A Simple Box Model of Stochastically Forced Thermohaline Flow

PAOLA CESSI

Scripps Institution of Oceanography, La Jolla, California and Instituto FISBAT-CNR, Bologna, Italy

(Manuscript received 17 May 1993, in final form 3 January 1994)

ABSTRACT

A modified Stommel two-box model is considered as a minimal representation of the buoyancy-driven ocean circulation. In the limit of fast temperature relaxation only the salinity evolves in time while the temperature is clamped to the prescribed ambient value. The box model has no *intrinsic* variability: just two linearly stable and one unstable equilibria. A finite perturbation is needed to shift the system from one stable equilibrium to the other. The minimum amplitude and duration in time of the perturbation are calculated.

A stochastic component of the freshwater flux forcing is then added to model the effect of changes in the global hydrological cycle due to the "weather." The stochastic forcing is a source of *extrinsic* time dependence. The salinity gradient obeys an equation analogous to the trajectory of a viscous particle in a double-welled potential, subject to Brownian agitation. If the amplitude of the stochastic driving is above a certain threshold, then there is a finite probability of switching from one stable equilibrium to the other. The threshold variance and the average residence time in each equilibrium are calculated. For timescales on the order of the average residence time or longer, the box model behaves like a random telegraph process.

The stochastic driving also induces a "rattling" around each steady equilibrium whose frequency is proportional to the curvature of the potential well at each equilibrium. The probability of being in each well can be calculated and, within each equilibrium, the box model behaves like an Ornstein–Uhlenbeck process.

Finally the spectrum of the salinity gradient is calculated analytically using standard approximations in stochastic processes. The approximate analytical results are in excellent agreement with those obtained by direct computation.

1. Introduction

The point of view of this work is that the thermohaline circulation in the ocean is the result of the interplay between the saline and thermal fluxes. Low surface temperatures in the polar regions, and high temperatures in the equatorial belt favor sinking at the poles and upwelling at the equator. This thermal forcing is opposed by the excess of precipitation and runoff over evaporation at high latitudes, which induces a freshwater flux opposing sinking near the poles.

Currently in the North Atlantic Ocean a deep thermally direct circulation prevails, but paleoclimatic records indicate that episodes of reduced circulation have occurred, the most recent being during the Younger Dryas (about 11 000 years BP) (Lehman and Keigwin 1992). One hypothesis is that the shutoff of the North Atlantic deep circulation was triggered by a pulse of meltwater that lasted about 1000 years (Fairbanks 1989). Numerical simulations of the thermohaline circulation also reveal an acute sensitivity of the deep transport, especially with respect to the freshwater flux (Weaver et al. 1993). A variety of models, ranging from complex coupled ocean–atmosphere simulations (Manabe and Stouffer 1988) to simple "box" models (Stommel 1961; Welander 1986), has shown that the competition between thermal and saline forcing results in multiple equilibria. In certain parameter ranges there are two stable solutions: one with strong and the other with weak (or even reversed) circulation, depending on differences in the initial conditions only.

A large body of work has shed much light on the issue of multiple *steady* equilibria [cf. the review article by Weaver and Hughes (1992)]. The question of *unsteady* thermohaline dynamics is a richer and more problematic one. It is useful to distinguish between *intrinsic* and *extrinsic* time dependence.

Intrinsic variability arises spontaneously, even when the external forcing is steady. A simple example is Welander's (1986) thermohaline loop. In certain parts of the parameter space all of the steady solutions become unstable and the model exhibits a limit cycle. A more complicated example is Winton and Sarachik's (1993) study of centennial variability in a GCM.[1]

Corresponding author address: Dr. Paola Cessi, Scripps Institution of Oceanography, University of California at San Diego, La Jolla, CA 92093-0230.

[1] Weaver and Sarachik (1991) document decadal variability in a different GCM. These relatively high-frequency variations eventually disappear and the model settles into a stable steady state, i.e., the variability is transitory. The Lorenz equations exhibit a similar phenomenology (metastable chaos).

Extrinsic time dependence is driven by unsteady forcing functions. An example is the centennial variability found by Mikolajewicz and Maier-Reimer (1990) when they force a global ocean model with stochastic (white noise in time) freshwater fluxes. The variability found in the model is a 50 Sv (Sv $\equiv 10^6$ $m^3\ s^{-1}$) pulsation in the transport of the circumpolar current. The pulsation is an event whose duration is roughly 320 years, but the time between these events seems to be random and varies between 200 and 1000 years [see Fig. 8 in Mikolajewicz and Maier-Reimer (1990)]. The model attenuates the high-frequency band of the forcing and the spectrum of the response is red, with a wide peak centered at 300 years. One possible interpretation of these results is that the white-noise forcing has excited a large-scale, weakly damped, linear eigenmode of the system. Indeed the pulsation is often characterized in these terms.

The simplest model consistent with this hypothesis is Hasselmann's (1976): the evolution of a single variable, the average temperature of the ocean, is governed by linear relaxation to a prescribed value. This is an attempt to parameterize the ocean–atmosphere feedback. If the prescribed temperature is a random variable with zero mean (white noise in time), then the response has a "red spectrum" bounded in amplitude at all frequencies. In this case there are no spectral peaks and the amplitude of the response is directly proportional to the size of the random forcing.

A second interpretation is that the pulsation found by Mikolajewicz and Maier-Reimer (1990) is triggered by some particularly large fluctuation in the stochastic forcing that pushes the system from one stable configuration to another; that is, the system is fundamentally nonlinear. In this case the response depends nonlinearly on the amplitude of the random forcing and on the parameters of the underlying deterministic dynamics. This perspective has been used in energy balance models of the atmosphere's climatic state by Sutera (1981) and Nicolis and Nicolis (1981) [cf. Ghil and Childress (1987) for a comprehensive review]. The same point of view is adopted here to study *extrinsic* nonlinear variability in the buoyancy-driven ocean circulation. The competing effects of heat and salt allow multiple equilibria, which can be randomly sampled in time if the fluctuations of the external forcing are large enough.

Hasselmann's model cannot be applied to study the evolution of randomly forced salinity concentration, because there is no direct ocean–atmosphere feedback that limits the value of salinity. Nonlocal exchanges between different latitudes are necessary to avoid unbounded accumulation of salt. The simplest model that illustrates how the thermohaline circulation transports properties latitudinally is the Stommel (1961) box model (see Fig. 1). Bryan and Hansen (1993) have examined the response of Stommel's box model to stochastic fluctuations of the thermal and saline forcing. The underlying deterministic system has two stable steady states and the random driving moves the temperature and salinity away from the initial equilibrium. In the regime considered by Bryan and Hansen (1993) the variance of the random forcing is not large enough to push the system to the alternate equilibrium: exponentially decaying fluctuations around the initial steady state are excited. White-noise forcing produces a response with a peakless red spectrum for both temperature and salinity. The thermohaline circulation ensures that the salinity spectrum is bounded at low frequencies. The temperature spectrum is saturated by both the ocean–atmosphere feedback and the exchange with other latitudes.

2. The box model

In this section we reexamine the box model of Bryan and Hansen (1993) in the regime where the stochastic forcing is of intermediate amplitude. Box models can be considered as very coarse finite difference approximations of the continuum system, with a box for every grid point. They are the simplest setting to study stochastically forced systems.

Figure 1 shows the coarsest possible "GCM": the model ocean has two grid points. It models the vertically averaged circulation in a single hemisphere. More boxes can be added to model interhemispheric exchanges (cf. Rooth 1982) or to "increase the resolution," but here we consider the minimal model that exhibits multiple equilibria. Box 1 (the low-latitude warm water) has temperature $T_1(t)$ and box 2 (the high-latitude cold water) has temperature $T_2(t)$. The salinities are $S_1(t)$ and $S_2(t)$, respectively. The density is related to the temperature and salinity by the linearized equation of state

$$\rho/\rho_0 = 1 + \alpha_S(S - S_0) - \alpha_T(T - T_0). \quad (2.1)$$

The subscript 0 denotes the reference state around which the linearization is done. We take $\rho_0 = 1029$

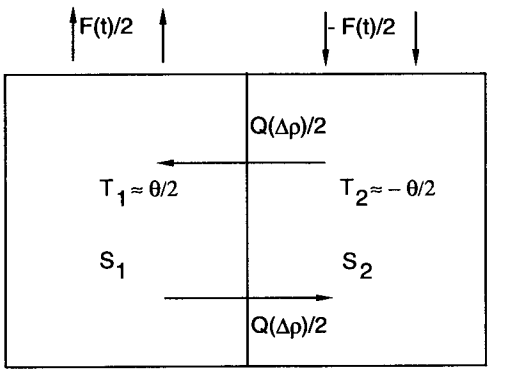

FIG. 1. The two-box model of Stommel (1961). The boxes represent two control volumes at different latitudes. Box 1 is the low-latitude box where the relaxation temperature is $\theta/2$, and box 2 is the high-latitude box where the relaxation temperature is $-\theta/2$.

kg m^{-3}, $T_0 = 5°C$, and $S_0 = 35$ psu, and we have $\alpha_S = 0.75 \times 10^{-3}$ psu^{-1} and $\alpha_T = 0.17 \times 10^{-3}\,°C^{-1}$. The conservation equations for the temperature and salinity are

$$\dot{T}_1 = -t_r^{-1}\left(T_1 - \frac{\theta}{2}\right) - \frac{1}{2}Q(\Delta\rho)(T_1 - T_2)$$

$$\dot{T}_2 = -t_r^{-1}\left(T_2 + \frac{\theta}{2}\right) - \frac{1}{2}Q(\Delta\rho)(T_2 - T_1)$$

$$\dot{S}_1 = \frac{F(t)}{2H}S_0 - \frac{1}{2}Q(\Delta\rho)(S_1 - S_2)$$

$$\dot{S}_2 = -\frac{F(t)}{2H}S_0 - \frac{1}{2}Q(\Delta\rho)(S_2 - S_1). \qquad (2.2)$$

The exchange of mass between the two boxes is modeled with the "exchange function" denoted by Q. We suppose that the exchange function depends only on the density difference between the two boxes, $\Delta\rho \equiv \alpha_S(S_1 - S_2) - \alpha_T(T_1 - T_2)$. Further we suppose that Q is positive definite so that the transport of temperature, salinity, and density are all downgradient. Specific models for Q are discussed below. The temperatures are forced using relaxation conditions to a prescribed value θ with a time constant t_r, while the salinities are forced with a prescribed flux $F(t)$. The asymmetry between heat and salt arises purely from the difference between these linear forcing functions: to model atmospheric feedbacks one uses relaxation forcing for T_1 and T_2 while the salinities are forced with a prescribed flux that represents imbalances between evaporation and precipitation plus runoff over each box. The depth of the model ocean is denoted by H.

Subtracting the equations in (2.2) gives a coupled pair of equations for the salinity and temperature differences between the boxes, $\Delta S \equiv S_1 - S_2$ and $\Delta T \equiv T_1 - T_2$,

$$\frac{d}{dt}\Delta T = -t_r^{-1}(\Delta T - \theta) - Q(\Delta\rho)\Delta T$$

$$\frac{d}{dt}\Delta S = \frac{F(t)}{H}S_0 - Q(\Delta\rho)\Delta S. \qquad (2.3)$$

The system above, together with the equation of state (2.1) and a model for $Q(\Delta\rho)$, is the final form of Stommel's two-box model.

Various models for the exchange functions are

$$Q_1 = t_d^{-1}, \quad Q_2 = t_d^{-1} + V^{-1}q|\Delta\rho|,$$

$$Q_3 = t_d^{-1} + V^{-1}q(\Delta\rho)^2. \qquad (2.4)$$

The choice Q_1 corresponds to diffusion on a timescale t_d. It makes (2.3) linear so that, if F is constant in time, there is a unique, steady, globally attracting solution. The choice Q_2 is Stommel's original model, which is based on a capillary tube analogy. With V the volume of the box, the transport of Poiseuille flow, q, is proportional to the pressure difference. This nonlinear exchange function leads to multiple equilibria: in certain parts of the parameter space, and with the flux F and the control volume V independent of time, (2.3) has three steady solutions and two of these are linearly stable. In fact, all initial conditions eventually fall into one of these two steady, stable attractors. Thus, with the exchange function Q_2 Stommel's model has no intrinsic time dependence. The choice Q_3 is qualitatively the same as Q_2: V is the volume of the box and q is proportional to the transport through advective processes. If F and V are constant, then there can be three steady equilibria, two of which are linearly stable. All initial conditions eventually find one of these two attractors, so that again there is no intrinsic time dependence. The model Q_3 will be used here because Cessi and Young (1992) have shown that, in a particular parameter range, Boussinesq convection leads to a nonlinear exchange function with that form.

The system (2.3) is best dealt with in nondimensional variables. We use the definitions

$$x \equiv \frac{\Delta T}{\theta}, \quad y \equiv \frac{\alpha_S \Delta S}{\alpha_T \theta}, \quad t = t_d t'. \qquad (2.5)$$

Equations (2.3) become

$$\dot{x} = -\alpha(x - 1) - x[1 + \mu^2(x - y)^2]$$

$$\dot{y} = p(t) - y[1 + \mu^2(x - y)^2]. \qquad (2.6)$$

We have denoted with $\alpha \equiv t_d/t_r$ the ratio of the diffusive timescale and the temperature relaxation timescale.

We now estimate the order of magnitude of the nondimensional parameters α, μ^2, and p. A typical value is $t_r \approx 25$ days, while the diffusion time is set to be $t_d = L^2/(\pi^2\kappa_H)$. [The factor π^2 arises because the gravest solution of the continuous diffusion equation, satisfying the condition of no lateral flux, is $\cos(\pi Y/L)$.] With a coefficient of horizontal diffusion $\kappa_H = 1000$ m^2 s^{-1} and a meridional scale $L = 8250$ km, the diffusion time is 219 years. This gives $\alpha \approx 3.6 \times 10^3$. The parameter $\mu^2 \equiv qt_d(\alpha_T\theta)^2/V$ corresponds to the ratio of the diffusive timescale, t_d, to the advective one, $t_a^{-1} \equiv q(\alpha_T\theta)^2/V$. The control volume is estimated as $V = L\delta_w H$, where δ_w is the typical width of the western boundary current. This choice is motivated by the hypothesis that the western boundary current transport, rather than the interior flow, determines the advective timescale. The transport [in this notation given by $q(\alpha_T\theta)^2$] is taken to be that of the deep western boundary current. With the present North Atlantic in mind an appropriate value for the transport is 12×10^6 m^3 s^{-1}. With $H = 4500$ m and $\delta_w = 300$ km, the advective timescale is $t_a \approx 29$ years, corresponding to a boundary current velocity of about 0.9×10^{-2} m s^{-1}. These estimates lead to $\mu^2 \approx 7.5$. If the control volume is estimated using the entire width of the North Atlantic, the advective timescale is $t_a = 320$ years, that

is, of the order of the diffusive timescale, a result at odds with the notion that the present day transport is dominated by advective processes.

The function

$$p \equiv \frac{\alpha_S S_0 t_d}{\alpha_T \theta H} F(t) \tag{2.7}$$

is the nondimensional freshwater flux and is decomposed in a time-averaged part $\bar{p}$ and a stochastic component, $p'(t)$, which fluctuates in time. The evaluation of this quantity is difficult because the observations of precipitation and evaporation over the oceans are inadequate. An estimate for the present North Atlantic freshwater flux, F, has been given by Schmitt et al. (1989). If the contribution of the marginal seas is included, $\bar{F} \approx 2.3$ m yr^{-1}. With $\theta = 20$°C, a reasonable value is $\bar{p} \approx 1$. The amplitude of the fluctuating part depends on the timescale considered. Schmitt et al. (1989), indicate that $p'/\bar{p}$ exceeds 3 on the seasonal timescale. The amplitude on much longer timescales is presumably smaller. However, here we will examine the case where $p'(t)$ is stochastic white noise, as in Stommel and Young (1993) and Bryan and Hansen (1993) [both papers used the Stommel (1961) original parametrization Q_2]. This choice is certainly unrealistic. However, the response of simple dynamical systems to colored noise is a subject of current research (e.g., Mannella and McClintock 1990), and would be beyond the scope of this article.

The system (2.6) can be further simplified because α is very large, while all the other parameters are O(1). In this regime the temperature is approximately clamped to the prescribed value and the solution of (2.6a) is

$$x = 1 + \mathrm{O}(\alpha^{-1}), \tag{2.8a}$$

and, to a first approximation, (2.2b) becomes

$$\dot{y} = -[1 + \mu^2(y-1)^2]y + \bar{p} + p'(t) + \mathrm{O}(\alpha^{-1}). \tag{2.8b}$$

The advantage of this limit is that, neglecting the corrections of $\mathrm{O}(\alpha)$, (2.8b) also describes the trajectory of a viscous particle [with no inertia] in the double-welled potential

$$V(y) \equiv \mu^2\left(\frac{y^4}{4} - \frac{2}{3}y^3 + \frac{y^2}{2}\right) + \frac{y^2}{2} - \bar{p}y, \tag{2.9}$$

subject to a Brownian force $p'(t)$. This is a classical and well-understood problem in stochastic processes, for example, Gardiner (1985). Its application [using the exchange function $Q_2(z)$] to the thermohaline problem has been examined by Stommel and Young (1993) in the range where the fluctuations are so strong that all the statistical properties of y are independent of the deterministic parameters μ and $\bar{p}$. Indeed the expected value of y is independent of the amplitude of $p'(t)$: this is why this is termed a "regulating" regime.

With no stochastic fluctuations, $p' = 0$, the salinity difference y always reaches one of the two stable equilibria located at the minima of the potential V [see Fig. 2]. The third steady state, located at the maximum of the potential V, is linearly unstable. In the limit of weak stochastic fluctuations, analyzed by Bryan and Hansen (1993), the system "rattles" around the initial equilibrium solution, but can never overcome the potential barrier associated with the unstable steady state.

Here we will consider the regime where the fluctuations are large enough to occasionally flip the system from one stable equilibrium to the other, yet not so large as to completely mask the underlying dynamical structure of the deterministic system. Thus we examine forcing whose stochastic component is of strength intermediately between that considered by Bryan and Hansen (1993) and Stommel and Young (1993).

3. Deterministic perturbations

In this section we determine the minimum amplitude and duration of a deterministic perturbation necessary to shift the system, initially in a stable state, to the alternate stable equilibrium. This is a useful exercise before proceeding with the statistical description of the stochastically forced system. The perturbation $p'(t)$ in (2.8b) is assumed to have the simple form

$$p'(t) = \begin{cases} 0, & t \leqslant 0 \\ \Delta, & t \leqslant \tau \\ 0, & t > \tau. \end{cases} \tag{3.1}$$

We assume that y is initially in the globally stable state, y_a, of Fig. 2; that is, we set $y(t<0) = y_a$. This corresponds to a state where the density difference between the boxes, and thus the "transport," is substantial. At $t = 0$ the perturbation is switched on. In order for y to

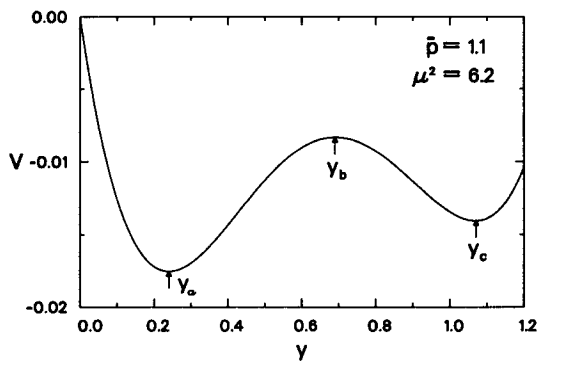

FIG. 2. The potential V defined in (2.9) as a function of the salinity gradient y. The steady states of (2.8b) with $p' = 0$ are extrema of V. The global minimum, located at $y_a \approx 0.24$ is linearly stable, as is the local minimum at $y_c \approx 1.07$. However, the latter is a metastable equilibrium, and a generic finite amplitude perturbation will move the system to y_a. The maximum, located at $y_b \approx 0.69$, is linearly unstable.

reach the alternate stable equilibrium, y_c, characterized by a weak density difference, it is sufficient that y is just over the potential barrier at the unstable point y_b after the perturbation is switched off at $t = \tau$. The first-order equation (2.8b) can be integrated to give

$$\int_{y_a}^{y_b} dy[\bar{p} + \Delta - y - \mu^2 y(y-1)^2]^{-1} = \int_0^{\tau} dt. \quad (3.2)$$

The integral in (3.2) gives the minimum duration of the perturbation, τ, as a function of its amplitude, Δ, and of the parameters $\bar{p}$ and μ. A typical graph of Δ versus τ, for fixed $\bar{p}$ and μ, is shown in Fig. 3. The most important point is that there is a critical amplitude, $\Delta_0(\bar{p}, \mu)$, below which the alternate equilibrium is never reached, even if the perturbation is applied for an infinitely long time.

The critical amplitude $\Delta_0(\bar{p}, \mu)$ is easily found by the graphic construction shown in Fig. 4. If the system is perturbed for an infinite time it will reach a new steady state, y'_a, corresponding to the value of the freshwater flux given by $\bar{p} + \Delta_0$. The new equilibrium is at the inflection point of the potential V shown in Fig. 4. The system can just overcome the potential barrier of the initial configuration, $V(y_b)$ in Fig. 2, if the new unstable equilibrium y'_b coincides with the perturbed stable state y'_a. In this case the potential well has one inflection point and one minimum, y'_c. If $\Delta > \Delta_0$, then the system moves past $y'_a = y'_b$ and "rolls down" to reach the alternate equilibrium y'_c. When the forcing is switched off the minimum moves from y'_c to y_c and the system is in a stable steady state. Straightforward algebra gives the expression for Δ_0:

$$\bar{p} + \Delta_0 = \frac{2}{27}\mu^2[1 + (1 - 3\mu^{-2})^{3/2}] + \frac{2}{3}. \quad (3.3)$$

Dimensional values can be restored using (2.7) for Δ. The duration τ is to be multiplied by the diffusion time,

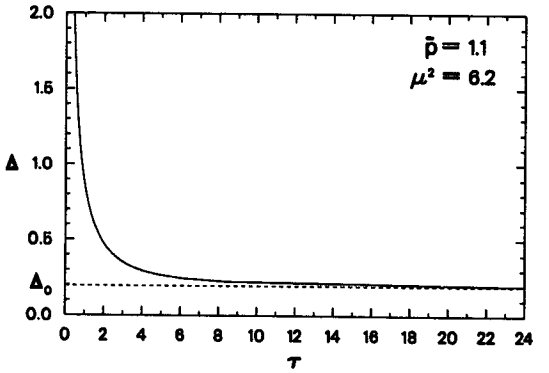

FIG. 3. The minimum amplitude of a perturbation, as a function of its duration, that will shift the system from the globally stable equilibrium y_a of Fig. 2 to the metastable state, y_c. The perturbation must exceed a critical amplitude, Δ_0, in order to displace the system from y_a, even if applied for an infinite time.

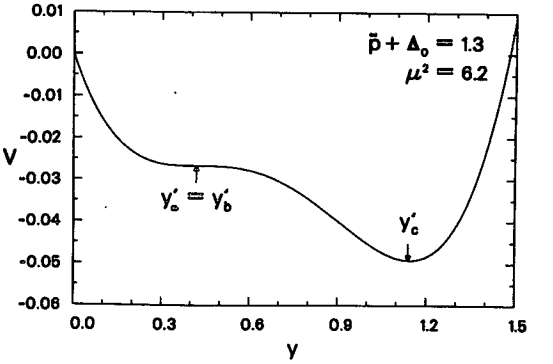

FIG. 4. The critical perturbation, of amplitude Δ_0, is found by requiring that the perturbed potential, shown here as a function of y, has one minimum and one inflection point. The perturbed potential is obtained by replacing $\bar{p}$ with $\bar{p} + \Delta$ in (2.9) and requiring that the global minimum y_a of Fig. 2 becomes the inflection point y'_a.

taken here to be $t_d = 219$ years. Here $\Delta_0 = 0.2$ corresponds to a high-latitude freshwater flux of 0.4 m yr^{-1}. Fairbanks (1989) estimates that the meltwater discharge rate preceeding the Younger Dryas event reached a maximum of 0.5 m yr^{-1} (assuming that it was uniformly distributed over half of the North Atlantic). This value is very close to the critical perturbation, Δ, obtained at $\tau = 4.6$ (corresponding to a dimensional value of 1000 years). Notice that Fairbank's estimate places the system very close to the critical amplitude below which no transition is possible. Thus, a possible interpretation of the Younger Dryas event is that a perturbation in the freshwater flux induced a transition to an alternate equilibrium with reduced meridional overturning [the numerical experiments of Wright and Stocker (1993) provide a more detailed framework for the same hypothesis]. The calculation of this section indicates that the duration of the perturbation, as well as its amplitude, is a crucial parameter determining whether a transition to the alternate state occurs.

4. Statistical description of the stochastically forced model

In this section we consider $p'(t)$ to be a random number, and the response y is also a stochastic variable. The state of the system is described in terms of its statistical properties by the probability density function [PDF] $\phi(y, t)$, which gives the fraction of salinity differences in the interval $(y, y + dy)$ at time t. Equation (2.8b) is discretized in time with the Euler scheme, and $p'(t)$ is randomly picked at every time step from a Gaussian distribution with zero mean and variance

$$\sigma^2 \equiv \langle p'^2 \rangle, \quad (4.1)$$

where the angle brackets indicate an ensemble average.

If dt is the size of the time step, then the autocorrelation function is

$$\langle p'(s)p'(s+t)\rangle = \begin{cases} (1 - t/dt)\sigma^2, & t \leqslant dt \\ 0, & t > dt. \end{cases} \quad (4.2)$$

In this way the correlation time of the stochastic forcing is much shorter than the timescale of the deterministic differential equation

$$\dot{y} = -V_y, \quad (4.3)$$

with V given by (2.9) and the PDF ϕ evolves according to the Fokker–Planck equation (cf. Gardiner 1985)

$$\partial_t \phi = \partial_y(V_y\phi) + D\partial_y^2\phi. \quad (4.4)$$

The diffusion coefficent, D, is determined by the statistical properties of p', and is given by

$$D = \int_0^\infty dt\langle p'(s)p'(s+t)\rangle = dt\sigma^2/2. \quad (4.5)$$

The PDF is found by solving the Fokker–Planck equation (4.4). A typical time series of the salinity difference is shown in Fig. 5. Mostly y fluctuates in the vicinity of the globally stable state y_a or, less often, in the neighborhood of the metastable state y_c. For a time series long enough to "forget" about the initial condition, the PDF of the states is well described by the stationary solution, ϕ_s, of the Fokker–Planck equation (4.4) given by

$$\phi_s(y) = N\exp(-V(y)/D). \quad (4.6)$$

The normalization coefficent N is such as to make the area under $\phi_s(y)$ equal to unity. The stationary distribution ϕ_s is in good agreement with the distribution of states obtained by direct computation (Fig. 6). Because D is much smaller than V except at the minima of V [cf. Fig. 2], the stationary PDF is bimodal, with two peaks near the stable equilibria. Moreover the peak centered at the globally stable state y_a exceeds that at the metastable state y_c by a factor of $\exp[(V(y_c) - V(y_a))/D]$.

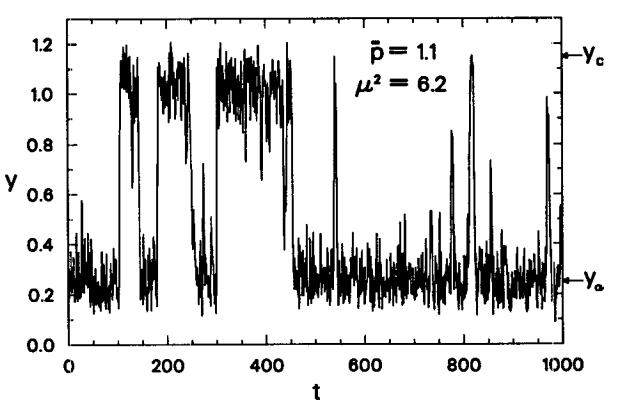

FIG. 5. A typical time series of y solution of (2.8b). At every time step, of size $dt = 3.33 \times 10^{-3}$, the stochastic forcing is randomly picked from a Gaussian distribution with zero mean and standard deviation $\sigma = 3.3$. The values of $\bar{p}$ and μ^2 are as in Fig. 2. The salinity difference y spends most of the time in the neighborhood of the stable states y_a and y_c.

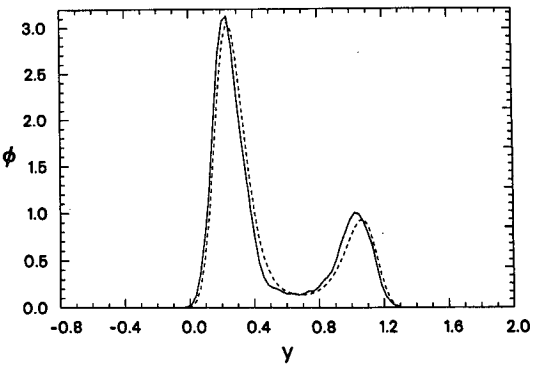

FIG. 6. The stationary PDF (4.6) (dashed line) is in good agreement with the PDF resulting from numerical simulation (solid line). To estimate the "true" PDF we have binned the values of salinity difference shown in Fig. 5.

Remarkably, much information about the low-frequency properties of the nonstationary PDF $\phi(y, t)$ can be obtained from the stationary distribution $\phi_s(y)$. For example, it is possible to determine how long $y(t)$ takes to leave the neighborhood of the left-hand well in Fig. 2. Because most of the time that a point takes to go from one stable equilibrium to the other is spent overcoming the potential barrier at the unstable point, the detailed specifications of the initial value in the left-hand well and of the final value in the right-hand well are unnecessary. Therefore all the points to the left of y_b can be treated as members of an ensemble and one can meaningfully define a "mean escape time" from the left-hand well to the right-hand well. It is approximately given by

$$\langle t_{a\to c}\rangle \approx \frac{1}{D}\int_{-\infty}^{y_b} dy\phi_s(y)\int_{y_a}^{y_c} dx[\phi_s(x)]^{-1}. \quad (4.7)$$

Similarly the mean escape time from the right-hand well to the left-hand well in Fig. 2 is given by

$$\langle t_{c\to a}\rangle \approx \frac{1}{D}\int_{y_b}^{\infty} dy\phi_s(y)\int_{y_a}^{y_c} dx[\phi_s(x)]^{-1}. \quad (4.8)$$

The approximations in (4.7) and (4.8) are accurate only if $D \ll V(y_b)$ (e.g., Gardiner (1985), and in this regime the integrals are easily calculated by Laplace's method

$$\langle t_{a\to c}\rangle \approx 2\pi[-V_{yy}(y_a)V_{yy}(y_b)]^{-1/2} \times \exp[(V(y_b) - V(y_a))/D],$$

$$\langle t_{c\to a}\rangle \approx 2\pi[-V_{yy}(y_c)V_{yy}(y_b)]^{-1/2} \times \exp[(V(y_b) - V(y_c))/D]. \quad (4.9)$$

Notice the strong dependence of the mean escape times on D, which is in turn proportional to the variance of the stochastic forcing as in (4.5).

The probabilities $N_a(t)$ and $N_c(t)$ of being in the left-hand and right-hand well, respectively, are defined as

$$N_a(t) \equiv \int_{-\infty}^{y_b} dy\phi(y, t)$$

$$N_c(t) \equiv \int_{y_b}^{\infty} dy\phi(y, t). \quad (4.10)$$

They are approximately governed by

$$\dot{N}_a(t) = -\omega_a N_a(t) + \omega_c N_c(t)$$

$$\dot{N}_c(t) = -\omega_c N_c(t) + \omega_a N_a(t). \quad (4.11)$$

The initial condition is $N_a(0) = 1$ [$N_c(0) = 1$] if the point is in the left-hand (right-hand) well initially, and $N_a(t) + N_c(t) = 1$. The decay rates toward the minima of the potential are inversely proportional to the mean escape times:

$$\omega_a \equiv (\langle t_{a\to c}\rangle)^{-1}, \quad \omega_c \equiv (\langle t_{c\to a}\rangle)^{-1}. \quad (4.12)$$

The approximate evolution equations (4.11), due to Kramers, state that the "bulk" probability to be in the neighborhood of a stable equilibrium is governed by the equation for a random telegraph process. It approximates the low-frequency behavior of $\phi(y, t)$ with a process where only two states are permitted: y_a with probability $N_a(t)$ or y_c with probability $N_c(t) = 1 - N_a(t)$. For times much longer than either mean escape times, $N_a(t)$ approaches the stationary probability N_{as} given by

$$N_{as} \equiv \int_{-\infty}^{y_b} dy\phi_s(y) = \frac{\langle t_{a\to c}\rangle}{\langle t_{a\to c}\rangle + \langle t_{c\to a}\rangle}. \quad (4.13)$$

Similarly $N_c(t)$ asymptotes to the stationary distribution

$$N_{cs} \equiv \int_{y_b}^{\infty} dy\phi_s(y) = \frac{\langle t_{c\to a}\rangle}{\langle t_{a\to c}\rangle + \langle t_{c\to a}\rangle}. \quad (4.14)$$

The equilibration to the stationary distributions is a slow process for weak stochastic forcing: the mean escape times (4.9) become very long if the diffusion becomes small. In a GCM it may not be possible to integrate for a time long enough to achieve a statistical steady state. Box models are simple enough to be integrated until their long-term statistics are reliable.

5. The spectrum of the solution

At low frequencies the salinity jumps between the two equilibria as described in the discussion surrounding (4.11). Thus we approximate the behavior of the system with a random telegraph process where only the states y_a and y_c are allowed. The time correlation function of the random telegraph process governed by (4.11) is (Gardiner 1985)

$$\langle y(t)y(s)\rangle = \omega_a\omega_c \frac{(y_a - y_c)^2}{(\omega_a - \omega_c)^2} \exp[-(\omega_a + \omega_c)|t - s|]. \quad (5.1)$$

The spectrum of y is the Fourier transform of the time correlation function (5.1) and is given by

$$\mathcal{S}(\omega) = \frac{2\omega_a\omega_c(y_a - y_c)^2}{(\omega_a + \omega_c)[(\omega_a + \omega_c)^2 + \omega^2]}[1 + O(\omega_a dt)]. \quad (5.2)$$

The approximate spectrum (5.2) is shown in Fig. 7 (dashed line) and, for small frequencies, is in excellent agreement with that obtained from direct computations (solid line).

Of course the random telegraph process is not a good approximation of the details of the fluctuations around the equilibrium states, but this "rattling" can be approximated by linearizing around each equilibrium. This is the approach of Bryan and Hansen (1993).

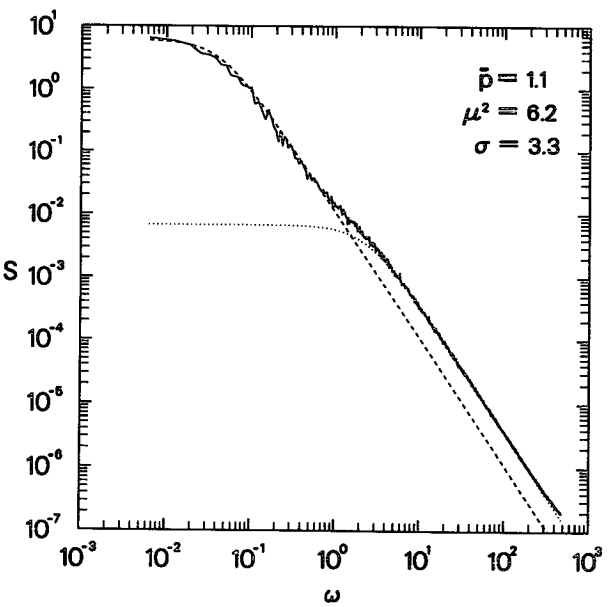

FIG. 7. The solid line shows the spectrum of y resulting from the average of 20 numerical solutions of (2.8b) for different realizations of the stochastic forcing. For each realization the parameters are the same as in Fig. 5. At low frequencies the evolution of the salinity difference is well approximated by a random telegraph process jumping between the two stable states $y = y_a$ and $y = y_c$ with average escape rates given in (4.12). For the parameters used here, the escape rates from the left-hand and right-hand well of Fig. 2 are $\omega_a = 1.07 \times 10^{-2}$ and $\omega_c = 3.20 \times 10^{-2}$, respectively. The analytic spectrum for this approximation is plotted as a dashed line. At high frequencies the salinity fluctuates around the equilibrium y_a with probability N_{as} [given in (4.13)] and around y_c with probability $(1 - N_{as})$. The decay rate toward the equilibria is calculated by linearizing around the steady states, and the spectrum is a weighted sum of the resulting Ornstein–Uhlenbeck processes (shown here as a dotted line). The decay rates toward the equilibria are $V_{yy}(y_a) = 2.32$ and $V_{yy}(y_c) = 1.94$.

The fluctuations around the equilibrium y_a can be approximated by an Ornstein–Uhlenbeck process:

$$\dot{y} = -V_{yy}(y_a)(y - y_a) + p'(t), \qquad (5.3)$$

while near the state y_c, the salinity evolves according to

$$\dot{y} = -V_{yy}(y_c)(y - y_c) + p'(t). \qquad (5.4)$$

The spectra of the linear equations (5.3) and (5.4) are easily calculated, as in Hasselmann (1976). Call $\mathcal{S}_a(\omega)$ the spectrum associated with the solution of (5.3). Then

$$\mathcal{S}_a(\omega) = \frac{\sigma^2 dt}{\omega^2 + V_{yy}(y_a)^2}. \qquad (5.5)$$

Similarly, the spectrum associated with the fluctuations around the equilibrium y_c, $\mathcal{S}_c$, is given by

$$\mathcal{S}_c(\omega) = \frac{\sigma^2 dt}{\omega^2 + V_{yy}(y_c)^2}. \qquad (5.6)$$

Thus, for frequencies of the order of the linear decay rates $V_{yy}(y_a)$ and $V_{yy}(y_c)$, the spectrum of $y(t)$ is approximately

$$\mathcal{S} \approx N_{as}\mathcal{S}_a + N_{cs}\mathcal{S}_c, \qquad (5.7)$$

where N_{as} and N_{cs} are defined in (4.13) and (4.14). Notice that the spectrum that results from considering the fluctuations within each potential well has been weighted by the stationary probability of being in that well in order to find the total spectrum. The spectrum (5.7) is plotted as a dotted line in Fig. 7 and for frequencies $\omega > O(1)$ it agrees very well with the spectrum obtained from the numerical solutions of (2.8b) (solid line).

Notice that both (5.2) and (5.7) are "red spectra" proportional to $(\Omega^2 + \omega^2)^{-1}$. However, they are the results of very different processes. At large frequencies the forcing excites fluctuations that decay exponentially in time and well described by the linearized Ornstein–Uhlenbeck equations (5.3) and (5.4). Indeed the spectrum in this range is proportional to the amplitude of the forcing. Low frequencies are dominated by a fundamentally nonlinear process and the dependence of the spectrum on the amplitude of the forcing is given by the factor $\omega_a\omega_c(\omega_a + \omega_c)^{-3}$ in (5.2). The variance σ^2, proportional to the diffusion D [cf. (4.5)], determines the height of the spectrum at low frequencies through the relation

$$\mathcal{S} \sim \omega_a\omega_c(\omega_a + \omega_c)^{-3} = (\langle t_{a\to c}\rangle\langle t_{c\to a}\rangle)^2(\langle t_{a\to c}\rangle + \langle t_{c\to a}\rangle)^{-3}, \qquad (5.8)$$

where the expressions for the mean escape times are given in (4.9). For weak diffusion, D, the dependence of the mean escape times is $\sim \exp(A/D)$, and the amplitude of the spectrum *increases* as the diffusion decreases. As the variance of the random forcing decreases the analytic approximation obtained by "patching" (5.2) to (5.7) becomes more accurate. At the same time numerical solutions of (2.8b) are very demanding because the saturation frequencies decrease rapidly as the diffusion D decreases. In Fig. 8 we show the analytic predictions for the spectrum when the variance σ^2 is half (dotted line) and twice (dashed line) of that used in Fig. 7 (shown again as a solid line). At high frequencies the amplitude of the spectrum scales linearly with σ^2. However in the low-frequency range, where nonlinearity is important, the amplitude of the spectrum increases when the variance decreases. Notice also that the saturation frequency increases by a factor of approximately 23 when the variance is halved.

The analytic approximation obtained by patching (5.2) and (5.7) fails when $V(y_b)/D$ is of order 1 or smaller. In that regime the jumps between equilibria to the other are frequent and the double-well structure of the potential is obscured. This is the regulating regime studied by Stommel and Young (1993).

6. Conclusions

A thermohaline box model driven by prescribed temperature and freshwater flux has been examined.

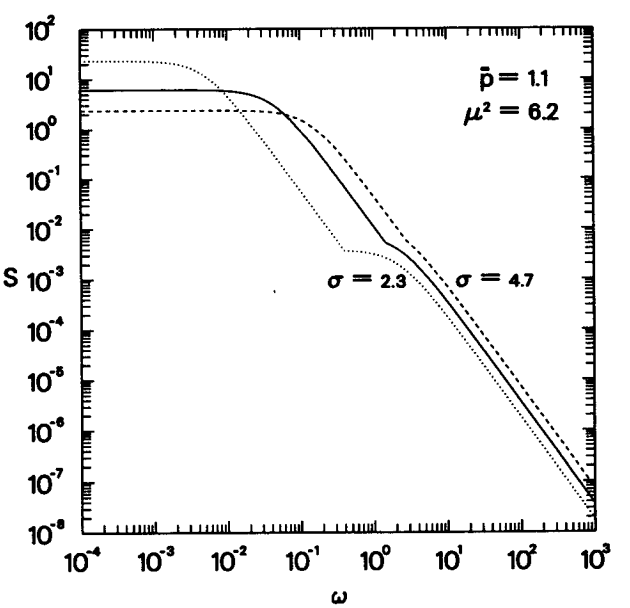

FIG. 8. The approximate analytic spectrum obtained by patching (5.2) and (5.7) for different values of the variance of the random forcing, σ^2. The solid line shows the spectrum for $\sigma = 3.3$, the same value used in Fig. 7. The dotted line is the spectrum for $\sigma = 3.3/\sqrt{2}$: the escape rates from the left-hand and right-hand well are $\omega_a = 4.61 \times 10^{-4}$ and $\omega_c = 4.50 \times 10^{-3}$, respectively. The smallest of the two determines the saturation frequency of the spectrum. The dashed line is the spectrum obtained for $\sigma = 3.3\sqrt{2}$: the escape rates are $\omega_a = 5.2 \times 10^{-2}$ and $\omega_c = 8.5 \times 10^{-2}$. In the high-frequency range, the amplitude of the spectrum increases linearly with the variance, while in the low-frequency range of the spectrum it decreases nonlinearly as the variance is decreased.

The dynamics are essentially the same as Stommel's (1961) model: the system always reaches one of two possible stable equilibria. A finite amplitude perturbation of the freshwater flux can shift the system into the alternate state, and the minimum amplitude depends on the duration of the disturbance. However, the disturbance must be above a critical amplitude, regardless of its duration, for a transition to occur. Using values of the parameters appropriate for the Younger Dryas event, the system can be switched from a state with substantial overturning to a mode with very weak transport. The suggestion that the Younger Dryas event can be characterized as a transition between equilibria has been advanced by Broecker et al. (1985) and has been documented in GCM experiments (e.g., Wright and Stocker 1993). The virtue of the box model is that the dependence on the parameters is easily established. The results of section 3 indicate that the freshwater perturbation preceding the Younger Dryas event [as estimated by Fairbanks (1989)] was of the same order of magnitude as the threshold below which no transition is possible.

We have then considered the forcing to have a stochastic component that induces occasional transitions from one equilibrium to the other, as well as fluctuations around each state. The evolution of the salinity gradient is analogous to the trajectory of a viscous particle in a double-well potential and subject to Brownian motion. This is a classical and well-understood problem in stochastic processes when the spectrum of the fluctuating forcing is white.

Not surprisingly the spectrum of salinity differences does not exhibit a peak at any frequency. Because the system does not have any *intrinsic* variability the white-noise forcing induces a red spectrum saturated at low frequencies. This does not necessarily imply that the underlying deterministic dynamics is *linear,* as in the cases examined by Hasslemann (1976) and Bryan and Hansen (1993). For low frequencies the nonlinear dynamics of the underlying deterministic system are important and the model behaves as a random telegraph process. In this range the spectrum is proportional to ω^{-2}. At high frequencies the deterministic system is essentially linear and again the spectrum has a ω^{-2} range. The two ranges are differentiated by the dependence on the amplitude of the stochastic forcing, σ.

At low frequencies the nonlinearity of the system is most important and the amplitude of the spectrum is proportional to the mean escape time $\langle t \rangle \sim \exp(A\sigma^{-2})$. For small amplitude of the random forcing the mean escape time becomes very long and this is a timescale that GCMs may not afford to resolve. In the high-frequency range, where the system is essentially linear, the spectrum's dependence on the random forcing's variance is linear. This proportionality dependence can be studied with GCMs.

In the example presented here the stochastic fluctuations are taken to be of the same amplitude at all frequencies and quite large: $\sigma/\bar{p} = 3$. Because the spectrum of freshwater fluxes anomalies is certainly colored, it is not possible to determine what value should be used for comparison with present day data. However, the analytic approximations for the mean escape times can be evaluated for any value of σ. Indeed the approximation (4.9) becomes more accurate as $\sigma \to 0$. For the parameters used in Fig. 7 the dimensional mean escape times can be obtained through multiplication by the diffusion time, estimated in section 2 to be $t_d = 219$ years. They are $\langle t_{c\to a} \rangle \times t_d \approx 6840$ years and $\langle t_{a\to c} \rangle \times t_d \approx 20\,470$ years. As noted earlier the mean escape times depend very strongly on the variance of the stochastic forcing. If σ^2 is decreased by a factor of 2 and all the other parameters are left unchanged, $\langle t_{c\to a} \rangle$ increases by a factor of 7 and $\langle t_{a\to c} \rangle$ by a factor of 23.

Determining the relation between the spectrum and the amplitude of the stochastic forcing in the low-frequency range requires long integrations with different values of the driving. This program is not feasible with GCMs, and the sources of uncertainty are numerous when analyzing observations. At present, simplified models of limited complexity may be the only viable tool to study the long-term statistical properties of randomly forced flows.

Acknowledgments. Numerous conversations with Bill Young and Bruce Cornuelle are gratefully acknowledged. The suggestions of the anonymous referees greatly improved this paper. Funding for this research is provided by the Commission of the European Communities through its MAST II Programme (MAS2-CT-92-0034).

REFERENCES

Broecker W. S., D. M. Peteet, and D. Rind, 1985: Does the ocean–atmosphere system have more than one stable mode of operation? *Nature,* **315,** 21–26.

Bryan, K., and F. C. Hansen, 1993: A toy model of North Atlantic climate variability on a decade to century time-scale. *The Natural Variability of the Climate System on 10–100 Year Time Scales,* U.S. Natl. Acad. of Sci.

Cessi, P., and W. R. Young, 1992: Multiple equilibria in two dimensional thermohaline convection. *J. Fluid Mech.,* **241,** 291–309.

Fairbanks, R. G., 1989: A 17,000-year glacio–eustatic sea-level record: Influence of glacial melting rates on the Younger Dryas event and deep-ocean circulation. *Nature,* **342,** 637–642.

Gardiner, C. W., 1985: *Handbook of Stochastic Methods for Physics, Chemistry, and the Natural Sciences.* Springer-Verlag, 442+xix.

Ghil, M., and S. Childress, 1987: *Topics in Geophysical Fluid Dynamics: Atmospheric Dynamics, Dynamo Theory, and Climate Dynamics.* Springer-Verlag, 485+xv.

Hasselmann, K., 1976: Stochastic climate models. Part I. Theory. *Tellus,* **28,** 289–305.

Lehman, S. J., and L. D. Keigwin, 1992: Sudden changes in North Atlantic circulation during the last deglaciation. *Nature,* **356,** 757–762.

Manabe, S., and R. J. Stouffer, 1988: Two stable equilibria of a coupled ocean–atmosphere model. *J. Climate,* **1,** 841–866.

Mannella, R., and P. V. E. McClintock, 1990: Noise in nonlinear dynamical systems. *Contemp. Phys.*, **31,** 179–194.

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

Nicolis, C., and G. Nicolis, 1981: Stochastic aspects of climatic transitions—Additive fluctuations. *Tellus,* **33,** 225–234.

Rooth, C., 1982: Hydrology and ocean circulation. *Progress in Oceanography,* Vol. 11, Pergamon, 131–149.

Schmitt, R. W., P. S. Bogden, and C. E. Dorman, 1989: Evaporation minus precipitation and density fluxes for the North Atlantic. *J. Phys. Oceanogr.*, **19,** 1208–1221.

Stommel, H., 1961: Thermohaline convection with two stable regimes of flow. *Tellus,* **13**(2), 224–230.

——, and W. R. Young, 1993: The average T–S relation of a stochastically forced box model. *J. Phys. Oceanogr.*, **23,** 151–158.

Sutera, A., 1981: On stochastic perturbation and long-term climate behavior. *Quart. J. Roy. Meteor. Soc.*, **107,** 137–151.

Weaver, A. J., and E. S. Sarachik, 1991: Evidence for decadal variability in an ocean general circulation model: An advective mechanism. *Atmos.-Ocean,* **29,** 197–231.

——, and T. M. C. Hughes, 1992: Stability and variability of the thermohaline circulation and its link to climate. *Trends in Physical Oceanography, Council of Scientific Research Integration—Research Trends Series,* Trivandrum, 56 pp.

——, J. Marotzke, P. F. Cummins, and E. S. Sarachik, 1993: Stability and variability of the thermohaline circulation. *J. Phys. Oceanogr.*, **23,** 39–60.

Welander, P., 1986: Thermohaline effects in the ocean circulation and related simple models. *Large-Scale Transport Processes in Oceans and Atmosphere,* J. Willebrand and D. L. T. Anderson, Eds., NATO ASI Series, Reidel, 163–200.

Winton, M., and E. Sarachick, 1993: Thermohaline oscillations induced by strong steady salinity forcing of ocean general circulation models. *J. Phys. Oceanogr.*, **23,** 1389–1410.

Wright, D. G., and T. F. Stocker, 1993: Younger Dryas experiments. *Ice in the Climate System,* W. R. Peltier, Ed., NATO ASI Series, Springer-Verlag.