Two-Way Feedback Interaction between the Thermohaline and Wind-Driven Circulations

DOUGLAS G. MACMYNOWSKI

Control and Dynamical Systems, California Institute of Technology, Pasadena, California

ELI TZIPERMAN

Earth and Planetary Sciences, and Division of Engineering and Applied Sciences, Harvard University, Cambridge, Massachusetts

(Manuscript received 22 October 2004, in final form 29 September 2005)

ABSTRACT

The thermohaline circulation (THC) affects the meridional atmospheric temperature gradient and therefore the atmospheric wind and the wind-driven ocean circulation. The wind-driven circulation (WDC), in turn, affects the THC by the advection of salinity anomalies into deep-water formation sites. This paper considers this two-way coupling between the WDC and THC using a simple box-type model and analysis tools from engineering feedback control. The two-way feedback can have a significant effect on the dynamics of the coupled system. For a reasonable choice of parameters, the feedback destabilizes the THC equilibrium for low freshwater forcing. For higher freshwater forcing, the feedback results in a new stable equilibrium instead of the large amplitude oscillation that develops without feedback. It is expected that the analysis approach used here may be broadly applicable to the study of feedback interconnections of other climate systems as well.

1. Introduction

The thermohaline circulation (THC) in the North Atlantic Ocean transports significant heat northward, and its variability has a significant impact on climate. The THC and its variability on interdecadal time scales have been studied using a hierarchy of models, from the Stommel box model (Stommel 1961) to full ocean–atmosphere general circulation models (e.g., Dijkstra 2005, and references therein). Similarly, variability in the wind-driven upper-ocean circulation (WDC) can also influence climate, and has been analyzed at depth (Berloff and Meacham 1997; Cessi and Ierley 1995; Dijkstra and Katsman 1997; Ghil et al. 2002; Meacham 2000; Nadiga and Luce 2001; Primeau 2002, 1998). The WDC and THC have mostly been analyzed separately, although there have been several recent analyses of the coupling between them. Pasquero and Tziperman (2004, hereinafter PT04) have analyzed the one-way interaction between the two by analyzing the effects of a fixed WDC on the variability of the THC. Van Veen (2003) has developed a model incorporating a fixed two-way feedback between a model of the atmosphere and a thermal overturning circulation but did not include full THC dynamics. Stommel and Rooth (1968) also added wind forcing to a simple box model to explore the dynamics of flow driven by competing bouyancy and wind stress forces.

The WDC and THC are fully coupled as a feedback system. The WDC advects salinity anomalies that, when advected to water-mass formation regions, can enhance or decrease the overturning circulation. THC variability leads to changes in the sea surface temperature (SST), changing the atmospheric circulation and hence the WDC. Thus, a full understanding of the system requires an understanding of the effects of this closed two-way feedback loop.

This feedback system is studied here using an enhanced version of the model of PT04, summarized briefly in the next section. As with Stommel's original box model, this model is as simple as plausible for capturing the coupled dynamics of both the WDC and THC. The analysis of the feedbacks between the THC

Corresponding author address: Douglas G. MacMynowski, Control and Dynamical Systems, California Institute of Technology, 1200 E. California Blvd., M/C 104-44, Pasadena, CA 91125.
E-mail: macmardg@cds.caltech.edu

and WDC is performed here using tools from control theory and engineering. New control tools and terminology are introduced and explained as they are used for the analysis. In essence, control theory is about the analysis of feedback systems. By representing the system under study as a set of input–output "black boxes" and feedbacks (e.g., see Fig. 5), much progress may be made to understand the role of these feedbacks, and a rich literature exists (e.g., Doyle et al. 1992; Franklin et al. 2002). The key insight from control theory is that much can be learned about the characteristics of the "closed loop" feedback system by studying the dynamics of the "open loop" system without the two-way feedback interconnection.

The new physical element introduced in this paper is the feedback from the THC to the WDC. We assume that the WDC strength varies because of changes in the atmospheric winds and that the wind, in turn, is affected by the meridional gradient of the SST. The SST, in turn, is advected by the circulation and therefore varies because of THC variability. We further assume that the WDC adjusts to the changed wind forcing with some specified time scale. PT04 found that when only the influence of a fixed WDC on the THC circulation is included, the thermally driven THC is stable for low freshwater (FW) forcing. As the freshwater forcing increases, this equilibrium becomes unstable and, for sufficiently strong WDC, exhibits self-sustained limit cycle oscillations that do not exist without the WDC.

We find here that for a reasonable estimate of the strength of the feedback of the THC on the WDC, there is a significant shift in the critical freshwater forcing (i.e., the location of the Hopf bifurcation point) leading to self-sustained variability. The model is less stable with the additional two-way feedback and requires a smaller FW forcing to be destabilized. In addition, a stable equilibrium appears for yet larger freshwater forcing, unlike the findings of PT04 in the presence of one-way influence from the WDC to the THC only. The existence of a stable equilibrium for large forcing is similar to the usual THC without the WDC (Marotzke 1989; Stommel 1961), although the nature of this equilibrium is different as the THC is not reversed; rather, it is a weakened thermally dominant THC that is stable with the combination of sufficiently large WDC and the dynamic influence of the new feedback considered here. Thus, the additional feedback from the THC to the WDC destabilizes the model at weak FW forcing and stabilizes it for strong FW forcing. Because both the strength of the newly added feedback and the time constant associated with the WDC adjustment are uncertain, we attempt here to understand the change in the dynamic behavior of the coupled system as a function of these parameters.

Three regimes are considered in the following. First, where the thermally driven equilibrium is stable, determining whether the feedback from the THC to the WDC is stabilizing or destabilizing follows directly from the linearization. The extent to which the bifurcation point shifts can be predicted from control tools using what is known as "root locus" arguments. Second, where the equilibrium is linearly unstable and a limit cycle exists, a low-order (a few ODEs) nonlinear model can be derived from the full model using empirical orthogonal functions (EOFs) and Galerkin projection (for methods of reducing complex models to low order ones, see also Timmermann et al. 2001). Describing functions (Gelb and Vander Velde 1968) can then be used to predict the limit cycle amplitude with and without the new feedback. Finally, at yet higher freshwater forcing values, we find that a new stable equilibrium appears as a result of the new feedback introduced here, and its dynamics are investigated.

We next describe the model (section 2), analyze the equilibrium solution (section 3), and the small amplitude self-sustained variability found beyond the first bifurcation point (section 4). The change from the large-amplitude limit cycle to the new stable equilibrium is examined in section 5, and we present conclusions in section 6.

2. Model

a. Fixed wind-driven circulation

The starting point for analyzing the feedback interconnection of WDC and THC is the model of PT04, shown schematically in Fig. 1, where the deep ocean is represented by two boxes (polar and midlatitudes) as in Huang et al. (1992) and Tziperman et al. (1994), while the surface ocean is represented by an annulus. The WDC strength (angular velocity of the flow around the gyre, Ω) is specified and is fixed in time. The THC flows meridionally through the surface layer and then through the deep boxes, driven by density gradients. This simple model is intended to capture the essential physics associated with both processes and the coupling between them, although it is clearly not quantitatively accurate. As with all simple models, the intent is to explore and understand the effect of new physical processes, in this case, the two-way feedback.

The location along the gyre is measured by an angle $0 < \theta < 2\pi$ increasing clockwise from $\theta = 0$ in the west.

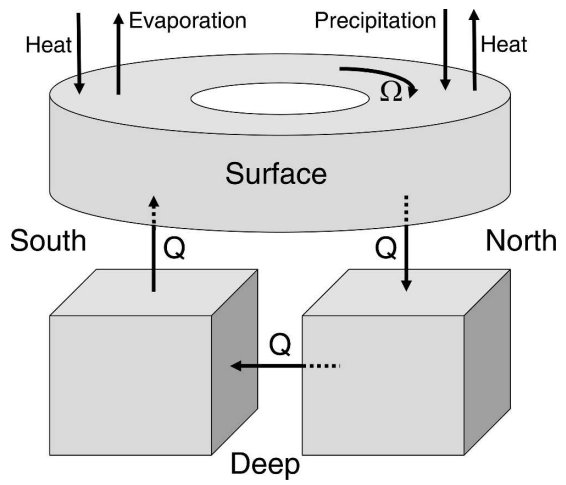

FIG. 1. Schematic of the model for coupled wind-driven and thermohaline circulations (from PT04). The system is forced by differential freshwater (salinity) forcing F_s in the northern and southern (midlatitude) sections.

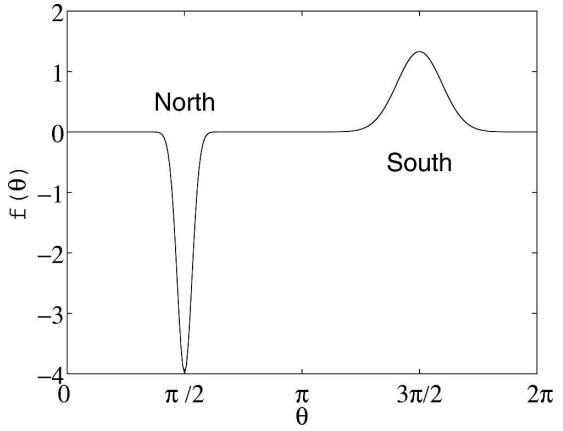

FIG. 2. Function $f(\theta)$ describing the distribution of upwelling and downwelling (from PT04).

The salinity $S(\theta)$ and temperature $T(\theta)$ of the surface layer are described by one-dimensional advection–diffusion partial differential equations. The temperature and salinity states T_d^N, T_d^S, S_d^N, and S_d^S of the northern and southern deep boxes are influenced by an advection by the THC and governed by the ordinary differential equations in (3)–(6). The density ρ is assumed to depend linearly on the temperature and salinity. The overturning circulation strength Q in (8) is a function of the meridional density gradient. This meridional density gradient is calculated from the deep densities, as well as the surface densities averaged over the northern and southern parts of the wind-driven gyre, and weighted by the upwelling/downwelling distribution $f(\theta)$. The total upwelling into the gyre (annulus) and the total downwelling from the annulus to the deep boxes are equal to the overturning Q. The spatial distribution of the upwelling and downwelling are given by the function $f(\theta)$, chosen to be two Gaussians distributed about the northern and southern point of the wind-driven gyre, with variances of σ_n and σ_s, respectively, as shown in Fig. 2. The system is forced by a restoring temperature T_R and freshwater flux F_S. One therefore obtains the following equations (PT04):

$$\frac{\partial S}{\partial t} + \frac{\partial(\tilde{\Omega}S)}{\partial\theta} - \frac{\kappa_S}{R_1R_2}\frac{\partial^2 S}{\partial\theta^2} = -F_S\sin\theta + \frac{2Qf(\theta)}{H(R_2^2 - R_1^2)} \times \begin{cases} S, & \theta < \pi \\ S_d^s, & \theta > \pi, \end{cases} \quad (1)$$

$$\frac{\partial T}{\partial t} + \frac{\partial(\tilde{\Omega}T)}{\partial\theta} - \frac{\kappa_T}{R_1R_2}\frac{\partial^2 T}{\partial\theta^2} = -\Gamma(T - T_R) + \frac{Qf(\theta)}{H(R_2^2 - R_1^2)} \times \begin{cases} T, & \theta < \pi \\ T_d^s, & \theta > \pi, \end{cases} \quad (2)$$

$$V_d^N\frac{dT_d^N}{dt} = -Q\left[\int_0^\pi T(\theta)f(\theta)\,d\theta + T_d^N\right], \quad (3)$$

$$V_d^N\frac{dS_d^N}{dt} = -Q\left[\int_0^\pi S(\theta)f(\theta)\,d\theta + S_d^N\right], \quad (4)$$

$$V_d^S\frac{dT_d^S}{dt} = Q(T_d^N - T_d^S), \quad (5)$$

$$V_d^S\frac{dS_d^S}{dt} = Q(S_d^N - S_d^S), \quad (6)$$

$$\rho/\rho_0 = -\alpha T + \beta S, \quad \text{and} \quad (7)$$

$$Q/Q_0 = \rho_d^N - \rho_d^S - \delta\int_0^{2\pi}\rho(\theta)f(\theta)\,d\theta. \quad (8)$$

The change in angular velocity around the gyre due to the addition/removal of flow from/to the deep boxes is

$$\tilde{\Omega}(\theta) - \Omega = \frac{2\pi Q}{V_s}\int_0^\theta f(\theta)\,d\theta, \quad (9)$$

where $V_s = \pi H(R_2^2 - R_1^2)$ is the volume of the surface layer.

Since the intent is to compare the dynamics of the coupled system with the new physics (two-way feedback) with the previously explored dynamics with fixed WDC, we have attempted where possible to maintain consistency with the parameters used in PT04. Parameter definitions and values used in the simulation here are given in Table 1 and are the same as those of PT04,

TABLE 1. Parameter definitions and values for the coupled WDC–THC model.

σ_n	$f(\theta)$ variance, polar	0.1 rad
σ_s	$f(\theta)$ variance, equatorial	0.3 rad
R_1	Annulus internal radius	5×10^5 m
R_2	Annulus external radius	3×10^6 m
H	Surface layer depth	500 m
D	Deep layer depth	3×10^3 m
V_d^N	Volume, north deep box	2.7×10^{16} m^3
V_d^S	Volume, south deep box	2.7×10^{16} m^3
κ_s	Salinity diffusivity	1×10^3 m^2 s^{-1}
κ_t	Thermal diffusivity	1×10^3 m^2 s^{-1}
β	Salinity expansion coef	7.61×10^{-4} psu^{-1}
α	Thermal expansion coef	1.5×10^{-4} (°C)$^{-1}$
Γ^{-1}	Thermal relaxation time	1 yr
Q_0	Reference water flux	6×10^{10} (m^3 s^{-1})/(kg m^{-3})
ρ_0	Reference density	1 kg m^{-3}
F_s	Salt flux	0–3 m yr^{-1}
T_R	Restoring temperature	$10 + 6 \sin\theta$ °C
$2\pi\Omega_0^{-1}$	Nominal WDC time scale	70 yr
$\tau_{\max}$	Max WDC response time τ	~20 yr

with the exception of an increased surface thermal relaxation time Γ. Note that while a 35-yr WDC circulation time might be more reasonable (e.g., a 1 cm s^{-1} average ocean flow with a length scale of 10^4 km), we have chosen to retain the 70-yr time scale used by PT04 for ease of direct comparison. Regardless of the choice of specific parameters, however, this simple model can only be expected to give qualitative, rather than quantitative results.

The PDEs for the surface temperature and salinity are discretized using central finite differencing, with M elements to yield a $2M + 4$ state system, including the surface and deep temperatures and salinities. These equations are simulated with a variable-order nondissipative solver in Matlab. Defining the full state vector $\mathbf{x} \in \mathbb{R}^{2M+4}$ as

$$\mathbf{x} = [T(\theta_1) \cdot \cdot \cdot T(\theta_M)\, S(\theta_1) \cdot \cdot \cdot S(\theta_M)\, T_d^S\, T_d^N\, S_d^S\, S_d^N]^{\mathrm{T}} \tag{10}$$

(temperature and salinity in all boxes and all grid points) then the resulting system can also be written in the following form, which is more useful for analysis:

$$\dot{\mathbf{x}} = \mathbf{X}(\Omega)\mathbf{x} + (\mathbf{C}_Q^{\mathrm{T}}\mathbf{x})\mathbf{Y}\mathbf{x} + \mathbf{Z}. \tag{11}$$

In (11), $\mathbf{X}(\Omega) = \mathbf{X}_0 + \mathbf{X}_1\Omega$ and $\mathbf{Y}$ are matrices and $\mathbf{Z}(F_s)$ captures both the surface freshwater forcing F_s in (1) and the temperature relaxation term ΓT_R in (2). Note that the thermohaline strength is assumed to be linear in density variations (8), which are in turn assumed linear in the temperature and salinity (7), and thus the (scalar) thermohaline circulation (Fig. 1) can be written as a linear function $Q = \mathbf{C}_Q^{\mathrm{T}}\mathbf{x}$ of the state vector, where each element of the column vector $\mathbf{C}_Q$ defines the influence of the corresponding state on the circulation. The nonlinear term in (11) represents the advection of temperature and salinity by the circulation. Next, we solve for the steady state (equilibrium) of (11), where $\dot{\mathbf{x}} = 0$, and denote it by $\mathbf{x}_e(F_s, \Omega)$. The perturbations about this equilibrium are then denoted by the vector $\boldsymbol{\xi} = \mathbf{x} - \mathbf{x}_e(F_s, \Omega)$, which satisfies

$$\dot{\boldsymbol{\xi}} = (\mathbf{X} + \mathbf{C}_Q^{\mathrm{T}}\mathbf{x}_e\mathbf{Y} + \mathbf{Y}\mathbf{x}_e\mathbf{C}_Q^{\mathrm{T}})\boldsymbol{\xi} + (\mathbf{C}_Q^{\mathrm{T}}\boldsymbol{\xi})\mathbf{Y}\boldsymbol{\xi}. \tag{12}$$

The linearized system can be obtained by dropping the quadratic term, and the local stability of any equilibrium determined from the eigenvalues of the matrix in parentheses in the first term.

For a WDC amplitude characterized by a fixed around-the-gyre advection time scale of $2\pi\Omega^{-1} = 70$ yr, the bifurcation behavior of the coupled WDC → THC system is shown in Fig. 3 (equivalent to Fig. 5 of PT04). Figure 3 also compares the bifurcation behavior with that without the wind-driven circulation. Without the WDC, the behavior is similar to that of many previous studies of the THC (e.g., Marotzke 1989; Stommel 1961) where, for a broad range of forcing values, there are two possible stable equilibria corresponding to thermally dominant (large positive flow) and salinity-dominated (small negative) circulation. With a sufficiently

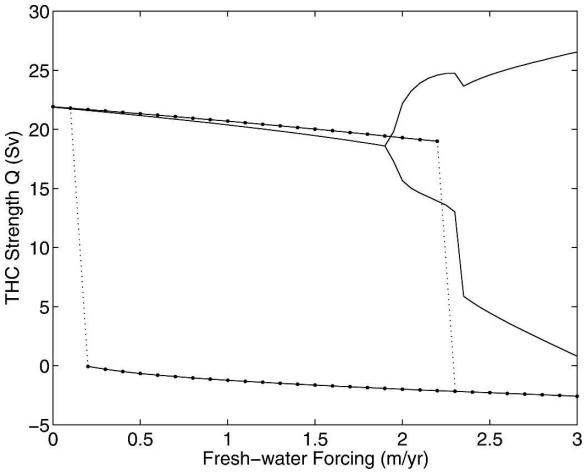

FIG. 3. Bifurcation behavior of THC as a function of freshwater forcing with fixed WDC corresponding to a 70-yr circulation time (solid line) and with no WDC (line with dots). With $2\pi\Omega^{-1} = 70$, the equilibrium is stable for small forcing, but a limit cycle develops for larger forcing ($F_s > 1.9$ m yr^{-1}). The maximum and minimum of each cycle are shown. The character of the limit cycle changes for larger forcing ($F_s > 2.35$). With no WDC, both thermally driven and salinity-driven equilibria exist for a broad range of FW forcing.

strong WDC, the salinity-driven reverse-flow THC equilibrium seems to no longer exist and, instead, a strong variability is found. The FW forcing at which the thermally dominant equilibrium loses stability is also reduced.

The WDC strength Ω in PT04 and in the above analysis is fixed in time. We next consider variations in Ω due to the effects of the THC on the meridional oceanic and atmospheric temperature gradient, and therefore on the wind stress and WDC.

b. *Including the feedback from the THC to the WDC*

Increased THC strength increases the sea surface temperature at high latitudes, decreasing the thermal gradient between mid and high latitudes. This results in a decrease in the wind strength and in a decrease in the WDC. Two parameters are needed to describe this coupling of the WDC to the THC: the amplitude ("gain" in control language) and the time lag associated with reaching an equilibration between the wind strength and the wind-driven circulation.

We assume that the WDC strength is restored toward a value that is directly proportional to the meridional SST gradient for small perturbations about a nominal state $\mathbf{x}_0$. We set $\mathbf{x}_0 = \mathbf{x}_e(F_s, \Omega_0)$ as the equilibrium state for a freshwater forcing of $F_s = 1.9$ m yr^{-1} and consider the effect of changing F_s, allowing the WDC to vary according to the parameterized feedback from the THC. Variations in SST used to calculate the changes in the WDC are assumed to be a factor of $\mu = 5$ larger than the variation in vertically averaged temperature in the 500-m-thick surface layer in the model, solved for by (2). This is because, in the presence of stratification, one expects the surface ocean to respond to changes in the circulation more strongly than the subsurface ocean, so the SST should vary more than the vertically averaged temperature in the upper 500 m. A reasonable measure of the meridional SST gradient in our model is

$$G = \frac{1}{2\pi}\int_0^{2\pi} \mu T(\theta)\sin\theta d\theta. \tag{13}$$

Since this is linear in the state $\mathbf{x}$, it can be written as $G = \mathbf{C}_G^{\mathrm{T}}\mathbf{x}$ with the first M elements of the column vector $\mathbf{C}_G$ being $\mu \sin\theta_i[\Delta\theta/(2\pi)]$.

The atmospheric winds equilibrate with the SST on time scales much shorter than the ocean dynamics, so this adjustment is assumed instantaneous here. The surface ocean responds to the changed shear stress with a time constant τ of the order of 10 yr [the first baroclinic-mode wave basin-crossing time; Anderson and Gill (1975)]. To explore the possible behavior that may exist with the additional feedback included, it is useful to consider a range of feedback parameter values. Here we consider adjustment time scales from zero to a time constant $\tau_{\max} = 20$ yr; this should be a sufficient range to capture the different phenomena that may result. The time evolution of the WDC strength, Ω, for perturbations about a mean state $\mathbf{x}_0$ characterized by a WDC strength of Ω_0, is therefore assumed here to be given by

$$\dot{\Omega} = \frac{1}{\tau}[\hat{\Omega}(x) - \Omega], \tag{14}$$

where the steady-state WDC for a given SST is given by

$$\hat{\Omega}(x) = \Omega_0\left[1 + k\frac{\mathbf{C}_G^{\mathrm{T}}(\mathbf{x} - \mathbf{x}_0)}{\mathbf{C}_G^{\mathrm{T}}\mathbf{x}_0}\right], \tag{15}$$

and where k is a nondimensional feedback gain (i.e., feedback amplitude), equal to one unless noted otherwise, while τ is the time constant of the WDC response to changes in the SST gradient G. For a vanishing response time, $\tau = 0$, the feedback law can be substituted into (11),

$$\dot{\mathbf{x}} = \mathbf{X}(\Omega_0)\mathbf{x} + (\mathbf{C}_Q^{\mathrm{T}}\mathbf{x})\mathbf{Y}\mathbf{x} + \mathbf{Z} + k\left(\frac{\Omega_0}{\mathbf{C}_G^{\mathrm{T}}\mathbf{x}_0}\right)\mathbf{X}_1\mathbf{C}_G^{\mathrm{T}}(\mathbf{x} - \mathbf{x}_0)\mathbf{x}, \tag{16}$$

and for $\tau \neq 0$ a similar equation can be developed for an augmented state vector $\mathbf{x}$ that includes Ω in addition to the temperature and salinity.

The system is simulated with $k = 1$ and with the response time of the WDC set to both $\tau = 0$ and $\tau = \tau_{\max}$ to explore the effects of the feedback changing the WDC as a function of the SST. The wind-driven circulation time scale without the feedback from the THC is again assumed to be $2\pi\Omega^{-1} = 70$ yr for ease of comparison with the fixed WDC results in PT04. The resulting THC strength as a function of freshwater forcing is shown in Fig. 4. Two key features are evident. First, for low freshwater forcing (1.5–1.9 m yr^{-1}), the THC is destabilized by the presence of the additional feedback from the THC and SST to the WDC such that oscillations appear for a smaller value of FW forcing. The destabilization extent strongly depends on the adjustment time of the WDC to changes in the SST; intermediate values of this time constant result in intermediate shifts in the bifurcation point.

A second key feature seen in Fig. 4 is the behavior at higher freshwater forcing where a large-amplitude limit cycle was found by PT04. In the presence of the new

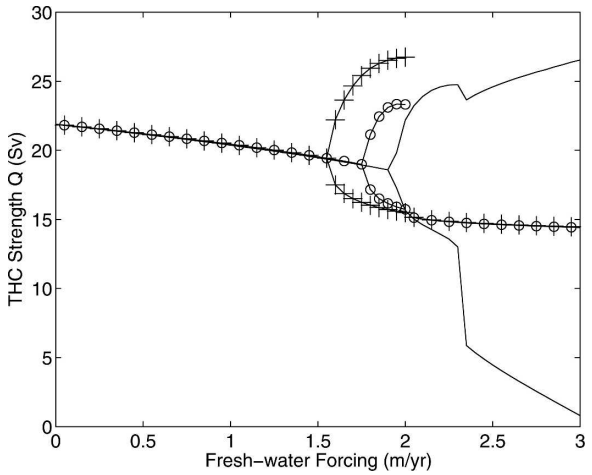

FIG. 4. As in Fig. 3 but with and without the feedback that allows the WDC to change as function of the SST. With this feedback, the behavior depends on the time constant for adjustment of the WDC to the meridional SST gradient, τ. Results are plotted for $\tau = 0$ (line with plus signs) and $\tau = \tau_{\max}$ (line with open circles). The solid line with no symbols is from Fig. 3, where the WDC is fixed. The presence of the new feedback destabilizes the THC for FW forcing of below 1.9 m yr^{-1}. Beyond the initial bifurcation point ($F_s \simeq 1.9$ m yr^{-1} without the feedback), there are self-sustained oscillations and the maximum and minimum of the limit cycle amplitude are plotted. For $F_s > 2$ the system with the feedback converges to a stable steady state, which does not exist without this feedback.

feedback from the THC to the WDC, the corresponding large-amplitude limit cycle is eliminated and a stable equilibrium exists. This effect does not appear to depend on the time constant of the feedback. As might be expected from the figure, the Hopf bifurcation, as one reduces the FW forcing on the new stable branch, is subcritical. As one reduces the feedback coupling, the bifurcation is supercritical for $F_s > 2.3$ and subcritical below that value.

To understand the effect of the additional feedback from the THC to the WDC, consider the following: A perturbation that weakens the THC results in an increase in salinity in the southern region and a decrease in the northern region. These anomalies act to reinforce the initial perturbation in the THC, while the corresponding temperature anomalies are stabilizing. At sufficiently large freshwater forcing, the salinity effect dominates and the THC becomes unstable (i.e., the advective instability feedback; Dijkstra 2005; Marotzke et al. 1988). With a fixed WDC, the increased salinity in the southern region is advected northward where it eventually counteracts the original THC perturbation. The balance between these two effects determines the stability of the system and the amplitude of the limit cycle that exists when the system is unstable.

The additional feedback between the THC and WDC modifies the WDC and alters this balance as follows: A small decrease in the THC results in an increase in the meridional temperature gradient at the surface, which strengthens the winds, and therefore also the WDC. Consider first the case in which the WDC is relatively weak in comparison with the THC, as occurs for our choice of parameters at low FW forcing and therefore high THC. In this case, the southward WDC on the eastern part of the gyre may be roughly equal but opposite to the northward surface flow of the THC, so the net flow is slow. The equilibrium state under these conditions has a high salinity east of the southern upwelling region. That is the case because the slow net flow on the eastern side spends longer time in the net evaporation region. The more rapid flow on the western side is fresher because it passes through the evaporation region faster. Consider now the advection of salt by the positive WDC anomaly excited by the decreased SST due to the decreased THC perturbation. This positive WDC anomaly transports the high eastern salinity in the above equilibrium state southward. At that point this salinity perturbations affects the THC in a way that reinforces the original perturbation decreasing the THC. This is therefore a positive feedback, leading to the destabilizing behavior observed at low FW forcing in Fig. 4.

The new feedback added in this paper, allowing the WDC to change as function of the SST, can be either stabilizing or destabilizing. This would depend on the relative strength of the WDC and THC via the location of the salinity maximum discussed above. The destabilizing feedback is most significant if the WDC change is in phase (simultaneous) with the THC perturbation, rather than being delayed.

It is possible to predict and understand many of the general characteristics of Fig. 4 for different feedback amplitudes (gains) k and time constants τ through the analysis of the open-loop system without the feedback from the THC to the WDC; this is the goal of the next three sections. We also note that some of the details of the above stabilizing and destabilizing effects of the new feedback may be specific to the very idealized annulus model that we use here for the WDC. Clearly, additional work is needed with a more detailed and realistic model.

3. Analysis of stable equilibrium

For sufficiently small values of the freshwater forcing, the thermally driven thermohaline circulation is stable. The linearization about this equilibrium provides a useful insight into the behavior of the equilib-

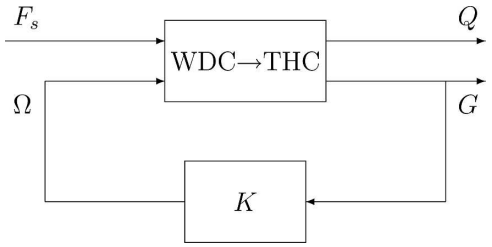

FIG. 5. Block diagram of feedback interaction; the upper block marked "WDC→THC" is the model of PT04 in which the WDC affects the THC but where its amplitude is specified and is constant in time. The lower block is the new feedback introduced in this study, from the THC to the WDC, allowing the WDC to change as a function of the SST.

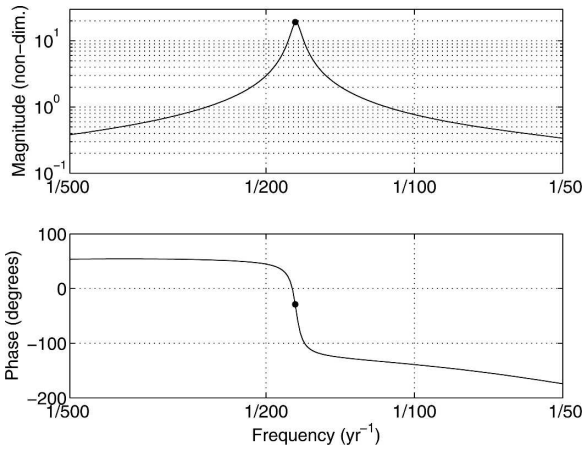

FIG. 6. Frequency response T_{gw} of linearization of the coupled WDC → THC system at $F_s = 1.9$ m yr^{-1}, from perturbations in WDC strength Ω to perturbations in SST gradient G resulting from THC. Magnitude is normalized by the equilibrium Ω_0 and G_0 around which linearization is done. At the frequency corresponding to maximum magnitude (marked by a dot on the two curves), the phase is −30°.

rium in the presence of a two-way feedback interaction between THC and WDC. We will use approximations to the frequency-domain response of the system without the feedback from the THC to the WDC in order to efficiently estimate the stability of the fully coupled system as a function of the feedback amplitude k and time constant τ, without resorting to simulation at each value of these feedback parameters. To do so, we first introduce some background terminology; more details are available in any standard controls textbook (Doyle et al. 1992; Franklin et al. 2002).

a. Root locus analysis

The strength Ω of the WDC and the freshwater forcing F_s, may be thought of as the inputs to the PT04 model, and the strength of the THC, Q, and meridional surface temperature gradient, G, are the outputs. The system is shown schematically in Fig. 5, where $K = K(k, \tau)$ represents the new feedback added here, allowing the WDC strength Ω to change as a function of the THC and the SST gradient G as in (14).

Our first objective is to characterize the stability of the two-way coupled system in terms of the properties of the one-way coupled system (i.e., without the feedback from the THC to the WDC). The input/output dynamics of a linear system can be described by the frequency-dependent amplitude and phase of the output in response to a unit amplitude input sinusoid, known as the "frequency response" or "transfer function" from input to output, see Franklin et al. (2002, section 3.1.2., p. 99).

We linearize the model around an equilibrium point $\mathbf{x}_0$ close to instability ($F_s = 1.9$ m yr^{-1}) where constant inputs Ω_0 and F_s lead to constant outputs $Q_0 = \mathbf{C}_Q^{\mathrm{T}}\mathbf{x}_0$, $G_0 = \mathbf{C}_G^{\mathrm{T}}\mathbf{x}_0$. Then, given a perturbation w to the WDC strength, $\Omega = \Omega_0 + w$, we obtain perturbations in the SST gradient, $G = G_0 + g$.

The transfer function is the ratio of the Laplace transforms of w and g so that $\hat{g}(s) = \hat{T}_{gw}(s)\hat{w}(s)$. To derive $\hat{T}_{gw}(s)$ we write our linearized Eq. (12) as $d\boldsymbol{\xi}/dt = \mathbf{A}\boldsymbol{\xi} + \mathbf{B}w$, where w is the "input" of the linearized model, and the SST gradient anomaly (the model output) is given by $g = \mathbf{C}\boldsymbol{\xi}$. Laplace transforming gives $\hat{g} = \mathbf{C}(\mathbf{I}s - \mathbf{A})^{-1}\mathbf{B}\hat{w} = \hat{T}_{gw}\hat{w}$. The magnitude and phase of $\hat{T}_{gw}$ are shown in Fig. 6 as a function of the complex frequency $s = i\omega$. The amplitude in Fig. 6 is nondimensionalized by G_0/Ω_0, and the phase $\phi(\omega)$ of $\hat{T}_{gw}(i\omega)$ is such that, if $w = \sin\omega t$, then $g = |\hat{T}_{gw}(i\omega)| \sin(\omega t + \phi(\omega))$.

The relationship between the transfer function and the corresponding time domain behavior can be examined through an example. If $\hat{T} = (s - \omega)[(s - \omega)^2 + b^2]^{-1}$ and the input is white noise $\hat{w} = 1$, then the output g is the inverse Laplace transform of $\hat{T}\,\hat{w}$, so $g(t) = \cos(\omega t)e^{bt}$. This transfer function is singular in the complex s plane at $s = b \pm i\omega$; these are referred to as the poles of the transfer function and are the same as the eigen-values of $\mathbf{A}$. The location of the poles determine both the frequency of the response, ω, and whether it is stable, $b < 0$, or unstable, $b > 0$. The behavior of these poles as a function of the model parameters is therefore of great value in providing intuition on the dynamics.

Remember that g and w are deviations from the steady state and are nonzero only for nonzero perturbations. To study the dynamics of these deviations, and therefore the stability of the system, we consider external perturbations (e.g., white noise, or periodic perturbations of different periods) added to the system as

follows. Let the introduced noise terms added to g and w be δ_g and δ_w. Since the system is linear, the response to the perturbations in $\hat{g}$ due to noise $\hat{\delta}_g$, and the response of the feedback path between the SST gradient G influencing WDC strength Ω, to noise δ_w, can be described in the Laplace domain by

$$\hat{g}(s) = \hat{T}_{gw}(s)\hat{w}(s) + \hat{\delta}_g(s) \quad \text{and} \tag{17}$$

$$\hat{w}(s) = \hat{K}(s)\hat{g}(s) + \hat{\delta}_w(s), \tag{18}$$

where $\hat{K}(s)$ is the transfer function of the assumed feedback (K in Fig. 5). Solving the above equations for the deviations from steady state g and w, in terms of the external noise terms $\hat{\delta}_g$ and $\hat{\delta}_w$, leads to the closed-loop dynamics (with the two-way feedback between the THC and the WDC) being described by

$$\begin{bmatrix} \hat{g}(s) \\ \hat{w}(s) \end{bmatrix} = [I - \hat{T}_{gw}(s)\hat{K}(s)]^{-1} \times \begin{bmatrix} I & \hat{T}_{gw}(s) \\ \hat{K}(s) & \hat{K}(s)\hat{T}_{gw}(s) \end{bmatrix} \begin{bmatrix} \hat{\delta}_g(s) \\ \hat{\delta}_w(s) \end{bmatrix}. \tag{19}$$

If $\hat{T}_{gw}$ and $\hat{K}(s)$ are both stable (i.e., both have poles in the left-half complex s plane only), then the stability of this closed-loop system is determined by the properties of the common factor $[I - \hat{T}_{gw}(s)\hat{K}(s)]^{-1}$ in front of the matrix on the rhs of (19) (Doyle et al. 1992). The stability of this system can be predicted from the dynamics of the loop-transfer function $\hat{T}_{gw}\hat{K}$. This loop transfer function physically represents our model, including the influence of the WDC on the THC and the dynamics that would influence the WDC in turn, but without the actual connection from the THC to the WDC. In other words, the loop-transfer function is the factor multiplying a signal as it travels once around the "loop" in Fig. 5.

If the loop transfer function $\hat{T}_{gw}\hat{K}$ is written as a rational polynomial function $kn(s)/d(s)$ of the Laplace variable s, then the closed-loop pole location as a function of the feedback gain k is given by the roots of the denominator of

$$(I - \hat{T}\hat{K})^{-1} = \frac{1}{1 - kn(s)/d(s)} = \frac{d(s)}{d(s) - kn(s)}. \tag{20}$$

Remember that both the feedback gain from the THC to the WDC, k, and the time constant with which the WDC equilibrates to changes in THC, τ, are uncertain. We therefore want to study the stability as a function of these parameters. As we shall shortly see, the advantages of analyzing the open-loop system rather than the closed-loop system are in both providing additional insight and in making the analysis more efficient.

The plot of the pole location as a function of k, also known as the root locus (chapter 5 in Franklin et al. 2002) is a useful tool for understanding the effects of the gain k on the stability. For $k = 0$ (no feedback), the poles are those of the open-loop system and, assuming this system is stable, the poles are in the left-half complex s plane. The closed-loop poles ($k > 0$) move away from the open-loop poles along a continuous trajectory as k increases. The angle in the complex s plane at which the poles move from their location for $k = 0$ is called the "departure angle." If this angle is 90°, for example, the poles move parallel to the imaginary axis and will not cross the imaginary axis; thus the system remains stable as k increases. If, on the other hand, the departure angle is 0°, the poles move directly toward the imaginary axis, and increasing k will eventually lead to the destabilization of the system. Thus the departure angle provides insight into how the stability of the closed-loop system depends on the feedback parameters (e.g., τ).

The exact root locus is obtained by calculating the roots of $d(s) - kn(s)$ as a function of k. Figure 7 shows the pole location as a function of the feedback gain (i.e., the root locus) obtained from the linearized system $T_{gw}K$ for response times of the WDC to the SST set to both $\tau = 0$ and $\tau = \tau_{\max}$. The nonzero response time is characterized by a larger departure angle, and therefore results primarily in a decrease in frequency of the response as the feedback amplitude varies, and the system remains stable for much higher gain. Thus, we find that the response time of the WDC to the winds and SST has a stabilizing effect. To better understand why this is true, an approximate analysis of the system is useful.

b. *Approximate root locus analysis*

The above general characteristics of the root locus, including the departure angle, can be obtained directly from the phase of $n(s)$ and $d(s)$ (see chapter 5 in Franklin et al. 2002). This allows us to predict the relevant features of Fig. 7 based on an approximate linearized analysis of the system without the feedback from the THC to the WDC, and without the need to explicitly calculate the poles for different values of the feedback amplitude k (which would be of value particularly for higher dimension systems). This involves approximating our two transfer functions, one describing the feedback from the WDC to the THC (PT04), and the other describing the new physical feedback added here from the THC to the WDC. The approximate analysis should

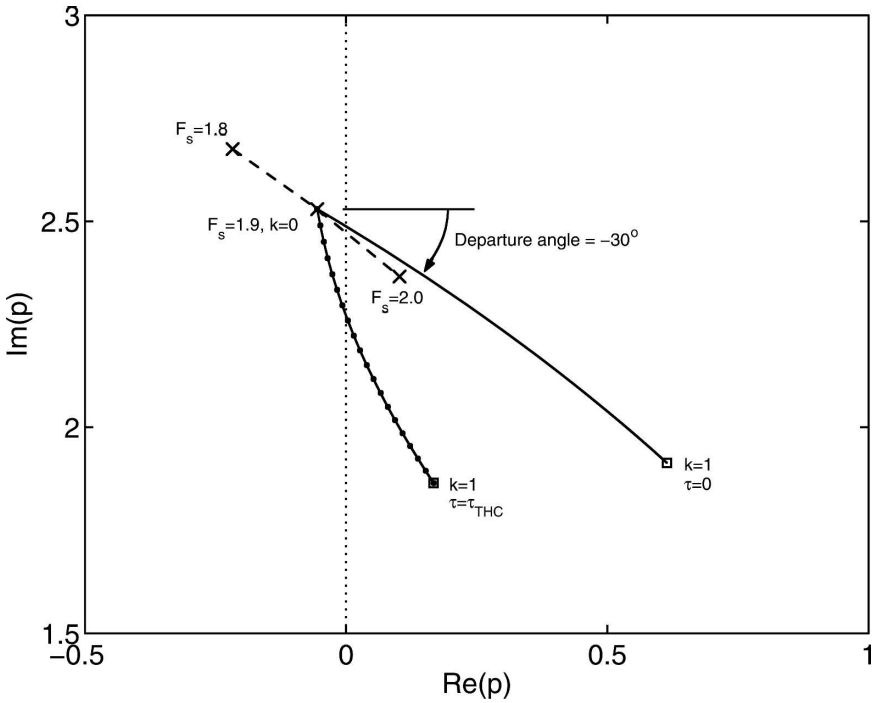

FIG. 7. Root locus of dominant poles (i.e., location of poles in the complex s plane as function of the amplitude of the feedback from the THC to the WDC) for linearized coupled WDC → THC system at $F_s = 1.9$ for gain (feedback amplitude) values between $k = 0$ and $k = 1$. Shown are the root locus for feedback time constant τ of zero (thin line extending into right-half-plane) and for feedback time equal to the that of the THC dynamics (line with dots). The departure angle is larger for a finite response time, τ, of the WDC to the SST (the dotted line). This indicates that a longer response time has a stabilizing effect on the coupled WDC → THC system. In addition, the pole location for varying F_s is shown with a dashed line, with the location for $F_s = 1.8$, 1.9, and 2.0 marked by Xs. The root locus is symmetric about the real axis; thus only the part above the real axis is shown here.

be especially useful when this methodology is applied to complex models in which the full analysis is not feasible.

First, note that near the instability threshold as a function of the freshwater forcing amplitude, the WDC → THC dynamics are damped oscillatory (Griffies and Tziperman 1995; Tziperman et al. 1994) and hence dominated by the dynamics of two complex-conjugate lightly damped (stable) poles; that is, two poles that are in the negative real part of the s plane. As the freshwater forcing increases, the damping of this dominant pole pair decreases and the poles move toward the imaginary axis. The Hopf bifurcation of the fully nonlinear model occurs when the damping of these poles reaches zero, and the linearized system loses stability when the poles move to the positive real axis half-plane. The WDC → THC transfer function T_{gw}, for the stable freshwater forcing regime near the instability threshold, is therefore close to that of a slightly damped second-order oscillatory system (a system of two first-order ODEs with two damped oscillatory eigenmodes). At the angular frequency of the THC oscillations past the first bifurcation point, ω_{THC}, the transfer function may therefore be approximated as

$$\hat{T}_{gw}(i\omega) \simeq |\hat{T}_{gw}(i\omega_{\mathrm{THC}})| e^{i\phi_T + i\pi/2} \frac{2\zeta\omega_{\mathrm{THC}}^2}{s^2 + 2\zeta\omega_{\mathrm{THC}} s + \omega_{\mathrm{THC}}^2}, \tag{21}$$

where ω_{THC} is the angular frequency of the poles at the point of instability. The denominator is characterized by having two poles, as required, with ζ being the damping and $\zeta\omega_{\mathrm{THC}}$ the decay rate (distance from the imaginary axis). The phase ϕ_T is the phase of $\hat{T}_{gw}(i\omega_{\mathrm{THC}})$, so the overall magnitude and phase of the transfer function at the resonant frequency are correct. For our model, the phase at the resonant frequency is approximately −30° (as marked by the dot on the two curves in Fig. 6) and the magnitude is approximately 19 G_0/Ω_0. This means that, if $\Omega(t)/\Omega_0$ included a small sinusoidal perturbation at this frequency, then the resulting normalized SST gradient $G(t)/G_0$ at this frequency

would lag the input signal by 30° and have a magnitude that is 19 times as large.

The transfer function $\hat{K}(s)$ of the feedback from the THC to WDC is obtained by taking the Laplace transform of the linearized version of the equation for the WDC amplitude, (14), using $G_0 = \mathbf{C}_G^{\mathrm{T}}\mathbf{x}_0$

$$\hat{K}(s) = \frac{\hat{w}(s)}{\hat{g}(s)} = k\,\frac{1}{\tau s + 1}\left(\frac{\Omega_0}{G_0}\right). \tag{22}$$

For frequencies close to ω_{THC}, the variation in $\hat{K}(s)$ with frequency is small relative to that of $\hat{T}_{gw}(s)$, which, as can be seen in Fig. 6, varies rapidly at that frequency. As a result, we can ignore the variations in $\hat{K}$ and approximate it as

$$\hat{K}(i\omega) \simeq k(\Omega_0/G_0)\alpha e^{i\phi_K}. \tag{23}$$

We are interested in the behavior for the normalized feedback gain k, varying between 0 and 1. The phase of $\hat{K}(s)$ at the frequency $s = i\omega_{\mathrm{THC}}$ is found from (22) to be $\phi_K = -\tan^{-1}(\tau\omega_{\mathrm{THC}})$ and depends on the assumed time constant τ. We have chosen $\tau_{\max}$ slightly less than $\omega_{\mathrm{THC}}^{-1}$, so that ϕ_K varies between 0° and −35° for τ between 0 and $\tau_{\max}$. To obtain the correct magnitude of the transfer function at the frequency ω_{THC}, we set $\alpha = |1 + i\omega_{\mathrm{THC}}\tau|^{-1}$, which varies between 0 and ~0.8.

Using the above approximations for $\hat{T}_{gw}$ (21) and for $\hat{K}$ (23), then for small feedback gain k it is straightforward to show (Franklin et al. 2002, 282–283) that the departure angle of the dominant pole is $\phi_T + \phi_K$ (measured counterclockwise as shown in Fig. 7). For small k, the damping ratio of the dominant poles as a function of both k and ϕ_K (obtained from τ) is given by

$$\frac{\zeta_{\mathrm{CL}}}{\zeta} = 1 - k\alpha\,\frac{\Omega_0}{G_0}\,|\hat{T}_{gw}(i\omega_{\mathrm{THC}})|\cos(\phi_T + \phi_K) \tag{24}$$

with instability occurring when $\zeta_{\mathrm{CL}} = 0$. These observations allow us to immediately estimate the stability of the closed-loop system for any (small) feedback gain k and feedback time lag τ. The departure angle with no feedback time lag is −30°. This means that, as the feedback amplitude k is increased, we expect the poles to move toward the positive real axis with a slight angle from the most direct path to the positive real half complex plane. This simply means that the feedback from the THC to the WDC, with no time lag ($\tau = 0$), is quite efficiently destabilizing. The additional feedback time lag $\tau \neq 0$ increases the departure angle to −65°, which is closer to being parallel with the imaginary axis, consistent with Fig. 7, meaning again that in this case changing the gain leads to a change of the frequency of the response but with less motion toward the unstable domain in the positive real half-plane.

The way that we represent the WDC response to the SST (14) is effectively a low-pass filter on the variation in the SST gradient. As the adjustment time is increased, the magnitude of the WDC response to high frequency THC or SST oscillations decreases, but also the phase of the response changes, lagging the THC/SST by a larger fraction of a cycle. The departure angle of the root locus depends on the phase relationship between oscillations in the WDC and oscillations in the THC at the natural frequency as follows. The THC and SST change due to WDC anomalies. This change then induces a change in the WDC as well, due to the new feedback introduced here. If this change to the WDC is in phase with the original WDC anomalies, the new feedback is strongly destabilizing, and vice versa. For large WDC readjustment times, this phase lag turns out to be large, nearly ¼ cycle, so the new feedback is not strongly destabilizing in this case. Thus, the effect of the new feedback on the THC stability is a *dynamic* effect associated with the relative phase of the WDC response. The analysis of the eigenvalue dependence via the root locus plot helps to understand the physics of the behavior that would not be immediately obvious based only on the simulation response in Fig. 4.

Thus, without the need to explicitly compute the stability as a function of time lag, it is immediately clear that, about the original stable equilibrium, the new feedback added in this paper will always be destabilizing. The bottom line is that this approximate analysis based only on computing the departure angle is consistent with the numerical calculation show in Fig. 7 and will be a very useful tool for a more realistic model for which the numerical calculation may be too expensive or difficult to perform.

Noting again that small perturbations to the stable steady state will result in damped oscillations, we now proceed to calculate the sensitivity of the damping rate (ζ) of the oscillations to changes in freshwater forcing. This is done using a linearization about equilibrium points for small perturbations in the freshwater forcing F_s. First, Fig. 7 shows the pole location for the WDC → THC system without the two-way feedback, as a function of the freshwater forcing (dash line and × marks). The estimate of damping variation due to both FW forcing ($\Delta\zeta/\Delta F_s$) and feedback amplitude [$\Delta\zeta/k$, obtained from (24)] leads to a predicted shift in the FW forcing at which the Hopf bifurcation occurs:

$$\Delta F_s^* = \left(\frac{\Delta\zeta}{\Delta F_s}\right)^{-1}\left(\frac{\Delta\zeta}{k}\right)k. \tag{25}$$

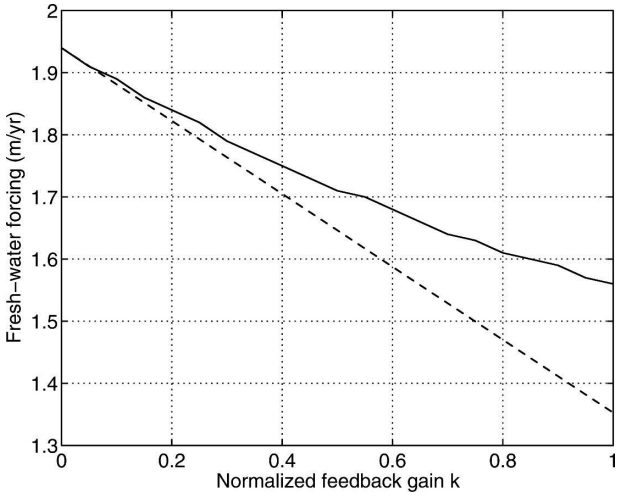

FIG. 8. Value of freshwater forcing F_s at which first bifurcation (transition from a stable THC to oscillations) occurs as function of feedback amplitude (gain) between 0 and 1. Shown are the values from full model (solid line) and a prediction using the linearized root locus arguments of Eq. (25) (dashed line).

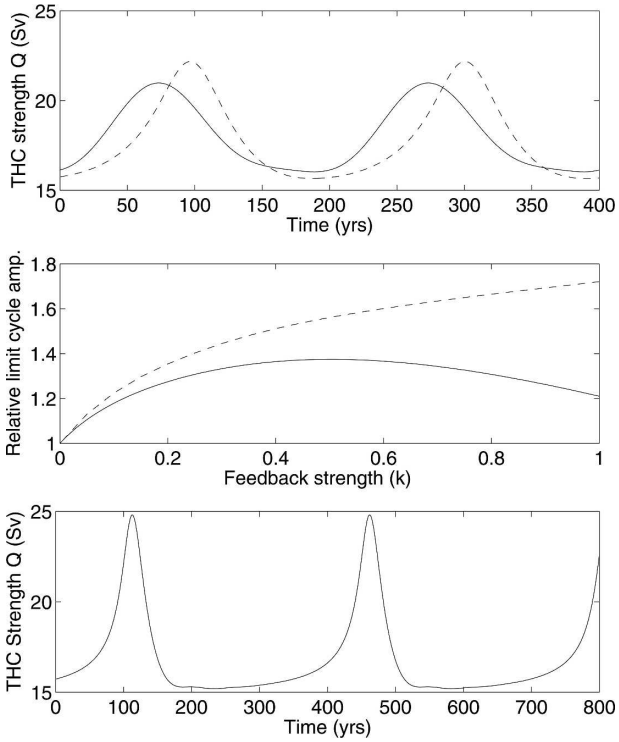

FIG. 9. Approximation of limit cycle behavior for freshwater forcing F_s = 2.0 m yr^{-1} using describing functions. (a) A comparison of the time history without the feedback from the THC with the WDC for the full solution (dashed line) and describing function approximation (solid line). (b) The trend in limit cycle amplitude with feedback gain for the full solution (dashed line) and describing function analysis (solid line). (c) The limit cycle characteristics for a normalized feedback gain of 0.5; note the different time scale.

This linear prediction, obtained through analysis at a single value of F_s, is compared in Fig. 8 with the actual shift in the bifurcation point with feedback for $\tau = 0$. Good agreement is obtained for modest feedback gains k, with the maximum error being approximately 30% at a feedback gain of unity. Even for this low-order model of WDC $\rightarrow$ THC dynamics, there is a significant computational savings in this analysis in comparison with a direct calculation of this sensitivity.

We have therefore demonstrated the value of the control tools for understanding small amplitude (linearized) behavior about the stable steady state. The analysis was performed at freshwater forcing values smaller than that leading to the first (Hopf) bifurcation and to self-sustained oscillations. We saw that, in this regime, approximate linear analysis efficiently predicts the effects of the feedback by which the THC strength affects the WDC strength as a function of the feedback gain and time constant, without resorting to full simulations at each of the parameter values. In the next section we use different control tools to analyze the self-sustained variability beyond this first bifurcation point.

4. Limit cycle analysis

For freshwater fluxes greater than roughly 2 m yr^{-1}, the model described in section 2 exhibits self-sustained oscillations (limit cycle behavior), as shown in the top panel in Fig. 9. This region of the parameter space is of particular interest to understanding climate variability. Approximate techniques for estimating the existence and amplitude of the limit cycle can be used to predict the variation in limit cycle amplitude as a function of the feedback parameters. However, these techniques are only computationally tractable for relatively low-order systems, so we first explore methods for model reduction using EOFs (Peixoto and Oort 1992), which may be used to reduce the behavior of complex models to low-order systems (i.e., a few ODEs) (Lumley 1970).

The behavior in Fig. 3 was obtained using a 256-point discretization of each of the PDEs describing the salt and temperature distribution around the surface annulus, resulting in a state vector made of 517 variables [256 $\times$ 2 (surface T, S) + 4 (deep T, S) + 1 (Ω)]. Near the bifurcation point, however, the behavior of the system is characterized by only a few active degrees of freedom and may therefore be described by a reduced order model composed of a few ODEs, simplifying subsequent analysis. This reduced model is obtained using EOFs of the simulation response to identify mode shapes, and Galerkin projection to identify the result-

ing nonlinear ODEs for the reduced order model. Related (and perhaps even more optimal) methods have been used in oceanography by Timmermann et al. (2001).

EOFs identify "optimal" basis functions $\boldsymbol{\phi}_i$ that minimize the residual error in fitting the observed data, with the norm of a vector $\mathbf{x}$ defined by an appropriate inner product $\|\mathbf{x}\| \equiv \langle \mathbf{x}, \mathbf{x} \rangle$ (e.g., Rowley et al. 2004). Frequently, the inner product $\langle \mathbf{x}_1, \mathbf{x}_2 \rangle = \mathbf{x}_1^{\mathrm{T}} \mathbf{x}_2$ is chosen for simplicity, but this does not provide any scaling or normalization of the state. One of the challenges associated with using EOFs as a basis for model reduction is that, while it captures those modes with the largest magnitude (as defined by the choice of inner product), there may be modes with small magnitude that nonetheless are critical for describing the future evolution of the system (Farrell and Ioannou 2001). This problem can be overcome using a balanced trunctation (Farrell and Ioannou 2001). Another approach by Rowley (2005) that still retains the computational advantages of EOFs is to use the "observability Gramian" in the inner product used in computing the EOF basis: $\langle \mathbf{x}_1, \mathbf{x}_2 \rangle = \mathbf{x}_1^{\mathrm{T}} \mathbf{W}_o \mathbf{x}_2$. The observability Gramian, $\mathbf{W}_o \in \mathbb{R}^{n \times n}$, relates the current state vector $\mathbf{x}(t) \in \mathbb{R}^n$ (in our case, the model temperature and salinity) to the evolution of the variables of interest $\mathbf{y}(t)$ (e.g., THC strength) over all future time; that is,

$$\mathbf{x}(t)^{\mathrm{T}} \mathbf{W}_o \mathbf{x}(t) = \int_t^\infty \mathbf{y}(r)^{\mathrm{T}} \mathbf{y}(r)\, dr. \tag{26}$$

By measuring the importance of the state explicitly in terms of its influence on the future evolution of the variables of interest, we ensure that the reduced-order model constructed using EOFs captures the dynamically relevant characteristics of the model rather than simply the most energetic states. This is essential to the validity of the reduced-order model. For a linear system $\dot{\mathbf{x}} = \mathbf{A}\mathbf{x}$ with output $\mathbf{y} = \mathbf{C}\mathbf{x}$, the observability Gramian defined in Eq. (26) is readily obtained as the solution to a Lyapunov equation $\mathbf{A}^{\mathrm{T}}\mathbf{W}_o + \mathbf{W}_o\mathbf{A} + \mathbf{C}^{\mathrm{T}}\mathbf{C} = 0$; for a derivation, see Franklin et al. (2002, p. 855).

The Lyapunov equation has been discussed in the context of data assimilation by Hoang et al. (1997), Tippett and Cohn (2001), and Tippett et al. (2000). Rowley (2005) also discusses data-based estimation of the Gramian appropriate for large systems where direct calculation of the Lyapunov solution is computationally infeasible. The observability Gramian is evaluated in our case near the stability boundary of the linearized system, with the THC strength Q as the system output [$\mathbf{y}(t)$ above]. Choosing this weighting by the Gramian gives an emphasis on those states with the largest contribution to the THC strength and improves the convergence of the EOFs in capturing the dynamics. Near the onset of the limit cycle we find that the first four EOFs capture 99% of the variance (defined via this inner product).

The change in basis of our model (from the values of temperature and salinity at each grid point to the EOF amplitudes) can be described by

$$\mathbf{x}(t) = \sum_{i=1}^{n} \boldsymbol{\phi}_i q_i(t), \tag{27}$$

where $\boldsymbol{\phi}_i \in \mathbb{R}^{2M+5}$, $i = 1 \ldots n$, and n is the number of basis functions (which are vectors of length equal to that of the state vector) that we choose to retain for a low-order representation of the full model dynamics. Because of the structure of our model, (11), the scalar coefficients of the Galerkin projection, $q_i(t)$, are straightforward to compute using inner products (Rowley et al. 2004). Define $\mathcal{X}_{ij} = \langle \boldsymbol{\phi}_i, \mathbf{X}\boldsymbol{\phi}_j \rangle$, $\mathcal{Y}_{ijk} = \langle \boldsymbol{\phi}_i, (\mathbf{C}_Q^{\mathrm{T}}\boldsymbol{\phi}_j)\mathbf{Y}\boldsymbol{\phi}_k \rangle$ and $\mathcal{Z}_i = \langle \boldsymbol{\phi}_i, \mathbf{Z} \rangle$. The resulting n state model for the system without the feedback from the THC to the WDC is

$$\dot{q}_i = \sum_{j=1}^{n} \mathcal{X}_{ij} q_j + \sum_{j=1}^{n} \sum_{k=1}^{n} \mathcal{Y}_{ijk} q_j q_k + \mathcal{Z}_i. \tag{28}$$

The corresponding terms with the feedback of the THC to the WDC can be similarly computed using (16). We typically retain $n = 4$ basis functions for the calculations shown below. This is sufficient for the dynamics in (28) to accurately reproduce the period, amplitude, and spatial characteristics of the limit cycle obtained with the full equations.

Once we have reduced our model from a $2M + 5 = 517$ equations for $\mathbf{x}(t)$ to $n = 4$ equations for $q_i(t)$, the amplitude of the limit cycle can be predicted using a "describing functions" procedure (Gelb and Vander Velde 1968), which is basically a weakly nonlinear approximation in which the nonlinear model evolution is "described" by a few sinusoidal terms. We seek solutions for $\mathbf{q}(t) \in \mathbb{R}^n$ of the form

$$\mathbf{q}(t) = \boldsymbol{\alpha} + \boldsymbol{\beta} \sin\omega t + \boldsymbol{\gamma} \cos\omega t + \boldsymbol{\delta} \sin 2\omega t + \boldsymbol{\varepsilon} \cos 2\omega t, \tag{29}$$

where $\mathbf{q}(t)$, $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, and so on, are vectors of dimension equal to the number of EOFs kept in the reduced system. Next, we substitute this form into the nonlinear equations in (28). The nonlinearities result in nonlinear sine and cosine terms, and identities such as $\sin^2\omega t = (1 - \cos 2\omega t)/2$ are used to reduce the expres-

sion to a linear series of sines and cosines of higher frequency. Next, we truncate the expressions obtained that way by neglecting all higher frequencies except for ω and 2ω. Finally, comparing coefficients for each of the harmonics (constant term which is independent of time, $\sin\omega t$, $\cos\omega t$, $\sin 2\omega t$, $\cos 2\omega t$, we obtain five vector equations for the elements of $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, $\boldsymbol{\delta}$ and $\boldsymbol{\varepsilon}$, which together form a set of $5n$ equations.

Including ω, there are $5n + 1$ unknowns; however, since we are not interested in an absolute time reference, one of the elements of any one of the unknown coefficient vectors may be set to zero without loss of generality before solving for the rest of the elements. This leads to $5n$ equations for the $5n$ unknown coefficients of $\mathbf{q}(t)$ in (29). This method of reducing the nonlinearities to a set of harmonics is called in the control literature "describing function analysis," the method for obtaining the resulting set of equations is known as the "harmonic balance," and together these allow us to estimate the presence and characteristics of limit cycles.

Frequently, only the component of the response at the frequency of the input is retained (i.e., only the $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$ terms above), and thus the approach can be thought of as a quasi linearization where each nonlinearity is replaced by a frequency- and amplitude-dependent gain. To capture the highly nonsinusoidal characteristics of the WDC $\rightarrow$ THC limit cycle, herein we retain the first harmonic of the response as well.

This method is essentially the same as the weakly nonlinear approximation used by Eccles and Tziperman (2004), which is also equivalent to various averaging techniques such as used in the appendix of Jin (1997). Nontrivial solutions describe limit cycles, which are self-sustained finite amplitude, periodic solutions to the model equations. An example illustrating the ability of this technique to capture limit cycle solutions of the original simulation is shown in the top panel in Fig. 9; the first four EOFs were used ($n = 4$) and the amplitudes of the fundamental and first harmonic (i.e., the ω and 2ω terms) solved for as described above. For the WDC $\rightarrow$ THC system without the new feedback added in this work ($k = 0$), the time history of the THC strength computed from the full nonlinear simulation for two cycles of the limit cycle is compared with the predicted limit cycle reconstructed from the solution $\mathbf{q}(t)$ identified from the "describing function" approach. While clearly an approximation, both amplitude and period are predicted with reasonable accuracy, despite retaining only a few spatial basis functions and a few terms in the harmonic balance.

The set of basis functions used in the top panel in Fig. 9 was obtained from the simulation without the feedback from the THC to the WDC. We can now modify the reduced-order model for $q_i(t)$ based on these same basis functions in order to predict the behavior of the limit cycle for nonzero values of the new feedback introduced in this paper (nonzero feedback gain, k). The reduced-order model with feedback is created as in (28), but with $\mathcal{X}$, $\mathcal{Y}$, and $\mathcal{Z}$ computed using the model with feedback in (16) rather than (11). For each value of k, the coefficients of the reduced-order model are updated and the limit cycle solution estimated, as described above. The middle panel in Fig. 9 shows the amplitude of the limit cycle based on this calculation using the reduced-order model versus a calculation based on the full model equations. The trend in limit cycle behavior with feedback is clearly adequately predicted for small values of k up to about 0.4. For larger gain, the set of basis functions is not sufficiently rich to capture the behavior of the dynamics, and more importantly, the limit cycle is less well represented by including only harmonics up to 2ω. The bottom panel in Fig. 9 shows the actual limit cycle for a normalized feedback gain of 0.5; not surprisingly, there is a significant error in using the describing function and harmonic balance approach to estimate the limit cycle amplitude at this gain.

The approach used in this section can do a reasonable job of predicting the parameter dependency of limit cycle amplitude and frequency provided that (i) the basis set used is sufficiently rich to capture the dynamics throughout the range of parameters explored and (ii) the number of harmonics kept is adequate to represent the limit cycle temporal behavior. We saw that, for the current example, these conditions are satisfied for feedback amplitude of less than about 0.5.

These methods of analyzing periodic solutions can also be applied to the simplification and understanding of periodic behavior of complex models, such as general circulation models.

5. Large-amplitude limit cycle

Without the feedback interaction allowing the THC to influence the WDC amplitude, the model undergoes a second transition near $F_s = 2.35$ m yr^{-1} from a moderate amplitude, long period ($\sim$200 yr) limit cycle to a large amplitude, shorter period ($\sim$50 yr) limit cycle (PT04). In this latter regime, the THC is nearly shut-off in part of the cycle and oscillates between this and a strong THC state. In the presence of the new feedback from the THC to the WDC considered in this work, however, the corresponding regime ($F_s > 2$) is characterized by a stable steady-state THC that is thermally

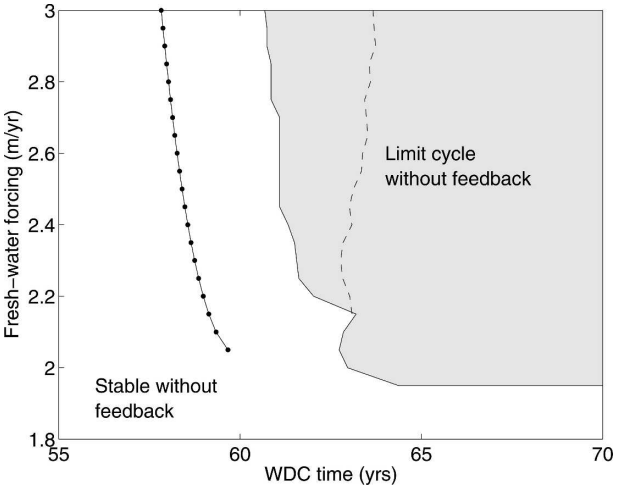

FIG. 10. Understanding the new stable equilibrium arising when the WDC is allowed to respond to changes in the SST ($k \neq 0$). The stability boundary without this feedback from the THC to the WDC is plotted as a function of WDC time scale and FW forcing: the shaded region corresponds to limit-cycle solutions, while the white area is characterized by a thermally dominant stable steady-state solution. Also plotted (line with dots) is the WDC as function of the FW forcing for the new stable thermally dominant stable equilibrium that appears at large FW forcing in the presence of the new feedback from the THC to the WDC. The nominal WDC strength has a 70-yr time scale, but the feedback increases the mean WDC strength, which brings the model into the stable region. The dashed line is the steady WDC time at the minimum stabilizing feedback gain, indicating that there is also a dynamic stabilizing effect.

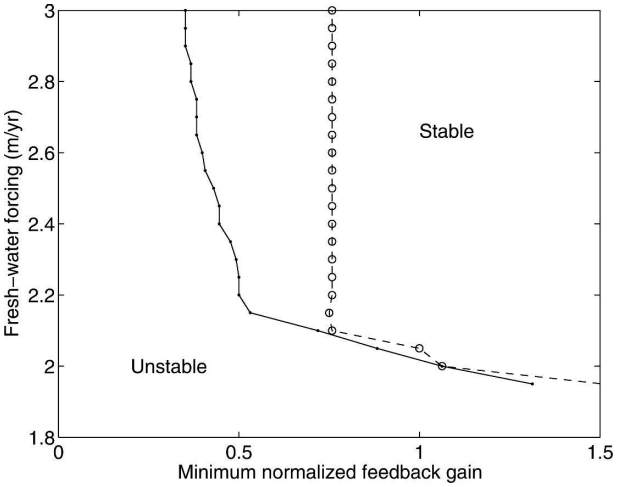

FIG. 11. Minimum amplitude k of the feedback from the THC to the WDC, for which the stable thermally dominant equilibrium at large FW forcing values exists as a function of the FW forcing. Shown for feedback time constant $\tau = 0$ (line with dots) and for $\tau = \tau_{\max}$ (line with circles).

dominant and only slightly weaker than the weak FW stable state (Fig. 4).

The primary mechanism underlying this stabilization results from an increase in the average strength of the wind-driven circulation due to the newly introduced feedback, with a stable equilibrium existing for sufficiently rapid WDC. This equilibrium also exists without feedback, for sufficiently rapid WDC. This can be seen from Fig. 10 in which the stable and limit cycle regimes are plotted for fixed WDC (without the new feedback) as a function of both the WDC circulation time and the FW forcing. For $2\pi\Omega^{-1} = 70$ yr the equilibrium loses stability for $F_s > 1.9$, but for larger Ω (shorter WDC circulation time) remains stable as FW forcing is increased. Note that the WDC axis of Fig. 7 in PT04 was inadvertently flipped about $2\pi\Omega^{-1} = 70$ yr, and with this correction also shows the stable region for sufficiently large Ω. For a feedback amplitude of $k = 1$, the WDC strength is shown as a function of FW forcing by the line with dots in Fig. 10.

The above explanation complements the more traditional physical mechanism given in section 2b in terms of the advection of salinity anomalies around the gyre. Together, the control analysis and the physical mechanism provide a fuller understanding of the model behavior.

We reiterate that, in addition to causing a steady change in the WDC, the feedback also responds to unsteady perturbations and therefore affects the stability of the mean flows. Specifically, we explained in section 2b that the feedback may be *stabilizing* if the THC is weak and the WDC strong. This is consistent with the location of the dashed line in Fig. 10. The solid line indicates the boundary between stable steady-state solutions (unshaded) and a time variable limit cycle solution (shaded) in the absence of the new feedback. This boundary with the new feedback included is shown by the dashed line. Clearly the feedback has made the stable region larger, indicating that it can indeed be stabilizing. This stabilization is due to two factors. First, the feedback makes the mean WDC stronger. Second, there is the dynamic stabilization involving the advection of salinity anomalies similar to that described earlier, but in the regime where the THC is weak and the WDC is strong.

The minimum normalized feedback gain required for stability, as a function of FW forcing, is shown in Fig. 11 for feedback time constants $\tau = 0$ and $\tau = \tau_{\max}$. As expected, the system remains stable for much lower feedback strengths if the WDC response is more nearly in phase with the THC perturbation that induced it. These stabilizing effects of the feedback that allows the WDC to change as function of the SST and THC may again be explained and predicted from a linearized analysis and a root locus plot, as in section 3. Note that

such an analysis is obviously incapable of determining whether the bifurcation at the stability boundary will be super- or subcritical.

6. Conclusions

While many previous studies have considered the separate variability and dynamics of the thermohaline circulation and the wind-driven circulation, the two are clearly strongly coupled. In this paper we have used a simple extension of the coupled WDC → THC model of PT04 to explore the dynamics of the fully coupled system. The new element that was added here is that the amplitude of the wind-driven circulation is determined via feedback from the sea surface temperature gradient, which is in turn influenced by the thermohaline circulation. This simple model illustrates that this feedback mechanism has the potential to significantly alter the dynamics of the coupled system.

For small freshwater flux where a steady-state solution exists (i.e., below the first Hopf bifurcation), the effect of the new feedback considered here is destabilizing. That is, one finds that self-sustained oscillations replace the steady-state solution for a smaller freshwater forcing than without this new feedback. Similarly, the amplitude of the limit cycle obtained for freshwater flux slightly higher than the bifurcation value is increased under feedback.

At higher freshwater forcing still, the WDC → THC system with a fixed WDC makes a transition into a large-amplitude limit cycle (PT04). We find here that allowing the WDC to change as function of the SST significantly affects this behavior and, instead of a very large amplitude THC oscillation, we find a new *stable* equilibrium point, which corresponds to a somewhat weaker thermally dominant THC. This new steady state exists even for fairly large freshwater forcing values and is purely the result of the adjustment allowed here of the WDC to changes in the SST.

We have analyzed the behavior of the coupled WDC → THC using both traditional oceanographic physical arguments and using methods borrowed from control engineering. We feel that these tools have the potential to be useful in many oceanographic and climate variability phenomena where feedbacks between different subsystems play a dominant role. Some of the physical mechanisms found here for the role of the two-way feedback between the THC and WDC are clearly specific to the very idealized representation of the two circulations that we have employed here. We feel, however, that the interesting results found here regarding the importance of properly representing the two-way feedback justify and motivate a further examination using more realistic models of the WDC and THC.

Acknowledgments. This work was supported by the James McDonnell Foundation. We thank Claudia Pasquero for her assistance with the model and additional advice. The comments of two anonymous reviewers helped to significantly improve the presentation.

REFERENCES

Anderson, D. L. T., and A. E. Gill, 1975: Spin up of a stratified ocean with application to upwelling. *Deep-Sea Res.,* **22,** 583–596.

Berloff, P. S., and S. P. Meacham, 1997: The dynamics of an equivalent-barotropic model of the wind-driven circulation. *J. Mar. Res.,* **55,** 407–451.

Cessi, P., and G. R. Ierley, 1995: Symmetry-breaking multiple equilibria in quasi-geostrophic, wind-driven flows. *J. Phys. Oceanogr.,* **25,** 1196–1205.

Dijkstra, H. A., 2005: *Nonlinear Physical Oceanography: A Dynamical Systems Approach to the Large Scale Ocean Circulation and El Niño.* 2d ed. Springer, 532 pp.

——, and C. A. Katsman, 1997: Temporal variability of the wind-driven quasi-geostrophic double gyre ocean circulation: Basic bifurcation diagrams. *Geophys. Astrophys. Fluid Dyn.,* **85** (3–4), 195–232.

Doyle, J. C., B. A. Francis, and A. R. Tannenbaum, 1992: *Feedback Control Theory.* Macmillan, 227 pp.

Eccles, F., and E. Tziperman, 2004: Nonlinear effects on ENSO's period. *J. Atmos. Sci.,* **61,** 474–482.

Farrell, B. F., and P. J. Ioannou, 2001: Accurate low-dimensional approximation of the linear dynamics of fluid flow. *J. Atmos. Sci.,* **58,** 2771–2789.

Franklin, G. F., J. D. Powell, and A. Emami-Naeini, 2002: *Feedback Control of Dynamic Systems.* 4th ed. Prentice Hall, 910 pp.

Gelb, A., and W. Vander Velde, 1968: *Multiple-input Describing Functions and Nonlinear System Design.* McGraw-Hill, 655 pp.

Ghil, M., Y. Feliks, and L. U. Sushama, 2002: Baroclinic and barotropic aspects of the wind-driven ocean circulation. *Physica D,* **167** (1–2), 1–35.

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

Hoang, S., P. Demey, O. Talagrand, and R. Baraille, 1997: A new reduced-order adaptive filter for state estimation in high-dimensional systems. *Automatica,* **33,** 1475–1498.

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

Jin, F.-F., 1997: An equatorial ocean recharge paradigm for ENSO. Part I: Conceptual model. *J. Atmos. Sci.,* **54,** 811–829.

Lumley, J. L., 1970: *Stochastic Tools in Turbulence.* Academic Press, 194 pp.

Marotzke, J., 1989: Instabilities and multiple steady states of the thermohaline circulation. *Oceanic Circulation Models: Combining Data and Dynamics,* D. L. T. Anderson and J. Willebrand, Eds., NATO ASI Series, Vol. 284, Kluwer, 501–511.

——, P. Welander, and J. Willebrand, 1988: Instability and mul-

tiple steady states in a meridional-plane model of the thermohaline circulation. *Tellus,* **40A,** 162–172.

Meacham, S. P., 2000: Low-frequency variability in the wind-driven circulation. *J. Phys. Oceanogr.,* **30,** 269–293.

Nadiga, B. T., and B. P. Luce, 2001: Global bifurcation of Shilnikov type in a double-gyre ocean model. *J. Phys. Oceanogr.,* **31,** 2669–2690.

Pasquero, C., and E. Tziperman, 2004: Effects of a wind-driven gyre on thermohaline circulation variability. *J. Phys. Oceanogr.,* **34,** 805–816.

Peixoto, J. P., and A. H. Oort, 1992: *Physics of Climate.* American Institute of Physics, 520 pp.

Primeau, F., 1998: Multiple equilibria of a double-gyre ocean model with super-slip boundary conditions. *J. Phys. Oceanogr.,* **28,** 2130–2147.

——, 2002: Multiple equilibria and low-frequency variability of the wind-driven ocean circulation. *J. Phys. Oceanogr.,* **32,** 2236–2256.

Rowley, C. W., 2005: Model reduction for fluids, using balanced proper orthogonal decomposition. *Int. J. Bifurcation Chaos,* **15,** 997–1013.

——, T. Colonius, and R. M. Murray, 2004: Model reduction for compressible flows using POD and Galerkin projection. *Physica D,* **189** (1-2), 115–129.

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

——, and C. Rooth, 1968: On the interaction of gravitational and dynamic forcing in simple circulation models. *Deep-Sea Res.,* **15,** 165–170.

Timmermann, A., H. U. Voss, and R. Pasmanter, 2001: Empirical dynamical system modeling of ENSO using nonlinear inverse techniques. *J. Phys. Oceanogr.,* **31,** 1579–1598.

Tippett, M. K., and S. E. Cohn, 2001: Adjoints and low-rank covariance representation. *Nonlinear Processes Geophys.,* **8** (6), 331–340.

——, ——, R. Todling, and D. Marchesin, 2000: Low-dimensional representation of error covariance. *Tellus,* **52A,** 533–553.

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.

van Veen, L., 2003: Overturning and wind-driven circulation in a low-order ocean-atmosphere model. *Dyn. Atmos. Oceans,* **37,** 197–221.