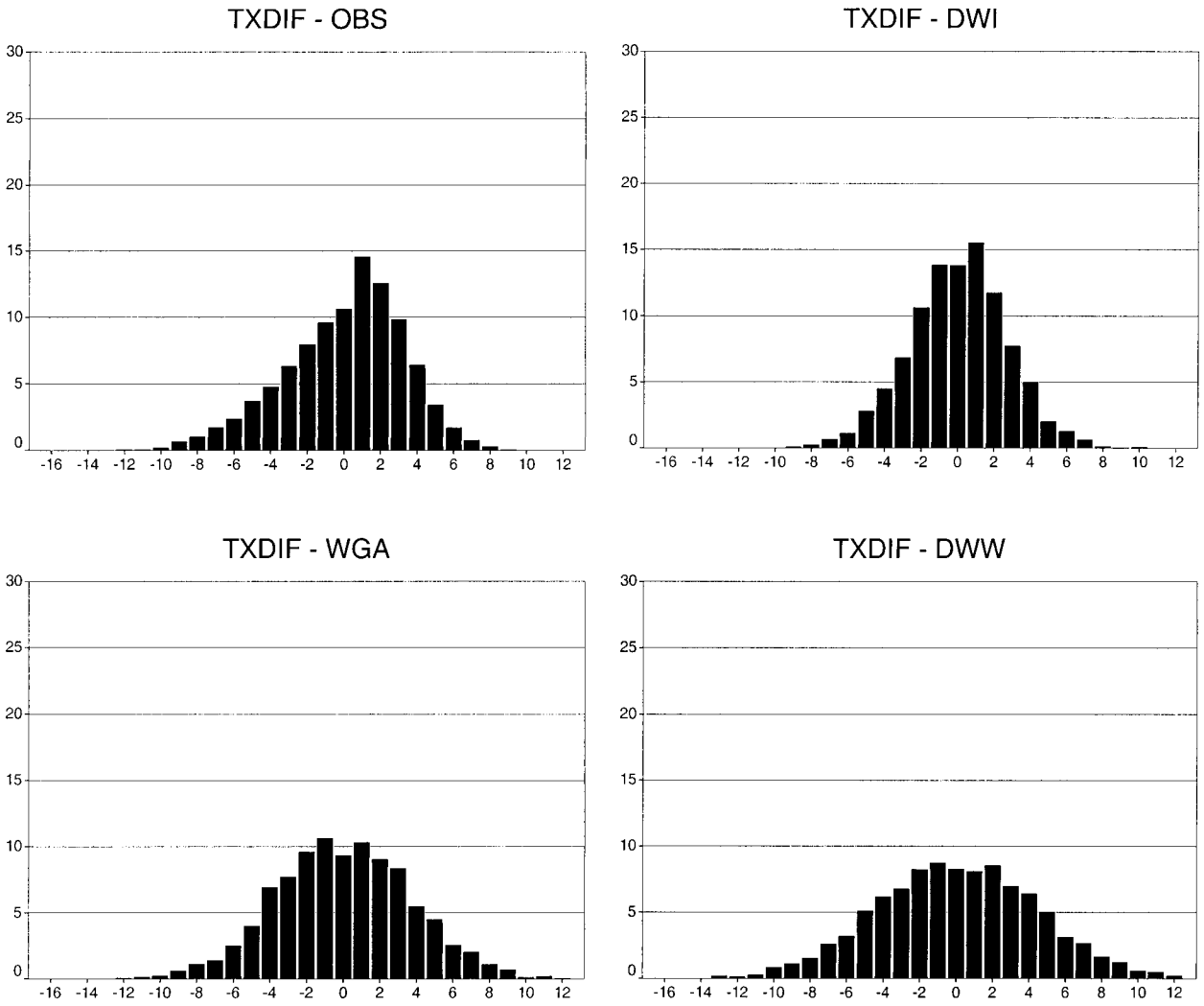


FIG. 7. Histograms of day-to-day change in maximum temperature in summer (binned into 1°C intervals) at Strážnice for selected time series. On the horizontal axis is temperature difference in degrees Celsius, on the vertical axis is frequency in percent. See Table 1 for definitions of time series.

in the two seasons are close to each other, which is unrealistic. For downscaling from both GCMs in winter, persistence is strongly underestimated at three stations (Neuchâtel, Switzerland; Kostelní Myslová and Prague, Czech Republic), whereas at the others, it is reproduced fairly well. Weather generator produces temperatures that are too variable; the incorporation of annual cycle of correlations improves the persistence in winter but further lowers it in summer. The GCMs are too persistent in summer (especially ECHAM), whereas in winter, they reproduce persistence rather successfully in general.

Standard deviations of day-to-day temperature changes (Fig. 4) are related to lag-1 autocorrelations through variance of temperature. Since by definition the variance of downscaled and stochastically generated series is identical to that observed the performance of the downscaling and weather generator is the same for both measures of day-to-day variability. A striking difference between lag-1 autocorrelations and standard deviation of day-to-day change, however, appears for the CCCM model. Although it reproduces persistence relatively well, only with a minor overestimation in summer and even a slight underestimation in winter, the variance of its day-to-day temperature change is very low. This is because of a strong underestimation of standard deviation of temperature, which in winter results from a very limited ability of CCCM to cross the freezing point during the annual cycle (Palutikof et al. 1997), in central Europe manifested by a lack of temperatures below 0°C. For example, at the station of Strážnice, Czech Republic, CCCM simulates only 17% of days in Nov–Mar with minimum temperature below −0.5°C, compared with 58% in observations (Fig. 5). Temperatures close to zero dominate in CCCM: the interval between −0.5° and +0.5°C contains almost 44% of all days, but 9.6% only

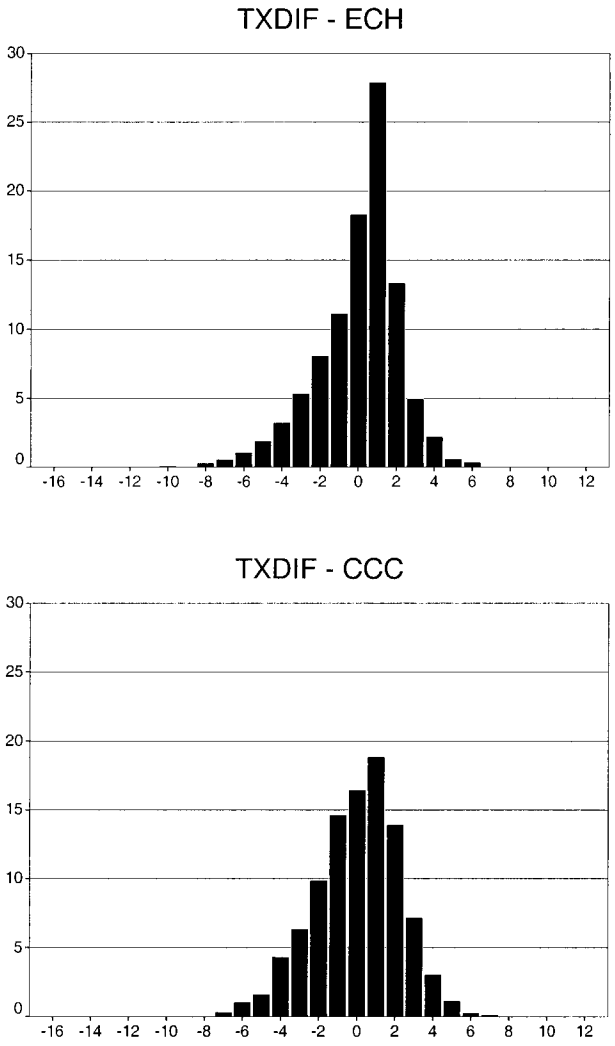


FIG. 7. (*Continued*)

in the observed. These features are common to all stations except Neuchâtel.

The distributions of observed day-to-day temperature change are negatively skewed in summer and positively skewed in winter (Fig. 6). The skewness is best captured by the CCCM model. ECHAM underestimates it, exhibiting negative values even in winter. The weather generator and downscaling with white noise yield skewness close to zero. The downscaling with inflation produces a slightly negative skewness of day-to-day temperature changes in both seasons, which is significantly different from zero at the majority of stations. The downscaling from GCMs leads to a negative skewness in summer whereas in winter, the reproduction of this characteristic is inconsistent among stations.

The negative skewness of observed day-to-day maximum temperature change in summer is illustrated in the histogram for the Strážnice station in Fig. 7: the most frequent day-to-day change is a slight warming, not a zero change, and strong coolings occur more often than strong warmings. A good reproduction of these features by the GCMs is apparent, as well as the symmetry of the downscaled and stochastically generated distributions. Temperature changes in GCMs are confined to a narrower range of values, the number of large changes being strongly underestimated. An opposite drawback appears for the white noise downscaling (DWW) series: it yields unrealistically large day-to-day temperature changes.

The histograms of day-to-day winter minimum temperature change at Strážnice are shown in Fig. 8. In contrast to summer, slight coolings dominate in the observed distribution and strong warmings are more frequent than strong coolings, which results in a positive skewness. The symmetry of day-to-day temperature changes for the downscaled and stochastically generated series is apparent as well as an exaggerated range of values for the white noise downscaling. The narrowness of the GCM-simulated distributions becomes extreme for CCCM where more than half of the day-to-day changes are within 0.5°C. For the ECHAM minimum temperatures, slight warmings are more frequent than slight coolings, which is exactly opposite to what is observed.

4. Analysis of prolonged extreme events

Many sectors of socioeconomic activities are particularly vulnerable to long-lasting extreme events, such as periods of persisting hot weather or strong frosts, which impose stress on plants, animals, as well as humans. It is therefore important to know with what degree of accuracy such events are reproduced. In this section, we introduce and analyze heat waves (HWs) and cold waves (CWs), together referred to as prolonged extreme temperature events (PETEs).

a. Definition

There are several possible ways of defining heat and cold waves (e.g., Macchiato et al. 1993; Karl and Knight 1997; Domonkos 1998). Here we adopt the definition by Huth et al. (2000): the heat wave is the longest continuous period (i) during which the maximum temperature reached at least T_1 in at least three days, (ii) whose mean maximum temperature was at least T_1, and (iii) during which the maximum temperature did not drop below T_2. The definition follows the general perception of what a HW is: it allows two periods of hot days separated by a slight drop of temperature to compose one heat wave but two periods of hot days separated by a pronounced temperature drop below T_2 (e.g., due to a cold front passage) are treated as separate heat waves. The threshold temperatures T_1 and T_2 are set to 30° and 25°C, respectively, in accordance with climatological practice in the Czech Republic. The minimum duration of a HW required by the definition is three days.

The definition of a cold wave in minimum temper-

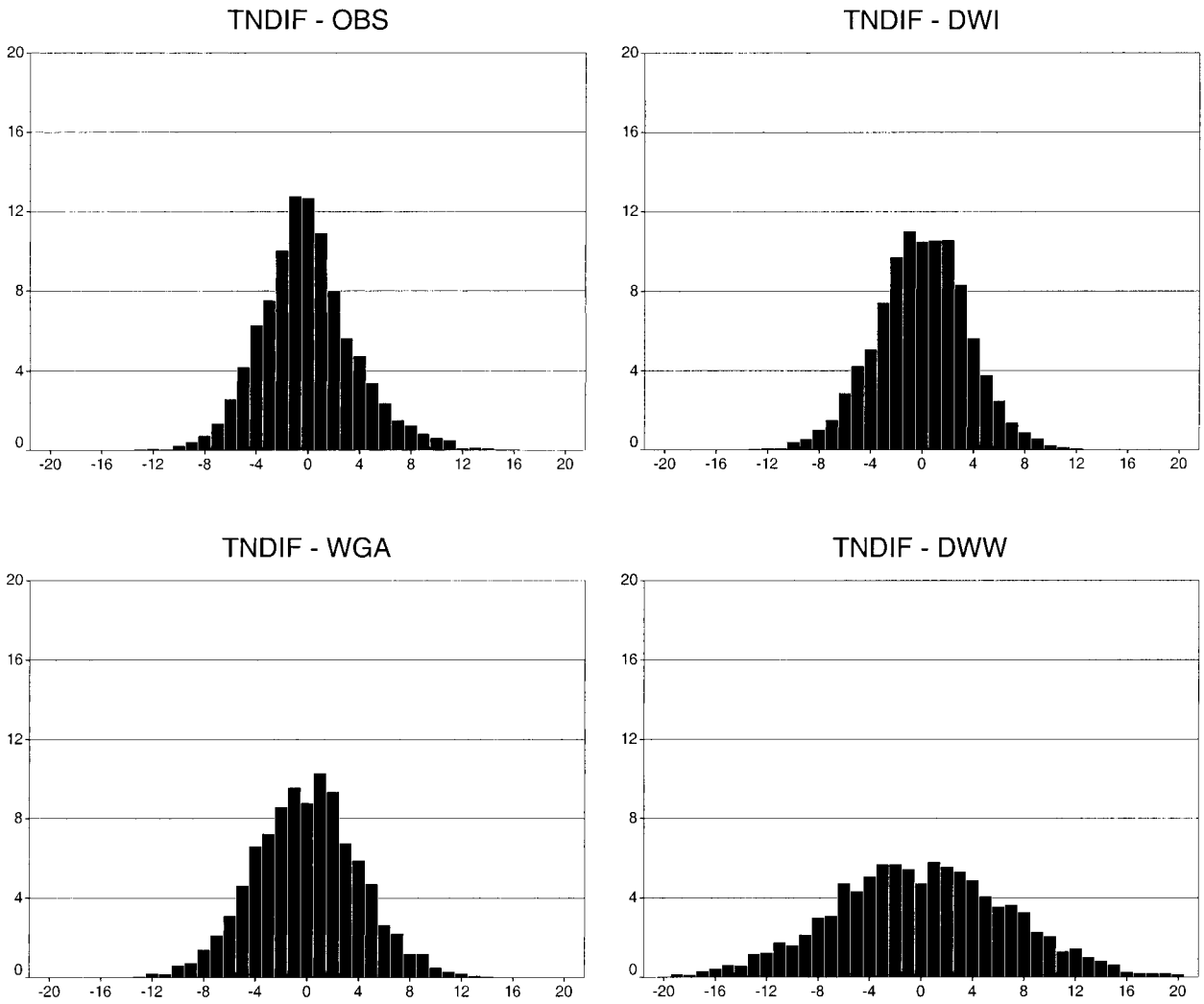


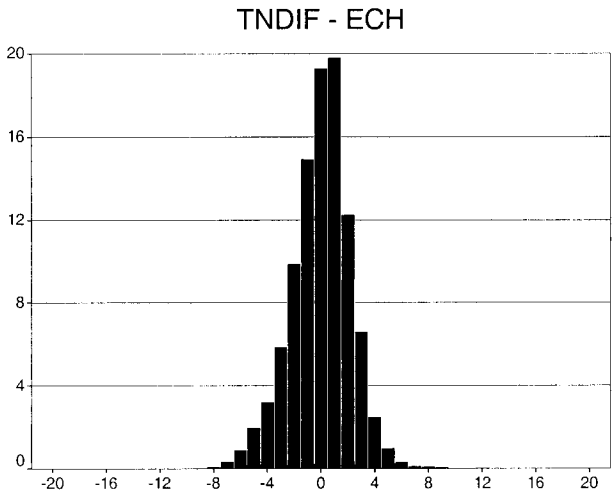
FIG. 8. Same as in Fig. 7 except for minimum temperature in winter. Note a different vertical scale for the CCC series.

ature series is analogous, only with reversed inequalities. The threshold temperatures are selected so that their percentiles in the empirical probability distribution of minimum temperatures correspond approximately to the percentiles of the HW thresholds in the maximum temperature distributions. This criterion led us to setting $T_1 = -12$°C and $T_2 = -5$°C. The probability of exceeding (dropping below) threshold temperatures T_1 and T_2 in the observed maximum (minimum) temperature distribution, averaged over all the examined stations, is 4.2% (4.9%) and 24.7% (21.2%), respectively. The days with temperatures exceeding (dropping below) threshold T_1 in summer (winter) are hereafter referred to as T_1 days. To account for the GCMs' temperature bias and allow a fair comparison with observations and other methods, the ECHAM and CCCM temperature distributions were adjusted to have the observed mean and standard deviation. Since CCCM simulates winter temperatures unrealistically, it was omitted from the analysis of CWs. It is important to note that no PETEs occur outside the analyzed periods at any station in any dataset, that is, all HWs occur in May–September and all CWs in November–March.

b. Frequency

The mean annual frequency of observed PETEs (Fig. 9) reflects the climatic conditions of the individual sites. The HWs are rare events at Hamburg, which is the northernmost station under strong maritime influence, and at the most elevated station, Kostelní Myslová, whereas they occur most frequently at the most continental site, Strážnice, situated in the lowlands. The CWs are least frequent at Neuchâtel, the southernmost station, and most frequent (occurring more than once a year on average) at the three Czech stations.

Both GCMs are quite successful in simulating the frequency of PETEs, but remember that their good per-



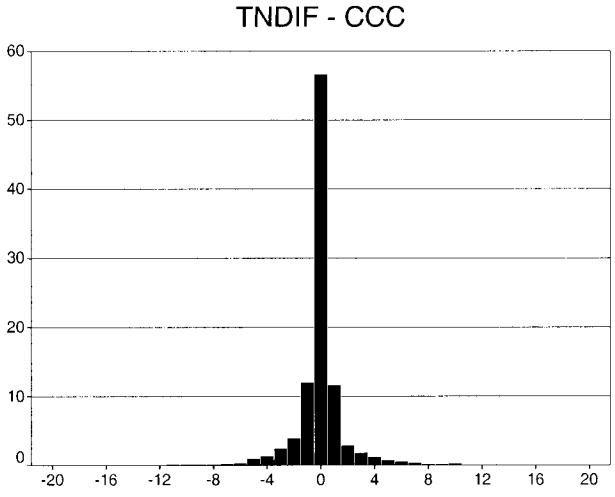
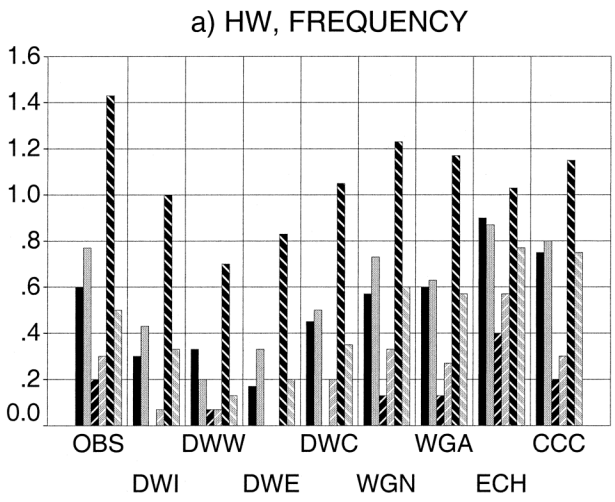


FIG. 8. (*Continued*)



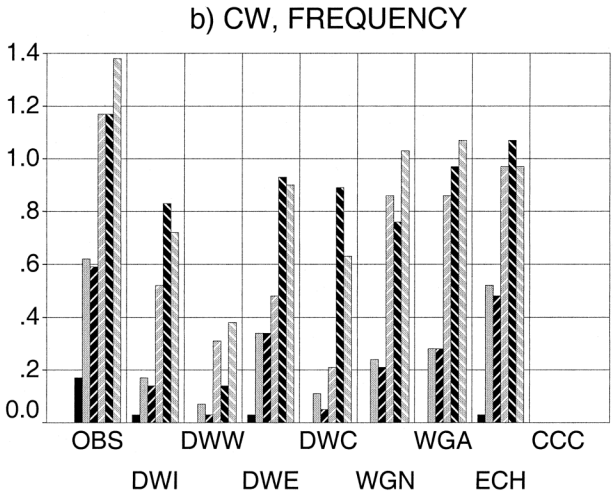


FIG. 9. Mean annual frequency of PETEs: (a) heat waves in summer and (b) cold waves in winter. Legends as in Fig. 3; in (b) the CCC series are omitted.

formance is achieved due to the adjustment to the observed mean and standard deviation. The unadjusted time series, taken directly from the models, would yield different numbers of PETEs because of their misestimated mean and variance. All the downscaling methods underestimate the frequency of PETEs. The underestimation is most severe for the DWW: its time series manifest an extremely high interdiurnal variability, which prevents them from staying in a period of high or low temperatures for long enough time. The weather generator is perhaps the best method in simulating the frequency of HWs, whereas it considerably underestimates the frequency of CWs (although not as seriously as the downscaling does).

Observed HWs at Hamburg and Kostelní Myslová, and CWs at Neuchâtel occur less than once in three years on average. Other characteristics of HWs and CWs cannot be determined reliably there, and we decided to exclude these stations from further analysis.

c. Other characteristics

The mean duration of a single PETE is displayed in Fig. 10. Observed CWs are longer than HWs at all stations (including those omitted because of the scarcity of events), which is reproduced by the downscaled series only. Looking at the two seasons separately, a satisfactory correspondence with the observed duration in summer is achieved by CCCM, the downscaling with inflation and from both GCMs, and the weather generator. The duration of HWs in ECHAM is highly overestimated. In winter, ECHAM performs best; the downscaling with inflation tends to produce CWs too long, whereas the other methods mostly underestimate their duration. The white noise addition makes both HWs and CWs too short, which results from extremely large day-to-day temperature variations.

The severity of a PETE is characterized by the highest

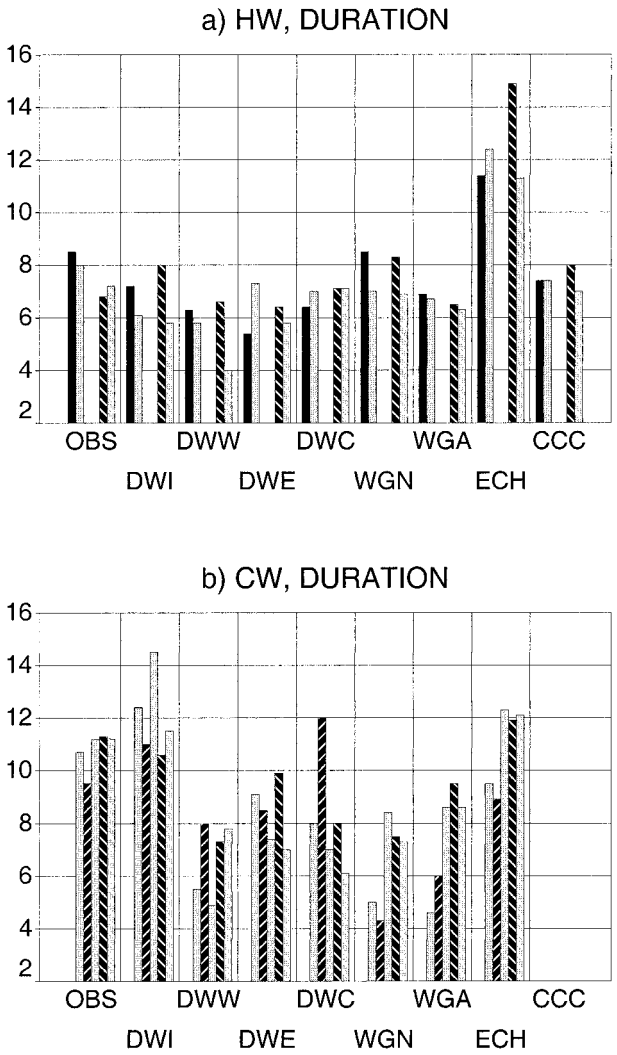


FIG. 10. Mean duration (in days) of (a) heat waves and (b) cold waves. The stations with a low incidence of PETEs are not shown. Legends as in Fig. 3; in (b) the CCC series are omitted.

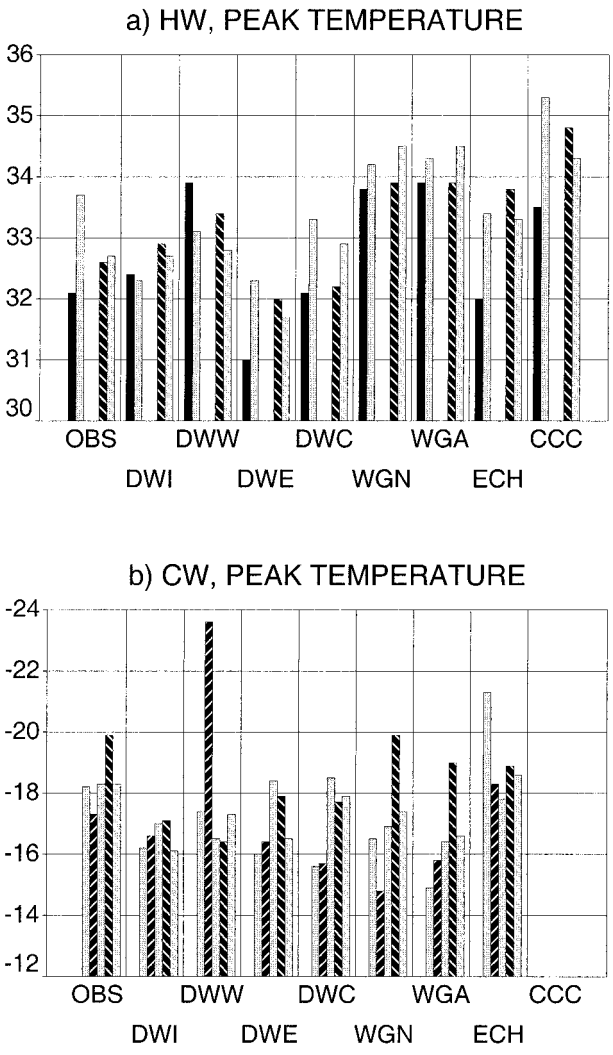


FIG. 11. Mean peak temperature (in °C) for (a) heat waves and (b) cold waves. Otherwise as in Fig. 10.

(lowest) temperature recorded during the event (Fig. 11). HWs are simulated too hot by CCCM and the weather generator; ECHAM and the downscaling methods (except for DWE with the peak too cool) show no bias. All the methods except ECHAM tend to underestimate the severity of cold waves. (There are only three CWs in the DWW series at Hamburg, so the peak of almost −24°C is not representative.)

A typical observed HW begins around 20 July (Fig. 12) shortly before the annual temperature cycle attains its maximum. The HWs are best positioned in time by the weather generator. The CCCM and ECHAM GCMs shift the start of HWs by about one and two weeks later, respectively. The downscaling from observed data (DWI and DWW) places HWs also a bit late. Observed CWs typically start between 10 and 15 January. ECHAM simulates them to occur a few days earlier except at Hamburg, whereas all the downscaling methods shift the typical date of their start at the Czech stations toward the end of January and beginning of February.

Another indicator of the validity of methods is the inclusion of T_1 days into PETEs, defined as the ratio of the number of T_1 days included in PETEs to the total number of T_1 days. Figure 13 shows that relatively more thermally extreme days occur outside HWs than outside CWs. In ECHAM in summer, too few T_1 days are isolated, that is, not involved in a HW. All other methods tend to underestimate more or less the inclusion of T_1 days in HWs. The inclusion of T_1 days in CWs is best approximated by ECHAM, still acceptable (though underestimated) in the DWI series, and strongly underestimated by the other methods. The downscaling with white noise yields the worst results, with a majority of T_1 days isolated in both seasons, which is a consequence of the high interdiurnal variability. Worth noting is also the effect of the annual cycle of lag-0 and lag-1 cor-

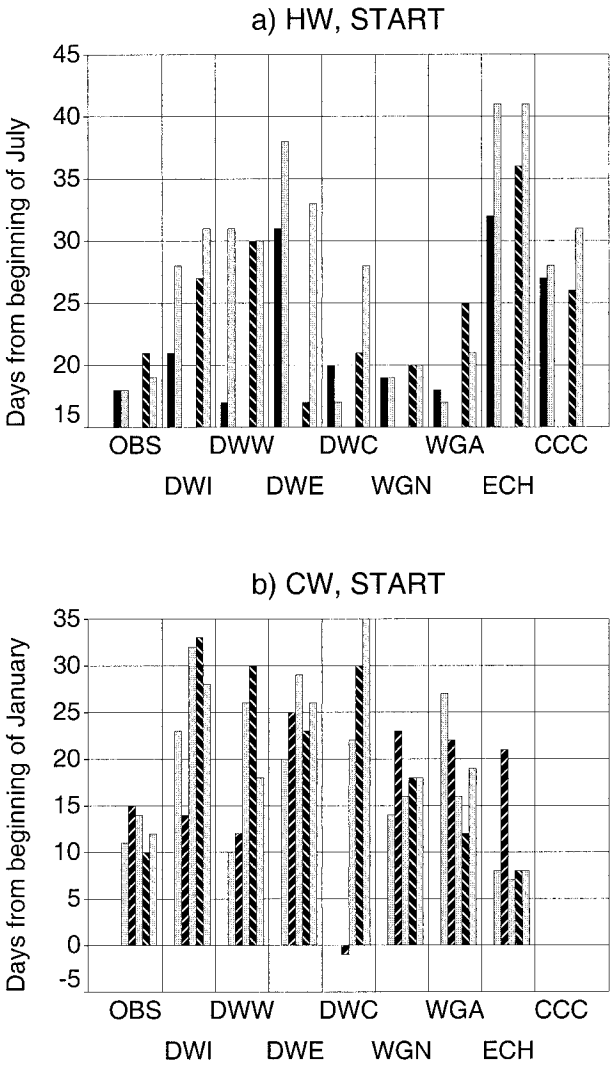


FIG. 12. Mean date of start of (a) heat waves and (b) cold waves. Numbers on the ordinate indicate the number of days from the beginning of Jul and Jan, respectively: e.g., for HWs 20 means 20 Jul. Otherwise as in Fig. 10.

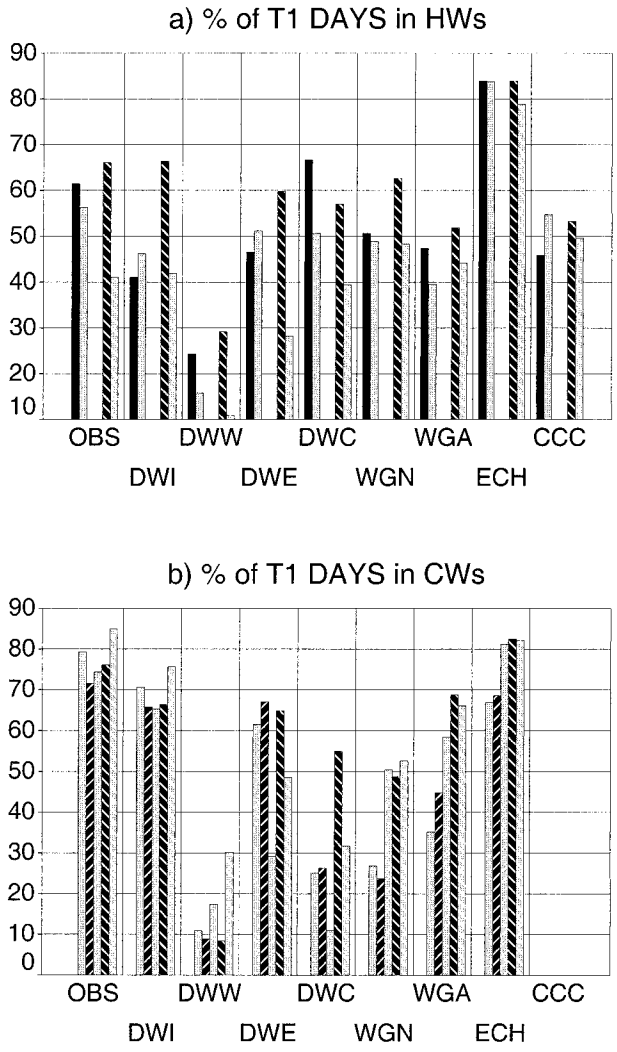


FIG. 13. Inclusion of T_1 days into (a) HWs and (b) CWs in percent. For definition see the text. Otherwise as in Fig. 10.

relations in the weather generator: the inclusion of T_1 days into PETEs is higher (that is more realistic) in winter for WGA but in summer for WGN, although in the former case it is still far from being acceptable.

5. Discussion

In this section we identify the sources of behavior of temperature and PETE characteristics in the simulated time series, and in particular, the reasons for their correspondence with or dissimilarity from the observations.

a. Presence or absence of physics

The degree of agreement of simulated temperature and PETE characteristics with observations depends on many factors. The first to discuss is whether the methods involve physical processes and what this implies for the time structure of temperature series.

The two physical processes, most important for the temperature and PETE statistics, are the surface radiation balance and atmospheric fronts. If a situation with a positive radiation balance in summer, such as an anticyclone, persists unchanged for a few days, the presence of physics ensures that the radiative warming takes place and a slight daytime temperature increase is observed. The opposite holds for a persistent situation with a negative radiation balance in winter: in the absence of a change in large-scale midtropospheric flow, the physics causes a radiative cooling, resulting in decreases in nighttime temperature. The surface radiation balance is thus the cause of the observed prevalence of slight day-to-day warmings (coolings) in summer (winter). In European midlatitudes, the passages of strong cold atmospheric fronts (the strength of a front meaning the

TABLE 2. Lag-1 autocorrelation of predictors that most frequently enter the regression equations of winter minimum temperature, as observed (OBS) and simulated by the ECHAM and CCCM models. The predictors are identified by the variable, *Z* standing for 500-hPa heights and *T* for 1000-/500-hPa thickness; and lat and long of the grid point.

Predictor	OBS	ECHAM	CCCM
Z 65.6°N, 28.1°E	0.852	0.845	0.878
Z 54.4°N, 5.6°E	0.805	0.826	0.812
Z 37.7°N, 16.9°W	0.815	0.817	0.761
Z 37.7°N, 5.6°E	0.813	0.825	0.806
T 54.4°N, 5.6°W	0.601	0.538	0.557
T 48.8°N, 22.5°E	0.771	0.757	0.621
T 32.1°N, 16.9°W	0.764	0.845	0.713

TABLE 3. The share of observed variance of minimum temperature in winter at six stations explained by the observed predictors (OBS) and predictors simulated by the ECHAM and CCCM GCMs.

	OBS	ECHAM	CCCM
NEU	0.71	0.78	0.81
WUR	0.67	0.64	0.59
HAM	0.72	0.69	0.62
KOS	0.67	0.91	0.83
STR	0.61	0.62	0.59
PRA	0.67	0.75	0.73

temperature contrast across it) are responsible for a majority of large coolings in summer and those of warm fronts for large warmings in winter; in contrast, strong warm fronts in summer and strong cold fronts in winter are rare events. The radiation balance and passages of atmospheric fronts are thus the main sources of the skewness of the day-to-day temperature change distributions. The radiation balance appears to govern also the frequency and duration of PETEs: the radiative warming (cooling) helps temperatures to reside above (below) the thresholds for a longer time, thereby supporting the formation of PETEs as well as making them last longer.

Of the features expected due to the presence of physics, we observe in both GCMs for example negatively skewed day-to-day temperature changes in summer and PETEs occurring with approximately correct frequencies. Some expected features appear in one GCM only because of model-specific deficiencies in the treatment of physical processes in GCMs, which are discussed in the next section. The absence of explicit physical information in downscaled and stochastically generated series results in near-zero skewness of day-to-day temperature changes, a shorter duration of HWs (except for the weather generator) and CWs (except for the DWI series), and an underestimated frequency of PETEs (except for HWs in the weather generator).

b. Inadequacies of physics in GCMs

Both the GCMs in both seasons produce too low an interdiurnal variability, that is, a narrower range of day-to-day temperature change values. As the lag-1 autocorrelations and the inclusion of T_1 days into PETEs suggest, the reasons for this differ between the two models: the day-to-day temperature change is constrained by an excessive persistence in ECHAM, and by too narrow a range of temperature values in CCCM. The more negative values of day-to-day temperature skewness in ECHAM relative to the observed in both seasons, which are manifested by the excess of slight warmings over slight coolings even in winter, suggest that the problem is most probably in ECHAM's radiative heating that is too strong to be realistically counteracted by other physical processes. In CCCM, a severe distortion of winter minimum temperatures with too few freezing days, too many days with temperatures around zero, and extremely low day-to-day temperature changes reminds one of conditions over a water surface rather than land. The reason is in the land surface scheme, which requires all soil water to freeze/thaw before the ground temperature is allowed to cross the freezing point (Palutikof et al. 1997; Laprise et al. 1998). This deficiency is also manifested in unrealistically high 20-yr return values of minimum temperature over central and western Europe (between −10° and 0°C; Zwiers and Kharin 1998). Recent analyses indicate that the problems reported here persist in the new coupled version of the model, CGCM1 (Kharin and Zwiers 2000).

Another inadequacy, the shift of a HW occurrence in ECHAM of about 2 weeks later, has already been noticed by Huth et al. (2000). It stems from the shift in the annual temperature maximum from July to August, which appears to be a common feature of many GCMs over continents (Mao and Robock 1998). Smaller shifts in the HW beginnings in CCCM and in the CW beginnings in ECHAM have an analogous cause in the shift of the relevant extremes of the annual temperature cycle.

c. Downscaling

The linearity of the link between large-scale upper-air fields and local temperature in downscaling results in fact in a direct transfer of some statistical properties of the upper-air variables to the surface ones. For this reason, the temperature persistence, which manifests almost no seasonality in the series downscaled with inflation, reproduces the negligible seasonality of lag-1 autocorrelations of 500-hPa heights (Gutzler and Mo 1983). The direct transfer of properties from upper-air fields to surface temperature is also likely responsible for a tendency of PETEs to occur too late in the series downscaled from observations. Whereas the observed temperatures in central Europe attain their annual minimum at about half of January and maximum at the end of July (Nemešová and Kalvová 1997), the annual extremes of 500-hPa heights over most of central and western Europe, from where the most important predictors for the regression equations are selected, tend to occur

later, in the beginning of February and August (Volmer et al. 1984).

Another issue to discuss in connection with downscaling is the failure of the downscaling from both GCMs in reproducing interdiurnal temperature variability in winter at Neuchâtel, Kostelní Myslová, and Prague where the DWE and DWC minimum temperatures have a very low lag-1 autocorrelation and an overestimated standard deviation of the day-to-day temperature change. The error is not regionalized: Strážnice behaves in a different way than the two close stations (Kostelní Myslová and Prague) but similarly to two distant stations (Hamburg and Würzburg, Germany). The source of the error is not in the procedure of enhancing the variance to fit the observed one, since it does not affect persistence; and it is not in the selection of predictors either, since there is no apparent difference between the stations affected and those unaffected by the error in the predictors selected and in their importance for the regression model. Moreover, the day-to-day variability of 500-hPa heights and 1000-/500-hPa thickness is reproduced quite accurately by both GCMs (Table 2), so the error does not stem from an incorrect reproduction of the time structure of predictors. The error appears to be connected with the degree to which the observed temperature variance is explained by the simulated predictors, that is, with the factor by which the downscaled series are multiplied to fit the observed variance. Table 3 (third and fourth columns) shows that the better the simulated circulation approximates the observed variance, the worse the time structure of the downscaled series. Two facts are worth noting in this context: first, the sites where a large share of observed temperature variance is explained by the *simulated* predictors are the same for both GCMs; and second, the share of observed variance explained by the *observed* predictors is distributed among stations in a different way (see the second column in Table 3), varying from site to site much less widely. The error, therefore, appears to be concealed in the transfer of information from simulated large-scale fields, which, for some unclear reason, fails at some of the stations but succeeds at the others. The problem may be in the inconsistency that the *observed* regression equations are applied to the *simulated* large-scale fields.

Let us turn to the effect of how the unexplained variance in downscaling is treated. The white noise addition leads to an unrealistically low persistence. The error is larger for winter minimum temperatures than for summer maxima because large-scale predictors explain less temperature variance in winter, and the added noise is therefore stronger. The excessive variability of the DWW series results in the inability of temperatures to reside above (below) the thresholds for a long enough time. As a consequence, the number of PETEs is extremely low in the DWW series, and if they occur they are too short. Moreover, only a very small portion of T_1 days are chained to form PETEs, that is, the majority of thermally extreme days are isolated. The white noise addition is not the only way to enhance the variance, indeed, but any more realistic treatment of the missing variance than the inflation is, which would consist in adding a stochastic process independent of large-scale forcing, necessarily implies an increase in day-to-day variability. Our analysis shows that this would impair the simulation of PETEs relative to the inflation approach, which itself is not much successful in this respect. Although the physical grounds of the variance inflation are flawed, we recommend the inflation as the least biased approach to be used in downscaling studies where the time structure and prolonged extreme events are important.

d. Weather generator

The lag-1 autocorrelations of stochastically generated series are too low in both seasons. This seems to be in contradiction with the fact that autocorrelations are among the generator's parameters and should, therefore, have been replicated accurately. The explanation is as follows. In calculating the parameters of the generator, the correlations and autocorrelations of both TMAX and TMIN are derived from the series that were normalized separately for wet and dry days. During the generation process, the generator produces normalized anomalies first, which are then turned into a temperature scale by multiplying by the standard deviations and adding to the means, both conditioned on the precipitation occurrence. The day-to-day changes of generated temperatures are thus a result of superposition of the first-order autoregressive model of normalized temperature and the first-order Markov chain of precipitation occurrence, which implies a suppression of lag-1 autocorrelations relative to the original autoregressive model. Moreover, the first-order Markov chain underestimates the persistence of the precipitation occurrence series (Dubrovský 1997). Both effects lead to the enhancement of the day-to-day variability and underestimation of persistence in the generated (both WGA and WGN) temperature series.

The difference in performance between the WGN and WGA series can be understood from the annual cycles of the lag-1 correlations of maximum and minimum temperatures (Fig. 1). The correlations attain their maxima in winter and minima in summer. This makes the day-to-day temperature variability in winter lower and persistence higher in the series where the annual cycle of lag-1 correlations is implemented (WGA), relative to that using the annual mean of lag-1 correlations (WGN); the opposite holds in summer, exactly as shown in Figs. 3 and 4. Since the temperature persistence is underestimated by the weather generator in general and the annual cycle of correlations does not influence the source of the underestimation, its inclusion acts to suppress this bias in winter, but to enhance it in summer. In accordance with a general effect of persistence on PETEs, the inclusion of annual cycle of lag-1 correla-

tions makes HWs shorter, they reach higher temperatures, and a lower fraction of T_1 days is included in them. On the contrary, CWs in the WGA series are longer and more frequent relative to WGN, and higher fraction of T_1 days is included in them. We can state that the effect of inclusion of annual cycle of lag-1 autocorrelations into the weather generator leads to the improvement of several temperature and PETE statistics in winter, but, rather paradoxically, to the deterioration of the same characteristics in summer.

6. Conclusions

In this paper, three approaches to constructing site-specific daily temperature series, namely, the direct GCM output, statistical downscaling, and the weather generator, were examined for their ability to reproduce the temporal structure of the series and prolonged extreme temperature events (PETEs). The simulated series have been compared against observations at six sites in central Europe. We have arrived at the following major conclusions.

- None of the methods of constructing site-specific daily temperature series appears to be able to reproduce the majority of statistics of day-to-day temperature change and PETEs correctly. Nevertheless, the ECHAM GCM output adjusted for the observed mean and variance approaches the demands most closely in winter, whereas in summer, the weather generator without the annual cycle of correlations appears to perform best.
- The causes of an incorrect reproduction of the examined temperature characteristics include (i) the absence of physical processes, particularly surface radiation balance and atmospheric fronts, in the downscaling and weather generator approach; (ii) inadequacies in treatment of some physical processes in the GCMs; (iii) the linearity of downscaling, imposing a direct transfer of properties of large-scale fields used as predictors to the surface temperature series; and (iv) a conjunction of the autoregressive model and Markov chain in the generation process of maximum and minimum temperature by the weather generator together with an underestimation of precipitation persistence by the first-order Markov chain implemented in it.
- The white noise addition, which is an alternative approach to the variance inflation in adjusting the variance in downscaled series, leads to temperature series that are too variable. It is therefore unsuitable if one is concerned with time structure and prolonged extreme events. Although the inflation is a physically questionable concept, it yields temperature time series much closer to reality.
- The inclusion of the annual cycle of lag-0 and lag-1 correlations into the generator does not lead to an overall improvement in the simulation of day-to-day temperature variability. The reason is a general underestimation of persistence by the weather generator: whereas in winter the inclusion of annual cycle of autocorrelations enhances the persistence, thereby improving its reproduction, in summer the inclusion of annual cycle results in a further suppression of persistence, that is, in a deterioration of its reproduction.

Acknowledgments. The authors are grateful to M. Beniston, University of Fribourg, Switzerland; A. Kästner, German Weather Service, Offenbach, Germany; H. Österle, Potsdam Institute of Climate Impact Studies, Potsdam, Germany; R. Schweitzer, University of Colorado, Boulder, Colorado; F. Zwiers, Canadian Centre for Climate Modelling and Analysis, Victoria, Canada; and the staff of the Czech Hydrometeorological Institute in Prague, Czech Republic, for preparing and providing the datasets. The study was supported by the Grant Agency of the Czech Republic under Projects 205/96/1670, 205/97/P159, and 205/99/1561.

REFERENCES

Buishand, T. A., and J. J. Beersma, 1993: Jackknife tests for differences in autocorrelation between climate time series. *J. Climate*, **6,** 2490–2495.

DKRZ, 1993: The ECHAM3 Atmospheric General Circulation Model. Deutsches Klimarechenzentrum Rep. 6, Hamburg, Germany, 184 pp.

Domonkos, P., 1998: Statistical characteristics of extreme temperature anomaly groups in Hungary. *Theor. Appl. Climatol.*, **59,** 165–179.

Dubrovský, M., 1997: Creating daily weather series with use of the weather generator. *Environmetrics*, **8,** 409–424.

——, Z. Žalud, and M. Št'astná, 2000: Sensitivity of CERES-Maize yields to statistical structure of daily weather series. *Climatic Change*, **46,** 447–472.

Giorgi, F., and L. O. Mearns, 1991: Approaches to the simulation of regional climate change: A review. *Rev. Geophys.*, **29,** 191–216.

Gutzler, D. S., and K. C. Mo, 1983: Autocorrelation of Northern Hemisphere geopotential heights. *Mon. Wea. Rev.*, **111,** 155–164.

Hay, L. E., R. L. Wilby, and G. H. Leavesley, 2000: A comparison of delta change and downscaling GCM scenarios for three mountainous basins in the United States. *J. Amer. Water Resour. Assoc.*, **36,** 387–397.

Hayhoe, H. N., 2000: Improvements of stochastic weather data generators for diverse climates. *Climate Res.*, **14,** 75–87.

Hulme, M., Z. C. Zhao, and T. Jiang, 1994: Recent and future climate change in East Asia. *Int. J. Climatol.*, **14,** 637–658.

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

——, and J. Kyselý, 2000: Constructing site-specific climate change scenarios on a monthly scale using statistical downscaling. *Theor. Appl. Climatol.*, **66,** 13–27.

——, ——, and L. Pokorná, 2000: A GCM simulation of heat waves, dry spells, and their relationships to circulation. *Climatic Change*, **46,** 29–60.

Kalnay, E., and Coauthors, 1996: The NCEP/NCAR 40-Year Reanalysis Project. *Bull. Amer. Meteor. Soc.*, **77,** 437–471.

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

——, A. Raidl, A. Trojáková, M. Žák, and I. Nemešová, 2000: Canadian Climate Model—Air temperature over the European re-

gion and in the Czech Republic (in Czech). *Meteor. Zprávy*, **53,** 137–145.

Karl, T. R., and R. W. Knight, 1997: The 1995 Chicago heat wave: How likely is a recurrence? *Bull. Amer. Meteor. Soc.*, **78,** 1107–1119.

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

Kattenberg, A., and Coauthors, 1996: Climate models—Projections of future climate. *Climate Change 1995: The Science of Climate Change*, J. T. Houghton et al., Eds., Cambridge University Press, 285–357.

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

Kidson, J. W., and C. S. Thompson, 1998: A comparison of statistical and model-based downscaling techniques for estimating local climate variations. *J. Climate*, **11,** 735–753.

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

Macchiato, M., C. Serio, V. Lapenna, and L. La Rotonda, 1993: Parametric time series analysis of cold and hot spells in daily temperature: An application in southern Italy. *J. Appl. Meteor.*, **32,** 1270–1281.

Mao, J., and A. Robock, 1998: Surface air temperature simulations by AMIP general circulation models: Volcanic and ENSO signals and systematic errors. *J. Climate*, **11,** 1538–1552.

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

Mearns, L. O., F. Giorgi, L. McDaniel, and C. Shields, 1995: Analysis of variability and diurnal range of daily temperature in a nested regional climate model: Comparison with observations and doubled CO_2 results. *Climate Dyn.*, **11,** 193–209.

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

Nemešová, I., and J. Kalvová, 1997: On the validity of ECHAM-simulated daily extreme temperatures. *Stud. Geophys. Geod.*, **41,** 396–406.

——, ——, J. Buchtele, and N. Klimperová, 1998: Comparison of GCM-simulated and observed climates for assessing hydrological impacts of climate change. *J. Hydrol. Hydromech.*, **46,** 237–263.

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

Semenov, M. A., and E. M. Barrow, 1997: Use of a stochastic weather generator in the development of climate change scenarios. *Climatic Change*, **35,** 397–414.

Skelly, W. C., and A. Henderson-Sellers, 1996: Grid box or grid point: What type of data do GCMs deliver to climate impacts researchers? *Int. J. Climatol.*, **16,** 1079–1086.

Solow, A. R., 1988: Detecting changes through time in the variance of a long-term hemispheric temperature record: An application of robust locally weighted regression. *J. Climate*, **1,** 290–296.

Trigo, R. M., and J. P. Palutikof, 1999: Simulation of daily temperatures for climate change scenarios over Portugal: A neural network model approach. *Climate Res.*, **13,** 45–59.

Volmer, J. P., M. Déqué, and D. Rousselet, 1984: EOF analysis of 500 mb geopotential: A comparison between simulation and reality. *Tellus*, **36A,** 336–347.

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

Wilby, R. L., L. E. Hay, and G. H. Leavesley, 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.

Wilks, D. S., 1992: Adapting stochastic weather generation algorithms for climate change studies. *Climatic Change*, **22,** 67–84.

Winkler, J. A., J. P. Palutikof, J. A. Andresen, and C. M. Goodess, 1997: The simulation of daily temperature time series from GCM output. Part II: Sensitivity analysis of an empirical transfer function methodology. *J. Climate*, **10,** 2514–2532.

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

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