STATISTICAL CONSIDERATIONS IN THE EVALUATION
OF CLIMATIC EXPERIMENTS WITH GENERAL
CIRCULATION MODELS

John A. Laurmann

and

W. Lawrence Gates

December 1975

P-5503
The Rand Paper Series

Papers are issued by The Rand Corporation as a service to its professional staff. Their purpose is to facilitate the exchange of ideas among those who share the author's research interests; Papers are not reports prepared in fulfillment of Rand's contracts or grants. Views expressed in a Paper are the author's own, and are not necessarily shared by Rand or its research sponsors.

The Rand Corporation
Santa Monica, California 90406
Numerical calculations made with general circulation models contain short-term fluctuations that need to be taken into account when considering the significance of long-term climatic means. In particular, when such models are used in the study of climatic change resulting from altered forcing, quantitative evaluation of the effects of the fluctuations, as measured by the variance of model generated climatic estimates, is required. The fluctuations are of large magnitude, and necessitate extended simulations or repeated Monte-Carlo calculations in order to determine accurate values of climatic means.

This report develops a rationale for planning the extent of climate simulation trials needed to establish the significance of any proposed climatic change experiment. The topic is treated from an elementary statistical viewpoint and the report is in part written to serve as an introduction to statistical questions that are important for climate modeling. Conclusions concerning the planning of climatic change experiments include a quantitative prescription for determining the extent of model calculations required to establish the presence of a climatic change between two numerical experiments at any chosen confidence level.
CONTENTS

PREFACE ...................................................... iii

Section
1. INTRODUCTION ......................................... 1

2. STATISTICAL TOOLS AND CONCEPTS FOR THE PRIMARY
   ANALYSIS OF CLIMATIC CHANGE EXPERIMENTS ........ 5
   2.1. Dealing with Trends and Cycles .................. 7
   2.2. Dealing with Persistence ......................... 9

3. EXPERIMENTAL DESIGN ............................... 19
   3.1. Use of the t-test of Significance ............... 19
   3.2. The Choice of Δ ................................ 22
   3.3. Discussion of Approximations ................... 24

4. EXPERIMENTAL ANALYSIS ............................ 28
   4.1. Evaluation Based on the Difference Between
        Random Perturbations of the Same
        Climatic State ..................................... 32
   4.2. Evaluations Based on Differencing the Control
        and Experiment Runs, Assuming No Change
        in Cycles or Trends ............................... 35
   4.3. Spatial Averaging ................................ 36

5. SUMMARY OF EXPERIMENTAL PROCEDURES ........... 37
   5.1. Determination of the Required Running Time
        for a Climatic Experiment ....................... 37
   5.2. Evaluation of Climatic Experiment Results ...... 39
       5.2.1. Control well known .......................... 39
       5.2.2. Differencing between control and
                climatic experiment ....................... 40
   5.3. Possible Difficulties ............................ 40
       5.3.1. Persistence time $T_0$ greater
                than the running time ($T^*$) .......... 40
       5.3.2. Changed cycles or trends in the
                climatic experiment ....................... 41
       5.3.3. Non-independence of random
                perturbation runs .......................... 41
       5.3.4. Variance estimates too large ............... 41

ACKNOWLEDGMENTS ........................................... 43

REFERENCES ................................................. 44
1. INTRODUCTION

General circulation models that attempt to simulate the dynamics of the earth's atmospheric motions exhibit a basic feature that leads to practical difficulties in their use for climatic change simulation. This characteristic, which is a reflection of real atmospheric behavior, arises from the fact that short term, day-to-day fluctuations in the elements describing atmospheric change are of the same order of magnitude as the longer term interannual changes (exclusive of the regular seasonal variations). Estimates of climatic average states—monthly, seasonal, or longer—are thereby corrupted by the presence of these short term fluctuations which are climatically uninteresting. The rest of this report will amplify upon this point and describe approaches to dealing with the situation.

If atmospheric simulations could be made with sufficient refinement, so as to yield representations that were correct in day-to-day detail for many months at a time, sufficiently long term integrations of the model equations would yield estimates of climatological means to any desired accuracy. However, and quite apart from the computational expenses likely involved in running such a detailed model if it were to exist, experience has shown that difficulties of simulation fidelity limit the practical range of predictability on a daily basis to a few weeks at most. After this time, although it is believed that the behavior of the atmosphere as described by the model is correct in a general way, its detailed chronology is not, and all phase coherence with the initial starting state of the model is lost. In long term simulations we are therefore dealing with a model whose averaged statistical output alone is relevant to the real world.
The necessity for treating any model experiment in statistical terms is in itself not worrisome, but the fact that the magnitude of the short time period changes are extremely large compared with typical interannual changes in monthly or seasonal averages, means that detection of the informational content of the latter in the presence what is now the "noise" of the former, can be a difficult and lengthy process. The logic of this observation can be readily appreciated by noting that a typical (non-diurnal) change in temperature accompanying a weather front at mid-latitudes can be tens of degrees, much larger than the interannual changes in average seasonal temperatures. Only by averaging the data over sufficiently long times can such interannual changes be distinguished from statistical fluctuations caused by the day-to-day changes. Later in this report we shall discuss the quantitative implications of this fact, covering in particular the design of climatic change experiments. Suffice it to say here that model running times will have to be made long enough to reduce the variance of climatic means arising from the short term fluctuations to manageable proportions.* When the magnitude of the climatological shifts being simulated are of "normal" magnitude (i.e., similar to naturally occurring interannual changes), the length of run needed is likely to prove a major impediment to drawing valid conclusions. Only when major "catastrophic" climatic events are studied can the statistical effects of the short term fluctuations be relegated to a minor role.

*As we shall describe later, these short term fluctuations are not describable as "white" noise. Their persistence ("red" noise) characteristics means that even larger running times are required to achieve this reduction in variance.
This report will deal in introductory fashion with the statistical aspects of using general circulation models that arise from the above considerations, aiming towards a reasonably simple approach to the design of climatic simulation experiments. We shall have as primary aims:

(a) the assessment of the order of magnitude of the scope of numerical experiments required for statistical accuracy in climate modeling.

(b) the application of what is presently known of atmospheric statistics to the climate simulation problem.

(c) the derivation of a logical testing procedure for the determination of the significance of climatic change simulation experiments.

Experience in the statistical properties of climatic simulations with general circulation models is quite limited; in our discussions, and in particular in our recommendations for item (c) above, we will have to make some basic, unproven assumptions. The major one of these is the existence of a stochastically stationary solution of the atmospheric motions* as described by the model. There are two reasons for this; firstly, all the statistical tools—in particular the statistical significance tests—require reduction to a stationary, independent sample set. Secondly, with the computer capacity currently available or available in the foreseeable future, general circulation models are limited in standard practice to a few months or seasons of (simulated) running time. Significant non-stationarity on a time scale of a month or longer would probably render the general circulation model unusable.

---

*We take it for granted that known regularities (i.e., the diurnal and seasonal cycles) have been extracted from the climatic time series.
as a tool for climatic studies.* Our approach at present is thus to assume this is not so, though verification is needed at some stage. A statistical significance testing procedure can be devised for such an objective, but will not be considered here.

The rest of this report will cover briefly the more elementary statistical tools needed to deal with the problems we have outlined, going into detail only for the particular approaches that it is felt are needed at the present stage of development of understanding the characteristics of climatic modeling.

---

*But we need to emphasize the adjective "significant." Enough climatic simulations have been made to satisfy us that approximate stationarity exists, and the issue then reduces again to deciding if a possible non-stationary effect is large enough to mask the presence of a systematic climatic change that the model could be producing.
2. STATISTICAL TOOLS AND CONCEPTS FOR THE PRIMARY ANALYSIS OF CLIMATIC CHANGE EXPERIMENTS

The philosophy adopted here for the coverage of statistical questions is based on a goal for design of numerical experiments that will readily demonstrate the presence, or absence, of a significant climatic change consequent on a change in forcing of the model. We will thus try to avoid having to deal with marginal situations that might require highly refined (and probably controversial) statistical test procedures. Rather, the objective will be to produce experimental design criteria that will avoid such situations. It follows from this tenet that our statistical analyses will not be unduly sophisticated, although we will try to be precise in the statements of assumptions made and the limitations of the methods, so that these can be reconsidered as necessary for refinement.

A further restriction on the scope of the analysis should also be mentioned now. We shall develop the recommended procedures for point values of the model variables, studying only the behavior of these as a time series during the running period of the model. This restriction to autocorrelations alone, in contrast to the study of spatial correlations, omits what we expect eventually to prove to be highly relevant and important processes for understanding the physical basis of the models statistical properties. However, if local estimates of possible climatic changes are being sought, the prescriptions we shall give will be minimum requirements for model performance. Grosser questions concerning global or regional behavior will likely require different and

*To be defined by the experimenter.
perhaps less constraining experimental procedures, as well as more
difficult considerations of the spatial statistical properties of the
model.

In order to apply the statistical test criteria, we shall discuss
three general requirements. The first of these is statistical station-
arity of the time series being produced, as already mentioned. The
basic technique we shall employ for dealing with this is to treat only
the differences between a standard "normal" climatic run of the model
and a perturbed state. Secondly, in evaluating standard deviations of
averaged variables, statistical interdependence, if any, of the samples
making up the means has to be accounted for. A principal part of what
follows will deal with phenomenon of "meteorological persistence" as
it affects this aspect of climatological error assessment. Finally,
most statistical tests require either a normal distribution of variables,
or a sufficiently large number of samples for which the test statistic
is evaluated. We know the former is not the case for most meteorologi-
cal variables, and, because of the persistence, the effective sample
size may be small. In marginal situations, therefore, we may find it
difficult to assess possible climatic change. However, as mentioned at
the start of this section, a primary aim is to develop criteria for nu-
merical experiment procedures that give unambiguous results, so that the
more difficult nonparametric statistical techniques for refined testing
of non-normal small samples will hopefully not be needed.

The rest of this section will cover these three points in more de-
tail, with the quantitative application to the climatological case being
relegated to Section 3.
2.1. **Dealing with Trends and Cycles**

Climatic change experiments with general circulation models are usually envisaged as comparisons between a "normal" climatic state and one modified by an alteration (usually time independent) of the forcing of the atmospheric model by a change in boundary conditions. If both the control and climatic change experiment could be run long enough, presumably a time series for the model output variables that is statistically stationary could be established. It is immediately clear, however, that if, for example, monthly means are being sought, the seasonal changes (if simulated by the model) demand many years of model running to obtain repeated samples of the same external forcing values. For this reason alone, therefore, as long as we are not interested in the seasonal cycle for its own sake, it is important to eliminate this cyclical feature from consideration. We may do this either by subtracting out the seasonal changes, assuming these to be known, or by dealing only with differences between control and climatic experiment runs. The latter technique eliminates not only the seasonal march, but also diurnal oscillations and any other trends that might exist and remain unaltered in the two cases. Since the variances (and higher order moments), as well as the averages can change between the control and perturbation, we should be aware that in treating only the differences any statistical evaluations will correspond to joint properties of the two runs. Although this may not be a handicap in the application of statistical significance tests, it is likely that there will be an interest in the absolute statistical properties for the climatic experiment case. If this is so, either we need to treat the non-stationary time series for control
and perturbation individually, or we can make use of the fact that we will have much more extensive knowledge of the control run than the perturbation. In this event we can subtract from the perturbation these relatively well-known results of the control run. The latter should have negligible residual short-term statistical fluctuations, but will still contain all regular trends and cycles, so that the difference between this averaged control run and a perturbation experiment will contain statistical properties of the perturbation alone.

The averaging process described here will likely be obtained by performing repeated runs with random perturbations on the initial conditions for the control run. This process is necessary if the simulation includes the long term changes in forcing resulting from the seasonal march, in contrast to models that employ a fixed sun angle, in which case the necessary smoothing can be obtained by running the model for a long enough period. The method of randomly perturbing the initial state in order to build up a large enough ensemble for establishing statistical significance is already established as standard for the Rand model, and we will assume it to be used in all runs, including control and climatic experiment. Section 4 deals with calculation of the statistical properties using the differencing technique. As will be discussed here, dealing only with differences also can give better estimates of errors.

In running general circulation models for simulating climatic conditions, a normal objective is to try to run the models long enough to establish a steady average state (allowing for the regular seasonal shifts). For initial conditions close to a real atmospheric state it is thought that about two weeks is needed to reach such a statistically
steady condition. In most model runs we are therefore talking about a month or two of running time, with enough repeats with randomly perturbed initial conditions to reduce the variance to the desired values. However, it is possible that small systematic non-stationarity exists in such a running period (even after extraction of diurnal and seasonal changes), and we have not as yet built up sufficient knowledge of numerical results to determine if this is so. Hopefully, as experience grows we can make a better evaluation.

The construction of a large ensemble of test samples using the random perturbation approach can certainly establish the relative shift of a climatic average compared with a trend (if such exists), but only for the period under investigation. Hence, if a much longer term trend or cycle is induced (comparable to a biennial oscillation), we will be unaware of it unless long term integrations are performed. Here we shall assume that the differences between runs form statistically stationary time series.

2.2. Dealing with Persistence

We have already indicated that most statistical significance criteria require independence of the test samples. In addition, the well known formula

\[ \sigma(\overline{x}_i) = \sigma(x_i)/\sqrt{N} \]  \hspace{1cm} (1)

for the reduction in the standard deviation of the mean \( \overline{x}_i \) of \( N \) observations of a variable \( x_i \), as compared with the sample standard deviation \( \sigma(x_i) \), requires that the \( N \) values be statistically independent. Since
it is precisely by estimating such time averages \((x_i)\) now forming the climatic time series) that we intend to assess the effect of climatic experiments, it is crucial that correct assumptions are made concerning the statistical independence or interdependence of the climatic time series.

In climatic experiments which yield a stationary time series, the samples \(x_i\) of a climatic variable \(x\) can theoretically be taken at arbitrarily small intervals. The requirement of independence for use of Eq. (1), for example, means that the magnitudes of \(x_i\) are randomly distributed at all time scales, i.e., if the time series is spectrally analyzed, all frequencies are equally likely, and consists only of so-called "white" noise. In fact neither the simulation nor the real atmosphere has this ideal characteristic, and in fact we find that successive values of all atmospheric variables are statistically dependent. Although no theories have yet appeared that synthesize the detailed motions into a statistical description of these dependencies, empirical evidence as to their nature does exist. This meteorological feature is so important for climatological assessment that a brief description is given here.

Following a number of interpretations of observations of time series of local meteorological variables, Lorenz (1973) extended these statistical analyses to a hemispheric basis. All of the studies seem to have revealed only one type of statistical structure relating successive values of the stationary part of meteorological time series. This was termed "red" noise by Lorenz, since its spectral form is dominated by low frequencies. In mathematical language red noise corresponds to the simplest form of Markov process in which
\[ x_{i+1} = kx_i + \varepsilon(t) \]  

(2)

where \( k \) is a constant, and \( \varepsilon(t) \) is a stationary random variable consisting of white noise only. Such a relationship has often been used in meteorological literature as a description of meteorological "persistence," and connotes non-independence of sequential values in the time series.

This simple law leads to simple statistical analyses, and if the data approximate it sufficiently well, it is therefore a favored approach. However, \( \sigma^2(\overline{x}_i) \) -- the most important quantity to evaluate for assessing the results of climatic experiments -- can be computed from the experimental output without the assumption of red noise, or indeed any other specific statistical process.* Thus, as proven for example by Anderson (1971, p. 44), we have the general expression

\[ \sigma^2(\overline{x}_i) = \frac{\sigma^2}{N} \sum_{\nu=-(N-1)}^{N-1} \left( 1 - \frac{|\nu|}{N} \right) \rho_\nu(x) \]  

(3)

for the variance \( \sigma^2(\overline{x}_i) \) of the mean over \( N \) samples of the variable \( x \) in terms of the "lag" \( \nu \) and serial correlation coefficient \( \rho_\nu \). Here

\[ \rho_\nu(x) = \frac{1}{N-\nu} \sum_{i=1}^{N-\nu} (x_i - \mu)(x_{i+\nu} - \mu) / \sigma^2(x) \]  

(4)

where \( \sigma^2 \) is the sample variance of \( x_1 \) defined as

*Though it should also be stated that all statistical significance tests based on estimated values of \( \sigma(\overline{x}_i) \) also require some information on the nature of the statistical distribution of the variable \( x \).
\[
\sigma^2(x) = \frac{1}{N} \sum_{i=1}^{N} (x_i - \mu)^2
\]  

(5)

We should note here that Eqs. (3) through (5) are all estimates of the properties of the complete (infinite population) time series, of which the \(x_i\), \(i=1, \ldots, N\) are a part. In fact these particular formulas can be shown to be "unbiased" estimates, i.e., the distribution of a series of independent measurements of \(x\) inserted into the equations will be symmetric about the population means. In practice the mean \(\mu\) will also be estimated as

\[
\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i
\]

from partial information, and will not be known exactly. Use of \(\bar{x}\) in place of \(\mu\) in Eqs. (4) and (5) leads to biased estimates, and to difficulties in the statistical analysis which we shall discuss further in Section 4. For the time being we shall treat only the properties of the total population, in contrast to those of a limited sample drawn from it.

The continuous equivalent of the discrete equation (3) for \(\sigma^2(\bar{x})\) is (Lee, 1960, p. 286)

\[
\sigma^2(\bar{x}) = \frac{\sigma^2}{T} \int_{-T}^{T} \left(1 - \frac{\tau}{T}\right) \rho(x) d\tau
\]

(6)

*By computing expected values by averaging over fictitious ensembles, or repeated trials of the indicated mathematical operations with randomly selected sequences of the \(N\) values of \(x\).*
where the lag of the correlation is now \( \tau \) time units, and \( T \) is the sampling time corresponding to the \( N \) samples in Eq. (3). Note that for large sampling times and \( T \gg \tau \), (6) approaches

\[
\sigma^2(\bar{x}) = \frac{\sigma^2}{T} \int_{-\infty}^{\infty} \rho_{\tau} \, d\tau
\]  

(7)

Once we are satisfied from the statistical test procedures available that the series describing the climate experiment is stationary, (3) or (6) give valid estimates of the standard deviation of the time series. In particular, if we apply them to the difference between a control run of the model and a climatic experiment (and random perturbations of these), they will then give the standard deviation of the mean climatic shift produced by the experiment. A simple trial for checking that an adequate length of the time series has been used to calculate the climatic mean is to estimate its variance for progressively larger time intervals. The variance should decay toward zero as this time increases.

Equations (3) or (6) applied to the calculation of variance in the mean of a set of interdependent samples can be also used to compute an "effective" number of independent samples \( N' \) for the sample length used. We define this by assuming that the relationship

\[
\sigma^2(\bar{x}_1) = \frac{\sigma^2}{N}
\]  

(8)

for \( N \) independent measurements (obtainable from (3) by putting \( \rho_v = 0 \), \( v \geq 1 \)) is applicable when \( N \) is replaced by a fictitious number \( N' \) such that
\[ \sigma^2(\bar{x}_1) = \frac{\sigma^2}{N'} \] (9)

where, from (3)

\[ N' = \left( \frac{\sigma(x)}{\sigma(x)} \right)^2 \]

\[ = N \left[ \sum_{\nu = -(N-1)}^{N-1} \left( 1 - \frac{|\nu|}{N} \right) \rho_{\nu} \right]^{-1} \] (10)

Analogously, if we assume the \( N' \) samples are equally spaced in time, we can define an effective time \( T_o \) between independent samples by

\[ N'T_o = T \] (11)

so that (10) gives

\[ T_o = \frac{T}{N} \sum_{\nu = -(N-1)}^{N-1} \left( 1 - \frac{|\nu|}{N} \right) \rho_{\nu} \] (12)

An equation corresponding to (12) for the continuous case is, from Eqs. (6) and (9)

\[ T_o = \int_{-T}^{T} \left( 1 - \frac{|\tau|}{T} \right) \rho_{\tau} d\tau \]

\[ \rightarrow \int_{-\infty}^{\infty} \rho_{\tau} d\tau \text{ for } T \rightarrow \infty \] (13)

Although we would recommend the use of the above equations as the general practice to follow for climate experimentation, in the design
of climate experiments we need to be able to make (at least rough) quantitative assessments of the running times needed to reduce the variances of the means to a desired level. We need therefore to be able to estimate \( \sigma(\bar{x}) \), or the parameters \( N' \) and \( T_0 \), and for this purpose we need information on the nature of the serial correlation \( \rho_T \). To obtain this we assume the numerical output of the model approximates red noise, and follow the description given by Leith (1973).

Considering the case of a stationary series with zero mean, the Markov process, described by Eq. (2), implies that

\[
x_{i+1} x_1 = x_1 (k x_1 + \varepsilon)
\]

\[
x_{i+j} x_1 = k^j x_1^2 + x_1 F(\varepsilon)
\]

where \( \varepsilon = (1 + k + k^2 + \cdots + k^{j-1}) \). Taking an ensemble average (denoted by \( < > \)) of these equations we get

\[
< x_{i+1} x_1 > = k < x_1^2 >
\]

\[
< x_{i+j} x_1 > = k^j < x_1^2 >
\]

(14)

since we assume \( < x_1 > = 0 \). Equation (14) states that the red noise covariance \( < x_{i+j} x_1 > \) at lag \( j \) is related to the covariance at lag 1 \( < x_{i+1} x_1 > \) by

\[
\frac{< x_{i+j} x_1 >}{< x_1^2 >} = \frac{< x_{i+1} x_1 >}{< x_1^2 >}
\]

(15)

From the usual definition of the correlation coefficient (for a series with zero mean),
\[ \rho_j = \frac{< x_i^j x_i >}{\sigma^2} \]

and since \( \sigma^2 = < x_i^2 > \), (15) may be written

\[ \rho_j = \rho_j^1 = k_j^1 \quad , \quad (16) \]

using (14).

The continuum limit of (16) is simply

\[ \rho_T = k_T \quad (17) \]

or

\[ \rho_T = e^{-\alpha T} \quad (18) \]

where \( \alpha \) is a constant equal to \( -\phi_n K \).

We can now estimate the standard deviation for continuous or discrete samples in the presence of red noise using Eqs. (3) and (6). In the discrete case we have

\[ \sigma^2(x_i) = \frac{\sigma^2}{N} \sum_{\nu = -(N-1)}^{N-1} \left( 1 - \frac{|\nu|}{N} \right) \rho_1^\nu \quad (19) \]

which to good approximation for large \( N \) can be shown to equal (see Mitchell et al., 1966)

\[ \sigma^2(x_i) = \frac{\sigma^2(x)}{N} \frac{1 + \rho_1}{1 - \rho_1} \quad (20) \]

which from Eq. (9) yields the effective number of independent samples.
\[ N' = N \left( \frac{1 - \rho_1}{1 + \rho_1} \right) \]  \hspace{1cm} (21)

or a time between effectively independent samples
\[ T_o = t_o \left( \frac{1 + \rho_1}{1 - \rho_1} \right) \]  \hspace{1cm} (22)

where \( t_o = T/N \) is the sampling time. Here we note \( \rho_1 = k = e^{-\alpha} \).

The corresponding results for continuous sampling are derivable by direct integration, introducing (18) into (6). There results (c.f., Leith, \textit{ibid})

\[ \frac{\sigma^2(\bar{x})}{\sigma^2(x)} = \frac{2}{\alpha \tau} \left( 1 - \frac{1}{\alpha \tau} \left( 1 - e^{-\nu \tau} \right) \right) \]  \hspace{1cm} (23)

Using (9), this equation defines an effective sample size \( N' \), and using (11) an effective sampling time \( T_o \). For large \( T \), (23) reduces to

\[ \frac{\sigma^2(\bar{x})}{\sigma^2(x)} = \frac{2}{\alpha T} \]  \hspace{1cm} (24)

and from (10) and (11)

\[ \frac{\sigma^2(\bar{x})}{\sigma^2(x)} = \frac{T_o}{T} = \frac{2}{\alpha T} \]

so that

\[ T_o = \frac{2}{\alpha} \]  \hspace{1cm} (25)
We note that this result is also obtainable directly from the asymptotic relationship, Eq. (7), valid for the red noise spectrum, by putting $\rho_\tau = e^{-\alpha \tau}$. 
3. EXPERIMENTAL DESIGN

In this section we shall assume that the time series of the model output for climatic change is both stationary and has zero mean. Later, in Section 4, we shall discuss specific methods for obtaining such an output for analysis.

3.1. Use of the t-test of Significance

We wish to determine the length of model run required for detectability of a climatic change of the variable $x$ from an assumed average $\mu_0$, resulting from a change in forcing of the model. Estimates of $\mu_0$ are generally available from the ensemble of runs that are performed with standard climate models. We also make an assumption that the standard deviation for the perturbed climate equals that of the control. We do not expect this to be true when significant climatic changes occur, but it is unlikely to produce large errors in the design requirements dealt with here. With these assumptions, the appropriate test statistic is student's $t$ ratio (Hoel, 1962, p. 275):

$$t = \frac{\bar{x} - \mu_0}{s} \sqrt{N - 1}$$  \hspace{1cm} (26)

where $\bar{x}$ is the estimate of the mean $\mu$ from $N$ measurements of $x$, and

$$s^2 = \frac{\sum (x - \bar{x})^2}{N}$$ is an estimate of the standard deviation $\sigma$, and

*For the long term output. The observed mean with any specific set of samples for such a series will not (necessarily) be zero, but will approach zero as the sample number approaches infinity.*
\[ N' = \text{effective number of independent samples forming } \bar{x} \]

(allowing for persistence in the time series).

The number of degrees of freedom required in using tables of the \( t \) function (26) is \( N' - 1 \) rather than \( N' \). Tabulated values of this function* enable us to determine the number of samples \( N' \) required to establish

\[ |\mu - \mu_o| \geq \text{a chosen } \Delta \]  

(27)

with a selected percentage probability, where \( \mu \) is the average of \( x \) in a climatic change experiment. The results of this procedure are given in Fig. 1, where we have plotted the required sample number \( N' \) against \( k \), where

\[ k = \frac{s^2}{\Delta^2} \]  

(28)

with \( \Delta = |\bar{x} - \mu_o| \)

for a series of different percentage confidence levels.

We note that in using (26), we are assuming all errors arise from uncertainty in the perturbation run, not the control. If this is not the case, more samples \( N' \) will be required than shown in Fig. 1. Generally speaking, if the number of samples is the same for both the control and perturbation, this doubles the total variance of the estimate of \( \Delta \), and means that running times have to be twice those shown in Fig. 1 to achieve equivalent confidence levels.

*For example, see Hoel, ibid., p. 402.
Fig. 1—The number of independent samples $N'$ required for a given confidence level as a function of the ratio $k = s^2/\Delta^2$ where $s^2$ is the experimental variance and $\Delta$ is the difference between the experimental and population mean.
3.2. The Choice of $\Delta$

Figure 1 gives the number of samples needed to achieve a specified confidence level as a function of the parameter $k = (s/\Delta)^2$. We wish to discuss whether there exists a rational choice for a particular value of $k$ to use in the model experiments.

Suppose we are primarily interested in year-to-year climatic variation in contrast to longer period changes. As a minimum requirement for practical relevance, we are therefore concerned with interannual climatic shifts that are at least as large as the climatic noise associated with the particular climatic averaging period that is being considered.* The latter, when considering year-to-year events, could be monthly, seasonal, or annual means. This criterion requires that

$$\Delta \geq \sigma_T$$

(29)

where $\sigma_T$ is the standard deviation of the mean (for both control and perturbation) over the chosen averaging period.

Let us choose an effective persistence time $T_o$, which, either with a red noise assumption, or with a sufficiently long sampling time $T$, means that Eq. (29) can be written (see Eq. (11))

$$\Delta \geq \sigma_T = \sqrt{\frac{T_o}{T}} \sigma$$

(30)

and when we approximate $\sigma$ in (30) by the unbiased estimate $s\sqrt{N'/N' - 1}$, the student $t$ ratio becomes

*Indeed, in the real atmosphere, where, of course, only one realization of a particular year is possible, only changes of this magnitude are detectable in a single year.
Fig. 2—The number of experiments n required to achieve a given confidence level as a function of the ratio $T/T_o$, where $T$ is the averaging time and $T_o$ is the effective independent sampling interval.
\[ t = \frac{\Delta}{\sigma_T} \sqrt{\frac{T_0}{T}} N'^{1/2} \]

\[ = \frac{\Delta}{\sigma_T} \sqrt{\frac{T_r}{T}} = \frac{\Delta}{\sigma_T} \sqrt{n} \]

where \( T_r = T_0 N' \) is the running time required to get \( N' \) independent samples, and \( n \) is the ratio of this time to the averaging time \( T \). Smoothed values of \( n \) required to obtain different confidence levels are plotted in Fig. 2.

Figure 2 shows a weak dependence of the results on values of the parameter \( T/T_0 \), the ratio of the climatic averaging period to time between effectively independent samples. In the limit \( T/T_0 \to \infty \), we have the result that the number of climatic averaging periods required to reach a selected confidence level in detecting a change equal to the noisiness of the average is independent of the actual noise level. For example, under these circumstances, in order to obtain 95 percent certainty of detecting a climatic shift of this magnitude we need a perturbation model run of about four times the climatic averaging period.

3.3. Discussion of Approximations

In the above analysis we have assumed the normality of distribution of the climatic series and the equality of variances of control and perturbation runs. Neither of these assumptions are likely to affect any of the prescriptions for optimizing the experimental design, since the latter are not in themselves error assessments. After the model outputs have been made, the establishment of accurate measures of error and statistical significance can be a much more critical task, especially
when the variance of the supposed climatic changes is close to the noise level. As a general principle of good design, we would hope that any climatic change that we wish to demonstrate will be obvious when the experiment is correctly designed, so that the difficult and sometimes questionable statistical techniques for detecting small signals in a noisy environment need not be resorted to.

The t-test is a strictly valid test for small samples \((N')\) only if the underlying distribution is Gaussian, or, more accurately and less specifically, if \(\bar{x} - \mu_0\) is distributed normally and \(s\) is distributed as \(\chi^2\). The central limit theorem states that the latter is exactly true as \(N' \to \infty\), no matter what the nature of the distribution of \(x\).

In application, it seems that a value of \(N'\) between 10 and 30 is the lower bound for general validity of the t-test. However, this criterion is based on a judgment for accurate estimates of statistical significance, whereas an error as large as 10 percent, say, in the appropriate choice of \(N'\) in an experimental design is quite minor. In general we can claim that the experimental design requirements are much coarser than those for quantitative evaluation of experimental results.

Since we have no a priori information on the nature of the underlying statistical distribution, the only recourse, if we are not satisfied with the above assurances of adequacy of the t-test, is to use a nonparametric test (see, for example, Bradley, 1968). Because of the less analytic nature of such tests, it is not clear, however, whether similar estimates of required sampling times can be made with them. We prefer to leave this question until results of perturbation runs on a scale dictated by the t-test analysis have been obtained.
The t-test as given in Eq. (26) requires identity of the variances in both control and perturbation runs. We believe the correction for this assumption (which we know is indeed false) is minor for experimental design purposes, unless we are searching for very large departures from the control, in which case the design task itself is a minor one.

There is, however, an alternative to the t-test, when the variances are unequal, namely, use of the Behrens–Fisher statistic, which in its most general form refers to a comparison between the population means \( \mu_x \) and \( \mu_y \) of two series with population standard deviations \( \sigma_x \) and \( \sigma_y \). The test statistic in this case under the null hypothesis \( \mu_x = \mu_y \) is

\[
t = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma_x^2}{N_x} + \frac{\sigma_y^2}{N_y}}} \tag{32}
\]

with number of degrees of freedom

\[
\nu = \frac{\left(\frac{\sigma_x^2}{N_x}\right) + \left(\frac{\sigma_y^2}{N_y}\right)}{\left(\frac{\sigma_x^2}{N_x}\right)^2 + \left(\frac{\sigma_y^2}{N_y}\right)^2} - 2 \tag{33}
\]

where \( N_x \) and \( N_y \) are the (effective) number of (independent) samples for the perturbation and control runs, and the standard deviations \( \hat{\sigma}_x \) and \( \hat{\sigma}_y \) are unbiased estimates [i.e., \( \hat{\sigma}_x^2 = \frac{N_x}{N_x - 1} \sum_{i=1}^{N_x} (x_i - \bar{x})^2 \)].

When the control mean is well known, \( \sigma_y/\sqrt{N_y} \) is small, and in the limit,

\*

*See Hoel, *ibid.*, p. 279.*
\[ t \sim \frac{\bar{X} - \mu}{\sigma_x / \sqrt{N_x}} \]

and

\[ \nu \sim N_x - 1 \]  \hspace{1cm} (34)

as in the standard t-test, though in our application in Section 3.2 we took \( \hat{\sigma}_x \) to correspond to the standard deviation of the control run \( \hat{\sigma}_y \), assumed equal to that for the perturbation. Since we have no a priori knowledge of \( \sigma_x \), this in fact is the only procedure available for the design of experimental conditions. In carrying through a complete a posteriori significance test, the t statistic given in (6) can in fact be used, though the qualifications of normality discussed in the above will apply.
4. EXPERIMENTAL ANALYSIS

We consider here in more detail the methods for developing the stationary series and the climatic change information it is designed to detect, assuming the number of model runs have been decided upon following the rationale developed above.

In climatic change experiments we desire to establish the change in mean conditions of a (number of) climatic variables as compared with a "normal" climate established in control runs. The climatic time average of interest will likely be recognized prior to the experiment, but results with varying averaging times may also be important in order to study the reduction in variance of the mean as the averaging time increases. Geographical averages as well as local values will no doubt be desired, but the variances of local quantities are normally larger than geographically averaged ones, so we may find computation time limitations restricting the significance we may be able to give to local climatic change estimates.

If experiment running times are adequate we would expect not to have to conduct careful statistical significance tests on the output. But we will still need to compute the variance of the new climatic state, which may very well change from that for the control run. To calculate the variances, and most particularly the variance of a time averaged output (or the variance of a change in output), we have to contend with the persistence characteristic of the atmosphere. There are three possibilities which may be mentioned:

1. If from lag correlation estimates, we can establish that the persistence corresponds to red noise, it is easy to calculate
from the first few lag correlations the effective number of independent samples (Eq. (21)), and hence the variance of the time average relative to the sample variance.

2. We can compute the time average for different time sections of the data, provided the latter are longer than the averaging time of interest (or if repeated runs are available). In a sense this is the most direct approach, but we need to know that the averaging time is larger than the persistence time $T_0$, and there is no way of assuring ourselves of this without other tests.

3. We can use serial correlation calculations to estimate the standard deviation of the mean, using Eq. (4). This is applicable no matter what the statistical character of the random variations. The method involves computing lag correlations $\rho_\tau$ which should decrease with increasing lag $\tau$ (unless a trend or oscillation exists in the data), so that only a restricted number of these need be used in the summation for the calculation of $\sigma(\bar{x}_i)$.

As a general procedure we would recommend procedure (3), though if in the process of calculating the correlations we find the noise to be well represented by the red spectrum, we can shortcut the calculations for $\sigma(\bar{x})$ by using the formulae based on red noise. There is, however, a basic difficulty in using any of the equations which arises from the error we make in estimating the population mean $\mu$ from a restricted sample of the total time series. For example, use of the conventional expression
\[ r_\nu = \frac{1}{N - \nu} \sum_{i=1}^{N-\nu} (x_i - \bar{x}_i)(x_{i+\nu} - \bar{x}_i)/s^2(x) \]  

(35)

with

\[ \bar{x}_i = \frac{1}{N} \sum_{i=1}^{N} x_i \]  

(36)

\[ s^2 = \frac{1}{N} \sum_{i=1}^{N} (x_i - \bar{x}_i)^2 \]  

(37)

as an estimate of \( \rho_\nu \), instead of the more precise formula (4) (in which \( \nu \) replaces \( \bar{x}_i \)) leads to a systematic underestimate of the true lag correlation \( \rho_\nu \) which for large sample number \( N \gg 1 \), and small correlations at large \( N \), \( \rho_\nu \ll 1 \), is approximately given by (Andersen, \textit{ibid}, p. 463)

\[ r_\nu = \rho_\nu - \frac{1}{N} \sum_{\nu=-(N-1)}^{N-1} \left(1 - \frac{\nu}{N}\right) \rho_\nu \]  

(38)

\[ \to \rho_\nu - \frac{1}{N} \sigma^2(x), N \to \infty \]  

(39)

Equation (38) shows that then the resulting bias is a constant independent of lag, and is more serious for small correlation coefficients (i.e., at large lags). The nature of this error is such as to completely eliminate the utility of the use of Eq. (6) for calculating the variance of the mean if Eq. (4) is used for \( \rho_\tau \) unless very large samples are available or the persistence is weak. Apart from increasing the sample size, the only technique for dealing with this difficulty is to attempt
to determine the spectral character of the noise. This again may be 
easy to do if the latter is red, but otherwise is very troublesome. 
Fortunately, in our case there is an alternative since we have avail-
able the a priori hypothesis that the governing equations for the 
model behavior do possess statistically stationary solutions. Hence, 
differences between sample runs of the same climatology, chronologi-
cally matched so as to eliminate diurnal and seasonal effects, repre-
sent a sample taken from a statistically stationary time series with 
a known value of the mean, r, namely zero. The biasing difficulty in 
analyzing the absolute values of the climatological series can there-
by be avoided by considering differences and applying equations (35) 
and (37) with $\bar{x} = 0$. The samples required to derive such a desired 
time series are normally obtained by random perturbations of some suf-
fi ciently early initial state of the model. Alternatively we can em-
ploy a longer model run, extract from it the cyclic effects and trend 
(if any), and form differences over equal climatic time averaging pe-
riods from the resultant stationary series, separated by intervals 
larger than the persistence time $T_o$. Either of these approaches re-
quires some consideration of the degree of persistence in the statistics.

Such series can be developed for both the climatic experiment and 
the control (the "normal" climate), and the effects of the climatic 
change then judged by the comparison of the means and standard devia-
tions of the two outputs. However, we can also consider the more di-
rect approach of treating the differences between the control and the 
climatic experiment as forming the time series of primary interest. In 
this case, not only do we have to deal with the problem of bias for a
series with non-zero mean, but we also need to assume that the climatic experiment does not induce a new trend or cyclic effect (for example, shifting the phase of the diurnal cycle), so that stationarity is not voided. The following deals with the equations to be used for estimating errors from the time series formed by these differencing techniques.

4.1. **Evaluation Based on the Difference Between Random Perturbations of the Same Climatic State**

Assuming the long-term mean (infinite sample) to be exactly zero and the series to be stationary, for the differenced variable \( \Delta = x_{1} - x_{1} \) of \( N \) samples, the unbiased estimates for the variance, the serial correlation and the variance of the mean over a finite time \( T \), are respectively given by

\[
s^2(\Delta) = \frac{1}{N} \sum_{i=1}^{N} \Delta_i^2
\]

(40)

\[
r_{\nu} = \frac{1}{N-\nu} \sum_{i=1}^{N-\nu} \frac{\Delta_i \Delta_{i+\nu}}{s^2}
\]

(41)

\[
s^2(\bar{\Delta}) = \frac{s^2}{N} \sum_{\nu=-(N-1)}^{(N-1)} \left(1 - \frac{\nu}{N}\right) r_{\tau}
\]

(42)

We can readily relate these properties to the original time series which we define as

\[
x_{i} = x_{i} + X(t)
\]

(43)

\(< x_{i} > = 0, < x_{i} > = X(t)\)
where \(< >\) denotes the ensemble average, so that \(X(t)\) represents the non-stochastic trends and cycles. Thus \(\sigma^2(x_1) = \sigma^2(x'_1) + \sigma^2(x)\), and in terms of the difference variable \(\Delta\) used above,

\[
x_{i1} - x_{i2} = x'_{i1} - x'_{i2} = \Delta_i
\]

Equations (40) - (42) directly yield the statistical properties of the variable \(x'\), as we now demonstrate. The variance estimate (40) in terms of \(x'\) has the ensemble average (i.e., the expected value)

\[
\frac{\langle (x'_{i1} - x'_{i2})^2 \rangle}{\langle \Delta^2 \rangle} = \frac{\langle \Delta^2 \rangle}{\langle \Delta^2 \rangle}
\]

where the bar refers to the time averaging operator \(\frac{1}{N} \sum_{i=1}^{N} \). Thus

\[
\langle \Delta^2 \rangle = \frac{\langle x'^2_{1} + x'^2_{2} - 2x'_{1}x'_{2} \rangle}{\langle x'^2_{1} \rangle + \langle x'^2_{2} \rangle - 2 \langle x'_{1}x'_{2} \rangle}
\]

\[
= 2 \sigma^2(x') - 2 \langle x'_{1}x'_{2} \rangle
\]

since \(\sigma(x'_{1}) = \sigma(x'_{2}) = \sigma(x')\). Moreover, \(x'_{1}\) and \(x'_{2}\) were selected at random and are uncorrelated, so that the last term vanishes and

\[
\langle \Delta^2 \rangle = 2 \sigma^2(x')
\]

(45)

It follows that \(\bar{\Delta}^2\) is an unbiased estimate of (twice) the variance of the stationary part of the parent series. It is interesting to note that the expression
\[ s^2(x) = \frac{(x - \bar{x})^2}{N} \]

for the variance of \( x \) is not unbiased, as can be shown by evaluating the ensemble mean \( < s'^2 > \). In fact

\[ < s^2(x) > = \sigma^2(x) - \sigma^2(\bar{x}) \] (46)

which cannot be reduced further without information on the nature of persistence of the series; \( \sigma^2(\bar{x}) \) is given by

\[ < s^2 > = \frac{N'\sigma^2}{N' - 1} \]

where \( N' \) is the number of effectively independent samples.

It also readily follows from (43) that

\[ < \Delta^2(\bar{x}) > = < s^2(\bar{\Delta}) > \] (47)

\[ = 2\sigma^2(\bar{x}') \]

and that

\[ < \rho(\Delta) > = \rho(x') \] (48)

the lag correlation of the parent series.

The results (45), (47), and (48) can be applied to both the control and climatic experiment runs. Because of the statistical independence the variance of the difference between the means of such runs is
\[
\sigma_{c-e}^2 = \sigma_c^2 + \sigma_e^2
\]
\[= \frac{1}{2} \left( \Delta_c^2 + \Delta_e^2 \right) \tag{49}\]

where \(c\) and \(e\) denote control and experiment, respectively.

4.2. Evaluations Based on Differencing the Control and Experiment Runs, Assuming no Change in Cycles or Trends

If we assume we are treating a series of differences \(\Delta_{c-e}\) of control and climatic experiments, with different population means and variances, \(\mu_c, \mu_e, \sigma_c^2\) and \(\sigma_e^2\), respectively, an analysis similar to that used above shows that

\[< s^2(\Delta_{c-e}) > = \sigma_c^2(x'_c) + \sigma_e^2(x'_e) \tag{50}\]

\[< r(\Delta_{c-e}) > = \frac{\rho(x'_c) + \rho(x'_e)}{\sigma_c^2(x'_c) + \sigma_e^2(x'_e)} \tag{51}\]

\[< s^2(\Delta_{c-e}) > = \sigma^2(x'_c) + \sigma^2(x'_e) \tag{52}\]

where the left hand sides of (50) - (52) are to be calculated with the aid of (40) - (42). These equations give unbiased estimates of the combined variances and correlations of the stationary parts of the two series. Since

\[\sigma^2(\Delta_{c-e}) = \sigma_c^2(x'_c) + \sigma_e^2(x'_e) \tag{53}\]

Equation (52) is a direct estimate of the variance of the change in climatic mean.
4.3. Spatial Averaging

The techniques described above for making error estimates are applicable to any differenced variables of the model, including spatially averaged quantities. Since a practical difficulty in climatic experimentation is the length of running time demanded for obtaining accuracy, it may often prove to be easier to establish the existence of a climatic change using spatial averaging in order to reduce the variance. Although this can be done by dealing with the statistics of the averaged variable, it is also possible to consider the effect of averaging the estimates of local changes over wider areas in order to reduce the uncertainty. In general, these two methods lead to different results, since spatial correlations are not taken into account in the second procedure. The second procedure will normally yield less persistence, since the spatial coherence of the moving baroclinic disturbances past a given location is believed to provide the main component of local persistence. In addition, and especially when treating global averages, globally coherent motions of the atmosphere, such as regular vacillations of the planetary wave system, could produce longer time correlations than would be apparent for a local quantity.

The relationship between the two areal averages for a variable $x_i$ is given by

$$\sigma^2(x_i) = \sigma^2(\bar{x}_i) + (x_k - \bar{x}_i)(x_e - \bar{x}_i)$$  \hspace{1cm} (54)

where the bar now refers to spatial averaging. The last term is a measure of the covariance over the averaging region, and is of the order of $k\sigma^2(x_i)$ where $k$ is the ratio of the areal extent of coherence to the total averaging area.
5. SUMMARY OF EXPERIMENTAL PROCEDURES

For ease of reference, we summarize here the steps proposed in planning and evaluating numerical experiments in climatic change.

5.1. Determination of the Required Running Time for a Climatic Experiment

Assuming we have at our disposal a satisfactory model for the normal climate to use as the control, we compute its statistical properties (if these are not already known) as follows.

From the model output for at least the duration of the proposed climatic experiment, plus at least one random perturbation run, we compute the serial correlation at lag \( v \) for the difference between the two outputs over the total length of the experiment \( T^* \) from Eq. (41):

\[
\rho_v(\Delta) = \frac{1}{N^* - v} \sum_{i=1}^{N^* - v} \frac{\Delta_i \Delta_{i+v}}{s^2} \quad (55)
\]

\[
s^2(\Delta) = \frac{1}{N^*} \sum_{i=1}^{N^*} \Delta_i^2 \quad (56)
\]

where \( N^* \) is the total number of samples in the time \( T^* \).

A display of (55) is sometimes useful in suggesting the possible types of statistical noise in the data. In particular, if it is found that

\[
\rho_v = (\rho_1)^v \quad (57)
\]
to within the accuracy of the estimates of \( \rho_v \) (the variance of which is approximately \((1 - \rho_v^2)/(N - v)\)), then, according to (16), red noise may be assumed. The time between effectively independent samples is thus (Eq. 22)

\[
T_o = t_0 \left( \frac{1 + \rho_1}{1 - \rho_1} \right)
\]

(58)

where \( t_0 \) is the sampling interval.

If no such simple result holds, the variance of the mean over time \( T \) (in which there are \( N \) samples) is given by (42) as

\[
s^2(\overline{\Delta}) = \frac{s^2(\Delta)}{N} \sum_{v=-(N-1)}^{N-1} \left( 1 - \frac{|v|}{N} \right) r_v(\Delta)
\]

(59)

\[
= \frac{s^2(\Delta)}{N} \left( 1 + 2 \sum_{v=1}^{N-1} \left( 1 - \frac{v}{N} \right) r_v(\Delta) \right)
\]

(60)

Alternatively, we can calculate \( T_o \) from (10) - (11) as

\[
T_o = T \frac{s^2(\overline{\Delta})}{s^2(\Delta)}
\]

(61)

The number of effectively independent samples \( N' \) in the averaging time \( T \) is then given by (58) or (61) as

\[
N' = T/T_o
\]

(62)

Using the approach described in Section 3, and having decided upon the confidence level we wish to attach to the detection of a specific
magnitude of climatic change, we may use Fig. 1 to determine $N'$, the number of independent samples required. If we find that $T = N'T_o < T^*$, only one random perturbation need be used. If $T > T^*$, we need to select a sufficient number ($n$) of random perturbations for the climatic experiment such that

$$T \leq nT^*$$

5.2. **Evaluation of Climatic Experiment Results**

As discussed earlier, there are different situations to be considered, depending on the quality of information available for the "normal" climate simulation.

5.2.1. **Control well known**. In this case the statistics of the climate experiment are determined from a series of random perturbation runs, using the equations given in Section 5.1 as applied to differenced variables, and the variance of the climatic mean is evaluated. The estimate for the variance of the random part of the variable is then given by (47) as

$$\sigma^2_{e}(\bar{x}') = \frac{1}{2} \Delta^2_{e,n}(\bar{x})$$

(63)

where $\Delta^2_{e,n}(\bar{x})$ is the estimated variance for $n$ perturbation runs, so that

$$\Delta^2_{e,n}(\bar{x}) = \frac{1}{n} \sum_{i=1}^{n} \Delta^2_{e,i}(\bar{x})$$

(64)

The variance of the difference between the averages of the control and experiment is estimated by (49) as
\[ \sigma_{c-e}(\bar{x}) = \sigma_c^2(\bar{x}') + \sigma_e^2(\bar{x}') \] (65)

Note that in (65) we have put \( \sigma_{c-e}^2(\bar{x}) = \sigma_{c-e}^2(\bar{x}') \), i.e., we have assumed no trends or regular cycles to be induced in the climatic experiment that were not in the control.

5.2.2. **Differencing between control and climatic experiment.** If knowledge of the variance of the control is poor, additional runs will be required. In this case we can carry through the computational analysis procedure for control and experiment separately, as described in 5.2.1 for the climatic experiment run. The amount of computation can be reduced by performing the variance calculations on the difference between the control and experiment outputs. This will give valid estimates for the variance of the difference between the means, provided no regular cycles or trends are induced in the climate experiment that do not exist in the control. In this case the variance \( \Delta_{c-e}^2(\bar{x}) \) will be a direct measure of the error in the observed change in the means (c.f. Section 4.2).

5.3. **Possible Difficulties**

A number of assumptions on the nature of the climatic time series have been made in these descriptions for handling simulated climatic data. Here we briefly indicate how these can affect the results.

5.3.1. **Persistence time \( T_o \) greater than the running time \( T^* \).**

This possibility should be revealed by significant correlations (greater than the probable error) at large lag times. In this case, an estimate of the persistence time \( T_o \) will not be possible, but Eq. (60) for the variance of the mean still holds and can be used to calculate \( N' \) for the total running time \( T^* \) of a single run.
5.3.2. **Changed cycles or trends in the climatic experiment.** As noted in Section 5.2.2, if these occur the control and climatic experiment have to be treated separately, unless the regular features are first extracted from the time series. Indicators of this situation are the presence of high correlations at regular intervals and an apparent non-zero asymptote for infinite lag correlations.

5.3.3. **Non-independence of random perturbation runs.** This is a more difficult problem to deal with, since its presence may not be evident from study of the serial correlations. For instance, if the phase and amplitude of a random wave is maintained in both the original and the supposedly randomized perturbation, this random component will not be seen in the differences between the two runs. The serial correlations and the variance estimates, made according to the prescription given in 5.1, will then be in error. The only way to determine with certainty that the randomization does indeed produce a statistically independent sequence is to study the *absolute* values of the output, or to undertake a cross-correlation analysis of the two series. We might note that the impact of this effect can be diminished by choosing as an initial state for the random run conditions obtained by randomizing at some other (late) running time of the model (but allowing for the seasonal and diurnal effects).

5.3.4. **Variance estimates too large.** We have assumed that the number of model runs could be adjusted to give a variance of the mean small enough to clearly delineate possible climatic changes. If the output shows larger variance than expected, and hence a large uncertainty

---

* Presumably because the amplitude of the random perturbation was too small.
as to the significance of the change in the mean (as might result, for example, if the climatic experiment had larger persistence than expected), we can either perform further test runs, or adopt a more comprehensive statistical significance testing method. A complete description of procedures for the latter go beyond the intent of this report, though we have given the appropriate test statistic for the significance of the difference between two means with different variances in Eq. (33); this can be applied to the results derived from the control and climatic experiments in a relatively simple and straightforward manner. However, it does require a sufficiently large sample size to be valid.
ACKNOWLEDGMENTS

We would like to thank the members of the Rand Climate Program for their interactions during the senior author's stay at Rand, and to thank Carolyn Huber and Walter Matyskiela for their assistance in programming the regression analysis methods described here.
REFERENCES


