Jackknife Tests for Differences in Autocorrelation between Climate Time Series

T. A. BUISHAND AND J. J. BEERSMA

Royal Netherlands Meteorological Institute (KNMI), De Bilt, the Netherlands*

10 June 1992 and 3 May 1993

ABSTRACT

Two tests for differences in the lag 1 autocorrelation coefficient based on jackknife estimates are proposed. These tests are developed for the pooled sample of all daily values in a certain calendar month (e.g., all January data). Jackknife estimates of the autocorrelation coefficients and their standard errors from such a sample are obtained by omitting each year once and recomputing the autocorrelation estimates. Monte Carlo results for several distributions show that the critical values of the test statistics can be based on the Student's t distribution. Regional analogs of these test statistics are derived from the jackknife estimate of the mean lag 1 autocorrelation coefficient for the sites of interest. In a similar way one can get a single test statistic for a season or the whole year. As an illustration it is shown that the lag 1 autocorrelation coefficients of the simulated daily temperatures of the Canadian Climate Centre second-generation general circulation model are significantly below those of the observed temperatures at De Bilt for most seasons. Over western Europe there is no statistical evidence of differences in autocorrelation between the $1 \times CO_2$ and $2 \times CO_2$ runs of this model.

1. Introduction

Autocorrelation is a measure of the strength of linear dependence between successive values in a climate time series. A consequence of climate change can be that this dependence might become stronger or weaker. This may have a serious impact on society, because it affects the frequency and duration of extreme events. Thus, there is a need to test for differences in autocorrelation coefficients in climate experiments, and it is also important to know how faithfully general circulation models (GCMs) can reproduce the autocorrelation in observed records.

Autocorrelation coefficients of GCM simulated time series have rarely been computed. A few exceptions are Reed (1986), Rind et al. (1989), Wilson and Mitchell (1987), and Mearns et al. (1990). The latter two removed the autocorrelation from their data (prewhitening) to test for a change in variability as suggested by Katz (1988). The autocorrelation coefficients themselves have, however, never been subjected to statistical tests.

In this paper a procedure is presented for testing for differences in autocorrelation coefficients. The procedure requires the estimated autocorrelation coefficients with their standard errors. In this method it is not necessary to specify the form of the underlying distribution nor to model the autocorrelation structure.

* This work is partly supported by the Dutch National Research Program on Global Air Pollution and Climate Change (NOP).

Corresponding author address: Dr. T. A. Buishand, Royal Netherlands Meteorological Institute, P.O. Box 201, 3730 AE De Bilt, the Netherlands.

First the estimation of autocorrelation coefficients for seasonally varying data is considered. The jackknife is introduced as a technique to reduce bias and to obtain standard errors of the estimated autocorrelation coefficients. This leads to simple statistics for testing for differences in autocorrelation coefficients of two independent time series. The method is then extended to test for a regional change in autocorrelation with data from several locations. The use of the tests is illustrated with an observed record of daily temperatures in the Netherlands and with data generated by the Canadian Climate Centre (CCC) second-generation GCM (McFarlane et al. 1992; Boer et al. 1992) for the present CO_2 level and a doubling of the CO_2 level.

2. Estimation of autocorrelation coefficients

For ease of presentation it is assumed that time series of daily values are available. The lag k autocorrelation coefficient, ρ_k, then represents the correlation between two values separated by an interval of $k-1$ days. The estimation of the autocorrelation coefficients should be done with some care, because the systematic annual cycle may introduce serious bias. A popular method is to estimate the autocorrelation coefficients for each calendar month separately. Let $x_{i,j}$ denote the value on day i ($i = 1, \cdots, n$) of a certain month in year j ($j = 1, \cdots, J$). Then the lag k autocovariance is estimated as

$$c_k = \frac{1}{N_k} \sum_{j=1}^{J} \sum_{i=1}^{n-k} (x_{i,j} - \bar{x})(x_{i+k,j} - \bar{x}), \quad k = 0, 1, \cdots \tag{1}$$

TABLE 1. Empirical means of the estimates $r_{k,\mathrm{JACK}}$ and r_k of the autocorrelation coefficient ρ_k for an AR1 process with $\rho_1 = 0.8$ ($J = 10$, $n = 30$, 5000 simulations).

lag	ρ_k	$E(r_{k,\mathrm{JACK}})$	$E(r_k)$
1	0.800	0.800	0.789
2	0.640	0.641	0.621
5	0.328	0.327	0.297
10	0.107	0.107	0.076

where $\bar{x}$ is the mean over all daily values

$$\bar{x} = \frac{1}{nJ} \sum_{j=1}^{J} \sum_{i=1}^{n} x_{i,j}. \tag{2}$$

Quite often N_k is set equal to the total number of days, $N_k = nJ$ for all k (Katz 1982; Trenberth 1984). In this paper, however, $N_k = (n - k)J$, which results in less biased estimates. The values of the test statistics in section 4 do not depend on the choice of N_k. For $k = 0$, c_k represents the sample variance. The estimated autocorrelation coefficients are obtained as

$$r_k = c_k / c_0, \quad k = 0, 1, \cdots. \tag{3}$$

In the jackknife method the required statistic is recomputed a large number of times after successive deletion of a group of observations from the entire dataset (Efron 1982). To obtain simple tests it is necessary that the data in different groups are almost independent. For the climatic data considered here, it is therefore most appropriate to delete a complete year each time. The jackknife estimate of ρ_k then reads

$$r_{k,\mathrm{JACK}} = Jr_k - (J - 1)r_{k(\cdot)}, \quad k = 0, 1, \cdots, \tag{4}$$

where

$$r_{k(\cdot)} = \frac{1}{J} \sum_{j=1}^{J} r_{k(j)}, \tag{5}$$

with $r_{k(j)}$ the estimate of ρ_k after omitting the data for year j. The use of the jackknife results in almost unbiased estimates of ρ_k. Table 1 demonstrates this for 10-year samples of 30 daily values generated by a normal first-order autoregressive (AR1) process with $\rho_1 = 0.8$. For the AR1 process $\rho_k = \rho_1^k$, $k = 0, 1, \cdots$.

To derive the jackknife estimate, it is convenient to rewrite Eq. (1) as

$$c_k = \frac{1}{N_k} \{P_k - S[S_{L,k} + S_{R,k} - (n - k)S/n]/(nJ)\}, \quad k = 0, 1, \cdots, \tag{6}$$

where

$$P_k = \sum_{j=1}^{J} \sum_{i=1}^{n-k} (x_{i,j}x_{i+k,j}), \quad S = \sum_{j=1}^{J} \sum_{i=1}^{n} x_{i,j},$$

$$S_{L,k} = \sum_{j=1}^{J} \sum_{i=1}^{n-k} x_{i,j}, \quad S_{R,k} = \sum_{j=1}^{J} \sum_{i=1}^{n-k} x_{i+k,j}.$$

The estimate $r_{k(j)}$ in the right-hand side of Eq. (5) is then easily obtained by subtracting the contribution of year j from the quantities P_k, S, $S_{L,k}$, and $S_{R,k}$.

3. Standard errors of estimated autocorrelation coefficients

The jackknife is not only a method of bias reduction. The recomputed statistics $r_{k(j)}$ can also provide a distribution-free estimate of the standard error. The jackknife estimate of the variance of r_k is given by (Efron 1982)

$$\hat{\sigma}^2_{k,\mathrm{JACK}} = \frac{J - 1}{J} \sum_{j=1}^{J} [r_{k(j)} - r_{k(\cdot)}]^2. \tag{7}$$

For the generated 10-year samples of the AR1 process in the previous section, Table 2 compares the mean of $\hat{\sigma}_{k,\mathrm{JACK}}$ with the standard deviations $\sigma(r_{k,\mathrm{JACK}})$ and $\sigma(r_k)$ of $r_{k,\mathrm{JACK}}$ and r_k. The table also presents these results for an AR2 process with the same value of ρ_1 as the AR1 process (0.8) but with a much smaller value of ρ_2 (0.45 instead of 0.64). The statistic $\hat{\sigma}_{k,\mathrm{JACK}}$ yields an almost unbiased estimate of $\sigma(r_k)$. It slightly underestimates $\sigma(r_{k,\mathrm{JACK}})$, in particular for large lags. This is consistent with Efron (1982) where Eq. (7) is considered to provide an estimate of var(r_k) rather than of var($r_{k,\mathrm{JACK}}$).

There are marked differences between the values of the standard errors for the two autoregressive processes in Table 2. For the AR2 process the standard error of r_1 is only 60% of that for the AR1 process.

TABLE 2. Empirical mean of the estimate $\hat{\sigma}_{k,\mathrm{JACK}}$ compared with the empirical standard deviations of $r_{k,\mathrm{JACK}}$ and r_k for an AR1 process with $\rho_1 = 0.8$ ($\rho_2 = 0.64$) and an AR2 process with $\rho_1 = 0.8$ and $\rho_2 = 0.45$ ($J = 10$, $n = 30$, 5000 simulations).

	AR1 process			AR2 process		
Lag	$E(\hat{\sigma}_{k,\mathrm{JACK}})$	$\sigma(r_{k,\mathrm{JACK}})$	$\sigma(r_k)$	$E(\hat{\sigma}_{k,\mathrm{JACK}})$	$\sigma(r_{k,\mathrm{JACK}})$	$\sigma(r_k)$
1	0.038	0.039	0.038	0.022	0.023	0.023
2	0.065	0.067	0.065	0.053	0.055	0.054
5	0.108	0.115	0.107	0.097	0.102	0.098
10	0.134	0.147	0.135	0.112	0.117	0.113

For the jackknife method it is not necessary to specify a time series model. Resampling of the residuals of a fitted autoregressive model can also provide an estimate of the standard error of the autocorrelation coefficients (Efron and Tibshirani 1986). This estimate generally has a smaller variance than our jackknife estimate. The estimate from a fitted time series model is, however, sensitive to model misspecification (Künsch 1989) and is more difficult to obtain.

4. Testing for differences at a single site

Table 2 shows that the estimated lag 1 autocorrelation coefficient has a much smaller standard error than the higher-order autocorrelation coefficients. A systematic difference in the values of ρ_1 for two climates (e.g., a 1 × CO_2 run and a 2 × CO_2 run) is therefore better detected than differences in the other autocorrelation coefficients. In addition tests for differences at more than one lag do not give independent results because the autocorrelation estimates at neighboring lags are correlated. For these reasons only the lag 1 autocorrelation coefficient is considered. Let $\rho_1(\mathrm{I})$ and $\rho_1(\mathrm{II})$ be the theoretical lag 1 autocorrelation coefficients for climate I and climate II, respectively. A test for differences in autocorrelation can then be based on their jackknife estimates $r_{1,\mathrm{JACK}}(\mathrm{I})$, $r_{1,\mathrm{JACK}}(\mathrm{II})$ and the jackknife estimates $\hat{\sigma}_{1,\mathrm{JACK}}(\mathrm{I})$, $\hat{\sigma}_{1,\mathrm{JACK}}(\mathrm{II})$ of the standard errors of the autocorrelation estimator from Eq. (7). A possible test statistic is

$$T_1 = \frac{r_{1,\mathrm{JACK}}(\mathrm{II}) - r_{1,\mathrm{JACK}}(\mathrm{I})}{[\hat{\sigma}^2_{1,\mathrm{JACK}}(\mathrm{I}) + \hat{\sigma}^2_{1,\mathrm{JACK}}(\mathrm{II})]^{1/2}}. \quad (8)$$

A similar statistic was used by Miller (1968) to test for differences between the variances of two independent samples. Another possibility is (Davis 1979)

$$\tilde{T}_1 = \left[\frac{JK(J+K-2)}{J+K}\right]^{1/2} \times \left\{\frac{r_{1,\mathrm{JACK}}(\mathrm{II}) - r_{1,\mathrm{JACK}}(\mathrm{I})}{[J(J-1)\hat{\sigma}^2_{1,\mathrm{JACK}}(\mathrm{I}) + K(K-1)\hat{\sigma}^2_{1,\mathrm{JACK}}(\mathrm{II})]^{1/2}}\right\} \quad (9)$$

with J and K the number of years for climate I and climate II, respectively. The statistic $\tilde{T}_1$ is the analog of the two-sample Student's t-statistic for testing differences in the mean. Statistics T_1 and $\tilde{T}_1$ are the same if $J = K$, but they are different for unequal sample sizes. Under the null hypothesis, $\rho_1(\mathrm{I}) = \rho_1(\mathrm{II})$, the statistic $\tilde{T}_1$ is approximately a Student's t-variable with $J + K - 2$ degrees of freedom. The quality of the approximation depends on J, K, the group size n, and the probability distribution of the climate variable involved. It is not valid, however, to derive the critical values of T_1 from this Student's distribution if $J \neq K$. Under the null hypothesis it is expected that this statistic behaves approximately as

$$\frac{\chi(1/J + 1/K)^{1/2}}{\{\chi^2_{J-1}/[J(J-1)] + \chi^2_{K-1}/[K(K-1)]\}^{1/2}} \quad (10)$$

where χ is a standard normal variable, χ^2_{J-1} and χ^2_{K-1} are two independent chi-squared variables with $J - 1$ and $K - 1$ degrees of freedom, also independent of χ. To make use of Student's t-distribution, the sum of the two chi-squared variables in the denominator of Eq. (10) is approximated as

$$\frac{\chi^2_{J-1}}{J(J-1)} + \frac{\chi^2_{K-1}}{K(K-1)} \approx \frac{\chi^2_d}{d}(1/J + 1/K). \quad (11)$$

The mean of both sides equals $1/J + 1/K$. The number d of effective degrees of freedom is chosen such that the variances are also equal, giving

$$d = \frac{(J+K)^2}{K^2/(J-1) + J^2/(K-1)}. \quad (12)$$

This method of approximation is known as the Satterthwaite method (Gaylor 1988). Note that $d = J + K - 2 = 2(J - 1)$ if $J = K$, whereas d tends to be smaller than $J + K - 2$ if $J \neq K$. The approximation of Eq. (10) by a Student's t-variable with d degrees of freedom works quite well (Tanburn 1938; Welch 1949). For noninteger d the t distribution is still defined through its relation with the beta distribution. The critical values of the test statistic can be obtained numerically or by interpolation in a table of Student's t distribution.

The approximation of the null distributions of T_1 and $\tilde{T}_1$ by the Student's t distribution has been checked by a Monte Carlo experiment with different AR1 pro-

TABLE 3. Empirical significance levels of two-sided tests for a difference in the lag 1 autocorrelation coefficient ρ_1 (n = 30; 2500 simulations for J = 10, K = 30; 5000 simulations in the other cases). The critical values $C_{0.10}$, $C_{0.05}$ and $C_{0.01}$ of the test statistic are based on the Student's t distribution. For the generated AR1 processes ρ_1 = 0.8, whereas for the AR2 processes ρ_1 = 0.8 and ρ_2 = 0.45. N refers to the normal distribution, E to the exponential distribution, and L to the Laplace distribution.

Process	Statistic	J	K	$C_{0.10}$	$C_{0.05}$	$C_{0.01}$
N AR1	$T_1, \tilde{T}_1$	5	5	.099	.047	.010
N AR1	$T_1, \tilde{T}_1$	10	10	.100	.054	.013
N AR1	T_1	5	15*	.099	.045	.008
N AR1	$\tilde{T}_1$	5	15	.120	.062	.016
N AR1	T_1	10	30*	.097	.053	.014
N AR1	$\tilde{T}_1$	10	30	.100	.049	.017
N AR2	$T_1, \tilde{T}_1$	5	5	.100	.051	.011
N AR2	$T_1, \tilde{T}_1$	10	10	.102	.049	.010
E AR1	$T_1, \tilde{T}_1$	5	5	.135	.069	.017
E AR1	T_1	5	15*	.120	.054	.009
E AR1	$\tilde{T}_1$	5	15	.151	.084	.023
L AR1	$T_1, \tilde{T}_1$	5	5	.109	.058	.014
L AR1	T_1	5	15*	.112	.053	.011
L AR1	$\tilde{T}_1$	5	15	.126	.069	.019

* d = 6.89 (J = 5, K = 15) and d = 15.47 (J = 10, K = 30), critical values obtained from Gardiner and Bombay (1965).

cesses and the normal AR2 process in Table 2. The main results of this experiment are summarized in Table 3. The generation of the AR1 processes with exponential and Laplace (double exponential) distributions was based on Lawrance (1981). For the normal and Laplace distributions the empirical significance levels are close to the nominal values. The largest discrepancies occur with the use of $\tilde{T}_1$ when $J = 5$ and $K = 15$. For small samples of unequal lengths the test based on $\tilde{T}_1$ becomes progressive. This has also been found with a similar jackknife test statistic for testing for differences in variances (Brown and Forsythe 1974; Boos and Brownie 1989). For the generated exponential processes the null distributions of T_1 and $\tilde{T}_1$ deviate more from the Student's t distribution. The exponential distribution is, however, a very skewed distribution. Climate variables with strong autocorrelation generally exhibit less skewness.

In the light of the results in Table 2 some simulations were also done with the autocorrelation estimate r_1 in the numerator of the test statistics instead of $r_{1,\mathrm{JACK}}$. The discrepancies in the exponential cases then disappear, but the empirical significance levels are significantly below the nominal values for the 5-year samples ($J = K = 5$) from the normal AR1 process. Simulation further indicates that Jenkins's arc sine transformation of the autocorrelation estimate (Kendall et al. 1983, 48.20) has little effect.

5. Increasing the power of the test by combining several samples

For GCM runs of limited length the statistics T_1 and $\tilde{T}_1$ can only detect quite substantial differences in autocorrelation. For two normal AR1 processes with $\rho_1(\mathrm{I}) = 0.8$ and $\rho_1(\mathrm{II}) = 0.7$, the probability of a significant result at the 5% level is about 0.35 if $J = K = 10$ (group size $n = 30$). A more powerful test is possible by combining the samples at several grid points within a region. Regional analogs of the test statistics T_1 and $\tilde{T}_1$ are derived as follows. For each grid point the estimates $r_1, r_{1(1)}, \ldots, r_{1(J)}$ are calculated. Taking averages over the grid points results in $\bar{r}_1, \bar{r}_{1(1)}, \ldots, \bar{r}_{1(J)}$. The jackknife estimates in the test statistic T_1 (or $\tilde{T}_1$) are then obtained by applying Eqs. (4), (5), and (7) to these average autocorrelation estimates. The value from Eq. (4) then represents an estimate $\bar{r}_{1,\mathrm{JACK}}$ of the mean lag 1 autocorrelation coefficient $\bar{\rho}_1$ for the grid points under consideration. This estimate is almost unbiased. The increase in power results from the fact that $\bar{r}_{1,\mathrm{JACK}}$ has a smaller standard error than the autocorrelation estimate for a single grid point.

An alternative regional autocorrelation estimate would be obtained by averaging the lag 1 autocovariance and variance estimates separately over the grid points. This regional estimate can, however, strongly differ from $\bar{\rho}_1$ when the second-order moments vary over the grid. Because the estimate then has no easy interpretation it is not considered further.

It should be noted that $\bar{\rho}_1$ is not necessarily equal to the lag 1 autocorrelation coefficient ρ_1^* of the spatial average of the data. The latter also depends on the lag 1 cross-correlation coefficients between the various samples. There is no guarantee that a jackknife estimate of ρ_1^* will have a smaller standard error than the autocorrelation estimates at the individual grid points.

By a similar averaging procedure one can get a single test statistic for the December, January, February (DJF) period and other seasons, as well as for the whole year. In the latter case, the group of deleted observations contains a number of days that are directly adjacent to those left in. It is unlikely that this will impair the use of the jackknife procedure, because only a limited number of the daily values in two subsequent years are significantly correlated. A simultaneous test on the results for individual calendar months is also possible using the binomial distribution or by means of Fisher's test (Sneyers 1990). These methods, however, assume independence among the values of the test statistics in successive months.

6. Results for daily mean surface air temperatures

In this section the proposed method is illustrated with the 1961–1990 temperature record at De Bilt (52°06′N, 5°11′E) and generated time series over Europe of the CCC GCM. Ten-year samples for both the $1 \times CO_2$ and $2 \times CO_2$ simulated climates were made available over a grid of 3.75° × 3.75°.

Table 4 presents autocorrelation estimates for the De Bilt record and the GCM $1 \times CO_2$ run at two grid points. Because the time series have unequal lengths, both T_1 and $\tilde{T}_1$ are given. For the GCM data the autocorrelation coefficients are systematically lower than those for the observed temperatures. This leads to very significant differences in the lag 1 autocorrelation coefficients averaged over the year. For the individual sea-

TABLE 4. Tests for differences in the lag 1 autocorrelation coefficients of daily temperature observations at De Bilt (1961–1990) and daily temperatures for the $1 \times CO_2$ climate of the CCC GCM (10 years). B refers to De Bilt, W and E refer to the nearest grid point west and east of De Bilt, respectively. The DJF values for De Bilt are based on the 29 complete winters in the 30-year period.

Statistic	DJF	MAM	JJA	SON	Year
$\bar{r}_{1,\mathrm{JACK}}$ (B)	0.85	0.82	0.79	0.78	0.81
$\bar{r}_{1,\mathrm{JACK}}$ (W)	0.73	0.73	0.73	0.69	0.72
T_1(B, W)	−3.46†	−2.24*	−2.39*	−3.42†	−5.39†
$\tilde{T}_1$(B, W)	−4.17†	−3.26†	−2.27*	−4.51†	−6.96†
$\bar{r}_{1,\mathrm{JACK}}$ (E)	0.75	0.77	0.73	0.71	0.74
T_1(B, E)	−3.23†	−1.65	−2.15*	−2.51*	−4.96†
$\tilde{T}_1$(B, E)	−3.63†	−2.15*	−2.27*	−3.30†	−5.85†

* (†) differences significant at the 5% (1%) level.

sons the underestimation is also significant, except for the spring temperatures of the grid point east of De Bilt when the statistic T_1 is used. The result for the statistic $\tilde{T}_1$ is often more significant than that for T_1. A Monte Carlo experiment with normal AR1 processes shows that the former statistic is slightly more powerful in the situation of Table 4. This is caused not only by the larger number of degrees of freedom, but also by the fact that the shortest samples have the smallest lag 1 autocorrelation coefficients.

To test for differences in autocorrelation between the generated $1 \times CO_2$ and $2 \times CO_2$ temperature data an area of 25 grid points around the Netherlands (extending from 40.8° to 59.4°N latitude and 5.6°W to 13.1°E longitude) was considered. The area includes Great Britain, Denmark, France, Germany, and parts of Spain and Italy. Sixteen grid points have been designated as land points, the other nine are sea points. Land and sea points were treated separately. The autocorrelation estimates are summarized in Table 5. For both the land and the sea points the standard error of the autocorrelation estimate is reduced by a factor 1.7 as a result of spatial averaging. This reduction is comparable to that obtained by averaging over three consecutive calendar months. Table 5 shows that the statistics based on the regional seasonal averages are only significant for the land points in winter (DJF) and for the sea points in summer (JJA). According to the binomial distribution the joint result for the four seasons is not significant at the 5% level. There is also no statistical evidence of differences in the annual mean of the lag 1 autocorrelation coefficient. This agrees with results in Rind et al. (1989) for a transient climate experiment.

TABLE 5. Tests for differences in the lag 1 autocorrelation coefficients between the $1 \times CO_2$ (C) and $2 \times CO_2$ (D) daily temperatures of the CCC GCM (both 10 years) for an area of 25 grid points around the Netherlands.

Statistic	DJF	MAM	JJA	SON	Year
		Land grid points (16)			
$\bar{r}_{1,\text{JACK}}$ (C)	0.70	0.72	0.74	0.68	0.71
$\bar{r}_{1,\text{JACK}}$ (D)	0.63	0.70	0.76	0.70	0.70
T_1	−2.63*	−0.37	0.77	0.92	−0.85
		Sea grid points (9)			
$\bar{r}_{1,\text{JACK}}$ (C)	0.81	0.84	0.87	0.86	0.85
$\bar{r}_{1,\text{JACK}}$ (D)	0.85	0.84	0.91	0.88	0.87
T_1	1.00	−0.19	2.26*	1.23	1.49

* differences significant at the 5% level.

7. Discussion

In the present paper it is demonstrated that the jackknife can be used to reduce the bias in autocorrelation estimates of climate time series and to calculate the standard errors of these estimates. The examples in Tables 4 and 5 show that tests based on the jackknife perform quite well. From Table 4 it is seen that these tests are able to detect differences in the range of 0.05 to 0.10 between the seasonal autocorrelation estimates of the 30-year observed temperature record and a 10-year GCM simulation. Table 5 indicates that changes in autocorrelation larger than 0.05 would generally be significant at the 5% level when 10-year samples over a region of $\sim 2.5 \times 10^6$ km^2 are combined.

The method is not suitable for very short GCM simulations. For sample sizes smaller than those in Table 3, Student's t distribution does not always provide a good approximation to the critical values of the test statistics T_1 and $\tilde{T}_1$. But, even if this distribution can be used, the critical values will be large as a result of lack of degrees of freedom. Unfortunately, averaging of monthly values does not lead to more degrees of freedom. Because of this and the relatively large standard errors the tests are then unable to detect meaningful changes in autocorrelation.

An extension of the jackknife procedure of estimating standard errors to the case of a stationary sequence $x_1, \ldots, x_N$ of correlated data has been proposed by Künsch (1989). In this extension an estimate of the lag 1 autocorrelation coefficient is obtained from regression of x_{i+1} on x_i ($i = 1, \cdots, N-1$) and jackknifing takes place through deletion of a single pair (x_i, x_{i+1}) or a group of l consecutive pairs. The standard error is obtained by an expression similar to Eq. (7). Simulations as in Table 2 indicate that the method performs quite well even for small sample sizes (e.g., $J = 2$ and $n = 30$). The choice of the group size l is important. For an AR1 process $l = 1$ is optimal, but for other processes Künsch (1989) demonstrates that a serious bias can be incurred with this choice. Because a relatively small amount of data is deleted for jackknifing there are sufficient degrees of freedom. It is not clear for which sample sizes Student's t distribution or the normal distribution can be used in tests for differences in autocorrelation with this method. A closely related bootstrap procedure is also presented in Künsch (1989). This procedure does not make use of a fitted time series model.

Our jackknife approach requires much less computer time than other resampling techniques, especially when use is made of Eq. (6) to obtain the various autocorrelation estimates. Other statistical properties of climate time series can also be investigated with this technique. In particular, a jackknife estimate of the variance is easily obtained from the estimation procedure of the autocorrelation coefficients. Tests for differences in variability can therefore be conducted in the same way. It is sometimes possible, however, to obtain more powerful tests by prewhitening the data first. A paper on this topic is in preparation.

Acknowledgments. We would like to thank the Canadian Climate Centre for providing the GCM output.

REFERENCES

Boer, G. J., N. A. McFarlane, and M. Lazare, 1992: Greenhouse gas-induced climate change simulated with the CCC second-generation general circulation model. *J. Climate,* **5,** 1045–1077.

Boos, D. D., and C. Brownie, 1989: Bootstrap methods for testing homogeneity of variances. *Technometrics,* **31,** 69–82.

Brown, M. B., and A. B. Forsythe, 1974: Robust tests for the equality of variances. *J. Amer. Statist. Assoc.,* **69,** 364–367.

Davis, W. W., 1979: Robust methods for detection of shifts of the innovation variance of a time series. *Technometrics,* **21,** 313–320.

Efron, B., 1982: *The Jackknife, the Bootstrap and Other Resampling Plans.* Society for Industrial and Applied Mathematics, 92 pp.

——, and R. Tibshirani, 1986: Statistical methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statist. Sci.,* **1,** 54–77.

Gardiner, D. A., and B. F. Bombay, 1965: An approximation to Student's *t*. *Technometrics,* **7,** 71–72.

Gaylor, D. W., 1988: Satterthwaite's formula. *Encyclopedia of Statistical Sciences, Vol. 8,* S. Kotz, N. L. Johnson, and C. B. Read, Eds., Wiley and Sons, 261–262.

Katz, R. W., 1982: Statistical evaluation of climate experiments with general circulation models: A parametric time series modeling approach. *J. Atmos. Sci.,* **39,** 1446–1455.

——, 1988: Statistical procedures for making inferences about climate variability. *J. Climate,* **1,** 1057–1064.

Kendall, M., A. Stuart, and J. K. Ord, 1983: *The Advanced Theory of Statistics, Vol. 3, 4th edition.* Charles Griffin, 780 pp.

Künsch, H. R., 1989: The jackknife and the bootstrap for general stationary observations. *Ann. Statist.,* **17,** 1217–1241.

Lawrance, A. J., 1981: Some autoregressive models for point processes. *Point Processes and Queuing Problems,* P. Bartfai and J. Tomko, Eds., Elsevier North Holland, 257–275.

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.

Means, 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 changes in variability in $2 \times CO_2$ experiments. *J. Geophys. Res.,* **95,** 20 469–20 490.

Miller, R. G., 1968: Jackknifing variances. *Ann. Math. Statist.,* **39,** 567–582.

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.

Sneyers, R., 1990: On the statistical analysis of series of observations. World Meteorological Organization. Tech. Note No. 143, WMO No. 415, 192 pp. (original French version published in 1975). P.O. Box 2300, CH 1211, Genève 2, Switzerland.

Tanburn, E., 1938: Note on an approximation used by B. L. Welch. *Biometrika,* **29,** 361–362.

Trenberth, K. E., 1984: Some effects of finite sample size and persistence on meterological statistics. Part I: Autocorrelations. *Mon. Wea. Rev.,* **112,** 2359–2368.

Welch, B. L., 1949: Further note on Mrs. Alpin's tables and on certain approximations to the tabled function. *Biometrika,* **36,** 293–296.

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