Comparison of General Circulation Model and Observed Regional Climates: Daily and Seasonal Variability

DAVID A. PORTMAN,* WEI-CHYUNG WANG,† AND THOMAS R. KARL**

*Atmospheric and Environmental Research, Inc., Cambridge, Massachusetts
†Atmospheric Sciences Research Center, SUNY, Albany, New York
**NOAA/NESDIS/NCDC, Asheville, North Carolina

(Manuscript received 24 September 1990, in final form 5 July 1991)

ABSTRACT

Validation of general circulation model (GCM) current climate simulations is important for further GCM development and application to climate change studies. So far, studies that compare GCM output with observations have focused primarily on large-scale spatial averages of the surface climate variables. Here we discuss two approaches to compare output of individual GCM grid boxes with local station observations near the surface and in the free troposphere. The first approach, proposed by Chervin, involves the application of standard parametric statistical analysis and hypothesis testing procedures. The second approach is nonparametric in the sense that no ideal distributions are postulated a priori to ascertain significance of the difference of mean temperature or the ratio of the temperature variance between model grid boxes and local stations. Instead, station observations are first subjected to a bootstrap technique and then used to define a unique set of distributions and confidence limits for each GCM grid box.

To demonstrate the usefulness of the two approaches, we compare daily and seasonal gridbox temperatures simulated by the National Center for Atmospheric Research (NCAR) Community Climate Model (CCM1) with station temperatures at the surface, 850-mb, 500-mb, and 300-mb levels for three different areas in the United States. We find that although CCM1 gridbox temperatures are mostly cooler than station temperatures, they are equally variable. For all grid boxes, gridbox-to-station differences decrease with height and vary with time of year. We conclude that the techniques presented here can provide useful comparisons of GCM regional and local observed temperatures. Application to other variables and GCMs is also discussed.

1. Introduction

General circulation models (GCMs) are based on numerical solutions of fundamental equations governing the dynamical and physical processes of the earth's climate system. GCMs are used to understand the global climate system as well as to assess regional distributions of possible future climate changes due to the greenhouse effect (Schlesinger and Gates 1980; Hansen et al. 1988; Manabe et al. 1990; Mitchell et al. 1989; Washington and Meehl 1984, 1989). Understanding how well GCMs reproduce the regional and local climates can guide improvement and development to increase their predictive capabilities (DOE 1985; WMO 1987; CES 1989; WMO/UNEP 1990).

Of the studies performed to validate GCM output on regional scales, most have considered only spatially averaged observations (e.g., Grotch 1988; Rind et al. 1989; Hunt and Gordon 1989; Mearns et al. 1990). Since model predictions will be used in the future to study climate changes on smaller scales, it is important to consider to what extent the GCMs are capable of simulating the variability of local climates such as those observed at individual stations.

Only a few studies have attempted to directly compare output of individual GCM grid boxes with local station observations. For example, Reed (1986) and Wilson and Mitchell (1987) compared daily near-surface air temperature and precipitation produced by a 3-year integration of the U.K. Meteorological Office's five-layer GCM with station observations over western Europe. They found that local geographical and topographical conditions, which were subgrid scale and thus remained unaccounted for in their model, can complicate near-surface intercomparisons of model output with observations.

In a related study, Wigley et al. (1990) used statistical methods to explore relationships of local surface temperature and precipitation data with regionally averaged values of surface temperature, precipitation, sea level pressure, and 700-mb height for a 32-station network over the state of Oregon. They found that despite difficulties of comparing near-surface mean values, it is possible to obtain useful local information by examining the large-scale variability.

Karl et al. (1990) used multivariate statistical meth-

Corresponding author address: David A. Portman, Atmospheric and Environmental Research, Inc., 840 Memorial Dr., Cambridge, MA 02139.

ods to relate a group of free-atmosphere variables (predictors) simulated for individual grid points of the Oregon State University (OSU) two-level atmospheric GCM with surface-based variables (predictands) observed at nearby local stations. They found that GCM free-atmosphere statistics that have been standardized using the model's internal means and variances can be used to infer the observed local surface climates. Long-range weather forecasting studies have used similar methods to investigate the climate inversion problem (Klein 1985; Klein and Yang 1986; Weare and Hoeschele 1983; Klein and Bloom 1987) as well as to formulate the Model Output Statistics (MOS) and Perfect Prog (PP) techniques (Klein 1982; Glahn 1985).

Here we discuss two approaches to compare output of individual GCM grid boxes with local station observations at the surface and in the free troposphere. The first approach, which was proposed by Chervin (1980, 1981), involves the application of standard statistical analysis and hypothesis testing procedures. The second approach, which we propose here, is nonparametric in the sense that no ideal distribution types are postulated a priori to ascertain significance of the difference of mean temperature or the ratio of temperature variance between model grid boxes and local stations.

To demonstrate the usefulness of the two approaches, we compare daily and seasonal temperatures simulated by the National Center for Atmospheric Research (NCAR) Community Climate Model (CCM1) with daily and seasonal temperatures observed at individual stations. Since we expect that surface geography and topography will cause results to vary both with height and between grid boxes, we perform comparisons for the surface and free troposphere and for three separate regions in the United States. Levels and regions selected for study as well as descriptions of the model simulation and observational data are detailed in section 2. Statistical methods are described in section 3. Results are presented in sections 4 and we summarize our conclusions in section 5.

2. Data

a. GCM output

The NCAR Community Climate Model (CCM1) is an atmospheric general circulation model with 12 vertical layers and 4.5° latitude × 7.5° longitude horizontal resolution. It includes solar and terrestrial radiation, realistic geography and topography, ice and snow cover over land and oceans, interactive cloud formation, and land surface characteristics such as soil moisture. Detailed descriptions of the CCM1 are presented in Williamson et al. (1987). The data used in this study came from a 10-yr simulation that included an annual cycle (but no diurnal cycle) and interactive surface hydrology.

We chose to compare temperature data of the following three CCM1 grid boxes, each covering an area of approximately 640 × 500 km^2 in a different region of the United States:

TABLE 1. NCAR CCM1 grid boxes and observational stations selected for study.

Name	Latitude (North)	Longitude (West)	Height (m)
Northwest			
Grid point	46°40′	120°00′	941
(A) Spokane, WA	47°40′	117°25′	721
(B) Portland/Salem, OR	44°57′	123°01′	61
Central			
Grid point	37°48′	97°30′	556
(A) Topeka, KS	39°20′	95°41′	270
(B) Dodge City, KS	37°45′	100°02′	790
Southeast			
Grid point	33°20′	82°30′	116
(A) Charleston, SC	32°48′	79°58′	13
(B) Atlanta, GA	33°45′	84°23′	312

1) Northwest—coastal and mountainous portions of the states of Washington, Oregon, and Idaho. The Pacific Ocean lies to the west.

2) Central—east of the Rocky Mountains, covering parts of the states of Kansas, Missouri, and Oklahoma. This region is landlocked and relatively flat.

3) Southeast—South Carolina and parts of four other states, including portions of the Appalachian mountain range, the Atlantic coastal plain, and the Atlantic Ocean to the south and east.

Gridbox elevations and coordinates of the central grid points are listed in Table 1. Relative positions are shown in Fig. 1.

b. Observations

We also chose to use daily surface and upper air temperatures of the following first order rawinsonde stations: 1) Spokane, Washington; 2) Portland, Oregon; 3) Topeka, Kansas; 4) Dodge City, Kansas; 5) Charleston, South Carolina; and 6) Atlanta, Georgia. Thus, in each GCM grid box, there are two stations located east and west of the central grid point. Station positions and surface elevations are listed in Table 1 and the relative locations are indicated in Fig. 1. Observational data were obtained on tape from the National Climatic Data Center (NCDC). To be consistent with the CCM1 output, all data for 29 February (leap day) were omitted from the observational records.

For both CCM1 grid boxes and local stations, we computed the daily mean temperatures for the surface, 850-mb, 500-mb, and 300-mb levels. For the CCM1 grid boxes, twice-daily (12-h) temperatures were averaged to yield daily means for 10 years beginning with day 319 (1 December) of the first model year. For local stations, measurements taken at 12 UTC and 00 UTC were averaged to yield daily means for 30 years, 1957–1986. In the next section, we discuss how these daily

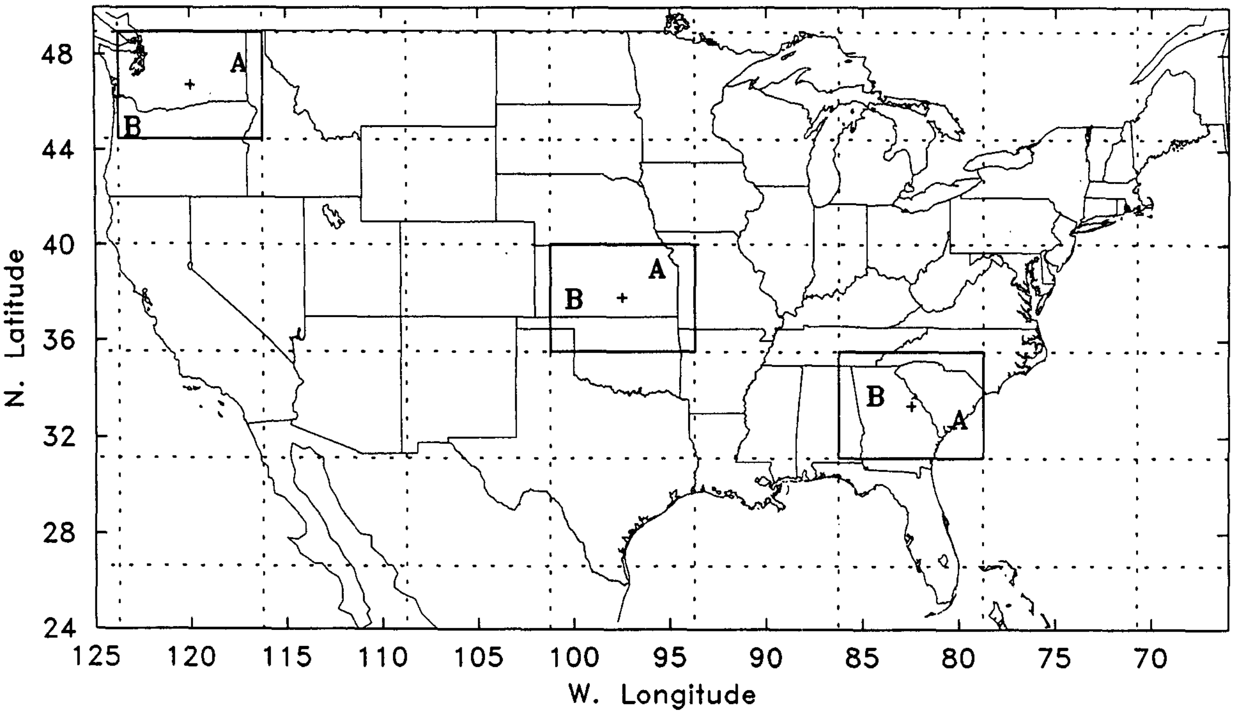

FIG. 1. United States map showing locations of NCAR CCM1 grid boxes and observational stations. In each grid box, two first-order rawinsonde stations (A and B) lie east and west of the central grid point. Latitudes, longitudes, and surface elevations of stations and grid points are summarized in Table 1.

data can be used to compare the simulated output of each CCM1 grid box with the local station observations.

3. Methodology

In the discussion that follows, we present two approaches to compare CCM1 regional gridbox output with local station observations. The first approach, proposed by Chervin (1980, 1981), uses standard parametric analysis and hypothesis testing procedures to evaluate agreement of GCM and observed temperatures. The second approach is nonparametric in the sense that no ideal statistical sample distribution is postulated.

Chervin (1980) proposed the use of certain sample test variates to study the difference of mean climate state and the ratio of climate variance between two sample populations. Since these test variates facilitate overall assessment and intercomparison of results among different grid boxes, seasons, and atmospheric levels, we have decided to use them here. To test equality of the variances, the test variate

$$r_2 = \frac{\sigma_a}{\sigma_b} \tag{1}$$

is computed, where σ_a and σ_b are the standard deviations (variances) of the time series of two stations or of one station and one GCM grid box, a and b. Test variate r_2 resembles the F statistic. If $r_2 = 1$, then the variance of a equals the variance of b. Larger (smaller) values of r_2 indicate that the variance of a may be greater (less) than the variance of b.

In the Chervin approach, the null hypothesis to be tested states that the time series of a and b exhibit equal variance. If we assume, after Chervin, that there are 9 degrees of freedom (10 data values were used to compute each σ value), then we may accept this null hypothesis provided that r_2 is within the range

$$0.56 < r_2 < 1.78. \tag{2}$$

Otherwise, we may reject the null hypothesis and assume that the two time series *do not* exhibit equal temperature variance. The limits of the range in (2) were taken from Table 1 of Chervin (1981) and define the a priori 90% statistical confidence interval for the F test.[1]

To test for equality of the means, the test variate

$$r_1 = \frac{(\mu_a - \mu_b)}{(\sigma_a^2 + \sigma_b^2)^{1/2}} \tag{3}$$

[1] Chervin based his choice of such a narrow range on the advice of Wadsworth and Brian (1960) concerning the possible erroneous judgement in a test for equality of variances that is performed in conjunction with a test for equality of the means (see Chervin 1980, p. 1905, for further elaboration on this point).

is computed, where μ_a and μ_b are the mean temperatures of the time series of two stations or of one GCM grid box and one station, a and b. Test variate r_1 resembles the Cochran and Cox approximation of Student's t-statistic for the case of unequal variances. If $r_1 = 0$, the mean temperature of a equals the mean temperature of b. Larger (smaller) r_1 values indicate that the mean temperature of a may be greater (less) than the mean temperature of b.

In the Chervin approach, the null hypothesis to be tested states that the mean temperatures of the two time series are equal. If we assume that there are 9 degrees of freedom, then we may accept this null hypothesis, provided that r_1 is in the range

$$-0.71 < r_1 < 0.71. \quad (4)$$

Otherwise we may reject the null hypothesis and conclude that the mean temperatures of the two time series are *not* equal. The limits of the range in (4) were taken from Table 2 of Chervin (1981) and define the a priori 95% statistical confidence interval for Student's t-test. Note that this acceptance region for r_1 is less narrow than the acceptance region defined above for r_2. Despite the dependence of r_1 on σ, results of the r_1 test remain valid even if the variances of the two time series are unequal.

According to the Chervin approach, if there is good agreement between model and observations, then r_2 should fit the F distribution and r_1 should fit Student's t-distribution. Due to the obvious mismatch of scale between the GCM grid boxes and the local stations, however, we should probably not expect ideal agreement. Therefore, let us consider a second comparison approach.

In each CCM1 grid box, the two stations (labeled A and B in Fig. 1) are located near opposite sides, east and west of the central grid point. Therefore, we shall assume that the difference of temperature mean or variability between the two stations represents the maximum difference of temperature mean or variability across the grid box. Then, to compare the temperature variance within the grid box, we can generate a sample r_2 population for station A over station B (and vice versa) and designate the 5–95 percentile range of the resulting r_2 distribution as the 90% confidence interval. Thus, the null hypothesis to be tested states that the ratio of temperature variance for a GCM grid box and any of the local stations is less than or equal to the maximum ratio of observed temperature variance for any two points in that grid box. If r_2 is within the acceptance region, we shall accept the null hypothesis and conclude that the model can reproduce the observed local temperature variance.

Similarly, to compare the mean temperature, for each grid box we shall generate a sample r_1 population for station A minus station B (and vice versa) and designate the 5–95 percentile range of the resulting r_1 distribution as the 90% confidence interval. Thus, the null hypothesis to be tested states that (the absolute value of) the difference of mean temperature between a GCM grid box and any of the local stations is less than or equal to (the absolute value of) the difference of observed mean temperature between any two points in that grid box.[2] If r_1 is within the acceptance region, we shall accept the null hypothesis and conclude that the model can reproduce the observed local mean temperature.

Since we had observations for only 30 years, we decided to use a resampling procedure known as the bootstrap (Efron 1982; Diaconis and Efron 1983) to increase the sample size of the data. First, 10 numbers, corresponding to the 10 years between 1957 and 1986, were selected at random with replacement permitted. The 365 consecutive daily mean temperatures of the 10 years were then combined, one year at a time in order of selection. By repeating this procedure, we were able to generate 100 10-year-long time series of the daily mean temperature for each station and atmospheric level. For each time series we calculated interannual statistics (10-yr mean μ and standard deviation σ) for four seasons (winter, spring, summer, and autumn) and for 365 days of the year.

Since temperature data are usually serially correlated, resampling procedures such as the bootstrap must be used with caution (Solow 1985; Zwiers 1990). Original time series should be examined and, if possible, adjusted to account for long-term trends. However, the long-term daily and seasonal statistics used in this study were calculated on an interannual basis, so serial correlation effects would be small. We assumed, therefore, that the 10-yr mean and variance statistics produced using the bootstrap represent normally distributed sample populations.

We did not apply the bootstrap to construct temperature time series for the CCM1 grid boxes. This particular model simulation was unforced (except for the annual cycle) and had reached equilibrium. Furthermore, the available model output covers just 10 years. Therefore, we just assumed that the gridbox statistics based on this output represent normally distributed sample populations.

For each season and atmospheric level, we computed 100 r_1 and r_2 values to compare each GCM grid box with the two local stations. Seasonal statistics permit comparisons of GCM output with observations on a seasonal basis and daily statistics permit examination of the annual and seasonal fluctuations of these comparisons in greater detail. Thus, the median (50th per-

[2] To retain as much information as possible concerning the differences between GCM and observed temperature mean and variability (e.g., grid boxes warmer or cooler than stations), the results were *not* expressed in terms of the absolute value.

centile) of r_1 and r_2 values computed for each day of the year were retained to construct daily series for the entire year, from 1 December to 30 November. Standard spectral techniques (Panofsky and Brier 1968) were then applied to remove periodicities of less than 90 days (3 months) from each daily series. Results are presented in the next section.

4. Results

To determine results of each comparison, we consider both magnitude and range of the statistics for each season and atmospheric level. If most of the r_1 (or r_2) values are within the confidence intervals, the result of the comparison is negative and we conclude that there is probably no difference of mean (or variance) between grid box and station. If, however, most of the r_1 (or r_2) values are outside of the confidence intervals, the result is positive and we conclude that a significant difference of mean (or variance) probably exists.

a. Seasonal temperature

Seasonal results of test variate r_2 are presented in Figs. 2, 3, and 4, for the northwest, central, and southeast grid boxes, respectively. Box plots depict the 10th,

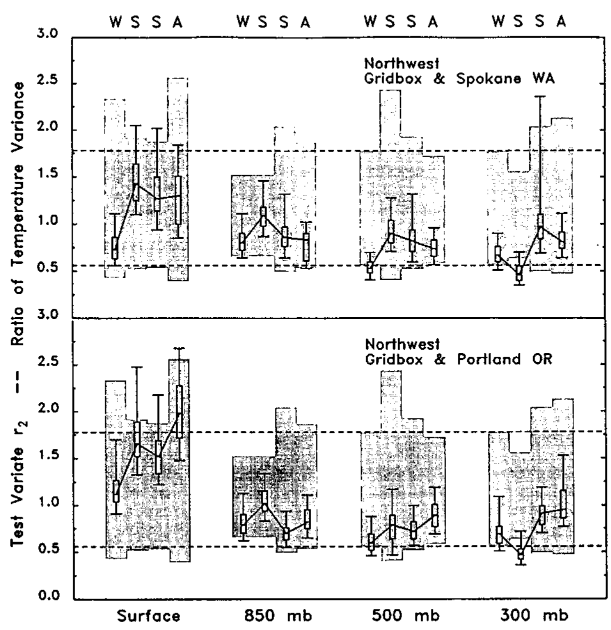

FIG. 2. Seasonal results of test variate r_2, ratio of temperature variance, for northwest grid box compared with Spokane, WA (top) and Portland, OR (bottom). Boxplots depict the 10th, 30th, 50th (median), 70th, and 90th percentiles of results for winter (W), spring (S), summer (S), and autumn (A) at the surface, 850-mb, 500-mb, and 300-mb levels. Dashed lines indicate a priori 90% confidence interval boundaries of the Chervin comparison approach and shaded background bars depict a priori 90% confidence intervals of the nonparametric comparison approach.

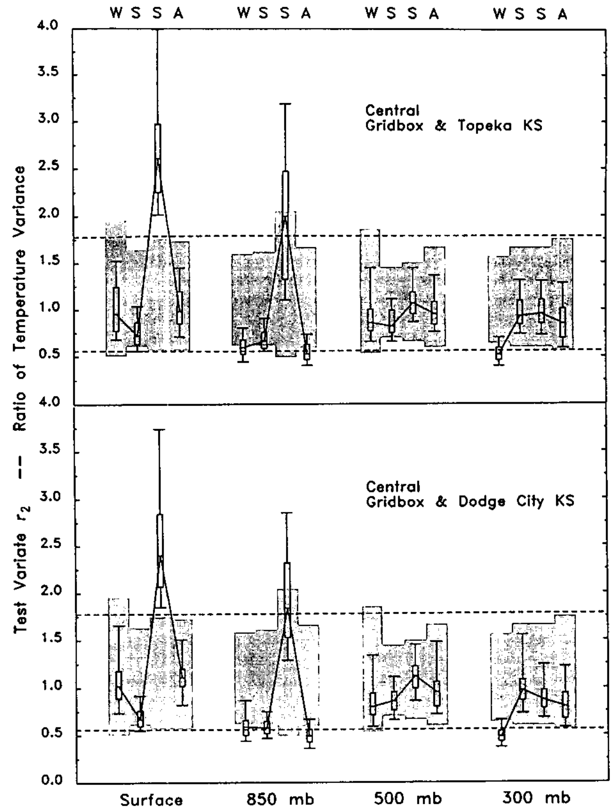

FIG. 3. Same as in Fig. 2, except for central grid box compared with Topeka, KS (top) and Dodge City, KS (bottom).

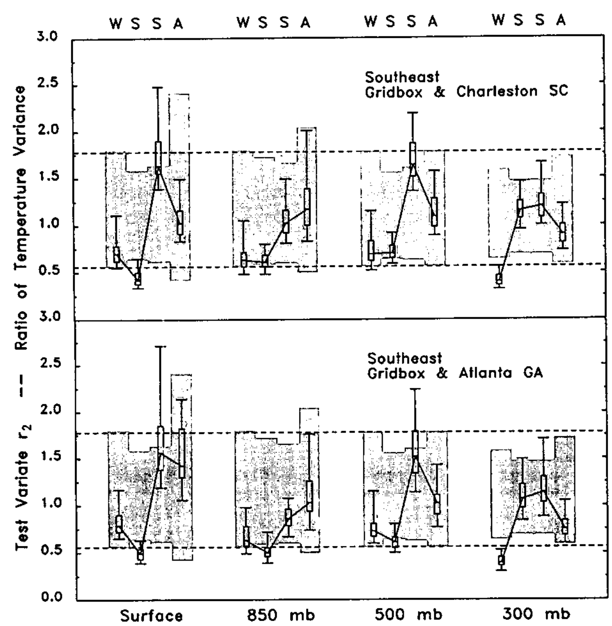

FIG. 4. Same as in Fig. 2, except for southeast grid box compared with Charleston, SC (top) and Atlanta, GA (bottom).

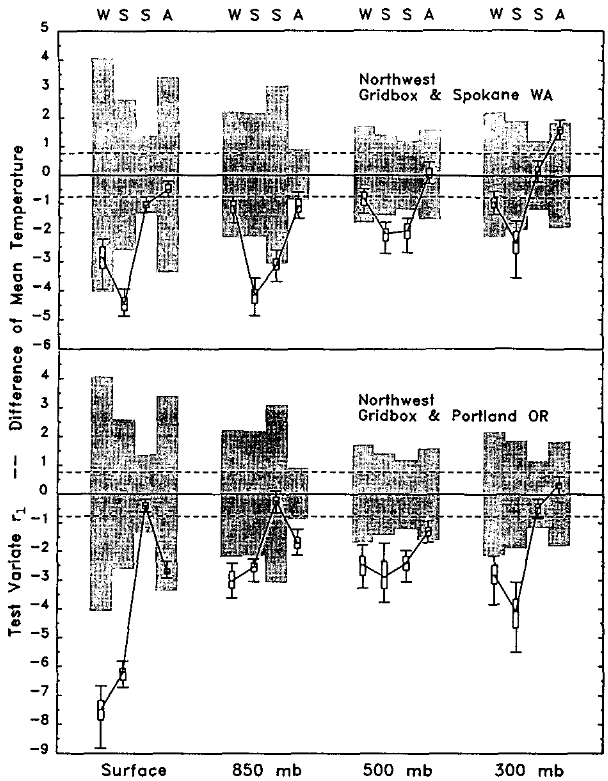

FIG. 5. Seasonal results of test variate r_1, standardized difference of mean temperature, for northwest grid box compared with Spokane, WA (top) and Portland, OR (bottom). Box plots and confidence intervals are as in Fig. 2, except for the confidence intervals of the Chervin approach (dashed lines) which indicate the a priori 95% confidence interval here (see text).

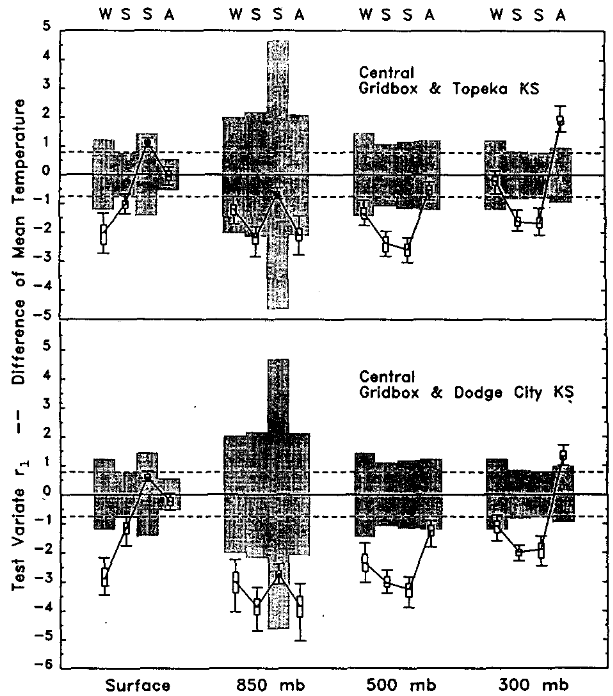

FIG. 6. Same as in Fig. 5, except for central grid box compared with Topeka, KS (top) and Dodge City, KS (bottom).

30th, 50th (median), 70th, and 90th percentiles of the results to compare CCM1 gridbox temperature variance with the temperature variances of local stations for each season and atmospheric level. Statistical confidence intervals are shown for the Chervin approach and for the second nonparametric approach. On the whole, these r_2 results remain well within the acceptance region boundaries of the two comparison approaches. Important outliers are found at the 300-mb level of the northwest grid box (Fig. 2) during spring, at the surface of the central grid box (Fig. 3) during summer, and at the 300-mb levels of both the central grid box and the southeast grid box (Fig. 4) during winter.

Seasonal results of test variate r_1 are presented in Figs. 5, 6, and 7, respectively, for the northwest, central, and southeast grid boxes to compare CCM1 gridbox mean temperatures with station mean temperatures for each season and atmospheric level. As in the case of r_2, confidence interval boundaries of the two comparison approaches are also shown. We find that although CCM1 grid boxes are generally cooler than the

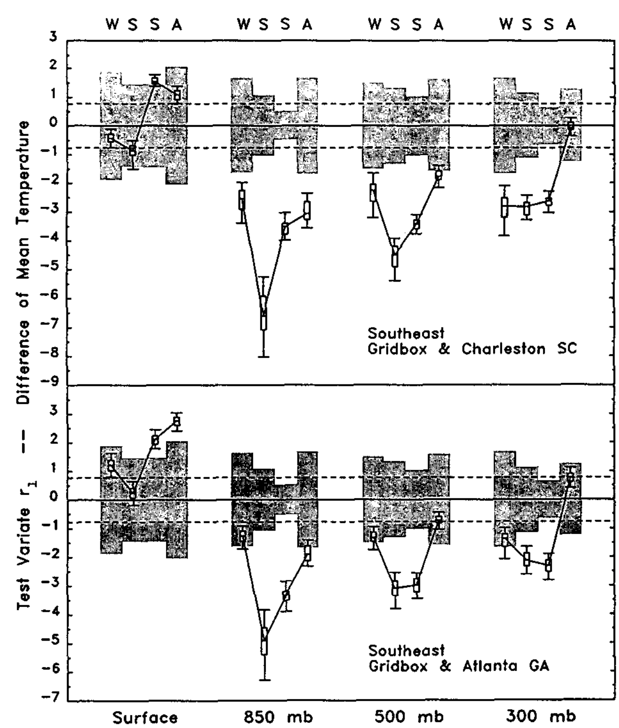

FIG. 7. Same as in Fig. 5, except for southeast grid box compared with Charleston, SC (top) and Atlanta, GA (bottom).

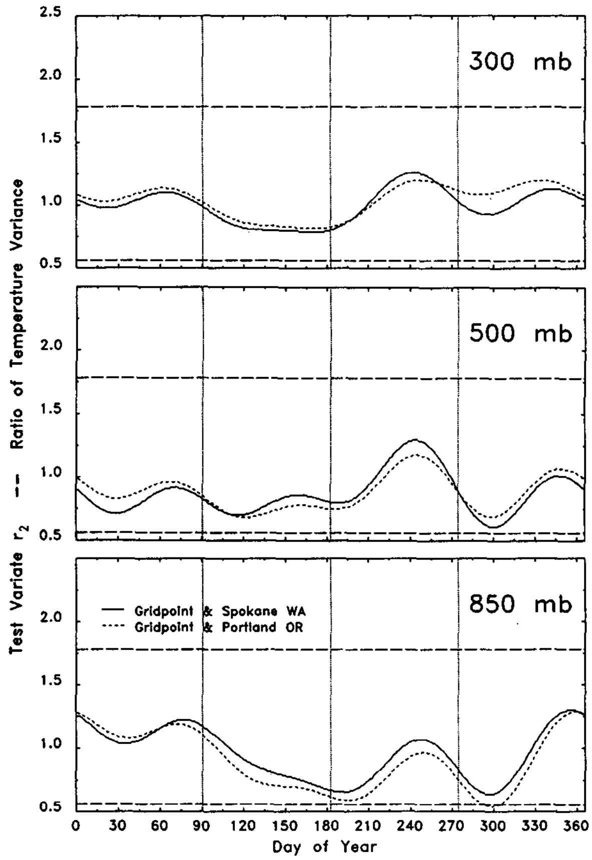

FIG. 8. Daily results of test variate r_2, ratio of temperature variance, at the 300-mb (top), 500-mb (center), and 850-mb (bottom) levels for northwest grid box compared with Spokane, WA (solid) and Portland, OR (dashed). Dotted vertical lines divide year (1 December–30 November) into four seasons. Dashed horizontal lines show a priori 90% confidence interval boundaries of the Chervin comparison approach.

local stations, statistical significance of differences is very much a function of which comparison approach is used. If we rely on the Chervin approach to evaluate the results, we find that significant differences of mean temperature between CCM1 grid boxes and local stations exist everywhere, for all levels and seasons. However, if we rely on the nonparametric approach to evaluate the comparison results, we find that the model actually reproduces the observed local mean temperatures quite well for many levels and seasons.

At the surface, greatest absolute differences are found between the northwest grid box and Portland, Oregon (bottom of Fig. 5) during winter–spring. This result is not surprising, given the complex terrain of the Pacific Northwest region and the large difference of elevation between the station and the grid box (see Table 1). In the free troposphere, gridbox-to-station temperature differences decrease with height and significant differences favor the spring and summer seasons. These patterns are clearly seen, for example, in the southeast grid box (Fig. 7), where r_1 values exceed six standard deviations at 850 mb.

b. Daily temperature

Daily results of test variate r_2 are presented in Figs. 8, 9, and 10, respectively, for the northwest, central, and southeast grid boxes. These results reflect in finer detail many of the interseasonal variations of the ratio of temperature variance that were previously noted concerning Figs. 2–4. Each of the curves shown in these figures was obtained by removing short-term periodicities (of less than 90 days) from a *noisy* 365-day time series. Since the variability of these *smoothed* results is relatively small, very few exceed the 90% confidence interval of the nonparametric comparison approach. Thus, only the confidence interval boundaries of the Chervin approach are shown in these figures.

Examining the curves, we note the range of values over the course of the year as well as dominant frequencies and peak amplitudes. We find that the ratio of gridbox-to-station temperature variance remains near 1.0, indicating good agreement between grid box

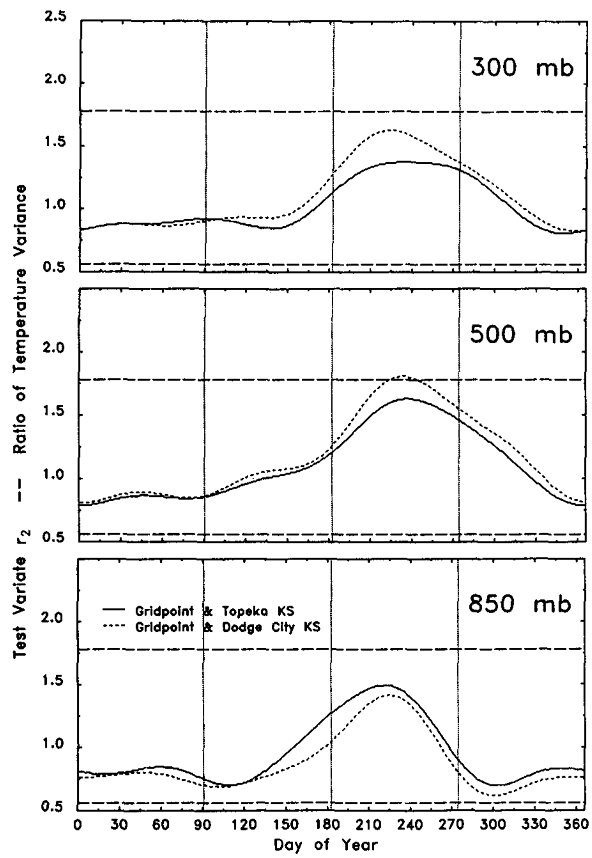

FIG. 9. Same as in Fig. 8, but for central grid box compared with Topeka, KS (solid) and Dodge City, KS (dashed).

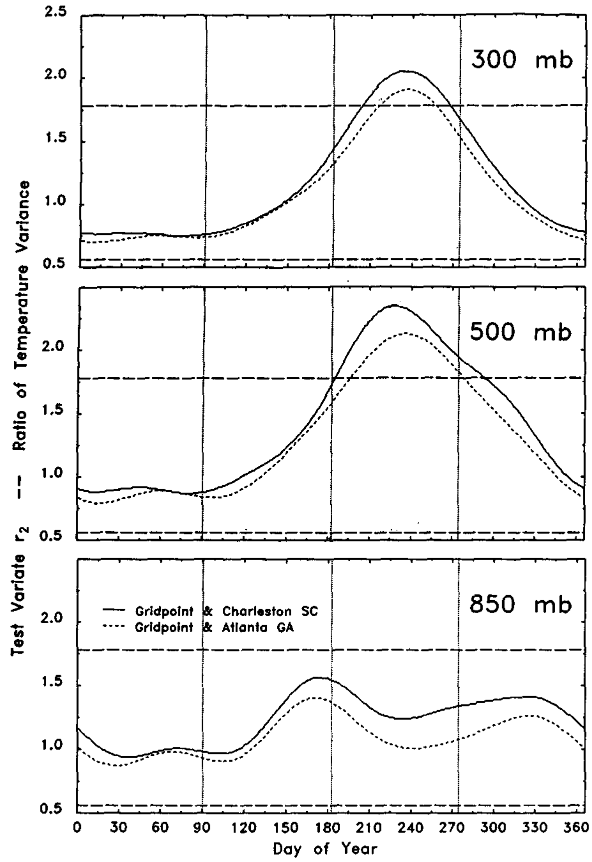

FIG. 10. Same as in Fig. 8, but for southeast grid box compared with Charleston, SC (solid) and Atlanta, GA (dashed).

and stations during most of the year. Minima (not significant) occur at the 850-mb level of the northwest grid box (Fig. 8) and maxima occur at the 500-mb level of the southeast grid box (Fig. 10). The maximum amplitude during summer increases from west to east across the United States.

Daily results of test variate r_1 are presented in Figs. 11, 12, and 13, respectively, for the same three grid boxes and reflect many of the interseasonal variations of the difference of mean temperature noted above for Figs. 5–7. As in the case of r_2, since the variability of these *smoothed* r_1 results is small, only the confidence interval boundaries of the Chervin comparison approach are shown in the figures.

We find that CCM1 grid boxes are cooler than local stations at the surface and 850-mb levels. However, at the 500-mb and 300-mb levels, this situation reverses during summer and autumn, when temperatures of some of the CCM1 grid boxes actually exceed the temperatures of some of the local stations. For example, at the 850-mb level of the central grid box (Fig. 12), the model is cooler than the two stations (i.e., r_1 is negative) for the entire year. At the 300-mb level, however, the CCM1 grid box becomes warmer than the local stations from late summer through midautumn.

5. Conclusions and discussion

We have presented two approaches to compare output of individual grid boxes in a GCM with local station observations. The first approach, which was proposed by Chervin (1980, 1981), involves the application of standard statistical analysis and hypothesis testing procedures. The second approach is nonparametric in the sense that no ideal distribution types are postulated a priori to ascertain statistical significance of differences between model output and observations. Rather, station observations are first subjected to a bootstrap technique and then used to define a potentially unique set of distributions and confidence limits for each model grid box.

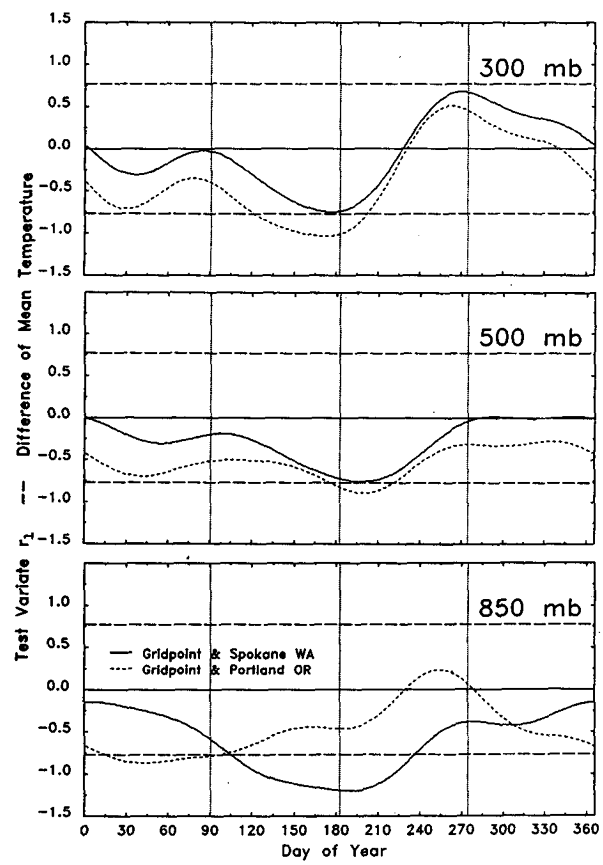

FIG. 11. Daily results of test variate r_1, standardized difference of mean temperature, at the 300-mb (top), 500-mb (center), and 850-mb (bottom) levels for northwest grid box compared with Spokane, WA (solid) and Portland, OR (dashed). Dotted vertical lines are as in Fig. 8. Solid horizontal lines depict zero and dashed horizontal lines show a priori 95% confidence interval boundaries of the Chervin comparison approach.

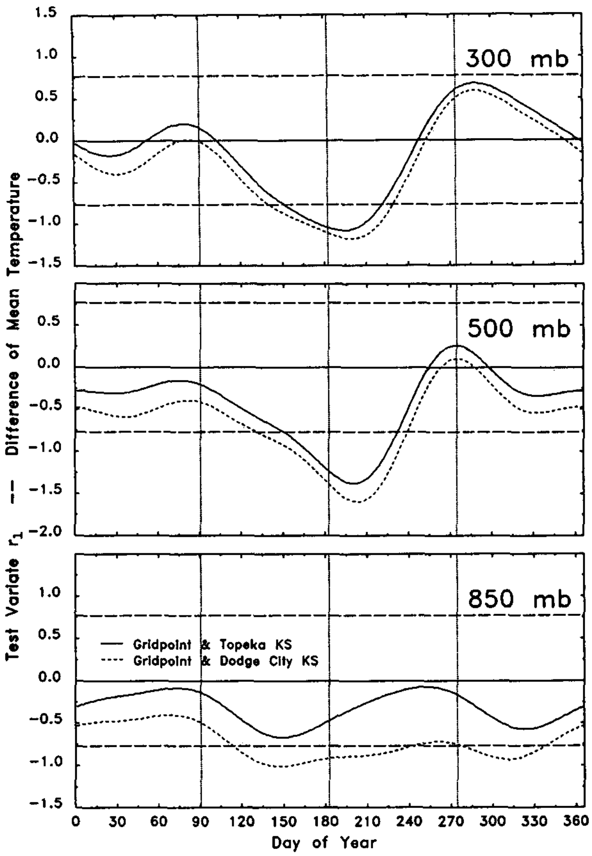

FIG. 12. Same as in Fig. 11, but for central grid box compared with Topeka, KS (solid) and Dodge City, KS (dashed).

To demonstrate the usefulness of the two approaches, daily and seasonal temperatures simulated by the National Center for Atmospheric Research (NCAR) Community Climate Model (CCM1) were compared with observations. Temperatures of individual CCM1 grid boxes were compared with local station temperatures for the surface and free troposphere for three separate regions in the United States. Statistical significance of the difference of mean temperature and the ratio of temperature variance between grid boxes and stations were assessed by means of two test variates (r_1 and r_2).

In general, we found that the gridbox temperatures of this particular CCM1 simulation are cooler than the local station temperatures but equally variable. In all grid boxes, the greatest differences of mean temperature between model output and observations were found to occur near the surface (surface and 850 mb). At higher levels (500 mb and 300 mb), there was close overall agreement for both means and variances. Of course, signs and magnitudes of differences can vary within and between grid boxes. In the northwest grid box, for example, since there were large differences in the observed climate across the grid box, agreement or lack thereof largely depended on which stations were selected to perform the comparisons.

Comparisons of the daily temperature indicated how the difference of mean temperature (r_1) and the ratio of temperature variance (r_2) can fluctuate over the course of the year. For example, at the upper levels (500 mb and 300 mb) of the southeast grid box, the model temperature variance exceeded the observed temperature variance during summer and the model mean temperature was cooler than the observed mean temperatures during spring. However, for both mean and variance, there was good agreement during the rest of the year.

Given the mismatch of time and space scales between model and observations, should we expect the climates of the GCM grid boxes to match the station climates? If stations and grid points are separated by some distance so that they occupy different latitudes, longitudes, and surface elevations, we certainly cannot expect absolute agreement. However, we also cannot simply assume that the agreement is poor. After all, here we have found better agreement among some sta-

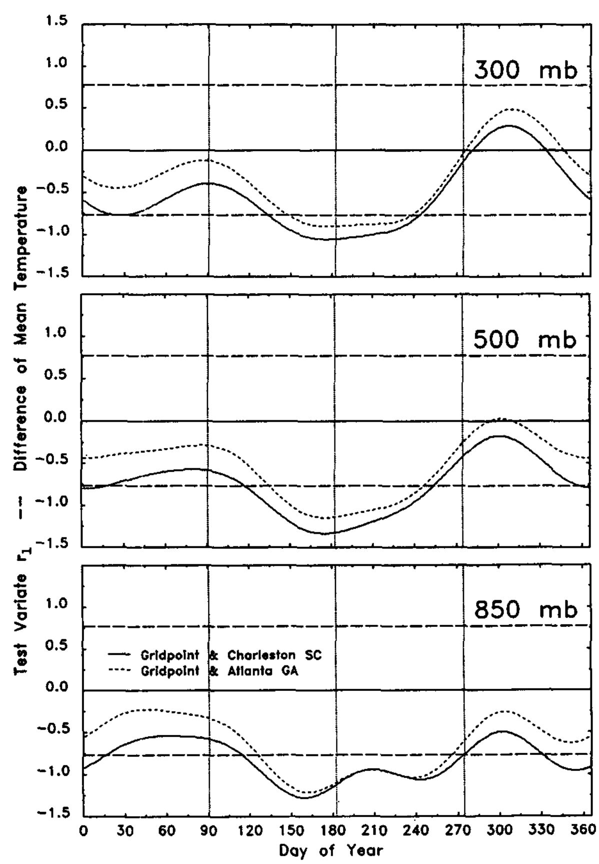

FIG. 13. Same as in Fig. 11, but for southeast grid box compared with Charleston, SC (solid) and Atlanta, GA (dashed).

tions than between grid points and stations separated by less than half the distance.

Let us consider a limiting case in which stations and grid boxes coincide; in each grid box, the stations occupy the same latitude, longitude, and elevation as the central grid point. Even here, differences would probably arise due to certain local space and time factors that remain unaccounted for in the global model. If we desire to use the results of regional comparisons to validate the model, then we must ascertain whether the criteria used to test agreement between model and observations are flexible enough to allow for interregional, subgrid-scale climate variations. We believe that our nonparametric approach, which uses differences of the observed climate to define unique test criteria for each grid box, provides the desired flexibility.

Test variate statistics such as r_1 and r_2 discussed in this study can be used to reveal information concerning systematic differences between the GCM simulated and observed climates for certain geographical and climatological situations. However, the three United States grid boxes considered here represent a very small sample of all possible situations. Investigation of additional grid boxes and stations should probably be conducted before any general conclusions are made.

Further study is also required to determine how systematic differences between GCM output and observations are related to the specific models considered. For example, output might be obtained from several GCMs run for the same length of time and with some of the same boundary conditions. One such effort, the Atmospheric Model and Intercomparison Program (AMIP), an activity of the World Climate Research Program's Working Group on Numerical Experimentation (WGNE 1990), is underway as of this writing. Running the GCM simulations for 20 or 30 years, instead of just 10, would also permit more definitive comparisons.

Although only surface and upper-air temperatures were compared in this study, the same diagnostic approaches and procedures can also be used to compare other parameters such as surface pressure, geopotential height, and moisture. However, GCM–observation comparisons can tell us only to what extent the models are capable of simulating regional and local observed climate characteristics. Actual improvement of the GCMs will have to rely on sensitivity studies. Even so, modelers should include statistical techniques such as those presented here among the standard diagnostics to be used while tuning the models.

Acknowledgments. The authors thank Dr. David L. Williamson of NCAR for providing some of the references that describe the bootstrap technique. This work was supported by a grant from the Atmospheric and Climatic Division, Office of Health and Environmental Research, U.S. Department of Energy (DE-FG02-86-ER60422).

REFERENCES

CES, 1989: *Our Changing Planet: The FY 1990 Research Plan, A report by the Committee on Earth Sciences.* U.S. Geological Survey, Reston, VA, 60 pp.

Chervin, R. M., 1980: On the simulation of climate and climate change with general circulation models. *J. Atmos. Sci.*, **37**, 1903–1913.

——, 1981: On the comparison of observed and GCM simulated climate ensembles. *J. Atmos. Sci.*, **38**, 885–901.

Diaconis, P., and B. Efron, 1983: Computer-intensive methods in statistics. *Scientific American*, **248**, 116–130.

DOE, 1985: *Projecting the Climatic Effects of Increasing Carbon Dioxide.* M. C. MacCracken and F. M. Luther, Eds. U.S. Department of Energy, Washington D.C., NTIS DOE/ER-0237, 381 pp.

Efron, B., 1982: *The Jackknife, the Bootstrap, and Other Resampling Plans,* CBMS-NSF Regional Conference Series in Applied Mathematics, Vol. 38, Society for Industrial and Applied Mathematics, 104 pp.

Glahn, H. R., 1985: Statistical weather forecasting. *Probability, Statistics, and Decision Making in the Atmospheric Sciences.* A. H. Murphy and R. W. Katz, Eds., Westview Press, 289–336.

Grotch, S. L., 1988: *Regional Intercomparisons of General Circulation Model Predictions and Historical Climate Data,* U.S. Department of Energy, Washington, D.C., DOE/NBB-0084, 291 pp.

Hansen, J., I. Fung, A. Lacis, D. Rind, S. Lebedeff, R. Ruedy, and G. Russell, 1988: Global climate changes as forecast by the Goddard Institute for Space Studies three dimensional model. *J. Geophys. Res.*, **93**, 9341–9364.

Hunt, B. G., and H. B. Gordon, 1989: Diurnally varying regional climatic simulations. *Int. J. Climatol.*, **9**, 331–356.

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

Klein, W. H., 1982: Statistical weather forecasting on different time scales. *Bull. Amer. Meteor. Soc.*, **63**, 170–176.

Klein, W. H., 1985: Space and time variations in specifying monthly mean surface temperature from the 700 mb height field. *Mon. Wea. Rev.*, **113**, 277–290.

——, and H. J. Bloom, 1987: Specification of monthly precipitation over the United States from the surrounding 700 mb height field. *Mon. Wea. Rev.*, **115**, 2118–2132.

——, and R. Yang, 1986: Specification of monthly mean surface temperature anomalies in Europe and Asia from concurrent 700 mb monthly mean height anomalies of the Northern Hemisphere. *J. Climatol.*, **6**, 463–484.

Manabe, S., K. Bryan, and M. D. Spelman, 1990: Transient response of a global ocean–atmosphere model to a doubling of atmospheric carbon dioxide. *J. Phys. Oceanogr.*, **20**, 722–749.

Mearns, L. O., S. H. Schneider, S. L. Thompson, and L. R. McDaniel, 1990: Analysis of climate variability in general circulation models: Comparison with observations and change in variability in $2 \times CO_2$ experiments. *J. Geophys. Res.*, **95**, 20 469–20 490.

Mitchell, J. F. B., C. A. Senior, and W. J. Ingram, 1989: CO_2 and climate: A missing feedback? *Nature*, **341**, 132–134.

Panofsky, H. A., and G. W. Brier, 1968: *Some Applications of Statistics to Meteorology.* The Pennsylvania State University, 224 pp.

Reed, D. N. 1986: Simulation of time series of temperature and precipitation over eastern England by an atmospheric general circulation model. *J. Climatol.*, **6**, 233–253.

Rind, D., R. Goldberg, and R. Ruedy, 1989: Change in climate variability in the 21st century. *Clim. Change*, **14**, 5–37.

Schlesinger, M. E., and W. L. Gates, 1980: The January and July performance of the OSU two-level atmospheric general circulation model. *J. Atmos. Sci.,* **37,** 1914–1943.

Solow, A. R., 1985: Bootstrapping correlated data. *Math. Geol.,* **17,** 769–775.

Wadsworth, G. P., and J. G. Bryan, 1960: *Introduction to Probability and Random Variables.* McGraw-Hill, 231–286.

Washington, W. M., and G. A. Meehl, 1984: Seasonal cycle experiments on the climate sensitivity due to a doubling of CO_2 with an atmospheric general circulation model coupled to a simple mixed layer ocean. *J. Geophys. Res.,* **89,** 9475–9503.

——, and ——, 1989: Climate sensitivity due to increased CO_2: Experiments with a coupled atmosphere and ocean general circulation model. *Climate Dyn.* **4,** 1–38.

Weare, B. C., and M. A. Hoeschele, 1983: Specification of monthly precipitation in the western United States from monthly mean circulation, *J. Clim. Appl. Meteor.,* **22,** 1000–1007.

WGNE, 1990. *Research Activities in Atmospheric and Oceanic Modelling,* Rep. No. 14, WMO, Geneva.

Wigley, T. M. L., P. D. Jones, K. R. Briffa, and G. Smith, 1990: Obtaining subgrid-scale information from coarse-resolution general circulation model output. *J. Geophys. Res.,* **95,** 1943–1954.

Williamson, D. L., J. T. Kiehl, V. Ramanathan, R. E. Dickinson, and J. J. Hack, 1987. *Description of NCAR Community Climate Model (CCM1).* Tech. Note, NCAR/TN-285+STR, National Center for Atmospheric Research, Boulder, CO, 112 pp.

Wilson, C. A., and J. F. B. Mitchell, 1987: Simulated climate and CO_2-induced climate change over Western Europe. *Climate Change,* **10,** 11–42.

WMO, 1987. *Clouds in Climate II, A WCRP workshop on modeling and observations.* Columbia, Maryland, World Climate Research Program, World Meteorological Organization, Geneva, 155 pp.

WMO/UNEP, 1990: *Scientific Assessment of Climate Change.* Intergovernmental Panel on Climate Change, World Meteorological Organization, Geneva, 200 pp.

Zwiers, F. W., 1990: The effect of serial correlation on statistical inferences made with resampling procedures. *J. Climate,* **3,** 1452–1460.