JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 100, NO. C6, PAGES 10,693–10,725, JUNE 15, 1995

Ocean variability and its influence on the detectability of greenhouse warming signals

Benjamin D. Santer,[1] Uwe Mikolajewicz,[2] Wolfgang Brüggemann,[3] Ulrich Cubasch,[4] Klaus Hasselmann,[2] Heinke Höck,[2] Ernst Maier-Reimer,[2] and Tom M. L. Wigley[5]

Abstract. Recent investigations have considered whether it is possible to achieve early detection of greenhouse-gas-induced climate change by observing changes in ocean variables. In this study we use model data to assess some of the uncertainties involved in estimating when we could expect to detect ocean greenhouse warming signals. We distinguish between detection periods and detection times. As defined here, detection period is the length of a climate time series required in order to detect, at some prescribed significance level, a given linear trend in the presence of the natural climate variability. Detection period is defined in model years and is independent of reference time and the real time evolution of the signal. Detection time is computed for an actual time-evolving signal from a greenhouse warming experiment and depends on the experiment's start date. Two sources of uncertainty are considered: those associated with the level of natural variability or noise, and those associated with the time-evolving signals. We analyze the ocean signal and noise for spatially averaged ocean circulation indices such as heat and fresh water fluxes, rate of deep water formation, salinity, temperature, transport of mass, and ice volume. The signals for these quantities are taken from recent time-dependent greenhouse warming experiments performed by the Max Planck Institute for Meteorology in Hamburg with a coupled ocean-atmosphere general circulation model. The time-dependent greenhouse gas increase in these experiments was specified in accordance with scenario A of the Intergovernmental Panel on Climate Change. The natural variability noise is derived from a 300-year control run performed with the same coupled atmosphere-ocean model and from two long (>3000 years) stochastic forcing experiments in which an uncoupled ocean model was forced by white noise surface flux variations. In the first experiment the stochastic forcing was restricted to the fresh water fluxes, while in the second experiment the ocean model was additionally forced by variations in wind stress and heat fluxes. The mean states and ocean variability are very different in the three natural variability integrations. A suite of greenhouse warming simulations with identical forcing but different initial conditions reveals that the signal estimated from these experiments may evolve in noticeably different ways for some ocean variables. The combined signal and noise uncertainties translate into large uncertainties in estimates of detection time. Nevertheless, we find that ocean variables that are highly sensitive indicators of surface conditions, such as convective overturning in the North Atlantic, have shorter signal detection times (35–65 years) than deep-ocean indicators (≥100 years). We investigate also whether the use of a multivariate detection vector increases the probability of early detection. We find that this can yield detection times of 35–60 years (relative to a 1985 reference date) if signal and noise are projected onto a common "fingerprint" which describes the expected signal direction. Optimization of the signal-to-noise ratio by (spatial) rotation of the fingerprint in the direction of low-noise components of the stochastic forcing experiments noticeably reduces the detection time (to 10–45 years). However, rotation in space alone does not guarantee an improvement of the signal-to-noise ratio for a time-dependent signal. This requires an "optimal fingerprint" strategy in which the detection pattern (fingerprint) is rotated in both space and time.

[1]Program for Climate Model Diagnosis and Intercomparison, Lawrence Livermore National Laboratory, Livermore, California.
[2]Max-Planck-Institut für Meteorologie, Hamburg, Germany.
[3]Institut für Logistik und Transport, Universität Hamburg, Hamburg, Germany.
[4]Deutsches Klimarechenzentrum, Hamburg, Germany.
[5]National Center for Atmospheric Research, Boulder, Colorado.

Paper number 95JC00683.
0148-0227/95/95JC-00683$05.00

1. Introduction

Until recently, most of our information concerning the response of the climate system to greenhouse gas (GHG) forcing has been obtained from experiments that have investigated the equilibrium response to a step function doubling of CO_2 [e.g., *Schlesinger and Mitchell*, 1987; *Mitchell et al.*, 1990]. Most of these experiments were performed with an atmospheric general circulation model (GCM) coupled to a model of the oceanic mixed layer [*Manabe and Stouffer*, 1980; *Hansen et al.*, 1984; *Washington and Meehl*, 1984] and generally integrated for 2 or 3 decades. Given the rapid response time of the oceanic mixed layer ($\leq$10 years), such simulations can be integrated to a statistical equilibrium state in a relatively short time. Signal-to-noise properties of the experiment are then investigated for the equilibrium state, using stationary time samples from both control and perturbation experiments [e.g., *Wigley and Jones*, 1981; *Barnett*, 1986; *Santer et al.*, 1991].

With improved computer capabilities, it is now possible to use state-of-the-art coupled atmosphere-ocean models to examine the more relevant question of the climate response to time-dependent changes in greenhouse gases [*Stouffer et al.*, 1989; *Washington and Meehl*, 1989; *Manabe et al.*, 1991; *Cubasch et al.*, 1992]. These models simulate the uptake and transport of heat by the deep ocean, which has characteristic timescales of centuries or longer [*Mikolajewicz et al.*, 1990]. The time dependence of the forcing and the slow response of the deep ocean introduce a time dimension to the problem of estimating signal-to-noise properties. The problem is to determine whether the trend in the signal (or its anticipated time evolution) can be distinguished statistically from the internally generated noise of the coupled ocean-atmosphere system on the relevant decade to century timescales. Both paleoclimate data [e.g., *Crowley and North*, 1991] and model studies [*Mikolajewicz and Maier-Reimer*, 1990; *Lorenz*, 1991; *Wigley and Raper*, 1991; *Zebiak and Cane*, 1991] indicate that the amplitude of temperature changes due to internally generated natural variability can be large on these timescales. In order to detect a time-evolving climate signal due to changes in GHG concentrations, it is therefore crucial to obtain reliable estimates of the low-frequency noise of natural climatic variability. Such estimates can be obtained from four sources, which we consider briefly below: instrumental and paleoclimate data records, noise-forced energy balance models (EBMs), fully coupled atmosphere-ocean GCMs (A/OGCMs), and stochastically forced ocean GCMs (OGCMs).

The first source, paleoclimate records with annual resolution, is available from a number of different proxies, such as corals, laminated sediments, ice cores, and tree rings [e.g., *Dansgaard et al.*, 1975; *Bradley*, 1985; *Bradley and Jones*, 1992; *Briffa et al.*, 1990, 1992; *Crowley and North*, 1991]. However, unraveling the history of climatic variability contained in such records is not a straightforward task. The spatial coverage is poor for paleoclimate data with annual resolution, and it is difficult to date and cross-check the climate information from different locations or different proxy sources, or to extract the climate signal from the noise introduced by other, nonclimatic factors. The reconstruction of a satisfactory, spatially coherent picture of climate variability over the past 1000 years is not yet possible and remains a challenge for the future.

Simple numerical models provide a second means of investigating the low-frequency noise properties of the climate system. Some of the first studies of noise behavior on timescales of decades to centuries involved stochastic forcing of energy balance models [e.g., *Hasselmann*, 1976; *Lemke*, 1977] or idealized models of atmospheric flow [*Lorenz*, 1984]. More recently, *Wigley and Raper* [1990, 1991] used an EBM with an upwelling-diffusive ocean and atmospheric stochastic forcing in order to examine the internally generated variability of globally averaged temperature, and *Kim and North* [1991] considered the temporal and spatial variability of surface temperature in a noise-forced zonal EBM.

These models introduce strong physical simplifications in order to achieve sufficient computational efficiency, thus making it possible to perform long integrations and detailed studies of the sensitivity to varying parameterizations. While such models reproduce the principal characteristics of the red natural variability spectrum, they are not able to simulate realistically processes such as the horizontal and vertical transport of heat, salt, and momentum in the global ocean, which probably govern the century timescale variability of the real climate system. Ideally (in the absence of real world data series of sufficient length), it would be desirable to determine the low-frequency natural variability by performing long ($\geq$1000 years) integrations with fully coupled atmosphere-ocean GCMs, the third method of estimating natural climate variability mentioned above. Owing to computational limitations, however, state-of-the-art coupled models generally have been integrated for periods of the order of 100–500 years only [*Stouffer et al.*, 1989; *Cubasch et al.*, 1992; *Delworth et al.*, 1993; *Manabe and Stouffer*, 1993], which are too short for reliable estimates of the century timescale natural variability. Only one coupled model control integration of length 1000 years has been published [*Stouffer et al.*, 1994]. Although coupled models may become the most reliable means of estimating natural climate variability once longer simulations are generally available, we have not used these data in our estimates of ocean variability.

As a compromise we have used the fourth, intermediate method: the estimate of the low-frequency noise from long (>1000 years) stochastically forced experiments with uncoupled ocean GCMs. These results yield adequate statistics for determining the variability on the 10- to 100-year timescales relevant for time-dependent greenhouse warming experiments with fully coupled models. A shortcoming of such estimates of low-frequency noise, however, is that it is not clear whether the low-frequency variance information derived from the ocean model alone is a good approximation to the results that would have been obtained in a long integration with a fully coupled ocean-atmosphere model. Although the important dynamics of the slow component of the climate system (the ocean) is correctly represented and the effective forcing by the atmosphere can be well simulated as white noise, ocean-only models do not include the potential feedback loops through the atmosphere. Another uncertainty arises through the spatial structure of the stochastic forcing, especially for the heat and net fresh water fluxes, since reliable observed data sets with global coverage and information on interannual variability do not exist for these quantities. A big advantage, however, over paleoclimate records is that data from stochastically forced ocean GCMs

are available for the entire global ocean, both over long periods and at high temporal resolution.

In this study we use estimates of the century timescale noise from two long stochastic forcing experiments performed by *Mikolajewicz and Maier-Reimer* [1990, 1994] (hereinafter referred to as MM90 and MM94, respectively). The two integrations differ in terms of the applied forcing. In the first experiment, the Hamburg large-scale geostrophic (LSG) ocean GCM [*Maier-Reimer et al.*, 1993] was forced for 3800 years by temporally white but spatially correlated anomalies of fresh water fluxes, superimposed on the prescribed climatological fresh water fluxes. We refer to this experiment hereinafter as SF ("stochastic freshwater flux"). In the second experiment, SALL ("stochastic forcing in all components"), the LSG ocean model was forced for 8000 years with white noise variability in fresh water fluxes, heat fluxes, and wind stress.

We use the noise information from these integrations to determine the significance of trends in ocean variables in several time-dependent greenhouse warming experiments recently performed with the ECHAM-1/LSG coupled atmosphere-ocean model. (The ECHAM-1 atmospheric GCM is a low-resolution version of the numerical weather-forecasting model developed at the European Centre for Medium-Range Weather Forecasts that was modified in Hamburg for climate applications.) The first experiment ("SCENA," for scenario A) starts with 1985 equivalent CO_2 concentrations, and the model is then integrated for 100 years with the GHG forcing specified in scenario A ("business as usual") of the Intergovernmental Panel on Climate Change (IPCC) [*Houghton et al.*, 1990]. The second, 150-year experiment ("EIN" for early industrialization) begins in 1935 and uses observed changes in equivalent CO_2 from 1935–1985 and scenario A forcing after 1985. The two integrations provide some insight into the magnitude of the "cold start" error [*Hasselmann et al.*, 1993] and its impact on detection time estimates (the cold start error is estimated by *Hasselmann et al.* [1993] but is not corrected in the SCENA run). We also consider a suite of three 50-year Monte Carlo experiments, each with identical scenario A forcing but starting from different initial conditions. This allows us to investigate uncertainties in our estimate of the signal that are related to imperfect knowledge of initial conditions. Details of these coupled-model experiments are given by *Cubasch et al.* [1992, 1994, 1995]. A pattern analysis of the atmospheric response in the SCENA integration and two further greenhouse warming experiments is presented by *Santer et al.* [1994].

The signal uncertainties illustrated by the SCENA, EIN, and Monte Carlo experiments show that there is a need for distinguishing between detection period and detection time. We can determine the particular date at which we will be able to detect some climate change signal only if we are confident that we have a reliable computation of the climate change signal in absolute time (i.e., in calendar years). This we do not have, neither for the cold-start-affected SCENA integration nor for the EIN integration, which does not have a true preindustrial start date. Furthermore, the lesson learned from the Monte Carlo experiments is that it is not enough to know one particular initial state and carry out one simulation. To define the statistical properties of the true time-evolving signal, we need to carry out an ensemble of long ($\geq$100 years) runs for different initial conditions in order to identify the model-generated internal variability that is superimposed on the climate change signal. Such ensembles are not presently available. In the following, we stress that the term "signal" is used to denote an estimate of the signal, with superimposed noise, rather than the true, underlying signal process.

To factor out signal evolution uncertainties, we therefore distinguish between detection period and detection time. We define detection period as the length of the climate time series that must be available in order to detect a given linear trend (growth rate $\overline{\beta}$) in the presence of the natural climate variability (at a specified significance level). Detection period is expressed in model years and is independent of reference time and the real time evolution of the signal. It depends only on $\overline{\beta}$, the noise properties of the natural variability time series, and the stipulated significance level.

Alternately, detection time is computed for an actual time-evolving signal from a greenhouse warming experiment. The time evolution of the signal need not be linear and can have any form. Thus it depends on both changes in the mean signal trend with time (which are specific to the assumed radiative forcing scenario) and the noise properties of the natural variability time series and is a function of the start date of the experiment (and the magnitude of the cold start error, if this has not been corrected). We will express detection times in model years relative to a reference time of 1985. This enables us to compare ocean signals in the SCENA and EIN experiments on a common basis. We caution, however, against interpreting "1985" too literally as the true calendar date, which would place unwarranted confidence in the precise details of the predicted signal evolution.

The present study should be viewed as an initial attempt to consider the time dimension of the signal-to-noise problem in a coupled ocean-atmosphere GCM with time-dependent GHG forcing. Previous investigations that addressed the time-dependent signal-to-noise problem in a greenhouse warming context have used signal information from a simple energy balance model [*Briffa et al.*, 1990] or from pseudo-transient experiments performed with an ocean GCM [*Mikolajewicz et al.*, 1993]. The focus of our investigation is on methodological aspects. We stress that there are (and will continue to be) model-dependent uncertainties in defining an enhanced greenhouse effect signal and in determining the magnitude and spatial structure of the low-frequency internally generated natural variability. The approach used here, to estimate the decade to century timescale noise from an idealized "ocean only" natural variability experiment, must be regarded as a temporary solution pending the availability of long ($\geq$1000 years) control integrations using either ocean models with realistic linear atmospheric feedback or fully coupled atmosphere-ocean GCMs.

Although the final section of this paper makes some attempt to follow the philosophy of *Hasselmann* [1979, 1993] in optimizing the signal-to-noise ratio, it does so only in a very simplified form. The correspondences with Hasselmann's strategy are threefold. First, we have reduced the dimensionality of the multivariate problem. Instead of considering the multivariate ocean signals and noise in the full grid point–time space, we have reduced the dimensionality to the order of 20 by using only a few characteristic integral properties of the ocean circulation. This enables us to obtain reliable estimates of the noise properties, since the effective number of independent time samples in the two long sto-

chastic forcing experiments examined here is at least several times larger than the reduced spatial dimension. Second, we have employed the concept of a "fingerprint." The fingerprint defines the direction in which the GHG signal is expected to lie. In our case it is a multivariate vector, containing information on the equilibrium response to GHG forcing of different ocean variables at large spatial scales (global or individual ocean basins). The fingerprint used here can be thought of as a simple multivariate pattern filter that is applied to both the ocean GHG signal and the ocean natural variability noise data in order to enhance signal-to-noise ratios (see section 5.4). Third, we have attempted to enhance the signal-to-noise ratios still further by optimizing the fingerprint. Optimization involves filtering out (spatial) components of the signal that are similar to the dominant patterns of natural variability noise [*Hasselmann*, 1979].

Implementation of the full optimal detection strategy outlined by *Hasselmann* [1993] would involve optimization of the combined space-time structure of a four-dimensional multivariate fingerprint relative to the noise. In the present study the optimization of the fingerprint is in space only. In other words, in attempting to detect linear trend signals, we have not applied an optimized time filter which depresses frequency bands with a large relative noise contribution. The application of the full "optimal fingerprint" strategy will be considered in a future paper.

In practice, it will be difficult to obtain observations for many of the spatially integrated ocean circulation indices that we use here. It is unlikely that measurement technology over the next several decades will significantly improve this situation. This is not a severe constraint here, since our focus is on methodological aspects of the signal-to-noise problem. We note also that the use of spatially integrated quantities facilitates a multivariate analysis of the model's ocean circulation and offers a rather simple way of examining the covariant behavior of different characteristic ocean variables and different ocean basins.

An important set of measurable integral properties of the ocean circulation is acoustic travel times for various combinations of transmitting and receiving stations [*Munk and Forbes*, 1989]. Although we have not explicitly considered travel times in this paper, our analysis should shed some light on the potential usefulness of these variables for detecting anthropogenic climate change [cf. *Mikolajewicz et al.*, 1993].

This paper is organized as follows. Section 2 begins by introducing the integrated ocean diagnostics we shall be using: the spatial averages (ocean basin or global scale) of such quantities as heat and fresh water fluxes, rate of deep water formation, salinity, temperature, transport of mass, and ice volume. We then describe and compare salient features of the oceanic variability in the uncoupled SF and SALL experiments.

Section 3 focuses on results from natural variability and greenhouse warming experiments performed with the fully coupled ECHAM-1/LSG atmosphere-ocean model [*Cubasch et al.*, 1992]. We first consider a 300-year control simulation (hereinafter referred to as CTL) and compare its variability with the variability simulated in the uncoupled SF and SALL experiments. We then discuss the coupled model's ocean response to the time-dependent GHG forcing specified in the SCENA and EIN experiments. Section 3 concludes with an analysis of the ocean response in three Monte Carlo experiments, each with identical GHG forcing but starting from different initial conditions.

In section 4 we compute detection periods and detection times for univariate data (individual spatially averaged ocean variables). This requires calculation of ε, the standard error of linear trends on timescales of 10–100 years, i.e., the standard deviation of the sampling distribution of linear trends in the presence of natural variability. We use ocean data from the SF and SALL experiments in order to compute ε, and we show how differences in ocean variability and signal uncertainties translate into different estimates of signal detection period and time.

Section 5 addresses the question of whether the use of a multivariate detection vector increases the probability of early detection of an ocean greenhouse warming signal. We first compute the empirical orthogonal functions (EOFs) of the spatially averaged, multivariate ocean time series in the two stochastic forcing experiments and then show how these modes provide information about the covariance between different circulation indices at different depths and spatial locations in the ocean. We next project the multivariate ocean greenhouse warming signal and natural variability noise onto a common fingerprint pattern, which defines the expected signal direction. Signal-to-noise ratios are then computed for the resulting EOF amplitude time series using the method outlined in section 4. Finally, we enhance the signal-to-noise ratio by rotating the fingerprint in the direction of low-noise components of the SF and SALL experiments. A brief summary and conclusions are given in section 6.

2. OGCM Stochastic Forcing Experiments

This section introduces the ocean diagnostics employed in the present study and the two stochastically forced ocean-only experiments used to derive the level of noise associated with natural variability in the models. Details and previous applications of the LSG OGCM used in these experiments are given by *Maier-Reimer et al.* [1993].

2.1. Ocean Diagnostics

To reduce the spatial dimension of the signal-to-noise problem, we consider integral properties that characterize the state of the ocean general circulation. For the computation of most of these quantities, the ocean is divided into four horizontal boxes. The first box covers the North Atlantic and the Arctic Ocean north of 30°N (including the Mediterranean), the second covers the North Pacific north of 30°N, and the third covers the Southern Ocean south of 30°S. The last box covers the tropical region of all oceans between 30°N and 30°S (For computing stream function, an additional box was defined, consisting of all Indian Ocean grid points north of 30°S). Quantities integrated over these boxes are the heat and fresh water fluxes between ocean and atmosphere (positive values indicating net heat and fresh water flux gains by the ocean, corresponding to warming and freshening of the surface, respectively), sea ice volume (relevant only in the North Atlantic and Southern Ocean), and potential energy loss due to convection as an indicator of the formation rate of North Atlantic Deep Water (NADW) and Antarctic Bottom Water (AABW).

The thermohaline circulation is described by the net inflow or outflow at depths below 1.5 km between the Atlantic,

Table 1. Comparison of Oceanic Mean States and Variability in Three Model Experiments

Ocean Variable	Abbreviation	Model		
		SF	SALL	CTL
Temperature at 4-km depth, globally averaged, °C	T4-MEA	2.3 ± 0.13	2.9* ± 0.04	2.9* ± 0.06
Salinity at 4-km depth, globally averaged, ppt	S4-MEA	34.8 ± 0.006	34.8 ± 0.004	34.7* ± 0.007
Mass transport through the Drake Passage, 10^6 m^3 s^{-1}	PSIDRA	119.0 ± 22.6	144.9 ± 3.0	101.6 ± 3.2
Ice volume (North Atlantic), 10^3 km^3	ICE-NA	20.7 ± 0.3	16.7* ± 1.0	35.5* ± 8.5
Ice volume (Antarctic), 10^3 km^3	ICE-AA	37.9 ± 10.7	76.1* ± 5.0	46.2 ± 19.7
Loss of potential energy by convection (North Atlantic), 10^{13} W	CEN-NA	80.9 ± 6.0	97.0 ± 16.8	99.5* ± 8.0
Loss of potential energy by convection (Antarctic), 10^{13} W	CEN-AA	228.3 ± 102.1	365.3 ± 25.3	249.6 ± 10.3
Heat fluxes (North Atlantic),† 10^{12} W	HFL-NA	−632.5 ± 42.6	−646.6 ± 59.7	−882.4* ± 47.5
Heat fluxes (North Pacific),† 10^{12} W	HFL-NP	−142.9 ± 51.2	−128.4 ± 58.2	−239.9 ± 43.9
Heat fluxes (30°N to 30°S),† 10^{12} W	HFL-TR	2175.6 ± 58.8	2817.4* ± 132.9	2682.1* ± 89.1
Heat fluxes (Antarctic),† 10^{12} W	HFL-AA	−1384.5 ± 523.5	−2042.0 ± 126.8	−1557.6 ± 52.4
Stream function at 1.5-km depth (Atlantic),‡ 10^6 m^3 s^{-1}	ME15-A	−17.3 ± 3.5	−18.2 ± 1.0	−22.6 ± 1.1
Stream function at 2.5-km depth (Atlantic),‡ 10^6 m^3 s^{-1}	ME25-A	−8.4 ± 4.9	−12.0 ± 1.3	−12.8 ± 1.1
Stream function at 1.5-km depth (Pacific),‡ 10^6 m^3 s^{-1}	ME15-P	21.3 ± 6.2	37.0 ± 1.2	27.8 ± 2.6
Stream function at 2.5-km depth (Pacific),‡ 10^6 m^3 s^{-1}	ME25-P	18.0 ± 8.8	39.7 ± 1.2	24.6 ± 2.5
Stream function at 1.5-km depth (Indian Ocean),‡ 10^6 m^3 s^{-1}	ME15-I	13.5 ± 1.9	17.1 ± 0.6	22.4* ± 0.7
Stream function at 2.5-km depth (Indian Ocean),‡ 10^6 m^3 s^{-1}	ME25-I	11.5 ± 2.3	13.5 ± 0.7	19.2* ± 0.7
Fresh water fluxes (North Atlantic),§ 10^3 m^3 s^{-1}	WFL-NA	120.4 ± 42.6	120.8 ± 16.9	147.8 ± 16.8
Fresh water fluxes (North Pacific),§ 10^3 m^3 s^{-1}	WFL-NP	118.1 ± 51.2	118.2 ± 20.7	96.7 ± 17.0
Fresh water fluxes (30°N to 30°S),§ 10^3 m^3 s^{-1}	WFL-TR	−779.3 ± 58.8	−781.5 ± 52.1	−901.4 ± 44.6
Fresh water fluxes (Antarctic),§ 10^3 m^3 s^{-1}	WFL-AA	543.8 ± 35.9	541.1 ± 21.5	625.8 ± 28.8

Means and standard deviations are computed from time series of spatially and globally averaged ocean diagnostics. Results are taken from the 3800-year SF stochastic forcing experiment [*Mikolajewicz and Maier-Reimer*, 1990], the 8000-year SALL stochastic forcing experiment performed by *Mikolajewicz and Maier-Reimer* [1994], and the 300-year CTL integration performed by *Cubasch et al.* [1992]. The SF and SALL experiments were both conducted with the Hamburg LSG OGCM, while the CTL integration was carried out with the fully coupled ECHAM-1/LSG atmosphere-ocean model.

*Mean state is more than 3 standard deviations larger or smaller than the corresponding mean state in the SF experiment (relative to the variability in the SF experiment). Note that the variability in the SF experiment is in many cases larger than in either the SALL or CTL experiments.

†Negative values denote net heat flux from ocean to atmosphere.

‡Negative values denote net southward transport. All values computed at 30°S.

§Negative values denote net fresh water flux from ocean to atmosphere.

Pacific, and Indian Oceans and the Southern Ocean at 30°S. Positive values in the Pacific and Indian Oceans indicate a net northward inflow, whereas the negative values in the Atlantic describe the outflow of NADW to the Southern Ocean. To monitor changes in the depth of the outflow, the same quantities are also computed for the transport below 2.5 km. Another important quantity characterizing the thermohaline circulation is the strength of the mass transport of the Antarctic Circumpolar Current (ACC). The properties of the deep water masses of the ocean are described by the globally averaged potential temperature and salinity at the 4-km level of the model.

These 21 integral quantities capture both the main features of the large-scale variations in the ocean circulation and of the thermohaline surface forcing. In this paper we present only the annual mean values, but all experiments described below were run with an annual cycle. A list of these integral quantities and their mean values for different natural variability experiments is given in Table 1.

2.2. SF Stochastic Forcing Experiment

In this experiment the Hamburg LSG OGCM was forced for 3800 years by temporally white (but spatially correlated) monthly anomalies of fresh water fluxes, superimposed on prescribed climatological fresh water fluxes [see MM90]. The amplitude of the fresh water flux forcing was chosen as 16 mm/month, approximately 20% of the observed global, annually averaged precipitation. The other surface forcing fields were taken from climatologies of monthly mean wind stress [*Hellerman and Rosenstein*, 1983] and monthly mean surface air temperature (Comprehensive Ocean-Atmosphere Data Set (COADS) [*Woodruff et al.*, 1987]). A Newtonian relaxation procedure was used in order to couple the temperatures in the surface layer to the COADS temperatures. Further details are given in MM90. The aim of the experiment was to examine the low-frequency response of the LSG OGCM to white noise forcing and test whether short-timescale atmospheric variability could be the origin of natural climate variability, as was proposed by *Hasselmann* [1976].

The ocean model demonstrated complex variability on a wide range of space scales and timescales. The response variance spectra of ocean circulation indices were typically red, with power increasing toward low frequencies for timescales up to several centuries. In order to study this variability, MM90 used both composites and a statistical technique known as principal oscillation pattern (POP) analysis [*Hasselmann*, 1988; *von Storch et al.*, 1988]. They found that the dominant eigenmode of the ocean circulation had a characteristic timescale of roughly 320 years. The signature of this mode was evident in the annual mean time series of a number of different spatially averaged ocean variables, such as mass transport through the Drake Passage, globally averaged deep-ocean temperature, and the formation rate of AABW (see Figure 1). The signals were strongest in the Southern Ocean. The eigenmode also affected circulation in the Pacific and Indian Oceans through interaction with the Antarctic Circumpolar Current (see also *Mikolajewicz and*

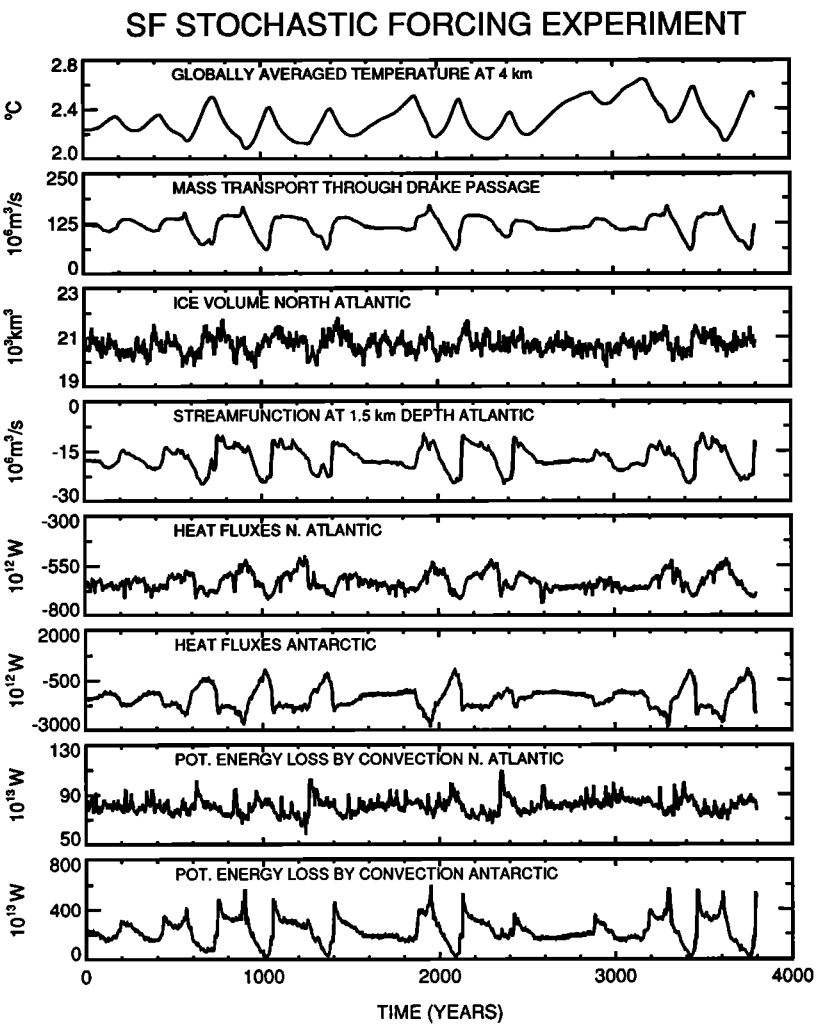

Figure 1. Time series of spatially averaged ocean variables from the 3800-year stochastic forcing experiment (SF) performed by *Mikolajewicz and Maier-Reimer* [1990]. All values are annual means. Note that the oceanic response to the fresh water flux forcing is characterized by variability on timescales of several centuries. This is the signature of the 320-year ocean circulation mode identified by Mikolajewicz and Maier-Reimer.

Maier-Reimer [1991]). MM90 interpreted the timescale of this mode as being related to the OGCM's characteristic flushing time for the Atlantic (~400 years). A subsequent study of the variability in the SF experiment by *Weisse et al.* [1994] revealed also the existence of a North Atlantic circulation mode with a characteristic timescale of 10–40 years.

To investigate the sensitivity of the ocean response to the amplitude of the forcing, *Mikolajewicz and Maier-Reimer* [1991] performed two further experiments in which they increased and decreased the amplitude of the forcing anomaly by a factor of 3 (from 16 mm month^{-1} to 48 and 5 mm month^{-1}, respectively). In the experiment with increased forcing amplitude, the ocean circulation rapidly switched into a new stable state characterized by permanent suppression of NADW formation. A reduction of the amplitude reduced the response significantly more than by the linear reduction factor of 3 and effectively eliminated the complex variability found in the original SF experiment. The 320-year eigenmode disappeared completely. This suggested the existence of some threshold value necessary for the excitation of the eigenmode. A recent suite of experiments by *Barnett et al.* [1993] attempted to identify this threshold by varying the amplitude and space-time coherence of the fresh water flux forcing. They found that triggering of the 320-year MM90 mode was highly sensitive to the amplitude of the forcing but not to its spatial coherence and that monthly anomalies of 30 mm month^{-1} were required to initiate the mode if no spatial correlation of the noise forcing was applied. These experiments also demonstrated that the period of the MM90 320-year mode was rather sensitive to the inclusion of a sea ice model.

2.3. SALL Stochastic Forcing Experiment

The original SF experiment was designed to look at the natural variability of the thermohaline circulation. Since preliminary experiments indicated that this was governed primarily by variations in the fresh water flux, variations in the heat flux and wind stress were omitted to establish a clear cause and effect relationship. Recently, MM94 investigated the impact of these other forcings by performing a

similar but longer 8000-year stochastic forcing experiment (SALL) in which the forcing consisted of uncorrelated monthly mean anomalies of all surface flux terms: momentum (wind stress), heat, and fresh water.

The forcing fields were taken from an experiment performed with the (T42 resolution) ECHAM-3 atmospheric GCM (AGCM), in which the atmospheric model was driven by the observed record of changes in sea surface temperature (SST) and sea ice distribution over the period 1979–1988. The integration was part of the Atmospheric Model Intercomparison Project (AMIP), an intercomparison of over 30 AGCMs driven by a standard set of observed boundary conditions [*Gates*, 1992].

The forcing in the SALL integration was constructed by random selection of individual months from the 10-year AMIP experiment. Thus for each January of the SALL integration, one of the 10 different AMIP Januaries was selected at random; a similar procedure was used for all other months. The anomalies of wind stress, fresh water fluxes and heat fluxes for the selected month of the AMIP experiment were then added to their respective climatologies and used as forcing for the OGCM (further experiments that included some correlation between successive months yielded essentially the same results).

An additional important difference between the two stochastically forced OGCM simulations was the treatment of the upper boundary condition for temperature. The SF experiment was run with relaxation of the model SST to observed near-surface air temperatures. Thus SST anomalies were damped out rather quickly with a time constant of 2 months (corresponding to a damping coefficient of 40 W m^{-2} $°C^{-1}$), whereas surface salinity anomalies tended to have a much longer lifetime and no direct damping owing to the prescription of climatological monthly mean fresh water fluxes. This type of mixed boundary condition qualitatively reproduces the nature of the air-sea interaction (damping of SST anomalies due to anomalous heat fluxes, whereas fresh water fluxes are independent of the surface salinity) and has been frequently used in recent experiments [e.g., *Bryan*, 1986; MM90; *Weaver and Sarachik*, 1991; *Power and Kleeman*, 1993]. However, simulations with this type of boundary condition generally underestimate SST variability. Recently, MM94 showed that a boundary condition for temperature consisting of prescribed climatological heat fluxes and subsequent weak damping of the model's SST gave a more realistic behavior of the uncoupled OGCM for climate simulations. In the SALL experiment this type of boundary condition was used with a damping time constant of 5 months.

The ocean variability simulated in the SF and SALL experiments shows considerable differences. Results for selected ocean circulation indices in the SALL experiment are given in Figure 2. In general, the SALL response is higher than the SF response in the 1- to 10-year timescale range, since variations on this timescale (e.g., El Niño–Southern Oscillation, or ENSO) are known to be governed mainly by wind stress fluctuations, which were not included in the SF simulation. The variability in the 1- to 10 year range tends to be underestimated in both experiments, however, since the LSG model is designed for the study of longer-term variability. In particular, it uses an implicit integration which damps out fast traveling waves (such as equatorial Kelvin waves) which play an important role in ENSO development.

We are more interested here in the longer-term variability. It is surprising that the SALL simulation, although driven by all three flux anomaly fields, generally exhibits a smaller response for longer periods. Thus the amplitude of the long-period variability in such circulation indices as mass transport through the Drake Passage and globally averaged deep-ocean temperature is much smaller than in the SF experiment (compare Figures 1 and 2) and their period is considerably longer than 300 years. In general, the SALL experiment exhibits a significantly lower level of ocean variability on the century timescale than the SF experiment, with the exception of ocean variables in the North Atlantic (ice volume, strength of NADW formation, and heat fluxes; see Table 1).

Clearly, even if the space-time evolution of an ocean greenhouse warming signal were perfectly known, the differences in the century timescale oceanic variability simulated in the two stochastic forcing experiments imply large uncertainties in estimates of the time period needed to detect such a signal.

Unfortunately, it is somewhat difficult to identify the source of the differences in ocean variability between the SF and SALL experiments, since these may be due to the differences in the boundary conditions for surface temperature and the associated difference in mean state, the amplitude and structure of the forcing patterns, and the fact that the stochasticity in the SALL experiment applies to all forcing terms, rather than to fresh water fluxes alone. However, the most likely cause for the reduced low-frequency response of the SALL experiment appears to be the weaker Newtonian coupling to the prescribed surface temperature, which increases the stability of the thermohaline circulation [see MM94; *Power and Kleeman*, 1994].

It should be noted that not only the variability but also the mean states of ocean variables (e.g., deep-ocean temperature, ice volume in the North Atlantic and Antarctic, and heat fluxes in the Antarctic) are different in the SF and SALL experiments (Table 1). Ice volume in the North Atlantic is much lower in the SALL experiment, while the reverse is true for ice volume in the Antarctic. The SALL integration also shows a more vigorous meridional circulation in the Pacific, larger heat fluxes from the atmosphere into the tropical ocean, and a 0.6°C warmer globally averaged deep-ocean temperature.

3. A/OGCM Signal and Noise Experiments

This section provides a brief summary of the experiments with the coupled A/OGCM (ECHAM-1/LSG) which are used in our signal-to-noise analysis of ocean data. Details are given by *Cubasch et al.* [1992, 1994, 1995], who also provide information relating to model validation and previous model applications.

3.1. Coupled Control Run

Both the SF and SALL experiments are idealized to the extent that they exclude atmospheric feedback loops, which can act to attenuate or amplify ocean circulation modes identified in the stochastic forcing experiments [see MM90]. Such feedbacks are automatically included in the 300-year CTL experiment performed with the fully coupled ECHAM-1/LSG A/OGCM, an extension of the original 100-year control run performed by *Cubasch et al.* [1992]. The atmospheric CO_2 concentration in this experiment was fixed at the level ob-

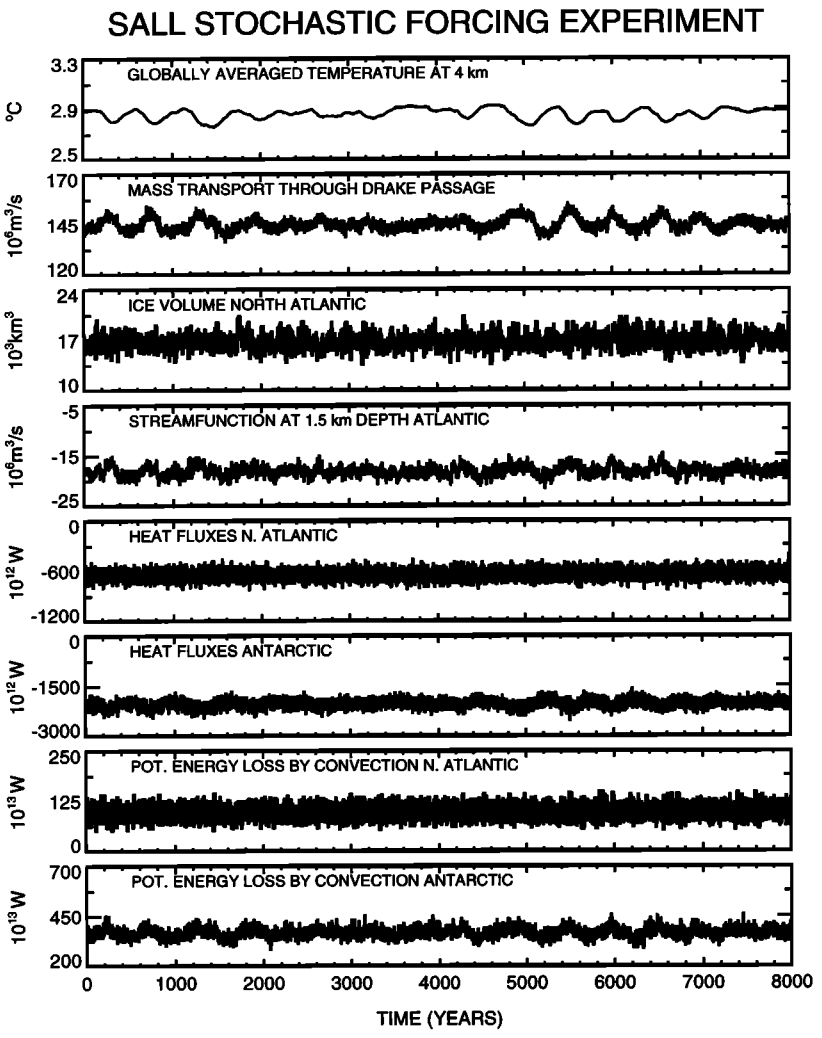

Figure 2. Time series of spatially averaged ocean variables from the 8000-year stochastic forcing experiment (SALL) performed by *Mikolajewicz and Maier-Reimer* [1994]. Results are annual means for the same ocean variables displayed in Figure 1. For details of the forcing, refer to section 2.3. The ocean variability simulated in the SALL experiment is very different from that generated in the original SF experiment. Although there is some long-period variability in such quantities as mass transport through the Drake Passage and globally averaged deep-ocean temperature, the amplitude of these oscillations is much smaller than in the SF experiment (compare Figure 1) and their period is considerably longer than 300 years.

served in 1985. To avoid climate drift of the coupled system, a flux correction scheme was employed [*Sausen et al.*, 1988]. The surface fluxes from the atmosphere into the ocean (heat, net fresh water flux, and wind stress) and the ocean surface data affecting the atmosphere (SST) were corrected both spatially and over the seasonal cycle. The correction terms can be regarded as constant anomaly fields which are added to the computed fluxes, enabling the atmosphere and ocean to receive the fluxes that they need in order to maintain a stable reference climate. For small perturbations that can be linearized, the constant flux corrections have no influence on the model response characteristics. The procedure reduces but does not completely eliminate climate drift.

The variability of near-surface air temperature in the first 100 years of the coupled control run has been described by *Cubasch et al.* [1992] and by *Santer et al.* [1994]. The latter investigation showed that while certain features of the variability in the control run could be interpreted as natural variability of the coupled model, other aspects of the variability were more consistent with residual climate drift caused by incomplete compensation in the flux correction procedure.

Results for selected ocean diagnostics from the CTL experiment are presented in Figure 3. The variables differ significantly in their temporal behavior. Surface quantities such as atmosphere-ocean heat exchange in the North Atlantic are relatively stationary, whereas Arctic ice volume and convection in the Southern Ocean show an initial drift. In the case of Southern Ocean convection, equilibration requires approximately 20 years. The equilibration time for Arctic ice volume is much longer, with ice volume increasing from an initial value of roughly 10×10^3 km^3 to an asymptotic value of approximately 40×10^3 km^3 after 100 years. The globally averaged temperature and salinity at 4-km depth showed an almost linear increase throughout the integration.

It is difficult to determine unambiguously whether the

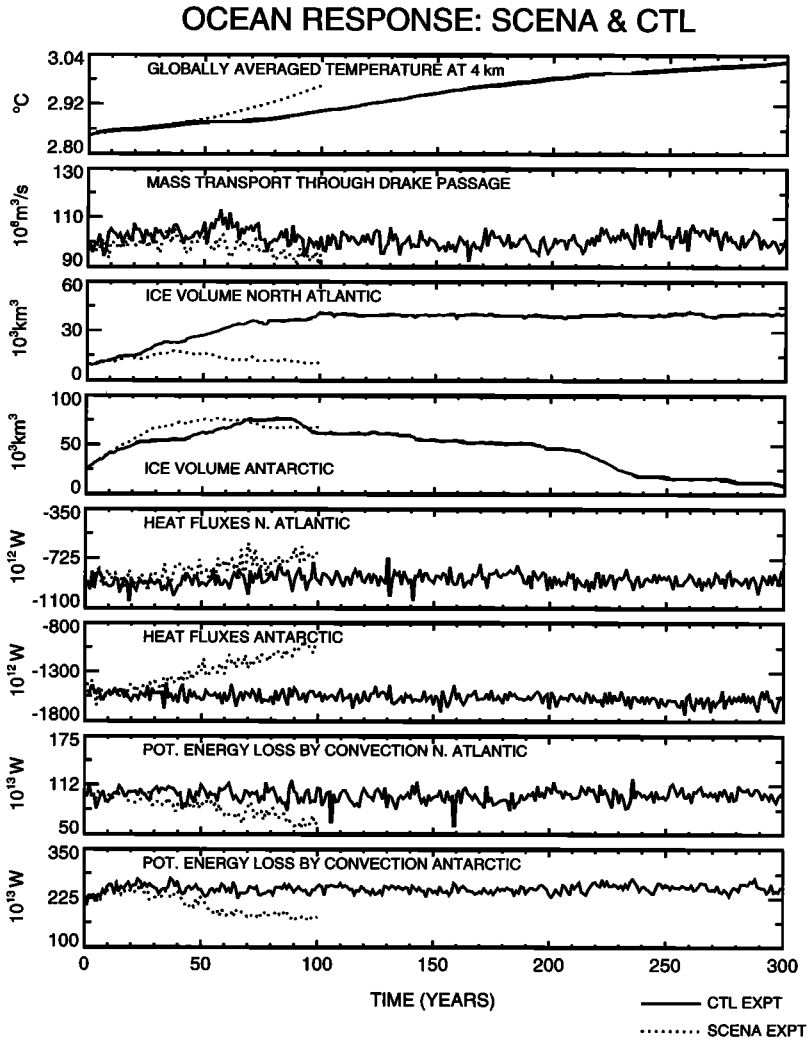

Figure 3. Time series of spatially averaged ocean variables from the 300-year control run (CTL, solid lines) and the 100-year scenario A greenhouse warming experiment (SCENA, dotted lines) performed with the ECHAM-1/LSG coupled atmosphere-ocean model [*Cubasch et al.*, 1992]. All values are annual means. In the control run, ice volume and globally averaged temperature show behavior consistent with a drift interpretation, while mass transport through the Drake Passage, ocean-atmosphere heat fluxes, and North Atlantic Deep Water (NADW) formation rate are quasi-stationary.

nonstationarity of ice volume and deep-ocean temperature and salinity represents bona fide natural variability of the fully coupled system or residual climate drift. The behavior of Arctic ice volume favors a drift interpretation, with slow equilibration over the first 80–100 years of the CTL experiment. This may be due to the fact that the flux correction procedure is conceptually not fully consistent with the nonlinear behavior of the system at the ice edge [*Cubasch et al.*, 1994]. The fact that incomplete flux compensation at the ice edge may also affect the properties of the newly formed deep and intermediate water provides one possible explanation for the almost linear increase in global mean deep-ocean temperature and salinity. We note, however, that the variability of deep-ocean temperature and salinity is not inconsistent with the amplitude of their fluctuations in the SF experiment (see Table 1), so that this behavior could also be due to natural variability rather than drift.

As in the SF-SALL intercomparison, the mean states in the SF and CTL experiments are very different for most ocean diagnostics (Table 1). The globally averaged temperature at 4-km depth is approximately 0.6°C warmer in the CTL experiment than in the SF integration. There is a tendency toward a more vigorous meridional circulation in the Atlantic, Pacific, and Indian Oceans in the coupled model control run, with associated increases in heat transfer from the ocean to the atmosphere in the North Atlantic, North Pacific, and Antarctic, and a stronger heat transfer into the ocean in the tropics.

3.2. Scenario A Experiment

The ECHAM-1/LSG coupled atmosphere-ocean model has also been used to perform a number of greenhouse warming experiments. In the SCENA integration [*Cubasch et al.*, 1992] the coupled model was forced by the time-dependent increases in equivalent atmospheric CO_2 concentration specified in the IPCC scenario A [*Houghton et al.*, 1990], which assumes unrestricted future emissions of greenhouse gases. The model was integrated for 100 years,

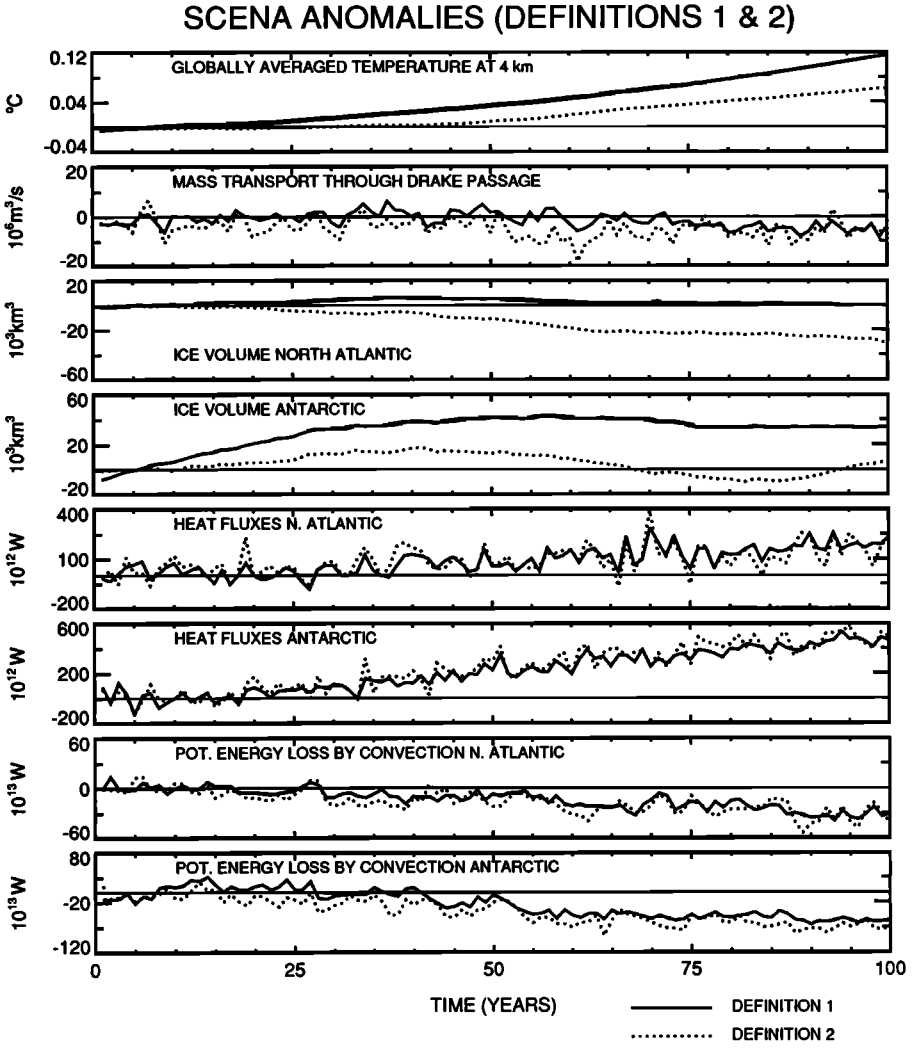

Figure 4. Greenhouse warming signals for spatially averaged ocean variables from the 100-year SCENA experiment [*Cubasch et al.*, 1992]. Results are for two different definitions of the signal. Definition 1 expresses the greenhouse warming signal relative to the smoothed initial state of the coupled model control run. Under definition 2 the signal is computed relative to the instantaneous state of the control run. All values are annual means. Because of the nonstationarity of ice volume and deep-ocean temperature (see Figure 3), the signals for these quantities are sensitive to the choice of definition.

corresponding to an increase of the equivalent atmospheric CO_2 concentration by a factor of 3 relative to the initial concentration in 1985.

The SCENA results for selected ocean diagnostics are shown in Figure 3, together with the previously discussed results from the 300-year CTL experiment. Some indices show clearly evolving signals relative to quasi-stationary behavior of the CTL experiment. Examples include the decreased strength of NADW and AABW formation rates and the consequent decrease in heat fluxes from the ocean to the atmosphere in the North Atlantic and Antarctic due to the suppression of upward convective heat flux [see *Mikolajewicz and Maier-Reimer*, 1991]. For other ocean diagnostics, such as deep-ocean temperature, the behavior of the coupled control run is nonstationary, and the precise definition of the signal is more problematic.

As was discussed by *Cubasch et al.* [1992] and *Santer et al.* [1994], the signal definition depends on assumptions regarding the correlation between the variability in the control run and the greenhouse warming experiment. If one assumes that the variability in the control run and response experiment are uncorrelated, it is appropriate to define the climate change signal by subtracting the initial state of the control experiment (definition 1). Alternately, one may assume that the control run changes represent a spurious residual drift or a long-term component of natural variability common to both experiments. In this case it is appropriate to subtract this common component and define the climate change signal as the instantaneous difference between the response experiment and control (definition 2). Subtraction of the long-term control run mean from a perturbation experiment [*Stouffer et al.*, 1994] also assumes uncorrelated variability in the two experiments, as in definition 1.

Figure 4 shows selected SCENA ocean signals according to the two definitions. The signals for indices which exhibit quasi-stationary behavior in the CTL experiment, such as NADW and AABW formation rates and heat fluxes in the North Atlantic and Antarctic, are relatively independent of

the choice of definition. In contrast, the signals for variables showing nonstationary behavior in the coupled control experiment are highly dependent on the way the signal is defined. For example, in both hemispheres the ice volume shows a large signal for one definition and zero signal for the other!

In view of the slow equilibration of Arctic ice volume in the CTL experiment (section 3.1), we regard definition 2 as more appropriate for this quantity. The CTL and SCENA similarities in the time evolution of the globally averaged deep-ocean temperature (Figure 3) and salinity (not shown) suggest that definition 2 might also be preferable for these ocean circulation indices. Since most other ocean variables are relatively insensitive to the choice of definition, we will use definition 2 except where explicitly stated.

3.3. Early Industrialization Experiment

The SCENA integration, with a start date in 1985, ignores the pre-1985 history of GHG forcing and assumes that the atmosphere and ocean are initially at equilibrium with respect to 1985 equivalent CO_2 concentrations. In the real world, pre-1985 GHG forcing may have lead to sequestering of heat in the intermediate and deep ocean [*Watts and Morantine*, 1991; *Kellogg*, 1993]. This previous warming is ignored in experiments with present-day start dates, leading to a retardation in global mean temperature increase (the so-called "cold start" error [*Cubasch et al.*, 1992; *Hasselmann et al.*, 1993]).

To investigate this error, *Cubasch et al.* [1995] recently repeated the SCENA experiment with a starting date in 1935 instead of 1985. This experiment (EIN) uses the observed GHG increases from 1935 to 1985 and GHG increases identical to those used in SCENA from 1985 to 2085. The experiment confirmed the linear cold start corrections of *Hasselmann et al.* [1993]. The earlier starting date of the EIN experiment reduces the impact of the cold start error but does not remove it completely.

In sections 4 and 5 we use data from the SCENA and EIN integrations to investigate the impact of the cold start error on detection time estimates for ocean signals. A qualitative comparison of the signals in the two experiments is shown in Figure 5. To compare the experiments on a common basis, the signal anomalies were defined relative to the average over the decade 1985–1994 of the respective experiment (i.e., years 1–10 of SCENA and 51–60 of EIN). For certain indices, such as deep-ocean temperature and strength of the ACC, the signal trends are clearly larger in the EIN experiment. The earlier starting date and longer forcing period have little apparent impact on other quantities, such as the NADW formation rate and ocean-atmosphere heat fluxes in the North Atlantic, which show very similar signal evolution in the two experiments. We cannot evaluate the robustness of such SCENA versus EIN similarities and differences in the signal evolution without performing further integrations to investigate the sensitivity of these results to initial condition uncertainties, as will be shown in the following section.

3.4. Monte Carlo Experiments

In order to investigate the statistical properties of a time-dependent greenhouse warming signal, *Cubasch et al.* [1994] recently performed an ensemble of three 50-year Monte Carlo experiments with the ECHAM-1/LSG coupled atmosphere-ocean model. These integrations used identical greenhouse gas forcing (the equivalent CO_2 increase from 1985 to 2034 specified in IPCC scenario A), but each started from different initial conditions of the *Cubasch et al.* [1992] CTL experiment, i.e., at years 30, 60, and 90. The *Cubasch et al.* [1994] study focused on atmospheric signals and showed that individual realizations exhibited substantial differences in the space-time structure of the 2-m temperature signal. This was due to the fact that different manifestations of natural variability noise were superimposed on the true, underlying GHG signal. Here we consider the statistical properties of the ocean signals in these Monte Carlo experiments.

Plate 1 compares the changes in globally averaged temperature at 4-km depth (top panel) and AABW formation rate (bottom panel) for the three Monte Carlo integrations (MC30, MC60, and MC90) and the SCENA integration. Deep-ocean temperature is an integral quantity which shows relatively small interannual variability, while the rate of AABW is strongly influenced by surface conditions and has much larger year-to-year variability. At the end of the SCENA integration, both variables exhibit clear signals. In the first 3–4 decades, however, the between-realization variability is as large as the ensemble average signal (MC-MEAN) after 40 years. The implication is that on these timescales, imperfect knowledge of initial conditions can translate into large uncertainties that make it difficult to define the signal unambiguously.

4. Univariate Estimates of Detection Period and Detection Time

In the preceding section we provided some preliminary comparison of the ocean variability in the SF, SALL, and CTL experiments. In this section we extend the comparison to the power spectra of selected ocean diagnostics and show how these differences in variability translate into different estimates of the standard error of a linear trend. Finally, we use univariate ocean data to estimate signal detection periods and times, and we show how both are affected by signal and noise uncertainties.

4.1. Spectra of Ocean Circulation Indices From SF, SALL, and CTL Experiments

The power spectra for various ocean circulation indices from the SF, SALL, and CTL experiments are shown in Plates 2a, 2b, and 2c. For the SF and SALL experiments, spectra were chunk averaged using nonoverlapping chunks of 1000 years in length (corresponding to 6 and 16 real degrees of freedom, respectively). For the CTL integration three nonoverlapping chunks with length of 100 years were used (6 degrees of freedom). Spectra were not normalized by their respective standard deviations.

Consider first the spectra for fresh water fluxes integrated over the entire North Atlantic and Arctic Ocean, which is representative of one of the forcing terms in the SF and SALL experiments (Plate 2a). In all three cases the spectra are almost white. For the stochastically forced OGCM experiments this is simply a reflection of the assumed lack of temporal correlation in the forcing. The coupled CTL experiment suggests that this is a valid assumption, at least in the North Atlantic.

Plate 2b shows spectra for one of the SF response variables, the atmosphere-ocean heat exchange in the North

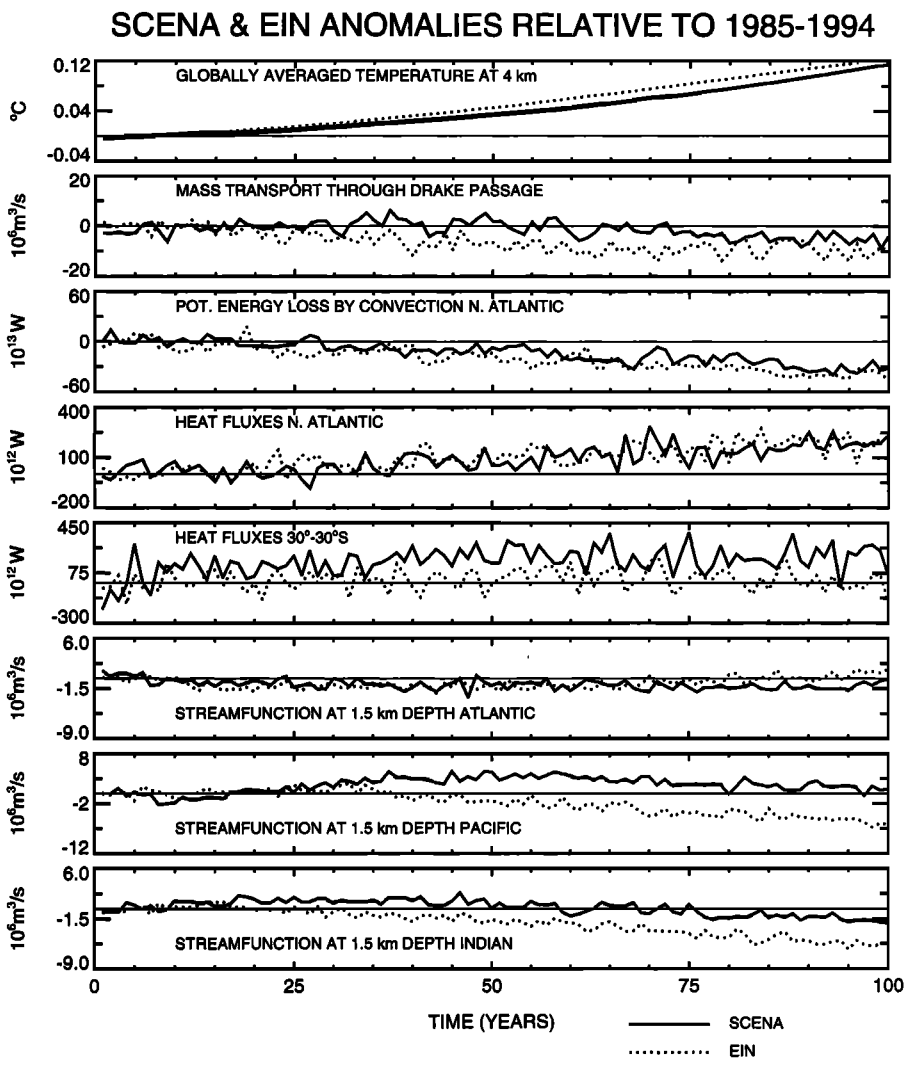

Figure 5. Greenhouse warming signals for spatially averaged ocean circulation indices from the 100-year SCENA experiment [*Cubasch et al.*, 1992] and from years 51–150 of the 150-year early industrialization experiment (EIN) [*Cubasch et al.*, 1995]. The start dates of the SCENA and EIN experiments were 1985 and 1935, respectively. The EIN experiment uses the observed greenhouse gas (GHG) increases from 1935 to 1985 and GHG forcing identical to SCENA from 1985 to 2085. To compare the experiments on a common basis, the signal anomalies were defined relative to the average over the decade 1985–1994 of the respective experiment. Results are expressed in model years rather than calendar years because we do not have a reliable estimate of the climate change signal in absolute time. The earlier starting date reduces the cold start error for such indices as deep-ocean temperature and mass transport through the Drake Passage.

Atlantic (in the SALL experiment, these fluxes are a mixed forcing-response term). This has a clear red response spectrum with a maximum at a period of approximately 320 years (Plate 2b). The corresponding spectra of the SALL experiment and the CTL simulation are white for timescales below 30 years, with some suggestion of reddening on timescales of 30–100 years. In the SALL experiment the atmosphere-ocean heat fluxes are a mixed forcing-response term. From the formulation of the boundary condition, the total heat flux consists of a prescribed component and a diagnostic term obtained from restoring SST to its climatological values. Because fluctuations of the prescribed forcing are white, the reddish component must therefore originate from changes in ocean circulation and the subsequent changes in the advection of heat.

A different picture emerges from the spectra of the mass transport of the ACC (Plate 2c). All spectra are red. The SF experiment has a discrete peak at approximately 320 years, which represents the signature of the salinity oscillator identified by MM90. The SALL spectrum has a peak at approximately 500 years, while the CTL integration is too short to identify any century timescale variance maxima.

In general, all ocean response terms involving deeper layers in the SF and SALL experiments show a monotonic increase in power as the period increases from 4 years to several hundred years, with discrete peaks at approximately 320 years and 500 years, respectively. The noise on timescales of less than 10–50 years is much larger in the SALL experiment than in the SF integration. The SF experiment in turn shows much greater variability on century timescales.

The ocean spectra from the CTL experiment are generally more similar to the SALL spectra than to the SF spectra in

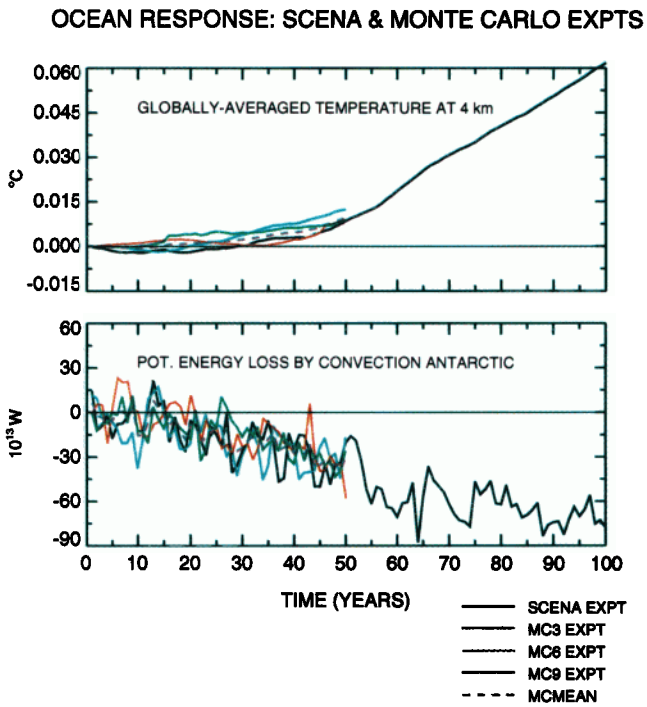

Plate 1. Changes in (top) globally averaged deep-ocean temperature and (bottom) Antarctic Bottom Water (AABW) formation rate in the original *Cubasch et al.* [1992] SCENA integration and in three "Monte Carlo" integrations (MC30, MC60, and MC90 [*Cubasch et al.*, 1994]) starting from different initial conditions of the CTL experiment. The mean of the three Monte Carlo experiments and the first 50 years of the scenario A integration is also shown (MCMEAN). At the end of the SCENA integration, both variables exhibit clear signals. In the first 3–4 decades, however, the between-realization variability is as large as the ensemble average signal after 40 years.

the frequency range where the three integrations overlap. Some of the CTL spectra are red (see, for example, Plate 2c), but as was noted in section 3.1, it is conceivable that part of this redness may be attributable to climate drift in the model.

4.2. Standard Error Estimates

In order to determine whether the linear trend in an ocean greenhouse warming signal β is statistically significant, we need to have information on ε, the standard deviation of the sampling distribution of the parameter which describes the

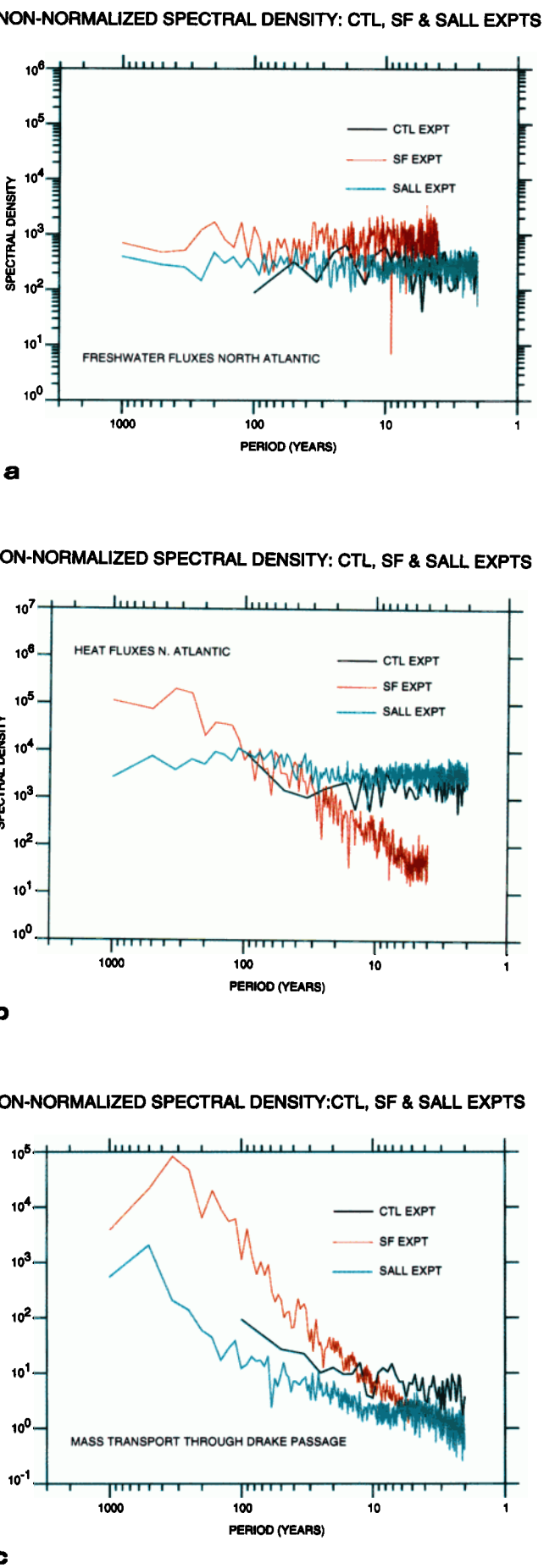

Plate 2. (Opposite) Nonnormalized power spectra for selected ocean variables from the CTL, SF, and SALL experiments. For the SF and SALL experiments, spectra were chunk averaged using three and eight nonoverlapping chunks of 1000 years in length, respectively. For the CTL integration three nonoverlapping chunks with length of 100 years were used. Results are for (a) fresh water fluxes and (b) atmosphere-ocean heat fluxes integrated over the entire North Atlantic and Arctic Ocean, and for (c) mass transport through the Drake Passage. For ocean response terms, the noise on timescales of less than 10–50 years is much larger in the SALL experiment than in the SF integration, while the SF experiment shows much greater variability on century timescales.

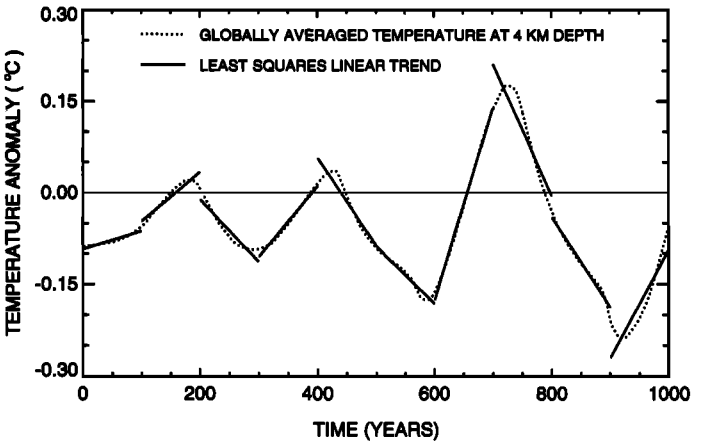

Figure 6. Least squares linear trends fitted to nonoverlapping 100-year chunks of a time series of deep-ocean temperature anomalies. Results are from the SF experiment.

linear trend. The standard error can be computed in either the time domain or the frequency domain [*Bloomfield and Nychka*, 1992]. Here we compute ε in the time domain.

Consider a time series $y(t)$, where $t = 1, \cdots, n$, from the SF or SALL experiment, where $y(t)$ is an integrated quantity such as deep-ocean temperature (because of the short integration length we do not use the CTL experiment to estimate ε). For a given chunk length L we use least squares regression to fit linear trends to each of m nonoverlapping chunks of this time series (Figure 6). For the SF experiment, for example, $L = 100$ and $m = n/L = 38$. The least squares trend for each chunk has a slope parameter $\beta(c)$, where $c = 1, \cdots, m$, and the standard error of the linear trend is then

$$\varepsilon = \left[(m-1)^{-1} \sum_{c=1}^{m} \beta(c)^2 \right]^{1/2} \quad (1)$$

In the following we show ε as a function of chunk length L, with L varying in the range 10, 20, $\cdots$, 100 years. Since the sample size m is small if L is large ($\geq$100) and chunks are nonoverlapping, we increased the sample size by using chunks that overlap by 1 year, so that $m = 3701$ for $L = 100$. The standard errors for nonoverlapping and overlapping chunks are very similar (Figure 7). The use of overlapping chunks provides little additional information but gives a much smoother picture of the decrease in ε with increasing chunk length. All subsequent standard error results therefore are based on overlapping chunks.

Figure 8 shows examples of the standard errors in the SF and SALL experiments. While both integrations present the same qualitative picture of a decrease in ε with increasing chunk length, the decay is generally more rapid for the SALL experiment. Quantitatively, the standard errors in the two experiments can differ in either direction by up to an order of magnitude. The difference is most pronounced for variables such as deep-ocean temperature, stream function in the Atlantic, and heat fluxes and ice volume in the Antarctic, for which standard errors are larger in the SF experiment. These results are consistent with the spectra of the SF and SALL ocean time series (section 4.1) and the generally higher variability of the ocean circulation in the SF experiment on timescales longer than approximately 10–30 years.

On shorter timescales, the standard errors in the SALL experiment are sometimes larger than those in the SF integration. This is seen most clearly in the atmosphere-ocean heat fluxes integrated over tropical oceans, which is a mixed forcing-response term in the SALL experiment.

The values of ε constitute the background noise level due to natural variability against which we must detect any ocean greenhouse warming signal. The large disparities between the SF and SALL standard errors translate into large uncertainties in estimates of the detection periods and detection times for ocean signals, as is discussed in the following sections.

4.3. Univariate Detection Period Estimates

In order to estimate detection periods, we first compute the SF and SALL standard errors ε_i as in (1), but now for chunk lengths $L_i = 10i$, where $i = 1, 2, \cdots, 10$. We then require a signal growth rate $\overline{\beta}$ for each ocean variable. This depends strongly on the ocean variable. An increase of 1% yr^{-1} is typical for some surface indices, while deep-ocean variables change much more slowly. We therefore choose $\overline{\beta}$ as the mean growth rate which yields the increase or decrease in a given ocean variable (according to our preferred definition 2) in the final decade of the SCENA experiment. We consider also growth rates $0.5\overline{\beta}$ and $2\overline{\beta}$ to reflect the large uncertainty range in linear signal evolution estimates.

The ratio $\overline{\beta}/\varepsilon_i$ is a measure of the signal strength relative to the natural variability noise. Assuming that the sampling distributions of linear trend parameters are normally distributed in both the SF and SALL experiments, we define the detection period T_p as the chunk length L_i for which the dimensionless quantity $\overline{\beta}/\varepsilon_i$ exceeds 1.96, the critical normal curve value appropriate for a stipulated significance level of 5% and a two-tailed test (we perform two-tailed tests because for many ocean variables, such as mass transport through the Drake Passage, we have no a priori information

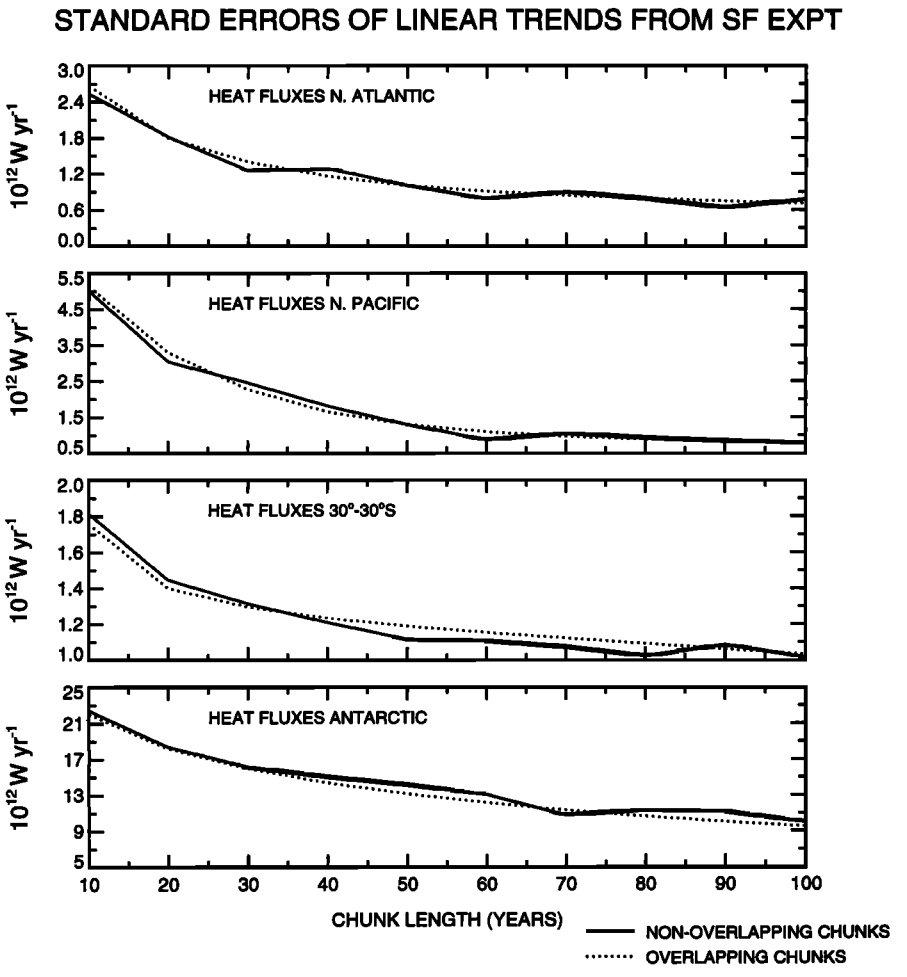

Figure 7. Standard error ε_i for four different spatially averaged ocean variables from the SF experiment. Standard errors are shown as a function of increasing chunk length L_i for both nonoverlapping chunks and chunks which overlap by 1 year. The overlapping chunks provide a smoother picture of the decrease in ε_i with increasing chunk length. The standard errors are a measure of the variance of linear trends.

on the direction of signal trends; i.e., we do not know whether they are likely to be positive or negative). In other words, the detection period is the chunk length for which the linear signal trend is nearly twice as large as the standard deviation of the sampling distribution of slope parameters for that chunk length.

Detection periods are highly sensitive to the natural variability noise differences in the SF and SALL experiments (Table 2). For the SF noise, less than one third (19 of 63) of the assumed growth rates were detectable within 100 years, while over two thirds (45 of 63) of the growth rates could be detected in the 10- to 100-year time frame relative to the SALL noise. These results are governed by the absolute differences in the standard errors (see Figure 10) and levels of ocean variability in the two stochastic integrations. Given the larger standard errors in the SF experiment, T_p is generally longer if noise is estimated from this integration. Exceptions are North Atlantic ice volume and NADW formation rate, where the SALL experiment has greater variability on timescales of 10–30 years.

Certain aspects of our results appear to be independent of the SF versus SALL standard error differences. First, indices that are highly sensitive to surface conditions, such as ice volume in the North Atlantic, ocean-atmosphere heat fluxes, and NADW formation rate, tend to have shorter detection periods than changes in deep-ocean temperature and salinity. Second, changes in mass transport in the Pacific and Indian Oceans have shorter detection periods than changes in transports from the Atlantic to the Southern Ocean. Third, detection periods can be long for the growth rates stipulated for fresh water flux terms, despite the fact that these quantities are white-noise-forcing terms in the SF and SALL experiments. This means that the changes in these indices at the end of SCENA (and hence the assumed growth rates) are relatively small.

The short detection periods for growth rates of North Atlantic ice volume (10–25 years) require some comment. This result is probably due to the prescription of atmospheric temperature, which is highly constrained in the SF experiment but less constrained in the SALL integration. This acts to suppress any variability directly associated with surface temperature. Of the ocean variables considered here, this will be manifested most clearly in the variability of sea ice. The detection periods and times for ice volume changes are therefore likely to be too optimistic.

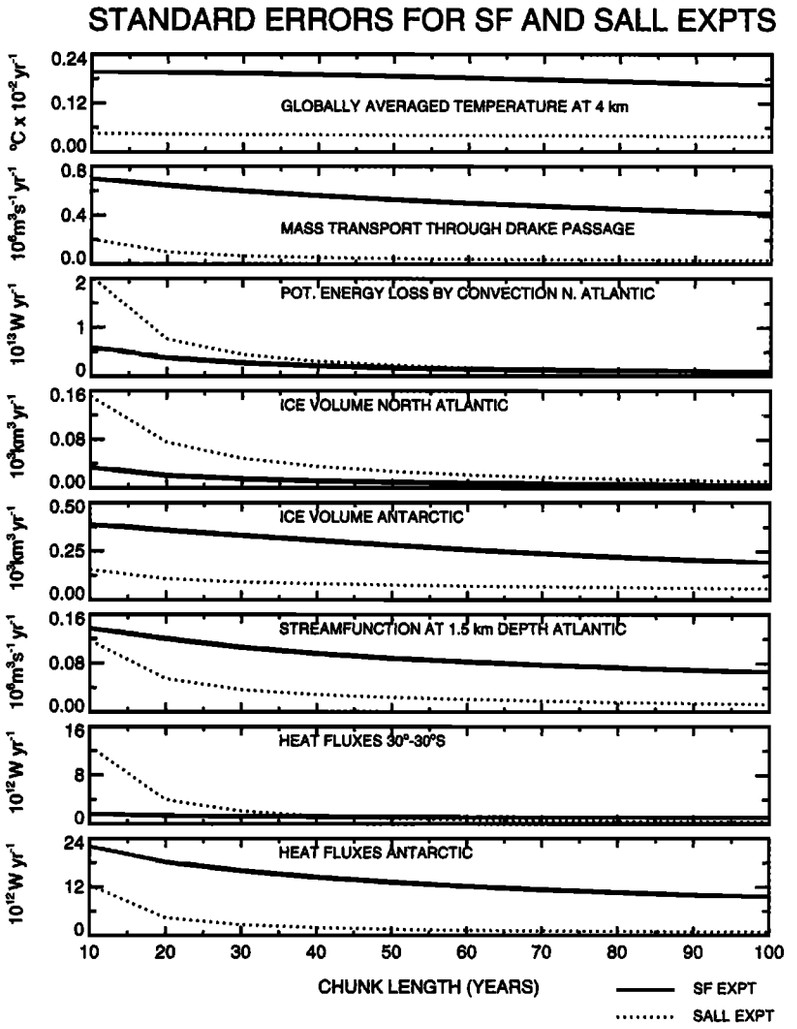

Figure 8. Standard errors ε_i for selected ocean diagnostics from the SF and SALL stochastic forcing experiments, shown as a function of increasing chunk length L_i. All values were computed using overlapping chunks. The standard errors in the two experiments can differ in either direction by up to an order of magnitude.

4.4. Univariate Detection Time Estimates

To compute detection time, we use the same SF and SALL standard errors ε_i employed in the calculation of detection period. Instead of assuming linear signal growth rates, we now use the real time-evolving signals in the SCENA and EIN experiments. The signals are described by the linear trend parameters β_i, with $i = 1, 2, \cdots, 10$, and are computed for intervals $L_i = 10i$ (i.e., for years 1–10, 1–20, $\cdots$, 1–100 of a given signal time series). Unlike $\overline{\beta}/\varepsilon_i$, the ratio β_i/ε_i now involves changes in both signal strength and noise levels as a function of increasing length of the time series.

In order to compare the SCENA and EIN experiments on a common basis, we computed the signal anomalies relative to the average over the decade 1985–1994 of the respective experiments (see Figure 5). Assuming Gaussian distributions for the linear trend parameters in the SF and SALL experiments, we define detection time T_d as the interval length L_i for which β_i/ε_i exceeds and remains above 1.96 (i.e., 5% significance for a two-tailed test). The qualification "remains above" is necessary because the signal trends are not constant with time (see Figure 5), so that β_i/ε_i may exceed 1.96 for a given interval but drop below this threshold for a longer interval. Because of the previously discussed problems associated with ice volume changes and the white noise nature of the SF and SALL fresh water flux terms, we have not computed detection times for these indices.

The detection times for the univariate SCENA and EIN signals show a behavior similar to the detection periods (Table 3). The SCENA and EIN signal trends could be detected within 100 years in only 7 of 45 cases for the SF noise, while the detection success, 31 out of 45 cases, was much higher for the SALL noise.

The effect of the cold start error on detection time is also evident. For most indices the EIN signals are detectable earlier than the SCENA signals. The largest improvements in detection time are for mass transport through the Drake Passage (70 years earlier) and transports in the Pacific and Indian Oceans (35–40 years earlier; see Table 3 and Figure 5).

Table 2. Detection Periods (in Years) for Univariate and Multivariate Ocean Data

Ocean Variable	Abbreviation	$\bar{\beta}$/yr	SF Noise			SALL Noise		
			$0.5\bar{\beta}$	$1.0\bar{\beta}$	$2.0\bar{\beta}$	$0.5\bar{\beta}$	$1.0\bar{\beta}$	$2.0\bar{\beta}$
Temperature at 4-km depth, globally averaged	T4-MEA	5.73×10^{-4} °C	N.D.	N.D.	N.D.	N.D.	N.D.	10
Salinity at 4-km depth, globally averaged	S4-MEA	3.73×10^{-6} ppt	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
Mass transport through the Drake Passage	PSIDRA	-4.72×10^{4} m^3 s^{-1}	N.D.	N.D.	N.D.	N.D.	N.D.	50
Ice volume (North Atlantic)	ICE-NA	-0.28×10^{3} km^3	10	10	10	25	15	10
Ice volume (Antarctic)	ICE-AA	1.13×10^{1} km^3	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
Potential energy loss by convection (North Atlantic)	CEN-NA	-0.33×10^{13} W	N.D.	55	25	100	65	40
Potential energy loss by convection (Antarctic)	CEN-AA	-0.72×10^{13} W	N.D.	N.D.	N.D.	N.D.	65	35
Heat fluxes (North Atlantic)	HFL-NA	1.59×10^{12} W	N.D.	80	25	85	50	30
Heat fluxes (North Pacific)	HFL-NP	1.02×10^{12} W	N.D.	N.D.	65	80	55	35
Heat fluxes (30°N to 30°S)	HFL-TR	0.56×10^{12} W	N.D.	N.D.	N.D.	N.D.	N.D.	70
Heat fluxes (Antarctic)	HFL-AA	4.91×10^{12} W	N.D.	N.D.	N.D.	65	35	20
Stream function at 1.5-km depth (Atlantic)	ME15-A	1.28×10^{4} m^3 s^{-1}	N.D.	N.D.	N.D.	N.D.	N.D.	100
Stream function at 2.5-km depth (Atlantic)	ME25-A	1.06×10^{4} m^3 s^{-1}	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
Stream function at 1.5-km depth (Pacific)	ME15-P	-6.35×10^{4} m^3 s^{-1}	N.D.	N.D.	N.D.	65	30	15
Stream function at 2.5-km depth (Pacific)	ME25-P	-6.00×10^{4} m^3 s^{-1}	N.D.	N.D.	N.D.	85	40	20
Stream function at 1.5-km depth (Indian Ocean)	ME15-I	-3.73×10^{4} m^3 s^{-1}	N.D.	N.D.	75	65	30	15
Stream function at 2.5-km depth (Indian Ocean)	ME25-I	-4.24×10^{4} m^3 s^{-1}	N.D.	N.D.	90	65	25	15
Fresh water fluxes (North Atlantic)	WFL-NA	0.26×10^{3} m^3 s^{-1}	N.D.	90	55	100	65	40
Fresh water fluxes (North Pacific)	WFL-NP	0.76×10^{3} m^3 s^{-1}	60	40	25	55	35	25
Fresh water fluxes (30°N to 30°S)	WFL-TR	-1.32×10^{3} m^3 s^{-1}	100	65	40	70	45	30
Fresh water fluxes (Antarctic)	WFL-AA	0.43×10^{3} m^3 s^{-1}	N.D.	N.D.	65	80	50	35
Unrotated fingerprint pattern			70	35	20	45	20	10
Rotated fingerprint pattern			45	20	10	55	20	10

Results are given for different noise experiments and signal growth rates. The detection period T_p is the length of a climate time series required until the ratio $\bar{\beta}/\varepsilon_i$ exceeds 1.96, indicating that the linear growth rate is significant at the 5% level or better. The growth rate $1.0\bar{\beta}$ yields the increase or decrease in a given ocean circulation index (according to definition 2) at the end of the 100-year SCENA experiment, so that $0.5\bar{\beta}$ and $2\bar{\beta}$ are one-half and twice the linear rates of change in SCENA, respectively. The standard errors ε_i are computed from two different natural variability noise experiments (SF and SALL). Detection times are given to the nearest 5 years, and "N.D." indicates that the prescribed linear trend could not be detected in at least 100 years. The final two rows give results for the multivariate analysis (see sections 5.4 and 5.5).

Table 3 also shows that the estimates of detection time T_d depend on the assumptions about the correlation between the SCENA and CTL variability, i.e., on the choice of definition 1 or 2 of the SCENA signal. This is not surprising in view of some of the results shown in Figure 4.

Note that heat fluxes integrated over tropical oceans yield a very optimistic detection time under definition 1 of the SCENA signal relative to the SALL standard errors (10 years; see Table 3). This is largely due to the strong drift in this quantity during the initial decade of the CTL integration (not shown). If the first 10 years of the CTL and SCENA experiment are excluded, the detection time for changes in tropical ocean-atmosphere heat fluxes increases from 10 to 40 years.

As was noted in section 3.4, a further source of signal uncertainty is related to our imperfect knowledge of initial conditions. Plate 3 illustrates this by showing β_i/ε_i relative to the SF standard errors for all signals (SCENA, the individual Monte Carlo experiments and the MCMEAN) and β_i/ε_i for the SALL standard errors and the SCENA signal only (to avoid complicating the diagrams, β_i/ε_i results for the Monte Carlo signals relative to the SALL standard errors are not shown). Consider results for NADW formation rate (Plate 3a). The SCENA signal for this quantity is detectable within 35 years (under the preferred definition 2), whereas the signals in the MC60, MC90, and MCMEAN experiments are not detectable within 50 years. This shows that the conclusion *Cubasch et al.* [1994] obtained for atmospheric variables is equally valid in an oceanic context: a single greenhouse warming integration of limited duration cannot provide insights into the statistical properties of the signal and may yield a misleading estimate of detection time. The Monte Carlo experiments indicate that the large local maxima and minima in β_i/ε_i for individual signals are not deterministic and tend to be smoothed out in the MCMEAN signal (see, for example, the results for deep-ocean temperature and strength of AABW formation, Plate 3b).

In summary, our results show that detection periods and detection times for univariate ocean greenhouse warming signals are highly sensitive to details of the decade to century timescale noise, which is in turn sensitive to the precise specification of the atmospheric forcing. Detection time estimates are also sensitive to signal uncertainties resulting from the cold start error, imperfect knowledge of initial conditions (and the possibly chaotic nature of the climate system), and assumptions regarding the relation between natural variability in the control and response experiments, which affect the definition of the signal. The implications of these results are discussed further in section 6.

5. Multivariate Estimates of Detection Period and Detection Time

In the preceding section we examined the greenhouse warming signals and associated detection periods and times for individual, spatially averaged ocean variables. We now consider whether we can achieve earlier detection of an ocean greenhouse warming signal by using a multivariate detection vector, rather than individual ocean variables only. This is the essence of the fingerprint approach to signal detection [*Madden and Ramanathan*, 1980; *MacCracken*

Table 3. Detection Times for Univariate and Multivariate Ocean Data

		SCENA					
		SF		SALL		EIN	
Ocean Variable	Abbreviation	Definition 1	Definition 2	Definition 1	Definition 2	SF	SALL
Temperature at 4-km depth, globally averaged	T4-MEA	N.D.	N.D.	60	N.D.	N.D. (N.C.)	35 (−25)
Salinity at 4-km depth, globally averaged	S4-MEA	N.D.	N.D.	25	N.D.	N.D. (N.C.)	30 (+5)
Mass transport through the Drake Passage	PSIDRA	N.D.	N.D.	N.D.	N.D.	N.D. (N.C.)	30 (−70)
Potential energy loss by convection (North Atlantic)	CEN-NA	50	35	65	60	50 (N.C.)	55 (−10)
Potential energy loss by convection (Antarctic)	CEN-AA	N.D.	N.D.	60	55	N.D. (N.C.)	55 (−5)
Heat fluxes (North Atlantic)	HFL-NA	65	N.D.	55	40	50 (−15)	45 (−10)
Heat fluxes (North Pacific)	HFL-NP	50	70	30	50	N.D. (+50)	45 (+15)
Heat fluxes (30°N to 30°S)	HFL-TR	N.D.	N.D.	10	N.D.	N.D. (N.C.)	N.D. (+90)
Heat fluxes (Antarctic)	HFL-AA	N.D.	N.D.	45	35	N.D. (N.C.)	30 (−15)
Stream function at 1.5-km depth (Atlantic)	ME15-A	N.D.	N.D.	N.D.	N.D.	N.D. (N.C.)	N.D. (N.C.)
Stream function at 2.5-km depth (Atlantic)	ME25-A	N.D.	N.D.	N.D.	N.D.	N.D. (N.C.)	N.D. (N.C.)
Stream function at 1.5-km depth (Pacific)	ME15-P	N.D.	N.D.	N.D.	65	N.D. (N.C.)	60 (−40)
Stream function at 2.5-km depth (Pacific)	ME25-P	N.D.	N.D.	N.D.	70	N.D. (N.C.)	65 (−35)
Stream function at 1.5-km depth (Indian Ocean)	ME15-I	N.D.	N.D.	80	60	N.D. (N.C.)	40 (−40)
Stream function at 2.5-km depth (Indian Ocean)	ME25-I	N.D.	N.D.	80	65	N.D. (N.C.)	40 (−40)
Unrotated fingerprint pattern		35	35	60	55	30 (−5)	30 (−30)
Rotated fingerprint pattern		10	10	45	10	25 (+15)	30 (−15)

Results are expressed in years relative to a reference date of 1985, and are given for different signal and noise experiments and different signal definitions. The detection time T_d is the length of a climate signal time series required until the ratio β_l/ε_l exceeds (and remains above) 1.96, indicating that the linear signal trend is significant at the 5% level or better. The ocean linear trend signals β_l are taken from the 100-year SCENA and 150-year EIN experiments, with start dates in 1985 and 1935, respectively. The standard errors ε_l are computed from two different natural variability noise experiments, SF and SALL. Detection times are given to the nearest 5 years, and "N.D." indicates that the time-evolving signal trend could not be detected in at least 100 years. The final two rows give results for the multivariate analysis (see sections 5.4 and 5.5). The EIN signal is defined relative to the decade 1985–1994 of the EIN experiment to facilitate comparison with the SCENA signal. The numbers in brackets in the final two columns indicate how much earlier (negative numbers) or later (positive numbers) the EIN signal can be detected relative to the SCENA signal; "N.C." denotes no change in detection time.

and Moses, 1982; *Barnett and Schlesinger*, 1987; *Wigley and Barnett*, 1990; *Hasselmann*, 1993].

5.1. Multivariate EOF Analysis

Consider as an example the normalized anomalies $\hat{z}(x, t)$ from the 3800-year SF stochastic forcing experiment

$$\hat{z}(x, t) = [z(x, t) - \bar{z}(x)]/s(x) \qquad x = 1, p; \quad t = 1, n \quad (2)$$

where $\bar{z}(x)$ is the 3800-year time average and $s(x)$ is the standard deviation of the time series. The discrete variable t denotes time in years, while x identifies the individual spatially averaged ocean variables in Table 1, excluding ice volume (because of the previously discussed drift problems) and the white noise forcing terms, so that $p = 15$ and $n = 3800$. The x index therefore denotes both different variables as well as different spatial averages for the same variable (e.g., ocean-atmosphere heat fluxes in the North Atlantic and Antarctic). Note that normalization is necessary, since the dimensions differ between variables.

The covariance matrix $c(x, y)$ is then defined as

$$c(x, y) = (n - 1)^{-1} \sum_{t=1}^{n} \hat{z}(x, t)\hat{z}(y, t) \qquad (3)$$

$$x = 1, p; \qquad y = 1, p$$

and represents the covariance between different ocean variables and different regions of the ocean, at very large spatial scales. The eigenvectors (EOFs) are defined by

$$\sum_{y=1}^{p} c(x, y)e_j(y) = l_j e_j(x) \qquad x = 1, p \qquad (4)$$

Since n is much larger than p, and since n is at least an order of magnitude larger than the dominant timescales of variability in the SF and SALL experiments, the EOFs can be well estimated. EOFs are normalized to form an orthonormal basis, so that

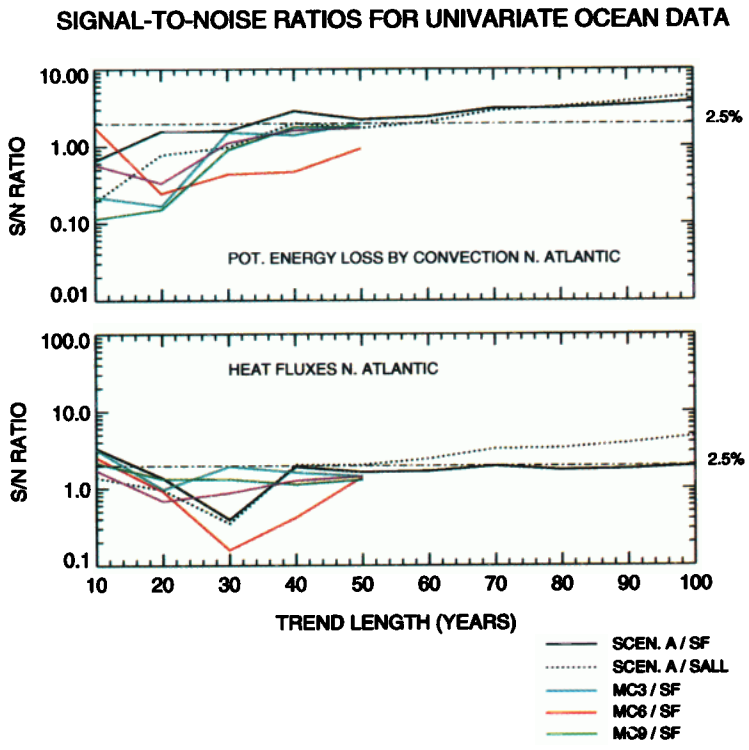

Plate 3a. Signal-to-noise ratios β_i/ε_i for ocean greenhouse warming signals from the scenario A and Monte Carlo experiments and standard errors from the SF and SALL stochastic forcing experiments. For a given ocean variable, the signal β_i is the slope parameter describing the linear trend in the ocean time series for chunk lengths 1–10 years, 1–20 years, etc. All signals are defined according to definition 2. The standard error ε_i is a measure of the variance of linear trends and is computed for chunk lengths corresponding to the signal length, i.e., for 10, 20, $\cdots$, 100 years. A signal-to-noise level of 1.96 represents the 5% significance threshold. Results are for NADW formation rate and atmosphere-ocean heat fluxes integrated over the North Atlantic.

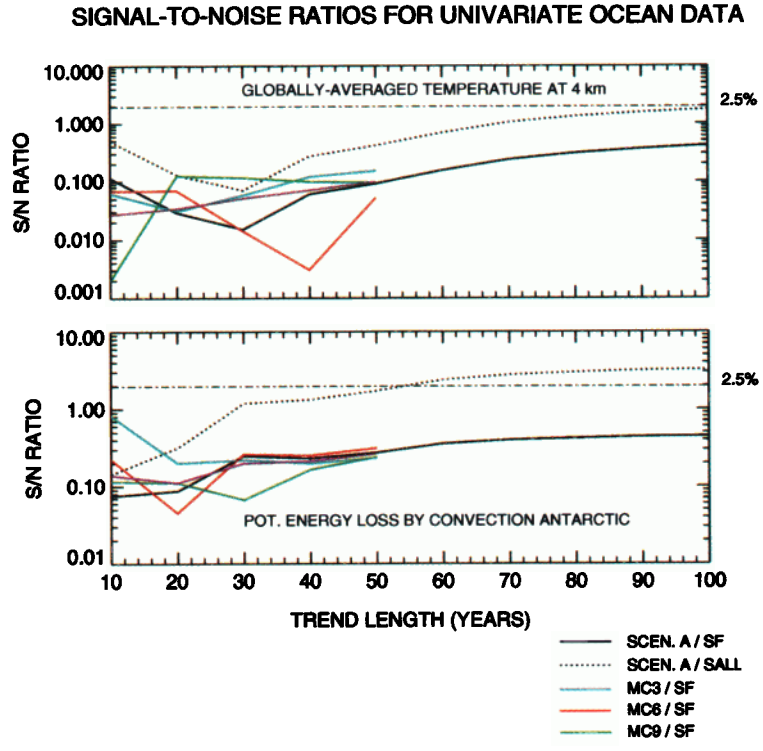

Plate 3b. As for Plate 3a, but for globally averaged temperature at 4-km depth and AABW formation rate. Note that the SF versus SALL differences in signal-to-noise ratio often exceed an order of magnitude and can mean the difference between detection or failure to detect the 100-year SCENA signal.

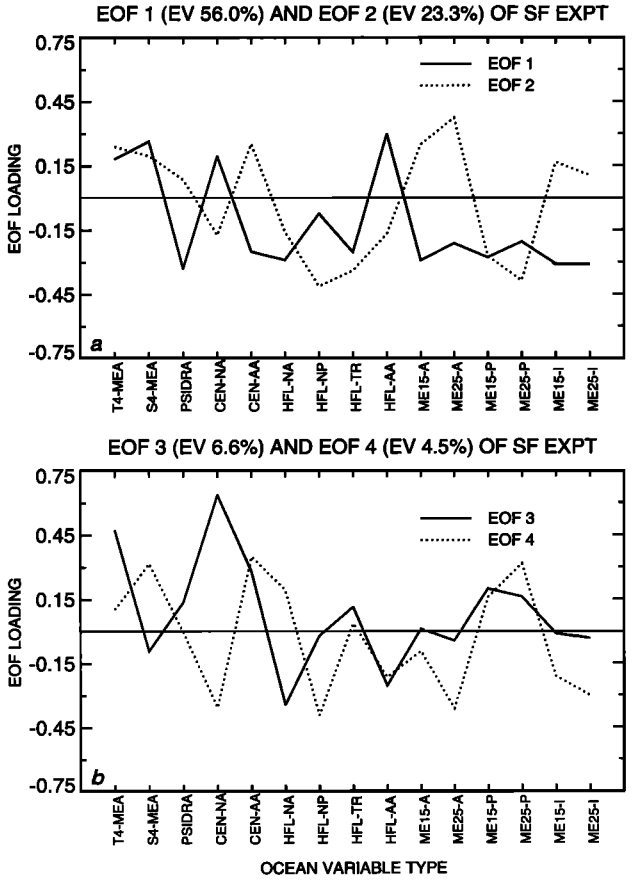

Figure 9. Multivariate empirical orthogonal functions (EOFs) 1–4 from the SF experiment. EOFs were computed from the 3800-year time series for 15 spatially averaged ocean circulation indices and provide information about the covariance between different ocean variables and different regions of the ocean (at the scale of individual ocean basins). The x axis indicates the ocean variable (see Table 1 for an explanation of the abbreviations). (a) EOFs 1 and 2 capture features of the 320-year "salinity oscillator" identified by *Mikolajewicz and Maier-Reimer* [1990, 1991]. (b) EOFs 3 and 4 may be related to "on" and "off" modes of the salinity oscillator (see section 5.2).

$$\sum_{j=1}^{p} e_j(x)e_j(y) = \delta_{xy} \qquad x = 1, p; \quad y = 1, p \qquad (5)$$

The EOF amplitudes $a_j(t)$ are defined by

$$a_j(t) = \sum_{x=1}^{p} \hat{z}(x, t)e_j(x) \qquad j = 1, p; \quad t = 1, n \qquad (6)$$

Most meteorological applications use the so-called S mode of EOF analysis [*Preisendorfer*, 1988], in which the first index of the input data set z is assumed to be spatial location and the second index runs over time. In this context, EOFs are spatial patterns and their amplitudes are time series that determine the relative importance of a particular pattern in a given year. In the mode of analysis used here, the EOFs carry some spatial information (at the scale of entire ocean basins) but also convey information about the relationships between different ocean variables.

5.2. Multivariate EOF Analysis: Results

SF experiment. Figure 9 shows EOFs 1–4 for 15 of the ocean time series from the SF stochastic forcing experiment. Consider first EOF 1. It has high positive or negative loadings on virtually all ocean variables, indicating that this mode is near-global in scale. The mode describes the anticorrelation between the strength of deep water formation: and that of atmosphere-ocean heat fluxes: increased NADW formation is linked with increased (i.e., larger negative) heat fluxes from the North Atlantic to the atmosphere ($r_{\text{CEN-NA:HFL-NA}} = -0.66$ for the normalized CEN-NA and HFL-NA time series). The same inverse relationship between heat fluxes and deep water formation rate applies in the Antarctic, where reduced formation of AABW results in reduced (i.e. smaller negative; see Figure 1) heat fluxes from the Southern Ocean to the atmosphere due to the suppression of upward convective heat flux ($r_{\text{CEN-AA:HFL-AA}} = -0.95$). NADW and AABW formation rates themselves show an inverse relationship: this was a prominent feature of the 320-year ocean circulation mode found in MM90. Our results suggest that EOF 1 is closely linked to this mode. The EOF 1 amplitude time series lends support to this hypothesis (see Figure 10a). It shows oscillations on timescales of several centuries, with maximum power at a period of roughly 320 years.

Other features of EOF 1 are also consistent with properties of the dominant MM90 mode, such as the strong correlation between mass transport through the Drake Passage and the AABW formation rate ($r_{\text{PSIDRA:CEN-AA}} = 0.79$ [see *Mikolajewicz and Maier-Reimer*, 1991]) and the anticorrelation between the strength of NADW formation and outflow from the Atlantic ($r_{\text{CEN-NA:ME15-A}} = -0.58$; see Figure 1). When NADW formation is increased, the outflow from the Atlantic at 30°S is also increased, as is indicated by the opposite signs of the loading on NADW formation rate and the loading on stream function at both 1.5- and 2.5-km depth in the Atlantic (note that negative values denote flow to the south in the sign convention used for stream function). EOF 1 also suggests that decreased mass transport through the Drake Passage is linked with decreased northward transport of mass in the Pacific and Indian Oceans ($r_{\text{PSIDRA:ME15-P}} = 0.73$ and $r_{\text{PSIDRA:ME15-I}} = 0.88$).

EOF 2 (Figure 9a) shows several features that are the inverse of EOF 1. It describes reduced NADW formation, with consequent reduction of outflow from the Atlantic (i.e., smaller negative values for stream function at 1.5- and 2.5-km depth in the Atlantic), increased AABW formation (with attendant increases in net heat flux from the Southern Ocean to the atmosphere, i.e., larger net heat flux values; see Figure 1), and increased mass transport through the Drake Passage. However, EOF 2 differs in other respects from the mirrored EOF 1, otherwise it would not be a separate EOF. In particular, it shows no anticorrelation between heat fluxes and convection in the North Atlantic, and an out-of-phase relationship between AABW formation rate and deep inflow into the Pacific.

EOFs 1 and 2 describe an oscillating pair of modes, as is evident from their corresponding amplitude time series (Figure 10a). This system is very similar to the oscillator identified by *Mikolajewicz and Maier-Reimer* [1991, Figure 7] in their POP analysis of vertical salinity profiles in the Atlantic. The present representation provides more informa-

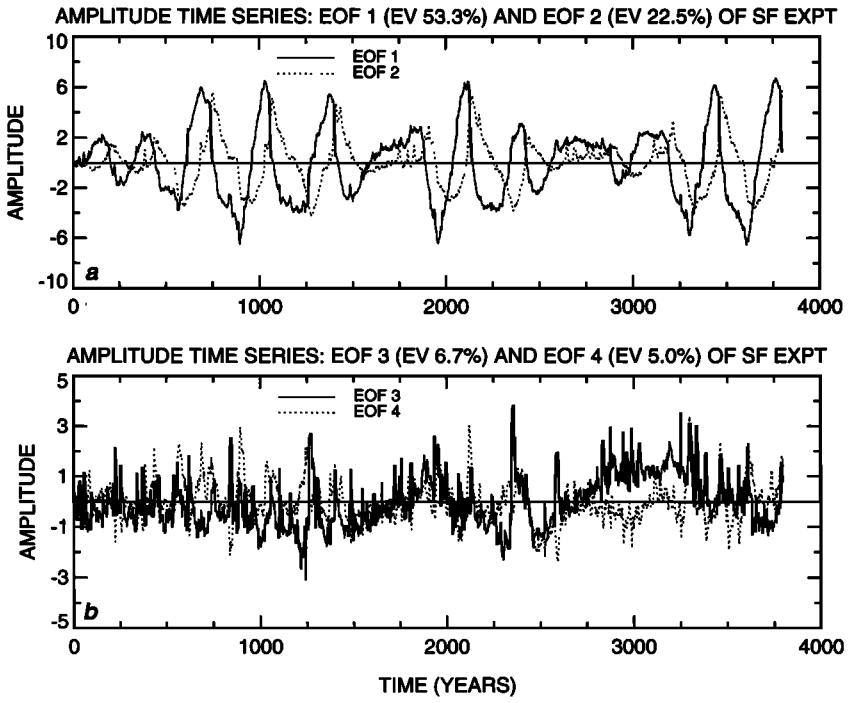

Figure 10. EOF amplitude time series 1–4 from the multivariate EOF analysis of the SF experiment. The time series give the amplitude of the EOF 1–4 patterns (see Figure 9) in the normalized multivariate ocean data. (a) Time series for EOFs 1 and 2 oscillate on timescales of several centuries, with maximum power at approximately 320 years. (b) Time series for EOFs 3 and 4 show fluctuations on the 1000-year timescale.

tion about the interrelationships between different integral ocean variables, while sacrificing spatial details.

EOF 3 (Figure 9b) principally describes simultaneous increases in NADW and AABW formation, with attendant increases in heat fluxes from the Southern Ocean and North Atlantic to the atmosphere (i.e., larger negative values). This mode has a weak indication of a spectral peak at a period slightly larger than 1000 years (results not shown). The long-period part of the corresponding amplitude time series (Figure 10b) shows some relation to the time series of POP coefficient amplitude shown by *Mikolajewicz and Maier-Reimer* [1991] and may be related to "on" and "off" modes of the salinity oscillator. EOF 4 is a mode in which there are anticorrelations between convective activity in the North Atlantic and Antarctic and between meridional transport in the Pacific and Indian Oceans. As in EOF 3, there is some evidence of power on the 1000-year time scale.

In summary, we conclude that EOFs 1 and 2 capture basic features of the salinity oscillator identified by MM90 and that some long-timescale variations of EOFs 3 and 4 may be related to the "on" and "off" modes of this oscillator.

SALL experiment. In order to compare the dominant modes of ocean variability in the two stochastic forcing experiments, we also computed the multivariate EOFs of the SALL integration (Figure 11). In both the SF and SALL experiments, the EOF 1 mode describes a reduction in the strength of the ACC, reduced flow from the North Atlantic into the Southern Ocean, and reduced flow from the Southern Ocean into the Pacific and Indian Oceans (compare Figures 9 and 11). The SF and SALL EOF 1 loading patterns are highly similar ($r = 0.84$). In SALL, however, EOFs 1 and 2 do not show an anticorrelation in NADW and AABW formation rates, which was a prominent characteristic of the salinity oscillator in the SF integration. EOFs 2–4 are quite different in the two runs.

The first four SF and SALL amplitude time series shown in Figures 10 and 12 illustrate some of the differences in the timescales of ocean variability in the two experiments. Although in both cases the low-frequency parts of the EOF 1 and 2 amplitude time series describe an oscillatory system, this is less clear in the SALL experiment than in the SF experiment. The spectra of the first four SALL amplitude time series (not shown) are red with discrete peaks at approximately 500 years, indicating that the dominant modes of the SALL experiment describe the long-period fluctuations in deep-ocean temperature and salinity, strength of the ACC, etc. (see Figure 2). This is some 50% longer than the timescale of the EOF 1 and 2 modes in the SF experiment.

More detailed analyses, making full use of pattern information rather than only spatially integrated quantities, will be required in order to determine whether the dominant modes of variability in the two stochastic forcing experiments are related to the same or similar physical mechanisms.

SCENA and 2 × CO_2 experiments. We also computed EOFs from the 15 spatially averaged ocean time series in the 100-year SCENA experiment and in a 100-year experiment with step function doubling of atmospheric CO_2 (2 × CO_2 [see *Cubasch et al.*, 1992]). Here we show results for the 2 × CO_2 experiment only, but note that the SCENA results are very similar. Signal anomalies were defined according to definition 2 and were normalized by both the standard deviations of the SF and SALL time series ($s(x)$ in (2)). This facilitates the direct comparison of signal and noise because it has the effect of reducing those components of the signal that are small relative to the variability in the SF or SALL

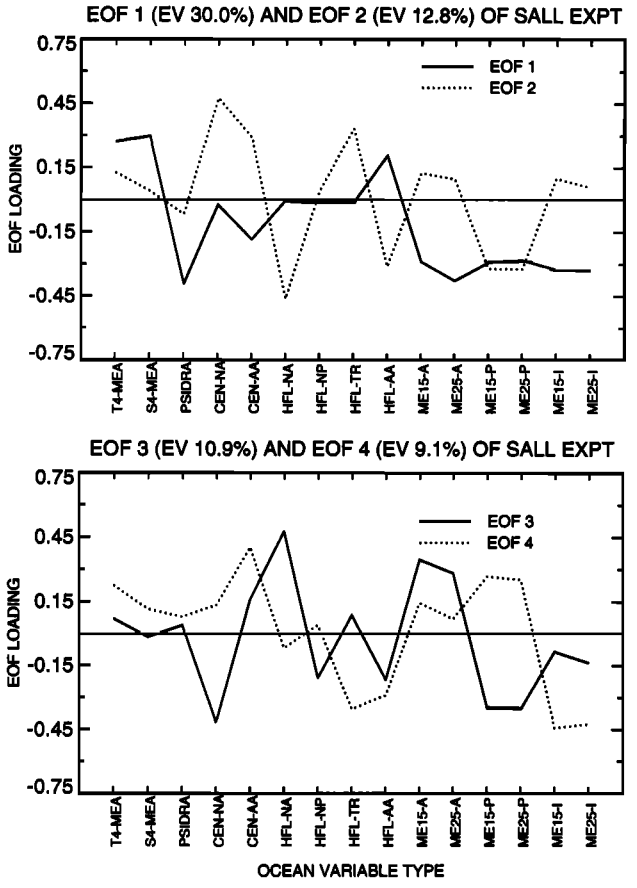

Figure 11. Multivariate EOFs 1–4 from the SALL experiment. EOFs were computed from the 8000-year time series for 15 spatially averaged ocean variables. The x axis indicates the ocean variable (see Table 1 for an explanation of the abbreviations). The dominant modes of variability are very different in the SF and SALL stochastic forcing experiments (compare Figure 9).

experiments. EOFs were then computed as described in section 5.1.

Since the variability in the SF and SALL experiments is very different, normalization of signal anomalies by the SF standard deviations series yields an EOF 1 pattern that is only weakly similar ($r = 0.57$) to the EOF 1 pattern obtained using SALL $s(x)$ values (see results for "unrotated fingerprint" in Figures 13a, b). In each case, however, the first mode is dominant, explaining ≥70% of the variance. This mode basically describes the change in mean state.

For the signal normalized with respect to the variability in the SF experiment (Figure 13a), the most pronounced feature of EOF 1 is the high negative loading on NADW formation rate, with attendant positive loadings on the heat fluxes from the ocean to the atmosphere in the North Atlantic. Note that although there are large reductions in the rate of AABW formation in both the $2 \times CO_2$ and SCENA experiments (by approximately 45×10^{13}W and 70×10^{13}W, reductions of nearly 20% and 30%, respectively, relative to the CTL mean state; see Figures 3 and 4 for SCENA results), these changes are not important in EOF 1 because they are small relative to the large variability of this quantity in the SF experiment (Table 1).

Normalization by the standard deviations from the SALL experiment yields an EOF 1 pattern with a more uniform distribution of loadings over all indices (compare Figures 13a and 13b). This mode reflects the reduced formation rate of both NADW and AABW, with attendant increases in heat fluxes from the ocean to the atmosphere in the North Atlantic and Antarctic, reduction in the strength of the ACC, and a decrease in the outflow of NADW from the Atlantic into the Southern Ocean (recall that southward transport is negatively defined, so positive values of stream function at 1.5- and 2.5-km depth in the Atlantic indicate smaller negative values and decreased southward transport). The strongest signature is the decreased transport from the Southern Ocean into the Pacific and Indian Oceans.

Mikolajewicz et al. [1990] have also used the LSG OGCM in two uncoupled "pseudo-transient" greenhouse warming experiments. In these integrations the ocean model was forced by equilibrium patterns of surface temperature change resulting from $2 \times CO_2$ runs performed with AGCMs with mixed-layer oceans. The time dependence was achieved by scaling these patterns using an exponential function and time constants of 40 and 80 years. The dominant EOF 1 modes computed from these experiments (not shown) are generally very similar to those obtained in the doubled-CO_2 and SCENA integrations.

5.3. Projections in Two-Dimensional EOF Space

Before proceeding to the calculation of detection periods and times for multivariate ocean data, it is useful to first visualize the evolution of the multivariate SCENA signal in the space of the SF and SALL EOFs (following *Preisendorfer* [1988] and *Santer et al.* [1994]). For this purpose we compute anomalies (according to definition 2) of the SCENA ocean data, normalize with both the SF and SALL noise as in (2), and then project the normalized anomalies onto the EOFs of the SF or SALL experiment, $e_j(x)$, to form the time series $b_j(t)$:

$$b_j(t) = \sum_{x=1}^{p} \hat{y}(x, t) e_j(x) \qquad j = 1, p; \quad t = 1, 100 \qquad (7)$$

where $\hat{y}(x, t)$ are the normalized anomalies from the SCENA experiment.

Plate 4a shows the projection of the normalized SF and SCENA anomaly data onto EOFs 1 and 2 of the SF experiment. This is a way of comparing the joint evolution of the multivariate signal and noise data in time and space [see *Preisendorfer*, 1988; *Santer et al.*, 1994]. Each symbol on the figure represents 1 year of the SCENA or SF experiment, and the symbols in consecutive years are linked by lines. In the SF experiment, these lines define the trajectory of the 320-year ocean circulation oscillator in the EOF 1–2 plane: the outer "loops" are traced when the oscillator is in "on" mode, and the trajectories collapse toward the center of the figure at times when the oscillator is in "off" mode. A similar picture was obtained by *Mikolajewicz and Maier-Reimer* [1991] in a scatterplot of the real and imaginary parts of the POP coefficients for their salinity oscillator.

If the multivariate ocean signal in the SCENA experiment showed no evidence of the types of interrelationships (between different ocean variables and different ocean basins) found in EOFs 1 and 2 of the SF integration, its projections on these EOFs would be close to zero.

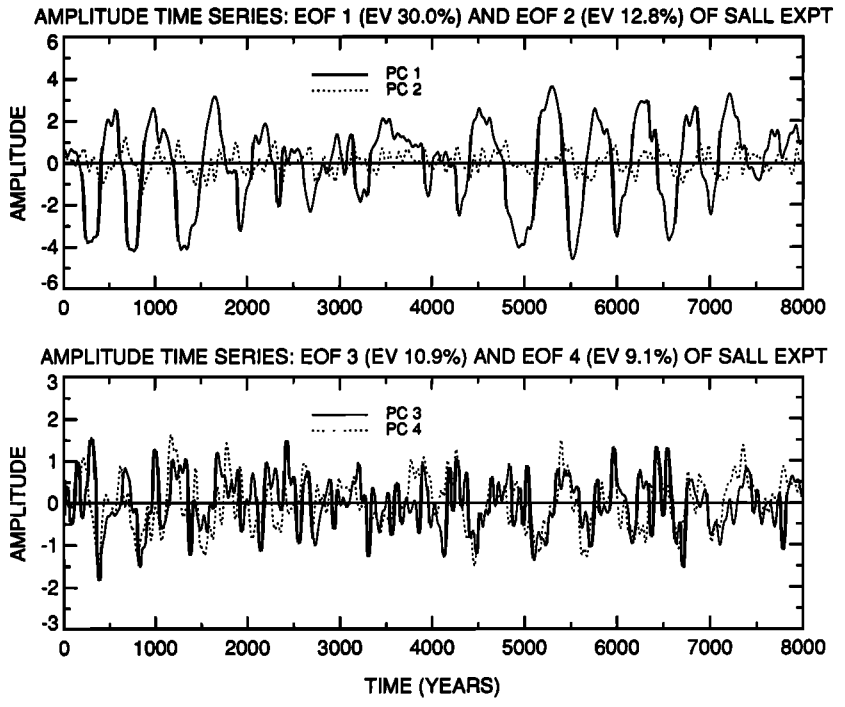

Figure 12. EOF amplitude time series 1–4 from the multivariate EOF analysis of the SALL experiment. The time series give the amplitude of the EOF 1–4 patterns (see Figure 11) in the normalized multivariate ocean data. Amplitudes for EOFs 1–4 oscillate on timescales of approximately 500 years. All time series have been low-pass filtered.

Clearly, this is not the case: the SCENA data have small but nonzero projections on both EOFs 1 and 2 (Plate 4a). This is not surprising, since EOFs 1 and 2 describe fundamental features of the ocean circulation (such as the coupling between NADW and AABW formation rates and ocean-atmosphere heat fluxes) which exist also in the greenhouse warming experiment. We conclude therefore that the SCENA signal and the SF variability have some common components, although pattern correlations between EOF 1 of SCENA and EOFs 1 and 2 of the SF experiment show that these common components are very small ($r_{SFe_1:SCENAe_1} = 0.16$; $r_{SFe_2:SCENAe_1} = 0.19$). The fact that the signal does not emerge from the "noise cloud" is due to both the weakness of the similarity between the dominant signal and noise patterns and the fact that the amplitude of changes in common components is much larger in the SF experiment than in the SCENA integration.

When projected onto EOFs 3 and 4 of the SF experiment, the SCENA data show much less overlap with the natural variability noise cloud (Plate 4b). Most of the separation is in the plane of EOF 3, where the scenario A projections are generally larger than the projections of the SF data. This is principally due to the fact that the NADW and AABW changes are in the same direction and are important components of EOF 3 in the SF experiment (see Figure 9b) and are also a prominent feature of the ocean signal in the SCENA experiment ($r_{SFe_3:SCENAe_1} = 0.69$). Thus the SCENA ocean signal should be easier to discriminate from the natural variability noise in the EOF 3 plane of the SF experiment, where the amplitude of the signal is larger than the amplitude of the noise.

A rather different result is obtained when the SCENA ocean data (normalized by the SALL noise) are projected onto EOFs 1 and 2 of the SALL experiment (Plate 5). Both signal and noise amplitude time series were low-pass filtered in order to better display the oscillatory behavior of the SALL integration in the EOF 1–2 plane and the development of the SCENA signal. The signal clearly emerges from the natural variability noise cloud after 50 years, with virtually all of the separation occurring in the EOF 1 plane. This suggests that unlike the SF case, the EOF 1 plane of the SALL integration is a direction in which the signal not only can be well represented ($r_{SALLe_1:SCENAe_1} = 0.69$), but also is considerably larger than the noise.

If signal and noise are unfiltered (not shown) the high-frequency noise is large and completely obscures the oscillatory behavior of the SALL integration. Even in the unfiltered data, however, there is clear separation of signal and noise after roughly 55–60 years.

5.4. Estimates of Detection Period and Time: Unrotated Fingerprint

The detection periods and times given in section 4 were univariate estimates computed for individual ocean circulation indices. In this section we consider whether earlier detection of an ocean greenhouse warming signal can be achieved using a multivariate approach.

In any multivariate approach one must generally attempt to filter out those signal components that are contaminated by natural variability noise. There are several sources of noise. First, for any individual climate variable, the "signal" in a single transient greenhouse warming experiment is not "pure" signal but rather signal plus some manifestation of the model's own natural variability noise. As was shown by

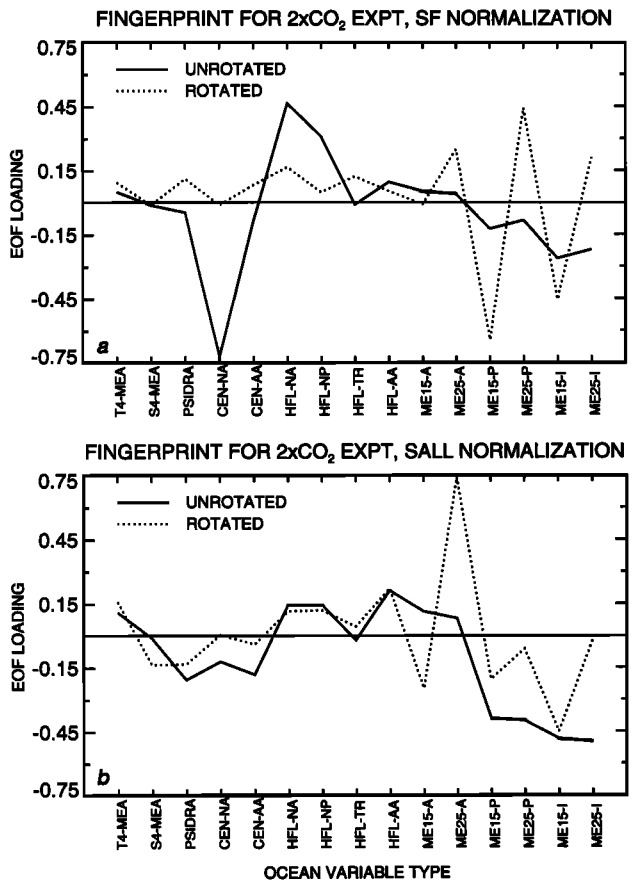

Figure 13. Unrotated and rotated fingerprint patterns, $f(x)$ and $f^*(x)$, for multivariate ocean data from a step function CO_2 doubling experiment. The x axis indicates the ocean variable (see Table 1 for an explanation of the abbreviations). The unrotated fingerprint is EOF 1 of the $2 \times CO_2$ integration, with normalization of ocean anomaly time series by the (a) SF and (b) SALL standard deviations. The unrotated fingerprint provides information about the direction of the expected ocean signal. The rotated fingerprints are skewed in the direction of the low-noise components of the SF and SALL experiments. Differences in the fingerprint structures in Figures 13a and 13b are attributable to differences in the variability in the SF and SALL experiments. For more detailed definitions of unrotated and rotated fingerprints, see sections 5.4 and 5.5, respectively.

Cubasch et al. [1994] (see also section 4.4), this noise can be reduced by averaging (in space and time) over an ensemble of initial condition realizations but will not be removed entirely. Second, the addition of each new variable to a multivariate detection vector adds finite amounts of both signal and noise. Whether the overall signal-to-noise ratio is enhanced or reduced by each new variable included depends on the relative amounts of signal and noise added. To reduce the noise from both sources, it is therefore necessary to apply some form of filtering. Without filtering, the signal detection problem may be as difficult as finding a needle in a haystack.

One way of improving signal detectability is by using a "fingerprint," which defines the direction in which the GHG signal is expected to lie. The fingerprint incorporates any a priori information that we may have about the multivariate GHG signal. This information may be derived from model results, physical intuition, or some combination of the two. The key point is that the fingerprint must be derived from an independent experiment or an independent model, i.e., an experiment different from the one for which we are trying to detect a signal. This means that in our study, which deals with model data only, we require three separate experiments in order to implement a fingerprint detection strategy: one for defining the fingerprint, one for defining the signal we wish to detect, and one for defining the natural variability noise properties.

The fingerprint represents an intelligent guess as to the nature of the expected greenhouse warming signal in the ocean. We cannot know the exact direction of the signal, since so far we lack any independent information on the climatic changes induced in the ocean by an enhanced greenhouse effect. If our guess is a good guess and is closer to the pure signal than to the signal plus noise, then the fingerprint acts as a noise filter, and its application can enhance signal-to-noise ratios. Suitable model-derived choices for a GHG fingerprint might be the average (over space and time) of a number of realizations starting from different initial conditions [*Cubasch et al.*, 1994], the time-averaged result from a single equilibrium response experiment (this study; also *Santer et al.* [1995]), or the spatial pattern of century timescale linear trends from a single transient experiment [*Hegerl et al.*, 1995]. In all three cases the aim of averaging over time and/or space is to reduce noise and obtain a better fingerprint.

This initial fingerprint may be optimized in various ways. For example, modification of the fingerprint by rotating it in a direction where the signal-to-noise ratio is maximized may improve signal detectability [*Hasselmann*, 1979, 1993]. Spatial rotation of the fingerprint is equivalent to optimizing a pattern filter in such a way as to reduce the contributions of those signal components that are similar to components of the dominant natural variability noise patterns. In this section we compute detection periods and times using an initial unrotated fingerprint, while the following section presents results for a rotated, optimized fingerprint.

In our case the initial fingerprint is a (time-invariant) multivariate pattern filter that is applied to the time-evolving ocean response to transient GHG forcing and to the selected natural variability noise experiment. The pattern that we use is the multivariate EOF 1 pattern of the *Cubasch et al.* [1992] $2 \times CO_2$ experiment. There are several assumptions underlying the selection of this particular fingerprint. First, we assume that the pseudo-equilibrium ocean response in the $2 \times CO_2$ experiment is similar to the response towards the end of the SCENA and EIN transient simulations (*Santer et al.* [1994] have showed that this assumption is justifiable for near-surface air temperature). Second, we assume that the $2 \times CO_2$ experiment, with no forcing changes as a function of time, yields the best (in the sense of least noisy) estimate of the expected signal direction. The third assumption is that EOF 1 of the $2 \times CO_2$ experiment captures most of the information on the ocean GHG signal.

Once the fingerprint has been decided upon, the multivariate detection problem can be reduced to a univariate problem. The multivariate anomaly time series from the SF and SCENA experiments, normalized by the standard deviations of the SF time series, are projected onto the initial unrotated fingerprint pattern

SCENA MULTIVARIATE OCEAN DATA PROJECTED ON SF EOFS 1 & 2

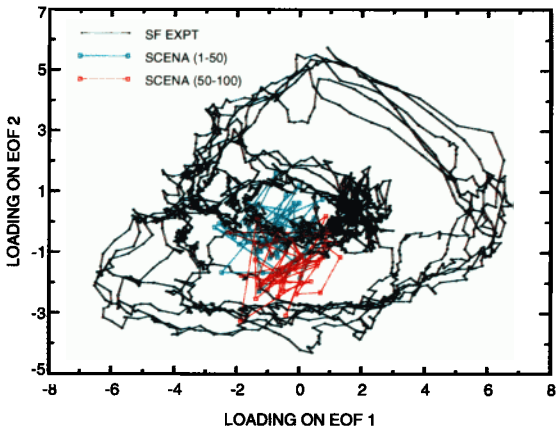

Plate 4a. Projection of normalized multivariate anomaly data from the SF and SCENA experiments onto multivariate EOFs 1 and 2 of the SF experiment. Each symbol on the figure represents 1 year of the SF or SCENA experiment; symbols in consecutive years are joined by lines. The first and last 50 years of the SCENA signal have been plotted in different colors to indicate the time evolution of the signal. In the SF experiment the outer "loops" define the trajectory traced by the 320-year ocean circulation oscillator in the EOF 1–2 plane. The SCENA multivariate ocean data has nonzero loadings on SF EOFs 1 and 2, indicating that both experiments have common components of variability in their overall general circulations. The fact that the signal does not emerge from the natural variability noise cloud indicates that the amplitude of changes in these common components is larger in the SF experiment than in the SCENA integration.

SCENA MULTIVARIATE OCEAN DATA PROJECTED ON SF EOFS 3 & 4

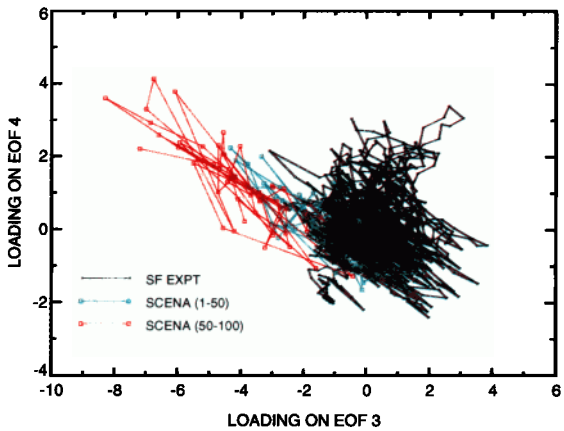

Plate 4b. As for Plate 4a, but for the projection onto multivariate EOFs 3 and 4 of the SF experiment. Note that the SCENA projections on EOF 3 are generally larger than the projections of the SF data. The signal clearly emerges from the natural variability noise cloud after 50 years.

$$u(t) = \sum_{x=1}^{p} \hat{z}(x, t) f(x) \qquad t = 1, \cdots, n \tag{8}$$

$$v(t) = \sum_{x=1}^{p} \hat{y}(x, t) f(x) \qquad t = 1, \cdots, 100 \tag{9}$$

where $\hat{z}(x, t)$ and $\hat{y}(x, t)$ are the normalized SF and SCENA anomalies, respectively, and $f(x)$ is the multivariate unrotated fingerprint pattern (EOF 1; see Figure 13a) derived from the $2 \times CO_2$ experiment. Note that in the following we operate primarily in the coordinate space of the SF variability but consider also results for the normalization of $\hat{z}(x, t)$ and $\hat{y}(x, t)$ by the SALL standard deviations.

As in the univariate case (sections 4.3 and 4.4), we can then compute the standard errors ε_i for $u(t)$, the actual signal trends β_i for $v(t)$, and a range of linear signal growth rates $\overline{\beta}$ based on the 100-year linear trend in $v(t)$. The signal-to-noise ratios $\overline{\beta}/\varepsilon_i$ and β_i/ε_i are again a function of the length of the time chunk or interval considered, L_i, with $L_i = 10i$, and $i = 1, 2, \cdots, 10$. Detection periods and times are then defined as in the univariate analysis. We also present detection time results for the EIN ocean signal (expressed relative to the decade 1985–1994) and for both definitions of the SCENA signal.

The multivariate analysis with an unrotated fingerprint reveals a sensitivity of detection period to the SF versus SALL ocean variability differences similar to that found in the univariate case. For all growth rates, detection periods are consistently shorter for the SALL noise (10–45 years) than for the SF noise (20–70 years; see Table 2). Detection periods for multivariate signals and noise are generally shorter than those obtained in the univariate analysis, at least for growth rates of $0.5\overline{\beta}$ and $1.0\overline{\beta}$ (and excluding again univariate results for North Atlantic ice volume because of drift problems).

Differences in the detection periods are due solely to the SF versus SALL differences in variability and in the fingerprint patterns onto which the multivariate noise is projected (Figure 13). SF versus SALL differences in detection time, however, are more difficult to interpret, since they incorporate the additional effect of different signal evolution rates, i.e., the evolution rates for the multivariate signal are not

SCENA MULTIVARIATE OCEAN DATA PROJECTED ON SALL EOFS 1 & 2

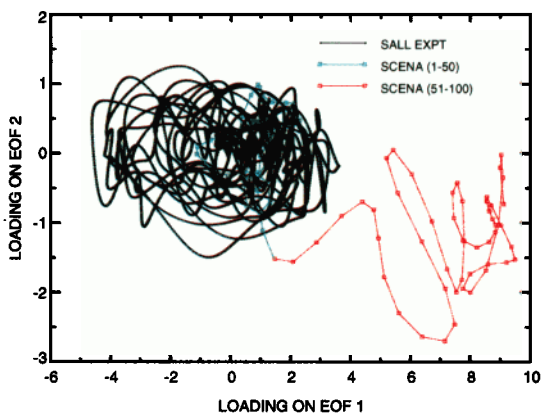

Plate 5. As for Plate 4a, but for the projection onto multivariate EOFs 1 and 2 of the SALL experiment. Both signal and noise data were low-pass filtered in order to better display the oscillatory behavior of the SALL integration and the development of the SCENA signal. The signal emerges from the natural variability noise cloud after 50 years, with most of the separation in the EOF 1 plane.

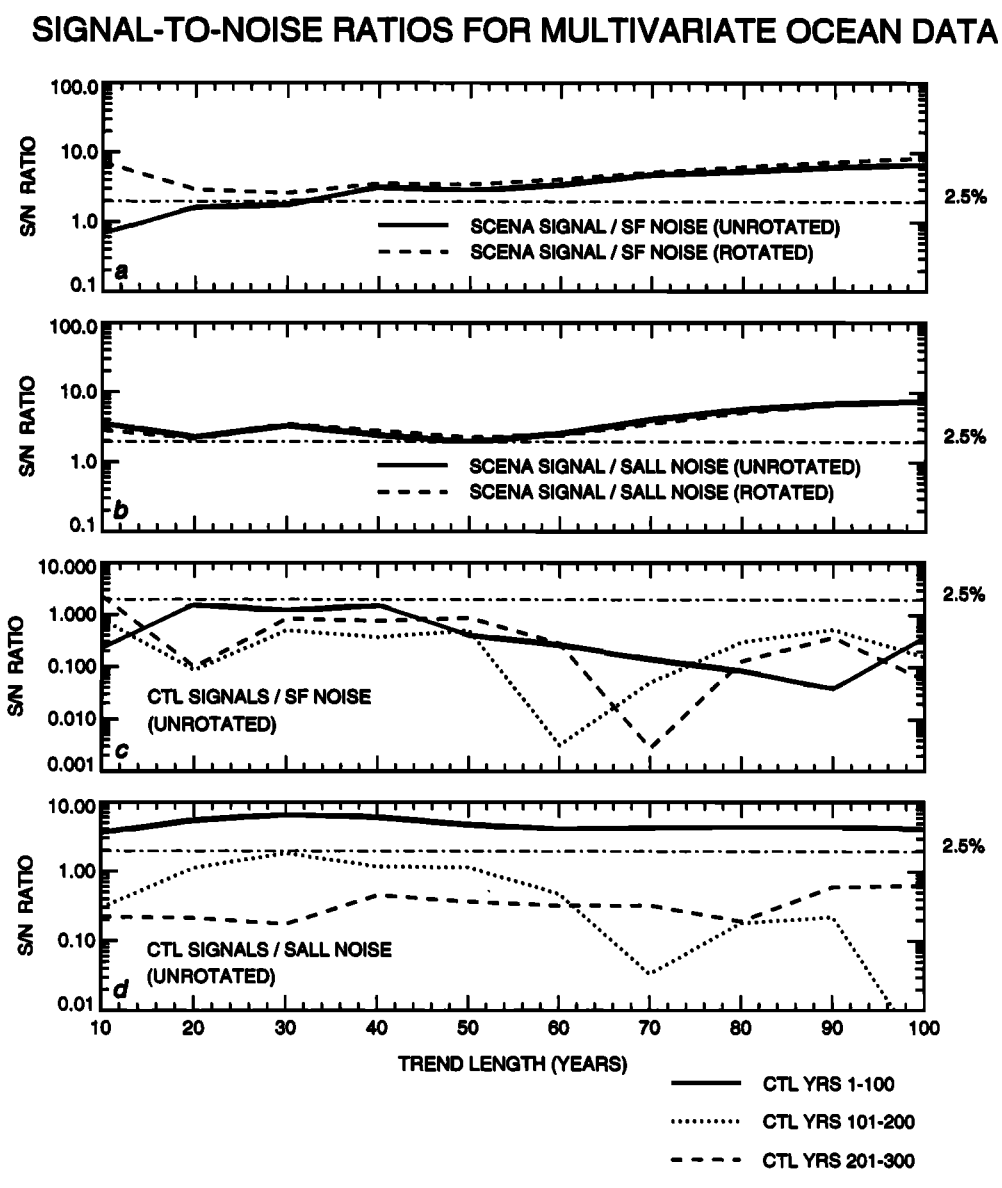

Figure 14. Signal-to-noise ratios β_i/ε_i for multivariate ocean data. The multivariate ocean signal from the SCENA experiment and the multivariate noise from the SF and SALL experiments were projected onto unrotated and rotated fingerprint patterns derived from a CO_2 doubling experiment (see Figure 13). The fingerprint patterns provide information about the direction of the expected greenhouse warming signal in the ocean. Rotation of the fingerprint leads to a substantial increase in β_i/ε_i in the case of the SF noise (Figure 14a) but produces only small changes in signal-to-noise ratio for the SALL noise (Figure 14b). If we regard the first 100 years of the coupled CTL experiment as the "signal" and project this signal onto the unrotated $2 \times CO_2$ fingerprint, the ocean circulation changes in the CTL would not be detectable relative to the SF noise (Figure 14c) but would be detectable after only 10 years relative to the SALL noise (Figure 14d). Linear trends in the second and third centuries of the CTL experiment are within the noise envelope defined by the SF and SALL experiments (Figures 14c and 14d).

identical in the coordinate spaces of the SF and SALL variability. In contrast to the detection period, there is no evidence that detection times are shorter if noise estimates are based on the SALL experiment, a result that must be due to differences in signal evolution (Table 3). Signal-to-noise ratios, however, tend to be higher using noise data from the SALL experiment (compare Figures 14a and 14b). The longer detection times for the SCENA signal relative to the SALL noise are due to a small signal evolution rate after 50 years, which causes the signal-to-noise ratio to dip just beneath the 5% significance threshold (Figure 14b).

In most cases, the multivariate analysis yields detection times that are shorter than or close to those obtained in the univariate analysis (Table 3). The exception is the SCENA signal and SALL noise, with detection times of 55–60 years. This result is due in part to the fact that the projection of the multivariate signal onto the unrotated fingerprint pattern has smaller signal-to-noise ratios than some of the individual components of the signal vector.

An interesting result is the reduction in detection time that is achieved by use of the EIN ocean signal rather than the SCENA signal. (It is appropriate to compare the EIN results

with the results for SCENA definition 1 in Table 3). For the unrotated fingerprint the reduction is from 35 to 30 years for the SF noise and from 60 years to 30 years for the SALL noise. One reason for the larger reduction in detection time for the SALL noise is that changes in Pacific and Indian Ocean transports are important components of the unrotated fingerprint (in the SALL coordinate space; see Figure 13b). The univariate analysis revealed that changes in these components are detectable 35–40 years earlier in the EIN experiment (Table 3). In contrast, the dominant component of the unrotated fingerprint in the SF coordinate space is the change in NADW formation (Figure 13a), which is not detectable earlier in the EIN experiment.

We can also ask whether the drift and/or low-frequency natural variability that characterizes the first 100 years of the CTL experiment would have been detectable relative to the SF and SALL natural variability noise. To address this question, we defined CTL anomalies according to definition 1. Anomalies were then normalized relative to the SF and SALL standard deviations and projected onto the $2 \times CO_2$ fingerprint patterns, as in (9). Detection times were then computed in the usual way. The ocean changes in the first 100 years of the CTL experiment (Figure 14c) are clearly below the threshold of detectability relative to the SF noise but would be detectable after only 10 years relative to the SALL standard errors (Figure 14d). The ocean changes in the second and third centuries of the CTL experiment are not significant, independent of the experiment chosen to estimate the noise.

5.5. Estimates of Detection Period and Time: Rotated Fingerprint

Finally, we consider whether we can further improve signal-to-noise ratios and achieve earlier detection by rotating the fingerprint in the direction of low-noise components of the two stochastic forcing experiments. We stress that the rotation takes into account only the spatial and not the temporal properties of the noise: we are not rotating the fingerprint in the direction of frequency bands where the noise has low energy [see *Hasselmann*, 1993].

The rotated fingerprint $f^*(x)$ is obtained by weighting the coefficients

$$\alpha_j = \sum_{x=1}^{p} f(x) e_j(x) \qquad j = 1, \cdots, p \tag{10}$$

for the representation of the $2 \times CO_2$ multivariate fingerprint $f(x)$ in the space of the EOFs $e_j(x)$ of the SF stochastic forcing experiment by the inverse of the noise eigenvalues l_j.

$$f^*(x) = \sum_{j=1}^{p} \alpha_j e_j(x) l_j^{-1} \qquad x = 1, \cdots, p \tag{11}$$

As in (8) and (9), we then project the multivariate anomaly time series from the SF and SCENA (or EIN) experiments onto the rotated fingerprint and use the resulting coefficient time series $u^*(t)$ and $v^*(t)$ to compute the standard errors and linear trends required to obtain the signal-to-noise ratio β_i/ε_i and detection time. We can also use $v^*(t)$ to derive the range of linear signal growth rates required for computing $\overline{\beta}/\varepsilon_i$ and detection period. We similarly compute $\overline{\beta}/\varepsilon_i$ and β_i/ε_i for signals, fingerprint, and noise expressed in the coordinate space of SALL variability.

The unrotated and rotated fingerprint patterns for the SF and SALL stochastic forcing experiments are shown in Figure 13. For the SF experiment the rotated and unrotated fingerprints are dissimilar ($r_{SFf:f^*} = 0.24$). The rotation is away from the direction of simultaneous changes in NADW formation and ocean-atmosphere heat fluxes in the North Atlantic (a high noise component) and toward changes in stream function in the Pacific and Indian Oceans. The most significant feature of the rotation is the change in sign of the deep Pacific and Indian inflow. The rotated fingerprint shows different signs for the inflows below 1500- and 2500-m depth. Thus the rotation leads to a detection pattern that monitors changes in the depth of the inflow (toward a deeper inflow) rather than changes in the absolute amount of the inflow.

In the SALL experiment (Figure 13b) the rotated and unrotated fingerprints show some similarity ($r_{SALLf:f^*} = 0.55$), suggesting that the fingerprint is already in a low-noise direction. The principal modification to the fingerprint is the higher weighting given to mass transport below 2.5 km in the Atlantic, again monitoring changes in the depth of the outflow. In the Pacific and Indian Oceans the weighting reduces the impact of deep inflow.

The detection periods for the optimized case are given in Table 2. Rotation of the fingerprint pattern yields significant decreases in detection period for the SF noise (from 20–70 years to 10–45 years) but no change or even a slight increase for the SALL noise (from 10–45 years to 10–55 years). This supports our result that in the coordinate space of the SALL variance, the unrotated fingerprint from the $2 \times CO_2$ experiment (see Plate 5 and Figure 13b) is already in a low-noise direction relative to the SALL noise. The large decreases in detection period for the SF noise are in part related to the differences in the eigenvalue spectra of the two experiments (Figure 15). Relative to the SALL experiment, the SF integration concentrates more of the total space-time variance in the first two modes and has much lower variance in the highest-numbered modes. The SALL integration has a much flatter eigenvalue spectrum. Rotation therefore has a greater effect in SF because of its larger range in l_j values as j increases. The fact that rotation of the fingerprint actually increases the detection period for one particular case ($0.5\overline{\beta}$, SALL noise) shows that rotation in space alone does not guarantee an improvement of the signal-to-noise ratio for a time-dependent signal.

Detection times for the rotated case are given in Table 3. For the SCENA signal, rotation substantially decreases detection times, both for the SF noise (from 35 years to 10 years) and for the SALL noise (from 55–60 years to 10–45 years). In the case of the SF noise this result is due to increases in β_i/ε_i at all chunk lengths (Figure 14a). For the SALL noise, rotation does not increase β_i/ε_i at all chunk lengths (Figure 14b), and the reduction in detection time is due to an increase of β_i/ε_i at a chunk length of 50 years. These results substantiate the conclusions from our analysis of detection period: rotation of the fingerprint tends to produce larger increases in signal-to-noise ratio for the SF noise than for the SALL noise.

For the EIN signal, optimization yields only a slight reduction in detection time for the case of SF noise (from 30 to 25 years) and has no impact in the case of SALL noise (30 years; Table 3). This illustrates again that for a time-evolving

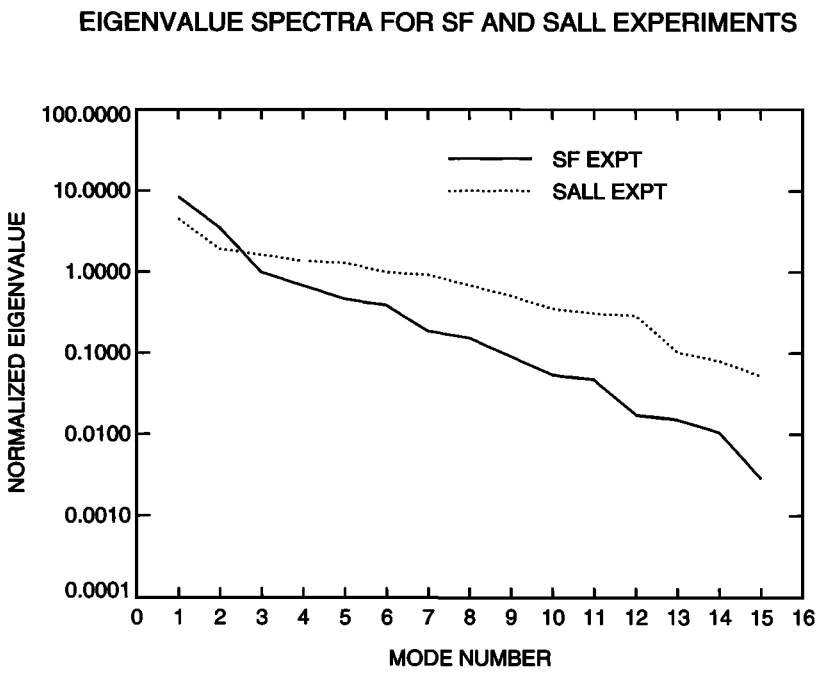

Figure 15. Eigenvalue spectra for multivariate EOFs from the SF and SALL experiments. EOFs were computed using normalized ocean data from the 3800-year SF integration and the 8000-year SALL integration. Note that the SF experiment concentrates more of the variance in the first two modes, while the SALL eigenvalue spectrum is much flatter.

signal, full space-time optimization will generally be needed to enhance the signal-to-noise ratios. The "optimized" detection time is longer for the EIN signal than for the SCENA signal in the case of the SF noise (25 years versus 10 years), but the reverse is true for the SALL noise (30 years versus 45 years; recall that the EIN results should be compared with definition 1 of the SCENA signal). As was noted in the preceding section, we suspect that the explanation for this result lies in the comparatively larger signals in the Pacific and Indian Ocean transport decreases in the EIN experiment. These components are all of the same sign in the rotated, SALL-normalized fingerprint (Figure 13b), and hence the EIN signal projects well onto this pattern, but they have different signs in the rotated, SF-normalized fingerprint, so that the EIN signal is distorted in this representation (Figure 13a).

5.6. Summary of Multivariate Detection Aspects

Our results suggest that a multivariate description of signal and noise and the use of a fingerprint pattern generally yields higher signal-to-noise ratios and shorter detection periods and times than a univariate approach, even when the fingerprint pattern is not optimized by rotation in the direction of low-noise components. Optimization of the signal-to-noise ratio by simple spatial rotation yields significant further improvements in detection period and time for the SF noise and SCENA signal but does not substantially improve the signal-to-noise ratios of the SCENA or EIN signals in the EOF coordinate space of the SALL noise. This implies that the unrotated fingerprint (when normalized by the SALL variability) is already in a low-noise direction. The EIN signal can be detected earlier than the SCENA signal in the unrotated case (independent of the experiment used to estimate the noise), but the benefit of the earlier start date is less clear in the rotated case. Finally, our results show that spatial rotation alone does not guarantee an improvement of the signal-to-noise ratio for a time-evolving signal. This can be achieved only with an optimization strategy that accounts for the full space-time characteristics of the signal, noise, and fingerprint.

6. Summary and Conclusions

In this study we have been concerned with the detection of oceanic greenhouse warming signals. The ocean responses we examined were the spatially averaged changes in such circulation indices as temperature, salinity, ice volume, heat, and fresh water fluxes; loss of potential energy by convection; and transports of mass. We currently lack the technology to obtain reliable measurements of these large-scale averages in the real world, so that the focus of this study was necessarily on methodology and the identification of sources of uncertainty in model estimates of ocean greenhouse warming signals and the noise levels of natural variability. However, the approach outlined here could have important practical applications for the proposed measurement of ocean basin acoustic travel times as indicators of climate change.

The analyzed ocean signals were taken from recent time-dependent greenhouse warming experiments [*Cubasch et al.*, 1992, 1994, 1995] with the Hamburg ECHAM-1/LSG coupled atmosphere-ocean GCM, in which the response of the climate system to the time-varying greenhouse gas increases specified in scenario A of the IPCC [*Houghton et al.*, 1990] was simulated. Two long simulations were carried out: SCENA, with greenhouse gas forcing from 1985 to 2085, and EIN, with forcing from 1935 to 2085. In addition, three 50-year so-called Monte Carlo simulations were made to study the impact of natural variability on the signal.

We first considered the signal-to-noise ratio of the linear trend β in an ocean signal from the greenhouse warming experiment to the standard error ε of the sampling distribution of linear trends due solely to internally generated variability of the coupled atmosphere-ocean system. To estimate ε, we used results from two long stochastic forcing experiments in which the uncoupled LSG ocean model was forced by noise superimposed on climatological boundary conditions.

In the first experiment, SF [*Mikolajewicz and Maier-Reimer*, 1990], the ocean model was forced by temporally white but spatially correlated monthly anomalies of fresh water fluxes, superimposed on prescribed climatological fresh water fluxes. In the second experiment, SALL [*Mikolajewicz and Maier-Reimer*, 1994], the stochastic component included heat fluxes and wind stress as well as fresh water fluxes. Another important difference was a reduction of the temperature feedback factor in this experiment by 2/5 (by use of a 5-month relaxation time constant instead of a 2-month constant for SF).

The ocean variability was significantly lower in the SALL experiment than in the SF experiment. For example, the variability in a key diagnostic of the ocean circulation, the transport of mass through the Drake Passage, was roughly a factor of 7 lower in the SALL experiment. There were also other qualitative differences between the two experiments. Variance spectra from the SF experiment were typically red for most circulation indices, with power increasing toward low frequencies for timescales up to several centuries, and a pronounced spectral peak at approximately 320 years, corresponding to the salinity oscillator identified by *Mikolajewicz and Maier-Reimer* [1990, 1991]. In the SALL experiment a discrete peak was found at 500 years for only a few circulation indices, and the spectra tended to be whiter: the noise on timescales less than 10–50 years was much larger in the SALL experiment than in the SF integration, while the reverse was true for variability on century time scales. The spectra for ocean diagnostics from a 300-year control run performed with a fully coupled atmosphere-ocean model [*Cubasch et al.*, 1992] were generally more similar to those of the SALL experiment in the frequency range where the integrations overlapped.

The differences in ocean variability in the SF and SALL experiments translate into large uncertainties in values of the standard error ε on timescales of 10–100 years. Values of ε for individual ocean indices in the two stochastic forcing experiments differ in either direction by up to an order of magnitude. Thus even if the space-time evolution of an ocean greenhouse warming signal were perfectly known, the uncertainties in standard error estimates result in large uncertainties in signal detectability.

To study the impact of such uncertainties on signal detectability, we introduced the concepts of detection period T_p and detection time T_d. The detection period T_p is defined as the length of a climate time series (a "chunk length") that must be available in order to detect a given linear trend in the presence of the model's natural climate variability (at some stipulated significance level). The detection period is defined in model years and is independent of reference time and the real time evolution of the signal. It is determined by the signal-to-noise ratio $\overline{\beta}/\varepsilon$ for some prescribed, time-invariant signal trend $\overline{\beta}$. We selected $\overline{\beta}$ to be the mean growth rate that yields the change in a given ocean variable at the end of the SCENA experiment.

In contrast, detection time T_d represents the actual time at which a particular time-dependent signal, with changing growth rate β, is detectable (again at some stipulated significance level). In the present case of simulated data, it depends on the experiment's start date. Since both the trend noise ε and the signal growth rate β are changing with time for a given global warming scenario, it is useful to first characterize the noise properties in terms of the detection period T_p for given $\overline{\beta}$ before considering the impact of changes in both on T_d. This enables one to determine whether an increase in signal-to-noise ratio is due primarily to changing noise properties or changes in the signal trend with increasing time.

We computed detection periods and times both for univariate data (individual circulation indices) and for multivariate signals and noise. For the univariate analysis we found strong sensitivity of T_p and T_d to the stochastic forcing experiment used to estimate the natural variability noise. The univariate detection period results indicate that for the SF noise, less than one third (19 of 63) of the assumed mean growth rates $\overline{\beta}$ were detectable at the 5% level within 100 years, while over two thirds (45 of 63) of the growth rates could be detected within 100 years for the SALL noise. Detection periods were consistently longer if the noise was estimated from the SF integration, with the exceptions of North Atlantic ice volume and strength of NADW formation (where the SALL experiment had greater variability on the 10- to 30-year timescale). Similarly, the univariate detection time results showed that the SCENA and EIN signals could be detected within 100 years in only 7 of 45 cases for the SF noise but in 31 of 45 cases for the SALL noise. In general, circulation indices that are highly sensitive to surface conditions, such as strength of NADW formation and ocean-atmosphere heat fluxes, tended to have shorter detection periods and times than changes in deep-ocean temperature and salinity.

We also showed that detection times are sensitive to uncertainties in the time evolution of a greenhouse warming signal. We illustrated signal uncertainties in three ways. First, we found that two alternative choices of the climate response signal, depending on different assumptions regarding the correlation of the variability in the coupled control run and the greenhouse warming experiment, had a large impact on univariate detection times. Second, by considering both the EIN and SCENA experiments (with start dates in 1935 and 1985, respectively), we were able to study the effect of the cold start error on univariate detection times. For most ocean variables, the EIN signals were detectable earlier than the SCENA signals, as expected, with the largest improvements for mass transport through the Drake Passage (70 years earlier) and transports in the Pacific and Indian Oceans (35–40 years earlier). Third, we examined the ocean signals in a suite of three 50-year greenhouse warming experiments with identical greenhouse gas forcing, but each starting from different initial conditions of the CTL integration, and thus with different manifestations of broadband natural variability superimposed on the true, underlying signal. Our results indicate that the "between realization" signal variability can be large for certain ocean circulation indices, so that a single greenhouse warming integration of

limited duration may yield a misleading estimate of detection time.

We also investigated whether it was possible to achieve shorter detection times for an ocean greenhouse warming signal by considering the time evolution of a multivariate detection vector rather than of individual variables. The multivariate analysis was carried out in the EOF space of the spatially averaged ocean circulation indices in the two stochastic forcing experiments. The EOFs carry some spatial information (at the scale of entire ocean basins), but, more importantly, convey information about the relationships between different ocean variables. The first two SF EOFs capture basic features of the "salinity oscillator" identified by *Mikolajewicz and Maier-Reimer* [1990] and clearly show that this mode is multivariate and global in scale. The dominant modes of variability of the SALL experiment have a somewhat different structure. In particular, the SALL EOF 1 vector is in a direction in which the multivariate ocean signal from SCENA can be reasonably well represented and is large relative to the low-frequency noise (see Plate 5).

Projection of the multivariate ocean data from the SCENA, SF, and SALL experiments onto an unrotated "fingerprint" pattern (obtained from an independent CO_2-doubling experiment) yielded detection times as short as 35–60 years for the SCENA and EIN signals. Our results suggest that a multivariate approach can yield higher signal-to-noise ratios and shorter detection periods and times than a univariate approach.

The signal-to-noise ratio can be further enhanced by rotating the fingerprint in the direction of low-noise components. With spatial rotation of the fingerprint, which is equivalent to a form of pattern filtering, we achieved detection times of only 10 years for the SCENA signal (according to our preferred definition 2), independent of the experiment used to estimate the noise. However, rotation of the fingerprint did not substantially improve signal-to-noise ratios for the SCENA or EIN signals in the coordinate space of the SALL noise. This suggests that the unrotated fingerprint (when normalized by the SALL variability) is already in a low-noise direction (see Plate 5). The EIN signal can be detected earlier than the SCENA signal in the nonoptimized case, independent of the experiment used to estimate the noise, but the benefit of the earlier start date is less clear in the optimized case.

A few examples in which spatial rotation of the fingerprint actually slightly increased the detection time demonstrate that spatial rotation alone does not guarantee an improvement of the signal-to-noise ratio for a time-evolving signal. A fully consistent optimization strategy requires rotation in space and time [*Hasselmann*, 1993]. Work on this general approach is currently in progress.

Our analysis highlights the need for improved estimates of long-term ocean variability. At present, it is difficult to obtain reliable estimates of century timescale ocean noise from paleoclimatic data [*Santer et al.*, 1993]. We must therefore rely on model simulations for such information. The SF and SALL stochastic forcing experiments indicate that the model-derived estimates of decade to century timescale noise are highly sensitive to the precise specification of the forcing. It is not clear at this time whether this sensitivity is due primarily to changes in the mean circulation, the mixed boundary condition, the amplitude, temporal coherence, or spatial pattern of the forcing, or whether it is due to the fact that the stochasticity in the SALL experiment applies to all forcing terms rather than fresh water fluxes alone. Further experiments, covering a range of amplitude values and space-time correlation scales for the forcing terms and including different flux feedback parameterizations, are required in order to better understand such sensitivity [*Mikolajewicz and Maier-Reimer*, 1991, 1994; *Barnett et al.*, 1993; *Power and Kleeman*, 1994].

Differences between the variability in fully coupled and uncoupled integrations constitute another source of uncertainty in model-based estimates of low-frequency noise. It is evident from the spectra of the coupled model control run (CTL) that the incorporation of full atmospheric feedback may significantly modify the spatiotemporal ocean variability simulated in idealized stochastic forcing experiments which neglect such feedback. Further sources of "noise uncertainty," which we have not discussed here, include possible model dependence of results, specifically the sensitivity to different resolutions [*Covey*, 1992] or parameterizations [*Zebiak and Cane*, 1991].

The reduction of such noise uncertainties will require a full program of numerical experimentation, as well as a concerted effort to construct a better paleoclimatic data base for "constraining" model noise estimates. The stochastic forcing experiments imply phase relationships between the variability displayed in different ocean basins, the North Atlantic and Antarctic, etc. Potentially, these relationships should be testable given appropriate paleoclimate data [*Crowley and Kim*, 1993].

Such noise validation studies will face a number of practical problems. One problem relates to deficiencies in the paleodata themselves. These include poor spatial coverage, dating uncertainties, possible modification of the low-frequency characteristics of the reconstructed climate variable by statistical manipulation of the data, and difficulties in extracting a climate signal from the noise introduced by nonclimatic factors. A further problem might be termed "scale incompatibility." Current global climate models perform poorly at detailed regional scales (i.e., 500 km or less); yet it is precisely these scales that paleodata represent best. Finally, validation exercises must at present compare apples and oranges: model noise represents unforced variability only, whereas low-frequency variability reconstructed from paleodata reflects the variance associated with a complex mixture of external forcing factors (solar, volcanic aerosols, etc.) and the internally generated variability of the coupled atmosphere-ocean system. A null result (little or no similarity between simulated and "observed" variability) is thus difficult to interpret unambiguously. Despite these formidable difficulties, we have no other choice but to use quality-controlled paleodata sets in order to build confidence in the low-frequency noise simulated by climate models.

In conclusion, we note that for an optimally rotated fingerprint pattern we estimated a detection time of 10 years for the SCENA ocean signal (at a 5% confidence level) relative to a starting date of 1985. If our SCENA experiment were taken literally, this would be today. Although the ocean circulation indices we considered have not been monitored in the past and cannot be readily monitored in the future, our results may have relevance for ocean acoustic travel times, which can be monitored [*Munk and Forbes*, 1989]. Our own findings, and those of *Mikolajewicz et al.* [1993], suggest that

a monitoring time of 1 or 2 decades may be sufficient to detect an enhanced greenhouse effect signal in acoustic travel times. The results of these studies are based on model signals and model noise. Their conclusions are burdened by very severe caveats regarding the realism of both the model-estimated decade to century timescale natural oceanic variability and the estimate of the true ocean signal from a limited number of realizations.

One further point should be mentioned in discussing the relevance of our purely model-based results for detection studies involving observed data. Our work has dealt with model signals in response to radiative forcing by CO_2 alone. The radiative effects of greenhouse gases other than CO_2 were considered together as equivalent CO_2 rather than being explicitly resolved. However, some recent evidence suggests that these gases may have climate signatures different from that of CO_2 [*Wang et al.*, 1991]. Furthermore, it is likely that emissions of anthropogenic sulfate aerosols have influenced regional or even global climate during this century [e.g., *Wigley*, 1989; *Charlson et al.*, 1992; *Karl et al.*, 1995]. For example, recent AGCM experiments show that the surface temperature response to combined forcing by CO_2 and sulfate aerosols is very different from the response to changes in CO_2 only [*Taylor and Penner*, 1994; *Roeckner et al.*, 1995; *Mitchell et al.*, 1995]. Thus it is likely that we have been searching for a "suboptimal" model-predicted signal, i.e., a signal that lacks important anthropogenic forcing components. This provides yet another reason for exercising caution in transferring the detection times estimated here to the real world.

We note, however, that model experiments that incorporate sulfate aerosols and greenhouse gases other than CO_2 would not necessarily yield signals with longer detection times in the observed data than "CO_2-only" signals. While longer detection times would be the likely result in univariate approaches dealing with changes in global averages (since sulfate-induced cooling compensates some of the CO_2-induced warming, yielding a smaller net temperature increase), shorter detection times are possible in multivariate approaches. This has been shown for surface air temperature data by *Santer et al.* [1995], who compared model-predicted CO_2-only and combined CO_2-sulfate aerosol signal patterns with observed data and found consistently higher signal-to-noise ratios for the latter case. By including other, non-CO_2 anthropogenic forcings, we may be able to reduce some of the regional-scale background noise against which we are attempting to detect a CO_2-only signal, thereby enhancing signal-to-noise ratios and achieving shorter detection times. Incorporating the effects of other anthropogenic forcing factors in a physically realistic way, and in a way that accounts for uncertainties in their space-time evolution, remains a challenge for the future.

Acknowledgments. This work was performed under the auspices of the U.S. Department of Energy, Environmental Sciences Division, by the Lawrence Livermore National Laboratory under contract W-7405-ENG-48. It was also sponsored by the Bundesministerium für Forschung und Technologie, the Commission of the European Community, the Max-Planck-Gesellschaft, the Freie and Hansestadt Hamburg, and the Deutsche Forschungsgemeinschaft (through Sonderforschungsbereich 318). The authors would like to thank the staff of the Deutsches Klimarechenzentrum and the Max-Planck-Institut für Meteorologie for their support, particularly Eduardo Zorita and Jin-Song Xu for supplying data from the coupled control run. Tim Barnett, Larry Gates, Jochen Segschneider, Sailes Sengupta, Ron Stouffer, and Karl Taylor provided many valuable suggestions and constructive criticisms.

References

Barnett, T. P., Detection of changes in the global tropospheric temperature field induced by greenhouse gases, *J. Geophys. Res.*, *91*, 6659–6667, 1986.

Barnett, T. P., and M. E. Schlesinger, Detecting changes in global climate induced by greenhouse gases, *J. Geophys. Res.*, *92*, 14,772–14,780, 1987.

Barnett, T. P., M. Chu, R. Wilde, and U. Mikolajewicz, Low frequency ocean variability induced by stochastic forcing of various colors, in *Fourth Symposium on Global Change Studies*, pp. 29–31, American Meteorological Society, Boston, Mass., 1993.

Bloomfield, P., and D. Nychka, Climate spectra and detecting climate change, *Clim. Change*, *21*, 275–287, 1992.

Bradley, R. S., *Quaternary Paleoclimatology*, 472 pp., Unwin Hyman, Boston, Mass., 1985.

Bradley, R. S., and P. D. Jones, *Climate Since A.D. 14.8*, 679 pp., Routledge, London, 1992.

Briffa, K. R., T. S. Bartholin, D. Eckstein, P. D. Jones, W. Karlén, F. H. Schweingruber, and P. Zetterberg, A 1,400-year tree-ring record of summer temperatures in Scandinavia, *Nature*, *346*, 434–439, 1990.

Briffa, K. R., P. D. Jones, T. S. Bartholin, D. Eckstein, F. H. Schweingruber, W. Karlén, P. Zetterberg, and M. Eronen, Fennoscandian summers from A.D. 500: Temperature changes on short and long timescales, *Clim. Dyn.*, *7*, 111–119, 1992.

Bryan, K., High latitude salinity effects and interhemispheric thermohaline circulations, *Nature*, *305*, 301–304, 1986.

Charlson, R. J., S. E. Schwartz, J. M. Hales, R. D. Cess, J. A. Coakley, J. E. Hansen, and D. J. Hofmann, Climate forcing by anthropogenic aerosols, *Science*, *255*, 423–430, 1992.

Covey, C., Behavior of an ocean general circulation model at four different horizontal resolutions, *PCMDI Rep. 4*, 29 pp., Lawrence Livermore Natl. Lab., Univ. of Calif., Livermore, 1992.

Crowley, T. J., and K.-Y. Kim, Towards development of a strategy for determining the origin of decadal-centennial scale climate variability, *Quat. Sci. Rev.*, *12*, 375–385, 1993.

Crowley, T. J., and G. R. North, *Paleoclimatology*, 339 pp., Oxford University Press, New York, 1991.

Cubasch, U., K. Hasselmann, H. Höck, E. Maier-Reimer, U. Mikolajewicz, B. D. Santer, and R. Sausen, Time-dependent greenhouse warming computations with a coupled ocean-atmosphere model, *Clim. Dyn.*, *8*, 55–69, 1992.

Cubasch, U., B. D. Santer, A. Hellbach, G. Hegerl, H. Höck, E. Maier-Reimer, U. Mikolajewicz, A. Stössel, and R. Voss, Monte Carlo climate change forecasts with a global coupled ocean-atmosphere model, *Clim. Dyn.*, *10*, 1–19, 1994.

Cubasch, U., G. Hegerl, A. Hellbach, H. Höck, U. Mikolajewicz, B. D. Santer, and R. Voss, A climate change simulation starting at an early time of industrialization, *Clim. Dyn.*, *11*, 71–84, 1995.

Dansgaard, W., S. J. Johnsen, N. Reeh, N. Gundestrup, H. B. Clausen, and C. U. Hammer, Climatic changes, Norsemen and modern man, *Nature*, *255*, 24–28, 1975.

Delworth, T., S. Manabe, and R. J. Stouffer, Interdecadal variability of the thermohaline circulation in a coupled ocean-atmosphere model, *J. Clim.*, *6*, 1993–2011, 1993.

Gates, W. L., AMIP: The Atmospheric Model Intercomparison Project, *Bull. Am. Meteorol. Soc.*, *73*, 1962–1970, 1992.

Hansen, J., A. Lacis, D. Rind, G. Russell, P. Stone, I. Fung, R. Ruedy, and J. Lerner, Climate sensitivity: Analysis of feedback mechanisms, in *Climate Processes and Climate Sensitivity*, *Geophys. Monogr. Ser.*, vol. 29, edited by J. E. Hansen and T. Takahashi, pp. 130–163, AGU, Washington D. C., 1984.

Hasselmann, K., Stochastic climate models, I, Theory, *Tellus*, *28*, 473–485, 1976.

Hasselmann, K., On the signal-to-noise problem in atmospheric response studies, in *Meteorology of Tropical Oceans*, edited by D. B. Shaw, pp. 251–259, Royal Meteorological Society, London, 1979.

Hasselmann, K., Optimal fingerprints for the detection of time dependent climate change, *J. Clim.*, *6*, 1957–1971, 1993.

Hasselmann, K., R. Sausen, E. Maier-Reimer, and R. Voss, On the cold start problem in transient simulations with coupled ocean-atmosphere models, *Clim. Dyn.*, *9*, 53–61, 1993.

Hegerl, G. C., H. von Storch, K. Hasselmann, B. D. Santer, U. Cubasch, and P. D. Jones, Detecting anthropogenic climate change with an optimal fingerprint method, *J. Clim.*, in press, 1995.

Hellerman, S., and R. Rosenstein, Normal monthly wind stress over the world ocean with error estimates, *J. Phys. Oceanogr.*, *13*, 1093–1104, 1983.

Houghton, J. T., G. J. Jenkins, and J. J. Ephraums (Eds.), *Climate Change: The IPCC Scientific Assessment*, 365 pp., Cambridge University Press, New York, 1990.

Karl, T. R., R. W. Knight, G. Kukla, and J. Gavin, Evidence for radiative effects of anthropogenic sulfate aerosols in the observed climate record, in *Aerosol Forcing of Climate*, edited by R. Charlson and J. Heintzenberg, John Wiley, New York, in press, 1995.

Kellogg, W. W., An apparent moratorium on the greenhouse warming due to the deep ocean, *Clim. Change*, *25*, 85–88, 1993.

Kim, K.-Y., and G. R. North, Surface temperature fluctuations in a stochastic climate model, *J. Geophys. Res.*, *96*, 18,573–18,580, 1991.

Lemke, P., Stochastic climate models, 3, Application to zonally-averaged energy models, *Tellus*, *29*, 385–392, 1977.

Lorenz, E. N., Irregularity: A fundamental property of the atmosphere, *Tellus*, *Ser. A*, *36*, 98–110, 1984.

Lorenz, E. N., Chaos, spontaneous climatic variations and detection of the greenhouse effect, in *Greenhouse-Gas-Induced Climatic Change: A Critical Appraisal of Simulations and Observations*, edited by M. E. Schlesinger, pp. 445–453, Elsevier, New York, 1991.

MacCracken, M. C., and H. Moses, The first detection of carbon dioxide effects: Workshop summary, 8–10 June 1981, Harpers Ferry, West Virginia, *Bull. Am. Meteorol. Soc.*, *63*, 1164–1178, 1982.

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

Maier-Reimer, E., U. Mikolajewicz, and K. Hasselmann, On the sensitivity of the global ocean circulation to changes in the surface heat flux forcing, *J. Phys. Oceanogr.*, *23*, 731–757, 1993.

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

Manabe, S., and R. J. Stouffer, Century-scale effects of increased atmospheric CO_2 on the ocean-atmosphere system, *Nature*, *364*, 215–218, 1993.

Manabe, S., R. J. Stouffer, M. J. Spelman, and K. Bryan, Transient responses of a coupled ocean-atmosphere model to gradual changes of atmospheric CO_2, I, Annual mean response, *J. Clim.*, *4*, 785–818, 1991.

Mikolajewicz, U., and E. Maier-Reimer, Internal secular variability in an ocean general circulation model, *Clim. Dyn.*, *4*, 145–156, 1990.

Mikolajewicz, U., and E. Maier-Reimer, One example of a natural mode of the ocean circulation in a stochastically forced ocean general circulation model, in *Strategies for Future Climate Research*, edited by M. Latif, pp. 287–318, Max-Planck-Institut für Meteorologie, Hamburg, Germany, 1991.

Mikolajewicz, U., and E. Maier-Reimer, Mixed boundary conditions in ocean general circulation models and their influence on the stability of the model's conveyor belt, *J. Geophys. Res.*, *99*, 22,633–22,644, 1994.

Mikolajewicz, U., B. D. Santer, and E. Maier-Reimer, Ocean response to greenhouse warming, *Nature*, *345*, 589–593, 1990.

Mikolajewicz, U., E. Maier-Reimer, and T. P. Barnett, Acoustic detection of greenhouse-induced climate changes in the presence of slow fluctuations of the thermohaline circulation, *J. Phys. Oceanogr.*, *23*, 1100–1109, 1993.

Mitchell, J. F. B., S. Manabe, T. Tokioka, and V. Meleshko, Equilibrium Climate Change, in *Climate Change: The IPCC Scientific Assessment*, edited by J. T. Houghton, G. J. Jenkins, and J. J. Ephraums, pp. 131–172, Cambridge University Press, New York, 1990.

Mitchell, J. F. B., R. A. Davis, W. J. Ingram, and C. A. Senior, On surface temperature, greenhouse gases and aerosols: Models and observations, *J. Clim.*, in press, 1995.

Munk, W., and A. M. G. Forbes, Global ocean warming: An acoustic measure?, *J. Phys. Oceanogr.*, *19*, 1765–1778, 1989.

Power, S. B., and R. Kleeman, Multiple equilibria in a global ocean general circulation model, *J. Phys. Oceanogr.*, *23*, 1670–1681, 1993.

Power, S. B., and R. Kleeman, Surface heat flux parameterization and the response of ocean general circulation models to high-latitude freshening, *Tellus*, *Ser. A*, *46*, 86–95, 1994.

Preisendorfer, R. W., *Principal Component Analysis in Meteorology and Oceanography*, *Develop. in Atmos. Sci.*, vol. 17, 425 pp., Elsevier, New York, 1988.

Roeckner, E., T. Siebert, and J. Feichter, Climatic response to anthropogenic sulfate forcing simulated with a general circulation model, in *Aerosol Forcing of Climate*, edited by R. Charlson and J. Heintzenberg, John Wiley, New York, in press, 1995.

Santer, B. D., T. M. L. Wigley, M. E. Schlesinger, and P. D. Jones, Multivariate methods for the detection of greenhouse-gas-induced climate change, in *Greenhouse-Gas-Induced Climatic Change: A Critical Appraisal of Simulations and Observations*, edited by M. E. Schlesinger, pp. 511–536, Elsevier, New York, 1991.

Santer, B. D., A. Berger, J. A. Eddy, H. Flohn, J. Imbrie, T. Litt, S. H. Schneider, F. H. Schweingruber, and M. Stuiver, How can paleodata be used in evaluating the forcing mechanisms responsible for past climate changes?, in *Dahlem Workshop on Global Changes in the Perspective of the Past*, edited by J. A. Eddy and H. Oeschger, pp. 343–367, John Wiley, New York, 1993.

Santer, B. D., W. Brüggemann, U. Cubasch, K. Hasselmann, H. Höck, E. Maier-Reimer, and U. Mikolajewicz, Signal-to-noise analysis of time-dependent greenhouse warming experiments, 1, Pattern analysis, *Clim. Dyn.*, *9*, 267–285, 1994.

Santer, B. D., K. E. Taylor, T. M. L. Wigley, J. E. Penner, P. D. Jones, and U. Cubasch, Towards the detection and attribution of an anthropogenic effect on climate, *Clim. Dyn.*, in press, 1995.

Sausen, R., K. Barthel, and K. Hasselmann, Coupled ocean-atmosphere models with flux corrections, *Clim. Dyn.*, *2*, 154–163, 1988.

Schlesinger, M. E., and J. F. B. Mitchell, Climate model simulations of the equilibrium climatic response to increased carbon dioxide, *Rev. Geophys.*, *25*, 760–798, 1987.

Stouffer, R. J., S. Manabe, and K. Bryan, Interhemispheric asymmetry in climate response to a gradual increase of atmospheric CO_2, *Nature*, *342*, 660–662, 1989.

Stouffer, R. J., S. Manabe, and K. Ya. Vinnikov, Model assessment of the role of natural variability in recent global warming, *Nature*, *367*, 634–636, 1994.

Taylor, K. E., and J. E. Penner, Response of the climate system to atmospheric aerosols and greenhouse gases, *Nature*, *369*, 734–737, 1994.

von Storch, H., T. Bruns, I. Fischer-Bruns, and K. Hasselmann, Principal oscillation pattern analysis of the 30- to 60-day oscillation in a general circulation model equatorial troposphere, *J. Geophys. Res.*, *93*, 11,022–11,036, 1988.

Wang, W.-C., M. P. Dudek, X.-Z. Liang, and J. T. Kiehl, Inadequacy of effective CO_2 as a proxy in simulating the greenhouse effect of other radiatively active gases, *Nature*, *350*, 573–577, 1991.

Washington, W. M., and G. A. Meehl, Seasonal cycle experiment on the climate sensitivity due to a doubling of CO_2 with an atmospheric general circulation model coupled to a simple mixed-layer ocean model, *J. Geophys. Res.*, *89*, 9475–9503, 1984.

Washington, W. M., and G. A. Meehl, Climate sensitivity due to increased CO_2: Experiments with a coupled atmosphere and ocean general circulation model, *Clim. Dyn.*, *4*, 1–38, 1989.

Watts, R. G., and M. C. Morantine, Is the greenhouse gas-climate signal hiding in the deep ocean?, *Clim. Change*, *18*, iii–vi, 1991.

Weaver, A. J., and E. S. Sarachik, Evidence for decadal variability in an ocean general circulation model: An advective mechanism, *Atmos. Ocean*, *21*, 1470–1493, 1991.

Weisse, R., U. Mikolajewicz, and E. Maier-Reimer, Decadal variability of the North Atlantic in an ocean general circulation model, *J. Geophys. Res.*, *99*, 12,411–12,421, 1994.

Wigley, T. M. L., Possible climate change due to SO_2-derived cloud condensation nuclei, *Nature*, *339*, 365–367, 1989.

Wigley, T. M. L., and T. P. Barnett, Detection of the greenhouse

effect in the observations, in *Climate Change: The IPCC Scientific Assessment*, edited by J. T. Houghton, G. J. Jenkins, and J. J. Ephraums, pp. 239–256, Cambridge University Press, New York, 1990.

Wigley, T. M. L., and P. D. Jones, Detecting CO_2-induced climatic change, *Nature*, *292*, 205–208, 1981.

Wigley, T. M. L., and S. C. B. Raper, Thermal expansion of sea water associated with global warming, *Nature*, *330*, 127–131, 1987.

Wigley, T. M. L., and S. C. B. Raper, Natural variability of the climate system and detection of the greenhouse effect, *Nature*, *344*, 324–327, 1990.

Wigley, T. M. L., and S. C. B. Raper, Internally-generated natural variability of global-mean temperatures, in *Greenhouse-Gas-Induced Climatic Change: A Critical Appraisal of Simulations and Observations*, edited by M. E. Schlesinger, pp. 471–482, Elsevier, New York, 1991.

Woodruff, S. D., R. J. Slutz, R. J. Jenne, and P. M. Steurer, A comprehensive ocean-atmosphere data set, *Bull. Am. Meteorol. Soc.*, *68*, 1239–1250, 1987.

Zebiak, S. E., and M. A. Cane, Natural climate variability in a coupled model, in *Greenhouse-Gas-Induced Climatic Change: A Critical Appraisal of Simulations and Observations*, edited by M. E. Schlesinger, pp. 457–469, Elsevier, New York, 1991.

W. Brüggemann, Institut für Logistik und Transport, Universität Hamburg, Von-Melle-Park 5, D-20146 Hamburg, Federal Republic of Germany.

U. Cubasch, Deutsches Klimarechenzentrum, Bundesstrasse 55, D-20146 Hamburg, Federal Republic of Germany.

K. Hasselmann, H. Höck, E. Maier-Reimer, and U. Mikolajewicz, Max-Planck-Institut für Meteorologie, Bundesstrasse 55, D-20146 Hamburg, Federal Republic of Germany.

B. D. Santer, Program for Climate Model Diagnosis and Intercomparison, Mail Stop L-264, Lawrence Livermore National Laboratory, P.O. Box 808, Livermore, CA 94550. (e-mail: bsanter@rainbow.llnl.gov)

T. M. L. Wigley, National Center for Atmospheric Research, Boulder, CO 80307.

(Received January 28, 1994; revised January 31, 1995; accepted February 21, 1995.)