The Role of Integration Time in Determining a Steady State through Data Assimilation

JOCHEM MAROTZKE

Center for Meteorology and Physical Oceanography, Department of Earth, Atmospheric, and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts

(Manuscript received 5 August 1991, in final form 10 March 1992)

ABSTRACT

The length of time an ocean model and its adjoint should be integrated in determining a steady state compatible with observed data is investigated. The starting point is based upon a suggestion that only one time step is required. This method fails to converge to an acceptable solution when applied to a general circulation model (GCM) of the North Atlantic. Using a very coarse resolution GCM in an idealized geometry, the problem is traced to the interplay of convective adjustment and the very short integration time.

The general assimilation technique is explored using a very simple model, a linear first-order equation with forcing and damping. The model is unable to provide a dynamical coupling between the forcing and the model response, owing to a mismatch of integration time and adjustment time scale. Coupling can be enforced in the simple linear model through a careful choice of weighting factors, a strategy excluded in the GCM due to the presence of very fast processes like convective adjustment. An integration over a sufficiently long time can avoid the problems encountered. Experiments with the idealized GCM prove successful for longer integrations, and a tentative upper limit of 50 years is given for inversions aiming at the main thermocline structure.

1. Introduction

Combining observed data with the conservation principles underlying an oceanic circulation model, or, in other words, the assimilation of observations into a numerical ocean model, has received increasing attention in the past few years. Data assimilation makes possible a systematic test of model dynamics against observations; if a model is found to be compatible with the data, one has obtained dynamically consistent estimates of the oceanic state and the surface fluxes that drive the ocean circulation. Efforts are under way at the Massachusetts Institute of Technology to determine whether a state of the North Atlantic Ocean can be estimated that is consistent both with the hydrographic data obtained during the early 1980s and with a general circulation model of the North Atlantic in dynamical steady state. A solution so obtained (given its existence, which is not a trivial determination) should provide us with improved estimates of some of the climatically most relevant quantities, such as the time-averaged thermocline structure and circulation, meridional transports of heat and fresh water, and air–sea exchanges of heat, fresh water, and momentum.

The North Atlantic inversion described in this paper is an attempt to extend the work described by Tziperman et al. (1990), using the same assimilation technique (the "adjoint method," described below) but using a more comprehensive dataset and a higher-resolution model. Some fundamental methodological problems arose, however, the nature and eventual tentative solution of which are the focus of this paper. For illustration, a review of the method is now given, followed by a brief account of some of Tziperman et al.'s (1990) results.

The assimilation technique makes use of an ocean model together with its "adjoint." The equations constituting the adjoint model can be derived as an application of the Pontryagin minimum principle in control theory (for derivations see, e.g., Talagrand and Courtier 1987; Thacker and Long 1988; Wunsch 1988). The basic idea can be stated as follows. Given a model prediction $\mathbf{x}^{\text{model}}$ for some ocean domain as a function of assumed initial and boundary conditions, the deviation of the model prediction can be measured from observed data $\mathbf{x}^{\text{obs}}$ by the means of a quadratic cost or objective function J defined as

$$J = \frac{1}{2}(\mathbf{x}^{\text{obs}} - \mathbf{x}^{\text{model}})^T \mathbf{W}_D(\mathbf{x}^{\text{obs}} - \mathbf{x}^{\text{model}}), \quad (1)$$

where $\mathbf{W}_D$ is a weighting matrix, ideally the inverse of the observation error covariance matrix. Our aim is to minimize the objective function, subject to the model constraints, which are enforced through a Lagrange multiplier technique. In other words, we want to find the combination of initial and boundary conditions that results in the best fit of the model time trajectory to the data. Given a first estimate of the independent

Corresponding author address: Dr. Jochem Marotzke, Center for Meteorology and Physical Oceanography, Department of Earth, Atmospheric and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, MA 02139.

variables, initial state and boundary forcing, and consequently a first-guess time history of the model, it may be shown that the integration of the adjoint equations, *backward in time*, yields the Lagrange multipliers used to augment J in Eq. (1); the multipliers constitute the gradient of the objective function with respect to all the independent (or control) variables (see, e.g., Thacker and Long 1988). The merit of the adjoint model thus resides in the gradient information it provides, which makes possible a very efficient minimization of the objective function, since the computational cost of an adjacent model run is usually no higher than that of a forward model run.

The aforementioned description alludes to a time-dependent problem. If, in contrast, a steady-state fit of a model to a climatological dataset is sought, the most straightforward manner of using an oceanic circulation model would be to employ a steady model that calculates the oceanic state as a function of the surface forcing fields, which are wind stress, heat fluxes, and freshwater fluxes, plus boundary conditions at open lateral boundaries. Most general circulation models currently in use, however, do not calculate the steady state directly as a function of the forcing variables. Instead, they perform, a time stepping from some arbitrarily chosen initial state until all transients have been damped by friction [e.g., see Bryan (1984) for details of the technique, and Marotzke and Willebrand (1991) for the uniqueness of solutions thus found]. Consequently, Tziperman and Thacker (1989) proposed the following strategy. The quality of the model fit to the data is measured with an objective function penalizing the quadratic deviation of the estimated fields from the data, weighted with an appropriate covariance matrix. To enforce steadiness, the cost function also includes a term penalizing the quadratic difference between the model state at the end of the integration period and the initial state. The objective function now reads:

$$J = \frac{1}{2}(\mathbf{x}^{\mathrm{obs}} - \mathbf{x}^{\mathrm{model}})^{\mathrm{T}}\mathbf{W}_D(\mathbf{x}^{\mathrm{obs}} - \mathbf{x}^{\mathrm{model}}) + \frac{1}{2}(\mathbf{x}_F^{\mathrm{model}} - \mathbf{x}_0^{\mathrm{model}})^{\mathrm{T}}\mathbf{W}_S(\mathbf{x}_F^{\mathrm{model}} - \mathbf{x}_0^{\mathrm{model}}), \quad (2)$$

where $\mathbf{x}_F$ is the model state at the end of the integration period, and $\mathbf{x}_0$ the initial state; $\mathbf{W}_S$ is a weight matrix specifying how stringently steadiness is demanded. As before, J is minimized subject to the model constraints, enforced through a Lagrange multiplier technique. The approach employs the conservation equations governing the numerical model as "strong constraints," whereas the demand for steadiness and a best fit with the data are "weak constraints." If a truly steady-state, time-independent model were fitted to the data, both the conservation equations and the steadiness demand would be strong constraints.

Tziperman and Thacker (1989) suggested carrying the integration over only a single time step, arguing that a steady state is sufficiently determined by demanding that the change over one time step be very small. From a first-guess initial state, the model steps forward in time for this one step, after which the cost function is calculated. Then the adjoint of the model is integrated backward for one time step, yielding the gradient of the cost function with respect to the independent variables: surface forcing and initial state. Using the efficiently determined gradient of the objective function, a conjugate gradient descent algorithm is employed to modify the control variables. This iterative method gives a better estimate for the initial state and the surface forcing for each subsequent iteration step; the procedure is continued until a combination of initial and boundary conditions is found that is optimal in the sense that it is both close to the data and yields a state that changes very little over the one time step under consideration. This solution is intended to be a steady-state best fit to the observations.

The procedure produced the correct state of a barotropic quasigeostrophic model in identical-twin experiments, where the "data" were chosen from a steady state obtained through a time-stepping procedure (Tziperman and Thacker 1989). Using the same strategy, Tziperman et al. (1992a) were also able to reconstruct, partially, the steady state of a general circulation model (GCM) in an identical-twin experiment. When applied to a model of the North Atlantic, using Levitus's (1982) climatology, however, the method did not find a meaningful solution. The optimization resulted in a state that was very close to the data but still far away from a steady state; the best estimate exhibited very large residual temporal drifts in the temperature and salinity fields. For example, there were time rates of change of temperature in the subsurface layer that amounted to more than 10°C y^{-1} in some areas. Consequently, Tziperman et al. (1990) concluded that there was no time-mean state of their model North Atlantic compatible with the data.

Another possibility, however, is that this result emerged not because the data and the model were incompatible with each other but because the optimization method had not succeeded in finding the minimum of the objective function. Consider Eq. (2), which shows that the cost function J can be written as the sum of two parts $J_{\mathrm{DATA}} + J_{\mathrm{STEADY}}$, where J_{DATA} denotes the contributions due to the deviations of the model from the data, and J_{STEADY} represents the penalty due to temporal drifts. In the general case where the constraints set by the data and the demand for steadiness are not exactly compatible, minimizing J means that a compromise is sought between the two competing sets of constraints. A compromise would be deemed reasonable if the respective penalties for deviating from the two sets of constraints were almost equally large, that is, J_{DATA} and J_{STEADY} should be approximately equal. If, additionally, both the residual data misfits and the residual temporal drifts match their a priori estimates as expressed through the weighting matrices, a compatible solution has been found. The solution of the North Atlantic inversion, however, as obtained by Tziperman et al. (1990), does not pass the first of the two tests. The ratio $J_{\rm STEADY}:J_{\rm DATA}$ is about 36:1, whereas a true compromise between the competing constraints would deviate more from the observed data, in order to make the model more steady.

Our own attempts at inverting North Atlantic data with the help of a GCM have revealed very similar problems and have motivated an investigation of how, in general, a steady-state fit of a model to observations can be found. We focus on the role of the integration time over which a prognostic model must be run. The paper is organized as follows. First, the practical difficulties in the North Atlantic model inversions will be described, suggesting that the mismatch between $J_{\rm STEADY}$ and $J_{\rm DATA}$ is the rule rather than the exception. Using a very coarse resolution GCM in an idealized geometry, we trace the problem to the interplay of convective adjustment and the very short integration time (section 2). The assimilation technique applied to the GCM inversions is then tested using a very simple model, a linear first-order equation with forcing and damping. Attention is focused on consequences of integrating over times shorter than the principal adjustment time scales of the model. Section 4 shows the results obtained through longer integrations of the GCM with the idealized geometry. A discussion follows in section 5.

2. General circulation model: Problems with the one time-step approach

The work described here started off by applying Tziperman et al.'s (1990) model and inversion strategy to a model with considerably higher spatial resolution, and a more recent dataset. The original aim was to produce an estimate of the North Atlantic in steady state, and within observation errors of the hydrography from the early 1980s, and the estimates of the air–sea fluxes of heat, fresh water, and momentum. A brief description of the inverse model and a summary of the North Atlantic experiments are given now.

The model extends from 9.5° to 59.5°N, 80° to 10°W with a resolution of 1° zonally, 2° meridionally, and 24 levels vertically. For the initial experiments all lateral walls are closed; in particular, there is no Straits of Florida. Apart from minor modifications, the model code of Tziperman et al. (1990, 1992a) is used in which the forward model is the GCM developed at the Geophysical Fluid Dynamics Laboratory (GFDL) at Princeton University (Bryan 1969; Cox 1984), in an abridged version with simplified momentum equations but the complete conservation equations for heat and salt, including a convection parameterization to eliminate static instability. The adjoint to the GFDL GCM was originally developed by W. Thacker, R. Long, and S. Hwang at the Atlantic Oceanographic and Meteorological Laboratory (AOML) in Miami; the simplifications to the model as used here are due to E. Tziperman. For a more complete model description, see Tziperman et al. (1990, 1992a) and Marotzke and Wunsch (1992). Experience suggests that any ocean model of comparable complexity would behave similarly in the optimization. Thus, the model used here serves as a prototype of any fairly realistic ocean model; its details are not essential to the present discussion.

The data consist of the objectively analyzed temperature and salinity fields compiled and described by Fukumori and Wunsch (1991) and Fukumori et al. (1991), the wind-stress data calculated from the European Centre for Medium-Range Weather Forecasts (ECMWF) analyses by Trenberth et al. (1989), the evaporation minus precipitation $(E - P)$ analysis by Schmitt et al. (1989), and the surface heat flux estimates by Isemer et al. (1989). The hydrographic and wind-stress data cover the period of the first half of the 1980s; the other two datasets use data compiled over a larger time span. If necessary, the data were subsampled on, or interpolated to, the model grid, resulting in one observation per model grid point.

The cost function to be minimized is formulated schematically in Eq. (2); both the data and the steadiness constraints are weak ones. It remains to specify the strength of the constraint that there be only small changes over the time step; that is, what weight factors $\mathbf{W}_S$ should be chosen. Tziperman et al. (1990) give a rationale for specifying the associated variances. If δT denotes a typical observation error at a certain depth, the data cannot exclude a drift of oceanic temperatures that is *slower* than $\delta T/\tau$, where τ is of the order of the time over which the measurements were taken. Thus, it is reasonable to demand that the equilibrium to be determined is steady only up to $\delta T/\tau$ (likewise for salinity). This concept can be translated into a recipe of how to specify $\mathbf{W}_S$, provided the weighting matrix for the data misfit part of the objective function is given. The model computes time rates of change of temperature and salinity over the one time-step integration, that is, changes in temperature and salinity divided by the length of the time step. These time rates of change are then extrapolated over time τ, $(\mathbf{x}_F^{\rm model} - \mathbf{x}_0^{\rm model})\tau/\Delta t$, where Δt is the time step. If the extrapolated drifts are smaller than the observation error, they cannot be detected from the data and hence should not be penalized heavily. Consequently, it is demanded through the weight matrices that the rms values of these extrapolated residual drifts should equal the rms errors ascribed a priori to the measurements. In terms of the matrices $\mathbf{W}_D$ and $\mathbf{W}_S$ introduced in Eqs. (1) and (2),

$$\mathbf{W}_S = \mathbf{W}_D\left(\frac{\tau}{\Delta t}\right)^2. \qquad (3)$$

During each minimization, $\mathbf{W}_D$, $\mathbf{W}_S$, and τ are held constant; furthermore, the larger τ, the more stringent is the demand for steadiness because, loosely speaking, τ marks the time that has to pass before a temporal drift can be detected; τ was chosen somewhat arbitrarily as 15 years in Tziperman et al. (1990).

The matrices $\mathbf{W}_D$ and $\mathbf{W}_S$ are assumed diagonal and normalized following Tziperman et al., that is, each diagonal element of $\mathbf{W}_D$ consists of the inverse observation error variance, multiplied by the volume of the grid cell under consideration and divided by the total volume of the model ocean [compared to $\mathbf{W}_D$, each element of $\mathbf{W}_S$ contains a very large additional factor $(\tau/\Delta t)^2$, according to Eq. (3)]. The normalization gives a simple rule of thumb for evaluating whether the outcome of an inversion actually represents a compatible solution.

The results of the North Atlantic experiments, which employ a τ varying between 1 and 15 years, can be summarized as follows:

1) Never did the minimization find a solution in which the data misfit contribution to the final value of the cost function was the same order of magnitude as the steady-penalty contribution. The larger of J_{STEADY} and J_{DATA} was always much larger than unity.

2) Which of J_{STEADY} and J_{DATA} was the dominant contributor to the final J depended on the first guess of the initial conditions.

3) The minimization progressed much further if static instability in the forward model was not removed by a convection parameterization.

The minimization method did not find a real compromise between the demand for steadiness and the data constraint, which could mean that the optimization failed to progress, as opposed to converged. The reason could be a very jagged surface of the objective function, caused by highly nonlinear processes in the physics of the model, such as convective overturning.

To investigate this notion further, a "small" version of the GFDL GCM (rectangular box, flat bottom, 6 × 6 interior horizontal grid points, 5° by 5° resolution, four levels) is used to make a large number of fast tests possible. A somewhat arbitrary set of "observed data" is defined (Fig. 1), meaning to reflect that water gets colder and fresher as one moves from lower to higher latitudes (Fig. 1a), the principal zonal structure of the subtropical gyre (Fig. 1b), and an idealized exponential profile with depth (Fig. 1c). Some reasonable values are assumed for the "observation error" of the synthetic data; the weight matrices $\mathbf{W}_D$ and $\mathbf{W}_S$ are constructed in a way analogous to the North Atlantic model. A deliberate decision was made not to use the results of a forward model run as data (identical-twin experiment), because both the method under investigation here had proved successful in twin experiments and the observed data are not likely to be fully compatible with the model physics. An identical-twin experiment poses too easy a test.

In addition to the data, a second reference point in state vector space is the steady state obtained through a forward integration, using some reasonable forcing functions (heat gain and freshwater loss linearly decreasing with latitude, a shift from easterly to westerly winds as one moves south to north). Any linear combination of these two points in state vector space defines a plausible initial guess for subsequent minimizations. The parameter μ specifying each linear combination is chosen such that $\mu = 0$ defines the data, and $\mu = 1$ the steady state obtained through the forward integration. The time τ over which steadiness is demanded and that enters $\mathbf{W}_S$ is set to 5 years. The result of the minimizations is very unsatisfactory: close-neighbor initial guesses produce very different final cost function values, which at best are as high as 250 (Fig. 2).

Turning off convective adjustment leads to a reduction in cost to about 7, the data misfit and the steady residuals having the same order of magnitude. Three different starting points, however, produce three different end points, and it takes more than 1500 iteration

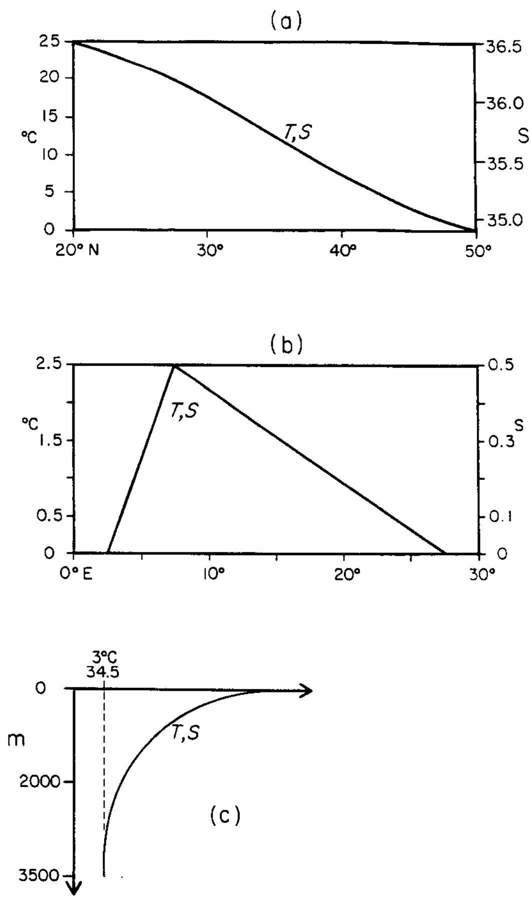

FIG. 1. Principal structure of (a) meridional, (b) zonal, and (c) depth dependence of the simulated data used for the box GCM.

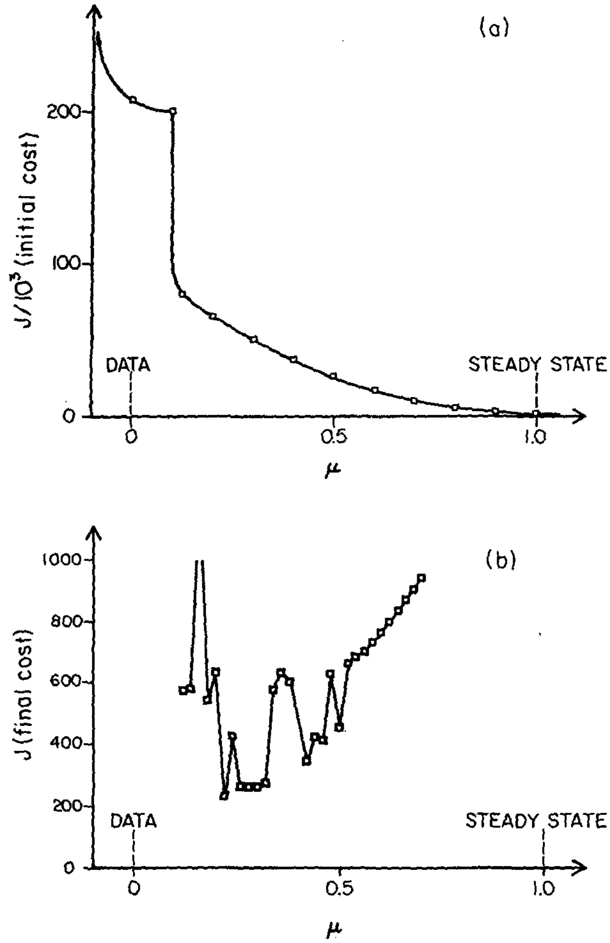

FIG. 2. Box GCM, with convective adjustment. (a) Initial and (b) final value of the cost function for various starting points, which were chosen by interpolating between the data displayed in Fig. 1 ($\mu = 0$) and one reference steady state, obtained through a forward integration ($\mu = 1$); $\tau = 5$ years.

steps to reach them. From the cost function evaluated between these three end points (marked A, B, C in Fig. 3), it seems that they form local minima in a "hilly" environment (Fig. 3). A complete linearization of the problem is achieved by turning off the dependence of the velocity field on T and S, in addition to turning off convective adjustment. That is, the first-guess flow field is not changed in subsequent iterations, and the inversion consists of a fit of the linear advection–diffusion equations to the data shown in Fig. 1. Unlike before, the same minimum is obtained from very different first-guess initial conditions, chosen again as linear combinations of the data and the steady state of the forward model run, with μ ranging from -10 to $+10$ (compare Fig. 2). The final cost function value is 35, much higher than in the nonlinear case, since the model cannot produce a changed velocity field to obtain a better fit.

Convective adjustment made any reasonable progress impossible. To understand this result, imagine two vertically adjacent boxes to be weakly, stably stratified. If the gradient of the cost function suggests increasing salinity in the upper box, the consequence is a decrease in cost even if the stratification becomes unstable. Static instability then leads to strong vertical mixing of the two boxes (details of which depend on which convective adjustment scheme is employed), and a drastic change occurs over just one time step, leading to a steplike increase in steady cost, since $\tau \gg \Delta t$. The model will thus refrain from an increase in top salinity that would cause convection. Such a behavior was observed for one grid box in a case where convection was switched on again, after a minimum had been found without convection. It is not quite clear if this argument is fully valid for a model with many degrees of freedom. The aforementioned experiments show, however, that a necessary physical ingredient of a general circulation model, convective adjustment, inhibits any satisfactory progress of the minimization, possibly because changes of the model were considered over only one time step. The model state cannot adjust to an increase in convective activity, although this might eventually lead to a better steady-state fit to the data. The role of the adjustment time scales, as compared to the integration time, will be investigated in a more general context in the next section, using a very simple dynamical model for which analytical solutions can readily be found. We will return to the GCM, and the interplay between convective overturning and the short integration times, in the discussion of the results obtained using the simple model.

3. Linear model with one degree of freedom

Following Willebrand (1991, private communication), a linear first-order equation with forcing and damping will be considered. Observations are assumed for the forcing and the model state, and both the exact

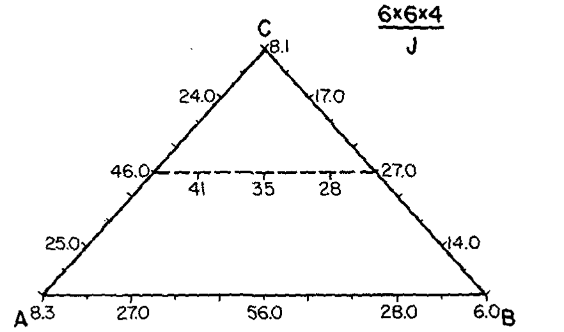

FIG. 3. Box GCM, without convective adjustment, $\tau = 5$ years. A, B, and C are the end points of minimizations with three different starting points. Shown is the cost function at points that were found by interpolating between the points A, B, and C (i.e., between the end points of three different minimizations).

solution of a steady-state fit and the solution obtained through minimization of a cost function analogous to Eq. (2) are calculated. Stated differently, the demand for steadiness is treated once as a strong constraint and once as a weak constraint. The dynamical model equation is always applied as a strong constraint. To obtain a more general result, the time span under consideration will not necessarily be confined to a single time step.

The model equation reads

$$\dot{u} = -\gamma u + f \tag{4}$$

(the forcing f is not to be confused with the Coriolis parameter), with a steady state given by

$$u = \frac{f}{\gamma}. \tag{5}$$

Given observations $u^{\text{obs}}, f^{\text{obs}}$, a best fit of the steady-state model Eq. (5) is found by setting to zero the partial derivatives with respect to u, f, and λ, of J, which is defined as

$$J = \frac{1}{2}\left(\frac{u - u^{\text{obs}}}{\sigma_u}\right)^2 + \frac{1}{2}\left(\frac{f - f^{\text{obs}}}{\sigma_f}\right)^2 + \lambda(u - f/\gamma); \tag{6}$$

λ is the Lagrange multiplier. One finds the optimal solution:

$$\hat{u} = \frac{\dfrac{f^{\text{obs}}}{\gamma} + u^{\text{obs}}\left(\dfrac{\sigma_f}{\gamma\sigma_u}\right)^2}{1 + \left(\dfrac{\sigma_f}{\gamma\sigma_u}\right)^2} \tag{7a}$$

$$\hat{f} = \gamma\hat{u} \tag{7b}$$

$$J_{\text{final}} = \frac{1}{2}\left(\frac{\hat{u}_1 - u^{\text{obs}}}{\sigma_u}\right)^2 + \frac{1}{2}\left(\frac{\hat{f} - f^{\text{obs}}}{\sigma_f}\right)^2 = \frac{1}{2\sigma_u^2}\frac{\left(\dfrac{f^{\text{obs}}}{\gamma} - u^{\text{obs}}\right)^2}{1 + \left(\dfrac{\sigma_f}{\gamma\sigma_u}\right)^2}. \tag{7c}$$

Alternatively, we can treat the demand for steadiness as a weak constraint, first integrating the time-dependent model up to some time t, giving

$$u_1 \equiv u(t) = u_0 e^{-\gamma t} + \frac{f}{\gamma}(1 - e^{-\gamma t}) \tag{8}$$

where $u_0 \equiv u(0)$. The cost function now contains a term penalizing the difference between u_1 and u_0:

$$J = \frac{1}{2}\left(\frac{u_1 - u^{\text{obs}}}{\sigma_u}\right)^2 + \frac{1}{2}\left(\frac{f - f^{\text{obs}}}{\sigma_f}\right)^2 + \frac{1}{2}\left(\frac{u_1 - u_0}{\sigma_\Delta}\right)^2 + \lambda\left[u_1 - u_0 e^{-\gamma t} - \frac{f}{\gamma}(1 - e^{-\gamma t})\right]. \tag{9}$$

Again, λ is the Lagrange multiplier. The one time-step approach is readily seen as a special case. In (8) u_1 is expressed as the explicit solution of the governing equation, instead of stating it in terms of a numerical scheme. If we assume $\gamma t \ll 1$ and replace t by Δt in (8), a Taylor expansion to first order reduces (8) to just one Euler forward time step.

The optimal solution is found by setting the partial derivatives of J with respect to λ, u_1, u_0, and f to zero, which yields, after some algebra,

$$\hat{u}_1 = \frac{\dfrac{f^{\text{obs}}}{\gamma} + u^{\text{obs}}\left[\dfrac{\sigma_f^2}{\gamma^2\sigma_u^2} + \dfrac{\sigma_\Delta^2}{\sigma_u^2}\dfrac{e^{-2\gamma t}}{(1 - e^{-\gamma t})^2}\right]}{1 + \dfrac{\sigma_f^2}{\gamma^2\sigma_u^2} + \dfrac{\sigma_\Delta^2}{\sigma_u^2}\dfrac{e^{-2\gamma t}}{(1 - e^{-\gamma t})^2}}. \tag{10}$$

Defining

$$D \equiv 1 + \frac{\sigma_f^2}{\gamma^2\sigma_u^2} + \frac{\sigma_\Delta^2}{\sigma_u^2}\frac{e^{-2\gamma t}}{(1 - e^{-\gamma t})^2}, \tag{11}$$

(10) and (11) give

$$\hat{u}_1 - u^{\text{obs}} = \frac{1}{D}\left[\frac{f^{\text{obs}}}{\gamma} - u^{\text{obs}}\right]. \tag{12}$$

It can readily be shown that

$$\hat{f} - f^{\text{obs}} = -\frac{1}{\gamma}\frac{\sigma_f^2}{\sigma_u^2}\frac{1}{D}\left[\frac{f^{\text{obs}}}{\gamma} - u^{\text{obs}}\right] \tag{13}$$

$$\hat{u}_1 - \hat{u}_0 = \frac{\sigma_\Delta^2}{\sigma_u^2}\frac{e^{-\gamma t}}{(1 - e^{-\gamma t})}\frac{1}{D}\left[\frac{f^{\text{obs}}}{\gamma} - u^{\text{obs}}\right]. \tag{14}$$

The final cost function

$$J_{\text{final}} = \frac{1}{2}\left(\frac{\hat{u}_1 - u^{\text{obs}}}{\sigma_u}\right)^2 + \frac{1}{2}\left(\frac{\hat{f} - f^{\text{obs}}}{\sigma_f}\right)^2 + \frac{1}{2}\left(\frac{\hat{u}_1 - \hat{u}_0}{\sigma_\Delta}\right)^2 \tag{15}$$

takes the value

$$J_{\text{final}} = \frac{1}{2}\frac{(f^{\text{obs}}/\gamma - u^{\text{obs}})^2}{\sigma_u^2 D}, \tag{16}$$

as can be seen using Eqs. (11)–(14). The relative contributions to the final value of J, from data misfit, forcing misfit, and steady misfit [as they appear in (15)], are (note that they sum to D)

$$\text{data misfit:} \quad 1, \tag{17a}$$

$$\text{forcing misfit:} \quad \frac{\sigma_f^2}{\gamma^2\sigma_u^2}, \tag{17b}$$

$$\text{steady misfit:} \quad \frac{\sigma_\Delta^2}{\sigma_u^2}\frac{e^{-2\gamma t}}{(1 - e^{-\gamma t})^2}. \tag{17c}$$

For any given σ_Δ, the solution (7) of the steady model can be recovered exactly for large enough γt. The steady misfit part (17c) of the final cost function will tend to zero in this limit. Ironically, however, the total final cost increases [the denominator D in (16) decreases].

Let us now turn to the special case $\gamma t \ll 1$, which includes the one time-step approach. As mentioned before, replacing t by Δt in Eq. (8) and a Taylor expansion to first order would reduce (8) to just one Euler forward time step.

Expanding the exponentials in the solution (10)–(14) to the lowest nonvanishing order yields

$$D \approx 1 + \frac{\sigma_f^2}{\gamma^2\sigma_u^2} + \frac{\sigma_\Delta^2}{(\gamma t)^2\sigma_u^2} \tag{18}$$

$$\hat{u}_1 \approx \frac{1}{D}\left\{\frac{f^{\text{obs}}}{\gamma} + u^{\text{obs}}\left[\frac{\sigma_f^2}{\gamma^2\sigma_u^2} + \frac{\sigma_\Delta^2}{(\gamma t)^2\sigma_u^2}\right]\right\} \tag{19}$$

$$\hat{u}_1 - \hat{u}_0 \approx (\gamma t)\frac{\sigma_\Delta^2}{(\gamma t)^2\sigma_u^2}\frac{1}{D}\left[\frac{f^{\text{obs}}}{\gamma} - u^{\text{obs}}\right]. \tag{20}$$

Obviously, the results depend on the choice of σ_Δ. Notice first that we set, according to Eq. (3),

$$\sigma_\Delta = (t/\tau)\sigma_u \tag{21}$$

because

$$W_D = \sigma_u^{-2} \tag{22a}$$

$$W_S = \sigma_\Delta^{-2}. \tag{22b}$$

Equation (21) again expresses the demand that the rms temporal drift during the integration time t, when extrapolated over time τ, be just the rms observation error. On specifying σ_Δ, or equivalently τ, three different regimes can be distinguished.

(i) $\sigma_\Delta \ll (\gamma t)\sigma_u$, equivalent to $\gamma\tau \gg 1$

This result occurs if steadiness is demanded over a time span much longer than the principal time scale of the process under consideration. The results reduce to the case of a steady model [Eqs. (7a)–(7c)]. In the experiments of section 2, $\tau = O(10\text{ yr})$; γ^{-1} could be the adjustment time scale of the surface fields to changes in the forcing.

(ii) $\sigma_\Delta \sim (\gamma t)\sigma_u$ or $\gamma\tau \sim 1$

Steadiness is demanded over the principal time scale of the model, which for $\gamma^{-1} \sim \tau \sim 10$ yr would be the adjustment time scale of the main thermocline. We obtain

$$D = 2 + \frac{\sigma_f^2}{\gamma^2\sigma_u^2} \tag{23}$$

$$\hat{u}_1 = \frac{1}{D}\left\{\frac{f^{\text{obs}}}{\gamma} + u^{\text{obs}}\left[\frac{\sigma_f^2}{\gamma^2\sigma_u^2} + 1\right]\right\} \tag{24}$$

$$\hat{u}_1 - \hat{u}_0 = (\gamma t)\frac{1}{D}\left[\frac{f^{\text{obs}}}{\gamma} - u^{\text{obs}}\right]. \tag{25}$$

Compared to the results from case (i), there is a relative downweighting of the forcing observation in the determination of $\hat{u}_1$. It is readily shown that there is a corresponding downweighting of u^{obs} in the determination of $\hat{f}$. It cannot generally be said which of the two results, case (i) or case (ii), is "better." It may be desirable to have a drift as expressed in (25) in the model solution. If the measurements are taken over the time τ, they produce no evidence that such a drift can be excluded.

(iii) $\sigma_\Delta \gg (\gamma t)\sigma_u$ or $\gamma\tau \ll 1$

Keeping τ as before, γ^{-1} stands for the very long adjustment times of the deep ocean, that is, centuries or millenia. Equation (18) is, in this limit,

$$D \approx \frac{\sigma_\Delta^2}{(\gamma t)^2\sigma_u^2} \gg 1. \tag{26}$$

From (19) and (13), respectively, neither $\hat{u}_1$ nor $\hat{f}$ deviate much from the observations:

$$\hat{u}_1 \approx u^{\text{obs}} \tag{27}$$

$$\hat{f} \approx f^{\text{obs}}. \tag{28}$$

There is essentially no dynamical coupling between the measurements of model state and forcing, respectively. We can improve neither of the measurements because we cannot match the principal adjustment time scale of the model.

In this simple model we can always make σ_Δ small enough to enforce a coupling between forcing and model response. We should, however, be aware of the fact that we can infer improved estimates of the surface fluxes only from the measurements that are influenced by the fast processes, in the sense of cases (i) and (ii). Measurements of the very slowly varying fields are likely to remain inactive in the inversion.

The aforementioned considerations can readily be extended to a model involving two linearly coupled layers. The upper layer is assumed to have a short adjustment time scale γ_1^{-1}, the lower layer has a long time scale γ_2^{-1}. The algebra is straightforward but very tedious and will not be presented here. The main conclusions can be summarized as follows. The time scale τ entering the steady penalty weights has to be compared to the time scales in each of the layers separately. If the two adjustment time scales are very different, it is possible for the upper layer to adjust in the sense discussed in case (i), whereas the lower layer does not. Specifically, if $\gamma_2\tau \ll 1$, one finds that the lower layer is essentially decoupled from the upper one and the surface forcing, as the inversion proceeds. The lower layer will not deviate from the observations, and the analysis for the upper layer can be performed along exactly the same path as for the one-layer model.

What can the linear models tell us about the GCM inversion? Let us assume that we can make a separation of time scales of several months for the surface, a decade

TABLE 1. Box GCM, experiments 1 and 2. Initial guess, integration time t, and time τ over which steadiness is demanded, for the different stages of the experiments.

	Experiment 1					Experiment 2		
	1a	1b	1c	1d	1e	2a	2b	2c
First guess	Data	Data	Expt. 1b	Expt. 1b	Expt. 1d	Data	Expt. 2a	Expt. 2b
t	10 days	2.7 years	16.4 years	54.8 years	274 years	2.7 years	54.8 years	274 years
τ	50 years	50 years	50 years	50 years	300 years	10 years	50 years	300 years

for the thermocline, and a millenium for the deep ocean. Applying the one time-step approach and demanding steadiness over 5–15 years, the preceding analysis suggests that we might expect equilibrium between surface fields and forcing, a drift in the thermocline, and a deep ocean that does not respond at all to changes in the estimated surface fluxes, but remains very close to the data.

In the linear model steadiness can be enforced by making the σ_Δ arbitrarily small. For the GCM this would cause at least two problems. There might be numerical instabilities because the steadiness parts of cost and gradient take very large numerical values; in addition, we saw in section 2 that with reasonable choices for τ in order to adjust the thermocline, convective adjustment inhibited a sensible progress in the minimization.

The solution (10)–(14), however, suggests an alternative if one wants to couple observations of the ocean interior to estimates of the surface fluxes: the residual temporal drift can be reduced not only by a small σ_Δ but also by a longer integration time, which would cause the exponential factor in (17c) to diminish rapidly. Instead of enforcing steadiness of these "slow modes" through the weights, their time scales would be resolved explicitly. Once the mismatch between integration and adjustment time scales is removed, the data misfit weights $\mathbf{W}_D$ and the steadiness weights $\mathbf{W}_S$ have the same order of magnitude [cf. Eqs. (3), (21), and (22)], and it can be expected that the surface of the objective function is much less jagged. Notice also that the discussion has been confined to the time scales of the problem; in an ocean circulation model, another aspect may enter because we also have to consider how information propagates spatially. For example, the model temperature somewhere in the deep thermocline may not be compatible with the data. Ultimately, one wants to adjust the surface fluxes so that the misfit vanishes. How long must an integration of the adjoint model last until the information about the model–data misfit at great depths reaches the surface and, conversely, how long does it take the forward model to give an improved estimate? Intuitively, it seems unlikely that, even without convective adjustment and with large steadiness weights, an integration over a few days should yield the desired coupling. It seems inevitable that the integration must be performed over much longer than just a single time step.

4. Box GCM calculations over more than one time step

Encouraged by the foregoing analysis, the model of Tziperman et al. (1990, 1992a) was extended to enable a computation over an arbitrary number of time steps for both the forward and the backward integration, thus testing whether longer integrations could actually give improved results. The surface fluxes are assumed invariant in time, and the objective function is the one represented symbolically in Eq. (2). Notice that there are a number of other possible choices for formulating the steady penalty, corresponding to different weights; the one adopted here was chosen for its simplicity.

The experiments are performed with the box (6 × 6) GCM already described in section 2, with the same artificial data (Fig. 1). In all experiments described below, the wind field is assumed to be perfectly known; so the control variables are the initial conditions and the surface heat and freshwater fluxes. The results of two experiments will be presented here, which differ in the parameterization of convective overturning and in the modeling strategy. Experiment 1 uses a vertical diffusion coefficient dependent on the Richardson number, with a maximum value of 50 cm^2 s^{-1} in case of static instability. In experiment 2, this formulation is changed to the standard GFDL convection scheme

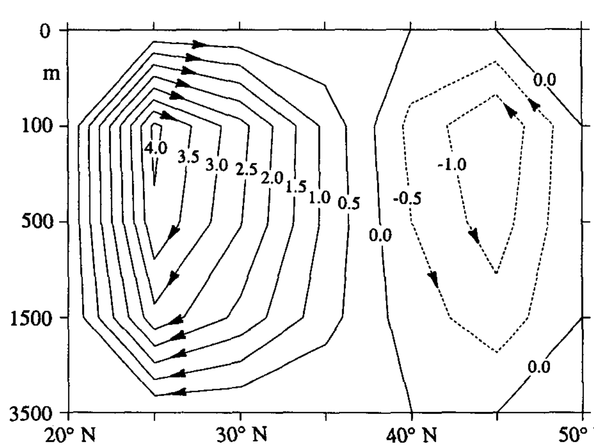

FIG. 4. Box GCM, experiment 1, meridional overturning streamfunction (Sv) as found by an almost purely geostrophic fit of the velocity field to the data displayed in Fig. 1, which serve as a first-guess initial state.

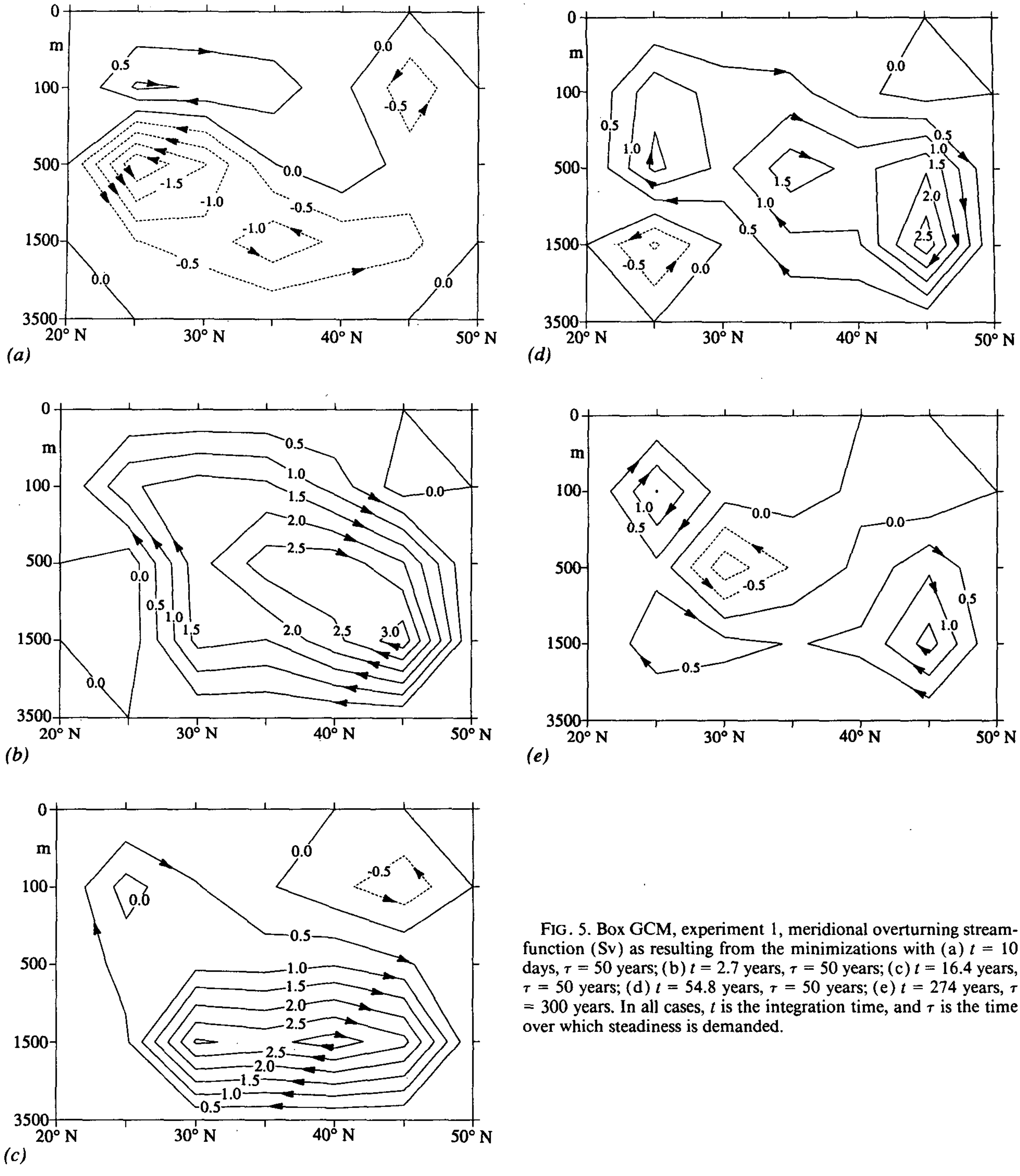

FIG. 5. Box GCM, experiment 1, meridional overturning streamfunction (Sv) as resulting from the minimizations with (a) $t = 10$ days, $\tau = 50$ years; (b) $t = 2.7$ years, $\tau = 50$ years; (c) $t = 16.4$ years, $\tau = 50$ years; (d) $t = 54.8$ years, $\tau = 50$ years; (e) $t = 274$ years, $\tau = 300$ years. In all cases, t is the integration time, and τ is the time over which steadiness is demanded.

that mixes neighboring boxes completely [but not the entire unstable part of the water column, e.g., Marotzke (1991)] in the case of static instability.

Experiments 1a and 1b (see Table 1) use the data as a first guess and demand steadiness over $\tau = 50$ years. The integration times are 10 days (1 time step) and 2.7 years in experiments 1a and 1b, respectively. The results are displayed primarily by showing the streamfunction of the meridional overturning, that is, the zonally integrated flow in the latitude–depth plane. Figure 4 shows the meridional overturning of the first-guess initial state, that is, it represents the zonally integrated geostrophic flow field derived from the data. We see a strong upwelling of 4 Sv (1 Sv $\equiv 10^6$ m^3 s^{-1}) at the equatorward boundary and a weaker one of 1 Sv at the poleward boundary, with downwelling in between. The one time-step minimization results in low-latitude sinking and rising motion everywhere else, except for the shallow, directly wind-driven circulation cells (Fig. 5a).

Extending each forward and backward integration over 2.7 years (experiment 1b) completely reverses the meridional circulation (Fig. 5b); there is now sinking at a rate of 3 Sv at high latitudes and rising everywhere else. Experiments 1c and 1d each take the solution of experiment 1b as the first guess and are run over 16.4 years and 54.8 years, respectively. Figures 5c,d show that we do not yet obtain a stable estimate; although the direction of the meridional overturning does not change, both the structure and the amplitude show clear differences. Using the result of experiment 1d as a first guess, experiment 1e integrates over 274 years and demands steadiness over 300 years. The result is a very sluggish thermohaline circulation (Fig. 5e) with no basinwide coherent pattern.

The most important lesson to learn from experiment 1 is that the result is very sensitive to the integration time. An objective measure of which of the five different minimizations gives the "best" result is given by the contributions to the final cost function. Figure 6 shows rms values of the temperature residuals, for each layer separately, and normalized with the variances entering the respective weights. If the integration is extended over only one time step (experiment 1a, Fig. 6a), the steady misfits are much larger than the data misfits,

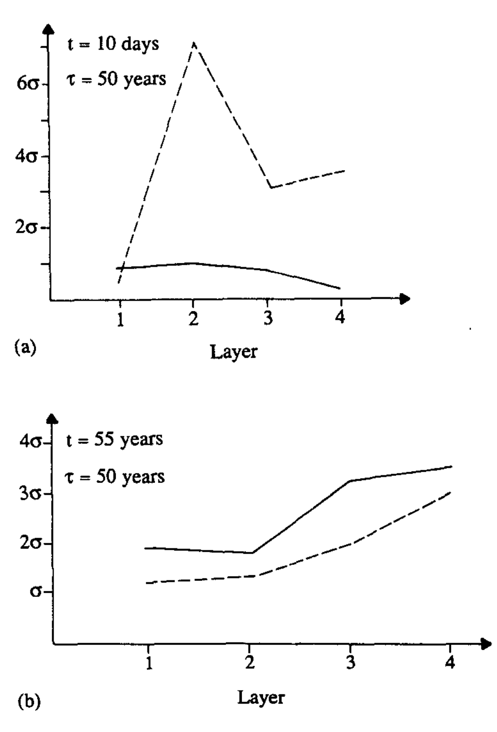

FIG. 6. Box GCM, rms values of the temperature residuals, for each layer separately and normalized with the a priori variances. Solid line: model data misfit, dashed line: steady penalty. (a) Experiment 1a, t = 10 days, τ = 50 years; (b) experiment 1d, t = 54.8 years, τ = 50 years.

except in the surface layer. In contrast, experiment 1d results in nearly equal data and steady misfits, for each layer individually (Fig. 6b). But one would have to reject the hypothesis that the data are compatible with a model changing significantly only over times as short as 50 years, because both the steadiness and misfit terms are too large; the rms values in the deepest layer, for example, are about three standard deviations. Experiment 1e (Fig. 5e) shows that an even longer integration, together with a more stringent demand on steadiness (t = 274 yr, τ = 300 yr), does not necessarily lead to a more acceptable solution. It appears that the deep-water formation processes are far too weak with the Richardson number parameterization of convective overturning. In the bottom layer, the rms temperature residual of the final estimate exceeds six standard deviations for both data and steady misfit terms (not shown).

The Richardson number convection parameterization seems to be inadequate for the fit of the model to the artificial data. Therefore, experiment 2 employs the standard GFDL convection scheme in case of static instability. Prior experience showed that convective adjustment makes progress more difficult, so the modeling strategy is modified (Table 1). Starting from the data as the first-guess initial state, the integration time t and the time τ entering the steady weights are increased to 2.7 years and 10 years, respectively (experiment 2a), then to, respectively, 54.8 years and 50 years (experiment 2b), then to 274 years and 300 years, respectively (experiment 2c). Experiments 2b and 2c take the results from 2a and 2b, respectively, as first-guess initial states.

Figure 7, again displaying the meridional overturning streamfunction, shows that the results differ drastically between experiments 2a and 2b, but that demanding steadiness over 50 years and 300 years, together with corresponding integration times, does not alter the solution much. In Fig. 8, the rms temperature residuals of experiment 2b are displayed, showing that a more detailed analysis would be required before one could conclude that model and data were incompatible. The employment of convective adjustment, demanding steadiness over 50 years (or more), and an integration time that matches τ yields the best (and roughly acceptable) model fit to the data represented in Fig. 1. Note that we can judge the success of the inversion only by the structure and magnitude of the residuals of the model fit to the data, an overall measure of which is the cost function (and its various parts). The meridional overturning shown in Fig. 7c most likely does not look very appealing, especially when compared to what conventional wisdom holds about the North Atlantic meridional circulation. The artificial data (Fig. 1) used here, however, chosen to represent a schematic of a generic subtropical gyre hydrography, have no information built into them to prevent the final solution from looking more like, for example, the North Pacific, with practically no overturning.

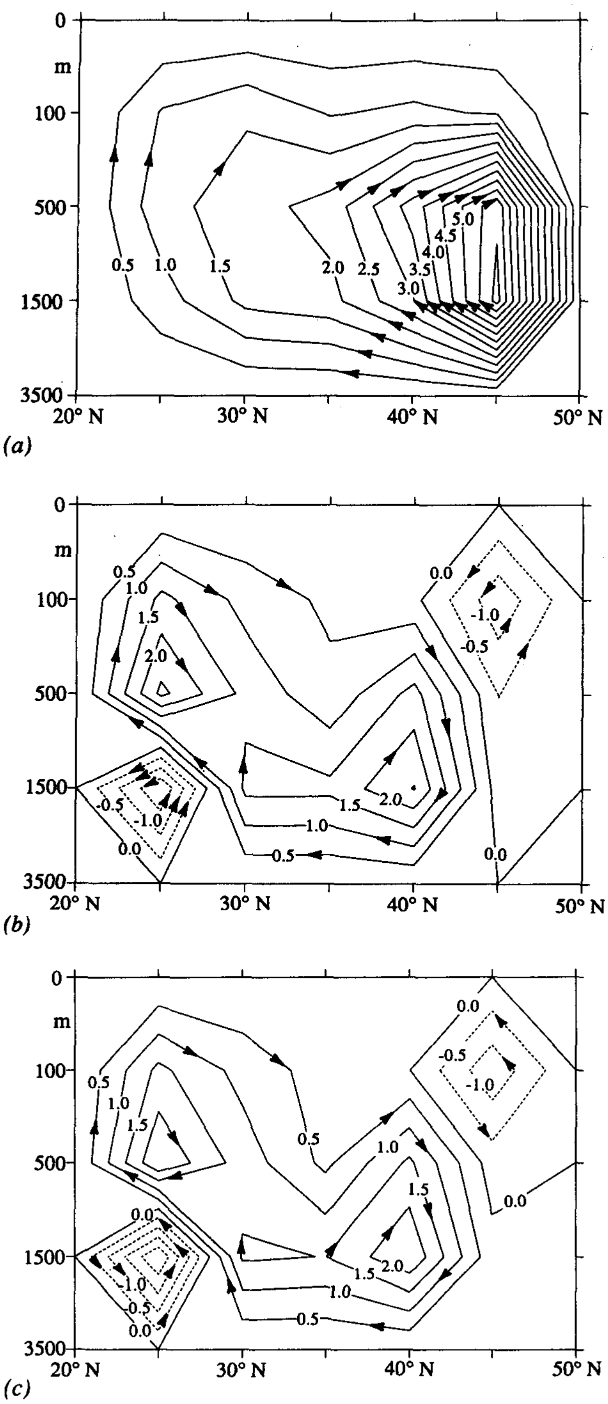

FIG. 7. Box GCM, experiment 2, meridional overturning streamfunction (Sv) as resulting from the minimizations with (a) $t = 2.7$ years, $\tau = 10$ years; (b) $t = 54.8$ years, $\tau = 50$ years; (c) $t = 274$ years, $\tau = 300$ years.

The minimization problem tackled in this section is a very difficult one, because the data were constructed using no dynamical principles at all. To pose an easier problem, some identical-twin experiments (results not shown here) are performed with the data taken from the steady state represented by $\mu = 1$ in Fig. 2 (see section 2). Convective adjustment is used, and the first-guess initial state is biased, compared to the "true" state, by specifying a $\mu < 1$. If $\mu = 0.9$, for example, this means that, compared to the true state, the bottom-layer initial-guess temperatures are biased by about 1 standard deviation of the assumed observation error. The time τ is set to 50 years in the experiments discussed here. Generally speaking, the results in the twin experiments are less sensitive to the length of the integration time. For a moderately biased first-guess initial state ($\mu = 0.8$ or 0.9), integration times of order 5 years and longer are successful in reproducing the correct state within error bars; however, integrations over one time step fail even in the experiments with a small bias. A large bias ($\mu = 0$, starting from the data as the first guess) does not lead to a compatible result, even for an integration time of 54.8 years.

The results from the identical-twin experiments are not described here in detail because, although they do give confidence in that an assimilation technique does in principle work (or not), they pose a misleadingly simple assimilation problem. In an inversion using real ocean data, we must face the possibility that the model physics are incompatible with the data, a case which is excluded in a (strictly) identical-twin experiment. Rather than adding noise to the data in a twin experiment, it was decided here to rely on artificial data that were not obtained through a forward model run.

5. Discussion

Integration time must roughly match the time scale of interest if one wants to obtain a meaningful minimum of objective functions as those considered here. Computations with the idealized box GCM and an artificial dataset (experiments 1 and 2) showed that results were sensitive to the length of the integration time. Specifically, in experiment 1, the one time-step approach yielded a meridional overturning that showed low-latitude sinking, in contrast to the longer runs that

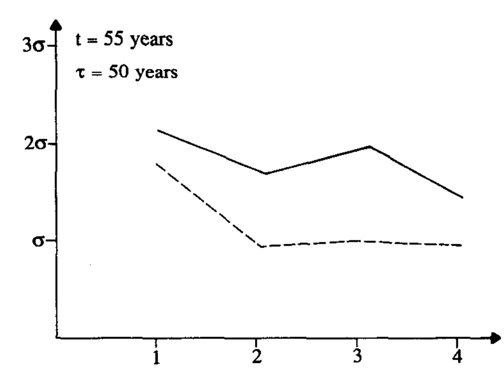

FIG. 8. Box GCM, rms values of the temperature residuals, for each layer separately and normalized with the a priori variances for experiment 2b, $t = 54.8$ years, $\tau = 50$ years. Solid line: model data misfit; dashed line: steady penalty.

formed deep water at high latitudes. A stable estimate was obtained if the integration was carried over 50 years and more (with a corresponding τ), provided an efficient convective adjustment scheme was used (experiment 2). Additionally, in the 50-year runs the residuals of data misfit and time drift, respectively, had the same order of magnitude.

In future attempts at performing a North Atlantic inversion with a GCM, the most important question is over how long we must integrate the model forward and backward, before we can confidently state whether the data are compatible with a steady-state assumption. We anticipate that, in a North Atlantic inversion, a time scale of interest will be that of the main thermocline since this is important for near-term climate problems. The time scale is thus set by either long internal Rossby waves (about 10 years), or by the ventilation of the thermocline (about 20 years), or both. Thus, we are forced to demand steadiness over at least 10 years, with a corresponding, or perhaps somewhat shorter, integration time.

The results presented here give hope that 5–10 years of integration may be sufficient, an estimate that is supported by the impression that the threshold of about 50 years found in experiments 1 and 2 represents an upper bound due to the severe nature of the minimization problem posed, and because the identical-twin experiments proved successful for O(5 yr) integration time. Substantially more than 10 years of integration would make the GCM inversion computationally very hard to perform; from preliminary tests with the North Atlantic model it can be estimated that one typical experiment would consume O(100) hours CPU on a CRAY-2, if the integration is carried out over 10 model years.

It may also be necessary to consider more carefully how information propagates in the model. How do the surface fluxes interact with, say, temperature and salinity in the lower thermocline? Is it possible to convey the information in less than the typical adjustment time, for example, the travel time of baroclinic Rossby waves or the ventilation time scale? This result is conceivable since a partial adjustment is performed several times at each conjugate-gradient iteration. Conversely, how does the information about a model data misfit reach the surface, in the adjoint model, to produce a better estimate of the boundary conditions? Answers to these questions must be found before one can hope to find the computationally least expensive way to perform the GCM inversion.

In designing a modeling strategy one is also faced with the question of how meaningful the concept of a strictly steady-state ocean is in the presence of real data. Quite apart from the fact that it is not clear how the equations describing spatial and temporal averages should be derived from the instantaneous conservation equations, the ocean almost certainly undergoes drifts on time scales of centuries or longer. It may not only be impossible but also inappropriate to try a GCM inversion while demanding steadiness over more than a few decades.

Note added in proof. Partly as a consequence of the results described here, Tziperman et al. (1992b) have revised some conclusions of Tziperman et al. (1990).

Acknowledgments. I am indebted to Eli Tziperman who made his inverse model code available to me, and to Bob Long, S. Hwang, and Carlisle Thacker for permitting the use of their adjoint model. Many colleagues helped to clarify the points discussed here, especially I wish to thank Carl Wunsch, Jürgen Willebrand, Jens Schröter, Dirk Olbers, Andreas Schiller, Detlef Stammer, Carlisle Thacker, and Eli Tziperman. Thanks go to Bud Brown for drawing the figures and Dorothy Frank for typing parts of the manuscript. This work was supported by Grant OCE-8823043 from the National Science Foundation.

REFERENCES

Bryan, K., 1969: A numerical method for the study of the circulation of the World Ocean. *J. Comput. Phys.*, **3**, 347–376.

——, 1984: Accelerating the convergence to equilibrium of ocean-climate models. *J. Phys. Oceanogr.*, **14**, 666–673.

Cox, M. D., 1984: A primitive equation, three-dimensional model of the ocean. GFDL Ocean Group Tech. Rep. No. 1, GFDL/Princeton University.

Fukumori, I., and C. Wunsch, 1991: Efficient representation of the North Atlantic hydrographic and chemical distributions. *Progress in Oceanography,* vol. 27, Pergamon, 111–195.

——, F. Martel, and C. Wunsch, 1991: The hydrography of the North Atlantic in the early 1980s. An atlas. *Progress in Oceanography,* vol. 27, Pergamon, 1–110.

Isemer, H. J., J. Willebrand, and L. Hasse, 1989: Fine adjustment of large scale air-sea energy flux parameterization by a direct estimate of ocean heat transport. *J. Climate,* **2**, 1173–1184.

Levitus, S., 1982: Climatological atlas of the World Ocean. NOAA Tech. Pap. 3, 173 pp.

Marotzke, J., 1991: Influence of convective adjustment on the stability of the thermohaline circulation. *J. Phys. Oceanogr.*, **21**, 903–907.

——, and J. Willebrand, 1991: Multiple equilibria of the global thermohaline circulation. *J. Phys. Oceanogr.*, **21**, 1372–1385.

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

Talagrand, O., and P. Courtier, 1987: Variational assimilation of meteorological observations with the adjoint vorticity equation. I. Theory. *Quart. J. Roy. Meteor. Soc.*, **113**, 1311–1328.

Thacker, W. C., and R. B. Long, 1988: Fitting dynamics to data. *J. Geophys. Res.*, **93**, 1227–1240.

Trenberth, K. E., J. G. Olson, and W. G. Large, 1989: A global ocean wind stress climatology based on ECMWF analyses. NCAR Tech. Note, NCAR, 93 pp.

Tziperman, E., and W. C. Thacker, 1989: An optimal control/adjoint equations approach to studying the oceanic general circulation. *J. Phys. Oceanogr.*, **19**, 1471–1485.

——, W. C. Thacker, R. B. Long, S. Hwang, and S. Rintoul, 1990: A North Atlantic inverse model using an oceanic general circulation model. Unpublished manuscript.

——, ——, ——, and ——, 1992a: Oceanic data analysis using a general circulation model. Part I: Simulations. *J. Phys. Oceanogr.*, **22**,

——, ——, ——, ——, and S. Rintoul, 1992b: Oceanic data analysis using an oceanic general circulation model. Part II: A North Atlantic model. *J. Phys. Oceanogr.*, **22**, 1458–1485.

Wunsch, C., 1988: Transient tracers as a problem in control theory. *J. Geophys. Res.*, **93**, 8099–8110.