4-1-2006

A Comparison of Data Assimilation Methods Using a Planetary Geostrophic Model

Edward D. Zaron
Portland State University, ezaron@pdx.edu

Citation Details

A Comparison of Data Assimilation Methods Using a Planetary Geostrophic Model

EDWARD D. ZARON

College of Oceanic and Atmospheric Sciences, Oregon State University, Corvallis, Oregon

ABSTRACT

Assimilating hydrographic observations into a planetary geostrophic model is posed as a problem in control theory. The cost functional is the sum of weighted model and data residuals. Model errors are assumed to be spatially correlated, and hydrographic station data are assimilated directly. Searches in state space and data space, for minimizing the cost functional, are compared to a direct matrix inversion algorithm in the data space.

State-space methods seek the minimizer of the cost functional by performing a preconditioned search in an N-dimensional space of state or control variables, where N is approximately 650 000 in the present calculations. Data-space methods solve the Euler–Lagrange equations for the extremum of the cost functional by working in an M-dimensional dual space, where M is the number of measurements. The following four solvers are compared: (i) an iterative state-space solver, with a naive diagonal matrix preconditioner; (ii) an iterative state-space solver, with a sophisticated preconditioner based on the inverse of the model's dynamical operators; (iii) an iterative data-space solver, with no preconditioning; and (iv) a direct, $M \times M$ matrix inversion, data-space solver. The best solver is the iterative data-space solver, (iii), which is approximately 10 times faster than the sophisticated preconditioned state-space solver, (ii), and 100 times faster than the direct data-space solver, (iv).

1. Introduction

The inference of oceanic flows from hydrographic observations has a long history in physical oceanography. The large-scale, time-averaged flow field, that is, the general circulation, is presumed to be in nearly geostrophic balance with the large-scale, time-averaged pressure field. As the pressure is related to density via hydrostatic balance, the thermal-wind relations for the vertical shear in the geostrophic flow have provided the central relationship between the flow field and the quantities observed by hydrographic measurements—conductivity, temperature, and depth—from which the density may be computed. This procedure of computing the geostrophic shear from observations of density is the "dynamic method." It determines the deviations of the velocity from an unknown reference velocity. The works of Wunsch (1978) and Stommel and Schott (1977) were seminal studies of systematic procedures for removing the indeterminacy of the velocity field obtained via the dynamic method.

Today, this early work is understood in a more general context. It is now understood that determining the general circulation fits naturally into the disciplines of control theory and statistical estimation theory. These more general frameworks seek to find that set of flow and hydrographic fields that are most consistent with a theoretical model (i.e., the equations of motion that govern the scales of interest) and a set of observations (i.e., the hydrographic fields and, in some cases, other observations). This wider framework has several advantages. First, the theory provides systematic procedures and tools for inferring observed fields at sites other than those where observations were made (Bretherton et al. 1976). Second, the theory of generalized inversion permits the estimation of quantities that were not directly observed, and it provides guidance for the design of observational programs (Bennett and McIntosh 1982). Last, this theory permits the systemic testing of ocean models (Bennett and Thorburn 1992). The application of these methods in oceanography is mature enough that several textbooks on this subject have been published (e.g., Bennett 1992; Wunsch 1996).

The general circulation problem is a significant computational challenge. The challenge is to find the hy-

Corresponding author address: Edward D. Zaron, Oregon State University, 104 Ocean Admin. Bldg., Corvallis, OR 97331.
E-mail: ezaron@coas.oregonstate.edu

drographic and velocity fields that are most consistent with hydrographic observations (CTD's and XBT's, primarily) and the equations of motion (expressed in an ocean model together with its boundary conditions and forcing functions). The solution of this challenge would consist of a set of smoothed and gridded hydrographic variables, improved estimates of the forcing functions and boundary conditions (i.e., air–sea fluxes), and an indication of the accuracy of the estimated fields. Attempts to determine the general circulation typically incorporate $O(100\ 000)$ or more individual observations and use ocean models with $O(10\ 000\ 000)$ or more computational nodes (e.g., Tziperman and Thacker 1989; Bogden et al. 1993; Marotzke and Wunsch 1993; Schlitzer 1993; Schiller and Willebrand 1995; Marotzke et al. 1999; Stammer et al. 2002). The present work grew out of an effort to quantitatively test a relatively simple ocean model based on the planetary geostrophic equations.

The computations reported below explore the relative efficiencies of solvers for the general circulation problem. Solvers may be classified as either "state space" or "data space" solvers. In state-space solvers, the minimum of the cost functional is sought by searching the space of hydrographic and velocity fields, the variables that define the state of the ocean. Data-space solvers seek to minimize the cost functional by searching in a space of dimension equal to the number of measurements. Formally, the state space is an infinite-dimensional space of functions, but in practice it is approximated on computer by some large-dimensional vector. The typical dimension, N, of this approximated state space is large, $O(10\ 000\ 000)$, while the dimension of the data space, M, is typically much smaller $O(100\ 000)$. The data-space approach works with the so-called dual variables (Courtier 1997a). Application of this approach to data assimilation in oceanography was first used by Egbert et al. (1994), and it was independently discovered by Amodei (1995).

It is found that an iterative data-space solver is more efficient than the state-space solvers, even when a sophisticated preconditioner is used in the state-space search. The measure of efficiency of the solvers is based on the count of executions of the model operator during the running of the algorithms. Care has been taken to implement the algorithms in such a way as to make a fair comparison; however, the results reported herein do not preclude the existence of particular implementations or special cases with different performance characteristics.

The next section briefly defines the test problem in the context of variational data assimilation (e.g., Courtier 1997b) and describes the ocean model and dataset

used in the comparisons. Then, the solvers are described and the computational work requirements of each are laid out in detail. In the last section, example calculations are shown that demonstrate the efficacy of the data-space approach.

2. Problem statement

a. The cost functional

The goal is to find a set of hydrographic fields, generically denoted by u (current, temperature, pressure, etc.), consistent with both the ocean model, which we denote formally as

$$\mathbf{L}u = \mathbf{F}, \tag{1}$$

and with an M-dimensional vector $\mathbf{d}$ of hydrographic data,

$$\mathbf{d} = \mathbf{H}u. \tag{2}$$

In Eq. (2), the operator $\mathbf{H} = [h_1, \ldots, h_M]^{\mathrm{T}}$ is a vector of M linear measurement operators that relate the observed data to the unknown oceanic state, u. Here $\mathbf{L}$ is shorthand for the dynamical operator, which is linear in the present study, and each h_i is assumed to be a linear functional that acts on u and produces a real number. Model inhomogeneities, including boundary data, are represented by $\mathbf{F}$. It is assumed that $\mathbf{L}$ and $\mathbf{F}$ are defined so that the operator $\mathbf{L}$ is invertable; that is, Eq. (1) is mathematically well posed.

Because of approximations and errors in the dynamical model, boundary condition data, and measurements, there is no reason to expect that there exists a u that satisfies both (1) and (2) simultaneously. Thus, one seeks the generalized inverse of the system, that is, the field u that minimizes the quadratic cost functional

$$J(u) = (\mathbf{L}u - \mathbf{F}) \cdot \mathbf{C}_{\mathbf{F}}^{-1} \cdot (\mathbf{L}u - \mathbf{F})$$
$$+ (\mathbf{H}u - \mathbf{d})^{\mathrm{T}} \mathbf{C}_{\mathbf{d}}^{-1} (\mathbf{H}u - \mathbf{d}). \tag{3}$$

The symbols $\mathbf{C}_{\mathbf{F}}$ and $\mathbf{C}_{\mathbf{d}}$ are the covariances for the model and data errors, respectively. The inverses of the covariances prescribe the importance or relative weight of the corresponding constraints in the cost functional. Superscript T stands for vector transpose, and $\cdot$ indicates an inner product over the appropriate space.

b. The ocean model

The above formulation is abstract and general, but it succinctly states the form of the minimization problem. To focus on the general circulation, the operator $\mathbf{L}$ is defined next.

It is assumed that the oceanic state is approximately

governed by the planetary geostrophic equations, a system suggested by Hasselmann (1982) for numerical models of the general circulation. Potential temperature, denoted by θ, is used as the vertical coordinate (Bleck 1978).

The model describes the steady, geostrophic motion of a layer of fluid on the surface of the earth. It is assumed that Fick's law provides an adequate description for the turbulent mixing of heat. The mean stratification (about which the model is linearized) is maintained by a small amount of uniform upwelling, which is determined by the global balance of deep-water production. Solutions of this system can be regarded as the steady response of damped, baroclinic Rossby waves to forcing at the upper, lower, and eastern boundaries. Pedlosky (1992) sought to explain the vertical layering of the abyssal circulation by considering the response of this system to surface Ekman pumping.

Because the planetary geostrophic system described below is linear, it is necessary to distinguish between the mean state and perturbation quantities. The mean state, about which the system is linearized, is indicated with the overbar, and the perturbation quantities are denoted *without* primes. For example, the height of an isotherm, a function of λ, ϕ, and θ, is given by $\overline{z}(\theta) + z(\lambda, \phi, \theta)$, where $\overline{z}(\theta)$ is the mean profile of depth as a function of temperature.

Spherical coordinates are used, with longitude λ, and latitude ϕ (Lorenz 1967). As mentioned above, potential temperature, θ, is the local vertical coordinate. The total velocity vector is (u, v, ω), where the velocity components are defined by

$$u = a\cos\phi\frac{d\lambda}{dt}, \quad v = a\frac{d\phi}{dt}, \quad \omega = \frac{d\theta}{dt}, \quad (4)$$

and a denotes the radius of the earth. The along-isotherm (adiabatic) velocity vector is denoted $\mathbf{u} = (u, v)$. Note that the conventional vertical velocity is

$$\frac{dz}{dt} = z_t + \mathbf{u}\cdot\boldsymbol{\nabla}z + \omega z_\theta, \quad (5)$$

rather than ω, which is used here. Buoyancy $\rho g/\rho_o$, is denoted by b, and is determined from a linear equation of state: $\rho = \rho_o(1 - \alpha\theta)$, with $\alpha = 0.210^{-3}(°C)^{-1}$. Here B is the geostrophic Bernoulli function (Welander 1971), which is related to pressure, buoyancy, and z by $B = (p/\rho_o) + \overline{b}z$. The Coriolis parameter is $f = 2\Omega\sin\phi$, where Ω is the earth's angular rotation rate. The operators $\boldsymbol{\nabla}$ and $\boldsymbol{\nabla}\cdot$ are, respectively, the horizontal gradient and divergence operators in spherical coordinates.

The horizontal momentum equations are simply geostrophic balance,

$$f\mathbf{k}\times\mathbf{u} = -\boldsymbol{\nabla}B; \quad (6)$$

the vertical momentum equation is the hydrostatic balance,

$$\overline{b}_\theta z = B_\theta; \quad (7)$$

the continuity equation is equivalent to conservation of volume,

$$\boldsymbol{\nabla}\cdot(\mathbf{u}\overline{z}_\theta) + (\omega\overline{z}_\theta)_\theta = 0; \quad (8)$$

and the energy equation expresses a simplified form of conservation of internal energy

$$\omega\overline{z}_\theta = -\left(\frac{K_V}{\overline{z}_\theta^2}z_\theta\right)_\theta. \quad (9)$$

Boundary conditions are specified at θ_B, the bottom isotherm,

$$z = z_B, \quad (10)$$

$$\omega\overline{z}_\theta = w_B, \quad (11)$$

at θ_T, the top isotherm,

$$z = z_T, \quad (12)$$

$$\omega\overline{z}_\theta = w_T, \quad (13)$$

and at $\lambda = \lambda_E$, the eastern boundary,

$$z = z_E, \quad (14)$$

$$B(\lambda_E, \theta_B) = B_E. \quad (15)$$

Lower-case Greek characters that appear as subscripts denote partial derivatives; Roman character subscripts are part of the variables' names.

In the language of the previous section, the abstract state variable and the variables in the planetary geostrophic system are related via $u = (\mathbf{u}, \omega, z, B)$, the forcing and boundary conditions are $\mathbf{F} = (z_B, w_B, z_T, w_T, z_E, B_E)$, and the operator $\mathbf{L}$ is expressed by Eqs. (6)–(15).

This form of the model and boundary conditions yields a mathematically well-posed problem. Details regarding the numerical implementation of the model will be presented below. The physical parameters used in the model may be found in Table 1; numerical parameters are in Table 2.

c. The observations

For the experiments reported below, two datasets are used. The first experiment uses an observing array consisting of 662 measurements of the 9.7°C isotherm depth, a proxy for the thermocline depth. Figure 1 shows the locations of the measurements along with the model domain. In the second experiment, an array of

TABLE 1. Physical parameters.

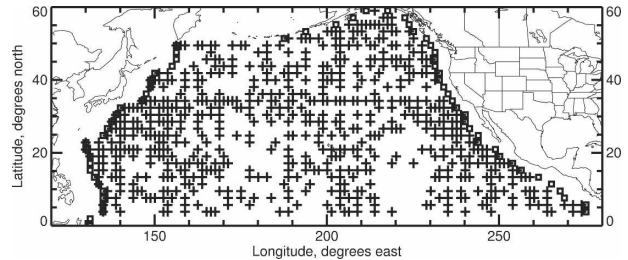| Parameter	Value
a, earth's radius	6×10^3 km
T_{day}, length of day	8.64×10^4 s
g, gravitational acceleration	9.8 m s^{-2}
K_V, vertical diffusivity	1×10^{-4} m^2 s^{-1}
L_{TC}, vertical length scale	538 m
$\bar{w}$, mean upwelling	6.6 m yr^{-1}

FIG. 1. Data locations for thermocline depth experiment. The 662 measurement locations used in the thermocline depth experiment are shown. The observations consist of the depth of the 9.6°C isotherm at each location. Squares indicate the zonal boundaries of the computational domain.

hydrographic stations spanning the subtropical gyre is used. The depths of 13 different isotherms were used, resulting in a total of 1143 observations. Figure 2 shows the locations of the hydrographic stations. The abstract measurement operator h_i simply corresponds to a point measurement of the z field at the ith location.

The data source for both experiments consists of a quality-controlled set of archived hydrographic casts (Reid and Mantyla 2006). The North Pacific was chosen as the site for these experiments because of the lack of deep convection and the general applicability of the planetary geostrophic system under these conditions.

Other observations are necessary to provide boundary conditions for the model. Ekman pumping velocity, w_T, is derived from the annually averaged wind stress data of Hellerman and Rosenstein (1983). It is assumed that mixing at the ocean bottom is zero, $w_B = 0$. The remaining three boundary conditions on z (at the top, bottom, and eastern boundaries) are set by performing an objective analysis of the entire Reid–Mantyla dataset; these smoothed data were extrapolated to the boundaries. The only remaining boundary condition is on the Montgomery potential on the eastern boundary; $B_E = 0$ is used.

Note that the inverse formulation admits errors in the hydrographic observations and the boundary data.

d. The error covariances

For these calculations, error is admitted in the hydrographic observations, the boundary conditions [Eqs.

(10)–(15)], and the continuity Eq. (8). The error in the continuity equation is assumed to arise from the linearization of the planetary geostrophic system; indeed, the mean flow speeds are comparable to the phase speed of the baroclinic Rossby waves. The momentum equations and the energy equation are assumed to hold without error.

As the emphasis of this paper is on the solution methods, the reader is referred to Zaron (1995) for a justification of the assumed errors. Table 3 lists the standard deviations of the observation errors. Table 4 lists the magnitude and correlation scale of the assumed errors in the boundary conditions and continuity equation. The "% error" column indicates that the magnitude of the assumed errors is based on a fixed percentage of the area average of the variable, computed from the first-guess solution.

The prior error covariances, as illustrated in Figs. 3 and 4, are neither homogeneous nor isotropic in spherical polar coordinates. The mathematical forms of the covariance functions have been chosen to ensure that the minimizer of (3) is an element of a reproducing kernel Hilbert space (Bennett and McIntosh 1982; Wahba 1990). The purpose of this is to ensure that the

TABLE 2. Numerical model parameters.

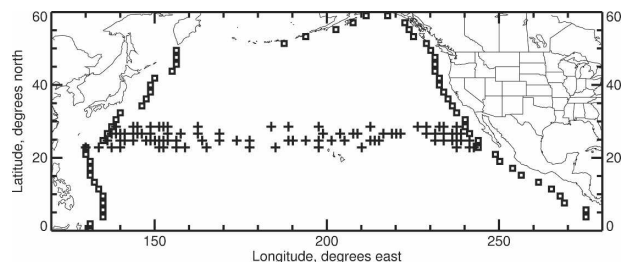| Parameter	Value
Model domain	North Pacific (1°–60°N 120°E–60°W)
$\Delta\lambda$, zonal resolution	1° (approx)
$\Delta\phi$, meridional resolution	2° (approx)
$\Delta\theta$, vertical resolution	0.5°C (approx)
nx, zonal grid points	128
ny, meridional grid points	32
n, vertical grid points	32
θ_B, bottom temperature	1°C
θ_T, top temperature	15°C

FIG. 2. Data locations for zonal section experiment. The sites of the zonal section station data are shown. At each location, as many as 13 isotherm depths are measured. There are a total of 1143 observations. The squares indicate the boundaries of the computational domain.

TABLE 3. Prior measurement errors.

Layer temperature (°C)	Uncertainty (m)
2.0	56
2.9	78
3.9	96
4.9	104
5.8	89
6.8	80
7.8	73
8.7	68
9.7	63
10.7	60
11.6	57
12.6	54
13.6	52

TABLE 4. Prior covariance values.

Term	L_{corr}	θ_{corr}	% error
Continuity equation, C^z	500 km	2°C	20
Top isotherm, C^{z_T}	250 km	N/A	20
Ekman pumping, C^{w_T}	1000 km	N/A	20
Bottom isotherm, C^{z_B}	250 km	N/A	20
Bottom w, C^{w_B}	500 km	N/A	20
Eastern boundary, C^{z_E}	250 km	2°C	20

inverse solution is a physically realizable field that does not contain delta functions, discontinuities, or other singularities.

3. Solution methods

In this section, the methods for finding the unique minimizer of (3) are briefly reviewed.

To provide a basis for comparing the solution methods, one must tally the computational work requirements of each. Rather than counting floating point operations (FLOPs), use is made of the fact that the various methods all must use one or more of the following operators: $\mathbf{L}$, $\mathbf{L}^{-1}$, $\mathbf{L}^*$, and $\mathbf{L}^{-*}$. Note that $\mathbf{L}$ is a partial-differential operator, $\mathbf{L}^{-1}$ is its inverse, $\mathbf{L}^*$ (another partial-differential operator) is the adjoint of $\mathbf{L}$, and $\mathbf{L}^{-*}$ is the inverse of $\mathbf{L}^*$. It is natural to count the work requirements of the search algorithms in units of "model integrations," where one model integration is the FLOP count for implementing $\mathbf{L}^{-1}$. The FLOP counts for the numerical implementations of $\mathbf{L}$, $\mathbf{L}^*$, and $\mathbf{L}^{-*}$ are nearly identical to the FLOP count for $\mathbf{L}^{-1}$; hence, the work required to use each operator is counted as one model integration. The work requirements for each solution method are then counted by adding up the number of times any one of these operators is used.

a. The representer expansion

The representer expansion is the basis for an elegant method of minimizing the cost functional. The elegance of the method lies in its ability to transform a system of coupled partial-differential equations (the Euler–Lagrange equations for an extremum of the cost functional) into a set of trivially coupled systems and a matrix inversion problem. To each measurement operator, $\mathbf{h}_m$, there corresponds a representer function, r_m, where

$m = 1, \ldots, M$. The minimizer of Eq. (3) may be written as the sum of a first guess and a linear combination of representers,

$$\hat{u}(\lambda, \phi, \theta) = u_{\mathbf{F}}(\lambda, \phi, \theta) + \sum_{m=1}^{M} b_m r_m(\lambda, \phi, \theta), \quad (16)$$

where the first guess, $u_{\mathbf{F}} = \mathbf{L}^{-1}\mathbf{F}$, is the exact solution of Eq. (1), and the b_ms are scalar representer amplitudes. The explicit dependence of $\hat{u}$, $u_{\mathbf{F}}$, and r_m on the spatial coordinates has been written out in (16) to emphasize that the representer expansion has transformed an infinite-dimensional problem (with unknown function $\hat{u}$) into a finite-dimensional problem (with M unknowns, b_m).

Each representer is obtained by solving the following system:

$$\mathbf{L}r_m = \mathbf{C}_{\mathbf{F}}\alpha_m, \quad (17)$$

$$\mathbf{L}^*\alpha_m = \mathbf{h}_m, \quad (18)$$

which may be derived from the Euler–Lagrange equations for the minimizer of Eq. (3). The representer amplitudes, $\mathbf{b} = (b_1, \ldots, b_M)$, are found by solving the $M \times M$ system of equations

$$(\mathbf{R} + \mathbf{C}_{\mathbf{d}})b = \mathbf{d}', \quad (19)$$

where $\mathbf{R}$ is the $M \times M$ representer matrix with columns

$$[\mathbf{R}]_m = \mathbf{H}r_m. \quad (20)$$

The $M \times 1$ vector $\mathbf{d}'$ is the difference between the data and the measurements of the first-guess field, $\mathbf{d}' = \mathbf{d} - \mathbf{H}u_{\mathbf{F}}$.

For future reference, the solution method just described shall be called REP. The computation work requirements of REP are $2M + 3$ model integrations [1 integration for the first guess, $2M$ integrations to solve Eqs. (17) and (18), and 2 more integrations to solve for $\hat{u}$]. Note also that an $M \times M$ linear system (19) must be inverted to use this method. This is the algorithm described in section 3.1.2 of Bennett (2002).

In realistic problems, the $M \times M$ system is too large to explicitly construct and invert. One solves Eq. (19) via an iterative method. For linear, positive-definite, symmetric systems, the conjugate gradient algorithm is
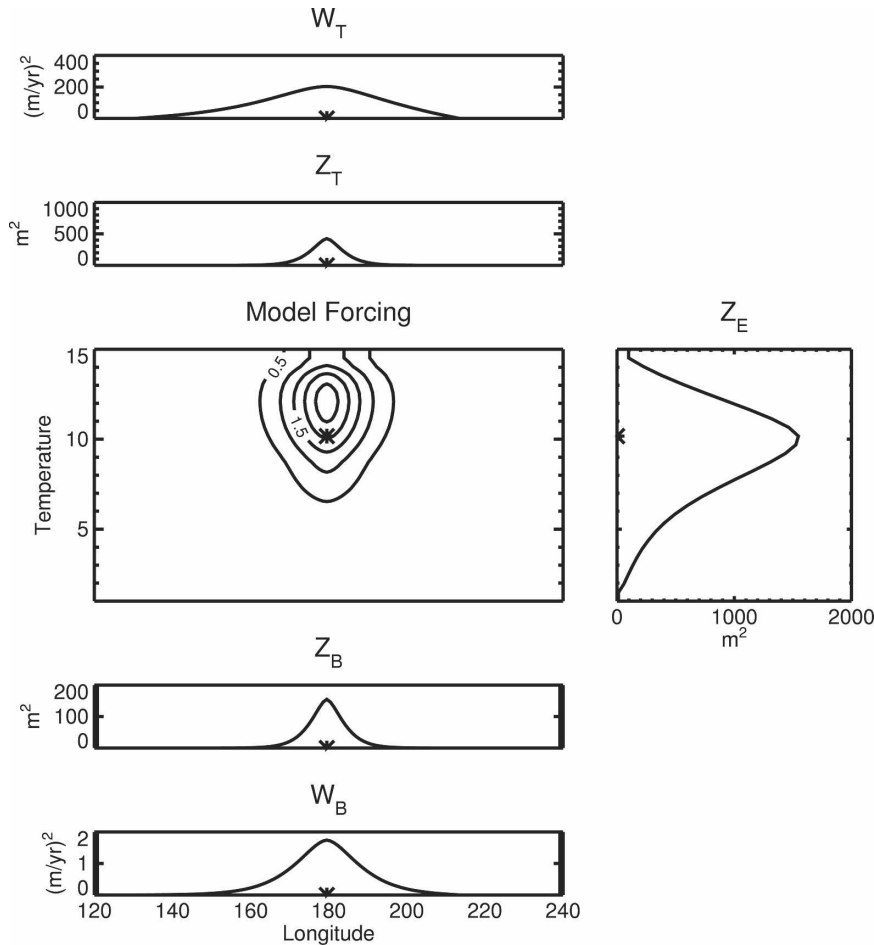
FIG. 3. Prior forcing covariances: zonal section. The prior model forcing error covariances at 25°N are shown. Each panel shows the covariance of the field with respect to the point marked with the asterisk (*). Units of m² (yr °C)$^{-2}$ are used in the central panel entitled "model forcing."

optimal (Luenberger 1973). This is the algorithm presented in section 3.1.4 of Bennett (2002). Let us call this method REP-CG. Each iteration of the conjugate gradient algorithm requires two model integrations, one forward ($\mathbf{L}^{-1}$) and one adjoint ($\mathbf{L}^{-*}$). To actually solve (19) will require some unknown number of conjugate gradient search steps, $N_{\text{REP-CG}}$. Once the optimal $\mathbf{b}$ vector is found, three more model integrations are required to obtain $\hat{u}$. The total work of this method is $2N_{\text{REP-CG}} + 3$.

The convergence properties of the conjugate gradient algorithm depend on the conditioning of the matrix ($\mathbf{R} + \mathbf{C_d}$). For comparison purposes, it will prove useful to express this matrix in a different form. Using Eqs. (17), (18), and (20) one may show that $\mathbf{R} = \mathbf{H}\mathbf{L}^{-1}\mathbf{C_F}\mathbf{L}^{-*}\mathbf{H}^{\text{T}}$. Multiplying this expression on the left and right by $\mathbf{C_d}^{1/2}$ (the Cholesky factorization of $\mathbf{C_d}^{-1}$) allows one to write the following identity:

$$\mathbf{R} + \mathbf{C_d} = \mathbf{C_d}^{1/2}(\mathbf{B}^{\text{T}}\mathbf{B} + \mathbf{I})\mathbf{C_d}^{1/2}, \qquad (21)$$

where $\mathbf{B} = \mathbf{C_d}^{1/2} \, \mathbf{L}^{-*}\mathbf{H} \, \mathbf{C_d}^{-1/2}$. The notation $\mathbf{L}^{-*}$ indicates the inverse of the adjoint model.

b. Direct minimization

The representer expansion solves the Euler–Lagrange equations for the extremum of the cost functional (3). In contrast, state-space methods seek the minimizer, $\hat{u}$, by working directly with J. When these algorithms are implemented on a computer, the formally infinite dimensional state space is reduced to some large, finite-dimensional, vector space. In this section, the symbols used to represent the infinite-dimensional operators $\mathbf{L}$, $\mathbf{L}^{-1}$, etc., are still used, but it should be understood that these symbols actually refer to the finite-dimensional approximations used in the state-space search algorithms.
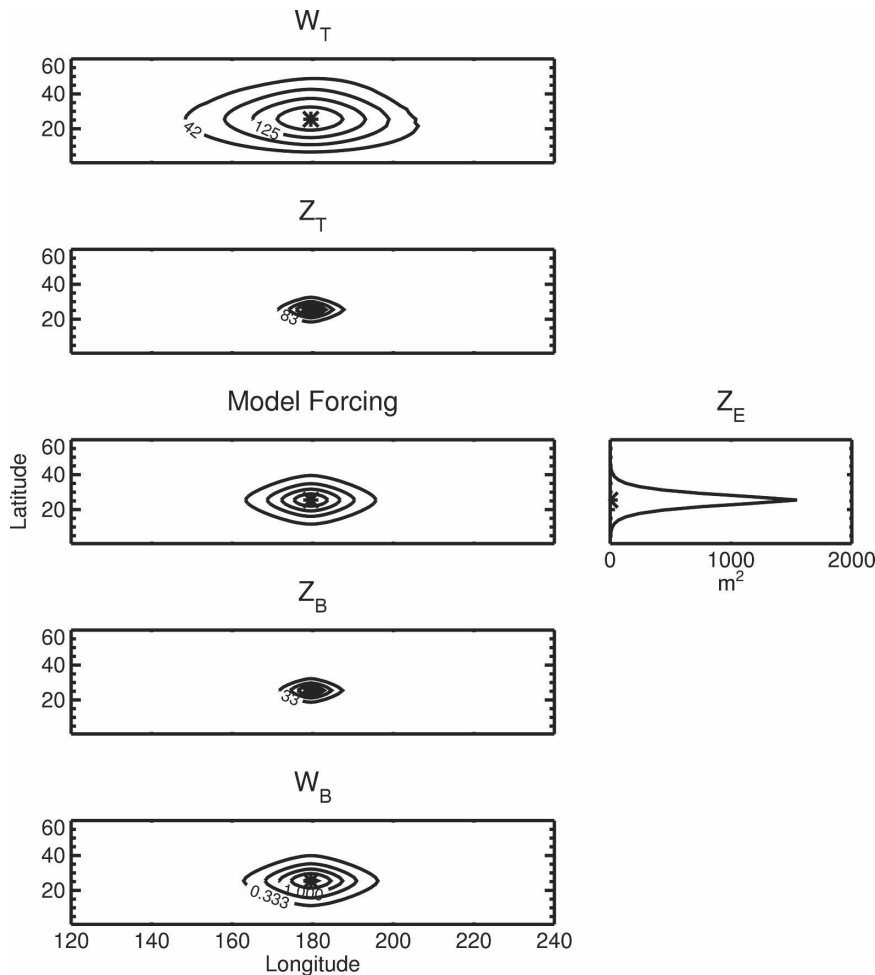
FIG. 4. Prior forcing covariances: plan view. The prior model forcing error covariances are shown in plan view. The units of the contoured fields are the same as those used in the previous figure. Once again, each panel shows the covariance of the field with respect to the point marked with the asterisk (*).

The Hessian matrix of second derivatives of the cost function in (3) is given by

$$\mathbf{S}_u = \mathbf{L}^*\mathbf{C_F}^{-1}\mathbf{L} + \mathbf{H}^\mathrm{T}\mathbf{C_d}^{-1}\mathbf{H}. \tag{22}$$

Because this is a symmetric and positive-definite operator, the conjugate gradient algorithm is still the optimal search algorithm. In practice, one finds that the ill conditioning of $\mathbf{S}_u$ is so severe that a preconditioner is necessary to achieve success with direct minimization methods. Two preconditioning methods are compared below.

The first method, J-CG1, attempts to minimize J using conjugate gradients with a simple rescaling of the variables. Specifically, the rescaling is done so that the diagonal elements of the rescaled Hessian are unity (Stammer et al. 2002). To obtain the preconditioning matrix (the inverse square root of the diagonal of $\mathbf{S}_u$), a significant amount of computational work is necessary. In the present case, 800 model integrations were necessary to evaluate the 650 000 diagonal entries of $\mathbf{S}_u$. In addition, each of the conjugate gradient steps requires two model integrations. The total computational work for this method is $2N_{\text{J-CG1}} + 800$, where $N_{\text{J-CG1}}$ is the number of search steps for this method.

The second method, J-CG2, preconditions the system by making the following change of variables:

$$\boldsymbol{v} = \mathbf{C_F}^{-1/2}\mathbf{L}u. \tag{23}$$

The $N \times N$ operator $\mathbf{C_F}^{-1/2}$ is the Cholesky factorization of $\mathbf{C_F}^{-1}$. In terms of $\boldsymbol{v}$, the new penalty function is

$$J(\boldsymbol{v}) = (\mathbf{H}\mathbf{L}^{-1}\mathbf{C_F}^{1/2}\boldsymbol{v} - \mathbf{d})^\mathrm{T}\mathbf{C_d}^{-1}(\mathbf{H}\mathbf{L}^{-1}\mathbf{C_F}^{1/2}\boldsymbol{v} - \mathbf{d}) + \boldsymbol{v}^\mathrm{T}\boldsymbol{v}. \tag{24}$$

The Hessian of J with respect to $\boldsymbol{v}$ is

$$\mathbf{S}_{\boldsymbol{v}} = \mathbf{I} + (\mathbf{HL}^{-1}\mathbf{C}_{\mathbf{F}}^{1/2})^{\mathrm{T}}\mathbf{C}_{\mathbf{d}}^{-1}\mathbf{HL}^{-1}\mathbf{C}_{\mathbf{F}}^{1/2}. \qquad (25)$$

Note that, under this transformation, the new Hessian is equal to the identity plus a rank-M matrix, that is,

$$\mathbf{S}_{\boldsymbol{v}} = \mathbf{I} + \mathbf{BB}^{\mathrm{T}}, \qquad (26)$$

where $\mathbf{B}^{\mathrm{T}} = \mathbf{C}_{\mathbf{d}}^{-1/2}\mathbf{HL}^{-1}\mathbf{C}_{\mathbf{F}}^{1/2}$ is an $M \times \mathrm{N}$ matrix. Writing $\mathbf{S}_{\boldsymbol{v}}$ in this form reveals that $\mathbf{S}_{\boldsymbol{v}}$ has at most $M + 1$ distinct eigenvalues (Golub and Van Loan 1989), the N eigenvalues of $\mathbf{I}$ are all unity, and $\mathbf{BB}^{\mathrm{T}}$ has at most M nonzero eigenvalues. The conjugate gradient algorithm should converge in at most $1 + M$ iterations. Each step of the conjugate gradient algorithm requires two model integrations. If the method requires $N_{\text{J-CG2}}$ searches, its work may be counted as $2N_{\text{J-CG2}}$ model integrations.

c. Implementation

The cost of the solution methods described above is expressed in units of model integrations. It must be understood that the cost of a model integration is only one of many possible measures to consider when comparing various inverse methods. To convince the reader that this measure is both fair and relevant, this section reviews the details of the implementation of the above solution methods.

To make the comparison fair, the same algorithms (in fact, the same computer subroutines) are used to implement the dynamical and statistical operators in the data-space and state-space methods. The ocean model defined by Eqs. (6)–(15) is implemented numerically on a computer, with the infinite-dimensional partial-differential operators approximated by finite-difference operators on a space of approximate dimension $N = 650\,000$. To explicitly form the operators $\mathbf{L}$, $\mathbf{L}^{-1}$, and their adjoints as $N \times N$ matrices would be impractical. Instead, these operators are implemented as functions; for example, the N state variables at model grid nodes are computed using an algorithmic definition of $\mathbf{L}^{-1}$. The operator $\mathbf{L}$ is hyperbolic (e.g., Courant 1962), with the zonal direction being the time-like dimension, which may be decoupled in the vertical by projection onto vertical modes. The operator $\mathbf{L}^{-1}$ is implemented as a zonal integral for each vertical mode and latitude.

The correct implementation of adjoint codes can be a challenging task, and several implementation tests outlined in Navon et al. (1992) were used to verify (Roache 1997) the code. First, the relationship between the forward and adjoint operators were tested with the definition of the adjoint; that is, the identity

$$(\mathbf{L}\boldsymbol{\eta}) \cdot \boldsymbol{\zeta} = \boldsymbol{\eta} \cdot (\mathbf{L}^*\boldsymbol{\zeta}) \qquad (27)$$

was verified to machine precision for random test fields $\boldsymbol{\eta}$ and $\boldsymbol{\zeta}$. Next, the operator composition was tested by computing the limit of the Gateaux derivative in the direction of the gradient of J (Thepaut and Courtier 1991). Finally, a check was made for agreement between the values of the penalty function computed from the state-space variables and the data-space variables [i.e., $J(\hat{u}) = \mathbf{b}^{\mathrm{T}}\mathbf{Rb} + \mathbf{b}^{\mathrm{T}}\mathbf{RC}_{\mathbf{d}}^{-1}\mathbf{Rb}$].

Tests were also conducted to verify the numerical convergence of the calculations, as a function of grid resolution. Because the representer equations involve the response of the model and adjoint operators to delta function forcing, it is not evident a priori that the solutions obtained from finite-difference approximations will be accurate. A 2×2 representer matrix for measurements at two widely separated sites was used to test for numerical convergence. The grid resolution was refined by factors of 2 until the values in the representer matrix changed by less than 1%. This resolution was used as the working resolution for the calculations.

The hyperbolic character of $\mathbf{L}$ has implications for the structure and implementation of the model error covariance, $\mathbf{C}_{\mathbf{F}}$. Bennett and Budgell (1987) show that the model forcing covariance should be chosen to yield mathematically regular solutions to the underlying partial differential equations, otherwise a solution of the inverse problem may produce an erroneous result. Subsequent grid refinements in an incorrectly posed inverse problem will reveal spurious discontinuities associated with the data sites. For the linearized planetary geostrophic system (6)–(15) a sufficient condition for the physical realizability of the solutions is that the Fourier transform of the covariance function decay like k^{-4}, where k is the one-dimensional wavenumber.

Note that it is necessary to implement convolutions with both the model forcing covariance, its square root [$\mathbf{C}_{\mathbf{F}}\alpha$ in Eq. (17), and $\mathbf{C}_{\mathbf{F}}^{1/2}\boldsymbol{v}$ in Eq. (24)], and the operator inverse of the covariance [$\mathbf{C}_{\mathbf{F}}^{-1}(\mathbf{L}u - \mathbf{F})$ in Eq. (3)]. These operators are implemented in a manner similar to that described in Purser et al. (2003). The square root of the one-dimensional inverse covariance kernel is the differential operator

$$\mathbf{C}_x^{-1/2}(x_1, x_2) = N_x \sigma_x^{-1}(x_1)\left(1 - \gamma_x^{-2}\frac{\partial^2}{\partial x_2^2}\right), \qquad (28)$$

where γ_x^{-1} is the correlation scale in the $\hat{x}$ direction, $\sigma_x(x)$ is the standard deviation at location x, and N_x is a normalizing coefficient (chosen to make the correlation unity at zero lag). This differential operator is approximated using centered differences, and $\mathbf{C}_x^{1/2}$ is stored as a tridiagonal matrix, which may be efficiently inverted. The inverse of the model forcing covariance is expressed as the product
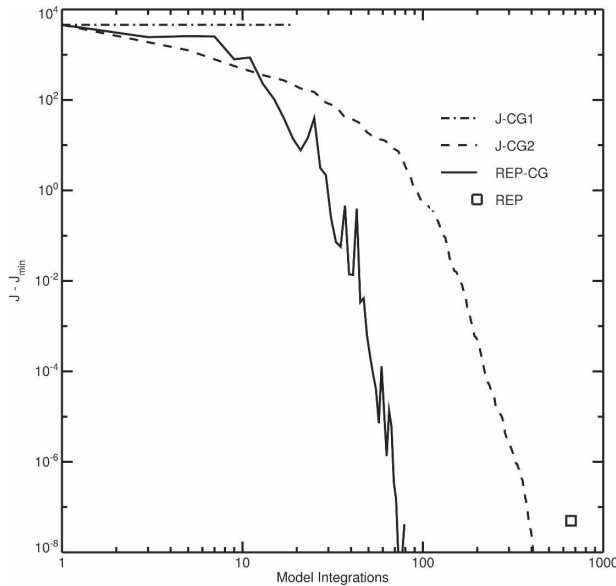
FIG. 5. Solution method comparison for thermocline depth. The plot compares the relative efficiency of the solution procedures listed in Table 5. The x axis measures the relative amount of work in units of model integrations, i.e., the number of times the model operator **L** is inverted. As mentioned in the text, model integrations and substitutions are equally expensive.

$$\mathbf{C_F}^{-1}(\lambda_1, \phi_1, \theta_1; \lambda_2, \phi_2, \theta_2) = \sigma^{-1}(\lambda_1, \phi_1, \theta_1)\mathbf{C}_\lambda^{-T/2}(\lambda_1, \lambda_2)$$

$$\mathbf{C}_\phi^{-T/2}(\phi_1, \phi_2)\mathbf{C}_\theta^{-1}(\theta_1, \theta_2)$$

$$\mathbf{C}_\phi^{-1/2}(\phi_1, \phi_2)\mathbf{C}_\lambda^{-1/2}(\lambda_1, \lambda_2)$$

$$\sigma^{-1}(\lambda_2, \phi_2, \theta_2), \tag{29}$$

which is positive definite and symmetric by construction. For reference, the one-dimensional (infinite domain) spatial correlation function that corresponds to Eq. (28) is

$$c_x(x_1, x_2) = (1 + |x_1 - x_2|\gamma_x)e^{-|x_1 - x_2|\gamma_x}. \tag{30}$$

For every pair of "model integrations" (i.e., the application of either **L**, $\mathbf{L}^{-1}$, or their operator adjoints) the operation count for the convolutions with the model forcing covariance is very similar in each of the above methods. For method J-CG1, the operator $\mathbf{C_F}^{-1}$ is used; the representer methods use the operator $\mathbf{C_F}$; and J-CG2 uses the operator $\mathbf{C_F}^{-1/2}$ twice.

Similarly, the cost of using the measurement operators is identical in each of the solution methods. The REP-CG method uses **H** and its adjoint to compute $(\mathbf{R} + \mathbf{C_d})\mathbf{b}$, via Eqs. (18) and (20). The direct minimization methods also use **H** and its adjoint in the gradient evaluations.

Finally, it is certain that the representer and direct minimization solvers implemented here could be modi-

TABLE 5. Solution methods.

Method	Work per step	One-time work
REP	2	$3 + (M \times M \text{ solve})$
REP-CG	2	3
J-CG1	2	800
J-CG2	2	0

fied so as to substantially alter their relative performance. To improve J-CG2, it may be possible to use a polynomial preconditioner or incomplete LU decomposition (Golub and Van Loan 1989) to form an approximate inverse of the Hessian (25). Likewise, several strategies exist for preconditioning the REP-CG solver (Bennett 2002). These are areas of ongoing research, and many practical and theoretical questions remain unanswered.

4. Results

The solution methods outlined in the previous section were compared in two experiments. The two experiments differed only in the number and distribution of observations. The first experiment used 662 thermocline depth measurements, as shown in Fig. 1. The second experiment used 1143 measurements at 13 depths across the subtropical gyre; the zonal band of station locations is shown in Fig. 2.

Figure 5 shows the progress of the minimization for the iterative solvers J-CG1, J-CG2, and REP-CG. The noniterative, direct solver, REP, is also placed on the graph at a point that indicates its computational work. It can be seen that the iterative representer solver, REP-CG, converged to the solution roughly 10 times faster than the preconditioned state-space search, J-CG2. The naive direct minimization, J-CG1, in which the variables were simply rescaled, made no significant progress compared with the other solvers. Similar performance is shown in Fig. 6, which used the zonal band of observations.

It should be emphasized that each successful solver obtained the same optimal solution. Because the ocean model and measurement operators are linear in this study, there is a unique minimizer of J.

We can gain a partial understanding of these results by considering the conditioning of the relevant Hessian matrices. First, the Hessian for the cost function in the naively preconditioned state-space search, J-CG1, $\mathbf{S}_u$ in Eq. (22), is likely enormous. To see this, note that $\mathbf{C_F}^{-1}$ and **L** are, respectively, fourth- and second-order differential operators. If we let $\hat{m}$ denote the maximum vertical wavenumber that may be represented on the
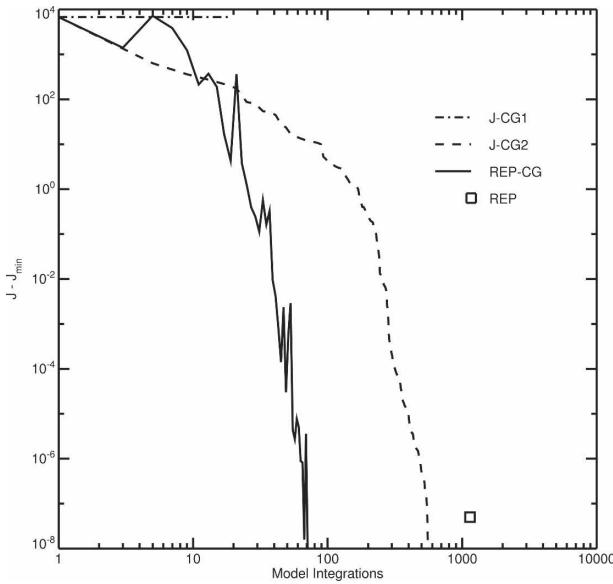
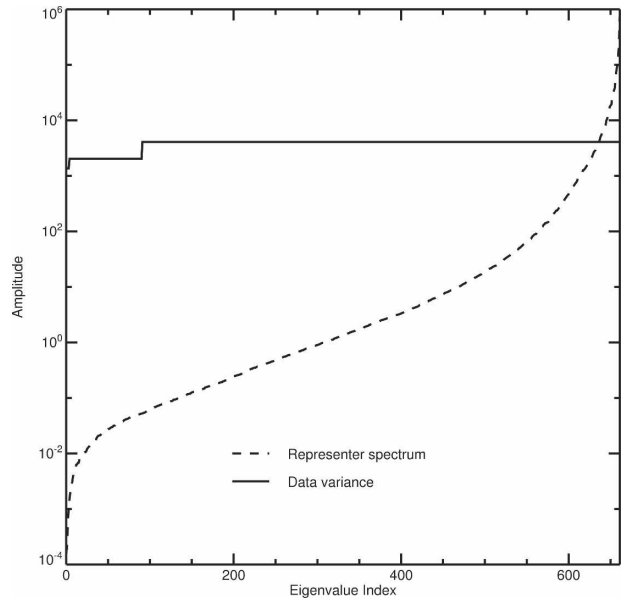FIG. 6. As in Fig. 5 but solution method comparison for zonal section.

FIG. 7. Eigenvalue spectrum for thermocline depth experiment. The plot shows that only about 60 of the representer matrix eigenvalues (of the 662 total) are comparable to or larger than the eigenvalues of $\mathbf{C_d}$. The data error covariance matrix, $\mathbf{C_d}$, is very nearly a rescaled version of the identity matrix. Because the $\mathbf{R} + \mathbf{C_d}$ matrix is approximately a rank-60 perturbation to the rescaled identity matrix, the REP-CG solver converged in about 60 iterates.

finite-difference grid, then the largest eigenvalue of $\mathbf{S}_u$ will scale like $\hat{m}^8$. Even with the state variables nondimensionalized by the preconditioner, the condition number based on just the vertical finite-difference operators will scale like $n^8 \approx 10^{12}$, where the number of vertical levels, n, is 32. This condition number is large enough that the descent method fails in double precision arithmetic (which nominally carries 15 digits). When there is enough data to constrain both the large and small scales, it is conceivable that the $\mathbf{H}^T \mathbf{C_d}^{-1} \mathbf{H}$ term will control the condition number of $\mathbf{S}_u$, but this situation only arises when the model dynamics are irrelevant to the inversion.

The conditioning of the REP-CG and J-CG2 solvers can be compared by looking at the expressions for the Hessian matrices in Eqs. (21) and (26). Note that the matrix B is the same in both cases. If the REP-CG solver had been preconditioned with the transformation $\hat{\mathbf{b}} = \mathbf{C_d}^{1/2}\mathbf{b}$, the condition numbers of the REP-CG and J-CG2 solvers would have been identical. This result is identical to that derived by Courtier (1997a) for time-dependent problems. Based on the conditioning of the Hessian, one would expect the performance of the data-space and state-space search algorithms to be comparable when the latter is preconditioned with the appropriate dynamical and covariance operators [as in Eq. (23)].

Why should it be that the data-space search, REP-CG, performed about 10 times faster than the preconditioned state-space search, J-CG2? Should one expect similar results with other models and datasets? Golub

and Van Loan (1989) point out that conjugate gradient algorithms generally solve matrix equations like $\mathbf{A}x = \mathbf{b}$, where $\mathbf{A}$ is a symmetric positive definite matrix, in $r + 1$ steps, when $\mathbf{A}$ is a rank r correction to the identity. Thus, the REP-CG and J-CG2 solvers must both converge to their solutions within M iterations.

Because the dimension of the data space is relatively modest in these examples, it is possible to perform a complete eigenvalue analysis of the $\mathbf{R}$ and $\mathbf{C_d}$ matrices. Figure 7 shows the spectrum of eigenvalues of these matrices in the "thermocline depth" experiment. The data covariance matrix, $\mathbf{C_d}$, is diagonal, so its eigenvalues are just its diagonal elements. Also, because the data variances span less than an order of magnitude, the $\mathbf{C_d}$ matrix is approximately a rescaled identity matrix. The representer matrix, $\mathbf{R}$, which may be interpreted as the covariance matrix of measurements of the prior model, has eigenvalues that span many orders of magnitude. Each eigenvalue is the variance of a mode of oceanic variability that is observable by the measurement array, an antenna array mode (Bennett 1985). Significantly, only about 60, or one-tenth, of the antenna array modes have variance comparable to the data variance. In other words, the matrix $\mathbf{R} + \mathbf{C_d}$ is roughly a rank-60 perturbation to the identity matrix. A

comparison with Fig. 5 reveals that the REP-CG1 algorithm converged in slightly more than 60 iterates. An analysis of the eigenvalue spectrum for the "zonal section" experiment (not shown) leads to similar observations.

The degree to which this analysis would apply to other models and datasets cannot be known a priori; however, it would seem that in most realistic cases there will be a significant fraction of the antenna array modes with variance well below the level of the data error variance. If this were not the case, it would suggest that the state variables at the data sites are dynamically uncoupled. In such instances there would seem to be little value in performing the data assimilation.

Inspection of Fig. 5 shows that the J-CG2 solver converged in fewer than M iterations, where $M = 662$. In fact, it outperformed the data-space solver within the first 10 iterations. Because the conditioning of the Hessian should be the same as for the REP-CG solver, there are evidently other causes for the different performance of the algorithms. Briefly, there are the following two possibilities:

- Finite-precision arithmetic and numerical round off errors: finite-precision arithmetic will cause the conjugate gradient search directions to loose the appropriate orthogonality relations during the iterations.
- Different initial states: the state-space and data-space searches begin from different initial states; the convergence of the conjugate gradient algorithm can be influenced by exactly how the initial state projects onto the eigenvectors of the associated Hessian matrix.

The problems due to finite-precision arithmetic are likely to become evident more quickly with the state-space solvers than the data-space solvers, due to the smaller dimensionality of the latter. However, experiments with the Broyden–Fletcher–Goldfarb–Shanno (BFGS) method (i.e., a quasi-Newton method; e.g., Nazareth 1979), which was recommended by Navon and Legler (1987), did not yield significant improvements. There is some evidence that the differing initial states may be a factor in the convergence of the algorithms. The phenomenon of superlinear convergence occurs when the initial state projects preferentially onto the eigenvectors associated with the largest eigenvalues of the Hessian matrix (Beckermann and Kuijlaars 2001). Under these conditions, the converge rate of the algorithm grows as the solution is sought in progressively better conditioned subspaces of the Hessian. The use of log–log coordinates in Figs. 5 and 6 exaggerates the impression of an increasing convergence rate in

both solvers. The REP-CG solver does indeed display superlinear convergence, but there is none at all for the J-CG2 solver. Unfortunately, a thorough analysis of the finite-precision and superlinear convergence aspects of the algorithms is beyond the scope of the present work.

5. Summary

The results of this study highlight the relative merits of data-space searches over state-space searches in the effort to determine the oceanic general circulation via variational or adjoint methods. The use of a relatively simple ocean model has permitted the intercomparison of four solution methods. In spite of its simplicity, the physics of the model are those that should govern the large-scale ocean circulation away from boundaries.

The iterated data-space solver (REP-CG) was able to obtain the unique minimum of the cost functional with the least computational work. A preconditioned state-space search (J-CG2) was found to be more efficient than a direct representer solver (REP), but it was still less efficient than REP-CG. While there are many measures of the efficiency of an algorithm, which depend on details of the implementation and underlying computer architecture, the count of model integrations is applicable across many implementations. It was shown that the count of model integrations scales with the number of convolutions in the algorithms, so identical results would have been obtained if the work had been measured in units of "error covariance convolutions." Perhaps other metrics will prove to be relevant if these results are extended to situations with much larger datasets.

The iterated representer method was found to be the most efficient solver for the Euler–Lagrange equations that define the minimum of the cost functional; however, the applicability of this solver to the planetary geostrophic system relied on the model's linear dynamics. What if the model had been nonlinear? When either the model or the measurement functionals are nonlinear, there are several options. One can use state-space search algorithms to directly minimize the cost functional. Sophisticated preconditioners are required, though these may be built up using, for example, limited-memory quasi-Newton methods (e.g., LeDimet et al. 1997). The large dimension of the state makes the application of these methods to realistic problems challenging. Compromises are possible, such as adjusting the stopping criteria of the search. For example, the minimization algorithm can be stopped once the value of $J(u)$ has been reduced to an "acceptable" value consistent with the postulated errors (Stammer et al. 2002).

To obtain the advantages of a data-space algorithm in the nonlinear case, one may consider a sequence of linear inversions, where each inverse problem is obtained by linearizing the model and/or measurement operators around the previous state estimate. There is little that can be said about the convergence of this kind of Picard, or functional, iteration, as it depends on details that are unique to the situation. When the iterations do not converge, step-length damping, the so-called creeping algorithm (e.g., Parker 1994), can stabilize the functional iteration. There has been some success with this method in applications to geophysical fluid flows (Bennett et al. 1997).

The results suggest that it is entirely feasible to assimilate hydrographic station data, rather than smoothed and gridded climatologies, into general circulation models. The flow patterns diagnosed from gridded climatologies such as Levitus (1982) are not steady; hence, many studies have sought to use inverse methods to determine the flow and hydrographic fields simultaneously. Perhaps the climatologies can now be produced by the methods described herein. Use of the REP-CG solver, in concert with preconditioners like those suggested in Bennett (2002), makes feasible the assimilation of very large datasets (e.g., Rosmond and Xu 2006).

REFERENCES

Amodei, L., 1995: Solution approchée pour un problème d'assimilation avec prise en compte l'erreur de modèle. *C. R. Acad. Sci.,* **321,** 1087–1094.

Beckermann, B., and A. Kuijlaars, 2001: Superlinear convergence of conjugate gradients. *SIAM J. Numer. Anal.,* **39,** 300–329.

Bennett, A., 1985: Array design by inverse methods. *Progress in Oceanography,* Vol. 15, Pergamon, 129–156.

——, 1992: *Inverse Methods in Physical Oceanography.* 1st ed. Cambridge University Press, 346 pp.

——, 2002: *Inverse Modeling of the Ocean and Atmosphere.* Cambridge University Press, 234 pp.

——, and P. McIntosh, 1982: Open ocean modeling as an inverse problem: Tidal theory. *J. Phys. Oceanogr.,* **12,** 1004–1018.

——, and W. Budgell, 1987: Ocean data assimilation and the Kalman filter: Spatial regularity. *J. Phys. Oceanogr.,* **17,** 1583–1601.

——, and M. Thorburn, 1992: The generalized inverse of a non-linear quasigeostrophic ocean circulation model. *J. Phys. Oceanogr.,* **22,** 213–230.

——, B. Chua, and L. Leslie, 1997: Generalized inversion of a global numerical weather prediction model, II. *Meteor. Atmos. Phys.,* **62,** 129–140.

Bleck, R., 1978: Finite-difference equations in generalized vertical coordinates. Part I: Total energy conservation. *Beitr. Phys. Atmos.,* **51,** 360–372.

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

Bretherton, F., R. Davis, and C. Fandry, 1976: A technique for objective analysis and design of oceanographic experiments applied to MODE-73. *Deep-Sea Res.,* **23,** 559–582.

Courant, R., 1962: *Methods of Mathematical Physics.* Vol. 2. Wiley-Interscience, 856 pp.

Courtier, P., 1997a: Dual formulation of four-dimensional assimilation. *Quart. J. Roy. Meteor. Soc.,* **123,** 2449–2461.

——, 1997b: Variational methods. *J. Meteor. Soc. Japan,* **75,** 211–218.

Egbert, G., A. Bennett, and M. Foreman, 1994: TOPEX/POSEIDON tides estimated using a global inverse model. *J. Geophys. Res.,* **99,** 24 821–24 852.

Golub, G., and C. Van Loan, 1989: *Matrix Computations.* 2d ed. Johns Hopkins University Press, 642 pp.

Hasselmann, K., 1982: An ocean circulation model for climate variability studies. *Progress in Oceanography,* Vol. 11, Pergamon, 69–92.

Hellerman, S., and M. Rosenstein, 1983: Normal monthly wind stress over the World Ocean with error estimates. *J. Phys. Oceanogr.,* **13,** 1093–1104.

LeDimet, F.-X., H.-E. Ngodock, B. Luong, and J. Verron, 1997: Sensitivity analysis in variational data assimilation. *J. Meteor. Soc. Japan,* **75,** 245–255.

Levitus, S., 1982: *Climatological Atlas of the World Ocean.* NOAA Prof. Paper 13, 173 pp. and 17 microfiche.

Lorenz, E. N., 1967: The nature and theory of the general circulation of the atmosphere. WMO Rep. 218, Tech. Publication 115, 161 pp.

Luenberger, D. G., 1973: *Introduction to Linear and Nonlinear Programming.* Addison-Wesley, 356 pp.

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,** 20 149–20 168.

——, R. Giering, Q. Zhang, D. Stammer, C. Hill, and T. Lee, 1999: Construction of the adjoint MIT ocean general circulation model and application to Atlantic heat transport sensitivity. *J. Geophys. Res.,* **104,** 29 529–29 548.

Navon, I., and D. Legler, 1987: Conjugate-gradient methods for large-scale minimization in meteorology. *Mon. Wea. Rev.,* **115,** 1479–1502.

——, X. Zou, M. Berger, P. Phua, T. Schlick, and F. LeDimet, 1992: Testing for reliability and robustness of optimization codes for large scale optimization problems. *Optimization Techniques and Applications,* K. H. Phua et al., Eds., Vol. 1, World Scientific, 445–480.

Nazareth, L., 1979: A relationship between the BFGS and conjugate gradient algorithms and its implications for new algorithms. *SIAM J. Numer. Anal.,* **16,** 794–800.

Parker, R. L., 1994: *Geophysical Inverse Theory.* Princeton University Press, 386 pp.

Pedlosky, J., 1992: The baroclinic structure of the abyssal circulation. *J. Phys. Oceanogr.,* **22,** 652–659.

Purser, R. J., W.-S. Wu, D. F. Parrish, and N. M. Roberts, 2003: Numerical aspects of the application of recursive filters to variational statistical analysis. Part I: Spatially homogeneous and isotropic Gaussian covariances. *Mon. Wea. Rev.,* **131,** 1524–1535.

Reid, J., and A. Mantyla, cited 2006: (Scripps) NODC deep ocean station profiles. NCAR Data Support Section. [Available online at http:dss.ucar.edu/datasets/ds543.0.]

Roache, P. J., 1997: Quantification of uncertainty in computational fluid dynamics. *Annu. Rev. Fluid Mech.,* **29,** 123–160.

Rosmond, T., and L. Xu, 2006: Development of NAVDAS-AR: Non-linear formulation and outer loop tests. *Tellus,* **58A,** doi:10.1111/j.1600-0870.2006.00148.x.

Schiller, A., and J. Willebrand, 1995: A technique for determination of surface heat and freshwater fluxes from hydrographic observations, using an approximate adjoint ocean circulation model. *J. Mar. Res.,* **53,** 433–451.

Schlitzer, R., 1993: Determining the mean, large-scale circulation of the Atlantic with the adjoint method. *J. Phys. Oceanogr.,* **23,** 1935–1952.

Stammer, D., and Coauthors, 2002: The global ocean state during 1992–1997, estimated from ocean observations and a general circulation model. *J. Geophys. Res.,* **107,** 3105, doi:10.1029/2001JC000888.

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.

Thepaut, J.-N., and P. Courtier, 1991: Four-dimensional variational data assimilation using the adjoint of a multilevel primitive-equation model. *Quart. J. Roy. Meteor. Soc.,* **117,** 1225–1254.

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

Wahba, G., 1990: *Spline Models for Observational Data.* SIAM, 169 pp.

Welander, P., 1971: Some exact solutions to the equations describing an ideal-fluid thermocline. *J. Mar. Res.,* **29,** 60–68.

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

——, 1996: *The Ocean Circulation Inverse Problem.* Cambridge University Press, 458 pp.

Zaron, E., 1995: The time-averaged circulation of the North Pacific Ocean: An analysis based on inverse methods. Ph.D. thesis, Oregon State University, 211 pp.