Advective Time Lags in Box Models

DOUGLAS A. KURTZE

Department of Physics, North Dakota State University, Fargo, North Dakota

JUAN M. RESTREPO

Mathematics Department and Program in Applied Mathematics, The University of Arizona, Tucson, Arizona

(Manuscript received 14 January 2000, in final form 27 September 2000)

ABSTRACT

A box model of the thermohaline circulation with mixed boundary conditions in which advective processes are incorporated via an explicit time delay mechanism is considered. The pipes that connect the subtropical and subpolar boxes have a finite volume and do not interact with the atmosphere or with the rest of the ocean except for channeling fluxes between the subtropical and subpolar regions. The configuration can be reduced to a two-box model, which, unlike the traditional Stommel model, incorporates finite-time advective processes. It is found that including a time lag leaves the haline dominant steady state stable, but the thermally dominant steady state, which is stable in Stommel's model, can have an oscillatory instability. However, this instability does not lead to sustained oscillations. Instead, it simply makes the circulation cross over to the stable haline dominant pattern. Even in part of the parameter range for which the thermally dominant state remains linearly stable, the time lag leads to a finite-amplitude instability so that a relatively small—but not infinitesimal—perturbation about the thermal state can switch the circulation to the haline state.

1. Introduction

Climate variability on timescales of the order of a decade or longer is a central issue in climate dynamics (for reviews, see Rahmstorf et al. 1996; Latif et al. 1996; Latif 1998; Sarachik et al. 1996). While this variability is not well understood, there is consensus among many ocean dynamicists that the thermohaline circulation plays a role in it [but see Latif et al. (1996), Latif (1998) for opinions to the contrary].

Models of the thermohaline circulation, of varying degrees of complexity, exhibit a multiplicity of solutions corresponding to different climatic states; they may also possess unsteady behavior in the form of oscillatory and temporally irregular solutions. Large-scale oceanic and climatic models are capable of reproducing features that resemble observed climate variations, but it is often a challenging task to extract from these models the basic physical mechanisms that lead to these climatic manifestations. Specifically, it can be difficult to assess whether features appearing in the computations are bona fide consequences of the physics built into the model, the result of parameterizations, or even numerical artifacts. For the purposes of identifying fundamental mechanisms, it is often expedient to appeal to highly idealized models. While these models do not permit direct comparison to data (which is scant to begin with), they are useful in that they offer a framework that, once understood, offers insight into the more complex and realistic geophysical problem.

Perhaps the best known of these idealized models of the thermohaline circulation is Stommel's two-box model (1961). This model pictures the ocean as consisting of two boxes connected by pipes at the surface and at the bottom. Each box is well mixed so that the water in it has uniform temperature and salinity, and hence uniform density. The boxes are insulated, except at the air–sea interface where the boxes exchange heat and freshwater with the atmosphere above. The atmospheric fluxes of salinity and heat tend to increase the density contrast between the two boxes. The pipes, however, permit a pressure-driven exchange of water between the two boxes, which tends to reduce the density contrast. The competition among these effects leads to a steady thermohaline circulation. This model is thought to capture the long-time dynamics of a zonally averaged ocean, in the absence of Coriolis and wind stress forces, consisting of subpolar and subtropical bodies of water. At the air–sea interface, one typically takes restoring boundary conditions on the temperature and a prescribed salinity flux between the boxes (although Stommel originally proposed restoring boundary conditions

Corresponding author address: Prof. Juan M. Restrepo, Mathematics Department, Building 89, The University of Arizona, Tucson, AZ 85721.
E-mail: restrepo@math.arizona.edu

on both temperature and salinity). As Stommel has shown, the model's variability results from a multiplicity of steady solutions or climate states. The model does not, however, exhibit oscillatory solutions (see Ruddick and Zhang 1996).

Many of the models that suggest that self-sustained oscillations are possible in the thermohaline circulation are variants of Welander's *salt oscillator* (1982). The salt oscillator was used to show that a rapid collapse of the circulation is possible if there is an interruption of convection, leading to an accumulation of freshwater. In the thermohaline cirulation setting, a generalization of this model was shown in Lenderink and Haarsma (1994) to have four different solutions, one of which was characterized by self-sustained oscillations. However, D. Kurtze and J. Restrepo (1999, unpublished manuscript) show that a Stommel box model with convective exchanges with a lower reservoir shows a different picture than that presented by Welander. Specifically, the Stommel exchanges ameliorate oscillations, producing a different set of dynamics.

This is not to say that the haline mechanism mentioned above is the only mechanism that can lead to climate variability in thermohaline circulation models. Greatbatch and Peterson (1995) related the Howard–Malkus loop to flows relevant to nonsteady thermohaline circulation solutions. Oscillations are also possible in a Stommel box model if there is strong *stochastic* forcing of the salt flux at the atmosphere–ocean interface (see Stommel and Young 1993; Cessi 1994). In this instance the variability of the atmospheric conditions that contribute to precipitation, and the resulting variability of salt fluxes at the boundary of the box model, is seen to induce variability of the thermohaline circulation as given by Stommel's box model. Unsteady solutions are possible, mostly oscillations about the mean state. However, if the oscillations are strong enough they can also effect a transition from one state of the model to another one.

Zhang (1991) suggested that time delays might be important in the thermohaline circulation. In Tziperman et al. (1994) a simple four-box model (without explicit time delays) was considered, which was found to exhibit an oscillatory instability in certain parameter ranges. Griffies and Tziperman (1995) interpreted the oscillatory solutions of this four-box model as being due to an effective time delay arising from the phase lags between temperature and salinity perturbations in the various boxes. They studied the model in the damped-oscillation regime, showing that stochastic driving in this regime can produce oscillations in the thermohaline circulation.

In this paper we propose and investigate the simplest possible box model that incorporates finite time delays to determine whether it exhibits nonsteady solutions. Moreover, we suggest that these finite time delays take into account meridional advective processes that are ignored in the standard Stommel model.

Time delays in differential equations are well known to lead to oscillatory solutions. For instance, Suarez and Schopf (1988) used a simple delay-differential equation to model ENSO oscillations, with later elaborations by Battisti and Hirst (1989). In Suarez and Schopf (1988) the delay term accounted for the finite time propagation of adjustments in the temperature due to advective processes, further suggesting that these advective processes could be related to equatorial waves.

The model proposed here consists of subtropical and subpolar boxes, as in the Stommel model, connected by ''pipes'' which have finite volume, and in which the water is not mixed. One of these pipes can be thought of as the ''subsurface'' pipe and the other as the ''lower'' pipe. We explicitly ignore any exchange of heat or salt between the pipes and the atmosphere or the rest of the ocean; the pipes act only as conduits for flow between the subtropical and subpolar boxes. Since the pipes have finite volume, and hence finite extent, and since the water they carry is not mixed, there is a finite transit time associated with the transport of heat and salt. Hence, the model takes into account the fact that at any instant in time the water emptying out of one of the pipes does not necessarily have the same temperature and salinity as the water entering the other side of the pipe. The configuration implies that the rate of transport or flux is determined by pressure gradients between the larger subtropical and subpolar boxes, which interact with the atmosphere via mixed boundary conditions. The system can then be reduced to a two-box system, which differs from the standard Stommel model in that it has delay terms. In the limit of pipe volume going to zero the delay also goes to zero and the model collapses to the standard Stommel model.

In the oceanic setting the transit of water from one region to another one takes a time on the order of a decade or more. This is comparable to the relaxation time for the water temperature in each of the regions or boxes (see, e.g., Zhang 1991; Huang et al. 1992; Marotzke 1996). We do not, however, identify the particular net meridional advective mechanisms that would play the climatological role; this study suggests that advection can play a role in climate variability, presents in detail how such variability would manifest itself, and provides a foundation upon which one can study and characterize the relevance of meridional transport to the thermohaline circulation. These results, of course, are direct consequences of the conceptual model proposed here.

In section 2 we propose a variation on the Stommel model that incorporates an advective time lag in the interbox flow and show that its steady states are the same as those of the Stommel model with mixed boundary conditions. In section 3 we examine the stability of the steady states and find that the advective time lag can lead to oscillatory instabilities of the thermally dominant steady circulation, which is stable in the Stommel model. This result suggests that self-sustained oscilla-

tions about this steady state might occur in the parameter range for which the oscillatory instability sets in. However, in section 4 we present numerical calculations that show that this is not the case. Instead, the instability leads to a transient oscillation, whose amplitude grows until finally the flow simply crosses over from the thermally dominant configuration to the stable, haline dominant steady state. In section 5 we present a nonlinear analysis valid in a particular region of parameter space that shows how this crossover occurs. Moreover, we show that even when the thermally dominant steady flow is linearly stable, there is a finite-amplitude instability that can still take the flow over to the haline dominant configuration and whose threshold amplitude is smaller than the threshold in the original Stommel model. Our results are summarized in section 6.

2. Model with advective time lags

Consider an ocean model with two boxes of equal volume V, with the water in each box having uniform temperature T, and salinity S. Each box exchanges heat with the atmosphere above it at a rate given by Newton's law of cooling. Denote the temperature of the atmosphere above box 1 by T_A and that above box 2 by T_B, with the boxes labeled so that $T_A > T_B$. The atmosphere also contributes freshwater fluxes. These are parameterized by virtual salt fluxes, from box 2 to box 1, at a rate denoted H_S, which we take to be constant. Water density differences between the two boxes generate flow exchanges at a volumetric rate RV through upper and lower pipes. The model described thus far is precisely Stommel's two-box model (1961), except with the restoring boundary conditions replaced by mixed boundary conditions. We now depart from Stommel's model by assuming that the water emptying out of each pipe at time $\tilde{t}$ has the same temperature and salinity as the water that entered that pipe at an earlier time $\tilde{t} - \tilde{\tau}$, rather than being identical to the water that is entering the pipe at time $\tilde{t}$. The flow rate R at time $\tilde{t}$, however, depends on the water densities of the two boxes at that same time. With these assumptions, the temperatures and salinities of the two boxes evolve according to

$$\frac{dT_1}{d\tilde{t}} = \gamma_T(T_A - T_1) + |R|[T_2(\tilde{t} - \tilde{\tau}) - T_1(\tilde{t})],$$

$$\frac{dT_2}{d\tilde{t}} = \gamma_T(T_B - T_2) + |R|[T_1(\tilde{t} - \tilde{\tau}) - T_2(\tilde{t})],$$

$$\frac{dS_1}{d\tilde{t}} = H_S + |R|[S_2(\tilde{t} - \tilde{\tau}) - S_1(\tilde{t})],$$

$$\frac{dS_2}{d\tilde{t}} = -H_S + |R|[S_1(\tilde{t} - \tilde{\tau}) - S_2(\tilde{t})], \tag{1}$$

where γ_T is the thermal relaxation rate. The exchange flow is taken to be driven by hydrostatic pressure differences across the pipes so that its rate R is proportional to the density difference between the two boxes,

$$R = k(\rho_2 - \rho_1). \tag{2}$$

The density of the water in each box is given in terms of its temperature and salinity by a linear equation of state,

$$\rho_i = \rho_0 + \beta_S S_i - \beta_T T_i, \qquad i = 1, 2. \tag{3}$$

Positive R represents a thermally dominant circulation in which water flows along the surface from box 1, which has the warmer atmosphere, to box 2, where it sinks because it is colder. Negative R is a haline dominant circulation, where the water in box 1 is denser because of its greater salinity.

It should be noted that total salt and heat do not seem to be conserved by Eq. (1). This is because the equations only account for the salt and heat *in the boxes* and do not count what is in the pipes. We could also write equations for the total salt and the total heat in each pipe; the resulting set of eight equations would automatically conserve total salt and total heat. However, we omit these extra equations from the model because the salinity and temperature profiles in the pipes do not affect the flow except through the effects already built into (1).

We now define dimensionless temperature and salinity differences between the boxes. They are

$$x \equiv 2k\beta_T\gamma_T^{-1}(T_1 - T_2)$$

$$y \equiv 2k\beta_S\gamma_T^{-1}(S_1 - S_2), \tag{4}$$

respectively. We also define a dimensionless exchange rate

$$q = 2\gamma_T^{-1}R. \tag{5}$$

The model equations then become

$$\dot{x} = \alpha - x - \frac{|q|}{2}[x(t - \tau) + x(t)],$$

$$\dot{y} = \mu - \frac{|q|}{2}[y(t - \tau) + y(t)], \tag{6}$$

where the overdot represents a derivative with respect to the dimensionless time $t = \gamma_T\tilde{t}$ and the time delay is $\tau = \gamma_T\tilde{\tau}$. The temperature forcing is measured by the dimensionless parameter

$$\alpha \equiv 2k\beta_T\gamma_T^{-1}(T_A - T_B) \tag{7}$$

and the salinity forcing by

$$\mu \equiv 4k\beta_S\gamma_T^{-2}H_S. \tag{8}$$

The parameter μ is positive in the physical context. Finally, the hydraulic relation (2) becomes simply

$$q = x - y. \tag{9}$$

Equations (6) and (9) form a closed set since the temperature and salinity *differences* between the boxes, rep-

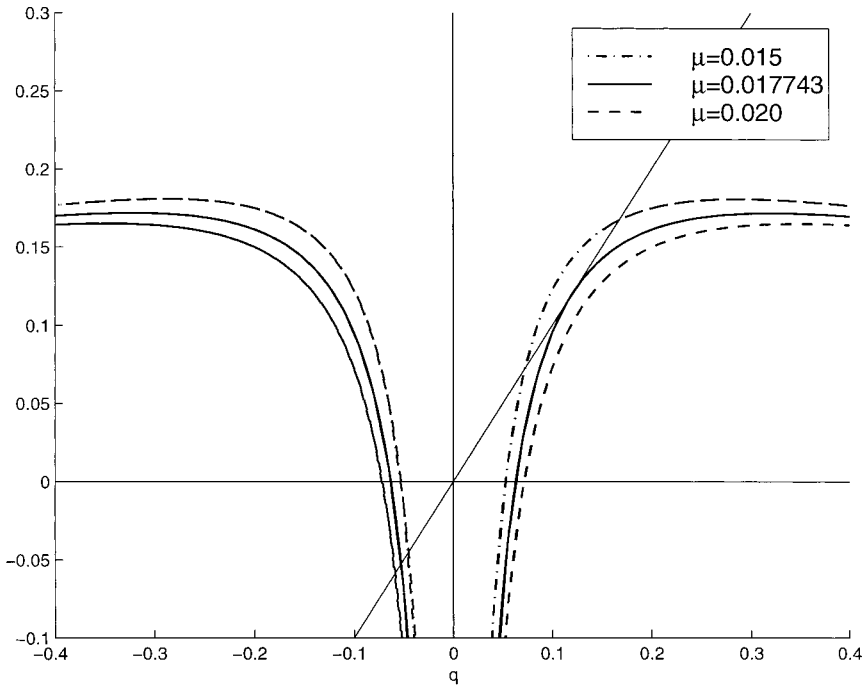

FIG. 1. The two sides of Eq. (12), with $f(q)$ plotted for $\alpha = 0.3$ and $\mu = 0.015$ (dot-dashed), $\mu = 0.017743$ (solid), and $\mu = 0.20$ (dashed). Steady-state solutions of the model exist for q values where $f(q)$ intersects the solid straight line.

resented by x and y, respectively, do not depend on, say, the *averages* of the temperatures or salinities.

The difference between this model and the standard Stommel model (with mixed boundary conditions) is that the exchange terms, those proportional to $|q|$, involve the properties of the water both at the current time t and at a past time $t - \tau$. The former accounts for the effect of the water leaving the box and entering one pipe at time t, the latter for the water emptying into the box from the other pipe at time t, which entered the pipe at $t - \tau$ and so has the temperature and salinity that the other box had at that time. We will investigate two versions of this model, with different assumptions about the time delay τ. In the first version, we will simply take τ to be a constant. In the second, we will think of the length of the pipe as being constant. The flow rate q should be proportional to the velocity of the fluid through the pipe, so integrating it over some time interval would give a result proportional to the distance traveled by the water during that interval. Thus in the second version of the model we will take τ at time t to be the smallest nonzero τ satisfying

$$\int_{t-\tau}^{t} q(t')\, dt' = \pm L \quad \text{or} \quad 0, \tag{10}$$

where L is a constant representing the length of the pipe. Setting the integral to zero is relevant in situations where the flow reverses direction so that the water entering each box actually left the *same* box a time τ earlier. During these periods, the exchange terms in the original Eq. (1) must change, with the $T_2(\tilde{t} - \tilde{\tau})$ term in the T_1 equation being replaced by $T_1(\tilde{t} - \tilde{\tau})$, and so forth. The effect of this is to replace the time-delayed terms $x(t - \tau)$ and $y(t - \tau)$ in the dimensionless equations (6) by their negatives.

Since this model is so close to the Stommel model, it comes as no surprise that it has exactly the same steady states. To find these we first set the time derivatives in (6) to zero and solve for x and y in terms of q, obtaining

$$x_{ss} = \frac{\alpha}{1 + |q|}, \qquad y_{ss} = \frac{\mu}{|q|}, \tag{11}$$

where the subscript denotes "steady state." We then substitute these into the hydraulic equation (9) to find an equation for q,

$$q = f(q) \equiv \frac{\alpha}{1 + |q|} - \frac{\mu}{|q|}. \tag{12}$$

This is a fixed point problem, whose solutions represent the steady states of the model.

The two sides of (12) are plotted in Fig. 1. There is always a solution with negative q for any $\mu > 0$ representing a haline dominant steady state; if μ is not too large, there are also two solutions with positive q, representing thermally dominant states.

From Fig. 1 we see that at the critical value of μ, where thermally dominant steady states first appear, the graphs of the two sides of (12) meet tangentially. We

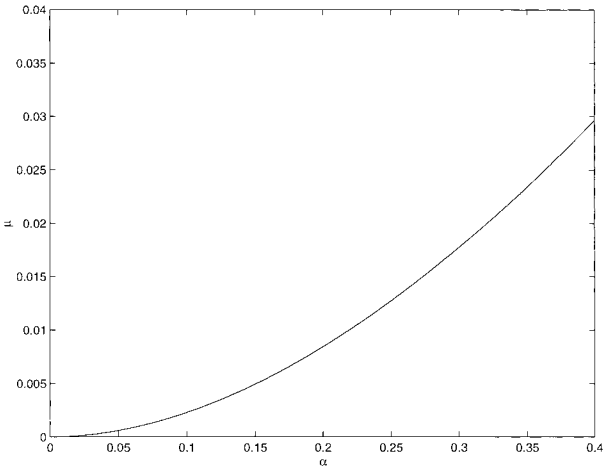

FIG. 2. Equation (13), defining the values of α and μ, which yield a single positive-q fixed point of Eq. (12). For α and μ above this curve, there is a single steady state; below it, there are three.

can find this critical μ by solving (12) and its q derivative simultaneously. This leads to parametric equations for α and μ,

$$\alpha = 2q(1 + q)^2, \qquad \mu = q^2(1 + 2q). \tag{13}$$

These equations define a curve in the α–μ plane, which marks the threshold at which thermally dominant steady states first appear. This curve is plotted in Fig. 2. Above it there is only a single steady state, a haline dominant state with negative q. Below it there are three steady states, the haline dominant one with negative q and two thermally dominant states with positive q. In the next section, we determine the stability properties of each of these steady states.

3. Stability analysis

To test the stability of a state (x_{ss}, y_{ss}) given by (11) and (12), we add perturbations to x and y,

$$x = x_{ss} + \delta x(t), \qquad y = y_{ss} + \delta y(t), \tag{14}$$

substitute into the evolution equations (6), and linearize in δ_x and δ_y. This yields

$$\dot{\delta x} = -\delta x - (\delta x - \delta y)x_{ss}\,\mathrm{sgn}(q_0) - \frac{|q_0|}{2}[\delta x(t - \tau) + \delta x(t)],$$

$$\dot{\delta y} = -(\delta x - \delta y)y_{ss}\,\mathrm{sgn}(q_0) - \frac{|q_0|}{2}[\delta y(t - \tau) + \delta y(t)], \tag{15}$$

where $q_0 = x_{ss} - y_{ss}$ is the flow in the unperturbed steady state.

Since the coefficients in these equations are constants, the solutions will have an exponential dependence on time. This suggests the *ansatz*

$$\delta x = \hat{x}\,\exp[(\sigma + i\omega)t],$$
$$\delta y = \hat{y}\,\exp[(\sigma + i\omega)t], \tag{16}$$

with σ and ω to be determined from the linearized equations. Solutions with positive growth rate σ correspond to unstable steady states. If a solution has nonzero ω, then $(\delta x, \delta y)$ oscillates as its amplitude grows (if σ is positive) or decays (if σ is negative). We generally expect that there will be an oscillatory solution to the full equations near the steady state if we find that the largest σ is small and positive, and the corresponding ω is nonzero.

Substituting (16) into the linearized evolution equations gives

$$0 = [z + 1 + x_{ss}\,\mathrm{sgn}(q_0)]\hat{x} - x_{ss}\,\mathrm{sgn}(q_0)\hat{y},$$
$$0 = y_{ss}\,\mathrm{sgn}(q_0)\hat{x} + [z - y_{ss}\,\mathrm{sgn}(q_0)]\hat{y}, \tag{17}$$

with

$$z \equiv \sigma + i\omega + \frac{|q_0|}{2}(1 + e^{-(\sigma + i\omega)\tau}). \tag{18}$$

In order for this pair of homogeneous linear equations to have a nontrivial solution, the determinant of the matrix of coefficients must vanish. This requirement yields the equation that determines z for a given steady state:

$$z^2 + (1 + |q_0|)z - \frac{\mu}{q_0} = 0. \tag{19}$$

If we are linearizing about a haline dominant steady state, so that q_0 is negative, then all coefficients in this quadratic equation are positive, so both solutions z are either negative or complex with negative real parts. If we are linearizing about a thermally dominant ($q_0 > 0$) steady state, then the constant term in the quadratic is negative, so one solution is negative and the other is positive. In this latter case it is not difficult to show that the positive root is, in fact, the flow rate in the *other* thermally dominant steady state. That is, the two positive solutions of Eq. (12) for q, which are the exchange flows in the two thermally dominant steady states, are q_0 and z.

From (18) we see that the real and imaginary parts of z are

$$\mathrm{Re}\{z\} = \sigma + \frac{|q_0|}{2}(1 + e^{-\sigma\tau}\cos\omega\tau),$$
$$\mathrm{Im}\{z\} = \omega - \frac{|q_0|}{2}e^{-\sigma\tau}\sin\omega\tau. \tag{20}$$

It is easy to see from the first of these that, whenever σ is positive, the real part of z is positive. Stating this the other way around, the steady state cannot be unstable unless the quadratic (19) has a solution z with positive real part. We know that this does not happen when q_0 is negative, so we conclude that the haline dominant steady state is always linearly stable.

We see then that only the thermally dominant, positive-q_0 steady states can be unstable. Therefore, we will henceforth restrict our attention to these states. Furthermore, it is only the positive root z of (19) that can result in an instability, and this z is the exchange flow rate in the other thermally dominant steady state that occurs at the given values of the control parameters α and μ. Since this z is real and positive, the second of (20) must then be zero.

According to (20), one way for z to be purely real is for ω to be zero. Perturbations of this type do not oscillate as they grow or decay. In this case, the expression (18) for z simplifies to

$$z = \sigma + \frac{q_0}{2}(1 + e^{-\sigma\tau}). \tag{21}$$

For $\sigma = 0$ the right side of this equation takes the value $z = q_0$, and it goes to $+\infty$ for large positive σ. Therefore if z is greater than q_0, there is some perturbation with real, positive growth rate σ. For given control parameter values α and μ, the two thermally dominant steady states have flow rates q_0 and z, so we have just shown that the one with the smaller flow rate is always unstable. This is the same result one finds in the original Stommel model: There the thermally dominant steady state with smaller q is the saddle point whose stable manifold forms the boundary separating initial conditions that evolve toward the haline dominant steady state from those that go to the other, stable, temperature dominant steady state. Here, with an advective time delay included, this aspect of the model is seen to persist.

We can use (20) with $\mathrm{Im}\{z\} = 0$ to map out the regions of the α–μ parameter space in which the thermally dominant steady state with the *larger* exchange flow, the one which is stable in the original Stommel model, becomes unstable due to the advective time delay. We are looking for those values of α and μ for which (20) yields a positive growth rate σ when q_0 and z are respectively the larger and smaller positive solutions of the steady-state equation (12) for the exchange flow q. To find these values, we regard σ and ω as parameters, and use (20) to find the corresponding q_0 and z,

$$q_0 = \frac{2}{\tau}e^{\sigma\tau}\frac{\omega\tau}{\sin\omega\tau},$$

$$z = \sigma + \frac{1}{\tau}(e^{\sigma\tau} + \cos\omega\tau)\frac{\omega\tau}{\sin\omega\tau}. \tag{22}$$

We then use (19) and (20) to write α and μ parametrically in terms of q_0 and z,

$$\alpha = (1 + q_0)(1 + z)(q_0 + z),$$

$$\mu = q_0 z(1 + q_0 + z). \tag{23}$$

For the resulting values of α and μ, there is a thermally dominant steady state with flow rate q_0, and infinitesimal perturbations about that steady state grow or decay with linear growth rate $\sigma + i\omega$, with σ and ω being parameter values we started with. For nonzero ω, perturbations oscillate as they grow or decay; it is not difficult to show from (22) that for nonzero ω and positive σ, we always have $q_0 > z$. For $\omega = 0$ we need to be a bit more careful since the second of (20) is then automatically satisfied, and instead of (22), we get (21) from the first of (20). We then use q_0 and σ as parameters. The result is simply that the part of the α–μ plane in which the thermally dominant steady state with the larger flow rate is linearly unstable is just the region between the curve (13), at which thermally dominant steady states cease to exist, and the curve obtained from (22) and (23) with $\sigma = 0$.

In terms of $\theta \equiv \omega\tau$, this latter curve is given by

$$q_0 = \frac{2}{\tau}\frac{\theta}{\sin\theta}, \qquad z = q_0\frac{1 + \cos\theta}{2},$$

$$\alpha = (1 + q_0)(1 + z)(q_0 + z),$$

$$\mu = q_0 z(1 + q_0 + z). \tag{24}$$

This region is plotted in Fig. 3. The stability boundary (24) meets the threshold of thermally dominant solutions, (13), tangentially at the point given by $q_0 = 2/\tau$, and for large α it is given by $\mu \approx (\pi/\tau)^2\sqrt{\alpha}$.

The curve marking the boundary between stable and unstable steady states is given by (24) with θ ranging from 0 to π. The instability that sets in for states on this boundary is oscillatory with $\omega = \theta/\tau$, so the period of oscillation $2\pi/\omega = (2\pi/\theta)\tau$ ranges from 2τ to infinity. Therefore, if the time delay in the model is on the order of decades, then the resulting oscillations have periods of decades or longer.

Note that nothing restricts the values of θ to be below π. Values between π and 2π are excluded since these would yield negative values for q_0 (which is restricted to be positive in this analysis), but the range from 2π to 3π is allowed, as is the range from $2n\pi$ to $(2n + 1)\pi$ for any integer n. These curves all lie within the unstable region—steady states beyond these curves have multiple oscillatory instabilities.

In all the above calculations we have implicitly assumed that τ is constant. However, very little changes if we use the other variant of the model, in which L is constant. This is because the steady-state values of x and y are independent of τ. Thus, if we take τ to be a function of time given implicity by (10) with a constant L, this introduces no additional terms into the linearized equations for δx and δy. The rest of the calculation is therefore unchanged. We need only replace τ everywhere by its steady-state value, which, from (10), is just L/q_0. The first equation in (24) then reduces to

$$\frac{\sin\theta}{\theta} = \frac{2}{L}, \tag{25}$$

whose solutions determine $\omega = \theta q_0/L$. Note that there is no solution for $L < 2$; instabilities cannot occur unless

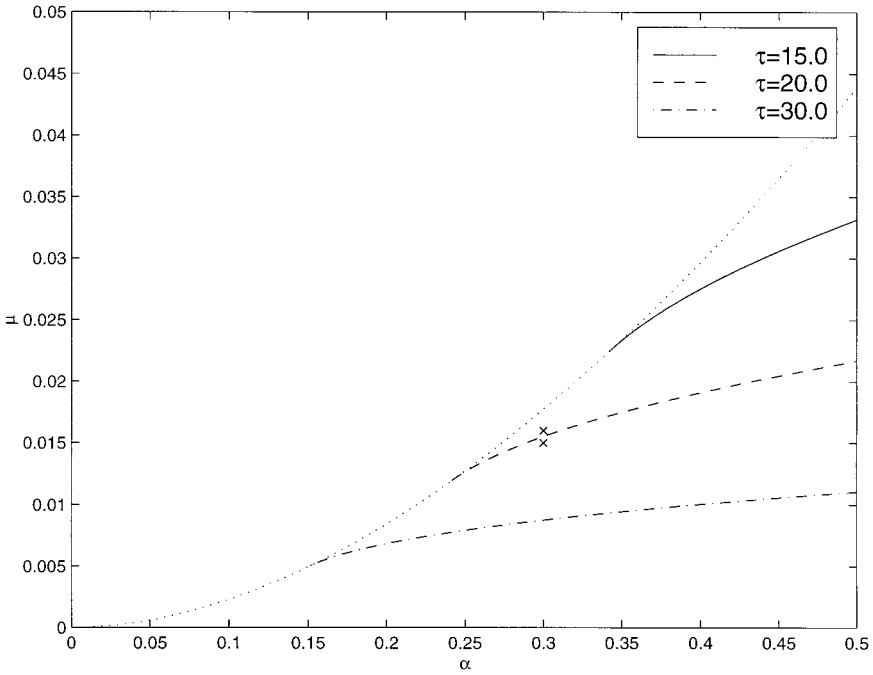

FIG. 3. The stability boundary given by (24) for several values of τ. Above the dotted line, no thermally dominant steady states exist. Thermally dominant steady states between the dotted line and the stability boundary are linearly unstable. The crosses mark points for which numerical calculations are presented in section 4.

L is large enough. For L above 2, there are only a finite number of possible values of θ, and the oscillation frequencies of perturbations about a steady state with flow rate q_0 are proportional to q_0. The proportionality constant is completely determined by the parameter L—it is independent of the temperature and salinity forcing parameters α and μ. The equations for the stability boundary then reduce to

$$\mu = q_0^2 \cos^2(\theta/2)\{1 + q_0[1 + \cos^2(\theta/2)]\},$$

$$\alpha = q_0 \cos^2(\theta/2)(1 + q_0)[1 + q_0 \cos^2(\theta/2)], \quad (26)$$

where θ is fixed and q_0 is used as a parameter. These boundaries are plotted in the α–μ plane in Fig. 4. Note that for $L > 2$, as we increase μ from small values, the thermally dominant steady state always suffers an oscillatory instability before it merges with the saddle state and disappears. In this respect, the behavior of the two versions of the model differs: we see from Fig. 3 that, when τ is fixed, the thermally dominant steady state remains stable as μ is increased until it disappears, provided α is small enough.

4. Model results

To explore the behavior of the model beyond the linear stability analysis of the preceding section, we present here the results of numerical integration of the model equations (6) and (9). The following examples were calculated using a straightforward implementation of a second-order explicit Runge–Kutta scheme. The scheme was checked for convergence, with and without delay terms.

We fixed $\alpha = 0.3$ and $\tau = 20$. For this value of α, thermally dominant steady states exist when μ is below 0.0177, and with the given value of τ, (24) puts the stability boundary at $\mu = 0.0155$. We show calculations for $\mu = 0.015$ and 0.016; these points are marked by crosses on the stability diagram, Fig. 3.

For reference, we show in Fig. 5 the well-known outcome when there is no delay, $\tau = 0$. The value of μ is 0.015. From (11) and (12) we find that the haline dominant steady state is at $(x, y) = (0.2870, 0.3322)$ with $q = -0.0452$, the stable thermally dominant steady state is at $(x, y) = (0.2570, 0.0896)$ with $q = 0.1674$, and the unstable thermally dominant steady state is at $(x, y) = (0.2798, 0.2075)$ with $q = 0.0723$. We used a time step of 0.25; halving the time step did not change the plots perceptibly. Initial conditions were taken along the line in the x–y phase plane, which connects the two thermally dominant steady states, 70%, 99%, and 101% of the way from the attractor to the saddle, in order from top to bottom. Note that the system settles down to a steady state in a time of order 50.

Since the right sides of the delay-differential equations depend on the values of x and y at a finite time in the past, initial conditions need to be specified for the entire time interval of length τ before the start of the calculation. We have simply taken them to be con-

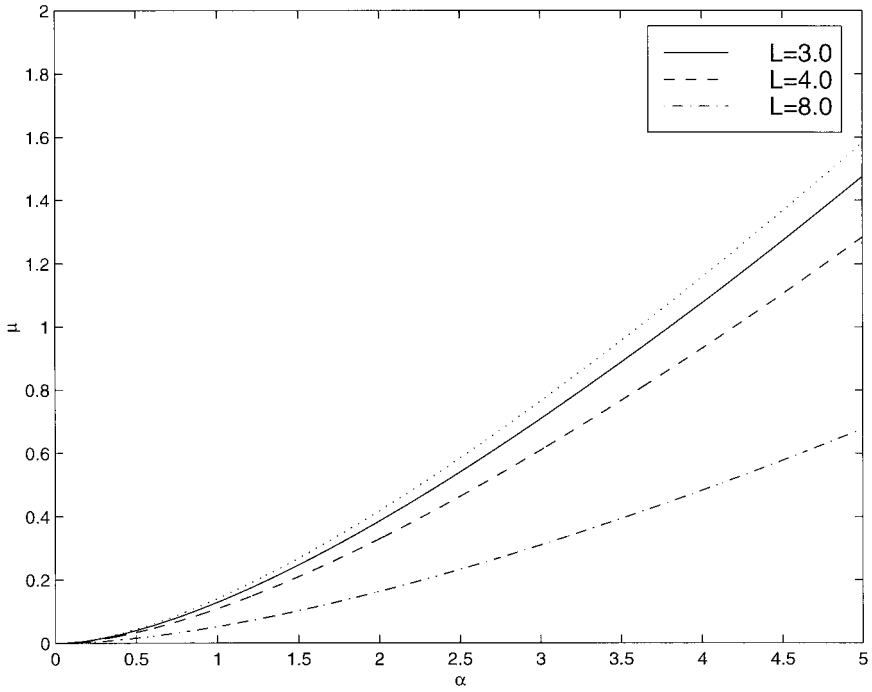

FIG. 4. The stability boundary given by (26) for the constant-L version of the model, for several values of L. The dotted line is the curve above which thermally dominant steady states do not exist. Thermally dominant steady states between it and the stability boundary are unstable.

stant before $t = 0$. In Fig. 6 we show the result of a calculation with $\tau = 20$ and $\mu = 0.015$, marked by the lower cross on Fig. 3. This is within the range in which the steady state with $q = 0.1673$ should be linearly stable. The initial x and y are taken to be along the line joining the two thermally dominant steady states, 30% of the way from the stable state to the saddle. We see the expected decaying oscillations. Note the different timescale between this run and the previous one without a time delay. The oscillations have a period of about 70, or 3.5τ, and decay slowly over times of order 1000. Solving equations (22) numerically for the linear growth rate of perturbations about the $q = 0.1673$ steady state gives $\sigma = -0.00096$ and $\omega = 0.0847$, so the linear stability analysis predicts that the period of oscillation should be $2\pi/\omega = 74.2$, and the time for the amplitude to decay by a factor of 2 should be $(\ln 2)/|\sigma| = 722$, consistent with the numerical results shown. The time

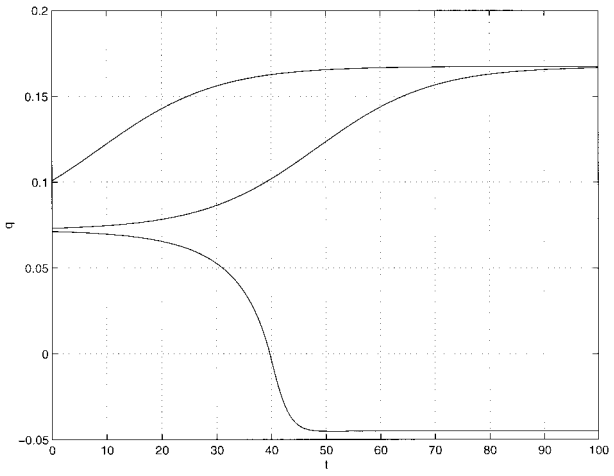

FIG. 5. Evolution of the flow rate for the model with no delay. The upper traces show the system reaching a thermally dominant state, the bottom trace a haline dominant state.

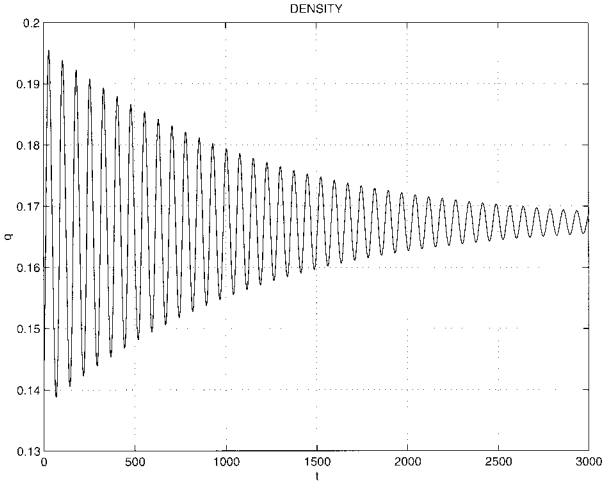

FIG. 6. Evolution of the flow rate for the model with delay, in the regime where the thermally dominant steady state is linearly stable.

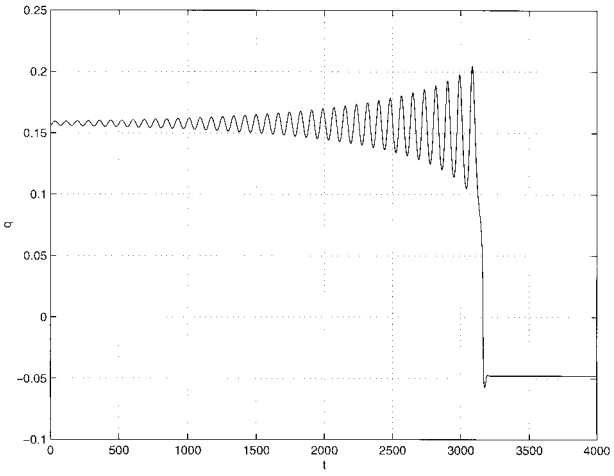

FIG. 7. Evolution of the flow rate for the model with delay, in the regime where the thermally dominant steady state is linearly unstable.

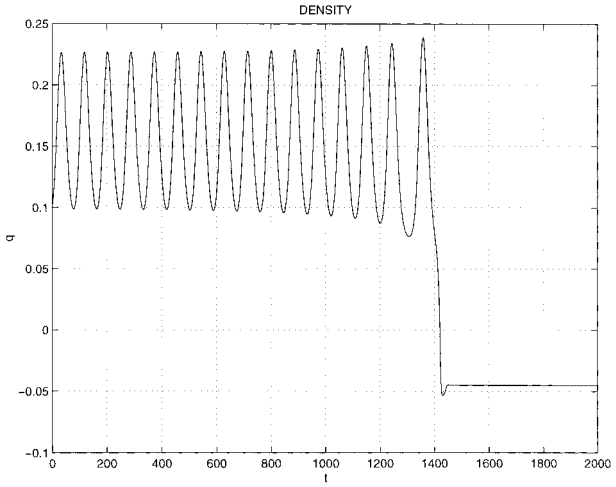

FIG. 8. Evolution of the flow rate for the model with delay, in the regime where the thermally dominant steady state is linearly stable. Here, the initial perturbation is large enough to make the system cross over to the haline dominant state.

step is 0.25; again, halving it makes no discernable difference to the plot.

Figure 7 shows the behavior for $\mu = 0.016$, marked by the upper cross in Fig. 3. This is in the unstable regime. Initial conditions were close to the thermally dominant steady state with the larger flow rate q. The time step was 0.1; halving this does not affect the plot. As expected, we see oscillations about the previously stable steady state that grow in amplitude. The linear stability result (22) predicts that the oscillations should have a period $2\pi/\omega = 81.4$, and the doubling time for the amplitude should be $(\ln 2)/\sigma = 685$. Instead of settling down to a self-sustained oscillation, however, the amplitude of the oscillation continues to grow until finally the flow rate drops below its value in the saddle state. At this point the oscillation abruptly stops, and the system quickly switches over to the haline dominant steady state. Thus although we might have expected the oscillatory instability found in the preceding section to be a sign that the model will make a transition from a steady state to an oscillatory state, we see that this is not what happens. The oscillations are a long-lived transient, presaging a transition from one steady state to another.

Another aspect of this behavior is shown in Fig. 8. Here we return to $\mu = 0.015$, where the thermally dominant steady state is linearly stable. However, we change the initial conditions from those used in Fig. 6: rather than starting 30% of the way to the saddle, we start 70% of the way there. Now instead of decaying back to zero, the amplitude of the oscillations about the linearly stable steady state actually grows, until finally the system crosses over to the haline dominant steady state, as it did when μ was in the unstable regime. Evidently, even though the thermally dominant steady state is stable against *small* perturbations, it is unstable to a finite amplitude perturbation. Instability under large perturbations is also present in the Stommel model: there, if the perturbation from one stable steady state is large enough to force the system into the unstable manifold of the saddle state, the system will indeed cross over to the other steady state. Here, however, the delay has made the threshold amplitude much smaller. Starting at the same initial state (x, y), but without the time delay, the system returns to the thermally dominant state, as was shown in Fig. 5. The time step here was 0.025, which again was small enough that halving the time step made no significant change to the plot.

5. Nonlinear analysis

The linear stability analysis in section 3 identifies the conditions under which a steady state becomes unstable, but it is not capable of predicting what then happens. In order to do this, and to shed light on the numerical results of the preceding section, we must include nonlinear terms in the calculation. We will focus our attention on the vicinity of some thermally dominant steady state given by (11) with $q = q_0 > 0$. In order to make use of the results of the linear stability analysis, we express $x - x_{ss}$ and $y - y_{ss}$ in terms of variables ξ and η, which decouple to *linear* order. Specifically, we write

$$x = \frac{\alpha}{1 + q_0} - (q_0 + z)\xi + (1 + z)\eta,$$

$$y = \frac{\mu}{q_0} - (1 + q_0 + z)\xi + z\eta, \tag{27}$$

with

$$q = x - y = q_0 + \xi + \eta. \tag{28}$$

As in section 3, q_0 and z are the flow rates in the two thermally dominant steady states, with $q_0 > z$, and α

and μ can be written parametrically in terms of q_0 and z via Eq. (23). Substituting these expressions into the governing equations (6), and using the parametric forms of α and μ, then leads to evolution equations for the new variables ξ and η,

$$\dot{\xi} = z\xi(t) - [q_0 + \xi(t) + \eta(t)]\frac{\xi(t) + \xi(t-\tau)}{2}, \quad (29)$$

$$\dot{\eta} = -(1 + q_0 + z)\eta(t) - [q_0 + \xi(t) + \eta(t)]\frac{\eta(t) + \eta(t-\tau)}{2}. \quad (30)$$

These equations are an exact transcription of the original equations of the model; no approximations have yet been made, other than restricting ourselves to the region of phase space for which the flow q is positive.

Being both nonlinear and nonlocal in time, Eqs. (29) and (30) are not analytically tractable. However, the linear stability analysis suggests regimes in which they can be simplified. For example, *near* the stability boundary (24) we know that the amplitude of the deviation from a steady state grows or decays *slowly* since the real part σ of the growth rate is small there. We can exploit this by writing ξ and η as oscillatory functions of time with amplitudes that vary on a slow timescale. We will not present this calculation here, but simply note that it accords with the numerical picture: when the steady state is linearly unstable, nonlinear terms cause the amplitude of the oscillation to grow even faster than the linear terms do; when it is linearly stable, there is still a finite threshold amplitude below which the amplitude decays but above which it grows. Rather, we will continue with a somewhat different calculation that provides a more detailed picture of the dynamics, but is only valid in a narrower range of parameter space.

From (18) we see that both real and imaginary parts of the linear growth rate, σ and ω, are small when both q_0 and z are close to $2/\tau$. Thus, in this parameter range, the whole function $\xi(t)$—not just its amplitude—varies slowly with time. The advantage of working in this parameter range is that the time delay τ is small compared to the timescale on which $\xi(t)$ varies, so the delayed term $\xi(t-\tau)$ can be approximated as a power series in τ. Motivated by these considerations, we now define a small parameter ϵ and a parameter d (of order unity) by

$$q_0 = \frac{2}{\tau}(1 + \epsilon^2), \qquad z = q_0\left(1 - \frac{3}{2}\epsilon^2 d\right). \quad (31)$$

We must now determine how the sizes of our variables ξ and η, and the timescales on which they vary, scale with ϵ when ϵ is small. The parameters ϵ and d directly measure the distance and direction in parameter space from (α, μ) to the point where the stability boundary meets the onset curve where thermally dominant steady states first appear. When ϵ is small, the distance is of order ϵ^2, the curve obtained by setting $d = 0$ and varying ϵ coincides with the onset curve, and the curve with $d = 1$ coincides with the linear stability boundary. Clearly the flow rates in the two thermally dominant steady states, q_0 and z, differ by order ϵ^2; from (11) we also find that the two states are separated by a distance in the x–y phase plane of the same order. Thus we expect ξ and η to have sizes of this order, at least until the crossover to the haline-dominated state occurs. To find the timescale on which the evolution takes place, we examine the linear stability result (18) with q_0 and z given by (31). Since z is equal to q_0 to leading order in ϵ, both σ and ω must be small; by expanding (18) in powers of σ and ω we then find that σ is of order ϵ^2 and ω is of order ϵ. Finally, by examining (30), we see that any deviation of η from zero on the right side of the equation is multiplied by a negative number of order unity. Thus η decays to zero on times of order unity, while ξ will oscillate with a period of order $\omega^{-1} \sim \epsilon^{-1}$, with an amplitude that varies on a timescale of order $\sigma^{-1} \sim \epsilon^{-2}$.

Since η decays on a timescale of order unity, while ξ varies only on slow timescales, η will have dropped essentially to zero by the time ξ starts to change appreciably. Thus, as we follow the oscillations of ξ, η will already be zero. From (27), this implies that—at least for small ϵ—the oscillation will occur along a straight line in the x–y phase plane, given by

$$(1 + q_0 + z)x - (q_0 + z)y - (q_0 + z)(1 + q_0 + z) = 0, \quad (32)$$

where we have used the parametric equations (23) for α and μ. This is the line in the phase plane that connects the two thermally dominant steady states. Even though ϵ is not small in the numerical calculations presented in the preceding section, this result still holds to rather good precision. For example, using the data plotted in Fig. 8, we find that the expression on the left always remains below 0.001 until the crossover to the haline dominant steady state occurs, at which point it jumps to about -0.6.

For the present analysis, the fact that η decays so rapidly means that we may drop it from the ξ equation:

$$\dot{\xi} = z\xi(t) - [q_0 + \xi(t)]\frac{\xi(t) + \xi(t-\tau)}{2}. \quad (33)$$

We now implement the scalings deduced for ξ by writing

$$\xi(t) = -\epsilon^2 q_0 f(2\epsilon t/\tau). \quad (34)$$

The time-delayed term in the ξ equation then becomes

$$\xi(t-\tau) = -\epsilon^2 q_0 f(2\epsilon(t-\tau)/\tau) = -\epsilon^2 q_0[f(2\epsilon t/\tau) - 2\epsilon f'(2\epsilon t/\tau) + \cdots], \quad (35)$$

where the prime denotes a derivative with respect to the slow time variable $2\epsilon t/\tau$. Substituting this expression

for ξ and the expressions (31) for q_0 and z into the evolution equation for ξ, and expanding in powers of ϵ, we get

$$f'' = -\frac{3}{2}df + f^2 + \epsilon\left(f' - ff' + \frac{2}{3}f'''\right). \quad (36)$$

We then substitute the derivative of the right side for the f''' on the right side and again expand in powers of ϵ, and so obtain

$$f'' = -\frac{3}{2}df + f^2 + \epsilon\left(1 - d + \frac{1}{3}f\right)f'. \quad (37)$$

To leading order in ϵ, (37) describes oscillations about the steady state $f = 0$, and the interaction between this steady state and the other thermally dominant steady state—Stommel's saddle point—at $f = 3d/2$. The phase portrait in the $(f, p \equiv f')$ plane is shown in Fig. 9. The unperturbed trajectories may be written in the form

$$p^2 = 2E - \frac{3}{2}df^2 + \frac{2}{3}f^3, \quad (38)$$

where E is a constant of integration used to label the trajectories. The fixed point at $f = 0$ has $E = 0$, and the fixed point at $f = 3d/2$ has $E = 9d^3/16$.

The overall effect of the order-ϵ correction in (37) is to break the closed trajectories seen on the $\epsilon = 0$ phase portrait. As is shown in the appendix, the perturbed trajectories spiral outward from the fixed point at $f = 0$ for

$$\int \left(1 - d + \frac{1}{3}f\right)\sqrt{2E - \frac{3}{2}df^2 + \frac{2}{3}f^3}\, df > 0, \quad (39)$$

where the integral runs between the two points where the unperturbed trajectory crosses the f axis. The integral can be evaluated in terms of elliptic integrals, but it is much more instructive to simply calculate it numerically. For d between 0 and 1, the range for which the thermally dominant steady state is linearly unstable, we find that the integral is positive for all E. This means that all trajectories spiral outward, including the one that passes through the saddle point at $f = 3d/2$. Thus in this range, the oscillatory instability of the fixed point at $f = 0$ does not lead to sustained oscillations. Instead, in the neighborhood of this fixed point an oscillation will be observed whose amplitude grows until the system crosses over to the (stable) haline dominant steady state.

When d is increased slightly above 1, the fixed point at $f = 0$ becomes linearly stable. However, the trajectories with larger amplitudes continue to spiral outward. There is a single trajectory that remains closed; it is an *unstable* periodic orbit that marks the threshold of a finite-amplitude instability of the critical point at $f = 0$. That is, in the very immediate neighborhood of $f = 0$ and closer than this trajectory, the evolution will lead back to $f = 0$, but farther out than this trajectory it will spiral even farther away until it crosses over to the haline dominant steady state. Of course, in the original Stommel model the thermally dominant steady state also has a finite-amplitude instability: if a fluctuation is large enough to reduce q below its value in the saddle-point state, this will cause the flow to cross over to the haline dominant steady state. Including an advective time delay in the model reduces the threshold of this instability when d is near 1. As d is increased further, the threshold amplitude increases until finally at $d = 14/13 = 1.077$ the orbit passing through the saddle state at $f = 3d/2$ starts to spiral inward.

For the constant-L version of the model the picture is substantially the same. Here the time evolution is slow for L slightly greater than 2. Thus we write

$$L = 2(1 + \epsilon^2) \quad (40)$$

and continue to use the expression in (31) for z. Writing

$$\xi(t) = -\epsilon^2 q_0 f(\epsilon q_0 t) \quad (41)$$

and carrying out the same calculations as for the constant-τ version of the model yields almost the same evolution equation for f:

$$f'' = -\frac{3}{2}df + f^2 + \epsilon\left(1 - d + \frac{4}{3}f\right)f'. \quad (42)$$

As before, all trajectories spiral outward for $0 < d < 1$, so when the $f = 0$ steady state is linearly unstable, the instability causes a crossover to the haline dominant configuration rather than self-sustained oscillations. For $d > 1$ the $f = 0$ steady state is linearly stable, but it is surrounded by an unstable periodic orbit that marks the threshold of a finite-amplitude instability, which again takes the flow over to the haline dominant state. The amplitude of this unstable periodic orbit becomes larger as d is increased until it contains the $f = 3d/2$ steady state for $d = 7/5$.

To see how the oscillatory instabilities lead to a crossover from the thermally dominant steady state to the haline dominant one, note that the flow is given by $q = q_0(1 - \epsilon^2 f)$. Thus as f oscillates about 0, the flow remains positive but alternately strengthens and weakens. As the amplitude of the oscillation grows, ultimately the flow weakens beyond the flow $q_0(1 - 3\epsilon^2 d/2)$ of the saddle-point steady state. At this point it crosses over to the haline dominant state, just as in the original Stommel model.

6. Discussion

We have investigated the consequences of including an advective time lag in a simple two-box Stommel model with mixed boundary conditions. We find that including a time lag leaves the haline-dominant steady state of the model stable, while the thermally dominant steady state can have an oscillatory instability. The latter state is stable in Stommel's model. Instead of leading

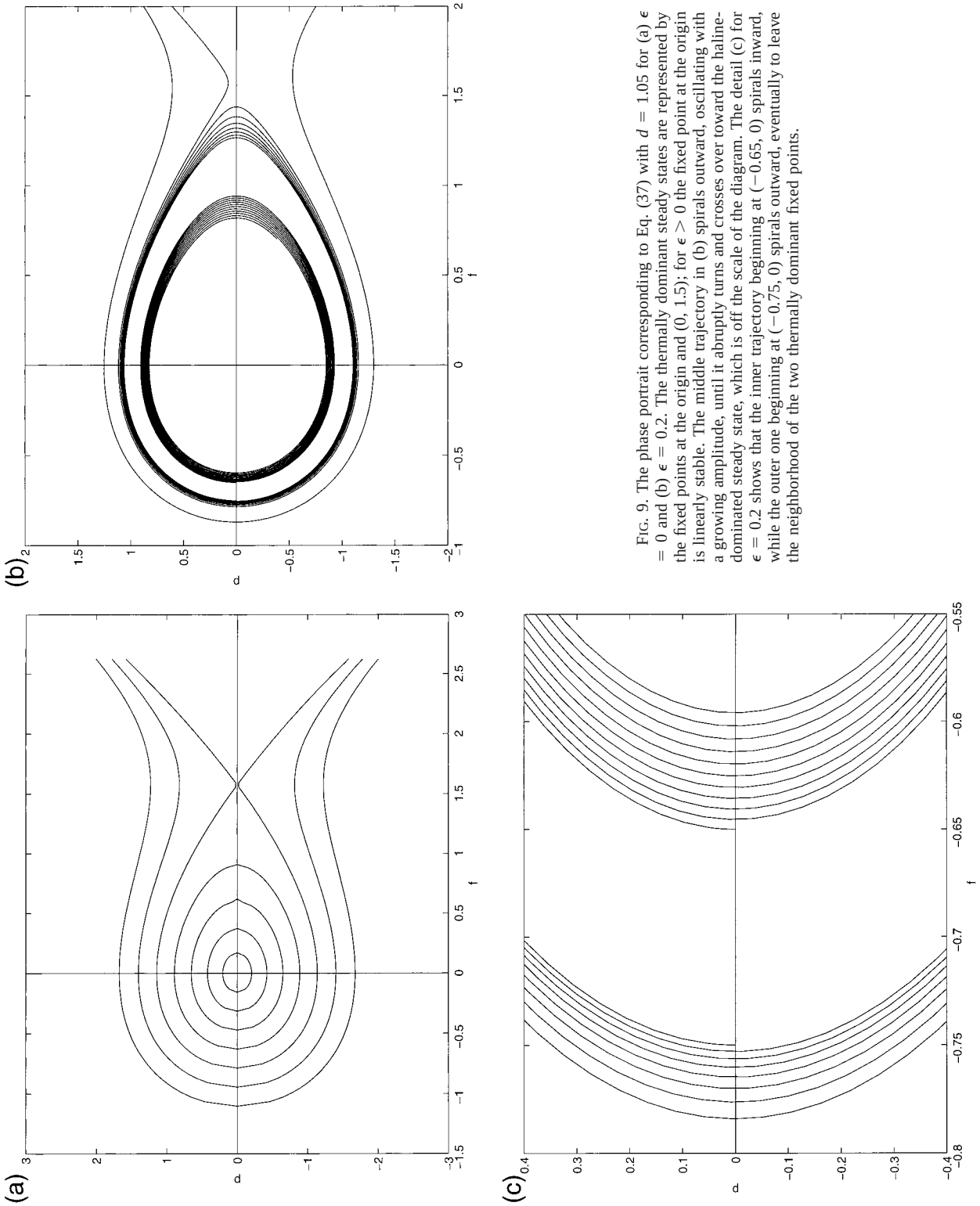

FIG. 9. The phase portrait corresponding to Eq. (37) with $d = 1.05$ for (a) $\epsilon = 0$ and (b) $\epsilon = 0.2$. The thermally dominant steady states are represented by the fixed points at the origin and (0, 1.5); for $\epsilon > 0$ the fixed point at the origin is linearly stable. The middle trajectory in (b) spirals outward, oscillating with a growing amplitude, until it abruptly turns and crosses over toward the haline-dominated steady state, which is off the scale of the diagram. The detail (c) for $\epsilon = 0.2$ shows that the inner trajectory beginning at (−0.65, 0) spirals inward, while the outer one beginning at (−0.75, 0) spirals outward, eventually to leave the neighborhood of the two thermally dominant fixed points.

to sustained oscillations, this instability simply makes the circulation cross over to the stable haline-dominant pattern. Even in part of the parameter range for which the thermally dominant state remains linearly stable, the time lag leads to a finite-amplitude instability, so that a relatively small—but not infinitesimal—perturbation about the thermally dominant state suffices to switch the circulation to the haline-dominant state.

In the language of bifurcation theory, the instability that we have uncovered is an inverted Hopf bifurcation. That is, for some range of a control parameter the system has a linearly stable steady state surrounded by an unstable periodic orbit. As the control parameter is varied, the periodic orbit shrinks around the steady state until, at some critical value, the periodic orbit merges with the steady state and disappears, leaving the steady state linearly unstable. This is in contrast with the more common forward Hopf bifurcation. There, as the control parameter is varied through its critical value, a stable steady state loses its stability and turns into a stable periodic solution.[1]

One way to view the delay terms in the model is as accounting for the fact that one should regard the ''pipes'' in the Stommel model as having finite length and volume so that the water emptying out of either pipe into one of the boxes at any given time does not necessarily have the same temperature and salinity as the water entering the other end of the pipe from the other box at that same time. In the oceanic setting, the transit of water between the regions represented by the two boxes takes a time on the same order—years to decades—as the relaxation time for water temperature in the boxes, so there is clearly the potential for the delay to be important.

In the Stommel model the reservoirs contain well-mixed fluids and these are connected by pipes at the top and at the bottom, through which excess pressure due to density differences in the boxes generate fluxes. Since the fluid is incompressible, any flux through one of the pipes is counteracted by a flux through the other pipe. However, since the pipes have zero volume, the consequences of density differences are instantaneous. In the model presented here, however, the pipes have finite volume and thus take into account the finite time that it might take for density differences to be finally realized. In its simplest guise the pipes do not interact with the rest of the system except, as in the Stommel model, to convey fluxes to and from the larger subpolar and subtropical boxes. Hence, it is suggested that the delay mechanism models unspecified meridional advective processes. Furthermore, the delay time is not measured absolutely but relatively: as a result of the conservation of salt and temperature, it represents the difference in time for these two reservoirs to reach mixed conditions. The delay model thus brings into the dynamics of the circulation, as modeled by box models, a second timescale, namely, the mixing timescale. In this setting the standard box model is seen as the case in which mixing effectively occurs instantaneously when compared to the timescales of the circulation.

Connections between the advective time delay model presented here and other models are sure to be found: when we examined the four-box model considered by Tziperman et al. (1994) and in (Griffies and Tziperman 1995) in the *unstable* oscillatory regime, we found that it behaves much as our model does. That is, in the oscillatory regime solutions do not exhibit a self-sustained oscillation, but simply cross over from the thermally dominant steady state to the haline dominant one. We would also expect to find in the model by Tziperman et al. (1994) the existence of a finite-amplitude instability in the *linearly* stable, oscillatory regime. Stochastic forcing of this model in this regime should eventually produce a fluctuation large enough to take the model past the threshold perturbation amplitude and so cause a crossover to the haline dominant state; the waiting time for this event should decrease as one moves the parameters closer to the onset of the oscillatory instability.

Acknowledgments. This paper benefited greatly from suggestions given by the anonymous reviewers. We are particularly grateful to them for suggestions that improved the climatological context of this work.

APPENDIX

We wish to determine whether $x(t)$ oscillates with a growing or decaying amplitude when x is given by the differential equation.

$$\ddot{x} = F(x) + \epsilon G(x, \dot{x}) + O(\epsilon^2). \tag{A1}$$

This equation may be rewritten as a first-order system,

$$\dot{x} = v, \qquad \dot{v} = F(x) + \epsilon G(x, v) + O(\epsilon^2). \tag{A2}$$

For $\epsilon = 0$, this problem is analogous to the Newtonian equation of motion of a particle in a one-dimensional potential $U(x)$. The energy integral is

$$\frac{1}{2}v^2 + U(x) = E, \tag{A3}$$

with

$$F(x) = -dU/dx. \tag{A4}$$

The energy E is a constant of motion that can be used to label the trajectories in the (x, v) plane. For $\epsilon = 0$, these trajectories are closed curves given by

$$v = \pm v_0(x) \equiv \pm\sqrt{2(E - U(x))}. \tag{A5}$$

Periodic solutions of the leading-order equations are represented by closed trajectories that run between turn-

[1] Applying the methods of section 5 and the appendix to the nonlinear ENSO model of Battisti and Hirst (1989) shows that their model does, in fact, have a forward Hopf bifurcation. This agrees with their numerical results, which show sustained oscillations.

ing points, zeros of $E - U(x)$, which we denote $x_-(E)$ and $x_+(E)$. For definiteness, we take $x_+(E) > x_-(E)$. When x is between these two turning points, $E - U(x)$ is positive so that $v_0(x)$ is real.

We now examine the effect of the order-ϵ term in (A1), focusing on the trajectory that passes through one particular point on the x axis, which is $x_-(E)$ for some energy E. In general, the perturbation can affect the upper and lower halves of the trajectory in different ways. We write the perturbed trajectory to first order in ϵ as

$$v = \pm v_0(x) + \epsilon f_\pm(x)/v_0(x), \tag{A6}$$

where f_+ and f_- are two functions of x that must be calculated; in order for the perturbed trajectory to pass through the point $[x_-(E), 0]$, it must be true that

$$f_+(x_-(E)) = f_-(x_-(E)) = 0. \tag{A7}$$

[In fact, f_+ and f_- must go to zero as x goes to $x_-(E)$ faster than v_0 does.] To derive the equation that $f_\pm$ must satisfy, we square (A6) and differentiate with respect to x; substituting in for v_0 yields

$$v\frac{dv}{dx} = \frac{d}{dx}\left(\frac{1}{2}v_0^2\right) + \epsilon\frac{df_\pm}{dx} = F(x) \pm \epsilon\frac{df_\pm}{dx}. \tag{A8}$$

But dv/dx is $\dot{v}/\dot{x}$, which is $\dot{v}/v$, so from (A1) we have

$$v\frac{dv}{dx} = \dot{v} = F(x) + \epsilon G(x, \pm v_0(x)) \tag{A9}$$

to order ϵ. Comparing these two expressions for $v\,dv/dx$, we see that $f_\pm$ must satisfy

$$\frac{df_\pm}{dx} = \pm G(x, \pm v_0(x)). \tag{A10}$$

Once (A10) is solved with the initial conditions (A7) for the functions $f_-$and f_+, the shape of the perturbed trajectory (A6) to first order in ϵ is obtained. The upper and lower parts of this perturbed trajectory will have zeros near the turning point $x_+(E)$ of the unperturbed trajectory, which can be located by solving

$$\begin{aligned} 0 = v^2 &= v_0^2(x) \pm 2\epsilon f_\pm(x) \\ &= 2[E - U(x) \pm \epsilon f_\pm(x)]. \end{aligned} \tag{A11}$$

Now suppose $E - U(x)$ has a *simple* zero at $x_+(E)$. (The generalization to a zero of higher multiplicity is straightforward.) Then, near $x_+(E)$, we have $E - U(x) \approx F[x_+(E)][x - x_+(E)]$. Thus the upper and lower parts of the perturbed trajectory cross the x axis at

$$x = x_+(E) \mp \epsilon f_\pm[x_+(E)]/F[x_+(E)]. \tag{A12}$$

Now follow the perturbed trajectory through one cycle, starting where its lower branch meets the x axis, with x given by the lower signs in (A12). Since we are following the half of the trajectory on which v is negative, x decreases until the trajectory crosses the x axis at $x_-(E)$. Then v turns positive, and we must follow the upper branch as x increases until the trajectory again crosses the x axis at a point given by the upper signs in (A12). If this latter zero is larger than the former, then the perturbed trajectory is not closed, but spirals outward. On the other hand, if the latter zero is smaller than the former, then the trajectory spirals inward. Thus we have found that to first order in ϵ, the trajectory spirals outward for

$$-f_+[x_+(E)]/F[x_+(E)] > f_-[x_+(E)]/F[x_+(E)]. \tag{A13}$$

Now by integrating (A10) for $f_\pm$ with the initial conditions (A7), expressions for $f_\pm[x_+(E)]$ are obtained as integrals of known functions. The condition for the trajectory to spiral outward is then

$$\begin{aligned} \frac{1}{F(x_+(E))}\int_{x_-(E)}^{x_+(E)} \{&G[x, \sqrt{E - U(x)}] \\ &- G[x, -\sqrt{E - U(x)}]\}\, dx < 0. \end{aligned} \tag{A14}$$

If the left-hand side is positive, then the trajectory spirals inward. In the cases that occur in the text, the perturbation G has the special form $G(x, \dot{x}) = G_1(x)\dot{x}$, so the condition for the trajectory to spiral outward is

$$\frac{1}{F(x_+(E))}\int_{x_-(E)}^{x_+(E)} G_1(x)\sqrt{E - U(x)}\, dx < 0. \tag{A15}$$

REFERENCES

Battisti, D. S., and A. C. Hirst, 1989: Interannual variability in a tropical atmosphere–ocean model: Influence of the basic state, ocean geometry and nonlinearity. *J. Atmos. Sci.*, **46,** 1687–1712.

Cessi, P., 1994: A simple box model of stochastically forced thermohaline flow. *J. Phys. Oceanogr.*, **24,** 1911–1920.

Greatbatch, R. J., and K. A. Peterson, 1996: Interdecadal variability and oceanic thermohaline adjustment. *J. Geophys., Res., Oceans*, **101,** 20 467–20 482.

Griffies, S. M., and E. Tziperman, 1995: A linear thermohaline oscillator driven by stochastic atmospheric forcing. *J. Climate*, **8,** 2440–2453.

Huang, R. X., J. R. Luyten, and H. M. Stommel, 1992: Multiple equilibrium states in combined thermal and saline circulation. *J. Phys. Oceanogr.*, **22,** 231–246.

Latif, M., 1998: Dynamics of interdecadal variability in coupled ocean–atmosphere models. *J. Climate*, **11,** 602–624.

——, A. Grötzner, M. Münnich, E. Maier-Reimer, and T. P. Barnett, 1996: A mechanism for decadal climate variability. *Decadal Climate Variability, Dynamics and Predictability*, D. L. T. Anderson and J. Willebrand, Eds., NATO Advanced Study Institute Series, Vol. I 44, 263–292.

Lenderink, G., and R. J. Haarsma, 1994: Variability and multiple equilibria of the thermohaline circulation, associated with deep water formation. *J. Phys. Oceanogr.*, **24,** 1480–1493.

Marotzke, J., 1996: Analysis of thermohaline feedbacks. *Decadal Climate Variability, Dynamics and Predictability*, D. L. T. Anderson and J. Willebrand, Eds., NATO ASI Series, Vol. I 44, Springer-Verlag, 334–378.

Rahmstorf, S., J. Marotzke, and J. Willebrand, 1996: Stability of the thermohaline circulation. *The Warmwatersphere of the North Atlantic Ocean*, W. Krauss, Ed., Borntraeger, 129–157.

Ruddick, B., and L. Q. Zhang, 1996: Qualitative behavior and non-

oscillation of Stommel's thermohaline box model. *J. Climate,* **9,** 2768–2777.

Sarachik, E. S., M. Winton, and F. L. Yin, 1996: Mechanisms for decadal-to-centennial climate variability. *Decadal Climate Variability, Dynamics and Predictability*, D. L. T. Anderson and J. Willebrand, Eds., NATO Advanced Study Institute Series, Vol. I 44, Springer-Verlag, 157–207.

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

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

Suarez, M. J., and P. S. Schopf, 1988: A delayed action oscillator for ENSO. *J. Atmos. Sci.,* **45,** 3283–3287.

Tziperman, E., J. R. Toggweiler, Y. Feliks, and K. Bryan, 1994: Instability of the thermohaline circulation with respect to mixed boundary conditions: Is it really a problem for realistic models? *J. Phys. Oceanogr.,* **24,** 217–232.

Welander, P., 1982: A simple heat-salt oscillator. *Dyn. Atmos. Oceans*, **6,** 233–242.

Zhang, L. Q., 1991: Simple models of thermohaline processes in the ocean. M.S. thesis, Dalhousie University, Halifax, NS, Canada, 178 pp.