Confidence Intervals of a Climatic Signal

YOSHIKAZU HAYASHI

Geophysical Fluid Dynamics Laboratory/NOAA, Princeton University, Princeton, NJ 08540

(Manuscript received 23 July 1981, in final form 13 May 1982)

ABSTRACT

In order to interpret climate statistics correctly, the definitions of climate change, signal-to-noise ratio and statistical significance are clarified.

It is proposed to test the significance of climate statistics by the use of confidence intervals, since they are more informative than merely testing the null hypothesis that the true response is zero. The confidence intervals of the mean difference, variance ratio and signal-to-noise ratio are formulated and applied to a climate sensitivity study.

It is also proposed to make a multivariate test of a response pattern by the use of joint confidence intervals, since they are more informative than merely testing the null hypothesis that the true response is everywhere zero. These intervals can also be applied to test the joint significance of the amplitude and phase of the seasonal cycles of a response.

1. Introduction

In order to estimate the standard error of time-average estimates of climatic means, Leith (1973) proposed to estimate the standard deviation (σ_T) of the finite-time mean. When many independent sample means are not available, σ_T can be estimated indirectly through the autocorrelation (Leith, 1973) or power spectra (Munk, 1960; Jones, 1975, 1976) of dependent daily data. In an effort to test the statistical significance of numerical experiments of climates, Shukla (1975) and Manabe and Hahn (1977) estimated the ratio of the time-mean difference between two experiments to the standard deviation of the time mean. However, it was not clear how large this estimated signal-to-noise ratio must be for the difference to be significant. Moreover, these two papers confused their *estimated* signal-to-noise ratio with the *true* signal-to-noise ratio defined by Leith (1973). The definitions of signal-to-noise ratio and statistical significance will be clarified in Section 2, extending Leith's (1973) discussion.

Subsequently, Chervin and Schneider (1976) showed that the criterion of the estimated signal-to-noise ratio for the statistical significance of mean difference is given by the t test (see Panofsky and Brier, 1968, p. 63; Mitchell, 1971, p. 63). Since then, the t test has been applied to climatic studies by Chervin *et al.* (1976), Laurmann and Gates (1977), Washington *et al.* (1977), van Loon and Rogers (1978), Julian and Chervin (1978), Holton and Tan (1980), Chervin (1980, 1981) and Keshavamurty (1982), while Warshaw and Rapp (1973) applied the analysis of "variance test" (see Panofsky and Brier, 1968, p. 66) which can test the equality of two or more variables.

Student's t is the ratio of sample mean difference to its sample standard deviation which is estimated directly from independent data without autocorrelation. The true standard deviation need not be known, since the distribution of t is known. This t is subject to a conventional null hypothesis test (see Fig. 1). If t exceeds its critical value of the t distribution, the null hypothesis of zero true difference is rejected at the specified probability level. In this case, the true difference is unlikely to be zero with a certain risk of false rejection (i.e., Type I error). However, even if the null hypothesis is not rejected, it is dangerous to conclude that the unknown true difference may be close to zero, since the probability of false acceptance (i.e., Type II error) can be larger than that of false rejection. Thus, this null hypothesis test alone is not satisfactory to claim that the unknown true difference between simulations and observations is likely to be small. The probability of false rejection can be evaluated by the "power of the test" (see Bendat and Piersol, 1971, p. 117). However, it is simpler and more informative to determine the range of hypothesized true values accepted by a null hypothesis test. For this purpose, the present paper proposes to determine confidence intervals.[1]

Recently, Hasselmann (1979a) and Storch (1982a,b) argued that a univariate significance test

[1] Chervin (1981) proposed the use of confidence intervals for testing the climatic mean itself but not the mean difference. He did not discuss the advantage of this test over a null hypothesis test.

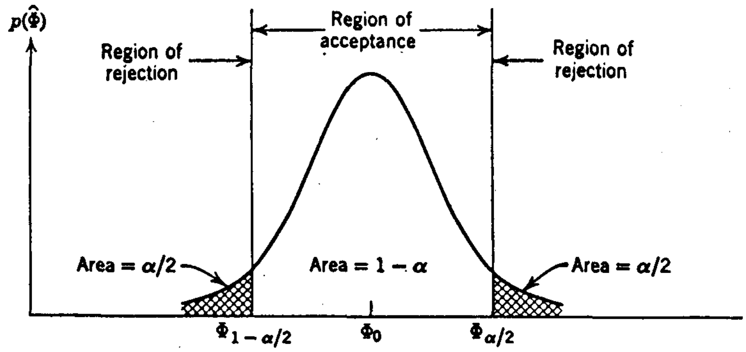

FIG. 1. Probability density of Φ around the true value Φ_0. When an estimate of Φ exceeds its critical value $\Phi_{\alpha/2}$, the null hypothesis that the true value is Φ_0 is rejected at the level of 100α percent (after Bendat and Piersol, 1971).

can be misleading when it is applied to a response pattern. When the true response is *everywhere close to zero*, a univariate null hypothesis will be *falsely* rejected at the 5% significance level for 5% of independent spatial data. In this case, the multivariate null hypothesis that the true response is zero everywhere will be *correctly* accepted at the same level on the basis of the joint probability distribution. Hasselmann showed that this null hypothesis tends to be more difficult to reject as spatial data increase, unless noise is somehow filtered out.

However, it can be argued that a multivariate test can also be misleading when the true response is *somewhere non-zero* (i.e., not extremely small). In this case the multivariate null hypothesis can be *falsely* accepted, whereas the univariate null hypothesis will be *correctly* rejected somewhere. Thus, when the multivariate null hypothesis is accepted, a further test must be made as to whether the response is *significantly* small everywhere. When this null hypothesis is rejected (i.e., the true response will be non-zero somewhere), it is more informative to further test as to whether the response is significantly positive in some regions and negative in other regions. For these reasons, the present paper introduces joint confidence intervals which approximate the multivariate confidence region. In Section 2, a signal-to-noise problem in climate studies is discussed. In Sections 3 and 4, individual and joint confidence intervals are formulated and applied to a climate sensitivity study. Summary and remarks are given in Section 5. The Appendix gives an interpretation of confidence intervals.

2. Signal-to-noise problem

In this section, the definitions of climatic change, signal-to-noise ratio and statistical significance are clarified.

a. Definitions of climatic change

The climatic variable X is defined as the mean $\bar{x}$ or moment $\overline{x'y'}$ averaged over some time interval T of interest. X can be partitioned as

$$X = \langle X \rangle + X^*, \tag{2.1}$$

where $\langle X \rangle$ is the ensemble (true) mean and X^* the deviation.

The difference (Δ) between two climatic states X_1 and X_2 responding to different external conditions is partitioned as

$$\Delta = X_2 - X_1 , \tag{2.2a}$$

$$= \Delta\langle X \rangle + \Delta X^*. \tag{2.2b}$$

On the other hand, the ensemble mean of the time variance of the climatic variable X is partitioned as

$$\langle V(X) \rangle = V(\langle X \rangle) + \langle V(X^*) \rangle, \tag{2.3}$$

where $V(Z)$ is the time variance of Z defined by

$$V(Z) = \overline{(Z - \bar{Z})^2}. \tag{2.4}$$

In (2.2b) and (2.3) $\Delta\langle X \rangle$ and $V(\langle X \rangle)$ are externally caused and are called a "signal" (Leith, 1973) and "forced variation" (Lorenz, 1979), respectively. On the other hand ΔX^* and $\langle V(X^*) \rangle$ are mainly caused internally and are called "noise" (Leith, 1973) and "free variation"[2] (Lorenz, 1979), respectively. The changes in $\langle X^{*2} \rangle$ or $\langle V(X^*) \rangle$ are, however, caused externally.

It should be remarked that Leith (1973) referred to X^* (rather than ΔX^*) as "noise." This is because his definition of Δ is given by

$$\Delta = X_2 - \langle X_1 \rangle, \tag{2.5a}$$

$$= \Delta\langle X \rangle + X_2^* . \tag{2.5b}$$

This definition, however, is not practical when the ensemble mean $\langle X_1 \rangle$ is not known.

The forced variation is simulated by a statistical climate model (Budyko, 1969; Sellers, 1969) with a parameterized eddy flux, while the free variation due

[2] Even if X itself is thermally or orographically forced, the change in X can be due to a free variation of the basic state.

to a random eddy flux resulting from weather fluctuations is also simulated by a stochastic-dynamical climate model (Hasselmann, 1976). Both these variations are intrinsically deterministic[3] and are simulated by a general circulation model (Manabe and Stouffer, 1980; Manabe and Hahn, 1981). However, when the governing physical equations of free variations are now known, the free variations can be regarded as a stochastic process for convenience.

b. Definitions of true signal-to-noise ratio

The true signal-to-noise ratio was defined by Leith (1973) as

$$\langle r_1 \rangle = \Delta\langle X\rangle / \sigma(X_2). \tag{2.6}$$

This definition follows from his definition of climatic change (2.5a).

Alternatively, it is proposed to define the true signal-to-noise ratio which is consistent with the definition (2.2b) of climatic change as

$$\langle r_2 \rangle = \Delta\langle X\rangle / \sigma(\Delta X), \tag{2.7a}$$

$$\approx \langle r_1 \rangle / \sqrt{2}. \tag{2.7b}$$

In deriving (2.7b) from (2.7a), X_1 and X_2 are assumed to be uncorrelated and have equal variances as

$$\sigma^2(X_2 - X_1)$$

$$= \sigma^2(X_2) + \sigma^2(X_1) - 2\,\mathrm{Cov}(X_2, X_1), \tag{2.8a}$$

$$= 2\sigma^2(X_2). \tag{2.8b}$$

On the other hand, Madden and Ramanathan (1980) have defined the true signal-to-noise ratio of an N-year average as

$$\langle r_3 \rangle = \frac{\Delta\langle X\rangle}{2\sigma(\{X\})}, \tag{2.9a}$$

$$= \frac{\langle r_1 \rangle}{2/\sqrt{N}}. \tag{2.9b}$$

Leith (1973) showed that with the increase of the true signal-to-noise ratio $\langle r_1 \rangle$ from 0 to +1, the probability increases from 33 to 72% for the finite-time mean climate to be regarded as above normal (above 0.43 σ_T). If $\langle r_1 \rangle$ or $\langle r_2 \rangle \geqslant 1.0$, the forced climatic change significantly contributes to the change Δ of the finite-time mean climate defined by (2.5) or (2.2). If the signal and noise happen to cancel each other, the signal itself may not be detectable during this period. Nevertheless, the signal has contributed to reducing the change of finite-time mean climate. In order for the signal to always be detectable, the true signal-to-noise ratio must be *much* larger than 1. Even if the true signal-to-noise ratio is small for a monthly mean climate, it may be large for a seasonal mean climate. This ratio will be larger, if a climatic mean is defined as a several years' monthly or seasonal average as in (2.9). In the following sections $\langle r_2 \rangle$ defined by (2.7a) will be referred to as the true signal-to-noise ratio.

c. Definitions of statistical significance

The ensemble mean of the climatic variable is hypothetical, since identical earths do not exist. However, the ensemble mean $\langle X\rangle$ of January mean X, for example, can be estimated by taking a composite means $\{X\}$ of many January means, if X is stationary and ergodic with respect to year-to-year variation.

The statistical significance of the sample mean difference $\Delta\{\hat{X}\}$ can be defined in the following four different ways:

1) $\Delta\{\hat{X}\}$ is large enough to reject the null hypothesis that $\Delta\langle X\rangle$ is zero.
2) $\Delta\{\hat{X}\}$ is likely to be close to its true value.
3) The estimated signal-to-noise ratio is large enough for the unknown true signal-to-noise ratio to be likely to exceed 1.
4) $\Delta\{\hat{X}\}$ at different places are jointly significant.

The second definition is more informative than the first definition, as will be discussed in Section 3. The third definition is more stringent than the first. The fourth definition is more stringent than individual significance, as will be discussed in Section 4. These definitions should not be confused with each other.

3. Individual confidence intervals

It is assumed that X_1 and X_2 are two independent Gaussian[4] processes with unknown true variances $\sigma^2(X_1)$ and $\sigma^2(X_2)$. If X represents one January mean for example, $\{X\}$ represents a composite of N January means.

a. Definition of t

Student's t for the difference (Δ) of two N-sample means (X_1, X_2) is defined by

$$t = \frac{\Delta\{X\} - \Delta\langle X\rangle}{S}, \tag{3.1}$$

where $\{Z\}$ is an estimator of the true mean $\langle Z\rangle$ defined by

$$\{Z\} = \sum_{i=1}^{N} Z_i/N. \tag{3.2}$$

[3] If a numerical experiment is repeated with exactly the same initial condition, the same X^* is reproduced, provided that the external condition is the same.

[4] Even if x itself is not Gaussian, the time mean X can be approximately Gaussian due to the Central Limit Theorem (see Jenkins and Watts, 1968, p. 64).

The denominator S of (3.1) is an estimator of the standard deviation of the sample mean difference $\Delta\{X\}$ (see Panofsky and Brier, 1968, p. 63) and is given by

$$S^2 = s^2(X_1 - X_2)/N, \quad (3.3a)$$

$$\approx \{s^2(X_1) + s^2(X_2)\}/N, \quad (3.3b)$$

where s is an estimator of the standard deviation defined by

$$s^2(Z) = \sum_{i=1}^{N} (Z_i - \{Z\})^2/(N-1). \quad (3.4)$$

The degree of freedom of t is given (see Wadsworth and Bryan, 1960, p. 259) by

$$n = 2N - 2 \quad \text{for} \quad \sigma(X_1) = \sigma(X_2), \quad (3.5a)$$

$$= N - 1 \quad \text{for} \quad \sigma(X_1) \neq \sigma(X_2). \quad (3.5b)$$

b. Confidence interval of mean difference

The critical value $t_{n,\alpha/2}$ which is the upper 100 $\alpha/2$ percentage point of the t distribution is defined by

$$\text{Prob}\{|t| \leq t_{n,\alpha/2}\} = 1 - \alpha. \quad (3.6)$$

The $100(1-\alpha)\%$ confidence interval is defined as the range of different hypothesized values ($\Delta\mu$) of the true value $\Delta\langle X\rangle$ accepted by the null hypothesis $\Delta\langle X\rangle = \Delta\mu$ at the $1-\alpha$ confidence level for the given estimates $\Delta\{\hat{X}\}$ and $\hat{S}$.

Such $\Delta\mu$ must satisfy

$$|\Delta\{\hat{X}\} - \Delta\mu|/\hat{S} \leq t_{n,\alpha/2}. \quad (3.7)$$

Thus the range of $\Delta\mu$ is given explicitly by

$$\Delta\mu \lessgtr \Delta\{\hat{X}\} \pm d, \quad (3.8)$$

where $\pm d$ is the confidence interval defined by

$$d = \hat{S}t_{n,\alpha/2}. \quad (3.9)$$

If the true standard deviation is known or $N > 30$, Eq. (3.9) can be replaced by

$$d = \sigma(\Delta X)t_{\infty,\alpha/2}/\sqrt{N}, \quad (3.10)$$

where $t_{\infty,\alpha/2}$ is equivalent to the percentage point of the Gaussian distribution.

According to a popular interpretation, the confidence interval is the range of *unknown* true values which are distributed[5] around a particular estimate with the probability of $1-\alpha$. This interpretation is justified in the Appendix.

c. Reliability ratio

Dividing (3.1) for $\Delta\langle X\rangle = 0$ by $t_{n,\alpha/2}$ and inserting (3.9) gives

[5] The *known* true value is not distributed by definition.

$$R \equiv \hat{t}/t_{n,\alpha/2} = \Delta\{\hat{X}\}/d, \quad (3.11)$$

where $\hat{t} = \Delta\{\hat{X}\}/\hat{S}$.

The "reliability ratio" R defined by (3.11) is interpreted (see Fig. 2) as follows:

(a) When $|R| \gg 1$, $\Delta\langle X\rangle$ is close to its estimate.

(b) When $|R| > 1$, $\Delta\langle X\rangle$ is significantly different from zero.

(c) When $|R| = 1$, the confidence interval is equal to $\Delta\{\hat{X}\}$.

(d) When $|R| < 1$ and d is small, $\Delta\langle X\rangle$ is also small.

(e) When $|R| \lessdot 1$ and d is large, $\Delta\langle X\rangle$ is not necessarily small.

As an example of applications, the proposed significance tests were applied to a climatic sensitivity study (Manabe *et al.* 1981; Wetherald and Manabe, 1981) based on a sector general circulation model with an idealized geography.

Fig. 3a shows the latitude–month distribution of the change of the 8 year average of the monthly-zonal mean soil moisture with the increase of carbon dioxide (CO_2) by a factor of 4. In middle latitudes, a positive response occurs in the winter half-year, while a negative response occurs in the summer half-year. However, even an 8 year average may not completely exclude internally caused climatic change which can occur even without the CO_2 increase.

Fig. 3b shows the 90% reliability ratio R defined by (3.11). During winter in high latitudes where R is large (~3), the estimated positive response is fairly reliable (close to many years' average). During the midlatitude summer season, the estimated value of the negative response is not reliable, but significantly different from zero ($R \approx -1.5$). This summer dryness is one of the highlights of this sensitivity study.

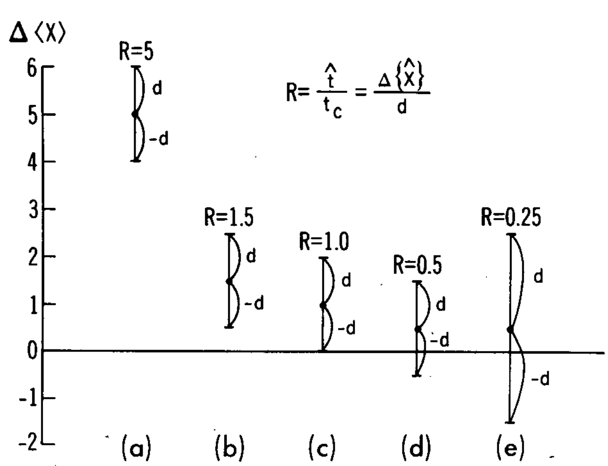

FIG. 2. Confidence intervals ($\pm d$) around the estimated mean difference. $R = \hat{t}/t_{n,\alpha/2} = \Delta\{\hat{X}\}/d$. If $|R| > 1$, the null hypothesis that the true difference is zero is rejected. If $|R| \gg 1$, the true value is close to its estimate (see text for details).

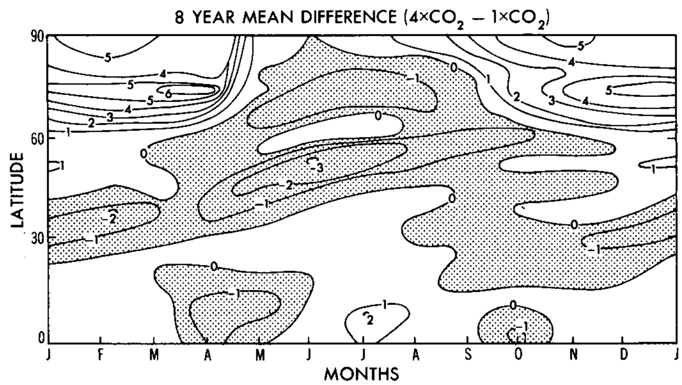

FIG. 3a. Latitude-month distribution of the difference ($4 \times CO_2 - 1 \times CO_2$) of the 8 year average of the monthly-zonal mean soil moisture (cm). Light shading denotes values < 0 (after Wetherald and Manabe, 1981).

Where $|R| < 1$, the true response is not significantly different from zero.

In order to examine whether or not the true response with $|R| < 1$ is small, Fig. 4 shows the latitude distribution of the July mean soil moisture difference with its 90% individual confidence limits given by (3.9). It is seen that the true response is *significantly* smaller than 1 cm around 40°.

d. Signal-to-noise ratio

As discussed in Section 2, the true signal-to-noise ratio $\langle r_2 \rangle$ is defined as

$$\langle r_2 \rangle = \Delta\langle X \rangle / \sigma(X), \tag{3.12a}$$

$$= \Delta\langle X \rangle [\sigma^2(X_1) + \sigma^2(X_2)]^{-1/2}. \tag{3.12b}$$

An estimated signal-to-noise ratio ($\hat{r}_2$) and its approximate confidence interval ($\hat{r}_2$) are given by dividing (3.8) by $\hat{S}\sqrt{N}$ and approximating $\langle r_2 \rangle$ by $\Delta\mu/(\hat{S}\sqrt{N})$ as

$$\langle r_2 \rangle \lesssim \hat{r}_2 \pm d_r, \tag{3.13}$$

where

$$\hat{r}_2 \equiv \hat{t}/\sqrt{N}, \tag{3.14}$$

$$d_r \equiv t_{n,\alpha/2}/\sqrt{N}. \tag{3.15}$$

It follows from (3.13) that

$$\text{if } |r_2| \gtrsim 1 \pm d_r, \quad \text{then} \quad |\langle r_2 \rangle| \gtrsim 1. \tag{3.16}$$

This criterion (3.16) is convenient in judging whether the true signal-to-noise ratio is significantly larger or smaller than 1. If the true standard deviation is known, $t_{n,\alpha/2}$ in (3.15) should be replaced by $t_{\infty,\alpha/2}$.

Fig. 5 shows the estimated signal-to-noise ratio $\hat{r}_2$ defined by (3.14) of the July mean soil moisture and its approximate 90% individual confidence intervals (3.15). According to this figure, it is not conclusive whether the true signal-to-noise ratio of the negative response in the midlatitude in July is larger or

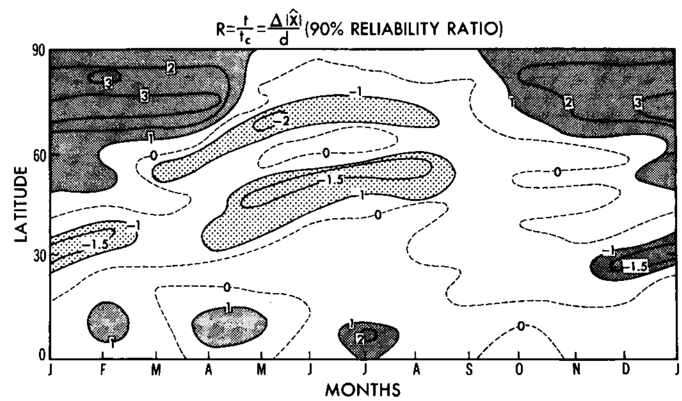

FIG. 3b. As in Fig. 3a, except for the 90% individual reliability ratio R. Dark shading >1, light shading <−1.

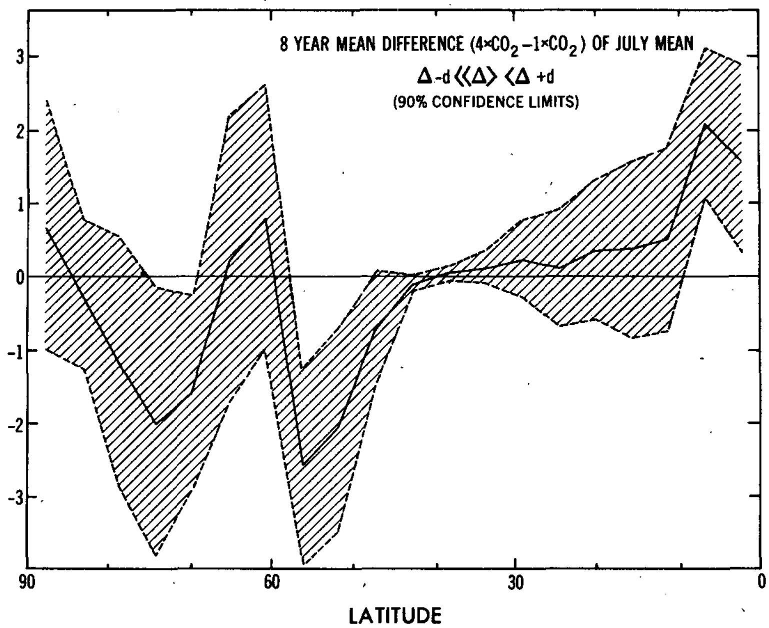

FIG. 4. Latitude distribution (solid line) of the difference ($4 \times CO_2 - 1 \times CO_2$) of July mean soil moisture (cm). The 90% individual confidence limits are indicated by dashed lines.

smaller than 1.0, although the signal itself is significantly different from zero. However, the signal-to-noise ratio in the subtropics is *significantly* smaller than 1 and there is little chance of floods or droughts as a result of the CO_2 increase.

e. Variance ratio

The $100(1 - \alpha)\%$ confidence limits of the ratio $\hat{F} = \hat{s}^2(X_2)/\hat{s}^2(X_1)$ between two estimated variances is given by the F distribution (see Spiegel, 1975, p. 197)

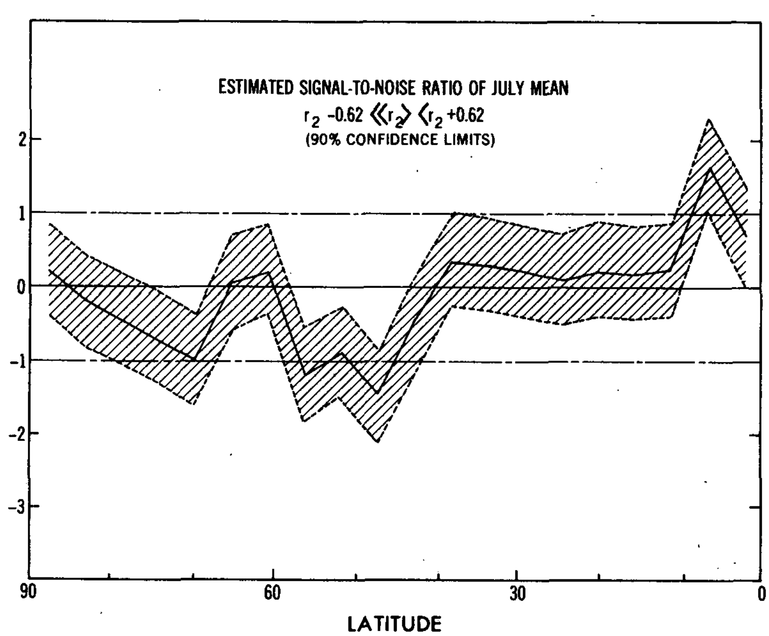

FIG. 5. Latitude distribution (solid line) of the estimated signal-to-noise ratio of July mean soil moisture. The 90% individual confidence limits are indicated by dashed lines.

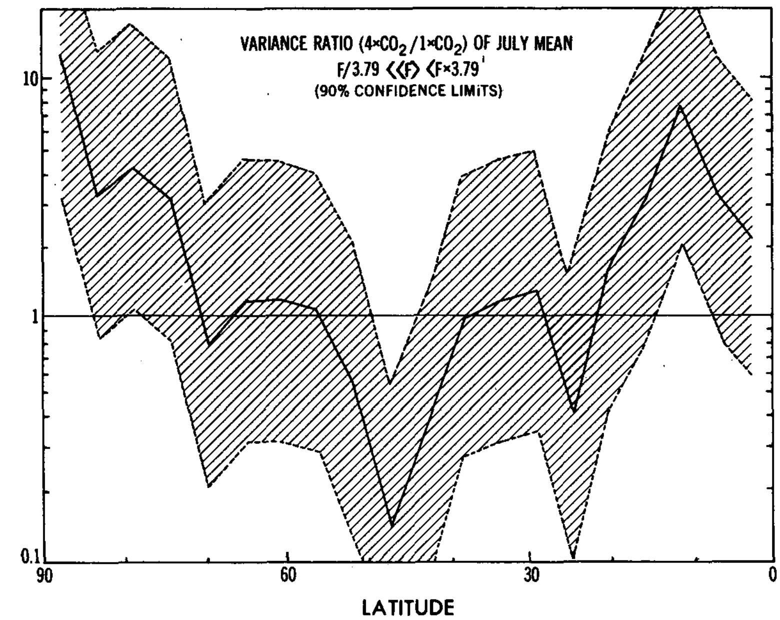

FIG. 6. Latitude distribution (solid line) of the ratio ($4 \times CO_2 - 1 \times CO_2$) of the variances of July mean soil moisture. The 90% individual confidence limits are indicated by dashed lines.

as

$$\hat{F}/F_{n,n,\alpha/2} \leqslant \langle F \rangle \leqslant \hat{F} \times F_{n,n,\alpha/2}\,. \qquad (3.17)$$

where $\langle F \rangle = \sigma^2(X_2)/\sigma^2(X_1)$ is the true variance ratio and $n = N - 1$. $F_{n,n,\alpha/2}$ and its reciprocal are the upper and lower $100\alpha/2$ percentage points, respectively.

It follows from (3.17) that

$$\left.\begin{array}{lll} \text{if} & \hat{F} > F_{n,n,\alpha/2}\,, & \text{then} \quad \langle F \rangle > 1 \\ \text{if} & \hat{F} < 1/F_{n,n,\alpha/2}\,, & \text{then} \quad \langle F \rangle < 1 \end{array}\right\}. \qquad (3.18)$$

This criterion (3.18) is convenient in judging whether the true variance ratio is significantly larger or smaller than 1.

Fig. 6 shows the latitude distribution of the estimated variances ratio $\hat{F}$ of the July mean soil moisture to measure the possible change $\langle F \rangle$ of the true variance with the CO_2 increase. The 90% individual confidence limits (3.17) indicate that the variance is not significantly altered, except for some latitudes (15, 45 and 90°). Thus, there is not a strong reason for adopting (3.5b) as the degree of freedom instead of (3.5a).

4. Joint confidence intervals

Even if the responses at different places are individually significant with some probability, the joint probability for the estimated response pattern to resemble the true response pattern decreases as independent spatial data increase. In order to test the joint significance of a response pattern, confidence intervals must be replaced by the confidence region (see Fig. 7) of multivariate space.

The $100(1-\alpha)$% confidence region (see Anderson, 1958 p. 108; Timm, 1975, p. 166; Kendall, 1975, p. 76) is defined as the range of the hypothesized true mean difference vector $\Delta\langle \mathbf{X} \rangle$ around its estimate $\Delta\{\hat{\mathbf{X}}\}$ accepted by a null hypothesis test based on the T^2 joint probability distribution. This region is represented by a multidimensional ellipsoid given by

$$T^2 = (\Delta\langle \mathbf{X} \rangle - \Delta\{\hat{\mathbf{X}}\})'\hat{\mathbf{S}}^{-1} \times (\Delta\langle \mathbf{X} \rangle - \Delta\{\hat{\mathbf{X}}\}) \leqslant T^2_{p,n,\alpha}\,, \qquad (4.1)$$

where $\hat{\mathbf{S}}$ is the estimated covariance matrix of $\Delta\mathbf{X}$ divided by the number of time data N and the prime denotes the transposed matrix.

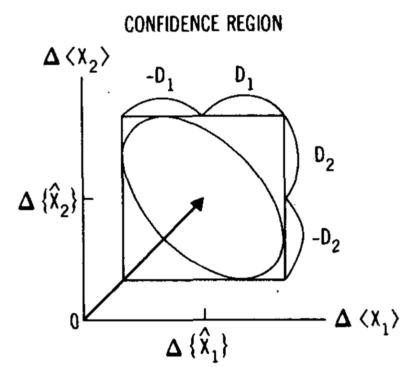

FIG. 7. Confidence region (ellipse) around the estimated mean difference vector. The confidence interval vector (D_1, D_2) consists of the widths of a box enclosing the confidence region.

TABLE 1. The 90% joint confidence limits of averaged latitudinal-seasonal distribution of the difference ($4 \times CO_2 - 1 \times CO_2$) of 8-year mean soil moisture (cm).

Latitude	October–March	April–September
60–90°	3.3 ± 2.2	0.3 ± 1.1
30–60°	−0.2 ± 1.1	−0.7 ± 0.8
0–30°	−0.1 ± 1.4	−0.7 ± 1.6

The critical value $T_{p,n,\alpha}$ is given by the T^2 distribution as

$$T^2_{p,n,\alpha} = \frac{np}{n-p+1} F_{p,n-p+1,\alpha}\,, \qquad (4.2)$$

where $F_{p,n-p+1,\alpha}$ is the upper $100\alpha/2$ percentage point of the F distribution; n is the degrees of freedom in time, and $n = 2N - 2$ for equal true covariance matrices of two means and $n = N - 1$ for non-equal covariances (see Anderson, 1958, p. 119); p is the effective number of spatially independent data, and $p \leq p_{max}$ where p_{max} is the number of statial data and should not exceed n. The values of $T^2_{p,n,\alpha}$ are tabulated in Timm (1975, p. 604). $T_{p,n,\alpha}$ increases with p and decreases with n.

For two variables $\Delta\mathbf{X} = (\Delta X_1, \Delta X_2)$ the confidence region takes the form of an ellipse (see Fig. 6) given by

$$\frac{1}{1-C_{12}^2}\left[\left(\frac{\Delta\langle X_1\rangle - \Delta\{\hat{X}_1\}}{S_1}\right)^2 + \left(\frac{\Delta\langle X_2\rangle - \Delta\{\hat{X}_2\}}{S_2}\right)^2 - 2C_{12}\left(\frac{\Delta\langle X_1\rangle - \Delta\{\hat{X}_1\}}{S_1}\right) \times \left(\frac{\Delta\langle X_2\rangle - \Delta\{\hat{X}_2\}}{S_2}\right)\right] \leq T^2_{p,n,\alpha}\,. \qquad (4.3)$$

If the origin of the coordinates $(\Delta\langle X_1\rangle, \Delta\langle X_2\rangle)$ is outside the confidence region, the null hypothesis that the true difference is zero everywhere is rejected. As spatial data increase, this null hypothesis becomes more difficult to reject, since the confidence region expands. Even if this null hypothesis is accepted, it does not follow that the true difference is likely to be zero everywhere, unless the confidence region is small.

a. Joint confidence intervals of mean difference

Since it is not possible to illustrate this region for many variables, it is proposed to replace this region by the "joint confidence intervals" D which are defined as the widths of multidimensional box enclosing this region. D is simply given by

$$D = \hat{S}T_{p,n,\alpha}\,. \qquad (4.4)$$

Although the probability level is a little larger than $1 - \alpha$, it is considered as $1 - \alpha$ for convenience. The joint confidence intervals derived here turn out to be a special case of "simultaneous confidence intervals" described in Timm (1975, p. 165).

In principle, the effective number p of spatially independent data can be determined as the degree of freedom of ρ^2 as

$$p = \frac{2\langle\rho^2\rangle^2}{\mathrm{Var}(\rho^2)}\,, \qquad (4.5)$$

where ρ^2 is distributed as χ^2 and defined (see Hasselmann 1979a,b; Lemke *et al.* 1980) as

$$\rho^2 = (\Delta\mathbf{X} - \Delta\langle\mathbf{X}\rangle)'\mathbf{C}^{-1}(\Delta\mathbf{X} - \Delta\langle\mathbf{X}\rangle)\,, \qquad (4.6)$$

where $\mathbf{C}$ is the true covariance of $\Delta\mathbf{X}$.

In practice, $\Delta\langle\mathbf{X}\rangle$, $\mathbf{C}$, $\langle\rho^2\rangle$ and $\mathrm{Var}(\rho^2)$ in the above expressions must be replaced by their estimates. If p is replaced by the number of all the spatial data for convenience, the joint confidence intervals are interpreted as the upper limit ($p = p_{max}$) of the true joint confidence intervals, while the individual confidence intervals are interpreted as the lower limit ($p = 1$).

In order to prove that the true response pattern of the soil moisture looks like the estimated pattern (Fig. 3a), the joint confidence intervals must be small. In order to reduce the effective number of spatial data and the joint confidence intervals, the responses (differences) of the model soil moisture have been averaged over three latitudinal bands as well as the winter (October–March) and summer (April–September) seasons and the standard deviation for each averaged value is estimated.

Table 1 shows the 90% joint confidence limits given by using (4.4) with $n = 14$ and $p = 6$, where the true covariances of the $1 \times CO_2$ and $4 \times CO_2$ experiments are assumed to be equal and p has been replaced by the number of the averaging segments. If p is overestimated, the joint confidence intervals are also overestimated. Based on Table 1, the joint null hypotheses listed in Table 2 are accepted. This means that it is jointly likely that the response is positive in the high-latitude winter and negative in the midlatitude summer and very small in the rest of the regions.

b. Joint confidence intervals of amplitude and phase

Since the conventional chi-square test for the power spectra is not applicable to deterministic oscillations, it is proposed to define the joint confidence intervals of the amplitude and phase of non-random Fourier components in the presence of random noise.

TABLE 2. Joint null hypotheses accepted by Table 1.

Latitude	October–March	April–September
60–90°	+	0
30–60°	0	−
0–30°	0	0

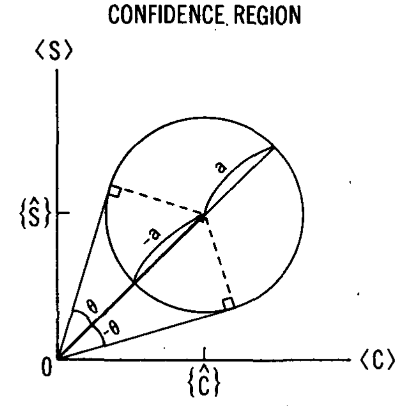

FIG. 8. Confidence region (circle) around the estimated mean vector ($\{\hat{c}\}$, $\{\hat{s}\}$) of the cosine and sine coefficients. The confidence intervals of the amplitude and phase are given by $\pm a$ and $\pm\theta$, respectively.

If the cosine and sine Fourier coefficients (c, s) are uncorrelated and have the same variance, their confidence region of these coefficients takes the form of a circle with the radius a given by

$$a = (\hat{S}_c\hat{S}_s)^{1/2}T_{2,n,\alpha}\,, \tag{4.7}$$

where $\hat{S}_c$ and $\hat{S}_s$ are the estimates of the standard deviation of the cosine and sine coefficients divided by $N^{1/2}$.

As illustrated by Fig. 8, the confidence intervals of the amplitude ($\pm a$) is given by the radius (a) and the phase ($\pm\theta$) is given by

$$\sin\theta = a(\{\hat{C}\}^2 + \{\hat{S}\}^2)^{-1/2}, \tag{4.8}$$

where $\{\hat{C}\}$ and $\{\hat{S}\}$ are the sample mean of the cosine and sine coefficients.

Fig. 9 shows the 90° confidence intervals of the amplitude (4.7) and phase (4.8) of seasonal cycles (1–5) of the soil moisture difference ($4 \times CO_2 - 1 \times CO_2$) at 72°N. The 8-year monthly mean data are subdivided into 8 sets and the 8-year mean and standard deviation of the cosine and sine coefficients of each year's data are estimated. It is seen that the annual cycle predominates and is significantly different from zero. The confidence limits of the amplitude and phase angle of the annual cycle are given by 4.1 ± 1.2 cm and 14 ± 17°, respectively. The joint confidence intervals of amplitude and phase are more appropriate than the standard deviations of amplitude and phase estimated by Angell and Korshover (1970).

5. Summary and remarks

The present study has clarified the definitions of climatic change, signal-to-noise ratio and statistical significance in order to interpret climatic statistics correctly.

It is proposed to test the significance of climatic statistics by use of the confidence intervals and reliability ratio. Confidence intervals are more informative than a single null hypothesis test, since they are defined as the range of hypothesized true values accepted by infinitely many null hypothesis tests. The confidence interval is commonly interpreted as the range of the unknown true values "distributed" around a particular estimate. The reliability ratio is defined as the ratio of an estimated mean difference to its confidence interval. When the reliability ratio is $\gg 1$, the estimate is close to its true value. When this ratio is < 1, the null hypothesis of the zero true value is not rejected. In this case the confidence in-

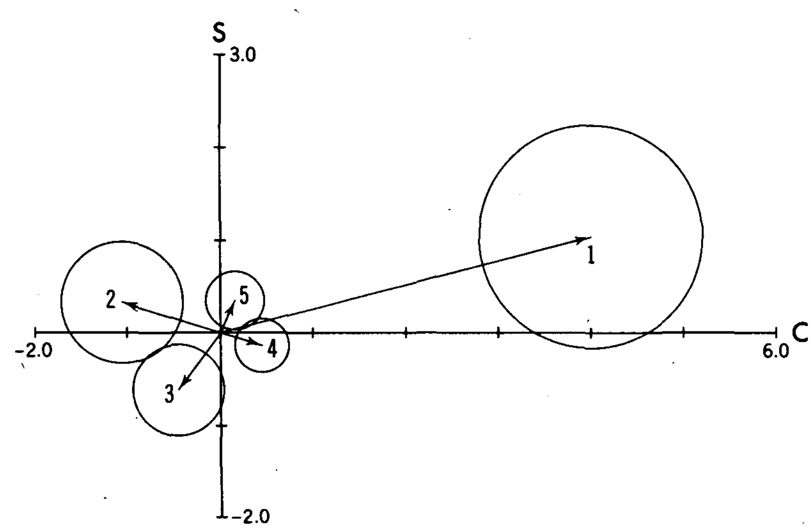

FIG. 9. The 90% confidence regions (circles) around the estimated (8-year mean) cosine (c) and sine (s) coefficients of seasonal cycles (1–5) of soil moisture difference ($4 \times CO_2 - 1 \times CO_2$) at 72°N. The length and direction of the vector (c, s) represent the amplitude and phase angle, respectively.

terval or the power of the test for a Type II error should be indicated to show whether or not the true value is *significantly* close to zero. Even if the signal is significantly different from zero, it may not be of climatic importance if the true signal-to-noise ratio is < 1. Unless the estimated signal-to-noise ratio is $\gg 1$, its confidence interval should be indicated to show whether or not the true ratio is > 1.

It is also proposed to make a multivariate test of a response pattern by the use of joint confidence intervals. When the null hypothesis of zero true response everywhere is rejected, the joint confidence intervals should be indicated to judge whether the response is significantly positive in some regions and negative in other regions. When this null hypothesis is accepted, the joint confidence intervals should be indicated to judge whether the response is *significantly* small everywhere. When dependent spatial data are regarded as independent, these intervals are overestimated.

Joint confidence intervals can also be applied to test the significance of the coefficients of a response pattern which is expanded with respect to such basis vectors (Hasselmann, 1979a) as to maximize the estimated multivariate signal-to-noise ratio. The joint confidence intervals of the amplitude and phase of seasonal cycles formulated in the present paper can be applied to test the significance of the response of atmospheric tides to a change in the excitation mechanism in the presence of noise. They can also be applied to the zonal Fourier coefficients of stationary planetary waves.

Finally, it should be stressed that the reliability of an estimated response pattern should be judged not only by the individual and joint significance tests, but also by its systematic and reasonable distribution and physical consistency.

Acknowledgments. The author is grateful to S. Manabe, R. T. Wetherald, P. Lemke, N. C. Lau and I. Held for valuable discussions and comments on the original manuscript. Prof. P. Bloomfield at the Department of Statistics of Princeton University kindly reviewed this paper. Thanks are extended to J. Kennedy for typing, P. Tunison for drafting and J. Connor for photographing.

APPENDIX

Interpretation of Confidence Intervals

Student's t for sample mean $\{x\}$ of an independent Gaussian variable x is defined as

$$t = \frac{\{x\} - \langle x \rangle}{s(x)/\sqrt{N}}, \tag{A1}$$

where $s(x)$ is an estimator of the standard deviation of x.

When the estimators $\{x\}$ and $s(x)$ are replaced by their known estimates[6] $\{\hat{x}\}$ and $\hat{s}(x)$, the unknown true values $\langle x \rangle$ and $\sigma(x)$ can be interpreted as "distributed" around their known estimates, although this interpretation is not found in standard text books.

In this case, t is also distributed as t as proven below. Eq. (A1) is rewritten (see Bendat and Piersol, 1971, p. 112) as

$$t = \frac{\{\hat{x}\} - \langle x \rangle}{\hat{s}(x)/\sqrt{N}}, \tag{A2a}$$

$$= \frac{(\{\hat{x}\} - \langle x \rangle)/\sigma(x)}{[x^2/(N-1)]^{1/2}}, \tag{A2b}$$

where

$$x^2 = \hat{s}^2(x)/\sigma^2(x), \tag{A3a}$$

$$= \sum_{n=1}^{N} (\hat{x}_i - \langle x \rangle)^2/\sigma^2(x) - (\{\hat{x}\} - \langle x \rangle)^2/\sigma^2(x). \tag{A3b}$$

The unknown true values of $[\{\hat{x}\} - \langle x \rangle]/\sigma(x)$ and x^2 are distributed as unit-Gaussian and chi-square, although the unknown true values of $\langle x \rangle$ and $\sigma(x)$ themselves are *not* distributed as Gaussian and chi-square, respectively. It follows that the above t is distributed as t, even if $\{x\}$ and $s(x)$ are replaced by their estimates.

This t distribution is given by

$$\mathrm{Prob}\{|t| \leq t_{n,\alpha/2}\} = 1 - \alpha. \tag{A4}$$

The confidence interval can be redefined as the range in which the unknown true value will fall with the probability of $1 - \alpha$.

This interval is determined from (A2a) and (A4) as

$$d = \hat{s}(x)/\sqrt{N} t_{n,\alpha/2}, \tag{A5}$$

which coincides with the confidence interval defined as the range of different hypothesized true values accepted by a null hypothesis test.

[6] An estimate is a realization of an estimator (random variable) and is not a random variable (see Jenkins and Watts, 1968, p. 93).

REFERENCES

Anderson, T. W., 1958: *An Introduction to Multivariate Statistical Analysis*. Wiley, 374 pp.

Angell, J. K., and J. Korshover, 1970: Quasi-biennial, annual, and semiannual zonal wind and temperature harmonic amplitudes and phases in the stratosphere and low mesosphere of the Northern Hemisphere. *J. Geophys. Res.*, **75**, 543–549.

Bendat, J. S., and A. G. Piersol, 1971: *Random Data: Analysis and Measurement Procedures*. Wiley-Interscience, 407 pp.

Budyko, M. I., 1969: The effect of solar radiation variations on the climate of the earth. *Tellus*, **21**, 611–619.

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.

——, and S. H. Schneider, 1976: On determining the statistical significance of climate experiments with general circulation models. *J. Atmos. Sci.*, **33,** 405–412.

——, W. M. Washington and S. H. Schneider, 1976: Testing the statistical significance of the response of the NCAR general circulation model to North Pacific ocean surface temperature anomalies. *J. Atmos. Sci.*, **33,** 413–423.

Hasselmann, K., 1976: Stochastic climate models. Part I: Theory. *Tellus*, **6,** 473–485.

——, 1979a: On the signal-to-noise problem in atmospheric response studies. *Meteorology of Tropical Oceans.*, Roy. Meteor. Soc., 251–259.

——, 1979b: Linear statistical models. *Dyn. Atmos. Oceans*, **3,** 501–521.

Holton, J. R., and H.-C. Tan, 1980: The influence of the equatorial quasi-biennial oscillation on the global circulation at 50 mb. *J. Atmos. Sci.*, **37,** 2200–2208.

Jenkins, G. M., and D. G. Watts, 1968: *Spectral Analysis and Its Applications.* Holden-Day, 525 pp.

Jones, R. H., 1975: Estimating the variance of time averages. *J. Appl. Meteor.*, **14,** 159–163.

——, 1976: On estimating the variance of time averages. *J. Appl. Meteor.*, **15,** 514–515.

Julian, P. R., and R. M. Chervin, 1978: A study of the Southern Oscillation and Walker Circulation phenomenon. *Mon. Wea. Rev.*, **106,** 1433–1451.

Kendall, M., 1975: *Multivariate Analysis.* Charles Griffin, 210 pp.

Keshavamurty, R. N., 1982: Response of the atmosphere to sea surface temperature anomalies over the equatorial Pacific and the teleconnections of the Southern Oscillation. *J. Atmos. Sci.*, **39,** 1241–1259.

Laurmann, J. A., and W. L. Gates, 1977: Statistical considerations in the evaluation of climatic experiments with atmospheric general circulation models. *J. Atmos. Sci.*, **34,** 1187–1199.

Leith, C. E., 1973: The standard error of time-averaged estimates of climatic means. *J. Appl. Meteor.*, **12,** 1066–1069.

Lemke, P., E. W. Trinkel and K. Hasselmann, 1980: Stochastic dynamic analysis of polar sea ice variability. *J. Phys. Oceanogr.*, **10,** 2100–2120.

Lorenz, E. N., 1979: Forced and free variations of weather and climate. *J. Atmos. Sci.*, **36,** 1367–1376.

Madden, R. A., and V. Ramanathan, 1980: Detecting climate change due to increasing carbon dioxide. *Science*, **209,** 763–768.

Manabe, S., and D. G. Hahn, 1977: Simulation of the tropical climate of an ice age. *J. Geophy. Res.*, **82,** 3889–3911.

——, and ——, 1981: Simulation of atmospheric variability. *Mon. Wea. Rev.*, **109,** 2260–2286.

——, and R. J. Stouffer, 1980: Sensitivity of a global climate model to an increase of CO_2 concentration in the atmosphere. *J. Geophys. Res.*, **85,** 5529–5554.

——, R. T. Wetherald and R. J. Stouffer, 1981: Summer dryness due to an increase of atmospheric CO_2 concentration. *Climate Change*, **3,** 347–386.

Mitchell, J. M., Jr. 1971: Climatic change. WMO Tech. Note, No. 79, 79 pp.

Munk, W. H., 1960: Smoothing and persistence. *J. Meteor.*, **17,** 92–94.

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

Sellers, W. D., 1969: A global climate model based on the energy balance of the earth-atmosphere system. *J. Appl. Meteor.*, **8,** 392–400.

Shukla, J., 1975: Effects of Arabian sea-surface temperature anomaly on Indian summer monsoon: A numerical experiment with the GFDL model. *J. Atmos. Sci.*, **32,** 503–511.

Spiegel, M. R., 1975: *Probability and Statistics.* McGraw-Hill, 372 pp.

Storch, H. V., 1982a: A remark on Chervin-Schneider's algorithm to test significance of climate experiments with GCM's. *J. Atmos. Sci.*, **39,** 187–189.

——, 1982b: Comparison of a sequence of model generated 500 mb topographies with climate. *Tellus*, **34,** 89–91.

Timm, N. H., 1975: *Multivariate Analysis with Applications in Education and Psychology.* Brooks/Cole, 689 pp.

van Loon, H., and J. C. Rogers, 1978: The see-saw in winter temperatures between Greenland and Northern Europe. Part I: General description. *J. Atmos. Sci.*, **106,** 296–310.

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

Warshaw, M., and R. R. Rapp, 1973: An experiment on the sensitivity of a global circulation model. *J. Appl. Meteor.*, **12,** 43–49.

Washington, W. M., R. M. Chervin and G. V. Rao, 1977: Effects of a variety of Indian Ocean surface temperature anomaly patterns on the summer monsoon circulation: Experiments with the NCAR general circulation model. *Pure Appl. Geophys.*, **115,** 1335–1356.

Wetherald, R. T., and S. Manabe, 1981: Influence of seasonal variation upon the sensitivity of a model climate. *J. Geophys. Res.*, **86,** 1194–1204.