Theory of Optimal Weighting of Data to Detect Climatic Change

THOMAS L. BELL

Laboratory for Atmospheres, Goddard Space Flight Center, Greenbelt, MD 20771

(Manuscript received 21 August 1985, in final form 18 February 1986)

ABSTRACT

A search for climatic change predicted by climate models can easily yield unconvincing results because of "climatic noise," the inherent, unpredictable variability of time-averaged atmospheric data. We describe a weighted average of data that maximizes the probability of detecting predicted climatic change. To obtain the optimal weights, an estimate of the covariance matrix of the data from a prior data set is needed. This introduces additional sampling error into the method. We show how to take this into account. A form of the weighted average is found whose probability distribution is independent of the true (but unknown) covariance statistics of the data and of the climate model prediction. A table of critical values for statistical testing of the weighted average is given, based on Monte Carlo calculations. The results are exact when the prior data set consists of temporally uncorrelated samples.

1. Introduction

Weather prediction, seasonal forecasting, and prediction of climatic change have as a goal foreknowledge of the atmosphere's future states. To this end, models are built reflecting as much as possible our understanding of the processes that govern the atmosphere's behavior and its interactions with land and ocean. Our understanding is tested by comparing the model predictions to what the atmosphere actually does. The nature of the comparison and the degree of success expected vary considerably, depending on how far into the future we try to predict.

General circulation models, on which weather forecasts are based, attempt to produce a picture of the atmosphere accurate on a spatial scale of hundreds of kilometers and a time scale of hours. The forecasts are sensitive to knowledge of the initial state of the atmosphere, and a forecast accurate out to a week is still cause for celebration. Verification of a weather forecast is based on a comparison of the forecast with the observed state of the atmosphere on a point by point basis. The forecast and the real atmospheric state are expected to agree with one another much better than two randomly selected states of the atmosphere.

Seasonal forecasts do not pretend to predict the future state of the atmosphere day by day, but attempt instead to say something about the probability distribution of the weather during a season. The forecast is less dependent on initial conditions of the atmosphere, though Shukla (1981) has offered some evidence that the atmosphere may retain information about its initial state out to a month or two. Much of the predictability of the atmosphere out to a season or more is expected to originate in the slowly varying "boundary conditions" such as the state of the oceans, soil moisture, snow cover, sea ice, and processes affecting radiative properties of the atmosphere. A seasonal forecast attempts to improve our odds of being right about the average weather during a season over what we can already guess from climatology. However, a single seasonal forecast, at least for the present, offers such wide latitudes in its predictions of what could happen that a history of many forecasts must be built up before a fair idea can be obtained of whether the forecasts are providing any extra information. Evaluation of seasonal forecasts is therefore a subtler problem than evaluating weather forecasts.

Climate forecasts are based on changes in boundary conditions, and they predict changes in the probability distribution of future weather. Initial conditions of the atmosphere are generally ignored in making these forecasts. Verification of climate forecasts has all of the problems that verification of seasonal forecasts has, aggravated by the scarcity of cases where climatic change is big enough to detect. We shall be primarily interested here in the problem of detecting climatic change when a model prediction is available to suggest what to look for and when the expected change, or "signal," is in danger of being obscured by the natural variability of time-averaged atmospheric data, the "climatic noise" characterized by Leith (1973). One of the best known problems of this sort is that of trying to detect the warming of the earth's surface believed to be occurring due to the increasing CO_2 and other trace gases in the atmosphere. Many decades from now the warming is expected to be so great that sophisticated statistical analyses of data will not be necessary to show it. However, we would like to see the first signs of it as soon as possible in order to begin checking our understanding of the process.

It will be helpful for the discussion that follows to recall briefly some of the probabilistic concepts associated with predicting atmospheric behavior. A more thorough treatment of these ideas may be found in the excellent discussions by Lorenz (1969) and Leith (1971, 1974).

Let us denote by $\xi(t)$ the set of fields completely describing the state of the atmosphere at a given time t. Although a great deal of effort is expended in gathering data about the atmosphere from all over the world, the data we collect are not perfectly accurate—measurements are never perfect—and in any case we are missing details of the state of the atmosphere too fine for our observational net to catch. We shall denote by $\mathbf{x}$ the vector of variables describing the atmosphere that can be obtained from our observing system. The variables $\mathbf{x}$ might, for instance, consist of temperature, moisture, and winds at the grid points of a general circulation model. The variables $\mathbf{x}$ tell us about the state of the atmosphere ξ at the grid points of the model (but with errors due to the inaccuracies in the observing system); they do not specify what happens in between. The variables evolve in time. Models of the atmosphere try to predict this evolution, based on the initial conditions $\mathbf{x}_0 \equiv \mathbf{x}(t_0)$ and, in the case of long-range forecasts, data about the state of the boundary conditions, like sea surface temperature, which we denote by the vector $\mathbf{b}_0$.

When a forecast is attempted at a given time t_0, many different initial states ξ of the atmosphere are consistent with what we know about the atmosphere from the observing network, $\mathbf{x}_0$. We can at best specify a probability distribution for the initial state $\xi(t_0)$. All of these possible initial states $\xi(t_0)$ would be effectively indistinguishable to our observing network, giving observations equal to $\mathbf{x}_0$ to within the errors of the measurement and data analysis scheme.

The dynamics of the atmosphere are unstable; there is good reason to believe that two nearly identical states of the atmosphere will, in general, evolve within a few weeks to states that appear to the casual observer virtually unrelated to each other. Differences initially confined to small spatial scales eventually show up on all scales. This is what makes weather prediction challenging. The dynamics of the atmosphere dictate how each of the possible states ξ evolves in time. The probability distribution for ξ spreads out due to instabilities in the dynamics. The observations $\mathbf{x}$ that each possible state would generate likewise diverge from each other with time.

We therefore introduce the probability distribution $P[\mathbf{x}(t)|\mathbf{x}_0, \mathbf{b}_0, t_0]$ for the possible observations $\mathbf{x}$ of the atmosphere that could follow from what we know about the initial state of the atmosphere, $\mathbf{x}_0$, and boundary conditions, $\mathbf{b}_0$. The very best model that we could hope to construct would predict P. We are far from being able to do that.

Models tell us something about $P[\mathbf{x}(t)|\mathbf{x}_0, \mathbf{b}_0, t_0]$. Exactly what they tell us depends on their complexity and on how far into the future they try to predict. Weather forecasting, for which $t - t_0$ is less than a week, integrates the dynamical equations of the atmosphere forward in time starting from $\mathbf{x}_0$, produces a single forecast $\mathbf{x}(t)$, and makes no attempt to estimate the uncertainty in the forecast. It is up to the individual forecaster to judge how much weight to give the numerical forecast based on his prior experience with the model. For forecasts of a few days the probability distribution remains comparatively narrow, and a single forecast is useful. Information about the width of the probability distribution P may be obtained with "Monte Carlo" forecasting (Leith, 1974; see also Epstein, 1969), in which multiple forecasts are made starting from many different initial conditions near $\mathbf{x}_0$. A variation on this approach recently investigated by Hoffman and Kalnay (1983) may make this method operationally feasible. (See also Dalcher et al., 1985.)

Because the probability distribution P is relatively narrow for the time scales of weather forecasting, comparisons of model predictions with atmospheric data can be quite specific. When there is a major discrepancy, say, between the 24-hour forecast of the 500 mb geopotential height over the central United States and the observed height field at that time, the modeler is as inclined to examine the parameters of his model as he is to ascribe the discrepancy to inadequate knowledge of the initial conditions.

As we move to prediction times $t - t_0$ characteristic of seasonal and climatic forecasting, $P[\mathbf{x}(t)|\mathbf{x}_0, \mathbf{b}_0, t_0]$ is becoming quite broad, comparable in fact to the climatic probability distribution typical of the season at the time t. We shall denote this climatic probability distribution by $P_0(\mathbf{x}, t)$. It represents the probability distribution of observable atmospheric variables, known to a certain extent from historical records. The explicit dependence of P_0 on time t is retained to allow for diurnal or seasonal variation of the climate. If our models were perfect, and we could run the models starting from a range of initial conditions and boundary conditions typical of the present-day earth, we would find that the predictions $P[\mathbf{x}(t)|\mathbf{x}_0, \mathbf{b}_0, t_0]$ average out to $P_0(\mathbf{x}, t)$; i.e., the model climate equals the observed climate. Seasonal forecasting tries to provide predictions P narrower than P_0 by taking advantage of the extra information $\mathbf{x}_0$, $\mathbf{b}_0$.

To compare the predictions P with actual atmospheric behavior is a more difficult task than is the case for weather predictions, because we can no longer afford to ignore the probabilistic aspect of seasonal and climatic forecasts. In fact, to make any progress at all we must begin to limit the number of variables to be studied. One reason for this has been described by Hasselmann (1979). He considers as an example (which

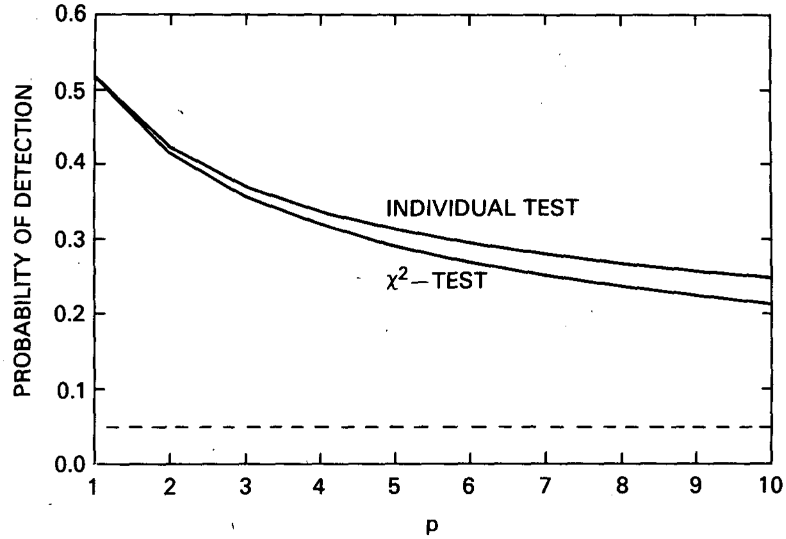

FIG. 1. The probability of detecting climatic change in p uncorrelated Gaussian random variables z_i, $i = 1, \cdots, p$, using two methods. Before the climatic change, all variables have zero mean and variance σ^2. After the climatic change, variable z_1 has mean 2σ; the remaining variables have mean zero. The first detection method checks whether the χ^2 variable $\sum z_i^2$ is significantly different from 0. The second detection method checks each variable individually to see if it is significantly different from zero. The proper way to conduct such a test is described by Livezey and Chen (1983). Both tests are conducted at the 0.05 significance level. The dashed line shows the probability of obtaining a (false) positive result if there were no climatic change.

we modify slightly) a case where p Gaussian random variables z_i $(i = 1, \cdots, p)$ vary independently of each other with mean 0 and variance σ^2. Suppose that the mean of the first variable z_1 has shifted by 2σ (we can always change to a new set of variables so that this is the case) and that we try to detect this "climatic change" in the system with a single observation of the variables **z**. One way to look for a change is to check whether the quantity $\sum_i z_i^2$, which is a measure of the distance of **z** from its (former) mean **0**, is unusually large. The range of values it would have if there were no change in the mean (which we shall refer to as the acceptance interval[1]) can be looked up in probability tables for a chi-squared variable with p degrees of freedom. A change in the mean is identified with confidence only if $\sum_i z_i^2$ lies outside this range.

We have plotted in the lower curve in Fig. 1 the probability that, if the mean of z_1 were actually 2σ instead of 0, $\sum_i z_i^2$ would fall outside the acceptance interval (at the 95% confidence level) for a χ^2 variable with p degrees of freedom. This is the probability that we would detect climatic change. Note that even when variable z_1 is tested alone ($p = 1$), there is still a 48% chance of failing to detect its climatic change at the 95% confidence level. As the number of variables tested increases, the probability of detection decreases. This is because the acceptance interval for the χ^2 test increases with the number of variables to take into account the increasing chance of large random excursions of some linear combination of the variables.

Another way we might look for a climate change in this example is to examine each variable individually to see if it is unusually far from its mean 0. Any one variable should fall within $0 \pm 1.96\sigma$ 95% of the time, if there were no climatic change. However, with p variables we must move the boundaries of the acceptance interval to values larger than $\pm 1.96\sigma$ so that the probability of *all p* variables lying inside their respective acceptance intervals is 95%. Livezey and Chen (1983) give a detailed discussion of this. If we then calculate how probable it is that we detect a climate change when the mean of z_1 shifts by 2σ, we obtain the upper curve in the plot shown in Fig. 1. Note again the decrease with increasing number of variables, although our chances of succeeding are better than with the χ^2 test. The reason for this improvement is that by testing each variable individually we are favoring detection of climatic change occurring in one variable alone (as hap-

[1] When $\sum_i z_i^2$ falls within the "acceptance interval," we need not conclude that we must accept the hypothesis that the climate is unchanged, as Hayashi (1982) has emphasized; only that, with the evidence at hand, we have no reason to reject the hypothesis (at the specified confidence level). This caveat applies throughout this paper to the use of the term *acceptance interval*.

pens in our example), whereas with the χ^2 test we are testing for a shift in the climate involving any linear combination of variables.

As this example illustrates, testing for a difference in the probability distribution P from P_0 must be done judiciously. If we examine too many variables we risk being unable to say anything with much confidence. In the example above it is obviously wisest to choose to concentrate on the first variable z_1. Part of the research effort in seasonal and climate forecasting is devoted to discovering what "z_1" is; that is, what variables are most likely to reveal significant differences from climatological behavior. These variables tend to describe time-averaged atmospheric behavior on large spatial scales, because dynamical instability of the atmosphere reduces the predictability fastest of shorter time scale, smaller spatial scale variations of the atmosphere. Moreover, the information we have about the boundary conditions $\mathbf{b}_0$ is usually for the more slowly varying components.

The approaches to verifying seasonal and climatic forecasts differ somewhat because of the different nature of the forecasts. Seasonal forecasts generally do not predict shifts in the mean of the probability distribution P away from the climatic mean more than one standard deviation. A forecast might indicate increased probability of higher seasonal temperatures, say, but nothing that had not been experienced several times in the previous 20 years. Verifying such a forecast would tend toward showing that the variance of the actual behavior of variables from the predicted behavior for the season is less than it would have been for a forecast based solely on climatic means. The test described by Preisendorfer and Mobley (1984) is in this spirit.

Predictions of climatic change, on the other hand, are usually based on single events, such as a change in the radiative properties of the atmosphere or characteristics of the surface. We cannot easily average over successive forecasts, as we can with seasonal forecasts, in order to make the comparison of model and data statistically stronger. To compare a model prediction of climatic change with what actually happens, we must be able to find a set of variables for which the individual climatic event is large enough to be detected over the natural variability.

The numerical example given above suggests that the fewer the number of variables we initially limit ourselves to the more likely we are to be successful. In that example it is obviously best to look for climatic change in the first variable alone. The choice is not so obvious when we look for climatic change in real data. Models typically predict small changes spread over the entire globe, and the best variable to look for evidence of the predicted change is some sort of global average of the data. Once a variable is chosen, methods described, for example, by Epstein (1982) or Katz (1982) can be explored for identifying climatic change in the variables. Livezey (1985) gives a useful review of recent work on statistical methods for evaluating climatic anomalies in general circulation model data.

It is the purpose of this paper to discuss a method of forming a weighted average of global data that maximizes the signal-to-noise ratio for a given climatic change. The method is closely related to a technique described by Hasselmann (1979) for discovering the pattern of climatic response to external forcing, in which a small number of "guesses" about the nature of the climatic response are statistically modified to obtain from observed anomalies as statistically significant an estimate of the pattern of climatic response as possible. Here the emphasis is more on comparing a single "guess" for the climatic response—a climate model prediction, for instance—to atmospheric data in a way that produces as statistically strong a conclusion as possible. Hannoschöck and Frankignoul (1985) describe an interesting attempt to implement Hasselmann's (1979) ideas with a climate model, but remark that the sampling theory for these techniques is not well developed. Some progress with the sampling theory described here for the optimal weighting method may be directly applicable to the problems encountered in using Hasselmann's (1979) approach.

The optimal weighting method was described in an earlier paper (Bell, 1982), and the reader may find helpful the derivation given there of the optimal weights. An alternative derivation using probabilistic arguments is given here, but the bulk of the present paper is devoted to making quantitative estimates of the effect on the method of having only a finite amount of data to work with when using the method.

In section 2 which follows we shall show that a properly weighted average of data maximizes the probability of detecting a change in the climate when 1) a model prediction of the change is available to guide in the choice of the weighting and 2) sufficient prior data are available to obtain good estimates of the covariance statistics of the variables included in the average. A critical value for the weighted average can be established such that if the weighted average of the data exceeds this critical value, then we can conclude that the climate has changed. In addition, the weighted average of the data can be tested to see whether it falls within an acceptance interval centered around the climate model prediction for the most probable value of the weighted average. [See Eq. (2.27).] If the climate model were incorrect and the climate had in fact not changed, this test would maximize one's chances of detecting the discrepancy. The weighted average picks out from among the variables for which there are data the "direction" in which the data are most likely, given the climatic noise, to show a significant difference depending on whether the climate conforms to the new climate predicted by the model or to the climate prior to the change. More detailed comparisons of the model with the data are of course desirable, but will be hampered by the increasing likelihood that differences between the two can be attributed simply to natural variability.

In section 3 we relax the assumption that unlimited prior data are available. The distribution theory for this problem is relatively undeveloped in the statistical literature, so far as we can tell, and so it is developed here. We show that with a finite amount of prior data the acceptance interval for detecting climatic change must be enlarged, and the chances of detecting climatic change are correspondingly reduced. It is shown that when data for p variables are used to form a weighted average, and N independent samples of prior data are used to estimate the covariance matrix of the variables, then the acceptance interval for detecting climatic change with optimally averaged data must be increased by approximately a factor $n/(n - p)$ (where $n = N - 1$) over what it would be if we had unlimited prior data. The approximation is quite accurate for most purposes.

In section 4 we show that because of the effect of the finite prior data set on the statistics of optimally weighted averages, the advantages of the optimal weighting can be nullified by including too many independently weighted variables in the average. A method for choosing how many variables to keep is described. Section 5 provides a summary of the optimal weighting approach and some conclusions. Two Appendices give some details of the statistical sampling theory calculations.

2. Optimal weighting of data to detect a predicted change in climate

a. Philosophy

We describe here a probabilistic approach to the derivation of the optimal weighting of data to detect a change in the climate. We assume that a climate model prediction is available to guide in the choice of the weighting, and that sufficient data are available, or a sufficiently advanced model, so that the covariance statistics are accurately known for the variables of interest. This approach allows us to investigate the effect on the weighting of uncertainties in the predicted climate change or of a change in the covariance statistics as the climate changes.

The philosophy behind the approach described here is not the only one that leads to the statistical quantity proposed at the end of this section to compare actual data to a climate model prediction. The problem can also be formulated in terms of classifying atmospheric data into one of two populations: one that is characteristic of the undisturbed climate and one that is characteristic of the altered climate predicted by the model. Both approaches lead to a quantity known in the statistical literature as the *discriminant function* introduced by Fisher (1936), a discussion of which may be found in Anderson (1958). Some applications of multiple discriminant analysis for meteorological problems have been described by Miller (1962).

Suppose that a climate model predicts climatic change over a certain span of time t_1 to t_2. For concreteness we may take this period to be 1 yr long. Data from that period are collected and some averaging of the data over seasons and geographical areas is done. (One might, for instance, average the station data over latitudinal zones.) The result would be values for p variables which we shall denote by Y_i, $i = 1, \cdots, p$, representing, for example, seasonally and zonally averaged surface temperature for the period t_1 to t_2.

In the absence of a model prediction, we expect the observations $\mathbf{Y}$ to occur with probability described by the climatic distribution $P_0(\mathbf{y})$, introduced in the previous section.[2] The climate model supplies us with an alternative probability distribution $P(\mathbf{y})$. We are therefore presented with two hypotheses $\mathscr{H}_0$ and $\mathscr{H}_1$ concerning the data $\mathbf{Y}$:

$\mathscr{H}_0$: the data $\mathbf{Y}$ are drawn from a population described by the climatic distribution $P_0(\mathbf{y})$;

$\mathscr{H}_1$: the data $\mathbf{Y}$ are drawn from a population described by the climate model prediction $P(\mathbf{y})$.

Our task, then, is to find evidence for or against one of the hypotheses by devising a suitable test of the data $\mathbf{Y}$. In formulating this test we shall make certain assumptions about the nature of this problem:

(i) The probability distributions P_0 and P are Gaussian. Since the data Y_i themselves represent averages over many measurements, this is a reasonable assumption, but it should be verified for the particular choice of variables. Some examples of such an examination of meteorological quantities may be found in Parthasarathy and Mooley (1978) and White (1980).

(ii) The probability distribution $P(\mathbf{y})$ is about as broad as $P_0(\mathbf{y})$, at least on global scales, and so it is essentially the shift in the mean that defines the change from P_0 to P. This differs from the case for verification of a seasonal forecast, where some additional help is to be gained from the narrowing of the distribution P relative to P_0.

(iii) The shift in the mean is not very strong, so we are advised to test for the change in the mean for only a few variables or risk being unable to say anything with much confidence, as was discussed in the Introduction. In order to maximize our chances of success, we shall reduce the problem to testing a single variable.

(iv) We have only one chance to make the test since the events producing climatic change generally do not repeat themselves: carbon dioxide levels will rise only

[2] We shall use upper case letters to denote quantities obtained from the data and lower case letters to denote characteristics of an underlying statistic population.

once, volcanos rarely inject aerosols into the stratosphere with the same spatial distribution, etc. In this respect also the problem we face differs from that of verifying seasonal forecasts, since each new season offers another opportunity to improve the statistics of the comparison of model and data.

The test we shall propose for the data $\mathbf{Y}$ will be designed to maximize our chances of rejecting confidently one of the hypotheses $\mathscr{H}_0$ or $\mathscr{H}_1$. As Hayashi (1982) has emphasized, it is in the nature of testing hypotheses about probability distributions that one can never be sure whether a hypothesis is correct. To show that a hypothesis is correct requires an infinite amount of data. One can only construct tests, based on experience or on physical insight into the problem, that are as likely as possible to reveal dissonance between the hypothesis and the data, if there is any.

A necessary condition for successfully discriminating between the two hypotheses is that the two probability distributions P_0 and P differ enough that if one of the hypotheses is correct then there is a reasonable chance that the other can be rejected. A climate model might in principle predict that the climate will not change much, and so offer no guidance how best to test the prediction. One then has no choice but to "go fishing" in the data to look for climatic anomalies that are not predicted by the climate model. Our approach can offer no help there. However, if the climate model predicts a change in the climate, then our approach can provide as definitive a test of the prediction as is possible given the size of the data set with which we have to work.

Let us form a weighted average of the data $\mathbf{Y}$ (a vector of p values):

$$A = \sum_{i=1}^{p} w_i Y_i = \mathbf{w}'\mathbf{Y}, \tag{2.1}$$

where primes indicate matrix transpose. The weights $\mathbf{w}$ are unspecified as yet. The probability $p(a)da$ that A will fall within a neighborhood da of a is sketched in Fig. 2 for the two hypotheses $\mathscr{H}_0$ and $\mathscr{H}_1$. It is convenient to assume the mean of the distribution $P_0(\mathbf{y})$ to be zero, as we have done in the figure. The separation between the two probability distributions and their widths depend on the weights $\mathbf{w}$.

The critical value a_c, the boundary of the acceptance interval for a test of hypothesis $\mathscr{H}_0$ at the 95% confidence level, is marked on Fig. 2. Hypothesis $\mathscr{H}_0$ is rejected (with 95% confidence) if $A > a_c$. If at the same time the value of A falls within the acceptance interval (at the 95% confidence level) appropriate to $\mathscr{H}_1$ (indicated by vertical dashed lines in Fig. 2), then we can conclude (with 95% confidence) that the climate *has* changed and that the new climate is consistent with the climate model prediction (insofar as we have tested it).

The more the two probability distributions overlap, the more likely is it that the value of A will be of little help in discriminating between the two hypotheses. We shall take the point of view that to verify a prediction of climatic *change* one needs first to find convincing evidence in the data that the climate has indeed changed, and in the direction indicated by the climate model. Otherwise any agreement between the data and the climate model prediction is open to the objection that it could be entirely fortuitous and have nothing to do with the physical mechanisms supposedly causing

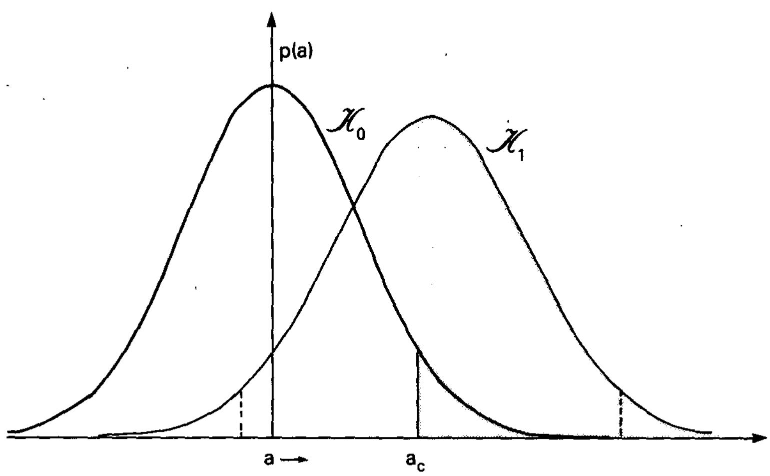

FIG. 2. The probability densities of the weighted average Eq. (2.1) under two hypotheses: $\mathscr{H}_0$, that the climate is unchanged; and $\mathscr{H}_1$, that the climate model prediction is correct. Climate change at the 5% significance level is indicated by values of $A > a_c$. The acceptance interval (at the 95% confidence level) of the climate model prediction is delimited by the vertical dashed lines.

the predicted change, since the model has presumably been tuned to reproduce the climatic situation before the climatic change.

b. Derivation of weights $\mathbf{w}$

We turn now to the problem of obtaining convincing evidence that the climate has changed, based on data $\mathbf{Y}$ for p variables. We shall return later to the problem of verifying that the climate change is consistent with the model. Our evidence for climate change will be a value of A, defined in Eq. (2.1), that exceeds the critical value a_c illustrated in Fig. 2. If hypothesis $\mathscr{H}_1$ is correct, the probability of obtaining such a value is equal to the area to the right of a_c under the curve marked $\mathscr{H}_1$ in Fig. 2. The means and widths of these curves depend on the weights $\mathbf{w}$. We therefore seek to maximize our chances of success by adjusting the weights.

Since the critical value a_c depends on the weights, we first determine this dependence. The probability distribution of

$$a = \mathbf{w}'\mathbf{y} \tag{2.2}$$

is dictated by the probability distribution $P_0(\mathbf{y})$, which we assume to be (multivariate) Gaussian:

$$P_0(\mathbf{y}) = (2\pi)^{-(1/2)p}\|\mathbf{\Sigma}_0\|^{-1/2}\exp\left[-\tfrac{1}{2}\mathbf{y}'(\mathbf{\Sigma}_0)^{-1}\mathbf{y}\right], \tag{2.3}$$

where $\mathbf{\Sigma}_0$ is the covariance matrix of the p variables $\mathbf{y}$,

$$(\mathbf{\Sigma}_0)_{ij} = \langle y_i y_j \rangle_{P_0}. \tag{2.4}$$

The angular brackets indicate an average over the population with probability distribution P_0. We shall denote probability distribution (2.3) in the standard way as

$$\mathbf{y} \sim \mathscr{N}(\mathbf{0}, \mathbf{\Sigma}_0), \tag{2.5}$$

meaning $\mathbf{y}$ is normally distributed with mean $\mathbf{0}$ and covariance $\mathbf{\Sigma}_0$. (We have assumed that the mean of $\mathbf{y}$ is zero to simplify the notation. If $\mathbf{y}$ were to have nonzero mean $\boldsymbol{\mu}_0$, $\mathbf{y}$ would simply be replaced by $\mathbf{y} - \boldsymbol{\mu}_0$ everywhere in this section. We shall return to using a nonzero $\boldsymbol{\mu}_0$ in section 3a.)

It is easily shown (see Anderson, 1958, for example) that a linear combination of normal variables is itself normally distributed. Consequently, variable a in Eq. (2.2) is distributed as

$$a \sim \mathscr{N}[0, (\sigma_0)^2], \quad \text{under } \mathscr{H}_0, \tag{2.6}$$

with

$$(\sigma_0)^2 = \mathbf{w}'\mathbf{\Sigma}_0\mathbf{w}. \tag{2.7}$$

The critical value a_c is therefore given by

$$a_c = \eta_c\sigma_0, \tag{2.8}$$

where η_c is a number depending on the level of significance desired for the test and can be obtained from standard statistical tables for the normal probability distribution (e.g., Beyer, 1968). For a one-sided test at the 5% significance level, the value $\eta_c = 1.65$ is appropriate. For a one-sided test at the 2.5% level, the value of η_c is the more familiar $\eta_c = 1.96$.

Next we need the dependence on the weights of the probability distribution of a under hypothesis $\mathscr{H}_1$. The probability distribution $P(\mathbf{y})$ is again assumed Gaussian:

$$\mathbf{y} \sim \mathscr{N}(\boldsymbol{\mu}, \mathbf{\Sigma}). \tag{2.9}$$

The climate model predicts the change $\boldsymbol{\mu}$ in the means of the variables $\mathbf{y}$ and could possibly predict a change in the covariance statistics to the new matrix $\mathbf{\Sigma}$, as well. The covariance $\mathbf{\Sigma}$ is defined as in Eq. (2.4) with P_0 replaced by P.

If the climate modeler were able to estimate the uncertainty in his prediction of the climatic shifts $\boldsymbol{\mu}$ as an error matrix $\mathbf{\Sigma}_\mu$ for the shifts, this information could still be represented by (2.9) with the covariance $\mathbf{\Sigma}$ composed of two parts:

$$\mathbf{\Sigma} = \mathbf{\Sigma}_y + \mathbf{\Sigma}_\mu, \tag{2.10}$$

where $\mathbf{\Sigma}_y$ represents the climatic noise—the covariance of the variables about the true (but imprecisely known) mean—and $\mathbf{\Sigma}_\mu$ represents the additional uncertainty in the prediction of $\mathbf{y}$ due to possible error in the climate model results. The quantity $\boldsymbol{\mu}$ in (2.9) would then be the most likely climatic shift. Some information about $\mathbf{\Sigma}_\mu$ would result naturally from statistical analysis of general circulation model estimates of $\boldsymbol{\mu}$, since there is always sampling error in these estimates due to the finite length of integration time of the numerical experiments. The climate modeler's confidence in how accurately his model captures the physics of the climate system is a less easily quantifiable component of $\mathbf{\Sigma}_\mu$.

For lack of such information in most cases, we shall later assume

$$\mathbf{\Sigma} = \mathbf{\Sigma}_0, \quad \mathbf{\Sigma}_\mu = \mathbf{0}; \tag{2.11}$$

that is, the covariance of the variables is not significantly affected by the climate change and that uncertainties in the climate model prediction for the shifts $\boldsymbol{\mu}$ can be neglected. However, we do not need to make these assumptions yet.

Given Eq. (2.9), it follows that the distribution of a would be

$$a \sim \mathscr{N}[\mathbf{w}'\boldsymbol{\mu}, \sigma^2] \quad \text{under } \mathscr{H}_1, \tag{2.12}$$

with

$$\sigma^2 = \mathbf{w}'\mathbf{\Sigma}\mathbf{w}. \tag{2.13}$$

Equations (2.12) and (2.13) describe the curve labeled $\mathscr{H}_1$ in Fig. 2. We can now write an expression for the probability if $\mathscr{H}_1$ is correct that the value of A defined in Eq. (2.1) would exceed the critical value a_c defined in Eq. (2.8):

$$p(a > a_c|\mathscr{H}_1) = (2\pi)^{-1/2}\sigma^{-1}\int_{a_c}^{\infty}\exp\left[-\tfrac{1}{2}(a - \mathbf{w}'\boldsymbol{\mu})^2/\sigma^2\right]da. \tag{2.14}$$

If $\mathscr{H}_1$ is correct, this is the probability we have of rejecting $\mathscr{H}_0$. It depends on $\mathbf{w}$. By changing integration variables in Eq. (2.14) to $x = (a - \mathbf{w}'\boldsymbol{\mu})/\sigma$ we can write it as

$$p(a > a_c|\mathscr{H}_1) = E(L), \qquad (2.15)$$

where

$$L = (a_c - \mathbf{w}'\boldsymbol{\mu})/\sigma, \qquad (2.16)$$

$$E(L) \equiv (2\pi)^{-1/2} \int_L^\infty \exp\left(-\frac{1}{2}x^2\right)dx$$
$$= [1 - \operatorname{erf}(L/\sqrt{2})]/2. \qquad (2.17)$$

Since $E(L)$ is a decreasing function of L, the probability $p(a > a_c|\mathscr{H}_1)$ in Eq. (2.15) is maximized by minimizing L, which occurs where

$$\frac{\partial L}{\partial w_i} = 0, \quad i = 1, \cdots, p. \qquad (2.18)$$

Substituting (2.8) and (2.13) into (2.16) and evaluating the derivatives in Eq. (2.18), we obtain equations for the weights:

$$\left[\frac{\eta_c}{\sigma_0}\boldsymbol{\Sigma}_0 + \frac{1}{\sigma^2}(\mathbf{w}'\boldsymbol{\mu} - \eta_c\sigma_0)\boldsymbol{\Sigma}\right]\mathbf{w} = \boldsymbol{\mu}. \qquad (2.19)$$

This equation, although apparently nonlinear in $\mathbf{w}$, can be solved easily if we assume Eqs. (2.11) and take into account the fact that the overall normalization of the weights $\mathbf{w}$ is irrelevant and not fixed by Eq. (2.19). In this case [given Eq. (2.11)] the solution to (2.19) would be the one obtained by Hasselmann (1979) and Bell (1982),

$$\mathbf{w} = c\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}, \qquad (2.20)$$

where c is an arbitrary constant. If we substitute Eq. (2.20) for the weights into Eqs. (2.7) and (2.13), we obtain an expression for the standard deviation of a:

$$\sigma_0 = \sigma = c(\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu})^{1/2}. \qquad (2.21)$$

c. Solution for weights for $\boldsymbol{\Sigma} \neq \boldsymbol{\Sigma}_0$

If the climate model prediction were sufficiently detailed that assumptions (2.11) would be inappropriate, then Eq. (2.19) can be solved iteratively numerically. One method of doing this is to fix the arbitrary normalization of $\mathbf{w}$ such that

$$\mathbf{w}'\boldsymbol{\mu} = \sigma^2. \qquad (2.22)$$

Then Eq. (2.19) can be written

$$\mathbf{w} = [(\eta_c/\sigma_0)\boldsymbol{\Sigma}_0 + (1 - \eta_c\sigma_0/\sigma^2)\boldsymbol{\Sigma}]^{-1}\boldsymbol{\mu}, \qquad (2.23)$$

where, in the first iteration, σ_0 and σ on the right-hand side would be evaluated from (2.7) and (2.13) using $\mathbf{w} = \boldsymbol{\Sigma}^{-1}\boldsymbol{\mu}$ to get a new estimate of $\mathbf{w}$ for the next iteration of the expression. When iteration of (2.23) has converged to a solution, the normalization of the solution may be changed to whatever suits the investigator. Note that, in general, the weights $\mathbf{w}$ depend on the significance level at which the test is planned, since the solution to (2.19) depends on η_c. However, for $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_0$ this dependence disappears, as in Eq. (2.20).

d. Probability of detection

Once the solution of Eq. (2.19) is obtained for the weights, it is interesting to know what the probability is of detecting climatic change; i.e., if the climate model is correct, what is the probability of the data confirming even the simplest prediction of the model: that the climate, as measured by A, has changed? This probability is given by Eq. (2.15) evaluated using the solution for the weights $\mathbf{w}$ just described. It is not, in general, a simple function of the climate model prediction. However, if we can assume $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_0$, as we did in order to write solution (2.20), the probability of detection reduces to a simple function of

$$\frac{\mathbf{w}'\boldsymbol{\mu}}{\sigma} = (\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu})^{1/2}, \qquad (2.24)$$

which is just the ratio of the model-predicted shift in the mean to the "climatic noise" for the average A. The square of (2.24) is just the signal-to-noise ratio discussed in Bell (1982). The larger the signal-to-noise ratio is, the larger is the probability of detecting climatic change,

$$p(a > a_c|\mathscr{H}_1) = E[\eta_c - (\boldsymbol{\mu}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\mu})^{1/2}], \qquad (2.25)$$

(assuming $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_0$).

It is informative to expand the climate change $\boldsymbol{\mu}$ in terms of the principal components, or empirical orthogonal functions (EOFs), of $\boldsymbol{\Sigma}$. If m_i is the amplitude of the ith EOF and λ_i is the corresponding eigenvalue of $\boldsymbol{\Sigma}$, Eq. (2.24) may be written

$$\left(\frac{\mathbf{w}'\boldsymbol{\mu}}{\sigma}\right)^2 = \sum_i \frac{m_i^2}{\lambda_i}. \qquad (2.26)$$

Thus an EOF that contributes only a small amount m_i to the climate signal $\boldsymbol{\mu}$ can contribute a great deal to the signal-to-noise ratio of the average A, if λ_i is small enough. This will be reflected in the optimal weights $\mathbf{w}$, which will emphasize such an EOF. This can also be a source of large errors, if either the prediction m_i is poor or the small eigenvalues λ_i are badly underestimated, as can happen when too little data are available to estimate the covariance matrix $\boldsymbol{\Sigma}$. The latter problem is solved in section 3.

e. Acceptance intervals for comparing model to data

Once evidence for climatic change has been found, there still remains the problem of comparing the climate model prediction to the data $\mathbf{Y}$. As Hayashi (1982) has emphasized, more information is conveyed by expressing climatic anomalies in terms of where they fall in an acceptance interval than in terms of simple acceptance or rejection of a hypothesis. The weighted average $A = \mathbf{w}'\mathbf{Y}$ should fall within the acceptance interval

$$\mathbf{w}'\boldsymbol{\mu} - \alpha_c\sigma < A < \mathbf{w}'\boldsymbol{\mu} + \alpha_c\sigma \qquad (2.27)$$

centered on the climate model prediction $\mathbf{w}'\boldsymbol{\mu}$, where σ is given by Eq. (2.13) and α_c is obtained from a table for the normal probability distribution and depends on the significance level desired for the comparison. The value of α_c in (2.27) appropriate for a two-sided test at the 95% confidence level is $\alpha_c = 1.96$. Note that α_c is distinguished here from η_c defined in Eq. (2.8), since the latter is used in a one-sided test of a Gaussian variable. They are both, of course, found in the same statistical table.

This comparison of the model with data has the advantage that it uses only one variable, and so avoids the pitfall described in the Introduction of testing too many variables at one time. The weights $\mathbf{w}$ tend to emphasize those variables Y_i for which the predicted shift μ_i is large compared with the noise level $\langle y_i^2\rangle^{1/2}$ for the variable, and so is, in some sense, most sensitive to the direction in which the climatic shift is likely to be significant.

By the same token, however, because Eq. (2.27) compares only one variable, albeit a sensitive one, to the climate model prediction, one is naturally curious to know how the model prediction compares to the actual atmospheric behavior in many other respects, to answer such questions as, Has the climate truly warmed more in the arctic regions than in the tropics? or, Has rainfall truly decreased more in the eastern United States than in the western? As was discussed in the Introduction, as one attempts more and more detailed comparisons of the model and the atmospheric data, the confidence one has in the conclusions reached will probably diminish (unless the climate model prediction is badly wrong). How far one can go in pursuing these details depends on the size of the climatic shift relative to the climatic noise.

3. Sampling error for optimally weighted data

a. Optimal weighting from a finite sample

In section 2 we derived a weighted average of data that, if the climate model prediction is correct, has more chance of revealing climatic change than any other linear average of the data. We assumed there that the covariance matrices $\boldsymbol{\Sigma}_0$ and $\boldsymbol{\Sigma}$ describing the probability distributions before and after the climate change are accurately known. In practice, these matrices are only imperfectly known. The matrix $\boldsymbol{\Sigma}_0$ can be estimated using data from a period prior to the climatic change. The matrix $\boldsymbol{\Sigma}$ will be assumed here to have changed so little from $\boldsymbol{\Sigma}_0$ that the approximation

$$\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_0 \qquad (3.1)$$

can be made. We will continue to ignore any uncertainties in the prediction $\boldsymbol{\mu}$ for the climatic shift, which might also cause $\boldsymbol{\Sigma}$ to differ from $\boldsymbol{\Sigma}_0$. [See Eq. (2.10).]

The covariance matrix $\boldsymbol{\Sigma}_0$ is defined in Eq. (2.4). To estimate it, let us suppose that we have data from N independent periods of time labeled by α, $\alpha = 1, \cdots, N$, prior to the period t_1 to t_2 over which climatic change is being investigated. For concreteness, each period can be imagined to be 1 yr in length. In each period α the data are averaged over time and space just as they were for the period t_1 to t_2 to produce a set of values $Y_i^{(\alpha)}$, from which we can obtain an estimate $\mathbf{S}$ of the covariance matrix $\boldsymbol{\Sigma}_0$:

$$(\mathbf{S})_{ij} = \frac{1}{N-1}\sum_{\alpha=1}^{N}[Y_i^{(\alpha)} - \overline{Y}_i][Y_j^{(\alpha)} - \overline{Y}_j], \qquad (3.2)$$

$$\overline{Y}_i = \frac{1}{N}\sum_{\alpha=1}^{N} Y_i^{(\alpha)}. \qquad (3.3)$$

The optimal weights can then be written in terms of estimate (3.2), using Eq. (2.20) and $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_0$, as

$$\mathbf{w} = c\mathbf{S}^{-1}\boldsymbol{\mu}, \qquad (3.4)$$

where c is an arbitrary normalization constant. We shall use $c = 1$ here.

It is convenient now to make the climatic mean $\boldsymbol{\mu}_0$ prior to the climatic change explicit in the weighted average Eq. (2.1), instead of assuming it equal to $\mathbf{0}$, as we have so far. [See the remarks after Eq. (2.5).] The change in the atmospheric variables following the climatic change is measured by $\mathbf{Y} - \boldsymbol{\mu}_0$, and the climate model predicts $\langle \mathbf{Y} - \boldsymbol{\mu}_0\rangle_P = \boldsymbol{\mu}$. The weighted average of the deviation of the data $\mathbf{Y}$ from the prior climatic mean, Eq. (2.1), should be written for nonzero $\boldsymbol{\mu}_0$ as

$$A = \mathbf{w}'(\mathbf{Y} - \boldsymbol{\mu}_0). \qquad (3.5)$$

We shall use prior data to estimate $\boldsymbol{\mu}_0$ just as we did to estimate $\boldsymbol{\Sigma}_0$, and replace $\boldsymbol{\mu}_0$ by its estimate $\overline{\mathbf{Y}}$, Eq. (3.3). The weighted average of the data $\mathbf{Y}$ appropriate to testing hypothesis $\mathscr{H}_0$ can thus be written in terms of the estimates (3.2) and (3.3) as

$$A = \boldsymbol{\mu}'\mathbf{S}^{-1}(\mathbf{Y} - \overline{\mathbf{Y}}). \qquad (3.6)$$

Although we cannot prove that this form of the average with $\boldsymbol{\Sigma}_0$ and $\boldsymbol{\mu}_0$ replaced by sample estimates is optimal for detecting climatic change (rejecting $\mathscr{H}_0$) when $\mathscr{H}_1$ is correct, it is a plausible choice to make, since it is simple to compute and agrees with (2.20) in the limit $N \rightarrow \infty$.

b. Distribution theory

To test hypothesis $\mathscr{H}_0$, we need to know the probability distribution that the quantity

$$a = \boldsymbol{\mu}'\mathbf{s}^{-1}(\mathbf{y} - \overline{\mathbf{y}}) \qquad (3.7)$$

would have under this hypothesis, where the statistical behavior of $\overline{\mathbf{y}}$, $\mathbf{s}$ and $\mathbf{y}$ are described by

$$\overline{\mathbf{y}} = \frac{1}{N}\sum_{\alpha=1}^{N}\mathbf{y}^{(\alpha)}, \qquad (3.8)$$

$$(\mathbf{s})_{ij} = \frac{1}{N-1} \sum_{\alpha=1}^{N} (y_i^{(\alpha)} - \bar{y}_i)(y_j^{(\alpha)} - \bar{y}_j), \quad (3.9)$$

$$\mathbf{y} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}), \quad (3.10)$$

$$\mathbf{y}^{(\alpha)} \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}), \quad \alpha = 1, \cdots, N. \quad (3.11)$$

For future reference, note that since $\mathbf{y}$ and $\mathbf{y}^{(\alpha)}$ are assumed independent of each other, it follows that

$$\mathbf{y} - \bar{\mathbf{y}} \sim \mathcal{N}[\mathbf{0}, (1 + 1/N)^{1/2}\mathbf{\Sigma}]. \quad (3.12)$$

The probability distribution of a in (3.7) is no longer Gaussian (except in the limit $N \rightarrow \infty$), and so we cannot so easily determine the critical value a_c such that if $A > a_c$ we will reject $\mathscr{H}_0$, say, with 95% confidence. The distribution of a depends on $\mathbf{\Sigma}$, which we know only approximately through Eq. (3.2).

In analogy with the Student's t-test, let us try to form a quantity independent of $\mathbf{\Sigma}$ by taking the ratio of a to an expression like Eq. (2.21) for its standard deviation, and define the variable

$$u \equiv (1 + 1/N)^{-1/2} \frac{a}{(\boldsymbol{\mu}'\mathbf{s}^{-1}\boldsymbol{\mu})^{1/2}} \quad (3.13)$$

$$= (1 + 1/N)^{-1/2}\boldsymbol{\mu}'\mathbf{s}^{-1}(\mathbf{y} - \bar{\mathbf{y}})/(\boldsymbol{\mu}'\mathbf{s}^{-1}\boldsymbol{\mu})^{1/2}. \quad (3.14)$$

The factor $(1 + 1/N)^{-1/2}$ is suggested by Eq. (3.12) and has been introduced for later convenience.

It is a remarkable fact that the random variable u has a statistical distribution independent of $\mathbf{\Sigma}$ and $\boldsymbol{\mu}$. It is shown in appendix A that u is distributed identically to the variable

$$v = \mathbf{e}'\mathbf{s}_x^{-1}\mathbf{x}/(\mathbf{e}'\mathbf{s}_x^{-1}\mathbf{e})^{1/2} \quad (3.15)$$

where $\mathbf{e}$ is an arbitrary unit vector in the p-dimensional variable space, $\mathbf{x}$ is a (p-dimensional) multivariate Gaussian variable with zero mean and unit covariance matrix, and $\mathbf{s}_x$ is (Wishart) distributed as the covariance matrix estimate from $n = N - 1$ samples of p multivariate Gaussian variables with (known) zero mean and unit covariance matrix. More precise definitions of these quantities may be found in Eqs. (A11) and (A15) in appendix A.

We shall now derive an approximate expression for critical values of this variable to be used in hypothesis testing. The approximate expression will be shown to be quite accurate for our purposes by analyzing Monte Carlo estimates of the statistics of u. But let us first discuss how these results will be used. Given the probability distribution for u, we can establish a value v_c such that $p(u > v_c|\mathscr{H}_0) = 0.05$, say. The corresponding value a_c from Eq. (3.13) is

$$a_c = (1 + 1/N)^{1/2}v_c(\boldsymbol{\mu}'\mathbf{S}^{-1}\boldsymbol{\mu})^{1/2}. \quad (3.16)$$

In the limit $N \rightarrow \infty$, when $\mathbf{\Sigma}_0$ is exactly known, this confidence limit approaches the limit in Eq. (2.8) and $v_c \rightarrow \eta_c$. Hypothesis $\mathscr{H}_0$ is rejected if A [defined in Eq. (3.6)] exceeds a_c. Thus, all the ingredients for testing hypothesis $\mathscr{H}_0$ are present in Eq. (3.16) except for v_c; $\boldsymbol{\mu}$ is supplied by the climate model and $\mathbf{S}$ is determined from prior data using Eq. (3.2). We turn now to evaluating v_c.

We will first obtain an approximation for the probability distribution of v in (3.15). Although v is not exactly normally distributed, we shall approximate it by a Gaussian. A Monte Carlo calculation, to be described later, will show that the Gaussian approximation is quite good for our purposes. The mean of the Gaussian is clearly zero, since $\mathbf{x}$ has zero mean and is statistically independent of $\mathbf{s}_x$. The width of the Gaussian is obtained by finding the variance of

$$\langle v^2 \rangle = \langle (\mathbf{e}'\mathbf{s}_x^{-1}\mathbf{x})(\mathbf{x}'\mathbf{s}_x^{-1}\mathbf{e})/(\mathbf{e}'\mathbf{s}_x^{-1}\mathbf{e}) \rangle, \quad (3.17)$$

where we have used the vector identities $\mathbf{A}'\mathbf{B} = \mathbf{B}'\mathbf{A}$ and the symmetry of matrix $\mathbf{s}_x$. Using $\langle \mathbf{x}\mathbf{x}' \rangle = \mathbf{1}$ from (A11) and the statistical independence of $\mathbf{x}$ and $\mathbf{s}_x$, we obtain

$$\langle v^2 \rangle = \langle (\mathbf{e}'\mathbf{s}_x^{-2}\mathbf{e})/(\mathbf{e}'\mathbf{s}_x^{-1}\mathbf{e}) \rangle. \quad (3.18)$$

This can be evaluated in the limit $n, p \rightarrow \infty$, p/n fixed; it is shown in appendix B that it has the value

$$\sigma_v \equiv \langle v^2 \rangle^{1/2} = 1/(1 - p/n), \quad p, n \rightarrow \infty. \quad (3.19)$$

The critical value v_c is therefore approximately given by

$$v_c \approx \eta_c\sigma_v = \eta_c/(1 - p/n), \quad (3.20)$$

where η_c is defined just as it was for Eq. (2.8). The approximation (3.20) becomes exact in the limit $n \rightarrow \infty$.

c. Monte Carlo simulation of the distribution

In order to obtain the analytical results above, we had to treat the distribution of v in (3.15) as approximately Gaussian and assume n large. By using Monte Carlo methods, we can avoid these assumptions and find out how good the approximation (3.20) is for small p, n.

To carry out the Monte Carlo study, we write the variable v in (3.15) in the form

$$v = \mathbf{g}'\mathbf{x}, \quad (3.21)$$

where the vector $\mathbf{g}$ is

$$\mathbf{g} = \mathbf{s}_x^{-1}\mathbf{e}/(\mathbf{e}'\mathbf{s}_x^{-1}\mathbf{e})^{1/2}. \quad (3.22)$$

The random vector $\mathbf{g}$ depends only on the random matrix $\mathbf{s}_x$ and is independent of $\mathbf{x}$. We make use of this fact by doing the integral over $\mathbf{x}$ in the expression (A17) for the probability density $p(v)$ given in appendix A to obtain

$$P(v) = \int \cdots \int [\prod_{\alpha=1}^{n} d\mathbf{x}^{(\alpha)}] \frac{1}{\sqrt{2\pi}g} \exp\left(-\frac{1}{2}v^2/g^2\right) \times \prod_{\alpha'=1}^{n} p_G[\mathbf{x}^{(\alpha')}] \quad (3.23)$$

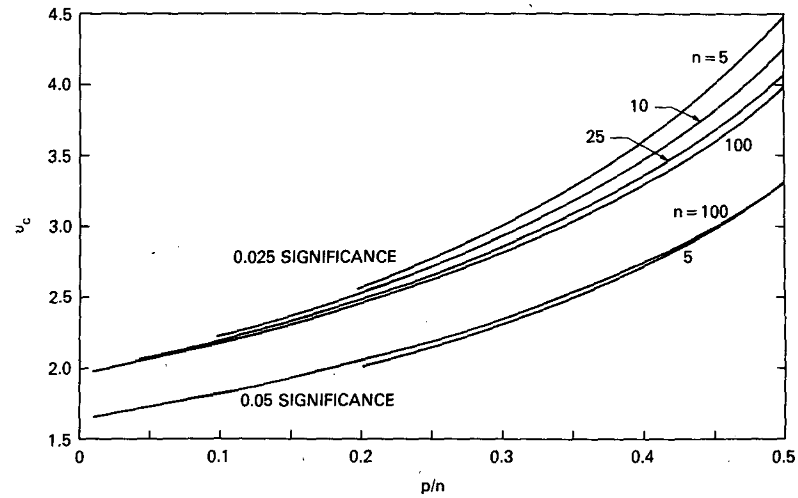

FIG. 3. Critical values $v_c(\beta)$ for p optimally weighted variables, with $N = n + 1$ prior independent samples available for estimating the covariance matrix used in obtaining the optimal weights. The critical values are obtained from Monte Carlo calculations described in the text and are fitted by smooth empirical curves, Eq. (3.27), for the two significance levels $\beta = 0.05$ and 0.025. The critical values $v_c(\beta)$ are used in Eq. (3.16) to obtain critical values for the weighted average A in Eq. (3.6).

where $g \equiv |\mathbf{g}|$. We can write Eq. (3.23) symbolically as

$$p(v) = \left\langle (\sqrt{2\pi}g)^{-1} \exp\left(-\frac{1}{2}v^2/g^2\right)\right\rangle_g, \quad (3.24)$$

where the notation $\langle \cdot \rangle_g$ means the average of the bracketed quantity over the distribution of vector lengths g generated by (3.22) when the covariance matrix $\mathbf{s}_x$ is estimated from n samples of normally distributed vectors $\mathbf{x}^{(\alpha)}$. [See Eqs. (A15) and (A11).] The critical value v_c is defined such that

$$\int_{v_c(\beta)}^{\infty} p(v)dv = \beta, \quad (3.25)$$

where β is the significance level desired for a test, $\beta = 0.05$, say. The value v_c depends on β, as is made explicit in Eq. (3.25). If Eq. (3.24) for $p(v)$ is substituted in Eq. (3.25), we obtain

$$\langle E[v_c(\beta)/g]\rangle_g = \beta, \quad (3.26)$$

where the function E is defined in Eq. (2.17).

A Monte Carlo approach was used to carry out the average over g indicated in Eq. (3.26). Random Gaussian vectors $\mathbf{x}^{(\alpha)}$ were generated, n at a time, to create sample covariance matrices $\mathbf{s}_x$; the vector $\mathbf{g}$ was evaluated using Eq. (3.22) for each sample covariance matrix, and the bracket operation in (3.26) was replaced by an average over many thousands of values of g thus generated. The value of $v_c(\beta)$ was then found, using Newton's method, that satisfied Eq. (3.26) for a given β.

The critical value $v_c(\beta)$ might, in general, depend on p and n, although the analytical result Eq. (3.20) depends only on the ratio p/n. In Fig. 3 we plot for $\beta = 0.025$ and 0.05 the results for $v_c(\beta)$ of the Monte Carlo study for various values of p and n vs the ratio p/n. The points for a given value of n are connected by smooth lines obtained from an analytical fit of the form

$$v_c = \frac{\eta_c}{1 - (p/n) + [\gamma_1 + \gamma_2(p/n) + \gamma_3/n]/n}, \quad (3.27)$$

where the values of η_c, γ_1, γ_2 and γ_3 depend on the significance level β. The constant η_c is just the critical value for testing a Gaussian variable, the solution of

$$E(\eta_c) = \beta, \quad (3.28)$$

and can be found in standard statistical tables. The constants $\gamma_1, \gamma_2, \gamma_3$ were obtained from a least squares fit to the Monte Carlo results and are given in Table 1. Note that they are small. The corrections to approximation (3.20) are therefore also small and, for practical purposes, negligible except for the very smallest values of n. To illustrate this more clearly, in Fig. 4 we plot $(1 - p/n)v_c$. On the scale of this plot, the uncertainties in the Monte Carlo computation can be made visible and are shown as one-standard-deviation

TABLE 1. Constants in the empirical fit, Eq. (3.27), for the critical values v_c of a test of the random variable v in Eq. (3.15).

Significance level (β)	Constants		
	γ_1	γ_2	γ_3
0.025	−0.2274	−0.4188	0.6264
0.05	0.0690	−0.2641	0.3378

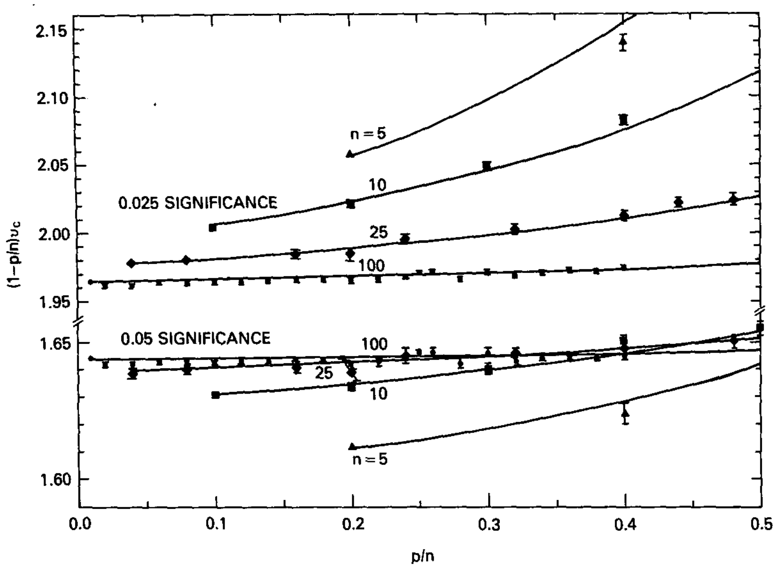

FIG. 4. The quantity $(1 - p/n)v_c(\beta)$ is plotted to show the accuracy of the Monte Carlo calculation and the fit (smooth curve) of the empirical curve, Eq. (3.27), to the Monte Carlo results. Note the change of vertical scale. One-standard-deviation error bars are errors estimated from the number of samples used in the Monte Carlo calculations.

error bars. The approximation (3.20) is seen to be accurate to a few percent for all but the smallest values of n, and the empirical fit (3.27) is accurate to a few tenths of a percent.

4. Power of the optimal weighting method

In section 3 we found that as the number p of variables for which optimal weights are determined increases, the critical value a_c, which must be exceeded by the average A [defined in Eq. (3.6)], likewise increases. [See Eqs. (3.16) and (3.20).] Thus, the inclination to try to extract as much signal from the data as possible, by using as many independent weights w_i as possible, must be weighed against the possible loss of detection power arising from the concomitant increase in the critical value a_c.

This problem is best illustrated by an example. Suppose that we wish to detect a predicted climatic warming of surface temperature during 1 yr. The surface temperatures are expanded in spherical harmonics, say, so that the variables Y_i, $i = 1, \cdots, p$, are the amplitudes (arranged in a suitable order) of the spherical harmonic expansion of surface temperature for the year in which the signal is being sought. The signal we wish to detect is $\mu = (\mu_1, \mu_2, \cdots, \mu_p)$ where μ_i is the ith amplitude of the spherical harmonic expansion of the climate model prediction. In this example, Y_1 might represent the globally averaged surface temperature for the year in question, and μ_1 the climate model prediction for the global warming for that year.

We suppose we have N years of prior data. To proceed further, we need an expression for the probability of detecting climatic change, like Eq. (2.15) but modified to take into account the finite amount of data in the prior data set.

Climate change is detected when the weighted average

$$A = \mu'\mathbf{S}^{-1}(\mathbf{Y} - \bar{\mathbf{Y}}) \tag{4.1}$$

exceeds the critical value

$$a_c = (1 + 1/N)^{1/2} v_c(\mu'\mathbf{S}^{-1}\mu)^{1/2}, \tag{4.2}$$

where v_c is given approximately by (3.27). The probability that this will occur can be derived following the steps of section 3, replacing assumption (3.10) by hypothesis $\mathscr{H}_1$:

$$\mathbf{Y} \sim \mathcal{N}(\mu, \mathbf{\Sigma}). \tag{4.3}$$

By repeating the steps of section 3, we find that Eq. (3.26) is replaced by

$$p(A > a_c|\mathscr{H}_1) = \langle E[(v_c - rg_1)/g]\rangle_{\mathbf{g}}, \tag{4.4}$$

with

$$r^2 = \mu'\mathbf{\Sigma}^{-1}\mu/(1 + 1/N), \tag{4.5}$$

where the average $\langle \cdot \rangle_{\mathbf{g}}$ is defined as in Eq. (3.24) and g_1 is the component of $\mathbf{g}$ in the direction $\mathbf{e}$,

$$g_1 = \mathbf{e}'\mathbf{g} = (\mathbf{e}'\mathbf{s}_x^{-1}\mathbf{e})^{1/2}. \tag{4.6}$$

Let us take as an example that the climate change vector has the values

$$\mu = (2, 0.75, 0.75, \cdots, 0.75), \tag{4.7}$$

and that the covariance matrix $\mathbf{\Sigma}$ is the unit matrix

$$\mathbf{\Sigma} = \mathbf{1}. \tag{4.8}$$

We assume we have $N = 26$ years of prior data from which to form the estimates $\bar{\mathbf{Y}}$ and $\mathbf{S}$ in Eqs. (3.2) and (3.3).

We are at liberty to include as many of the variables Y_i as we choose in forming the optimal average (4.1). Each additional variable brings some extra signal to be detected (as well as additional noise), but the threshold a_c which A must exceed increases as well. Our problem is to choose the number of variables p that maximizes the probability $p(A > a_c|\mathscr{H}_1)$ of detecting climatic change. A plot of the probability vs p for the assumptions (4.7) and (4.8) is shown as the solid curve in Fig. 5 computed from Monte Carlo estimates of Eq. (4.4) as in the previous section. The confidence level was set at 97.5%, which requires approximately a two-standard-deviation shift in the average A. For $p = 1$, which may be thought of as corresponding to examining only the globally averaged surface temperature, assumption (4.7) implies that the signal is exactly 2σ in size ($\sigma = 1$), and so the probability of detecting it is approximately 0.5, but slightly less because of our assumption that we have only 26 yr of prior data. The dashed curve is a plot of the probability of detecting climatic change with optimal weights determined from an infinite amount of data, so that we, in effect, know $\mathbf{\Sigma}$ and the prior mean of $\mathbf{Y}$ exactly. In that case (dashed curve) the probability of detection increases with each additional variable, whereas with a finite amount of prior data (solid curve) a point is reached ($p = 9$) beyond which the power of the optimal weighting begins to decrease; it is counterproductive to keep more than this number of variables.

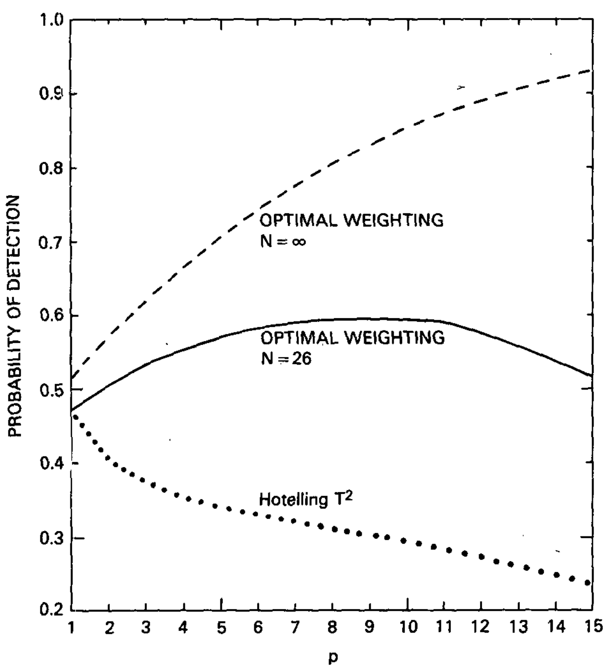

FIG. 5. Probability of detecting climatic change at the 0.025 significance level for p uncorrelated Gaussian random variables z_i, $i = 1, \cdots, p$, with unit variance, and a climatic shift in the means $\mu = (2, 0.75, 0.75, \cdots, 0.75)$. The dashed curve shows the detection probability of the optimal weighting method given an infinite amount of prior data. The solid curve gives the detection probability for the optimal weighting method with $N = 26$ samples of prior data. The values are from Monte Carlo calculations accurate to ± 0.007. The dotted curve shows the detection probability that Hotelling's T^2 test would yield.

For comparison, we also show in Fig. 5 (as the dotted curve) the probability of detecting a shift in the mean of $\mathbf{Y}$ using the standard multivariate Hotelling T^2 test, which tests whether the variable

$$T^2 = (\mathbf{Y} - \bar{\mathbf{Y}})'\mathbf{S}^{-1}(\mathbf{Y} - \bar{\mathbf{Y}}) \tag{4.9}$$

passes a critical value appropriate to this statistic. The test is described in Anderson (1958). It makes no assumptions about the nature of the signal μ being sought. As can be seen, even going from $p = 1$ to 2 results in a loss in the detection power of this test. The reason for this loss was explained in section 2. Much of it is due to the same phenomenon that caused the χ^2 test to fail rapidly with increasing p in Fig. 1, although some additional deterioration is due to the finite amount of prior data ($N = 26$) used in estimating $\mathbf{S}$ and $\bar{\mathbf{Y}}$ in (4.9).

As can be seen in this example, for a given amount of prior data there is an optimum choice for the number of variables used in forming the weighted average A. Unfortunately, this number depends in a detailed way on the exact values of the climatic shifts μ and the covariance matrix $\mathbf{\Sigma}$, and on the way the variables Y_i are chosen and ordered, and no general rule can be given for the best choice to make. Clearly the investigator must bring whatever a priori physical insight he can to the problem to choose variables Y_i beginning with those likely to have large climatic shifts μ_i relative to the noise level $[(\mathbf{\Sigma})_{ii}]^{1/2}$.

Carrying out Monte Carlo calculations to evaluate the power $P(A > a_c|\mathscr{H}_1)$ in (4.4) is cumbersome and is probably not necessary for many purposes since the variable u defined in Eq. (3.13),

$$u = \frac{\mu'\mathbf{s}^{-1}(\mathbf{y} - \bar{\mathbf{y}})}{(1 + 1/N)^{1/2}(\mu'\mathbf{s}^{-1}\mu)^{1/2}}, \tag{4.10}$$

is nearly Gaussian when N is large. We have already found in Eq. (3.19) that its width is approximately given by

$$\sigma_u \approx 1/(1 - p/n). \tag{4.11}$$

Its mean under hypothesis $\mathscr{H}_1$ [Eq. (4.3)] is given by

$$\langle u \rangle_P = (1 + 1/N)^{-1/2}\langle(\mu'\mathbf{s}^{-1}\mu)^{1/2}\rangle_P. \tag{4.12}$$

The quantity in angular brackets can be evaluated using a theorem for Wishart-distributed matrices given, for example, in Mardia et al. (1979) [Theorem (3.4.7)], that $n(\mu'\mathbf{\Sigma}^{-1}\mu)/(\mu'\mathbf{s}^{-1}\mu)$ is distributed as a χ^2_{n-p+1} variable. Using the χ^2 probability distribution we find

$$\langle(\mu'\mathbf{s}^{-1}\mu)^{1/2}\rangle_P = (\mu'\mathbf{\Sigma}^{-1}\mu)^{1/2}\left(\frac{n}{2}\right)^{1/2}\frac{\Gamma[(\nu-1)/2]}{\Gamma(\nu/2)} \quad (4.13)$$

$$\approx (\mu'\mathbf{\Sigma}^{-1}\mu)^{1/2}/(1-p/n)^{1/2} \quad (4.14)$$

with $\nu = n - p + 1$. Approximation (4.14) is good for large n. For large n, Eq. (4.12) can therefore be written

$$\langle u\rangle_P \approx r/(1-p/n)^{1/2}, \quad (4.15)$$

where r is defined in Eq. (4.5). Using the Gaussian approximation for the statistics of u, with mean (4.15) and standard deviation (4.11), we find that the probability of detecting a signal, Eq. (4.4), is approximated by

$$p(A > a_c|\mathscr{H}_1) \approx E(\eta_c - \langle u\rangle_P/\sigma_u), \quad (4.16)$$

where η_c is entirely determined by the confidence level at which the change is to be detected and is defined in Eq. (3.28). The Gaussian integral E is defined in Eq. (2.17). Thus the detection power is a strictly increasing function of

$$\frac{\langle u\rangle_P}{\sigma_u} \approx (1-p/n)^{1/2}(\mu'\mathbf{\Sigma}^{-1}\mu)^{1/2}/(1+1/N)^{1/2}. \quad (4.17)$$

To find the value of p that maximizes detection power, one can simply maximize (4.17). In the example used above, for $\mu = (2, 0.75, 0.75, \cdots, 0.75)$, $\mathbf{\Sigma} = \mathbf{1}$, and $N = 26$, the maximum is found with some simple algebra to occur at $p = 9.4$, which agrees well with the Monte Carlo calculation results in Fig. 5.

In practice one does not know the true covariance matrix $\mathbf{\Sigma}$ required to evaluate expression (4.17). However, using Eq. (4.14) we can replace $(\mu'\mathbf{\Sigma}^{-1}\mu)^{1/2}$ by its unbiased estimate $(1-p/n)^{1/2}(\mu'\mathbf{S}^{-1}\mu)^{1/2}$ and so obtain the estimated signal-to-noise ratio

$$\widehat{\left(\frac{\langle u\rangle_P}{\sigma_u}\right)} = \frac{(1-p/n)}{(1+1/N)^{1/2}}(\mu'\mathbf{S}^{-1}\mu)^{1/2}, \quad (4.18)$$

where the hat indicates that the quantity is estimated. One can therefore choose the best number of variables to work with in forming the optimal average A by maximizing expression (4.18) with respect to p.

5. Discussion and conclusions

We have discussed some of the problems of comparing climate model predictions with atmospheric data that arise from the probabilistic nature of the predictions. The results of such a comparison will be stronger the fewer the number of variables compared, at least when predictions of small climatic change are being investigated; this is the case when the first signs of global climatic change are being sought.

A weighted average of the data can be found that maximizes the chance of detecting a predicted climatic change. However, it requires knowledge of the covariance statistics of the variables in which the climatic change is being looked for, and prior data must usually be used to estimate these statistics. The implications of this for the optimal weighting approach have been examined here. If N is the number of independent samples used to estimate the mean and covariance statistics of the variables, the number p of variables for which optimal weights can be determined can at most be $N - 1$, and in general should be far fewer.

We summarize here a scheme for using the optimal weighting method as effectively as possible on a climatic change detection problem. A decision must first be made about how to prepare the data beforehand, so that a sequence of variables Y_i can be established such that as much of the climate signal as possible (relative to climatic noise) will occur in the first few variables. If global climatic change is predicted, projections of the data onto a basis system such as spherical harmonics or solutions of linearized dynamical equations should be considered.

A related consideration in choosing the variables is how well one trusts the climate model. For instance, if the model were to predict a large difference between temperatures at 10° and 20°N and one had some physical reason to suspect that this is due to a fault in the model, one may want to exclude from the list of the variables Y_i ones that can represent the temperature difference between these two latitudes. Otherwise the "optimal weights" might weight this temperature difference in the data too strongly. In principle, such a problem could be resolved by including directly in the equation determining the optimal weights an estimate of the uncertainty of the predicted climatic shifts, as was discussed in section 2. However, such estimates are usually unavailable in so concise a form as an error covariance matrix $\mathbf{\Sigma}_\mu$, and, at least for the present, the researcher must adjust his approach to the more qualitative estimates at hand.

Once a list of variables is chosen, their covariance statistics are determined from prior data. The best number of variables to keep is then found by computing the effective signal-to-noise ratio (4.18) for variables Y_i, $i = 1, \cdots, p$. The number p of variables that maximizes (4.18) is the best choice of the number of variables to form the optimal average from.

To detect climatic change, the optimally weighted average of the data $\mathbf{Y}$,

$$U = \frac{\mu'\mathbf{S}^{-1}(\mathbf{Y} - \bar{\mathbf{Y}})}{(1+1/N)^{1/2}(\mu'\mathbf{S}^{-1}\mu)^{1/2}} \quad (5.1)$$

is found, where we have divided by the factors suggested in Eq. (3.14) to create a variable whose probability distribution is independent of the true covariance statistics $\mathbf{\Sigma}$ and the climate change prediction μ. If $U > v_c(\beta)$, where a good approximation to v_c is given by Eq. (3.27), then climatic change (significant at the β level) has been found in the direction of the climate model prediction. Moreover, if the climate model is correct, the value of U should lie within the acceptance interval

$$M - v_c(\beta_1/2) < U < M + v_c(\beta_1/2) \quad (5.2)$$

where

$$M = \frac{(\mu' \mathbf{S}^{-1}\mu)^{1/2}}{(1 + 1/N)^{1/2}} \quad (5.3)$$

is an estimate of the climate change signal to be expected for the weighted average (5.1). The significance level β_1 for the acceptance interval in (5.2) is divided by two because v_c in Eq. (3.27) is defined for a one-sided test. Equation (5.2) provides a test of the climate model that emphasizes those variables with the largest signal-to-noise ratio.

The results in section 3 require that one know the number n of independent samples in the prior data set used to estimate covariance statistics. This is easy when the temporal correlation is not large; when temporal correlations are large, some of the analytical results obtained here can be inaccurate and one must resort to modeling the data with an autoregressive process to carry out the procedures described above. A description of such an effort will be given in a subsequent paper.

If at some point climate models could supply not only the predicted change in the means but could also provide good estimates of the covariance matrices $\mathbf{\Sigma}_0$ and $\mathbf{\Sigma}$ used in section 2, then optimal weights could be found using Eq. (2.23), which would not require the assumption that the covariance statistics do not change with the climate. Furthermore, the complexities introduced in sections 3 and 4 due to using a limited prior data set to estimate these statistics could be avoided. But this development will have to await substantial improvements in our ability to build models of the climate system.

Acknowledgments. Discussions with A. Abdullah, R. F. Cahalan, D. Chesters, A. Dalcher and G. R. North are gratefully acknowledged. Comments by E. S. Epstein and an anonymous reviewer were especially helpful.

APPENDIX A

Proof That the Distribution of the Ratio-Variable u in Eq. (3.14) for an Optimally Weighted Average is Independent of $\mathbf{\Sigma}$ and μ

The probability distribution of u, defined in Eq. (3.14), can be written as

$$p(u) = \int \cdots \int d\mathbf{y}[\prod_{\alpha=1}^{N} d\mathbf{y}^{(\alpha)}]\delta[u - (1 + 1/N)^{-1/2} \times \mu' \mathbf{s}^{-1}(\mathbf{y} - \bar{\mathbf{y}})/(\mu' \mathbf{s}^{-1}\mu)^{1/2}]P_0(\mathbf{y}) \prod_{\alpha'=1}^{N} P_0[\mathbf{y}^{(\alpha')}], \quad (A1)$$

where P_0 is the normal probability distribution in Eq. (2.3) and $\delta[u]$ is the Dirac δ function. Since $\mathbf{s}$ is independent of variable $\bar{\mathbf{y}}$, change variables in (A1) from the set $\{\mathbf{y}^{(\alpha)}, \alpha = 1, \cdots, N\}$ to the set $\{\sqrt{N}\bar{\mathbf{y}}, \mathbf{v}^{(\alpha)}, \alpha = 1, \cdots, N-1\}$ by an orthogonal transformation, and integrate over variable $\bar{\mathbf{y}}$. Equation (A1) then becomes

$$p(u) = \int \cdots \int d\mathbf{y}[\prod_{\alpha=1}^{n} d\mathbf{v}^{(\alpha)}] \times \delta[u - \mu' \mathbf{s}_v^{-1}\mathbf{y}/(\mu' \mathbf{s}_v^{-1}\mu)^{1/2}]P_0(\mathbf{y}) \prod_{\alpha'=1}^{n} P_0[\mathbf{v}^{(\alpha')}], \quad (A2)$$

where

$$n = N - 1, \quad (A3)$$

$$(\mathbf{s}_v)_{ij} = \frac{1}{n}\sum_{\alpha=1}^{n} v_i^{(\alpha)} v_j^{(\alpha)}. \quad (A4)$$

Using the orthogonal transformation $\mathbf{U}$ that diagonalizes $\mathbf{\Sigma}$,

$$\mathbf{U\Sigma U}^{-1} = \mathbf{\Lambda}, \quad (A5)$$

$$\Lambda_{ij} = \Lambda_i \delta_{ij}, \quad (A6)$$

and where $\mathbf{U}' = \mathbf{U}^{-1}$, define the square root of $\mathbf{\Sigma}$ as

$$\mathbf{\Sigma}^{1/2} \equiv \mathbf{U}'\mathbf{\Lambda}^{1/2}\mathbf{U} \quad (A7)$$

$$(\mathbf{\Lambda}^{1/2})_{ij} \equiv \Lambda_i^{1/2}\delta_{ij}. \quad (A8)$$

Change variables to

$$\mathbf{y} = \mathbf{\Sigma}^{1/2}\mathbf{x}, \quad (A9)$$

$$\mathbf{v}^{(\alpha)} = \mathbf{\Sigma}^{1/2}\mathbf{x}^{(\alpha)}, \quad \alpha = 1, \cdots, n, \quad (A10)$$

so that

$$\mathbf{x}, \mathbf{x}^{(\alpha)} \sim \mathcal{N}(\mathbf{0}, \mathbf{1}), \quad (A11)$$

where $\mathbf{1}$ is a $p \times p$ unit matrix. Equation (A2) then becomes

$$p(u) = \int \cdots \int d\mathbf{x}[\prod_{\alpha=1}^{n} d\mathbf{x}^{(\alpha)}]\delta[u - \mathbf{e}_0'\mathbf{s}_x^{-1}\mathbf{x}/ (\mathbf{e}_0'\mathbf{s}_x^{-1}\mathbf{e}_0)^{1/2}]p_G(\mathbf{x}) \prod_{\alpha'=1}^{n} p_G[\mathbf{x}^{(\alpha')}] \quad (12)$$

where $\mathbf{e}_0$ is a unit vector pointing in the direction of $\mathbf{\Sigma}^{-1/2}\mu$,

$$\mathbf{e}_0 = \mathbf{\Sigma}^{-1/2}\mu/(\mu'\mathbf{\Sigma}^{-1}\mu)^{1/2}, \quad (A13)$$

$$\mathbf{e}_0'\mathbf{e}_0 = 1, \quad (A14)$$

$$(\mathbf{s}_x)_{ij} = \frac{1}{n}\sum_{\alpha=1}^{n} x_i^{(\alpha)} x_j^{(\alpha)}, \quad (A15)$$

$$p_G(\mathbf{x}) = \prod_{i=1}^{p} (2\pi)^{-1/2} \exp\left[-\frac{1}{2}x_i^2\right]. \quad (A16)$$

Finally, since the statistics of $\mathbf{x}$, $\mathbf{x}^{(\alpha)}$ are independent of direction, because of Eq. (A11), we can find an orthogonal rotation matrix $\mathbf{R}$ chosen so that $\mathbf{e}_0 = \mathbf{Re}$, where $\mathbf{e}$ points in any direction we choose, and define new variables $\xi = \mathbf{R}^{-1}\mathbf{x}$, $\xi^{(\alpha)} = \mathbf{R}^{-1}x^{(\alpha)}$. Since the ξ's are dummy integration variables, we may replace them by $\mathbf{x}$'s, so that Eq. (A12) becomes

$$p(u) = \int \cdots \int d\mathbf{x}[\prod_{\alpha=1}^{n} d\mathbf{x}^{(\alpha)}]\delta[u - \mathbf{e}'\mathbf{s}_x^{-1}\mathbf{x}/(\mathbf{e}'\mathbf{s}_x^{-1}\mathbf{e})^{1/2}]p_G(\mathbf{x}) \prod_{\alpha'=1}^{n} p_G[\mathbf{x}^{(\alpha')}]. \quad (A17)$$

We may, if we like, choose

$$(\mathbf{e})_i = \delta_{1i}, \quad (A18)$$

so that $\mathbf{e}$ points in the direction of variable 1.

We have thus shown that the variable u is distributed as

$$v = \mathbf{e}'\mathbf{s}_x^{-1}\mathbf{x}/(\mathbf{e}'\mathbf{s}_x^{-1}\mathbf{e})^{1/2}, \quad (A19)$$

where $\mathbf{s}_x$ is the covariance matrix estimate from n samples, defined in Eq. (A15), for variables $\mathbf{x}^{(\alpha)}$ with $\mathbf{0}$ mean and unit variance as in (A11), where $\mathbf{x} \sim \mathcal{N}(\mathbf{0}, \mathbf{1})$, and $\mathbf{e}$ is an arbitrary unit vector. The statistical distribution of u is completely independent of $\mathbf{\Sigma}$ and $\boldsymbol{\mu}$!

APPENDIX B

Large *n* Limit for the Distribution of Optimally Weighted Data

To evaluate Eq. (3.18) in the large n limit, we make use of some results on principle components described by Cahalan (1983). The random matrix $\mathbf{s}_x$ in Eq. (A15) has principal components $\mathbf{e}^{(k)}$, and eigenvalues λ_k, $k = 1, \cdots, p$:

$$\mathbf{s}_x\mathbf{e}^{(k)} = \lambda_k\mathbf{e}^{(k)}. \quad (B1)$$

The eigenvalues are ordered in decreasing magnitude, $\lambda_k \geqslant \lambda_{k+1}$. The eigenvectors $\mathbf{e}^{(k)}$ are orthogonal and can point with equal probability in any direction, subject to their orthogonality. If we write Eq. (3.18) in terms of principal components, we obtain

$$\langle v^2 \rangle = \langle (\sum_{k=1}^{p} \epsilon_k^2\lambda_k^{-2})/(\sum_{l=1}^{p} \epsilon_l^2\lambda_l^{-1}) \rangle \quad (B2)$$

where

$$\epsilon_k = \mathbf{e}^{(k)\prime}\mathbf{e} \quad (B3)$$

is the component of vector $\mathbf{e}$ along eigenvector $\mathbf{e}^{(k)}$. Since the eigenvectors form a complete set,

$$\sum_{k=1}^{p} \epsilon_k^2 = \mathbf{e}'\mathbf{e} = 1. \quad (B4)$$

The brackets in (B2) now represent an average over all ϵ_k constrained to lie on the unit sphere by Eq. (B4) and an average over the distribution of eigenvalues of matrix $\mathbf{s}_x$. By expanding the denominator in (B2) in a power series about one,

$$(\sum_{l=1}^{p} \epsilon_l^2\lambda_l^{-1})^{-1} = \sum_{m=0}^{\infty} (1 - \sum_{l=1}^{p} \epsilon_l^2\lambda_l^{-1})^m, \quad (B5)$$

and considering each term in the expansion in the limit $p \rightarrow \infty$, it can be shown that the average over all orientations of the eigenvectors $\mathbf{e}^{(k)}$ in (B2) yields

$$\langle v^2 \rangle = \left\langle \left(\frac{1}{p}\sum_{k=1}^{p} \lambda_k^{-2}\right) \Big/ \left(\frac{1}{p}\sum_{l=1}^{p} \lambda_l^{-1}\right) \right\rangle + \mathcal{O}(1/p), \quad (B6)$$

where we have used $\langle \epsilon_k^2 \rangle = 1/p$ (for all k), which is easily derived from Eq. (B4) and the fact that the statistics of $\mathbf{s}_x$ are independent of direction.

The average over the eigenvalue spectrum in the limit $p, n \rightarrow \infty$ is carried out by using the fact that the eigenvalues in this limit approach a spectrum that depends only on p/n. Fluctuations of the spectrum from this limiting spectrum are of order $1/n$ and can be neglected in this limit. An average over any function of the eigenvalues $f(\lambda)$ can be written as

$$\lim_{\substack{p \rightarrow \infty \\ p/n \text{ fixed}}} \frac{1}{p}\sum_{k=1}^{p} f(\lambda_k) = \int_0^\infty d\lambda \rho(\lambda) f(\lambda), \quad (B7)$$

where $\rho(\lambda)$ is the probability density of the eigenvalues. Based on a method of Dyson's (1962), an explicit form for $\rho(\lambda)$ was derived by Cahalan (1983) and is given by

$$\rho(\lambda) = [-\lambda^2 + 2(1 + r^2)\lambda - (1 - r^2)^2]^{1/2}/(2\pi r^2\lambda),$$
$$\lambda_- \leqslant \lambda \leqslant \lambda_+, \quad (B8)$$

with

$$\lambda_\pm = (1 \pm r)^2, \quad r = (p/n)^{1/2}. \quad (B9)$$

The density $\rho(\lambda)$ vanishes for $\lambda < \lambda_-$ and $\lambda > \lambda_+$. By changing variables to

$$y = (\lambda - 1 - r^2)/2r, \quad (B10)$$

the average of powers of the eigenvalue can be written

$$\overline{\lambda^{-n}} \equiv \lim_{\substack{p \rightarrow \infty \\ p/n \text{ fixed}}} \frac{1}{p}\sum_{k=1}^{p} (\lambda_k)^{-n} = \frac{2}{\pi}\int_{-1}^{1} dy \frac{(1 - y^2)^{1/2}}{(1 + r^2 + 2ry)^{n+1}}. \quad (B11)$$

This can be evaluated using contour integration methods. We find

$$\overline{\lambda^{-1}} = (1 - p/n)^{-1}, \quad (B12)$$

$$\overline{\lambda^{-2}} = (1 - p/n)^{-3}. \quad (B13)$$

Substituting this result into (B6), we obtain

$$\langle v^2 \rangle = (1 - p/n)^{-2} + \mathcal{O}(1/n). \quad (B14)$$

REFERENCES

Anderson, T. W., 1958: *An Introduction to Multivariate Statistical Analysis.* Wiley and Sons, 126–153.

Bell, T. L., 1982: Optimal weighting of data to detect climatic change: Application to the carbon dioxide problem. *J. Geophys. Res.*, **87**, 11 161–11 170.

Beyer, W. H., Ed., 1968: *Handbook of Tables for Probability and Statistics,* 2nd ed. CRC Press, 125–134.

Cahalan, R. F., 1983: EOF spectral estimation in climate analysis. *Proc. Second Int. Meeting on Statistical Climatology,* Lisbon. Published by Instituto Nacional de Meteorologia e Geofísica, 4.5.1–4.5.7.

Dalcher, A., E. Kalnay, R. Livezey and R. N. Hoffman, 1985: Medium range lagged average forecasts. *Proc. Ninth Conf. on Probability and Statistics in Atmospheric Sciences,* Virginia Beach, Amer. Meteor. Soc., 130–136.

Dyson, F. J., 1962: A Brownian-motion model for the eigenvalues of a random matrix. *J. Math. Phys.,* **38,** 1191–1198.

Epstein, E. S., 1969: Stochastic dynamic prediction. *Tellus,* **21,** 739–759.

——, 1982: Detecting climatic change. *J. Appl. Meteor.,* **21,** 1172–1182.

Fisher, R. A., 1936: The use of multiple measurements in taxonomic problems. *Ann. Eugen.,* **7,** 179–188.

Hannoschöck, G., and C. Frankignoul, 1985: Multivariate statistical analysis of a sea surface temperature anomaly experiment with the GISS general circulation model I. *J. Atmos. Sci.,* **42,** 1430–1450.

Hasselmann, K., 1979: On the signal-to-noise problem in atmospheric response studies. *Meteorology Over the Tropical Oceans,* D. B. Shaw, Ed., Roy. Meteor. Soc., 251–259.

Hayashi, Y., 1982: Confidence intervals of a climatic signal. *J. Atmos. Sci.,* **39,** 1895–1905.

Hoffman, R. N., and E. Kalnay, 1983: Lagged average forecasting, an alternative to Monte Carlo forecasting. *Tellus,* **35A,** 100–118.

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

Leith, C. E., 1971: Atmospheric predictability and two-dimensional turbulence. *J. Atmos. Sci.,* **28,** 145–161.

——, 1973: The standard error of time-average estimates of climatic means. *J. Appl. Meteor.,* **12,** 1066–1069.

——, 1974: Theoretical skill of Monte Carlo forecasts. *Mon. Wea. Rev.,* **102,** 409–418.

Livezey, R. E., 1985: Statistical analysis of general circulation model climate simulation: Sensitivity and prediction experiments. *J. Atmos. Sci.,* **42,** 1139–1149.

——, and W. Y. Chen, 1983: Statistical field significance and its determination by Monte Carlo techniques. *Mon. Wea. Rev.,* **111,** 46–59.

Lorenz, E. N., 1969: The predictability of a flow which possesses many scales of motion. *Tellus,* **21,** 289–307.

Mardia, K. V., J. T. Kent and J. M. Bibby, 1979: *Multivariate Analysis.* Academic Press, p. 72.

Miller, R. G., 1962: *Statistical Prediction by Discriminant Analysis. Meteorol. Monogr.,* Vol. 4, No. 25, Amer. Meteor. Soc., 54 pp.

Parthasarathy, B., and D. A. Mooley, 1978: Some features of a long homogeneous series of Indian summer monsoon rainfall. *Mon. Wea. Rev.,* **106,** 771–781.

Preisendorfer, R. W., and C. D. Mobley, 1984: Climate forecast verifications, United States mainland, 1974–83. *Mon. Wea. Rev.,* **112,** 809–825.

Shukla, J., 1981: Dynamical predictability of monthly means. *J. Atmos. Sci.,* **38,** 2547–2572.

White, G. H., 1980: Skewness, kurtosis, and extreme values of Northern Hemisphere geopotential heights. *Mon. Wea. Rev.,* **108,** 1446–1455.