Max-Planck-Institut für Meteorologie

REPORT No. 18

$$\mathrm{SSAB} = \sum_{i=1}^{I} m_i \cdot (\bar{y}_{i..} - \bar{y}_{...})^2 + \sum_{i=1}^{I} \sum_{j=1}^{m_i} (\bar{y}_{ij.} - \bar{y}_{i..})^2$$

$$= \sum_{i=1}^{I} m_i \cdot \hat{\alpha}_i^2 + \sum_{i=1}^{I} \sum_{\ell=1}^{k_i} \hat{a}_{i\ell}^2 + \sum_{i=1}^{I} \sum_{j=1}^{m_i} \left\{ \bar{y}_{ij.} - \sum_{\ell=1}^{k_i} \hat{a}_{i\ell} \cdot p_{i\ell}(j) - \bar{y}_{i..} \right\}^2$$

ASPECTS OF THE STATISTICAL ANALYSIS OF CLIMATE EXPERIMENTS WITH MULTIPLE INTEGRATIONS

by

FRANCIS W. ZWIERS

HAMBURG, JUNE 1988

AUTHOR:

FRANCIS W. ZWIERS		CANADIAN CLIMATE CENTRE
DOWNSVIEW, ONTARIO, CANADA

MAX-PLANCK-INSTITUT
FÜR METEOROLOGIE
BUNDESSTRASSE 55
D-2000 HAMBURG 13
F.R. GERMANY

Telefon nat.: (040) 41 14 - 1
Telefon int.: +49 40 41 14 - 1
Telex-Nr.: 211092
Telemail: MPI.METEOROLOGY
Telefax nat.: (040) 41 14 - 298
Telefax int.: +49 40 41 14 - 298

Aspects of the Statistical Analysis of Climate Experiments with Multiple Integrations

Francis W. Zwiers
Canadian Climate Centre
Downsview, Ontario
Canada

Abstract

Many modern climate experiments consist of multiple General Circulation Model (GCM) integrations. Many experiments also contain climate integrations in which a climate equilibrium is not obtained for a considerable period of time, or in which the mean state slowly evolves with time in response to slowly changing external conditions (such as the concentration of CO_2. In this paper we describe some relatively simple statistical techniques which test for the presence of trend and take its effects into account when studying simulated inter-annual variability. The limitations of these methods, the assumptions which are implicit in their use, and techniques for developing tests of hypothesis are all discussed.

1. Introduction

The purpose of this paper is to describe a simple statistical framework within which it is possible to analyze several related climate simulations on inter-annual time scales. There is a need for such a formal structure because climate experiments which consist of several related General Circulation Model (GCM) integrations are becoming relatively common. Many coupled atmosphere/ocean GCM experiments have already been reported in the literature (Manabe and Bryan, 1969; Manabe, et al., 1975, 1979; Washington, et al., 1980; Gates, et al., 1985) and many are planned or in progress. Also, many El Nino Southern Oscillation (ENSO) and enhanced CO_2 experiments have already been conducted or are in the planning or execution phase. The design of such experiments invariably involves one or more control runs with uncoupled atmospheric GCM's as well as one or more experimental runs involving a coupled ocean (and perhaps cryosphere) and/or enhanced concentrations of CO_2.

One feature of the proposed statistical framework is that it takes the possibility of slowly evolving trends in the analyzed climates into account. Such trends are apparent in some coupled atmosphere/ocean climate simulations and are to be expected in CO_2 experiments in which the atmospheric composition is prescribed as a function of time. In the case of coupled atmosphere/ocean climate simulations, trends sometimes occur due to the long time require to obtain an equilibrium of the simulated climate system. For example, many coupled ocean/atmosphere climate simulations are conducted by bringing the component parts of the coupled system into some sort of equilibrium with a set of prescribed boundary conditions which are presumed to be representative of the mean boundary conditions when the systems are allowed to interact. When the individual equilibria have been obtained the coupling mechanism is turned on and the entire system is allowed to come into its own equilibrium. Depending upon the details of the particular system, this may take a considerable period of time. At some point along the path to equilibrium the components of the simulated system begin to interact in a realistic way and to simulate variation about the (still transient) mean state in a realistic way. At that point it becomes possible, and may be desirable, to analyze the variation simulated about the still transient mean.

We will describe statistical techniques for the analysis of simulated climate variability which are not invalidated by the presence of a gradual drift through a sequence of transient mean states. It is assumed that variation about this trend is representative of that which eventually occurs about the new equilibrium to which the simulated climate system is converging. This drift, or trend, will be represented in a polynomial manner for the sake of convenience. Such a representation has the advantages that it is relatively flexible in its form and that it is linear in the unknown trend coefficients. The latter is an important logistical consideration as it greatly simplifies the problem of coefficient estimation. While the representation of trend by a decaying exponential, a damped sinusoid or some other function which is non-linear in its unknown coefficients may be more elegant and aesthetically pleasing, coefficient estimation is not straight forward in such cases. This is particularly so in climate experiments because non-linear minimization procedures (which are necessary for coefficient estimation when trend models are non-linear in their coefficients) must be applied repeatedly for every field considered and at every grid point in the field.

In the discussion that follows we will have the following basic experimental design in mind. There will be a <u>control</u> climate simulation with an atmospheric GCM or perhaps a reference coupled atmosphere/ocean GCM. In addition, there will be one or more experimental climate simulations with perturbed climate models (such as a couple atmosphere/ocean model or a model in which CO_2 concentrations are increased - perhaps gradually). The purpose of the experiment is to study the effects of the perturbations on the simulated mean state and on the inter-annual variability of the simulated climate system. Note that the techniques which will be prescribed are not particular to simulated climate data. In fact, one may often want to include observed climate data as one of the control or experimental climate realizations.

In addition to simple questions regarding changes in means and inter-annual variance, there are also some additional questions which are of interest in the simulation of climate systems which can be addressed within the statistical framework described below. First, do the simulated climates display trends to new equilibria, perhaps as a result of coupling (with an oceanic GCM) or gradual presribed changes in atmospheric composition. The answer to this question fundamentally affects the analysis techniques which are used to answer subsequent questions. The form of the fitted trend model provides information about whether a simulated system has come to an equilibrium or the rate at which this is occurring.

Secondly, do the climate simulations which comprise the climate experiment display "potentially predictable" is the sense of Madden (1976) or Trenberth (1985). That is, within a particular climate simulation, is there more non-systematic variation about the trend on time scales of seasons or years than that which is a consequence of simple day-to-day variation alone.

Third, if such transient "signals" are simulated in several of the climate simulations, are they more or less frequent and/or more or less intense in one system than in another? Is the presence of such signals associated with enhanced inter-annual variability? The answers to these questions provide information about the relative importance of the atmosphere's internal dynamics and external forcing in the predictability of climate and the effects of boundary interactions on climate predictability. Recent work by Zwiers (1987) with the Canadian Climate Center (CCC) Atmospheric General Circulation Model (AGCM) suggests that the atmosphere's internal dynamics alone may generate potentially predictable signals on seasonal and inter-annual time scales.

Statistical tests which attempt to answer some of these questions for a collection of related climate simulations will be described in the sections that follow. As note above, careful attention will be paid to the statistical assumptions which are implicit in these statistical procedures (and which are commonly overlooked). These assumptions (regarding, for example, the independence of climate realizations) comprise the basic framework within which significance is assessed. The judgment of whether an outcome of an experiment is significant is made relative to a set of rigid and artificial statistical standards which are **derived** from the assumptions (i.e., the complete statistical model) implicit in the procedure used to assess significance. The caveat that the assumed statistical model does not match reality must always be borne in mind.

A second point which is important for the interpretation of results is that all the procedures described here are *univariate* statistical procedures which must be applied to individual climate variables realized at a single grid point or as a single expansion function coefficient. Difficulties with interpretation of a field of test results relating to its correlation structure as discussed by Livezey and Chen (1983), Storch (1982) and others must therefore be taken into account.

A final point to make is that the techniques described are not being proposed in a vacuum. They evolve from experience gained analysing the results of a recent Tropical-Ocean Global-Atmosphere (TOGA) experiment conducted at the Max Planck Institute for Meteorology in Hamburg. Preliminary results of the experiment are described by Latif and Biercamp (1987) and Zwiers and Storch (1987). A full report of the experiment is in preparation.

The techniques which are proposed below are extensions of standard analysis of variance (ANOVA) techniques (see for example, Sachs, 1982). ANOVA techniques applied in the time domain have been employed in the analysis of observed and simulated climate variability for a considerable period of time now. The idea that a time domain ANOVA might be a useful way in which to study the "potential predictability" of climate systems was first suggested by Jones (1975, 1976) and subsequently applied by Madden (1976), Madden and Shea (1978), Shukla and Gutzler (1983), Trenberth (1985), Zwiers(1987) and others. Madden (1983), Shukla (1983), Trenberth (1984a,b), and Zwiers (1987) discuss the statistical methodology used in such studies in detail. The methodology will be reviewed briefly in Section 2. Extensions of the methodology to systems displaying a climate trend and to groups of climate simulations are described in Section 3. The paper is concluded with a short discussion in Section 4.

2. Potential Predictability Studies

In a "potential predictability" study, variation observed or simulated in a particular season is partitioned into two components: inter-annual variability and intra-seasonal variability. The purpose of a "potential predictability" study is to determine whether inter-annual variance, either simulated or observed, is greater than would be induced by simple "day-to-day" fluctuations in the observed field. The excess variability, if it exists, is potentially predictable in the sense that it is not the consequence of short time scale "noise" and therefore may be predictable.

The methods which are used in potential predictability studies appear to quite different externally; studies have been carried out both in the "frequency domain" (Zwiers, 1987; Madden, 1976) and the "time domain" (Trenberth, 1985). They are in fact very similar and consist of a few basic steps. In all studies it is recognized that the observed time series is not stationary in the mean, even within seasons, and thus as a first step the annual cycle of the mean is estimated and subtracted from the observations. Some workers have also estimated the annual cycle of the standard deviation and scaled the observed time series with this estimate (Madden, 1976; Zwiers, 1987). The resulting time series are presumed to be approximately stationary, at least within seasons.

The second step is to compute seasonal (or sometimes monthly) means and corresponding inter-annual variances. The third step is to infer these same inter-annual variances from the observed intra-seasonal variation in the transformed time series. This is done by noting that the variance of a time average is given by

$$(1) \qquad \sigma^2(\bar{x}) = \int_{-\pi}^{\pi} \frac{\sin^2(n\lambda/2)}{n \sin^2(\lambda/2)} \cdot f_{xx}(\lambda) d\lambda$$

where λ is frequency in radians per observing interval, $f_{xx}(\lambda)$ is the spectral density function of the transformed time series and n is the length of a season expressed in observing intervals. Equivalently

$$(2) \qquad \sigma^2(\bar{x}) = \frac{1}{n} \left\{ C_{xx}(0) + 2 \sum_{\tau=1}^{n} (1-\tau/n)\ C_{xx}(\tau) \right\}$$

where $C_{xx}(\tau)$ represents the auto-covariance function of the transformed time series.

Calculations are done in the "frequency domain" if the variance of the seasonal mean is estimated using expression (1). In this case the spectral density function $f_{xx}(\lambda)$ is estimated from fluctations observed within a season and substituted into (1) to estimate the variance of the seasonal mean. This is repeated for each, say December, January, February (DJF), season in the observed time series, and the resulting collection of estimates is averaged to form a pooled estimate of the variance of the time average derived only from intra-seasonal variation.

The periodogram, $I_{xx}(\lambda)$, which is given by

$$(3) \qquad I_{xx}(\lambda) = \left| \sum_{t=1}^{n} \exp(\ i\lambda t) \cdot X_t \right| / 2\pi n$$

is frequently used to estimate the spectral density function within a particular season. The periodogram does not give a useful estimate of $f_{xx}(0)$ because the periodogram is proportional to the square of the mean of the transformed time series at zero frequency. Hence $f_{xx}(0)$ is estimated by extrapolating the periodogram ordinate at frequency $1/2\pi n$ to the origin. Madden (1976) calls this the *Low Frequency White Noise* extrapolation of the periodogram and justifies its use in the context of intra-seasonal variation which behaves as a red noise process. Such an extrapolation is in fact reasonable regardless of the stochastic character of intra-seasonal variation, as long as it behaves as a stationary, ergodic stochastic process, because all such processes have a spectral density function with zero slope at the origin. That is, at low enough frequency (long enough time scale) all stationary ergodic stochastic processes behave as white noise processes.

When calculations are done in the "time domain" the variance of the seasonal mean is estimated using expression (2). In this case the auto-covariance function $C_{xx}(\tau)$ is estimated from intra-seasonal variation and the estimate is substituted into (2) to estimate the variance of the time average. Trenberth (1984a,b) discusses various aspects of auto-covariance function estimation which are pertinent to potential predictability studies.

With either approach the fourth and last step is to form the ratio of the observed inter-annual variance with the estimate of inter-annual variance derived from intra-seasonal fluctuations alone, and to compare this ratio with critical values derived from an appropriate F-distribution. In the case of calculations performed in the frequency domain, appropriate critical values can be determined easily because a relatively informative asymptotic statistical theory is available (see Zwiers, 1987). On the other hand, determining critical values for the F-ratio computed in the time-domain calculation is more difficult because the relevant asymptotic statistical theory is less informative. The same sorts of difficulties described by Thiébaux and Zwiers (1984) in connection with difference of means tests also occur here.

Although not explicitly stated in most potential predictability studies, a simple statistical model is implicit in the sequence of calculations described above. This model is given by

$$(4) \quad y_{jt} = \mu + \delta_j + \varepsilon_{jt}, \; j=1, \ldots, m \text{ (years)}, \; t=1, \ldots, n \text{ (days)}$$

where index j runs over years and index t runs over days within a season. The coefficients δ_j are assumed to be independent Gaussian random variables with mean 0 and variance $\sigma^2(C)$. They represent the possibly predictable transient **climate signal**. The time series $\{\varepsilon_{jt}, t=1, \ldots, n\}$ are assumed to be independent realizations of a stationary stochastic process and represent intra-seasonal variation.

If we take time averages in (4) we see that

$$(5) \quad \bar{y}_{j.} = \mu + \delta_j + \bar{\varepsilon}_{j.} \;, \; j=1, \ldots, m.$$

Here and elsewhere we use the notation commonly employed in the statistical literature to represent means: an overbar indicates an average of some sort and the subscript(s) over which the average is computed is(are) indicated by a dot(s). From (5) we see that the variance of the seasonal mean is given by

$$(6) \quad \sigma^2(\bar{y}_{j.}) = \sigma^2(C) + \sigma^2(\bar{\varepsilon}_{j.})$$

This quantity is estimated directly by the inter-annual variance of the seasonal mean. The third step of the four step procedure described above results in a statistically independent estimate of the climate noise $\sigma^2(\bar{\varepsilon}_{j.})$. Thus the ratio of the two variance estimates will be larger than one if the variance $\sigma^2(C)$ due to the climate signal is different from zero. The ratio will be equal to one if there is no climate signal. The F-test described above looks for evidence that the ratio is greater than one.

To unravel the assumptions which are implicit in the F-test for potential predictabilty, it is useful to express model (4) somewhat differently. Consider time series

$$(7) \qquad y_i = \mu + \delta_i + \varepsilon_i$$

where i simply indexes days and $\{\delta_i\}$ and $\{\varepsilon_i\}$ represent statistically independent stationary stochastic processes. We could think of these two processes as independent red noise processes with different "decorrelation" times, amounts of persistence, or predictability. Equivalently, we could think of the spectral density function of $\{y_i\}$ as the sum of two spectral density functions, the first having greater curvature near the origin and dropping to zero faster than the second. Now suppose that process $\{\delta_i\}$ has a spectral density function which is essentially zero at frequencies greater than one cycle per season and is essentially flat at frequencies less than one cycle per year. Then (7) reduces to (4) if (7) is sampled daily for an n-day season once per year for m years. We see from this that there are several assumptions implicit in the F-test. These include:

(i) the climate signal and climate noise are stochastically independent. In particular, the stochastic characteristics of climate noise are unaffected by the state of the climate signal, something which is unlikely to be true, and there are no noise-to-signal feedbacks.

(ii) the climate signal contains no possibly predictable components at frequencies greater than one cycle per season. This automatically relegates phenomena such as the 30-60 day wave to climate noise, and dictates a fair degree of smoothness for the climate signal.

(iii) the climate signal does not persist to any great extent from one year to the next.

These observations regarding the limitations of potential predictability studies are not new: they have been made in various forms by Zwiers (1987), Trenberth (1985), Shukla (1983) and others. But stating the limitations in connection with specific statistical models such as (4) or (7) clarifies them considerably. Unfortunately the limitations noted above cannot easily be overcome. It would be necessary to construct statistical models which anticipate the form of the climate signal and the manner in which the presumably nonlinear interaction between signal and noise occurs. However, even with these limitations, simple models such as (4) and its extensions described below are very useful for analysing variability simulated by climate models.

3. Variability in Related Climate Simulations

Implicit in the techniques and models described above is the assumption that the simulated climate is stationary. This includes the assumption that that there is no trend or gradual transition to a new equilibrium in the simulated climate. With many systems this assumption is not satisfied and statistical techniques must be employed which take trend into account. One can envision at least two kinds of experiments in which this might be a problem: a single long integration of a coupled atmosphere/ocean climate model, and a pair of long integrations; one being a control climate simulation with an AGCM and the second being an experimental climate simulation with a coupled or otherwise perturbed GCM.

3.1 A single simulation with evolving mean state

We begin by considering a single long integration of a climate system. A suitable modification of (4) which takes the possibility of trend into account might have the form

$$(8) \qquad y_{jt} = \mu(j) + \delta_j + \varepsilon_{jt}, \quad j=1, \ldots ,m \text{ (years)}, \quad t=1, \ldots ,n \text{ (days)}$$

where the long term mean $\mu(j)$ is now a slowly varying deterministic function of time. The word "deterministic" is used in the sense that if statistically independent realizations could be obtained of the same climate simulation (perhaps by repeating the run from randomly perturbed initial conditions) the result would be a single ensemble of climate simulation realizations with time dependent ensemble mean $\mu(j)$. As noted in section 1, estimation and inference problems remain fairly simple as long as $\mu(j)$ has a form which is linear in its parameters. For most purposes a polynominal should be adequate. Thus we have

$$(9) \qquad y_{jt} = \mu + a_1p_1(j) + \cdots + a_kp_k(j) + \delta_j + \varepsilon_{jt}, \quad j=1, \ldots ,m \text{ (years)}, \quad t=1, \ldots ,n \text{ (days)}.$$

The trend is represented by the k'th degree polynomial $a_1p_1(j) + \cdots + a_kp_k(j)$ where p_1, p_2, ..., p_k are orthonormal polynomials of degree 1, 2, ..., k respectively defined on integers j=1, ...,m; μ represents the m-year seasonal mean; coefficients δ_j represent the transient climate signal and are assumed to be independent Gaussian random variables with mean 0 and variance $\sigma^2(C)$; and time series $\{\varepsilon_{jt}, t=1, \ldots,n\}$ are independent realizations of a stationary stochastic process representing intra-seasonal fluctuations. In addition to the three assumptions stated above, there are implicit in this model assumptions that the characteristics of the climate signal and noise are independent of the trend and that there is no interaction between the trend and other components of the climate. Again, our intuition tells us that this cannot be entirely the case. As suggested in Section 1, a certain amount of good judgment is required to avoid analyzing data obtained too soon after climate simulation is initiated.

3.1.1 A Partition of Observed Variation

Development of tests of hypotheses concerning the presence of trend and a climate signal are straight forward once a linear statistical model such as (9) has been adopted. We begin by partitioning total variability into inter-annual and intra-seasonal components. We have

$$(10) \quad \sum_{j=1}^{m} \sum_{t=1}^{n} (y_{jt} - \bar{y}_{..})^2 = n \cdot \left\{ \sum_{j=1}^{m} (\bar{y}_{j.} - \bar{y}_{..})^2 \right\} + \left\{ \sum_{j=1}^{m} \sum_{t=1}^{n} (y_{jt} - \bar{y}_{j.})^2 \right\}$$

$$= n \cdot SSA + SSE.$$

Total inter-annual variability may be further partitioned as

$$(11) \quad SSA = \sum_{\ell=1}^{k} \hat{a}_{\ell}^2 + SSR = SSD + SSR$$

where

$$(12) \quad SSD = \sum_{\ell=1}^{k} \hat{a}_{\ell}^2 , \qquad \hat{a}_{\ell} = \sum_{j=1}^{m} \bar{y}_{j.} \cdot p_{\ell}(j), \text{ and } \qquad SSR = SSA - SSD.$$

Each squared coefficient in (11) has one degree of freedom (df), and the residual sum of squares SSR has m-k-1 df. With the assumptions above all k+1 components of SSA are mutually statistically independent. Taking expectations using model (9) we see that

$$\text{(13)} \qquad \begin{aligned} E(\hat{a}_\ell^2) &= a_\ell^2 + \sigma^2(C) + \sigma^2(\bar{\varepsilon}_{j.}) \\ E(SSD) &= \sum_{\ell=1}^{k} a_\ell^2 + k\cdot\left[\sigma^2(C) + \sigma^2(\bar{\varepsilon}_{j.})\right] \\ E(SSR) &= (m-k-1)\cdot\left[\sigma^2(C) + \sigma^2(\bar{\varepsilon}_{j.})\right] \end{aligned}$$

3.1.2 A Test for Trend

Most of the information necessary to construct tests of hypotheses concerning trend and potential predictability is contained in (13). For example, to test whether or not there is a significant trend in the coupled simulation the suitable null and alternate hypotheses are

$$\text{(14)} \qquad H_o\colon a_1 = a_2 = \ldots = a_k = 0 \quad \text{vs.} \quad H_a\colon \text{at least one } a_\ell \neq 0$$

and an appropriate test statistic is

$$\text{(15)} \qquad F = \left\{ SSD/k \right\} / \left\{ SSR/(m-k-1) \right\}.$$

The numerator of the test statistic is chosen by noting that SSD is the only term in the partition of SSA whose expectation involves the parameters being tested. The expectation of the numerator is $\left\{ \sum_{\ell=1}^{k} a_\ell^2 / k \right\} + \sigma^2(C) + \sigma^2(\bar{\varepsilon}_{j.})$.

The denominator of the test statistic is chosen by looking for a mean sum of squares which is statistically independent of the numerator and estimates that portion of the expectation of the numerator which is not affected by the parameters under test regardless of whether or not the null hypothesis is true. In this case the denominator must be chosen to have expectation $\sigma^2(C)$ + $\sigma^2(\bar{\varepsilon}_{j.})$. When the null hypothesis is true, both components of the ratio estimate the same value. When it is false, the numerator estimates a quantity greater than that of the denominator and thus the null hypothesis is rejected when F is unusually large. The assumptions which have been made ensure that F has Fisher's F-distribution when H_o is true, and thus H_o is tested by comparing F with the upper quantiles of the F-distribution with k and m-k-1 degrees of freedom. Similarly, the significance of the ℓth degree component of the trend can be determined by comparing $\hat{a}_\ell^2/\{SSR/(m-k-1)\}$ with upper tail critical values from the F-distribution with 1 and m-k-1 df.

3.1.3 A Test for Potential Predictability

To test for potential predictability *taking trend into account* the appropriate hypotheses are

$$(16) \quad H_o: \sigma^2(C) = 0 \quad \text{vs.} \quad H_a: \sigma^2(C) > 0$$

and the appropriate test statistic is

$$(17) \quad F = \left\{ SSR/(m-k-1) \right\} / \hat{\sigma}^2(\bar{\varepsilon}_{j.})$$

where $\hat{\sigma}^2(\bar{\varepsilon}_{j.})$ is a frequency- or time-domain estimate of $\sigma^2(\bar{\varepsilon}_{j.})$, the variance due to climate noise. As above, both components of this ratio estimate the same quantity when H_o is true, and the numerator estimates a quantity greater than that estimated by the denominator when H_o is false. The assumptions and the appropriate asymptotic theory can be used to show that the distribution of F is approximately Fisher's F-distribution when H_o is true. Thus H_o is rejected when F is greater than the appropriate upper tail critical values determined from the approximating F-distribution.

3.1.4 Tests of the Mean and Variance

It is also a straight foreward matter to test that the simulated climate has an **apriori** specified inter-annual variance, say σ_I^2. The appropriate null and alternate hypotheses are

$$(18) \quad H_o: \sigma_I^2 = \sigma^2(C) + \sigma^2(\bar{\varepsilon}_{j.}) \quad \text{vs.} \quad H_a: \sigma_I^2 \neq \sigma^2(C) + \sigma^2(\bar{\varepsilon}_{j.})$$

and the appropriate test statistic is

$$(19) \quad \chi^2 = \text{SSR} / \sigma_I^2$$

When the null hypothesis is true this statistic has a Chi-squared distribution with m-k-1 df and thus the test is conducted by comparing the value of the test statistic with upper and lower critical values of this distribution.

It is a somewhat less straight foreward matter to test hypotheses concerning the long term mean because (i) statistical model (9) takes into account the possibility that the mean can change as a function of time; and (ii) it has done this by assuming, for the sake of mathematical and computational convenience, that the mean changes as a polynomial in time. This effectively precludes us from asking questions regarding the equilibrium, or asymptotic value of the mean unless we are willing to assume that an equilibrium has been obtained during the integration. We will test

(20) $H_o : \mu_{equil} = \mu_o$ vs. $H_o : \mu_{equil} \neq \mu_o$

where μ_{equil} represents the equilibrium mean by assuming that

$$\mu + \sum_{j=s}^{m} \left[a_1 p_1(j) + \cdots + a_k p_k(j) \right] / (m-s+1) \cong \mu_{equil}$$

for some s, $1 \le s \le m$. With this assumption an unbiased estimator of μ_{equil} is given by

(21) $$\hat{\mu}_{equil} = \hat{\mu} + \sum_{j=s}^{m} \left[\hat{a}_1 p_1(j) + \cdots + \hat{a}_k p_k(j) \right] / (m-s+1)$$

It can be shown that the variance of this estimator is given by

(22) $$\left\{ \frac{1}{m} + \sum_{\ell=1}^{k} \left[\sum_{j=s}^{m} p_\ell(j) \,/\, (m-s+1) \right]^2 \right\} \cdot \left\{ \sigma^2(C) + \sigma^2(\bar{\varepsilon}_{j.}) \right\}.$$

It follows that the appropriate statistic for testing (20) is

(23) $$T = (\hat{\mu}_{equil} - \mu_o) / \left[\left\{ \frac{1}{m} + \sum_{\ell=1}^{k} \left[\sum_{j=s}^{m} p_\ell(j) \,/\, (m-s+1) \right]^2 \right\} \cdot \frac{SSR}{(m-k-1)} \right]^{1/2}$$

which is distributed as a Student's t random variable with (m-k-1) df when the null hypothesis is true. The test is conducted by comparing (23) with the extreme quantiles of the appropriate t-distribution.

3.2 Paired simulations

The analysis of the previous section illustrates the manner in which trend might be taken into account in climate simulations, and the manner in which statistical tests are developed. The analysis was conducted in the context of a single simulation. However, one suspects that in most cases there will be related control and experimental simulations which may be profitably analyzed simultaneously. The simple one-way analysis of variance model described above can be expanded into a two-way model which employs information from several integrations. This expanded model provides a framework for more powerful tests of potential predictability in individual climates than could be obtained if each climate was regarded separately. It also provides the framework within which it is possible to develop several other tests including comparisons of means, inter-annual variance and climate noise.

An appropriate 2-way model might have the form

$$(24) \qquad y_{ijt} = \mu + \alpha_i + \sum_{\ell=1}^{k_i} a_{i\ell} \cdot p_{i\ell}(j) + \delta_{ij} + \varepsilon_{ijt}\,, \quad i=1, \ldots, I \ \text{(simulations)}$$

$$j=1, \ldots, m_i \ \text{(years)}$$

$$t=1, \ldots, n \ \text{(days)}$$

where i designates a climate simulation, j runs over m_i years within a simulation, and t runs over n days within a season. In this model coefficients α_i represent a deterministic simulation effect after trends in the individual climates have been taken into account. When there is no trend the α_i's represent differences between the overall mean μ and the individual climate means. The term $\sum_{\ell=1}^{k_i} a_{i\ell} \cdot p_{i\ell}(j)$ is a k_ith degree polynomial representation of trend in simulation i where $p_{i\ell}(j)$ are ℓth degree orthonormal polynomials defined on integers $j=1, \ldots, m_i$. A trend in the control simulation might, for example, reflect the gradual loss of mass which is experienced in some climate models or gradual changes in evaporation and radiation which occur over land areas in models with interactive soil hydrology. Terms δ_{ij} represent the transient climate signal and are assumed to be independent Gaussian random variables with mean 0 and variance $\sigma^2(C_i)$.

Time series $\{\varepsilon_{ijt}, t=1, \ldots, n\}$ represent intra-seasonal variations in the simulations. It is assumed that they are independent realizations of the same stationary stochastic process for all i and j. In other words, it is assumed that all climates behave in the same way at short time scales. This is not an unreasonable assumption unless, for example, a significant downward trend in simulated tropical sea surface temperature results in a more closely constrained and less energetic atmospheric simulation. In any case, this assumption can be checked by fitting model (9) separately for each simulation and then comparing estimates of the variance due to climate noise. This assumption provides the only connection in (24) between the simulations. The advantage of making this assumption is that subsequent tests for potential predictability will be more powerful than those available when (9) is applied to each simulation separately. This is because the estimate of variance due to climate noise will be based on information derived from all simulations.

3.2.1 Partition of Variation

With (24) it is possible to answer questions about equality of control and experimental means, trend, potential predictability, and whether it is present to same degree in all climates. As with (9), the development of statistcal tests begins with a partitioning of the total variation in the simulations. First the total variation is partitioned as follows into two terms: (i) inter-simulation plus inter-annual variation; and (ii) intra-seasonal variation.

$$(25)\quad \sum_{i=1}^{I}\sum_{j=1}^{m_i}\sum_{t=1}^{n}(y_{ijt}-\bar{y}_{...})^2$$

$$= n\cdot\left\{\sum_{i=1}^{I}\sum_{j=1}^{m_i}(\bar{y}_{ij.}-\bar{y}_{...})^2\right\}+\left\{\sum_{i=1}^{I}\sum_{j=1}^{m_i}\sum_{t=1}^{n}(y_{ijt}-\bar{y}_{ij.})^2\right\}$$

$$= n\cdot SSAB + SSE.$$

With the assumptions which were made above it can be shown that the two components of the partition are statistically independent sums of squares. Because all intra-seasonal variation is contained within the second term, independence of terms also implies that the estimate of the variance due to climate noise (a component of the second term) is statistically independent of any variation partitioned out of the first sum of squares. The first term is further partitioned as

$$(26)\quad SSAB = \sum_{i=1}^{I} m_i\cdot(\bar{y}_{i..}-\bar{y}_{...})^2 + \sum_{i=1}^{I}\sum_{j=1}^{m_i}(\bar{y}_{ij.}-\bar{y}_{i..})^2$$

$$= \sum_{i=1}^{I} m_i\cdot\hat{\alpha}_i^2 + \sum_{i=1}^{I}\sum_{\ell=1}^{k_i}\hat{a}_{i\ell}^2 + \sum_{i=1}^{I}\sum_{j=1}^{m_i}\left\{\bar{y}_{ij.} - \sum_{\ell=1}^{k_i}\hat{a}_{i\ell}\cdot p_{i\ell}(j) - \bar{y}_{i..}\right\}^2$$

which we will write as

$$(27)\qquad SSAB = SSM + \sum_{i=1}^{I} SSD_i + \sum_{i=1}^{I} SSR_i$$

where $SSM = \sum_{i=1}^{I} m_i \cdot \hat{\alpha}_i^2$, $SSD_i = \sum_{\ell=1}^{k_i} \hat{a}_{i\ell}^2$ and

$$SSR_i = \sum_{j=1}^{m_i} \left\{ \bar{y}_{ij.} - \sum_{\ell=1}^{k_i} \hat{a}_{i\ell} \cdot p_{i\ell}(j) - \bar{y}_{i..} \right\}^2$$

In these expressions

$$(28)\qquad \hat{\alpha}_i = \bar{y}_{i..} - \bar{y}_{...} \quad \text{and} \quad \hat{a}_{i\ell} = \sum_{j=1}^{m_i} \bar{y}_{ij.} \cdot p_{i\ell}(j).$$

The first term in (27), SSM, has I-1 df and reflects differences in climate means after trend has been taken into account. The next group of terms, SSD_i, have k_i df each and reflect variation due to trend. The last group of terms, SSR_i, have m_i-k_i-1 df each and are referred to as residual sums of squares. It can be shown that all components of this partition are statistically independent. Taking expectations using model (24) we find that

$$E(\hat{a}_{i\ell}^2) = a_{i\ell}^2 + \sigma^2(C_i) + \sigma^2(\bar{\varepsilon}_{ij.})$$

$$(29)\qquad E(SSD_i) = \sum_{\ell=1}^{k_i} a_{i\ell}^2 + k_i \cdot \left[\sigma^2(C_i) + \sigma^2(\bar{\varepsilon}_{ij.}) \right]$$

$$E(SSR_i) = (m_i - k_i - 1) \cdot \left[\sigma^2(C_i) + \sigma^2(\bar{\varepsilon}_{ij.}) \right]$$

3.2.2 Tests Concerning Trend

The expressions above provide most of the information necessary to construct tests of various hypotheses. In this subsection we will describe tests which address questions regarding the evolution of the mean states of the simulated climates. In all cases tests are constructed in the same manner as is illustrated in Section 3.1.

To test for a <u>significant trend in climate i</u> the appropriate hypotheses are

(30) $H_o\colon a_{i\ell}\ ,\ \ell=1,\ldots,k_i$ vs. $H_a\colon a_{i\ell} \neq 0$ for at least one ℓ.

The appropriate test statistic is

(31) $$F = \left\{ SSD_i/k_i \right\} / \left\{ SSR_i/(m_i-k_i-1) \right\}.$$

When the null hypothesis is true and all assumptions are satisfied (31) has an F-distribution with k_i and m_i-k_i-1 degrees of freedom. The test is conducted by comparing (31) with the upper quantiles of this distribution. To test whether the <u>l'th degree component of the trend in zero in climate i,</u> we compare $F = \hat{a}^2_{i\ell} / \left\{ SSR_i/(m_i-k_i-1) \right\}$ with the upper tail critical values of the F-distribution with 1 and m_i-k_i-1 degrees of freedom.

To test for <u>equality of the l'th degree components</u> of the trends in a pair of climates, say p and q, the appropriate hypotheses are

(32) $H : a_{p\ell} = a_{q\ell}$ vs. $H : a_{p\ell} \neq a_{q\ell}$

and the corresponding test statistic is

$$(33) \qquad F = \left\{ \hat{a}_{p\ell} - \hat{a}_{q\ell} \right\}^2 / \left\{ (SSR_p + SSR_q)/(m_p+m_q-k_p-k_q-2) \right\}$$

This statistic has an F-distribution with 1 and $m_p+m_q-k_p-k_q-2$ degrees of freedom when the null hypothesis is true and when both climates have the same inter-annual variance after trend is taken into account. Again, the test is conducted by comparing the test statistic with the upper quantiles of the null distribution. The test statistic is derived by noting that $E\left\{(\hat{a}_{p\ell} - \hat{a}_{q\ell})^2\right\} = E\left\{\hat{a}^2_{p\ell}\right\} + E\left\{\hat{a}^2_{q\ell}\right\}$.

To test for <u>equality of trends</u> in a pair of climates, say p and q, the appropriate hypotheses are

$$(34) \qquad H_o: a_{p\ell} = a_{q\ell} \text{ , } \ell=1, \ldots ,k \quad \text{vs.} \quad H_a: a_{p\ell} \neq a_{q\ell} \text{ for some } \ell$$

where k is the (common) degree of the fitted trend polynomial. The corresponding test statistic is

$$(35) \qquad F = \left\{ \sum_{\ell=1}^{k} (\hat{a}_{p\ell} - \hat{a}_{q\ell})^2/k \right\} / \left\{ (SSR_p + SSR_q)/(m_p+m_q-2k-2) \right\}$$

This statistic has an F-distribution with k and m_p+m_q-2k-2 df when the null hypothesis is true and when both climates have the same inter-annual variance after trend is taken into account. The test is conducted by comparing the test statistic with the upper quantiles of the null distribution. An obvious generalization allows one to compare k trend coefficients simultaneously or to compare the common trend coefficients in the event that both trend polynomials are not of the same degree.

3.2.3 A Test Concerning the Equality of Means

We now address the question of the equality of equilibrium means in a pair of simulated climates. As in subsection 3.1.4 above, this test is complicated somewhat by the fact that one or both of the climates may have a mean which is varying in time and that we are using a polynomial representation of this time dependency. As in section 3.1.4, we will assume that an equilibrium has been obtained obtained in both of the climate simulations, after say s_p and s_q years respectively. An appropriate set of hypotheses for our consideration is

(36) $$H_o: \mu_{equil,p} = \mu_{equil,q} \quad \text{vs.} \quad H_a: \mu_{equil,p} \neq \mu_{equil,q}.$$

With the assumption above, the equilibrium mean of climate p is given by

$$\mu + \alpha_p + \sum_{j=s_p}^{m_p} \left[\sum_{\ell=1}^{k_p} a_{p\ell} \cdot p_{p\ell}(j) \right] / (m_p - s_p + 1).$$

An unbiased estimator of this quantity is given by

$$\hat{\mu}_{equil,p} = \bar{y}_{p..} + \sum_{j=s_p}^{m_p} \left[\sum_{\ell=1}^{k_{\hat{k}}} \hat{a}_{p\ell} \cdot p_{p\ell}(j) \right] / (m_p - s_p + 1).$$

and has variance which is given by

$$\left\{ \frac{1}{m_p} + \sum_{\ell=1}^{k_p} \left[\sum_{j=s_p}^{m_p} p_{p\ell}(j) \;/\; (m_p - s_p + 1) \right]^2 \right\} \cdot \left\{ \sigma^2(C_i) + \sigma^2(\bar{\varepsilon}_{ij.}) \right\}.$$

If we now assume that the two climates in question have the same inter-annual variance, then it can be shown that an appropriate test statistic for (36) is given by

$$(37) \qquad T = \left[\hat{\mu}_{equil,p} - \hat{\mu}_{equil,q} \right] / S_{pool}$$

where

$$S^2_{pool} = \Bigg[\left\{ \frac{1}{m_p} + \sum_{\ell=1}^{k_p} \left[\sum_{j=s_p}^{m_p} p_{p\ell}(j) \;/\; (m_p - s_p + 1) \right]^2 \right\} \cdot SSR_p + \left\{ \frac{1}{m_q} + \sum_{\ell=1}^{k_q} \left[\sum_{j=s_q}^{m_q} p_{q\ell}(j) \;/\; (m_q - s_q + 1) \right]^2 \right\} \cdot SSR_q \Bigg] \Big/ \left[m_p + m_q - s_p - s_q + 2 \right]$$

When the null hypothesis is true this statistic has a Student's t-distribution with $(m_p+m_q-s_p-s_q+2)$ df. Unless we have *apriori* knowledge about the sign of the difference of means the test is conducted by comparing the test statistic against both upper and lower quantiles of the t-distribution. When such information is available, the comparison is made only against only one of these quantiles as dictated by the *apriori* known sign.

It should be noted that the tests for trend and equality of means do not use the connection between the paired coupled and uncoupled simulations. They use only information which is contained in the inter-simulation plus inter-annual variation sum of squares. These tests can be developed as easily by beginning with a separate version of (9) for each climate and making no assumption about a connection between the climates. On the other hand, tests of hyptheses which address questions regarding climate variability do exploit the connection which is assumed in (24). These tests are described next.

3.2.4 A Test of Potential Predictability

To test for <u>potential predictability in climate i</u> the appropriate hypotheses are

$$(38) \quad H_o: \sigma^2(C_i) = 0 \quad \text{vs.} \quad H_a: \sigma^2(C_i) > 0$$

and the appropriate test statistic is

$$(39) \quad F = \left\{ SSR_i/(m_i-k_i-1) \right\} / \hat{\sigma}^2_{pool}$$

where $\hat{\sigma}^2_{pool}$ is a frequency- or time-domain estimate of the variance due to climate noise constructed from all the climate simulations. If $\hat{\sigma}^2(\bar{\varepsilon}_{ij.})$ is an estimate of the variance due to climate noise in climate i which has ν_i df, then $\hat{\sigma}^2_{pool}$ is given by

$$(40) \quad \hat{\sigma}^2_{pool} = \sum_{i=1}^{I} \nu_i \cdot \hat{\sigma}^2(\bar{\varepsilon}_{ij.}) \Big/ \sum_{i=1}^{I} \nu_i$$

and has $\sum_{i=1}^{I} \nu_i$ df. When the null hypothesis is true the F-ratio is approximately distributed as an F random variable with (m_i-k_i-1) and $\sum_{i=1}^{I} \nu_i$ df. The test is conducted by comparing test statistic (39) with the upper quantiles of this distribution.

3.2.5 Comparision of Climate Noise Estimates

The test described in the previous subsection makes use of the statistical connection which is built into (24) by pooling the individual estimates of the variance due to climate noise. It is possible to check that this connection (the assumption that the variance due to climate noise is the same in all simulations) is valid by conducting yet another F-test to compare estimates of the variance due to climate noise in specific pairs of climates. In this case the appropriate hypotheses are

$$(41) \quad H_o: \sigma^2(\bar{\varepsilon}_{pj.}) = \sigma^2(\bar{\varepsilon}_{qj.}) \quad \text{vs.} \quad H_a: \sigma^2(\bar{\varepsilon}_{pj.}) \neq \sigma^2(\bar{\varepsilon}_{qj.})$$

and the appropriate test statistic is

$$(42) \quad F = \hat{\sigma}^2(\bar{\varepsilon}_{pj.}) \,/\, \hat{\sigma}^2(\bar{\varepsilon}_{qj.}).$$

When the null hypothesis is true this statistic is approximately F-distribution with ν_p and ν_q degrees of freedom. Unlike all previously described F-tests, the present test is *two sided* and the null hypothesis should be rejected if F is significantly less than *or* greater than 1. If the test is to be conducted at the α significance level (42) should be compared with both the $\alpha/2$ and $1-\alpha/2$ quantiles of the reference distribution. Most tables only give upper tail quantiles for the F distribution. Lower tail quantiles can be obtained by using the fact that $F_{\nu_p,\nu_q,\alpha/2} = F^{-1}_{\nu_q,\nu_p,1-\alpha/2}$.

As well as making comparisons between pairs of estimates of climate noise, it is also possible to apply Bartlett's Test (Sachs, 1982) to test the hypothesis that all simulated climates have the same variance due to climate noise.

3.2.7 Comparison of Inter-Annual Variances

Finally we describe a test concerning the equality of inter-annual variances in a pair of climates. The appropriate hypotheses are

$$(43) \quad H_o: \sigma^2(C_p) + \sigma^2(\bar{\varepsilon}_{pj.}) = \sigma^2(C_q) + \sigma^2(\bar{\varepsilon}_{qj.}) \quad \text{vs.} \quad \sigma^2(C_p) + \sigma^2(\bar{\varepsilon}_{pj.}) \neq \sigma^2(C_q) + \sigma^2(\bar{\varepsilon}_{qj.})$$

The appropriate test statistic is

$$(44) \quad F = \left\{ SSR_p/(m_p-k_p-1) \right\} / \left\{ SSR_q/(m_q-k_q-2) \right\}.$$

which has an F-distribution with m_p-k_p-1 and m_q-k_q-1 df when the null hypothesis is true. The test should be conducted by comparing (44) with both the upper and lower quantiles of this reference distribution because there is no *apriori* reason to suppose that one variance should be larger than the other. As in the previous subsection, we may also apply Bartlett's test to test the equality of all inter-annual variances simultaneously.

Note that comparisions of inter-annual variance may be regarded as comparisons of potential predictability if it is possible to assume that the climates under consideration have the same variance due to climate noise. When this is the case, expected differences are due entirely to differences in the variance of the transient climate signals, and hence are due to differences in potential predictability.

4. Discussion

Several statistical models and tests have been described which may be suitable for the study of variability simulated in groups of related climate simulations as well as in an individual climate simulation in which the mean has time dependent behaviour. The techniques which have been described are an extension of the methodology used in potential predictability studies and thus suffer from the same limitations. Many assumptions are made implicitly when these techniques are used which do not correspond to our knowledge of climate system mechanics. These techniques should thus be viewed as tools which can help to uncover evidence regarding the climate puzzle, but certainly not as final arbiters.

Statistical procedures are also limited simply by the manner in which they operate. Specifically, the null hypothesis is only rejected when there is sufficient evidence as measured by some test statistic. Failure to reject the null hypothesis *does not* indicate that it is true; it simply indicates that there was not enough evidence to reject the null hypothesis.

The techniques described above are relatively straight forward extensions of ordinary analyses of variance (see Sachs, 1982) which statisticians apply every day. The models described above have been used in a specific application (Latif and Biercamp, 1987; Zwiers and Storch, 1987) and it is partly the purpose of this paper to document those techniques. However, specific applications will no doubt require specific models for the analysis of variance for which specific tests will have to be derived. Thus it is also partly the purpose of this paper to illustrate by example the derivation of statistical tests in analysis of variance models. The principles employed in this paper are universal in such models.

Acknowledgements

I gratefully acknowledge that this work was partially supported by the Max Planck Institute for Meteorology.

5. References

Gates, W.L., Y.-J. Han, and M.E. Schlesinger, 1985: The global climate simulated by a coupled atmosphere-ocean general circulation model: preliminary results. In **Coupled Ocean Atmosphere Models**, J.C.J. Nihoul (ed), Elsevier, Amsterdam, 131-151.

Jones, R.H., 1976: On estimating the variance of time averages, **J. Appl. Meteor., 15**, 514-515.

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

Latif, M. and J. Biercamp, 1987: A Ten Year Simulation with a Coupled Ocean-Atmosphere General Circulation Model, **Research Activities in Atmospheric and Oceanic Modelling**, G.J. Boer (ed), WMO/ICSU World Climate Research Programme, Report No. 10, Geneva, 8.18-8.20.

Livezey, R.E. and W.Y. Chen, 1983: Statistical Field Significance and its Determination by Monte Carlo Techniques. **Mon. Wea. Rev., 111**, 46-59.

Madden, R.A., 1976: Estimates of the natural variability of time averaged sea level pressure, **Mon. Wea. Rev., 104**, 942-952.

Madden, R.A., and D.J. Shea, 1978: Estimates of the Natural Variability of Time-averaged Temperatures over the United States. **Mon. Wea. Rev., 106**, 1695-1703.

Madden, R.A., 1983: Reply: Comments on "Natural Variability and Predictability". **Mon. Wea. Rev., 111**, 586-589.

Manabe, S., and K. Bryan, 1969: Climate Calculations with a Combined Ocean-Atmosphere Model, **J. Atmos. Sci., 26**, 786-789.

Manabe, S., K. Bryan and M.J. Spelman, 1975: A global ocean-atmosphere climate model. Part I. The Atmospheric Circulation. **J. Phys. Oceanogr., 5**, 3-29.

Manabe, S., K. Bryan and M.J. Spelman, 1979: A global ocean-atmosphere climate model with seasonal variation for future studies of climate sensitivity. **Dyn. Atmos. Oceans, 3**, 393-426.

Sachs, L, 1982: **Applied Statistics, a handbook of techniques**, translated by Z. Reynarowych, Springer-Verlag, New York, 706 pp.

Shukla, J., 1983: Comment on "Natural Variability and Predictability". **Mon. Wea. Rev., 111**, 581-585.

Shukla, J., and D.S. Gutzler, 1983: Interannual variability and predictability of 500mb geopotential heights over the northern hemisphere, **Mon. Wea. Rev., 111**, 1273-1279.

Storch, H. von, 1982: A Remark on Chervin-Schneider's Algorithm to Test Significance of Climate Experiments with GCM's. **J. Atmos. Sci., 39**, 187-189.

Thiébaux, H.J., and F.W. Zwiers, 1984: The Interpretation and Estimation of Effective Sample Size, **J. Clim. Appl. Meteor., 23**, 800-811.

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

Trenberth, K.E., 1984b: Some effects of finite sample size and persistence on meteorological statistics. Part II: Potential predictability, **Mon. Wea. Rev., 112**, 2369-2379.

Trenberth, K.E., 1985: Potential Predictability of Geopotential Heights over the Southern Hemisphere, **Mon. Wea. Rev., 113**, 54-64.

Washington, W. M., A. J. Semtner, G. A. Meehl, D. J. Knight, T. A. Mayer, 1980: A general circulation experiment with a coupled atmosphere, ocean and sea ice model. **J. Phys. Oceanogr., 10**, 1887-1908.

Zwiers, F.W., 1987: A Potential Predictability Study conducted with an Atmospheric General Circulation Model. **Mon. Wea. Rev., 115**, 2957-2974.

Zwiers, F.W., and H. von Storch, 1987: Inter-Annual Variability Simulated in a Coupled Tropical Ocean Global Atmosphere General Circulation Model. **Research Activities in Atmospheric and Oceanic Modelling**, G.J. Boer (ed), WMO/ICSU World Climate Research Programme, Report No. 10, Geneva, 8.21-8.23.