Journal of Marine Research, 53, 433–451, 1995

A technique for the determination of surface heat and freshwater fluxes from hydrographic observations, using an approximate adjoint ocean circulation model

by Andreas Schiller[1,2] and Jürgen Willebrand[1]

ABSTRACT

A technique to determine the large-scale time-averaged ocean circulation from hydrographic observations and surface flux estimates is described. It is based on an inversion of the Bryan-Cox ocean general circulation model. We have constructed an approximate adjoint to that model which is computationally simpler and more economic than the exact adjoint. The optimization algorithm, although not optimal in a statistical sense, allows calculation of all state variables such that they are consistent with the equilibrium dynamics of the circulation model and agree as closely as possible with the observed data.

To verify the technique, we have performed identical twin experiments with the circulation model in an idealized geometry. It is found that in principle the true model state including surface fluxes can be recovered with acceptable accuracy, even if no information on the surface fluxes is available. Under ideal conditions, the resulting rms errors of the surface fluxes were as low as 3 W/m^2 and 0.2 m/year for heat and freshwater, respectively. Regions of deep-water formation due to convection show however larger errors on a small spatial scale, depending on the nonlinear, threshold-like nature of convective adjustment. The optimized solutions are distinctly sensitive to the integration time interval, with optimal values around three years.

The results suggest that the procedure is suitable to obtain a consistent description of the oceanic state, and in particular more accurate estimates of the air-sea heat and freshwater fluxes.

1. Introduction

The distributions of temperature and salinity in the ocean have long been a primary source of information for the large-scale ocean general circulation. The classical water mass analysis which is based only on the conservation of heat and salt has been successful for qualitative inferences. Attempts to determine circulation parameters more quantitatively have additionally involved some aspects of the dynamics, as e.g. geostrophy or an approximation to the vorticity balance (e.g. Stommel and Schott, 1977; Wunsch, 1978; Olbers *et al.,* 1985; Martel and Wunsch, 1993; Bogden *et al.,* 1993).

1. Institut für Meereskunde, Düsternbrooker Weg 20, 24105 Kiel, Germany.
2. Present address: Max-Planck-Institut für Meteorologie, Bundesstraße 55, 20146 Hamburg, Germany.

A complete dynamical description can only be obtained from three-dimensional numerical models which have become increasingly important tools to understand and interpret the ocean circulation and water mass distribution. However, prognostic models still have considerable inaccuracies, which besides insufficient resolution are mainly due to uncertainties in surface forcing and in subgrid-scale parameterizations. For that reason, models need to be systematically validated by ocean observations, and in particular hydrographic observations.

The concept of inverse modelling, or data assimilation, allows us to combine both model dynamics and observations in a systematic fashion. The main objectives which have been described (e.g., by Willebrand and Wunsch (1990)) are: (i) to obtain estimates of oceanic state variables, in particular of those parameters that are not easily and directly observable, (e.g. heat and freshwater transports and water mass formation rates, including an estimation of the error of all fields) and (ii) to establish consistency of the observed data within their error bounds with the model dynamics.

Recently, growing computer facilities have allowed the introduction of extensive assimilation schemes in oceanography employing full three-dimensional general circulation model (GCM) dynamics. The technique based on the adjoint equations which is described (e.g., by Le Dimet and Talagrand (1986) or Thacker and Long (1988)) has been particularly attractive for systems with a large number of degrees of freedom. This method which is inspired from optimal control techniques is based on standard least-squares procedure which minimizes the difference between model state and observations while satisfying the model dynamics exactly. The adjoint method has been used for data assimilation into primitive equation models with simplified momentum balance to determine the mean hydrography and circulation of the North Atlantic (Tziperman *et al.,* 1992a,b; Marotzke and Wunsch, 1993). A full primitive equation model and its adjoint have been applied to estimate the seasonal circulation of the Mediterranean (Bergamasco *et al.,* 1993).

Construction and operation of the full adjoint of a primitive equation GCM is extremely demanding in terms of computing and manpower, and was not feasible for us with the available resources. The simplified approach presented in this paper is based on the following considerations. Since the forward model describes the dynamical connections among oceanic fields and boundary fluxes, any approximations here potentially restrict the validity and credibility of the results. On the other hand, the adjoint (backward) model serves only to provide information about how the control variables have to be changed in order to reduce the model-data differences. Any approximations here may reduce efficiency of the minimization procedure or prevent finding the absolute minimum. However, the quality of the optimized state can always be tested independently, and an acceptable state is guaranteed not to violate any dynamical laws.

Consequently, we have constructed a simplified adjoint to the full Bryan-Cox forward circulation model. The momentum and vorticity equations in that model are

algebraically rather complex and cause the most difficulties for the adjoint model. As velocities are very nearly in geostrophic equilibrium, we have decided to limit the adjoint formulation to the heat and salinity conservation equations. The resulting approximate adjoint model is technically much less complex and computationally much more economic than the full adjoint. The obvious disadvantage is that this formulation does not permit use of direct current observations, and also prevents optimization of the wind field.

In this paper we describe the optimization procedure and report results from identical twin experiments to demonstrate that the technique is indeed suitable to calculate the thermohaline fluxes. In a companion paper, the method is applied to the North Atlantic circulation (Schiller, 1995).

2. Model description and approximate adjoint equations

The adjoint method will be presented in the simplified form which was used to solve the oceanographic inverse problem. The deviation of the model fields of temperature and salinity, T, S as well as heat- and freshwater flux H_T, H_S from the corresponding observations (index obs) is measured by a cost function J which subsequently will be minimized. Assuming Gaussian statistics, the cost function thus reads:

$$J = \frac{1}{2}\int\int \frac{(T^{\text{obs}} - T)^2}{\sigma_T^2}\,d\mathbf{x}dt + \frac{1}{2}\int\int \frac{(S^{\text{obs}} - S)^2}{\sigma_S^2}\,d\mathbf{x}dt + \frac{1}{2}\int\int \frac{(H_T^{\text{obs}} - H_T)^2}{\sigma_{H_T}^2}\,d\mathbf{a}dt + \frac{1}{2}\int\int \frac{(H_S^{\text{obs}} - H_S)^2}{\sigma_{H_S}^2}\,d\mathbf{a}dt + \frac{1}{2}\int \frac{(T^f - T^0)^2}{\sigma_{\Delta T}^2}\,d\mathbf{x} + \frac{1}{2}\int \frac{(S^f - S^0)}{\sigma_{\Delta S}^2}\,d\mathbf{x}. \tag{1}$$

The contributions to the cost function are interpreted as follows. The first two terms denote the model-data differences for temperature and salinity fields which are normalized by the respective standard deviations σ_T, σ_S of the observed fields, and integrated over time and over the model volume. A rigorous procedure would also involve covariances of the observed fields which are however difficult to estimate and would increase the computational burden and are therefore ignored.

The next two terms in (1) arise from the difference between model and observed surface heat and freshwater fluxes, again normalized by the errors σ_{H_T}, σ_{H_S} of the observed fluxes and integrated over time and over the surface area.

The last two terms measure the temporal changes of temperature and salinity during an integration, penalizing the difference between the final model state T^f, S^f and the initial state T^0, S^0 (Tziperman and Thacker, 1989). The necessity for these terms, and the choice of the weight factors $\sigma_{\Delta T}$ and $\sigma_{\Delta S}$ needs some explanation.

We are interested in the time-averaged circulation state so that in principle all variables should be independent of time. In practice, the model steady state is

obtained by integrating the time-dependent Bryan-Cox model forward in time until a steady state is reached. The baroclinic adjustment processes in the ocean take place on a timescale of years to decades, and the advective-diffusive adjustment of the thermohaline circulation requires several centuries. Computational requirements prevent us from computing the model's steady state for every iteration, and restrict the integration to a finite interval Δt. To overcome this problem, the additional constraints are added to the cost function that measure the quadratic difference of the temperature and salinity fields between the end and beginning of the integration period. Large differences force the next forward integration via the new estimates T^0, S^0 to be closer to a steady state.

The best choice for the interval Δt is not obvious. Tziperman and Thacker (1989) performed experiments with a quasigeostrophic barotropic model by an integration of the model one time step forward and the adjoint model one time step backward. However, as described in Marotzke (1992), with a primitive equation model it is necessary to carry out the integration over a much longer time span than just one time step. We have performed most experiments with a value of $\Delta t = 610$ days. The sensitivity to this choice will be discussed in Section 3 below.

The quantities $\sigma_{\Delta T}$, $\sigma_{\Delta S}$ represent the relative strength of the steady state constraint. We have chosen

$$\sigma_{\Delta T}^2 = \sigma_T^2 \tag{2}$$

and likewise for $\sigma_{\Delta S}$. This empirical choice means that a temporal temperature change of magnitude σ_T over the integration time interval Δt has the same weight in (1) as one standard deviation between model and observation.

We will now discuss the model equations and their simplifications in our adjoint model as well as the consequences for the applied minimization procedure. The model code is based on the primitive equations, including conservation of momentum, mass, heat and salt as described by Bryan (1969) and Cox (1984). The prognostic equations for this model are given by

$$E_{\mathbf{u}}(T, S, \mathbf{u}, \mathbf{x}, t) = \frac{\partial \mathbf{u}}{\partial t} + \mathbf{u} \cdot \nabla \mathbf{u} + w \frac{\partial \mathbf{u}}{\partial z} + f\mathbf{k} \times \mathbf{u} + \frac{1}{\rho_0} \nabla p - A_v \frac{\partial^2 \mathbf{u}}{\partial z^2} - \nabla \cdot (A_h \cdot \nabla \mathbf{u}) = 0 \tag{3}$$

$$E_T(T, S, \mathbf{u}, \mathbf{x}, t) = \frac{\partial T}{\partial t} + \mathbf{u} \cdot \nabla T + w \frac{\partial T}{\partial z} - \nabla \cdot (K_h \cdot \nabla T) - \frac{\partial}{\partial z}\left(K_v \frac{\partial T}{\partial z}\right) = 0 \tag{4a}$$

$$E_S(T, S, \mathbf{u}, \mathbf{x}, t) = \frac{\partial S}{\partial t} + \mathbf{u} \cdot \nabla S + w \frac{\partial S}{\partial z} - \nabla \cdot (K_h \cdot \nabla S) - \frac{\partial}{\partial z}\left(K_v \frac{\partial S}{\partial z}\right) = 0 \tag{4b}$$

The notations are standard and can be found elsewhere (e.g. Cox, 1984). The diagnostic equations are given by

$$\frac{\partial p}{\partial z} = -\rho g \tag{5}$$

$$\nabla \cdot \mathbf{u} = -\frac{\partial w}{\partial z} \tag{6}$$

$$\rho = \rho(S, T, p). \tag{7}$$

The momentum equation (3) is actually decomposed into a barotropic and baroclinic part, and the stream function of the vertically integrated volume transport is used as a prognostic variable. This aspect is however not relevant for our adjoint model and hence not described in detail.

The equations 3,4a and 4b are solved by forward time stepping with finite spatial differences. In the vertical dimension an implicit scheme for convective adjustment processes is used (Richtmyer and Morton, 1967; Cox, 1987), that allows K_v to take on a large value in case of static instability.

The fluxes of heat (H_T) and freshwater (H_S) (positive into the ocean) enter through the boundary conditions

$$\rho_0 c_p K_v T_z = H_T \tag{8a}$$

and

$$K_v S_z = H_S S_{\text{ref}}. \tag{8b}$$

Here c_p is the specific heat. The freshwater flux in Eq. (8b) is approximately translated into an equivalent salinity flux, with S_{ref} representing a fixed reference salinity. Note that this approach is different from the commonly used technique of Newtonian restoring boundary conditions.

Insulating boundary conditions on T and S are specified for the bottom and lateral walls.

The cost function has to be minimized, while simultaneously satisfying all dynamical constraints of the model. The resulting functional is the Lagrange function L with the model dynamics as "strong constraints":

$$\begin{aligned} L(T, S, \mathbf{u}, H_T, H_S, \lambda_T, \lambda_S, \lambda_{\mathbf{u}}) = J &+ \int\int \lambda_T(\mathbf{x}, t) \cdot E_T(T, S, \mathbf{u}, \mathbf{x}, t) d\mathbf{x}dt \\ &+ \int\int \lambda_S(\mathbf{x}, t) \cdot E_S(T, S, \mathbf{u}, \mathbf{x}, t) d\mathbf{x}dt + \int\int \lambda_{\mathbf{u}}(\mathbf{x}, t) \cdot E_{\mathbf{u}}(T, S, \mathbf{u}, \mathbf{x}, t) d\mathbf{x}dt. \end{aligned} \tag{9}$$

The purpose of introducing the Lagrange multipliers λ_T, λ_S, and $\lambda_{\mathbf{u}}$ in Eq. (9) is to transform the problem from a constrained minimization of the cost function J to an unconstrained minimization using only unconstrained derivatives as if all variables

were independent. Variation of the Lagrange function with respect to, e.g., temperature yields:

$$\delta_T L = \int\int d\mathbf{x}dt\left[\lambda_T \frac{\delta E_T}{\delta T} + \lambda_S \frac{\delta E_S}{\delta T} + \lambda_{\mathbf{u}} \frac{\delta E_{\mathbf{u}}}{\delta T} - \frac{(T^{\text{obs}} - T^f)}{\sigma_T^2}\right]\delta T = 0. \qquad (10)$$

(I) (II) (III) (IV)

Term (IV) denotes the cost function contribution. The only assimilation of observations is done at the end of the forward integration with an ocean assumed closer to an observed steady state than at the initial time.

In accordance with the motivation given in the introduction, we neglected the terms (II) and (III) in Eq. (10). Term (II) results from the dependence of the salinity budget from temperature (through a stability-dependent mixing or convection). This information is ignored for the adjoint model (but still included in the forward model). Term (III) results from the temperature-dependence of velocity (via hydrostatic pressure gradient), its influence on the adjoint is likewise ignored. Attempts to include this influence by approximating (III) with a geostrophic momentum balance and a linearized density equation did not result in any improvement, compared to neglecting (III) completely.

With this approximation, and after some integration by parts, from Eq. (10) one obtains the adjoint equation

$$-\frac{\partial \lambda_T}{\partial t} - \mathbf{u}\cdot\nabla\lambda_T - w\frac{\partial \lambda_T}{\partial z} - \nabla\cdot(K_h\cdot\nabla\lambda_T) - \frac{\partial}{\partial z}\left(K_v\frac{\partial \lambda_T}{\partial z}\right) - \frac{\partial}{\partial z}\left(\frac{\partial K_v}{\partial\left(\frac{\partial T}{\partial z}\right)}\frac{\partial \lambda_T}{\partial z}\frac{\partial T}{\partial z}\right) = -\frac{(T^{\text{obs}} - T^f)}{\sigma_T^2}. \qquad (11)$$

(*)

As a further approximation, we have neglected the term (*) in Eq. (11) which again results from stability-dependent mixing. Inclusion of this term would cause severe numerical problems in the adjoint equations in the case of static instability when the factor $\partial K_v/\partial T_z$ becomes very large. Ideally, this should give no contribution to the gradients because that large number is multiplied by the vertical temperature gradient which vanishes in case of vertical convection. However, numerical inaccuracies cause a nonvanishing contribution of the convective adjustment to the respective gradients. As a consequence, the gradients would lose their correct information content about the model-data misfits.

The main advantage of these simplifications is that Eq. (11) (and likewise for λ_S) is no longer dependent on $\lambda_{\mathbf{u}}$. Hence it is no longer necessary to solve the adjoint equations for the velocity unless one wants to include direct velocity observations or to optimize the wind forcing. A further advantage is a considerable reduction of the

computational cost and model code complexity. Incidentally, Eq. (11) is formally identical with the corresponding forward equations (except for the forcing term and the sign of the first order terms), and thus needs only minor coding efforts.

On the other hand, as a consequence of the approximations the Lagrange multipliers are only approximately correct. Hence we expect that the approximations may have adverse effects on the optimization algorithm described below.

The Lagrange multipliers are calculated by integrating Eq. (11) backwards in time. Physically, the Lagrange multipliers $\lambda_{T,S}$ are a measure for the sensitivity of the cost function to variations of the model variables. The right-hand side of Eq. (11) is the gradient of the cost function with respect to the state variables (T, S) and represents the forcing of the adjoint equation. The gradients of the Lagrange function with respect to the initial conditions (T^0, S^0) determine the direction in state space along which the control variables have to be changed to find a smaller value for the cost function. In accordance with the thermohaline boundary conditions of the forward model, there is neither advective nor diffusive transport of the Lagrange multipliers across the bottom or the lateral walls.

The gradients of the cost function with respect to the thermohaline surface fluxes are obtained through variation of the Lagrange function (Eq. 9), e.g.:

$$\frac{\delta L}{\delta H_T} = -\frac{(H_T^{\text{obs}} - H_T)}{\sigma_{H_T}^2} - \int \lambda_T \cdot \frac{1}{\rho_0 \cdot c_p \cdot z_1}\, dt, \tag{12}$$

with z_1 denoting the depth of the uppermost layer in the model.

To summarize, the procedure contains the following steps (Fig. 1): The model is integrated forward in time starting from an initial guess for the control variables (T^0, S^0, H_T, H_S). The initial velocity field for the first forward integration is taken from a steady state solution of the prognostic model. At the end of the integration time interval which will be specified later, the resulting final values T^f and S^f are used to compute the cost function. If no minimum of the cost function is found at a current iteration, the model-data misfits in the adjoint model are integrated backward in time determining the gradients and thus new estimates of the control variables. The gradients and the actual value of the cost function are passed to a mathematical descent algorithm to compute new estimates for T^0, S^0, H_T and H_S. With these new estimates another forward integration of the prognostic model is performed. This procedure is continued as long as the cost function continues to decrease.

The velocity field (including baroclinic and barotropic component), which will not be optimized in our approach is handled in the following way: In the forward model the first guess of the velocity field is taken from a steady state solution of a prognostic model integration. For subsequent iterations the velocity of the last time step t^f from the previous iteration serves as initial condition for the next forward run. Strictly speaking, the estimates for the velocity at the beginning of every iteration are not

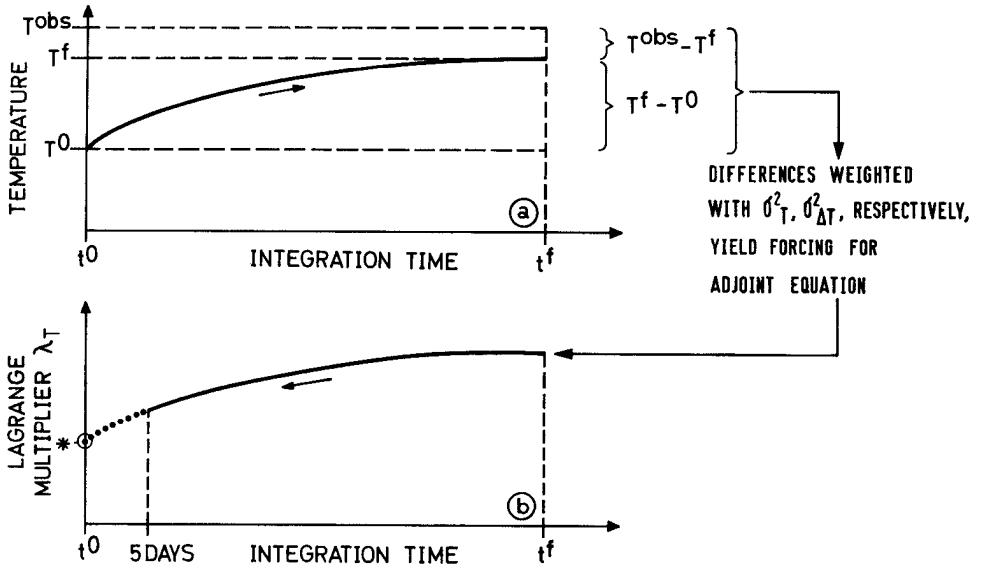

Figure 1. Schematic plots of temporal evolution of an arbitrary temperature point of the model: (a) prognostic model, (b) adjoint model. Notation: t^0: initial time, t^f: final time, T^{obs}: observed temperature, T^0: initial temperature, T^f: final temperature. Curve (a) describes the forward run starting from T^0 (from initial guess for first iteration, from new estimate for subsequent iterations). $T^f - T^{obs}$ denotes the model-data misfit computed at the end of the forward integration, $T^f - T^0$ denotes the difference between the final and initial temperature. Curve (b) describes the computation of the Lagrange multiplier λ_T. The velocity is kept fixed the last five days of backward integration (dotted part of curve (b)). At the end of the backward integration (*) the gradient is computed. See text for further explanations.

determined correctly. However, the velocity field adjusts geostrophically on a time scale of typically 3–4 days to a modified density field (Anderson *et al.*, 1979), so that incorrect estimates for the initial velocity become unimportant on longer time scales. The model ocean can be regarded as being always in a near-geostrophic equilibrium, except for the first few days of forward integration. The geostrophic adaptation might negatively influence the advective terms of the adjoint model during the last time steps of backward integration. For this reason, the velocity field is kept fixed during the last five days of backward integration (dotted λ_T in Fig. 1b).

3. Identical twin experiments

We have performed identical twin experiments where data extracted from a spin-up of the model serve as simulated observations. These artificially created observations are subsequently assimilated into the identical model. Identical twin experiments serve as a tool for exploring the performance of an inverse model, because observations and model dynamics are compatible with each other. For that

reason, identical twin experiments represent the best results one can possibly expect when using real data instead of simulated observations.

The actual model state of an OGCM with given values of the parameters is determined from the initial conditions and the boundary conditions, in particular, the fluxes of heat and freshwater which are among the least well known variables in models. In contrast, the three-dimensional distributions of temperature and salinity are known with comparably high accuracy. This suggests investigation of the capability of the inverse procedure to recover the thermohaline surface fluxes with the observations of temperature and salinity assumed to be almost perfect. The hardest test is to assume that there are no surface flux observations available. Technically, this strategy is reflected in the cost function (Eq. 1) by suppressing any model-data misfits for the surface fluxes (equivalent to assuming a very large error for the observed fluxes).

We have used a coarse resolution (3.75° × 4°) basin-scale model without topography. The model geometry is a sector 60° in longitude, the latitudinal range extending from the equator to 64N. There are 15 levels in the vertical with layer thicknesses ranging from 50 m near the surface to 500 m near the bottom. The parameters for mixing and diffusion are (in m^2/s): $A_h = 2.5 \cdot 10^5, A_v = 1.0 \cdot 10^{-4}, K_h = 1.0 \cdot 10^3, K_v = 0.5 \cdot 10^{-4}$. Model resolution, parameters for mixing and diffusion as well as the spin-up state have been chosen in accordance with a model configuration used by Marotzke (1990). The integration time was set to 610 days; the reason for this choice will be discussed later. The time steps are two hours for the momentum and vorticity balances and five days for the temperature and salinity equations. This asynchronous integration (Bryan, 1984) significantly reduces the amount of computing time and produces correct steady state solutions. Although we do not integrate our model to steady state this procedure gives reasonable results within the estimated steady state errors σ_T, σ_S in the cost function.

The control variables are the surface fluxes of heat and freshwater, as well as the initial conditions for temperature and salinity. The initial guesses and the observations for the hydrography have been chosen as the true values from the model's steady state, while the initial guesses for the surface fluxes were set to zero (i.e. nothing is known about the surface fluxes). The observation and steadiness errors have been set to small values (≈ 0.02°C and 0.005 psu at the surface with smaller values below) compared to observation errors computed by Levitus (1982). The low errors in temperature in salinity documentate the confidence one has in the hydrographic data. The wind stress field was the same as in the spin-up model run.

Solving for the unknowns (16 × 16 horizontal × 15 vertical grid points corresponding to about 4350 unknowns in T^0, S^0, H_T, H_S) required 57 iterations. Figure 2a displays the descent of the cost function as well as the different contributions to the cost function as a function of the iteration number. Although the hydrography has been started from the true values, after the first forward run the zero fluxes have

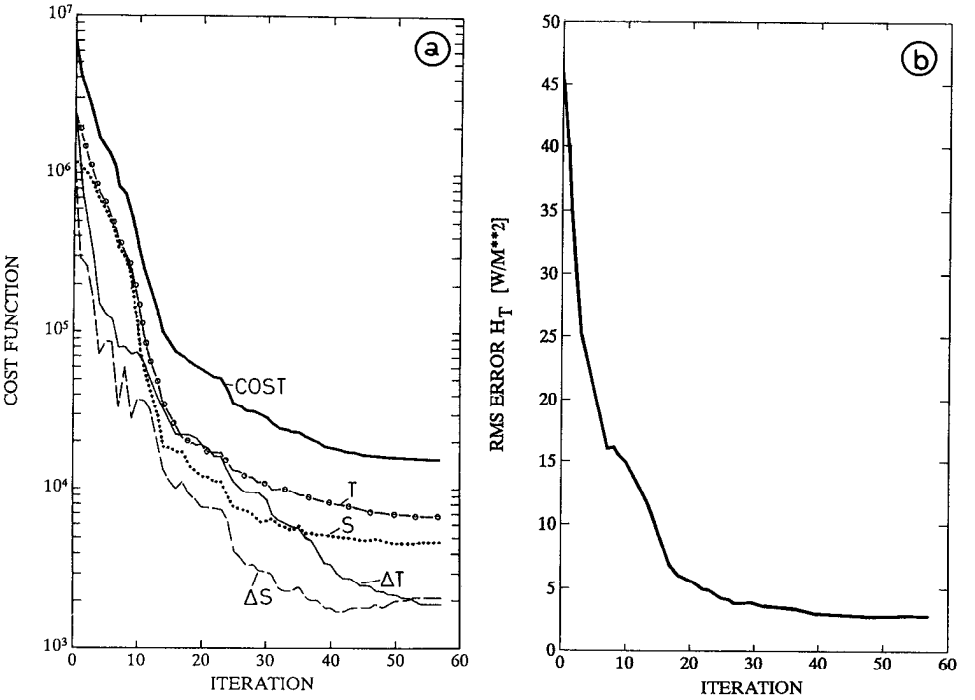

Figure 2. (a) Cost function and different cost function contributions for experiment without surface flux observations. Values are plotted vs. the iteration number. T: model-data misfit of temperature, S: model-data misfit of salinity, ΔT: steadiness misfit of temperature, ΔS: steadiness misfit of salinity. (b) rms-difference of the heat flux residual (W/m^2) plotted as function of the iteration number.

modified significantly the temperature and salinity fields. Thus, in the following iterations the hydrography had to be corrected together with the surface fluxes. All cost function contributions of T, S, ΔT, ΔS normalized by their errors are below one standard deviation, revealing that a consistent solution within the *a priori* error estimates has been found.

After the first 20 iterations the cost function has been reduced to one percent of its initial value. Another 37 iterations are performed, reducing the cost function further by nearly one order of magnitude. The residuals of the solution are due to the misfits of the hydrographic data and the steadiness-penalties, having the same order of magnitude (Fig. 2a). The residuals are influenced indirectly by the unknown surface fluxes, because only a reduced amount of information is available for the inverse model. However, even the rms-values of the surface fluxes show significant reductions. Figure 2b and Figures 3a,b demonstrate the way the errors in the surface heat fluxes are reduced (similar for freshwater fluxes). The strongest reduction in the rms-errors takes place within the first few iterations (Fig. 2b), increasing the number

of iterations reduces the errors only marginally. The surface fluxes are recovered such that the large-scale structure is rebuilt at first, while small-scale features become visible in the later iterations. This process can be understood with Figures 3a,b: after 3 iterations the regions with strong heat release in the area of the western boundary current and the northwestern part of the model domain are already visible. Heat uptake by the ocean can be seen in the central and southern part of the model area. A similar behavior is found for the freshwater flux. After 57 iterations the cost function cannot be reduced further by the optimization procedure and the area-averaged rms-errors for the heat and freshwater fluxes are 2.8 W/m^2 and 0.14 m/year, respectively.

Why are the errors not exactly zero? Possible reasons might be the approximate adjoint approach, which can result in systematic errors of the optimized parameters, and, additionally, the nonlinear convection parameterization. To examine this problem in more detail, Figure 4 shows the differences between the true and optimized surface heat fluxes. The largest deviations occur in the northwestern part of the model area with maximum local values of about ± 20 W/m^2, while the remaining areas show deviations smaller than 2 W/m^2. The northwestern area is dominated by permanent vertical convective activity that reaches the lowest model layers. The nonlinear formulation of the convection scheme used here suggests significant differences between the true and optimized heat fluxes (likewise for freshwater fluxes, which are not shown). Even regional averaging does not completely smooth out the differences with alternating signs. As a consequence of the surface flux deviations from their true values the hydrography also shows the strongest deviations in the northwestern region (Fig. 5). In all other areas the surface flux errors are relatively small and no fast adjustment process like convection exists to transfer the surface flux information into greater depths within the integration time of 610 days. Like the model-data misfits for the hydrography, the same holds for the steadiness misfits of the model. In Figure 6 the temporal differences for temperature in a depth of 1875 m are displayed. The errors of the surface fluxes are transported through vertical mixing to great depths and are detectable even in the time rate of change for the hydrography.

The choice of the integration time needed to obtain optimal results is not obvious. Tziperman *et al.* (1989) performed identical twin experiments, using a barotropic quasigeostrophic model, and the correct model state was found by carrying out only one forward time step of the model and one backward time step of the adjoint model. When we adopted their approach for our primitive equation model, the result was a model state that was far away from being steady. The contributions to the cost function by the steady penalties dominated over the model-data misfits. Especially along the western boundary current and the areas of strong vertical convection unacceptably large deviations from the steadiness demand appeared. The minimization could not find a compromise between the steady penalties and the observation

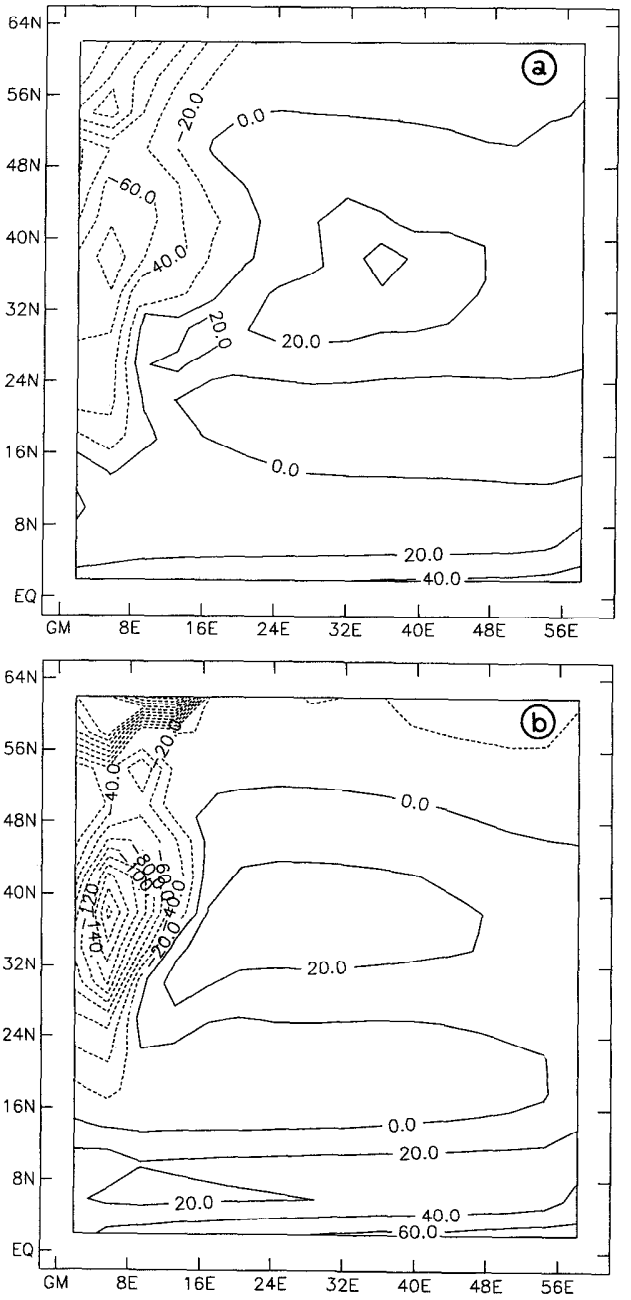

Figure 3. Optimized heat flux (W/m^2). Solution after 3 iterations (3a) and after 57 iterations (3b) (end of optimization).

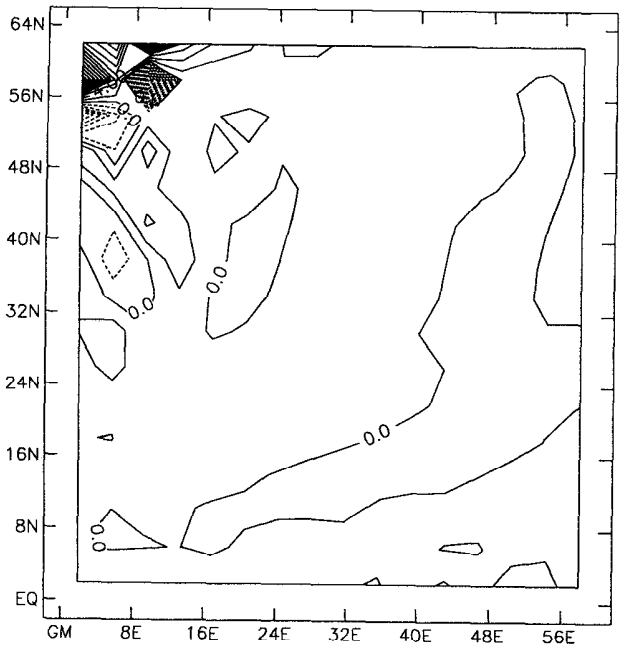

Figure 4. Difference between true and optimized surface heat flux. C.I. = 2.0 (W/m^2). Positive values denote a true heat flux that is larger than the optimized heat flux. Min./Max.: -20.0, $+22.0$ W/m^2.

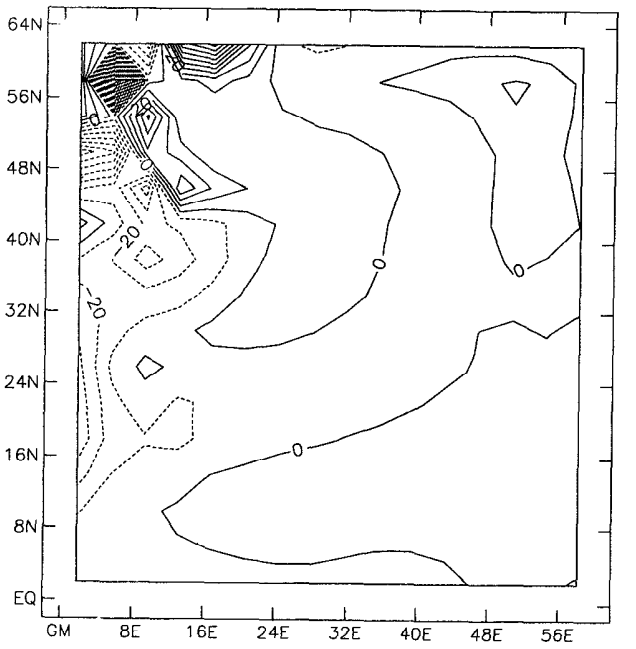

Figure 5. Difference between true and optimized salinity at 425 m depth. C.I. = 0.001 psu. Labels are scaled by 10^4. Min./Max.: $-0.018/+0.009$ psu.

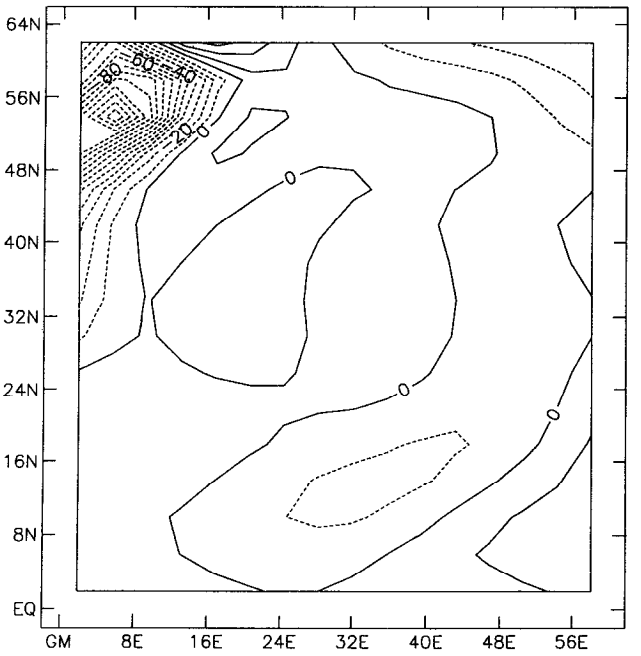

Figure 6. Temperature difference $T^f - T^0$ of the optimized model state at 1875 m depth. C.I. = 0.001°C. Labels are scaled by 10^4. Min./Max.: −0.014/+0.003°C.

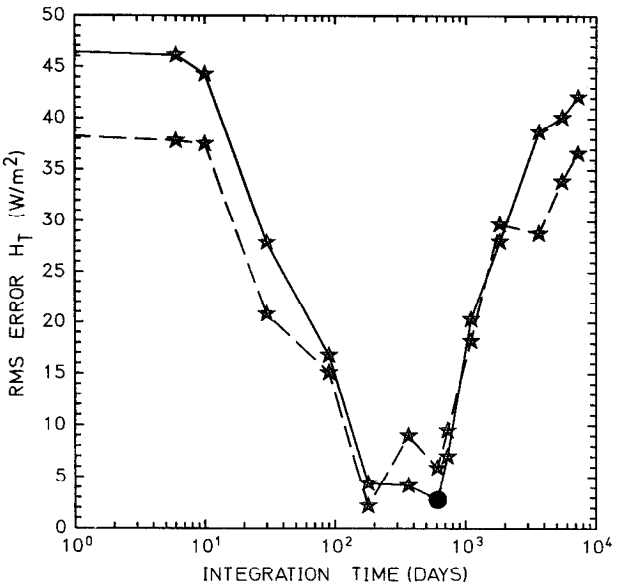

Figure 7. Rms-difference of the heat flux residual (W/m^2) plotted as function of the integration time. Initial guesses for the surface heat flux have been set to zero (solid line) and to zonal mean values of the steady state (dashed line). The stars denote the individual experiments, the dot denotes the single experiment discussed in the text (Fig. 2 to Fig. 6).

misfits, which would characterize a reasonable solution. It has been demonstrated by Marotzke (1992), that using only one time step must fail when applied to a primitive equation model. To summarize his results, the main reason was found in the convective adjustment processes that need more than one time step to adjust to a modified model state. For the same reason, Tziperman *et al.* (1992a) failed to find a reasonable solution in the northern, convectively active part of their primitive equation model, using again one time step for the forward and backward integrations. A model with many degrees of freedom and few convective grid points may show a reduced sensitivity to the problems reported above. Nevertheless, one has to be aware of the possibility that even a small number of convection points may prevent an effective minimization in addition to poor results in the convective region itself.

To determine the optimal integration time, two sets of experiments were performed. Using the same model configuration as described above, the initial guesses for the surface fluxes were either set to zero or to the zonal mean values of the spin-up run, and the optimization was performed for several different integration times. The resulting rms-errors for the heat fluxes are shown in Figure 7. The optimal integration time with the smallest errors in the surface heat flux lies in a temporal window between approximately 100 and 1000 days (an explanation will be given below). In this range the rms-errors vary between about 2 and 10 W/m^2, showing a strong increase in the error for shorter and longer integration times. A similar picture arises for the freshwater flux with minimum errors of less than 0.2 m/year in the time range of 100 to 1000 days.

Marotzke (1992) performed data assimilation experiments with a simple box GCM. Using data from the steady state of his model, the experiments proved successful for integration times between about 5 and 50 years, with the lower integration time being not far away from the optimal time range described above.

Tziperman *et al.* (1992b) have suggested reformulating the steadiness penalties in the cost function (Eq. 1). Investigating the role of ocean waves in the steady state demand, they proposed to involve the sum of squares of the differences between the initial temperature and the temperature at several different times between the initial and final states. Their analysis suggests an optimal integration time of the order of one to a few years.

Observation errors are inherent to all climatological hydrographic data sets. Concerning our inverse approach, errors in the hydrographic data might prevent an improvement of the surface fluxes. To simulate the noise in the hydrographic data, randomly generated Gaussian noise has been added to the model data that serve as simulated observations in our experiments. The model configuration is the same as before, i.e. no surface flux observations are used. To obtain some information about the dependence of the final rms-errors of the surface fluxes on the noise level in the hydrographic data we performed three experiments in addition to the one already described above. Table 1 shows the final rms-errors of the surface heat flux for

Table 1. Final rms-errors of surface heat flux H_T as a result of different noise amplitudes in the observed hydrography. The profiles of noise amplitudes (decreasing from top to bottom) for T, S have been chosen somewhat arbitrarily from error profiles of the Levitus data (Levitus, 1982) in the North Atlantic. The left column shows the noise amplitude of the surface layer temperature for the four experiments, the right column contains the final rms errors of the surface heat flux (similar results are obtained for the freshwater flux).

Noise amplitude in surface layer (°C)	rms-error H_T (W/m^2)
0	2.8
0.019	3.0
0.19	6.5
1.9	23.9

different noise amplitudes in the hydrographic data. The rms-errors of the surface fluxes are nearly insensitive to small noise amplitudes in the hydrography. However, for noise amplitudes similar to the errors in the surface layers of the Levitus data, i.e. 1°C and 0.5 psu, the rms-errors in the surface fluxes become larger than 20 W/m^2 (rms-error H_S: 0.5 m/year). As a result from these experiments without any surface flux observations, the rms-errors of the surface fluxes show a strong sensitivity to the order of magnitude in the hydrographic data errors. Reasonable estimates for the surface fluxes were obtained with hydrographic data at low noise levels (i.e. ≤ 0.1°C, 0.05 psu).

4. Discussion and conclusions

A simplified adjoint technique has been successfully applied to obtain optimal thermohaline surface fluxes and three-dimensional distributions of temperature and salinity. The present method can be considered as an alternative to a complete adjoint model whenever one is mainly interested in the optimization of thermohaline circulation parameters. It has been shown that with the simplified technique improved estimates for the surface fluxes of heat and freshwater can be achieved. Experiments performed under ideal conditions yielded rms-errors for the thermohaline surface fluxes that were up to an order of magnitude smaller than those known from observations so far. Therefore, it should be possible in principal to obtain better estimates of the oceanic surface fluxes when using real hydrographic data. Care has to be taken, however, for regions with strong convective activity and—to some degree—western boundary currents. The identical twin experiments showed that the adjoint method used here has difficulties to produce sensible results in these areas.

The surface fluxes represent the upper boundary condition for the hydrography (Eq. 8a,b) and are linked to the temperature and salinity fields in the adjoint model through Eq. (12). Although in our specific experimental setup the inverse model did not contain any surface flux observations, "recovering" the hydrography was accompanied by an improvement of the surface fluxes too. Moreover, increasing the

integration time to more than one timestep not only improves the estimates for the hydrography, but also the thermohaline surface fluxes, which were absent in the cost function of this experiment. This can be seen in Figure 7, where the errors are reduced at first with an increase in integration time.

The details of one experiment within the range of low rms-errors in Figure 7 have been discussed in Section 3. It is a surprising result that for longer integration times (larger than 1000 days) the errors increase again. The reasons are not obvious to us. It is likely that this increase is at least partly due to the initial guesses of the surface fluxes, which deviate too much from the true values of the model's steady state. An integration of the prognostic model for a long time range with the wrong fluxes yields a wrong model state far away from the data at the end of the forward integration. It is likely that there exist multiple local minima of the cost function (Marotzke, 1992), and the further the model moves away from the global minimum, the greater is the possibility of falling into a secondary minimum.

Additionally, we offer another possible explanation for the failure of the optimization to retrieve the initial and forcing fields. The wrong model state at the end of the forward integration results in large model-data misfits at the beginning of the backward integration of the adjoint model (recall that the observation misfits are computed at the beginning of the backward integration). For longer integration times the diffusion has to be considered such that locally adjacent Lagrange multipliers (i.e. model-data misfits) are strongly mixed and with it the gradients of the initial conditions lose their information about the original model-data misfit. Finally, the optimization fails to progress, so that the correct minimum of the cost function cannot be found. The loss of information in the adjoint integration concerns only the quality with which the initial conditions can be determined. In contrast, the gradients with respect to the boundary conditions are computed from integrals of the surface Lagrange multipliers over time, which tend not to zero for long integration times.

One could argue that an improvement in the results might be obtained when performing longer model integrations, i.e. when the data are not assimilated only once but repeatedly. One has to keep in mind that the first forward run of the model was started from the true hydrography. Consequently, the model state at the end of the first model run shows a larger deviation from the true hydrography than a model state some time steps after the beginning of the prognostic computation. For this reason, the gradients of the control variables are mainly determined by the large model-data misfits at the beginning of the backward integration of the adjoint model. Now the same thing happens as described before and the largest misfits determine the adjoint solution. Consequently, the minimization stalls and does not make any further progress toward lower cost function values. In both experiments shown in Figure 7, an integration time of more than about 1000 days did not give satisfactory results. It is important to note that the described effect of diffusion on time scales

down to a few years does not correspond to the diffusion time scale of a purely prognostic model, which is of the order of centuries or millenia. However, the effects of diffusion obviously may influence the accuracy of the gradients and thus the results of the minimization even on shorter time scales.

The results of the authors mentioned above together with our own experiences seem to suggest that the optimal integration time lies in a range of a few months up to some decades, depending on the kind and quality of the data. This integration time is sufficient to resolve the oceanic response within the surface layers and the main thermocline to modified boundary conditions. A shorter integration time prevents an adjustment of the model to the modified initial and boundary conditions. On the other hand, optimization experiments with long time integrations of a complex GCM may fail, if the experiments use a bad first guess. Probably, the optimization finds only local minima of the cost function or the increasing effect of diffusion leads to inaccurately determined gradients and therefore to wrong estimates of the control variables.

Application of the approximated adjoint model using real oceanographic data for the Atlantic Ocean will be discussed in Schiller (1995).

Acknowledgments. The experiments were performed as part of the Sonderforschungsbereich 133, "Warmwassersphäre des Atlantiks," financed by the DFG. Thanks go to Annegret Schurbohm for her help with the figures.

REFERENCES

Anderson, D. L. T., K. Bryan, A. E. Gill and R. C. Pacanowski. 1979. The transient response of the North Atlantic: Some model studies. J. Geophys. Res., *84,* (C8), 4795–4815.

Bergamasco, A., P. Malanotte-Rizzoli, W. C. Thacker and R. B. Long. 1993. The seasonal steady circulation of the eastern Mediterranean determined with the adjoint method. Deep-Sea Res., *40,* 1269–1294.

Bogden, P. S., R. E. Davis and R. Salmon. 1993. The North Atlantic circulation: Combining simplified dynamics with hydrographic data. J. Mar. Res., *51,* 1–52.

Bryan, K. 1969. A numerical method for the study of the circulation of the world ocean. J. Comput. Phys., *4,* 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, 3-dimensional model of the ocean. GFDL Ocean Group Tech. Rep. No. 1, GFDL/Princeton University.

—— 1987. GFDL Ocean Model Circular No. 8, GFDL/Princeton University.

Le Dimet, F. X. and O. Talagrand. 1986. Variational algorithms for analysis and assimilation of meteorological observations: Theoretical aspects. Tellus, *38A,* 97–110.

Levitus, S. 1982. Climatological atlas of the world ocean, NOAA Prof. Pap. 13, U.S. Govt. Print. Office, Washington, D.C., 173 pp.

Marotzke, J. 1990. Instabilities and Multiple Equilibria of the Thermohaline Circulation, Berichte aus dem Institut für Meereskunde an der Christian—Albrechts-Universität Kiel, Nr. 194, PhD thesis, 126 pp.

—— 1992. The role of integration time in determining a steady state through data assimilation. J. Phys. Oceanogr., *22,* 1556–1567.

Marotzke, J. and C. Wunsch. 1993. Finding the steady state of a general circulation model through data assimilation: Application to the North Atlantic Ocean. J. Geophys. Res., *98,* (C11), 20149–20167.

Martel, F. and C. Wunsch. 1993. The North Atlantic circulation in the early 1980s—An estimate from inversion of a finite difference model. J. Phys. Oceanogr., *23,* 898–924.

Olbers, D. J., M. Wenzel and J. Willebrand. 1985. The inference of North Atlantic circulation patterns from climatological hydrographic data. Rev. Geophys., *23,* 313–356.

Richtmyer, R. D. and K. W. Morton. 1967. Difference methods for initial-value problems, 2. ed., Interscience Publishers (John Wiley & Sons), 405 pp.

Schiller, A. 1995. The mean circulation of the Atlantic Ocean north of 30S determined with the adjoint method applied to an ocean general circulation model. J. Mar. Res., *53,* 453–497.

Stommel, H. and F. Schott. 1977. The beta spiral and the determination of the absolute velocity field from hydrographic station data. Deep Sea Res., *24,* 325–329.

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

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.

Tziperman E., W. C. Thacker, R. B. Long and S. M. Hwang. 1992a. Oceanic data analysis using a general circulation model. Part I: Simulations. J. Phys. Oceanogr., *22,* 1434–1457.

Tziperman E., W. C. Thacker, R. B. Long, S. M. Hwang and S. R. Rintoul. 1992b. Oceanic data analysis using a general circulation model. Part II: A North Atlantic model. J. Phys. Oceanogr., *22,* 1458–1485.

Willebrand, J. and C. Wunsch. 1990. Inversion of ocean circulation models, EOS Trans. AGU, *71,* 2–3, 5.

Wunsch, C. 1978. The North Atlantic general circulation west of 50W determined by inverse methods. Rev. Geophys. Space Phys., *16,* 583–620.

Received: *21 June, 1994;* revised: *31 January, 1995.*