Three- and Four-Dimensional Variational Assimilation with a General Circulation Model of the Tropical Pacific Ocean. Part I: Formulation, Internal Diagnostics, and Consistency Checks

A. T. WEAVER

Centre Européen de Recherche et de Formation Avancée en Calcul Scientifique/SUC URA 1875, Toulouse, France

J. VIALARD

European Centre for Medium-Range Weather Forecasts, Reading, United Kingdom, and Laboratoire d'Oceanographie Dynamique et de Climatologie/CNRS/IRD/UPMC/MNHN, Paris, France

D. L. T. ANDERSON

European Centre for Medium-Range Weather Forecasts, Reading, United Kingdom

(Manuscript received 20 May 2002, in final form 27 November 2002)

ABSTRACT

Three- and four-dimensional variational assimilation (3DVAR and 4DVAR) systems have been developed for the Océan Parallélisé (OPA) ocean general circulation model (OGCM) of the Laboratoire d'Océanographie Dynamique et de Climatologie. An iterative incremental approach is used to minimize a cost function that measures the statistically weighted squared differences between the observational information and their model equivalent. The control variable of the minimization problem is an increment to the background estimate of the model initial conditions at the beginning of each assimilation window. In 3DVAR, the increment is transported between observation times within the window using a persistence model, while in 4DVAR a dynamical model derived from the tangent linear (TL) of the OGCM is used. Both the persistence and TL models are shown to provide reasonably good descriptions of the evolution of typical errors over the 10- and 30-day widths of the assimilation windows used in the authors' 3DVAR and 4DVAR experiments, respectively.

The present system relies on a univariate formulation of the background-error covariance matrix. In practice, the background-error covariances are specified implicitly within a change of control variable designed to improve the conditioning of the minimization problem. Horizontal and vertical correlation functions are modeled using a filter based on a numerical integration of a diffusion equation. The background-error variances are geographically dependent and specified from the model climatology. Single observation experiments are presented to illustrate how the TL dynamics act to modify these variances in a flow-dependent way by diminishing their values in the mixed layer and by displacing the maximum value of the variance to the level of the background thermocline.

The 3DVAR and 4DVAR systems have been applied to a tropical Pacific version of OPA and cycled over the period 1993–98 using in situ temperature observations from the Global Temperature and Salinity Pilot Programme. The overall effect of the data assimilation is to reduce a large bias in the thermal field, which was present in the control. The fit to the data in 4DVAR is better than in 3DVAR, and within the specified observation-error standard deviation. Intermittent updating of the linearization state of the TL model is shown to be an important feature of the incremental 4DVAR algorithm and contributes significantly to improving the fit to the data.

1. Introduction

The El Niño–Southern Oscillation (ENSO) phenomenon is one of the main contributors to predictability on seasonal to interannual timescales (Barnett et al. 1993; Palmer and Anderson 1994). Accurate ENSO forecasts are thus a prerequisite to the development of a reliable dynamical seasonal forecasting system with a coupled ocean–atmosphere model (CGCM). Early studies (Cane et al. 1986; Latif and Flügel 1991; Balmaseda et al. 1994) showed that some ENSO forecasting skill could be obtained using only observed surface wind stress to initialize the ocean component of the forecasting system. More recent studies have shown that assimilating ocean observations to initialize CGCMs resulted in significantly better ENSO forecasts (Ji and

Corresponding author address: Dr. Anthony T. Weaver, CERFACS, 42 Ave. Gaspard Coriolis, 31057 Toulouse Cedex 1, France.
E-mail: weaver@cerfacs.fr

Leetma 1997; Rosati et al. 1997; Segschneider et al. 2000, 2001; Alves et al. 2002). Subsurface temperature observations from expendable bathythermographs (XBTs) and the Tropical Atmosphere Ocean (TAO) array (McPhaden 1993) were shown to be especially beneficial to forecast skill (Ji and Leetma 1997; Alves et al. 2002; Segschneider et al. 2001). Combining available observations and an ocean model into a dynamically consistent picture of the ocean state can also help to provide better insight into the processes determining ocean variability at various timescales.

In meteorology, optimal interpolation (OI) was the standard technique used for many years for producing initial conditions for numerical weather prediction (NWP; Gandin 1965; Rutherford 1972; Lorenc 1981). More recently, however, most of the major NWP centers have replaced their OI systems by variational assimilation systems (Parrish and Derber 1992; Courtier et al. 1998; Gauthier et al. 1999; Rabier et al. 2000; Lorenc et al. 2000). In variational assimilation, the analysis problem is defined by the minimization of a cost function that measures the statistically weighted squared differences between observations (including a model background state) and their model counterpart. The cost function is minimized with respect to selected control variables and this is done iteratively using a gradient descent method. Variational assimilation overcomes many of the limitations of OI: it allows for greater flexibility for assimilating different observation types (possibly nonlinearly related to the model state); it eliminates the need to split the analysis domain into subsections so that all observations can, in principle, influence the analysis at every model grid point; it provides a more general framework for using more sophisticated background-error covariance models; and it provides a clearer development path toward advanced, four-dimensional assimilation techniques. These advantages are equally relevant for oceanographic data assimilation.

In this paper, we describe three- and four-dimensional variational assimilation (3DVAR and 4DVAR) systems that have been developed for the rigid-lid version of the Océan Parallélisé (OPA) ocean general circulation model (OGCM) of the Laboratoire d'Océanographie Dynamique et de Climatologie (LODYC; Madec et al. 1998). One of the main motivations for developing the system is to produce ocean analyses for seasonal climate forecasting. In the present study, the 3DVAR and 4DVAR systems are applied to produce a reanalysis of the tropical Pacific Ocean over the period 1993–98 using in situ temperature observations. Here, and in the companion paper by Vialard et al. (2003, hereafter referred to as Part II), we evaluate the analyses by focusing on their statistical and physical properties, and their comparison with independent datasets, rather than their impact on climate forecasts.

Both the 3DVAR and 4DVAR systems have been designed following the incremental approach (Courtier et al. 1994). The control variable of the minimization problem is taken to be an increment to the background estimate of the model initial conditions at the beginning of a given assimilation window. In the cost function, the observations are compared to the sum of the background counterpart of the observations and an increment computed in observation space using a linear model. The fundamental difference between the 3DVAR and 4DVAR formulations lies in the level of sophistication of the linear model used to transport the state increment between observation times. In 3DVAR, a simple persistence model is used, whereas in 4DVAR a dynamical model based on the tangent linear (TL) of the OPA OGCM is used. The 4DVAR scheme involves substantially more development than does 3DVAR since an adjoint model must be derived for the linearized version of the OGCM in order to compute the gradient of the cost function with respect to the increment at initial time.

This particular incremental version of 3DVAR can be viewed as a limiting case of incremental 4DVAR in which the TL operator is replaced by the identity matrix. As in 4DVAR, observations can be assimilated at their appropriate measurement times since they are compared directly to the background state, which is propagated in time using the OGCM. For this reason, the scheme has been coined 3D-FGAT, for first guess at appropriate time (Fisher and Andersson 2001).[1] For example, the FGAT feature may be particularly important in the Tropics where an equatorial Kelvin wave can travel more than 2000 km in 10 days.[2]

In incremental 4DVAR, the dynamical model used to propagate the increment provides a time-dependent multivariate constraint on the analysis. Incremental 4DVAR is derived as an approximation to the complete 4DVAR problem in which the full nonlinear model (here an OGCM) is imposed as a constraint in the cost function with the model initial conditions taken as the control variables (Le Dimet and Talagrand 1986; Talagrand and Courtier 1987). In oceanography, most 4DVAR-related applications to date with OGCMs have concentrated on solving the complete problem directly, and in some cases using different or additional control variables (e.g., surface forcing fields; Tzipermann et al. 1992a,b; Greiner et al. 1998a,b, Greiner and Arnault 2000; Bonekamp et al. 2001). The incremental formulation was introduced in meteorology to overcome some important practical difficulties with solving the complete 4DVAR problem directly. In the latter, nonlinearities in the model constraint can significantly complicate the structure of the cost function and prevent its minimization at a

[1] FGAT was initially introduced in OI in the mid-1980s by D. Vasiljevic at the European Centre for Medium Range Weather Forecasts (ECMWF). The 40-yr atmospheric reanalysis project (ERA-40) at ECMWF employs an FGAT version of 3DVAR.

[2] Ten days is a typical window width used in OI-type ocean analysis systems such as the operational system at ECMWF (Alves et al. 2002). It is also the window width used for the 3DVAR experiments presented here.

reasonable computational cost using a gradient descent method. Furthermore, in order to compute a numerically accurate gradient, the adjoint of the *exact* TL of the full nonlinear model is required. When the nonlinear model contains discontinuous parameterizations or numerics, an accurate derivation of these models can be particularly difficult (Xu 1996).

The incremental algorithm should be viewed as a practical algorithm for approximately solving the complete problem. As constraints are linear, the cost function is quadratic and minimization with a gradient descent method is generally much more efficient. In the present study, our main objective is to reconstruct the large-scale, low-frequency component of the tropical ocean circulation, which is well known to be largely governed by linear wave dynamics (e.g., see Philander 1989). Therefore, a priori the linearity assumption in incremental 4DVAR would not appear to be a very restrictive one. More generally, however, nonlinear effects can be partly accounted for by introducing a feedback (outer loop) in the algorithm to update the basic state of the linear model with increments generated during minimization (the inner loop) (Courtier et al. 1994; Laroche and Gauthier 1998). In terms of technical development, the incremental formulation has a distinct advantage over the complete formulation as the derivation of the linear and adjoint models can be greatly simplified by smoothing or neglecting discontinuous parameterizations (Mahfouf 1999). For example, this latter point has been exploited in the present study to neglect changes in vertical diffusion coefficients associated with perturbations in temperature, salinity, and velocity. Finally, the computational cost of 4DVAR may also be significantly reduced by computing the increments at lower resolution than that of the full model (Rabier et al. 2000), although this is not an issue in our present system, which employs a relatively low resolution version of the OGCM.

The purpose of this paper is to give a thorough description of the current 3DVAR and 4DVAR systems and to study certain algorithmic and statistical aspects of the two systems. Validation of physical aspects of the analyses is given in Part II. The organization of the paper is as follows. Section 2 describes the general formulation of the incremental 3DVAR and 4DVAR problems. In section 3, the different system components are described. In section 4, several diagnostics are presented to highlight some important properties of the two systems. A summary is given in section 5.

2. Formulation of the 3DVAR and 4DVAR problems

a. Incremental formulation

The notation used in this paper closely follows the recommendations of Ide et al. (1997). Let $\mathbf{w}$ denote the ocean state vector. The components of $\mathbf{w}$ consist of those model variables that are to be estimated from observations to produce the analysis $\mathbf{w}^a$. A model forecast, initiated from a previous analysis, provides a prior or background estimate, $\mathbf{w}^b$, of $\mathbf{w}$. Since $\mathbf{w}^b$ will already be close in some sense to the ''true'' state we wish to estimate, it is convenient to formulate the estimation problem in terms of an increment, $\delta\mathbf{w}$, where

$$\mathbf{w} = \mathbf{w}^b + \delta\mathbf{w}. \tag{1}$$

The state vector is propagated in time by the ocean model:

$$\mathbf{w}(t_i) = M(t_i, t_{i-1})[\mathbf{w}(t_{i-1})], \tag{2}$$

where $M = M(t_i, t_{i-1})$ represents the nonlinear model operator acting on $\mathbf{w}(t_{i-1})$ between times t_{i-1} and t_i. Substituting (1) into (2) and expanding about $\mathbf{w}^b(t_{i-1})$ gives, to first order,

$$\mathbf{w}(t_i) \approx M(t_i, t_{i-1})[\mathbf{w}^b(t_{i-1})] + \mathbf{M}(t_i, t_{i-1})\delta\mathbf{w}(t_{i-1}), \tag{3}$$

where $\mathbf{M} = \mathbf{M}(t_i, t_{i-1})$ denotes a linear operator that acts on $\delta\mathbf{w}(t_{i-1})$ between times t_{i-1} and t_i. We define the prognostic model for the increment as

$$\delta\mathbf{w}(t_i) = \mathbf{M}(t_i, t_{i-1})\delta\mathbf{w}(t_{i-1}). \tag{4}$$

Now, let $\mathbf{y}_i^o$ denote the observation vector at time t_i. Denoting H_i as the observation operator at t_i, then the model equivalent of $\mathbf{y}_i^o$ can be written as

$$H_i[\mathbf{w}(t_i)] \approx H_i[\mathbf{w}^b(t_i)] + \mathbf{H}_i\delta\mathbf{w}(t_i), \tag{5}$$

where $\mathbf{H}_i$ is a linear operator that acts on the increment at t_i. Assuming that a time sequence of observation vectors is available over an interval $t_0 \le t_i \le t_n$, then the model estimates of the observations within this interval can be directly related to the model initial conditions since $\mathbf{w}(t_i) = M(t_i, t_0)[\mathbf{w}(t_0)]$, where $M(t_i, t_0) \equiv M(t_i, t_{i-1}) \circ \cdots \circ M(t_1, t_0)$. Thus,

$$H_i[\mathbf{w}(t_i)] = G_i[\mathbf{w}(t_0)] \approx G_i[\mathbf{w}^b(t_0)] + \mathbf{G}_i\delta\mathbf{w}(t_0), \tag{6}$$

where the combined operator $G_i = H_iM(t_i, t_0)$ is a generalized observation operator and $\mathbf{G}_i = \mathbf{H}_i\mathbf{M}(t_i, t_0)$ is its linearized counterpart with $\mathbf{M}(t_i, t_0) \equiv \mathbf{M}(t_i, t_{i-1}) \cdots \mathbf{M}(t_1, t_0)$.

In 4D variational assimilation, the analysis is defined as the state vector $\mathbf{w}^a = \mathbf{w}^a(t_0)$ that simultaneously minimizes the ''distance'' to the background state $\mathbf{w}^b = \mathbf{w}^b(t_0)$ and to the time sequence of observations $\mathbf{y}_i^o$ on $t_0 \le t_i \le t_n$. Distance is defined by an inner product (a cost function J^F) whose weighting metric takes into account the statistical accuracy of the background and observational information. Expressed as a function of the increment $\delta\mathbf{w} = \delta\mathbf{w}(t_0)$, which constitutes the control vector of the minimization problem, J^F may be written as

$$J^F(\delta\mathbf{w}) = \underbrace{\frac{1}{2}\delta\mathbf{w}^{\mathrm{T}}\mathbf{B}^{-1}\delta\mathbf{w}}_{J_b} + \underbrace{\frac{1}{2}\sum_{i=0}^{n}[G_i(\mathbf{w}^b + \delta\mathbf{w}) - \mathbf{y}_i^o]^{\mathrm{T}}\mathbf{R}_i^{-1}[G_i(\mathbf{w}^b + \delta\mathbf{w}) - \mathbf{y}_i^o]}_{J_o^F}, \quad (7)$$

where the matrices $\mathbf{B}$ and $\mathbf{R}_i$ contain estimates of the covariances of background and observation error, respectively. In (7) the observation errors are assumed to be uncorrelated in time and uncorrelated with the background error. The observation term (J_o^F) measures the fit between the observations and their model equivalent. The background term (J_b) penalizes the size of the increment (i.e., measures the fit to the background state). The analysis is given by $\mathbf{w}^a = \mathbf{w}^b + \delta\mathbf{w}^a$, where $\delta\mathbf{w}^a$ is the increment that minimizes J^F.

In the *incremental* formulation of variational assimilation (Courtier et al. 1994), (7) is approximated by a *quadratic* cost function J of $\delta\mathbf{w}$ by replacing $G_i(\mathbf{w}^b + \delta\mathbf{w})$ with its linearized counterpart (6). This results in an important practical simplification to the minimization problem, from one with potentially many minima due to the nonlinearity in G_i, to one with a unique minimum as guaranteed by the linearity of $\mathbf{G}_i$. The simplified cost function reads

$$J(\delta\mathbf{w}) = \underbrace{\frac{1}{2}\delta\mathbf{w}^{\mathrm{T}}\mathbf{B}^{-1}\delta\mathbf{w}}_{J_b} + \underbrace{\frac{1}{2}\sum_{i=0}^{n}(\mathbf{G}_i\delta\mathbf{w} - \mathbf{d}_i)^{\mathrm{T}}\mathbf{R}_i^{-1}(\mathbf{G}_i\delta\mathbf{w} - \mathbf{d}_i)}_{J_o}, \quad (8)$$

where the innovation vector

$$\mathbf{d}_i = \mathbf{y}_i^o - G_i(\mathbf{w}^b) = \mathbf{y}_i^o - H_i[\mathbf{w}^b(t_i)] \quad (9)$$

plays the role of an *effective* observation vector. Our 3DVAR (FGAT) and 4DVAR formulations differ principally in the choice of the linear operator $\mathbf{M}$ that is used in $\mathbf{G}_i$ to propagate the increment $\delta\mathbf{w}(t_i)$ in (4). In 3DVAR, the increment is evolved by a simple persistence model, which corresponds to setting $\mathbf{M} = \mathbf{I}$, the identity matrix. In 4DVAR, the increment is propagated by an approximate TL operator, $\mathbf{M} \approx (\partial M/\partial\mathbf{w})|_{\mathbf{w}=\mathbf{w}^b}$, the main approximation being introduced in the parametrization of vertical mixing as discussed in section 3b.

In practice, the cost function (8) is minimized approximately using an iterative gradient descent method. The increment is updated on each iteration using the gradient of the cost function with respect to the increment. The gradient of the J_o term with respect to the increment ($\nabla_{\delta\mathbf{w}}J_o$) can be obtained efficiently using the adjoint of the linear operator $\mathbf{G}_i$ (Le Dimet and Talagrand 1986; Thacker and Long 1988). A feedback between the linear and nonlinear models can then be introduced by allowing the basic-state trajectory of the linear model to be regularly updated with the most recent estimate of the state trajectory obtained during minimization (Courtier et al. 1994). The updates are performed on an *outer* loop of the assimilation algorithm, while the iterations of the actual minimization are performed within an *inner* loop [see appendix A in Weaver et al. (2002) for details on how the inner–outer loop algorithm is implemented in our current system]. With frequent updates, the accuracy of the linear model should improve and the minimum of the incremental cost function should be closer to that of the original (nonincremental) cost function (7) involving the nonlinear model (Laroche and Gauthier 1998). This provides a practical way of accounting for nonlinearities in the assimilation algorithm while retaining the computational advantages of a quadratic minimization problem.

In order to improve the convergence properties of the minimization, a preconditioning transformation is employed by which the cost function is redefined in terms of a nondimensional variable:

$$\mathbf{v} = \mathbf{U}^{-I}\delta\mathbf{w}, \quad (10)$$

where $\mathbf{U}$ is a rectangular matrix defined such that $\mathbf{B} = \mathbf{U}\mathbf{U}^{\mathrm{T}}$ and $\mathbf{B}^{-1} = (\mathbf{U}^{-I})^{\mathrm{T}}\mathbf{U}^{-I}$, the superscript $-I$ denoting generalized right inverse [i.e., $\mathbf{U}^{-I} = \mathbf{U}^{\mathrm{T}}(\mathbf{U}\mathbf{U}^{\mathrm{T}})^{-1}$]. The fact that $\mathbf{U}$ is rectangular, with $\dim(\delta\mathbf{w}) < \dim(\mathbf{v})$, is a feature that is particular to our current system and is related to the *rigid-lid* constraint used in the formulation of the ocean model. This point is discussed in more detail in section 3a and 3c. Introducing the transformation (10) directly into (8) leads to a simplified background term, $J_b = \mathbf{v}^{\mathrm{T}}\mathbf{v}/2$. The variational analysis problem is solved directly in $\mathbf{v}$ space, and then transformed back to model space using the generalized left inverse transformation:

$$\delta\mathbf{w} = \mathbf{U}\mathbf{v}. \quad (11)$$

The adjoint of (11) is used for computing the gradient of the J_o term in $\mathbf{v}$ space:

$$\nabla_{\mathbf{v}}J_o = \mathbf{U}^{\mathrm{T}}\nabla_{\delta\mathbf{w}}J_o. \quad (12)$$

The conditioning of the J_b term in $\mathbf{v}$ space is optimal in the sense that the Hessian of J_b (the matrix of second derivatives of J_b) is the identity matrix. For the special case of a single observation, the convergence of the minimization in $\mathbf{v}$ space is achieved in a single iteration using a gradient descent method with an exact line search.

The minimization routine used in this study is the

limited-memory quasi-Newton algorithm M1QN3 of Gilbert and Lemaréchal (1989). An exact line search has been employed with M1QN3 in order to improve the efficiency of the algorithm for quadratic minimization problems. The so-called warm start facility of M1QN3 is also employed when more than one outer iteration is performed in order to precondition the minimization using the information accumulated on the Hessian matrix during the preceding minimization.

b. The nonlinear analysis trajectory

The minimizing solution of the quadratic cost function (8) is a trajectory of increments $\delta\mathbf{w}^a(t_i)$ satisfying exactly the equations of the linear model (4) on $t_0 \leq t_i \leq t_n$. The corresponding model trajectory $\mathbf{w}^b(t_i) + \delta\mathbf{w}^a(t_i)$ is used in the J_o term to compare with the observations [Eq. (8)]. It will be convenient to refer to this trajectory as the *linear* analysis in order to distinguish it from the *nonlinear* ''analysis'' trajectory $\mathbf{w}^a(t_i)$ defined below. We will return to this point in section 4f and Part II of the paper.

There are several ways the analysis increment $\delta\mathbf{w}^a$ may be used to correct the trajectory of the nonlinear model. Two different approaches have been adopted in our 3DVAR and 4DVAR systems. Since our current assimilation system incorporates only temperature observations and relies on a univariate formulation of the background-error covariance matrix (section 3c), the 3DVAR produces an analysis increment for the temperature field only. A practical way of adjusting the nonanalyzed model fields while minimizing spurious adjustment processes is to apply the analysis increment gradually through a forcing term in the nonlinear model:

$$\mathbf{w}^a(t_i) = M(t_i, t_{i-1})[\mathbf{w}^a(t_{i-1})] + F(t_i)\delta\mathbf{w}^a, \quad (13)$$

where $\mathbf{w}^a(t_0) = \mathbf{w}^b$ and $F(t_i)$ is a weighting function defined such that $\Sigma_{i=0}^{n} F(t_i) = 1$ so as to conserve the time-integrated value of the analysis increment $\delta\mathbf{w}^a$. The forcing term can be shown to behave as a low-pass time filter (Bloom et al. 1996). In this study, the temperature increment is applied uniformly over the time window via a constant forcing $F(t_i) = 1/n$. A similar procedure has been adopted in the ECMWF ocean analysis system (Alves et al. 2002) and is illustrated schematically in Fig. 1a.

In contrast to the 3DVAR, the 4DVAR produces a multivariate analysis increment since the TL model dynamics act to couple the different increment variables. This dynamical coupling thus allows us to generate increments in velocity and salinity even if only temperature data are assimilated. Since these increments will be in approximate dynamical balance, in 4DVAR we choose to initialize the nonlinear model directly using the analysis at t_0:

$$\mathbf{w}^a(t_i) = M(t_i, t_0)[\mathbf{w}^a(t_0)], \quad (14)$$

where $\mathbf{w}^a(t_0) = \mathbf{w}^b + \delta\mathbf{w}^a$ (Fig. 1b). Finally, when cycling the 3DVAR or 4DVAR over an extended period, the analysis obtained from the trajectory at the end of the interval is taken to be the background state for a variational analysis performed on the following interval (Fig. 1).

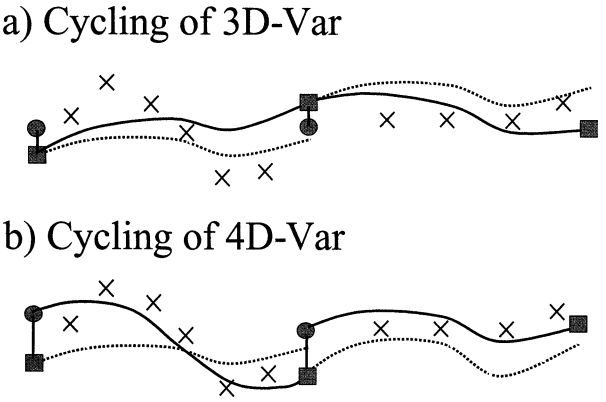

FIG. 1. Schematic representation of the cycling procedures used for (a) 3DVAR and (b) 4DVAR. Two cycles are illustrated in each sketch. The dotted (solid) curves correspond to the background (analysis) trajectory; the cross symbols denote observations. The shaded square (circle) at the beginning of each cycle denotes the background (analysed) initial state. Note that in 4DVAR the analysis increment (represented by the difference between the shaded square and circle) is applied directly to the background initial state to produce the analysis trajectory, whereas in 3DVAR it is applied gradually as a forcing to the model equations. This explains why in 3DVAR the analysis and background start from the same point at the beginning of each cycle.

3. Components of the assimilation system

a. The ocean model

The ocean model used in this study is the OPA OGCM of the Laboratoire d'Océanographie Dynamique et de Climatologie (Madec et al. 1998). The model solves the primitive equations for horizontal currents $\mathbf{u} = (u, v)$, potential temperature T_θ, and salinity S. The equations are formulated in orthogonal curvilinear z coordinates and discretized using finite differences on an Arakawa C grid. The basic configuration of the model is described in Vialard et al (2001). It covers the tropical Pacific Ocean from 30°N to 30°S, and 120°E to 70°W. The zonal resolution is 1° and the meridional resolution varies from 0.5° at the equator to 2° at the northern and southern boundaries. The model has 25 levels, with a resolution of 10 m in the upper 130 m, increasing to 1000 m in the bottom level. Realistic bathymetry is included using the Levitus (1982) land–sea mask.

At solid boundaries, conditions of no-slip and no-normal flux are applied on the velocity and tracer fields, respectively. At the ocean surface ($z = 0$), a rigid-lid and no-volume flux condition is applied (Roullet and Madec 2000). An obvious consequence of the rigid-lid condition is that, unlike in a free-surface model, there is no prognostic equation for sea surface height ($\eta =$

0). The no-volume flux condition is an additional requirement that the vertical velocity identically vanishes at the surface ($w = 0$ at $z = 0$) and, hence from the continuity equation, that the horizontal velocity field (u, v) is nondivergent (Bryan 1969). The latter condition is satisfied by defining the (u, v) field through a barotropic streamfunction ψ and a set of independent (subsurface) baroclinic velocities ($\hat{u}$, $\hat{v}$). This leads to some important technical difficulties in formulating the 4DVAR problem as detailed in appendix B of Weaver et al. (2002).

Surface fluxes of momentum, heat, and freshwater are prescribed at the ocean–atmosphere interface. The momentum flux is specified through weekly wind stress products from the European Remote Sensing (ERS) satellite's scatterometer (Grima et al. 1999). The heat and freshwater fluxes are specified as a daily climatology computed from the ECMWF (ERA-15) reanalysis (Gibson et al. 1997). The solar and nonsolar components of the heat flux are specified separately in order to allow penetration of the shortwave radiation in the upper ocean. A relaxation to weekly analyses of sea surface temperature (SST; Reynolds and Smith 1994) is applied through a Newtonian damping term added to the surface (nonsolar) heat flux. The relaxation coefficient is set to -40 W m^{-2} K^{-1}, which for a depth scale of 50 m corresponds to a restoring timescale of 1 month. No relaxation is applied to sea surface salinity. In the control integration only, in which no data are assimilated, a damping to Levitus climatological temperature and salinity is applied below the surface mixed layer outside the 10°S–10°N band.

b. The tangent-linear and adjoint models

The numerical codes of the TL and adjoint models have been derived directly from the numerical code of the nonlinear model by applying standard, hand-coding techniques (Talagrand 1991; Giering and Kaminski 1998). Some approximations have been introduced in the derivation of the TL and adjoint models, the most important one being in the parameterization of vertical diffusion, as described in more detail below. Another approximation relies on an intermittent storage of the basic-state trajectory to reduce computer memory requirements. In our experiments, the basic state has been stored once per day (every 16 time steps) and defined at intermediate times through linear interpolation. The impact of this approximation on the accuracy of the TL model was minor. Finally, to reduce the CPU cost of the adjoint integration, an approximation has been introduced in the adjoint of the elliptic solver that is applied on each time step to enforce the nondivergence constraint on the vertically integrated velocity. Details can be found in appendix B of Weaver et al. (2002).

Simplified vertical diffusion

Let α denote one of the prognostic state variables u, v, T_θ, or S. In the complete TL model, the tendency of a perturbation $\delta\alpha$ produced by vertical diffusion is described by the two-term partial differential equation:

$$\frac{\partial \delta\alpha}{\partial t} = \frac{\partial}{\partial z}\left(A_v \frac{\partial \delta\alpha}{\partial z}\right) + \frac{\partial}{\partial z}\left(\delta A_v \frac{\partial \alpha}{\partial z}\right), \qquad (15)$$

where $A_v = A_v(u, v, T_\theta, S)$ is the vertical diffusion coefficient and $\delta A_v = (\partial A_v/\partial u)\delta u + (\partial A_v/\partial v)\delta v + (\partial A_v/\partial T_\theta)\delta T_\theta + (\partial A_v/\partial S)\delta S$ is the perturbation of A_v resulting from perturbations of u, v, T_θ, and S. The first term on the right-hand side of (15) is a standard diffusion operator with coefficient A_v. In the incremental algorithm, A_v is updated on each outer iteration and held constant only during the inner iterations. The second term on the right-hand side of (15) is more problematic because of both theoretical and practical difficulties in computing the perturbation δA_v. First, a direct computation of δA_v would involve linearizing the turbulent kinetic energy (TKE) and enhanced vertical diffusion parameterization schemes used in OPA. These schemes are highly nonlinear and discontinuous so any attempt to linearize them directly would have to be done with considerable caution (Xu 1996; Zou 1997). Discontinuities are a well-known source of numerical noise in the TL model. One possibility for reducing such noise is to derive δA_v from a suitably smooth approximation to the original nonlinear scheme (Janiskova et al. 1999). Generally speaking, however, the problem is more complex than that of simply smoothing isolated discontinuities as the second term in (15) contains other generating mechanisms of spurious noise, which are present even in smoother parameterization schemes (Mahfouf 1999; Zhu and Kamachi 2000). The simplest way of avoiding these potential noise problems is to neglect the second term altogether by setting $\delta A_v \equiv 0$. This simplification, which has been used extensively in meteorology (e.g., Mahfouf 1999), is adopted here. Implementing a more elaborate linear physics parameterization was considered premature at this stage without first evaluating results using the simplified scheme.

c. The background-error covariance matrix

The background-error covariance matrix (**B**) plays an important role in determining the spatial structure of the analysis increment in both 3DVAR and 4DVAR. There are two basic difficulties in specifying **B**. First, given the sparsity of ocean observations, it is difficult, if not impossible, to obtain complete and accurate estimates of the covariances, even in the tropical Pacific, which is one of the best-observed ocean basins. Second, even if there were sufficient observations, the sheer size of **B** (roughly 5×10^{11} elements in our application) means that this matrix cannot be stored explicitly and so must be simplified. In practice, this is done by *modeling* the

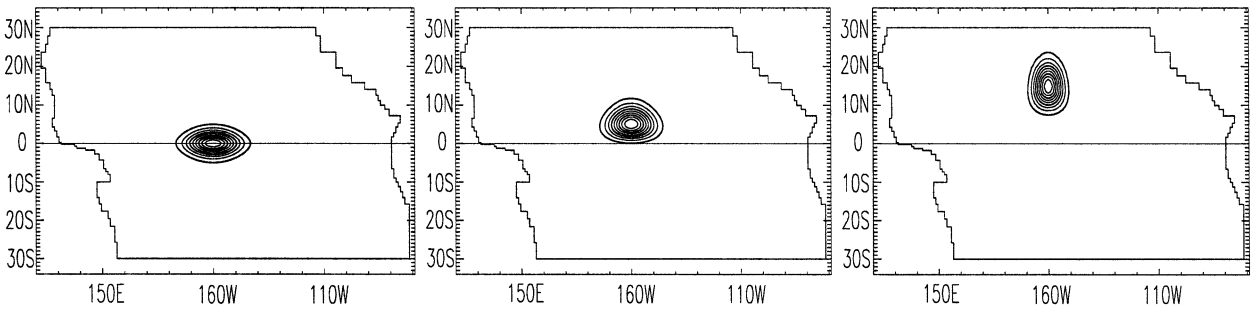

FIG. 2. The horizontal correlation functions for the tracer fields at latitudes of (a) 0°, (b) 5°N, and (c) 15°N. The contour interval is 0.1.

covariances using analytical functions or filters that depend on a limited number of tunable parameters and that are numerically efficient for large-scale problems.

In our current system, $\mathbf{B}$ is univariate (three-block diagonal with respect to horizontal velocity, temperature, and salinity) and includes a relatively simple correlation model. Here, $\mathbf{B}$ is constructed as a symmetric product of several operators:

$$\mathbf{B} = \underbrace{\mathbf{S}\boldsymbol{\Sigma}\boldsymbol{\Lambda}\mathbf{L}^{1/2}\mathbf{W}^{-1}\mathbf{L}^{T/2}\boldsymbol{\Lambda}\boldsymbol{\Sigma}\mathbf{S}^{T}}_{\mathbf{C}} \tag{16}$$

$$= \underbrace{(\mathbf{S}\boldsymbol{\Sigma}\boldsymbol{\Lambda}\mathbf{L}^{1/2}\mathbf{W}^{-1/2})}_{\mathbf{U}}\ \underbrace{(\mathbf{W}^{-1/2}\mathbf{L}^{T/2}\boldsymbol{\Lambda}\boldsymbol{\Sigma}\mathbf{S}^{T})}_{\mathbf{U}^{T}}, \tag{17}$$

where $\mathbf{L}$ is a 3D filtering operator that is *self-adjoint* with respect to the scalar product whose metric is the diagonal matrix $\mathbf{W}$ of volume elements; $\boldsymbol{\Lambda}$ and $\boldsymbol{\Sigma}$ are diagonal matrices of normalization factors and background-error standard deviations, respectively; and $\mathbf{S}$ is a simplification operator that maps the horizontal components of total velocity (u, v) into *independent* components $(\psi, \hat{u}, \hat{v})$ [for a detailed discussion see appendix B in Weaver et al. (2002)].

The underbraces in (17) highlight the preconditioning matrix $\mathbf{U}$ and its adjoint $\mathbf{U}^{T}$, which are needed in (11) and (12), respectively. Here, $\mathbf{U}$ is a rectangular matrix since the factor $\mathbf{S}$ is a mapping from a higher-dimensional space spanned by (u, v) into a lower-dimensional space spanned by $(\psi, \hat{u}, \hat{v})$. This explains why the generalized right inverse $\mathbf{U}^{-1}$ has been used in the transformation (10). The underbrace in (16) highlights that part of $\mathbf{B}$ corresponding to the correlation matrix $\mathbf{C}$. The univariate correlations in $\mathbf{C}$ are assumed to be approximately Gaussian and are modeled implicitly with the filter $\mathbf{L}$. The vertical (v) and horizontal (h) correlations are modeled separately using a 1D filter $\mathbf{L}_v$ and 2D filter $\mathbf{L}_h$. The 3D correlation model is then constructed from the product $\mathbf{L} = \mathbf{L}_v\mathbf{L}_h$. The diagonal normalization matrix $\boldsymbol{\Lambda}$ is needed to ensure that the variances (diagonal elements) of $\mathbf{C}$ are equal to unity. Various filters exist for modeling correlation functions but some are better suited than others depending on the application. The complex boundaries associated with coastlines imposes a particular constraint for oceanographic applications. For such applications, Laplacian- or diffusion-based filters are particularly well suited (Derber and Rosati 1989; Egbert et al. 1994; Weaver and Courtier 2001). Here, a diffusion-based filter has been used to model both the vertical and horizontal correlations; $\mathbf{L}_v$ is defined by an explicit time step integration of a 1D diffusion equation in the vertical direction, while $\mathbf{L}_h$ is defined by an explicit time step integration of a 2D diffusion equation over the sphere. The boundary conditions are chosen to be of Neumann type and are imposed directly within the finite-difference expression for the Laplacian using a land–ocean mask (Madec et al. 1998).

The correlation functions are made anisotropic and varied geographically by introducing a ''diffusion'' tensor in the Laplacian operator (Weaver and Courtier 2001). In particular, the tensor coefficients have been tuned to allow for longer correlation length scales near the equator in the zonal direction than in the meridional direction (Meyers et al. 1991; Kessler et al. 1996). Here, the horizontal length scales are taken to be a function of latitude and symmetric about the equator. The zonal and meridional length scales for the tracer fields have been defined to be 8° and 2°, respectively, at the equator, and 4° in both directions poleward of 20°N/S, with a linear transition between these values within the equatorial strip. The anisotropic ratio at the equator is consistent with the climatological observation statistics of Meyers et al. (1991), although the values of the actual length scales are somewhat smaller in our study. The values chosen here are broadly similar to those used in previous ocean data assimilation studies of the tropical Pacific (Smith et al. 1991; Behringer et al. 1998; Segschneider et al. 2001). The horizontal correlation functions for the tracer fields are illustrated in Fig. 2.

The horizontal correlations of the velocity field are modeled using a diffusion equation formed from the *vector* Laplacian rather than the scalar Laplacian used in the tracer diffusion model. The vector Laplacian has the desirable property of ensuring that the smoothing by the correlation function acts separately on the horizontal divergence and vorticity components of the velocity field (Madec et al. 1998). With the introduction of the anisotropic tensor, however, this is no longer

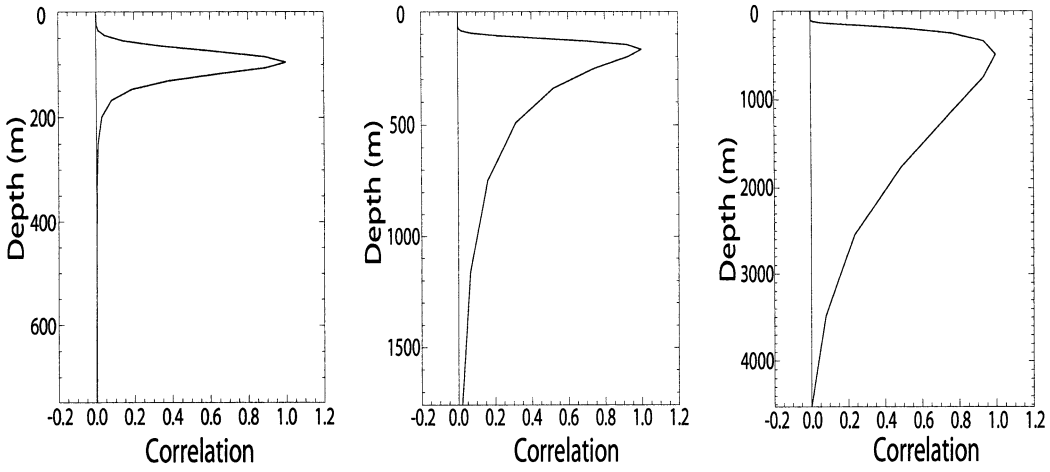

FIG. 3. The vertical correlation functions for the tracer fields at depths of (a) 96 (level 10), (b) 168 (level 15), and (c) 490 m (level 19). Note the different scales on the vertical axes.

strictly guaranteed. Geostrophy allows us to approximate the length scales of the vorticity correlations as $L_h^{\zeta} \approx 0.6L_h$, where L_h is the specified length scale of the tracer correlations (see Weaver et al. 2002). These smaller length scales have also been used for the divergence correlations. It is worth remarking that, while geostrophy has been used as a constraint for defining the length scales of the tracer and velocity correlation functions, our covariance model does not at present include a geostrophic balance constraint on the covariances themselves. Note also that the correlation model for the velocity background errors is not required in our univariate 3DVAR, which assimilates temperature information only.

The vertical correlations of the velocity field are modeled using a diffusion equation that acts separately on the components u and v. The vertical length scales for both the tracers and the velocity components are taken to be a function of depth with a dependence on the model's vertical resolution to provide adequate smoothing between model levels. At each model level, the vertical length scale is set to twice the depth of that level. This results in rather sharp correlations above the thermocline where the resolution is highest and much broader correlations below the thermocline where the resolution is coarsest (Fig. 3).

The background-error standard deviations are allowed to vary with each grid point and have been computed with respect to the climatological model mean obtained from a control run without data assimilation. This specification is based on the assumption that background errors are likely to be largest in regions of strong ocean variability (e.g., in the thermocline). The same background-error standard deviations are used at the beginning of each assimilation cycle.

d. Observations

The assimilation dataset consists of in situ temperature observations from the Global Temperature and Salinity Pilot Project (GTSPP) of the National Oceanographic Data Centre (NODC). This includes data from mainly TAO moorings and XBTs, and from a limited number of conductivity–temperature–depth (CTD) casts and drifting buoys. A manual quality control procedure was used to remove suspect data (Alves et al. 2002). Observations falling within the surface level of the model (between 0 and 10 m) were also discarded to avoid potential redundancy with the observed SST (Reynolds and Smith 1994) used in the Newtonian damping term during outer iterations. The in situ temperatures retained for assimilation were then converted into potential temperature (the prognostic model variable) using a standard conversion formula [Eq. (A3.13) in Gill (1982)] with a reference salinity of 35.0 psu.

The observation-error covariances are assumed to be uncorrelated in space and time. The error variances are set to $(0.5°C)^2$ for TAO data and $(1.0°C)^2$ for all other data to ensure that the generally higher quality TAO data have more weight in the analysis. The observation operator H consists of horizontal bilinear interpolation and vertical linear interpolation. For TAO data, H also includes a time averaging since these data are available as daily averages.

4. Evaluation of 3DVAR and 4DVAR: Internal diagnostics and consistency checks

The purpose of this section is to evaluate certain algorithmic and statistical properties of the 3DVAR and 4DVAR systems. The analyses themselves will be discussed in detail in Part II. First, in section 4a, the assimilation experiments are introduced. In section 4b, the validity of the linear assumption, which underlies the incremental formulation, is investigated. Convergence and optimality properties of the assimilation systems are then examined in sections 4c and 4d. In section 4e, some of the flow-dependent characteristics of the background-error statistics used implicitly in 4DVAR are examined

in a simplified framework with single observations. Finally, in section 4f, the statistics of the innovation and residual vectors are examined to assess the fit of the background and of the analysis to the assimilated observations.

a. Experimental setup

Both the 3DVAR and 4DVAR systems have been cycled over the period 1 January 1993–30 December 1998 using a 10- and 30-day assimilation window, respectively. A total of 60 (inner) iterations of the minimization were performed per cycle, with an outer iteration performed every 10 inner iterations in 4DVAR. No outer iterations were performed in 3DVAR since H_i is linear and $\mathbf{M} = \mathbf{I}$ is independent of the basic state. To provide a reference for evaluating 3DVAR and 4DVAR, an additional (control) experiment was performed in which no data were assimilated. In all experiments (hereafter referred to as EX3D, EX4D, and EXCL), the initial conditions on 1 January 1993 were generated from a 4-yr spinup of the model starting from rest and from Levitus (1982) climatological temperature and salinity. The wind stress forcing used for the first three years of the spinup was a climatology computed from the ERS wind stress products. The final year of the spinup was a transition year between ERS climatological and year 1992 products.

b. Validity of the linear increment models

Incremental variational assimilation is founded on the linear approximation (6). In this section, we wish to examine the validity of this approximation by checking the accuracy of both the persistence model used in 3DVAR on a 10-day window and the TL model used in 4DVAR on a 30-day window. The accuracy of each model can be assessed by comparing the time evolution of an initial perturbation $\delta\mathbf{w}$ in the nonlinear (*NL*) model with its evolution in the linear (*L*) model. The nonlinear perturbation can be computed from the finite difference,

$$\delta\mathbf{w}^{NL}(t_i) = M(t_i, t_0)(\mathbf{w}^b + \delta\mathbf{w}) - M(t_i, t_0)(\mathbf{w}^b), \quad (18)$$

while the linear perturbation is given by

$$\delta\mathbf{w}^{L}(t_i) = \mathbf{M}(t_i, t_0)\delta\mathbf{w}. \quad (19)$$

Ideally the initial perturbation $\delta\mathbf{w}$ should have structure and amplitude typical of background errors. The actual background errors are not known, however; so in order to apply this test a suitable proxy must be defined. In meteorology, it is commonly assumed that the background errors can be roughly approximated from differences in model forecasts initiated from analyses at different time lags (Rabier et al. 1998). Here we consider a similar approach in which an initial perturbation is derived from the difference between two model states obtained from free integrations of the model having different initial states. One of the initial states is taken to be the *analysis* state ($\mathbf{w}^a$) at the *end* of a 4DVAR assimilation interval [as defined by (14)], while the other initial state is taken to be the background state ($\mathbf{w}^b$) valid at the same time (see Fig. 1b). In the first example presented below, $\mathbf{w}^a$ and $\mathbf{w}^b$ are taken from the end of the second 30-day cycle (on 1 March 1993) of EX4D, with $\delta\mathbf{w} = \mathbf{w}^a - \mathbf{w}^b$ defining the initial perturbation in (18) and (19). The validity of the linear model was also investigated for other types of perturbations (analysis increments, differences between analyzed states at two different dates) and for other starting dates, and the results were qualitatively similar to those discussed below.

A meridional–vertical section of the perturbation temperature at 110°W is shown in Fig. 4a. The field is characterized by large perturbations of up to 4°C, appearing as a result of the temperature observations, which are assimilated during the second cycle. The perturbations after 10 and 30 days of integration in the nonlinear model are shown in Figs. 4b and 4c, respectively. Figure 4b allows us to check the assumption in 3DVAR that perturbations do not evolve significantly over 10 days outside the equatorial strip. The large-amplitude positive anomaly near 12°N and the smaller-amplitude negative anomalies near 16°N and 7°S are indeed well approximated by the 10-day persistence model (cf. Figs. 4a and 4b). The persistence assumption breaks down nearer the equator where the oceanic dynamical response is much more rapid. In the persistence model, the positive anomaly at 4°N is largely overestimated, while the negative anomaly at 7°N is largely underestimated. In contrast, Figs. 4c and 4d illustrate that the TL model is able to provide a good description of the perturbation in the equatorial waveguide at least 30 days ahead. The similarity of Figs. 4c and 4d may come as little surprise since it is well known that the large spatial scale, low-frequency response of the tropical oceans involves predominantly linear wave dynamics (Philander 1989).

At smaller spatial and temporal scales (<1000 km, <1 month), energetic motions in the central and eastern equatorial Pacific are dominated by tropical instability waves (TIWs) (Legeckis 1977). Because of the importance of nonlinear processes in ultimately limiting the growth of unstable waves, it is interesting to see if some of the differences between perturbations in the TL and nonlinear models can be linked to TIWs. TIWs are most energetic during the autumn months. If TIW activity is indeed a limiting factor of the TL approximation, then one would expect this approximation to be less valid during this time. To check this, the above experiment was repeated using a starting date of 27 September 1993 and a temperature perturbation defined as the difference between the 4DVAR analysis and background state at that time (i.e., at the end of the ninth 30-day cycle of EX4D). A zonal–vertical section of the difference at 4°N between the nonlinear and TL model-predicted temperature perturbations after 30 days is shown in Fig. 5.

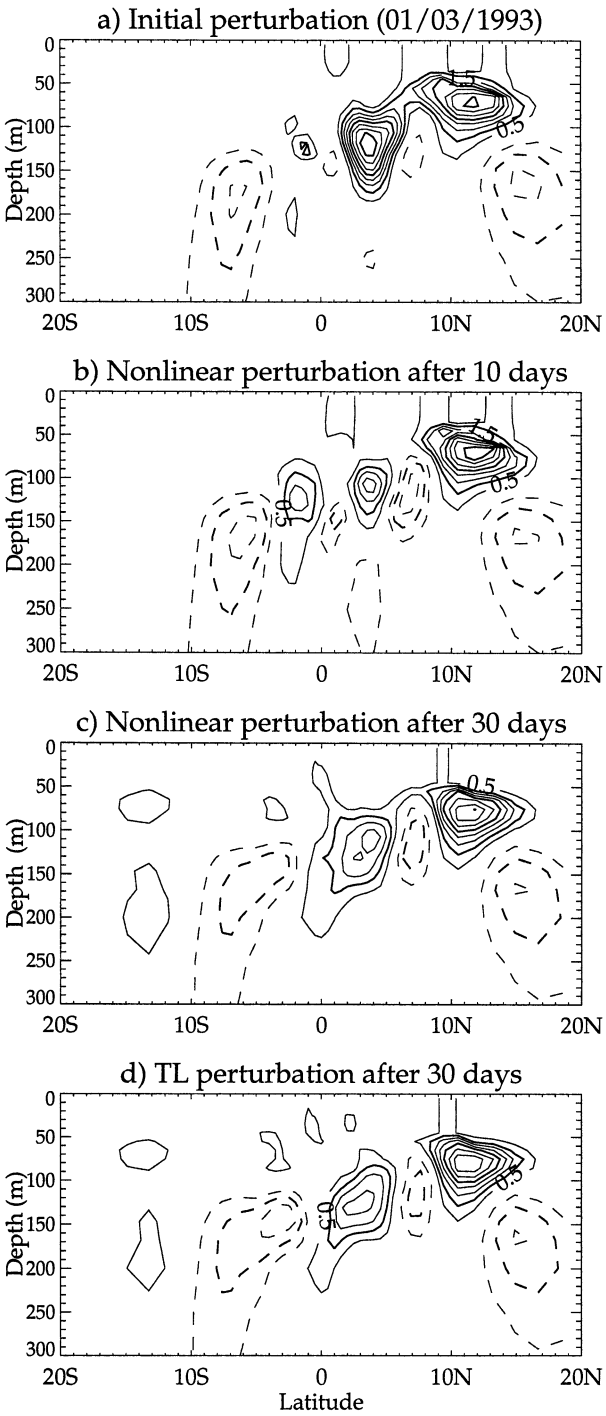

FIG. 4. (a) Meridional–vertical section at 110°W of a temperature perturbation defined as the difference between a 4DVAR analysis and the background state valid at the same time (1 Mar 1993). The perturbation after (b) 10 and (c) 30 days of evolution in the nonlinear model, and after (d) 30 days of evolution in the TL model. The contour interval is 0.5°C.

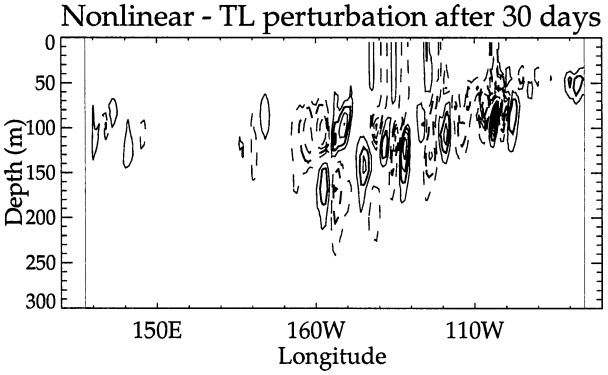

FIG. 5. Zonal–vertical section at 4°N of the difference between temperature perturbations in the nonlinear and TL models after 30 days. The initial perturbation in both experiments is defined as the difference between the 4DVAR analysis and the corresponding background state on 27 Sep 1993. The contour interval is 0.5°C.

West of 160°W, the TL approximation holds quite well. Most of the nonlinear behavior is located in the upper thermocline in the eastern and central Pacific, which is the area of largest thermal signal from TIWs. The differences between the nonlinear and linear perturbations are mostly associated with larger amplitudes of the linear perturbation. This confirms that nonlinear mechanisms play an important role in limiting the amplitude of the TIWs, and that the TL approximation is limited at their spatial scale. Over longer integration periods, the TIWs show up clearly as the dominant error signal (not shown), with the amplitude of the perturbations tending to be greatly overestimated in the TL model. This limitation is, however, only present at spatial scales of order 1000 km and during the active TIW season.

Strongly nonlinear processes associated with vertical mixing and convection are another factor contributing to the limitation of the TL approximation, particularly in the small vertical scales. For example, when a perturbation leads to a statically unstable water column in the nonlinear model, the water column will be mixed instantaneously as a result of enhanced vertical diffusion. This process will effectively reduce the amplitude of the perturbation in the nonlinear model. In the TL model, on the other hand, there is no linearized counterpart of this process and the perturbation will remain unchanged. This could explain the differences in the mixed layer seen in Fig. 5. Weaver et al. (2002) present additional diagnostics that provide further confirmation of these points.

In summary, the results from this section indicate that the 10-day persistence model used in 3DVAR is adequate for describing large-scale perturbations in off-equatorial regions but has some limitations closer to the equator where the dynamical adjustment timescales are shorter. In comparison, the 30-day TL model provides a good description of large-scale perturbations in both off-equatorial and equatorial regions. Smaller-scale, regionally confined structures associated with TIWs, how-

ever, are more problematic and our experiments suggest that they are possibly the main source of error in the TL integration. Vertical mixing and convective processes also tend to limit the accuracy of the TL model in the small vertical scales. Further research is required to determine whether the accuracy of the TL model can be improved using a more sophisticated parametrization of vertical diffusion, in particular one that accounts for (nonzero) perturbations in the vertical mixing coefficients. However, it will be shown in section 4f and in Part II that these limitations on the validity of the TL model are much less severe than appears at first sight and that in fact a very good fit to the data is possible.

c. Convergence properties

Each 10-day 3DVAR cycle required roughly 25 min of CPU time on a Fujitsu VPP700, while each 30-day 4DVAR cycle required roughly 5 h of CPU time (i.e., a factor of 4 more costly than 3DVAR). The convergence of the minimization in 3DVAR was relatively quick, requiring on average less than 25 iterations to reach the effective minimum of J. The convergence of the minimization in 4DVAR was generally slower than in 3DVAR. After the final (60th) iteration, the reduction of the norm of $\nabla_v J$ relative to its initial value was typically between three and four orders of magnitude compared to over six orders of magnitude in 3DVAR.

As a typical example, Fig. 6 (solid curve) shows the value of J from the second cycle of EX4D. The jumps in the solid curves in Fig. 6 occur after an outer iteration when the reference trajectory is reinitialized (here every 10 iterations). To illustrate the value of these outer iterations, the 4DVAR minimization on this particular cycle was repeated with only one outer iteration, that is, 60 minimization iterations with no trajectory updates (dashed curve in Fig. 6). Although it was found that convergence was more efficient without than with outer iterations, with a gain of approximately an order of magnitude in the reduction of the gradient norm, the reduction of the cost function in the experiment performed without outer iterations saturates at a level above the cost function with outer iterations. One has to be careful, however, in making direct comparisons between these two curves since, after the first outer iteration, the cost functions are no longer the same in the two experiments. It is more instructive to compare the final value of the full (nonapproximated) cost function J^F [Eq. (7)] for each experiment. These values are plotted in Fig. 6 with an asterisk (plus sign) for the experiment with (without) outer iterations. This figure provides a clear indication of the positive impact of the outer iterations. The final value of J^F with outer iterations is about half that without outer iterations. Furthermore, it is very close to the final value obtained with the approximated (quadratic) cost function (solid curve) and thus provides a good measure of the consistency of the incremental approach. In the absence of outer iterations, however, this consistency is lost as illustrated by the large discrepancy between the final value of J^F and that of the approximated cost function J (dashed curve).

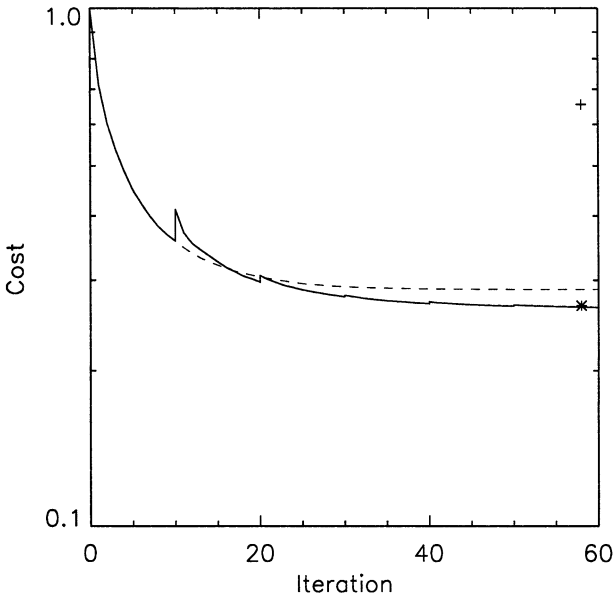

FIG. 6. The value of the cost function (J) as a function of the minimization iteration on the second cycle of EX4D (solid curve). As a sensitivity test, the second cycle of the 4DVAR experiment has been repeated with only one outer iteration (dashed curve), compared to five outer iterations used in the reference experiment. The plus sign (asterisk) indicates the final value of the nonapproximated cost function (7) for the experiment with one (five) outer iteration(s). For clarity, these symbols have been displaced slightly to the left of the right border; J has been normalized by its respective value at the start of the minimization and plotted on a logarithmic vertical axis.

In 3DVAR, 60 iterations was more than double the number actually needed to reach an acceptable level of convergence. Despite this potential to economize in 3DVAR, 60 iterations were retained simply to be consistent with the total number of iterations used in 4DVAR. In 4DVAR, convergence was slower but still reasonably efficient. Preconditioning techniques such as those described in Fisher and Andersson (2001) offer considerable scope for further improving the minimization efficiency. The outer iterations were clearly shown to be an essential feature of the algorithm. At present, however, it is not clear how many outer iterations are needed for the solution of the incremental problem to be a good approximation to the solution of the full problem. It is also not clear what combination of outer and inner iterations gives the best convergence rate. These issues are a matter for further research.

d. Optimality properties

The formulation of the variational assimilation problem relies on a number of hypotheses on the statistics of the background and observation errors. The validity of these hypotheses is an important factor in determining

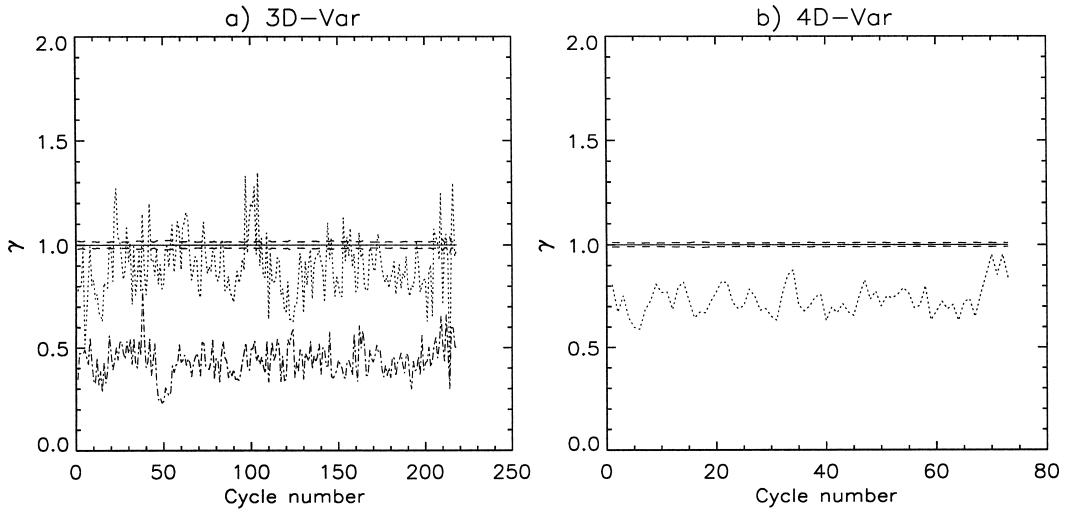

FIG. 7. The value of $\gamma = 2J_{\min}/p$ plotted as a function of the assimilation cycle during the period 1993–98 in (a) 3DVAR and (b) 4DVAR. On each cycle, $J_{\min}$ represents the value of the cost function at the end of the minimization and p the total number of observations assimilated through the J_o term. A total of 219 (73) 10-day (30 day) cycles were performed in 3DVAR (4DVAR). The *expected* value of $\gamma = 1$ (solid line) is plotted together with error bars (dashed lines) at $1 \pm \sigma_\gamma$, where $\sigma_\gamma = \sqrt{2/p}$ is the *expected* standard deviation of γ. The dotted curves in (a) and (b) correspond to the *actual* values of γ computed on each cycle in the reference experiments EX3D and EX4D. The dashed–dotted curve in (a) corresponds to the *actual* values of γ from a 3DVAR experiment in which the observation-error variance for the TAO data is set to twice the value used in EX3D.

the optimality of the analysis. One particularly simple diagnostic that can be used to check whether the statistics of **B** and **R** are consistent with the innovation vector is the value of the cost function at its minimum ($J_{\min}$), which, for a linear system, should, on statistical average, be equal to $p/2$ where p is the total number of assimilated observations (Tarantola 1987; Bennett et al. 2000). If we assume further that the background and observation errors are Gaussian then it can be shown (e.g., see section 4.3.6 in Tarantola 1987) that the expected variance of $J_{\min}$ is $\sigma^2_{J_{\min}} = E[(J_{\min} - E[J_{\min}])^2] = p/2$, where $E[\,]$ is the mathematical expectation operator. Therefore, provided **B** and **R** are correctly specified and that the system is quasi-linear, the value of $J_{\min}$ on each assimilation cycle should be $p/2$ within a standard deviation of $\sqrt{p/2}$.

In order to compare the actual values of $J_{\min}$ with its expected value, it is more convenient to consider the normalized quantity $\gamma = 2J_{\min}/p$ for which $E[\gamma] = 1$ and $\sigma^2_\gamma = E[(\gamma - E[\gamma])^2] = 2/p$. Here, σ^2_γ scales with the inverse of p so will be small when a large number of observations are assimilated. Figures 7a and 7b (dotted curves) show the actual values of γ as a function of the assimilation cycle in EX3D and EX4D. The expected value of $\gamma = 1$ (solid line) is also plotted together with the expected error bars (dashed lines) at $1 \pm \sigma_\gamma$. The average of the actual values of γ over all cycles in EX3D is close to its expected value of one ($\overline{\gamma} = 0.90$, where the overbar denotes cycle average) but in EX4D it is somewhat too small ($\overline{\gamma} = 0.73$). On average, the actual values of γ exceed its expected value by $9\sigma_\gamma$ in EX3D and $29\sigma_\gamma$ in EX4D. The expected error bars are slightly larger in EX3D than in EX4D since there were fewer observations assimilated during a 10-day 3DVAR cycle than during a 30-day 4DVAR cycle (roughly 10 000 observations were available every 10 days). Even so, the actual values of γ nearly always greatly exceed the expected error bounds in both experiments.

Consider the expression for $J_{\min}$ in terms of the innovation vector **d** (Tarantola 1987):

$$J_{\min} = \frac{1}{2}\mathbf{d}^{\mathrm{T}}(\mathbf{GBG}^{\mathrm{T}} + \mathbf{R})^{-1}\mathbf{d}, \tag{20}$$

where the generalized quantities $\mathbf{d} = (\ldots, \mathbf{d}_i^{\mathrm{T}}, \ldots)^{\mathrm{T}}$, $\mathbf{G} = (\ldots, \mathbf{G}_i^{\mathrm{T}}, \ldots)^{\mathrm{T}}$, and $\mathbf{R} = \mathrm{diag}(\ldots, \mathbf{R}_i, \ldots)$ have been introduced to simplify the notation. From (20) it is clear that the value of $J_{\min}$, or equivalently γ, may be decreased (increased) if the variances of either **B** or **R** are increased (decreased). The lower than expected value of γ in EX4D therefore could be a sign that either the background- or observation-error variances have been overestimated. Our prior estimate of the observation-error variances is very crude and in particular takes no account of possible representativeness error. For example, the value of 0.5°C used for the standard error of TAO data is only slightly larger than the documented estimate of TAO observation error, which accounts only for the instrumental component of that error (McCarty et al. 1997). The specified observation-error variances are thus probably underestimated and therefore acting to increase the value of γ. This effect is clearly demonstrated by the dashed–dotted curve in Fig. 7a, which shows the actual values of γ from a 3DVAR

experiment in which the standard error for TAO data was increased to 1.0°C. This led to a factor of 2 decrease in the cycle average of $\gamma(\overline{\gamma} = 0.44)$.

It is more likely that the background-error variances, which have been specified from the climatology of a model integration performed without data assimilation, are substantially overestimated. Whereas this specification might yield a reasonable estimate of the background errors for the first cycle, it is probably too large an estimate in later cycles, particularly in well-observed regions, where the data assimilated in previous cycles have acted to reduce the innovation vector. Generally speaking, however, it is safer to overestimate than underestimate the background-error variances to prevent the model from drifting too far from observations. It should be noted that a misspecification of the background- and/or observation-error correlations could also change the value of γ.

Another interesting feature in Fig. 7 is the rather large variance in γ between cycles. This is an indication that, contrary to what has been assumed, the background-error covariances are not stationary, in addition to being largely overestimated as already mentioned. From Fig. 7a, we see that the variability of γ is larger in EX3D than in the sensitivity experiment where the standard error used for TAO data was larger. This is understandable since the smaller standard error for TAO in EX3D means that γ will be more sensitive to the true variations in the background-error covariances between cycles.

While the $J_{\min}$ statistic gives some useful insight into the optimality properties of the system, it is not possible based on this information alone to correct unambiguously any misspecification of the background- and/or observation-error covariances. If the actual value of $J_{\min}$ is not equal to $p/2$, then this can be rectified simply by muliplying $\mathbf{B}$ and $\mathbf{R}$ by the factor $\gamma = 2J_{\min}/p$. This procedure will change the *absolute* values of the background-, observation-, and analysis-error variances, but will have no influence on the analysis increment itself. What is important then is the *relative* magnitude of the variances in $\mathbf{B}$ and $\mathbf{R}$ (the absolute values can be obtained by postmultiplication by γ). Adaptive procedures can be used to tune the variances in $\mathbf{B}$ and $\mathbf{R}$ using information on the mismatch between the expected and actual values of *subparts* of the cost function at its minimum (Desroziers and Ivanov 2001). To compute the expected minimum value of subparts of the cost function is more complicated, but practical methods do exist (Bennett et al. 2000; Desroziers and Ivanov 2001). We have not attempted to apply any of these methods in the present study but they do offer an interesting possibility for improving the variance estimates in the future.

e. Flow-dependent background-error variances

It is well known that, in the limit of a perfect, linear model, variational assimilation is equivalent to the Kalman filter in that, given identical input parameters, they produce the same analysis at the end of the assimilation window (e.g., see Courtier et al. 1994). Consider an assimilation window $t_0 \le t_i \le t_n$ in which a background state $\mathbf{w}^b$, with error-covariance matrix $\mathbf{B}$, is available at the beginning of the assimilation window and an observation vector $\mathbf{y}_n^o$, with error-covariance matrix $\mathbf{R}_n$, is available at the *end* of the window. Within the linear approximation, the error-covariance matrix $\mathbf{P}^b(t_n)$ for the background state $\mathbf{w}^b(t_n) = M(t_n, t_0)(\mathbf{w}^b)$ is obtained by evolving $\mathbf{B}$ using the linear model and its adjoint:

$$\mathbf{P}^b(t_n) = \mathbf{M}(t_n, t_0)\mathbf{B}\mathbf{M}(t_n, t_0)^{\mathrm{T}}, \tag{21}$$

where $\mathbf{P}^b(t_0) = \mathbf{B}$. In an extended Kalman filter, (21) would be used explicitly to transport the covariances forward in time. In incremental variational assimilation, on the other hand, this propagation is implicit in the global minimization process. For the example above, the variational analysis increment at t_0, obtained by minimizing (8) exactly, can be written as

$$\delta\mathbf{w}^a = \mathbf{B}\mathbf{M}(t_n, t_0)^{\mathrm{T}}\mathbf{H}_n^{\mathrm{T}}[\mathbf{H}_n\mathbf{M}(t_n, t_0)\mathbf{B}\mathbf{M}(t_n, t_0)^{\mathrm{T}} \times \mathbf{H}_n^{\mathrm{T}} + \mathbf{R}_n]^{-1}\mathbf{d}_n, \tag{22}$$

where $\mathbf{d}_n = \mathbf{y}_n^o - H_n[\mathbf{w}^b(t_n)]$. The analysis increment at t_n can be obtained by applying $\mathbf{M}(t_n, t_0)$ to both sides of (22) to yield, after inserting (21),

$$\delta\mathbf{w}^a(t_n) = \mathbf{P}^b(t_n)\mathbf{H}_n^{\mathrm{T}}[\mathbf{H}_n\mathbf{P}^b(t_n)\mathbf{H}_n^{\mathrm{T}} + \mathbf{R}_n]^{-1}\mathbf{d}_n, \tag{23}$$

where $\delta\mathbf{w}^a(t_n) = \mathbf{M}(t_n, t_0)\delta\mathbf{w}^a$. Equation (23) is the standard analysis step of an extended Kalman filter, which weights the background state at t_n using an error-covariance matrix predicted by (21).

Our FGAT version of 3DVAR can be viewed as a limiting case in which $\mathbf{M}$ is taken to be the identity matrix. This implies that $\mathbf{P}^b(t_n) = \mathbf{B}$, that is, that the background-error covariances at the end of the window (and at all intermediate times) are identical to those specified at the initial time. In incremental 4DVAR, on the other hand, $\mathbf{M}$ is the TL operator so that $\mathbf{P}^b(t_n)$ will be modified by dynamical processes acting within the window. In this section, we focus on how the TL dynamics act to modify the prior estimates of the background-error *variances* [the diagonal elements of $\mathbf{P}^b(t_n)$]. The TL dynamics will also modify the background-error correlations but a discussion of this feature is beyond the scope of the present paper.

Single-observation experiments provide a practical way of computing the effective background-error variances used implicitly in 4DVAR (Thépaut et al. 1993, 1996). For a single observation, $d = \mathbf{d}_n$ and $(\sigma^o)^2 = \mathbf{R}_n$ become scalars, and $\mathbf{h}^{\mathrm{T}} = \mathbf{H}_n$ becomes a vector of the same length as $\mathbf{w}$. The error variance of the background equivalent of the observation is then given by the scalar product $(\sigma_n^b)^2 = \mathbf{h}^{\mathrm{T}}\mathbf{P}^b(t_n)\mathbf{h}$. If we assume further that the observation is of one of the model state variables and that its location coincides with a model grid point, then $\mathbf{h}^{\mathrm{T}} = (0, \ldots, 0, 1, 0, \ldots 0)$, where the

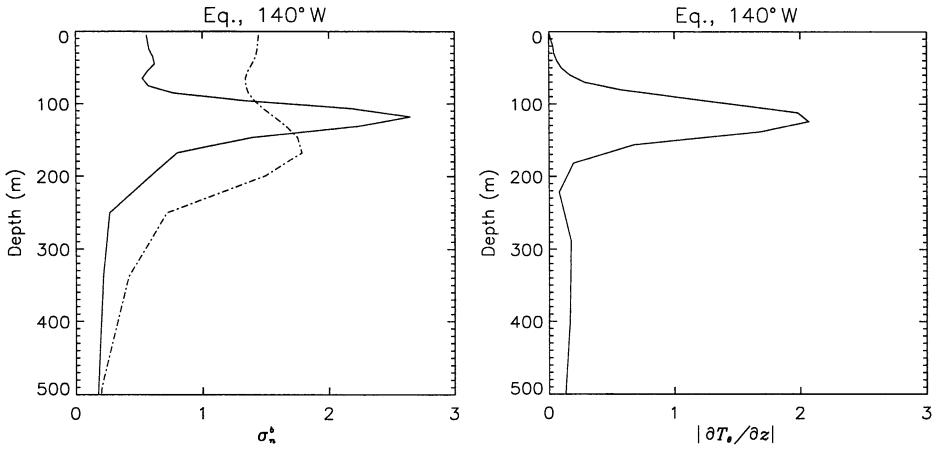

FIG. 8. (a) Vertical profiles of the background-error std devs (°C) used in 3DVAR and 4DVAR at the equator at 140°W. The dashed–dotted curves correspond to the standard deviations *specified* at the *beginning* of the assimilation window. In 3DVAR, these are also the *effective* standard deviations used at all future times within the window. The solid curves correspond to the *effective* standard deviations used in 4DVAR at the *end* of the 30-day window of the second cycle in EX4D. (b) The corresponding profile of $|\partial T_\theta/\partial z|$ at this location, computed from the 30-day mean of the background temperature state on the second cycle. The values of $|\partial T_\theta/\partial z|$ have been multiplied by a factor of 10 in order to be plotted with the same horizontal scale as in (a).

entry equal to one corresponds to that grid point, and $(\sigma_n^b)^2$ becomes an element of the diagonal of $\mathbf{P}^b(t_n)$. It is straightforward to deduce $(\sigma_n^b)^2$ by applying $\mathbf{h}^T$ to both sides of (23):

$$\mathbf{h}^T\delta\mathbf{w}^a(t_n) = (\sigma_n^b)^2\left[\frac{d}{(\sigma_n^b)^2 + (\sigma^o)^2}\right], \qquad (24)$$

which can be rearranged to give

$$(\sigma_n^b)^2 = \left[\frac{\mathbf{h}^T\delta\mathbf{w}^a(t_n)}{d - \mathbf{h}^T\delta\mathbf{w}^a(t_n)}\right](\sigma^o)^2. \qquad (25)$$

The term $\mathbf{h}^T\delta\mathbf{w}^a(t_n)$ is the value of the analysis increment at the time and gridpoint location of the single observation. Equation (25) thus provides the basis of an algorithm for systematically diagnosing the exact diagonal elements of $\mathbf{P}^b(t_n)$. To compute all diagonal elements of $\mathbf{P}^b(t_n)$ using (25) would be prohibitively expensive since (25) requires as many single observation experiments as the number of elements of $\mathbf{w}$. This algorithm is therefore only practical for computing a small subset of the σ_n^b, which is what is desired here. If an estimate of the complete diagonal of $\mathbf{P}^b(t_n)$ were required, then a more efficient, but approximate, algorithm based on randomization would be more appropriate (e.g., Andersson et al. 2000).

Figure 8 shows vertical profiles of the background-error standard deviations (σ_n^b) for temperature at the equator in the central Pacific (140°W). The dashed–dotted curve is the prior estimate of σ_n^b, which, as discussed earlier, was computed from a control experiment without data assimilation. In 3DVAR, the prior estimates are effectively used to weight the background state at all times within the assimilation window, while in 4DVAR they are used for weighting the background state at the *beginning* of the assimilation window. The prior estimate displays a maximum around the depth of the climatological thermocline where variability is greatest (around 170 m). The solid curve shows the profile of the effective σ_n^b used in 4DVAR at the end of the second 30-day cycle (3 March 1993) of EX4D (see Part II for a thorough description of this experiment). By repeating this experiment at different latitudes it was found that the flow dependency in the σ_n^b estimates is greatest near the equator, which is not surprising since dynamical effects have shorter timescales there. In 30 days, the ocean state can change significantly at the equator but much less so at higher latitudes. Figure 8a also shows that the TL dynamical processes tend to reduce the σ_n^b in the ocean mixed layer, particularly in the upper 70 m. Note that the Newtonian relaxation term for SST is contributing to this tendency by damping temperature perturbations in the top level of the TL model.

The TL dynamics also tend to move the level of maximum background-error variance to the level of the thermocline, as illustrated by comparing Fig. 8a with the profile of $|\partial T_\theta/\partial z|$ computed from the 30-day-averaged background temperature state on the second cycle. This tendency is physically sensible since the level of maximum variability of the thermal field, and thus of maximum likely error in the background state, is located at the level of the thermocline. It can also be noted that the TL dynamics can substantially increase the maximum value of the background-error variance near the equator. This is related to the fact that the background trajectory in this experiment has already felt the influence of observations assimilated during the previous

cycle through a tightening of the thermocline. This tighter, well-defined thermocline, relative to that of the control, leads to larger thermal signals and thus to an increase in the maximum value of σ_n^b by the TL dynamics. In some 3D assimilation systems (e.g., Behringer et al. 1998; Alves et al. 2002), the background-error variances have been parameterized by making them dependent on the vertical gradient of the background temperature field. Figure 8 suggests that relating σ^b to the background temperature gradient is dynamically sensible. Such a weighting could also be useful in 3DVAR and 4DVAR by introducing a flow dependence in the variances of **B** at the beginning of the window.

f. Fit to the observations

Statistics derived from the background minus observation vector $\{\text{BmO} \equiv H_i[\mathbf{w}^b(t_i)] - \mathbf{y}_i^o\}$ and analysis minus observation vector $\{\text{AmO} \equiv H_i[\mathbf{w}^a(t_i)] - \mathbf{y}_i^o\}$ can yield useful information about the internal consistency of the data assimilation system (Hollingsworth and Lönnberg 1989). Figures 9a and 9b show the 1993–98 averaged statistics of the BmO and AmO as a function of depth for all the assimilated TAO data in EX3D and EX4D. For reference, the average statistics of the difference between the control and TAO data are also shown, which for convenience will be referred to as the BmO of EXCL. In Fig. 9a the BmO curve of EXCL displays a large warm bias of about 2°C just below the thermocline. This bias is largely reduced in EX3D, with the maximum of the averaged BmO and AmO being about 0.3°C at 50 m. The bias is almost completely absent from EX4D: the average of AmO is very small, except at 250 m where the analysis is about 0.1°C too cold.

In Fig. 9b the rms of BmO of EXCL shows large differences both just below the thermocline, where there are large biases, and in the thermocline, where signals associated with the seasonal cycle and interannual variability are largest. These differences are substantially reduced in EX3D and EX4D. The rms of BmO in EX4D is similar to that of EX3D, with a maximum around 1.5°C at the level of the thermocline. On the other hand, the rms of AmO in EX4D is very much reduced, being less than 0.5°C over the entire water column, compared to the rms of AmO in EX3D, which is hardly smaller than that of BmO. The fit to the TAO data in EX4D is thus within the specified level of the observation error (0.5°C), which is not the case in EX3D. The larger AmO in EX3D is primarily an artifact of the gradual way the analysis increment is applied to the model. While minimizing nonphysical adjustment processes, this procedure ultimately degrades the fit to the data achieved by the linear analysis $\mathbf{w}^b(t_i) + \delta\mathbf{w}^a(t_i)$. This is illustrated in Fig. 9c, which shows that for EX3D the rms of the J_o residual, $\mathbf{H}_i\delta\mathbf{w}^a(t_i) - \mathbf{d}_i$, is considerably less than the AmO and is comparable to the specified observation error (0.5°C). An additional experiment was performed

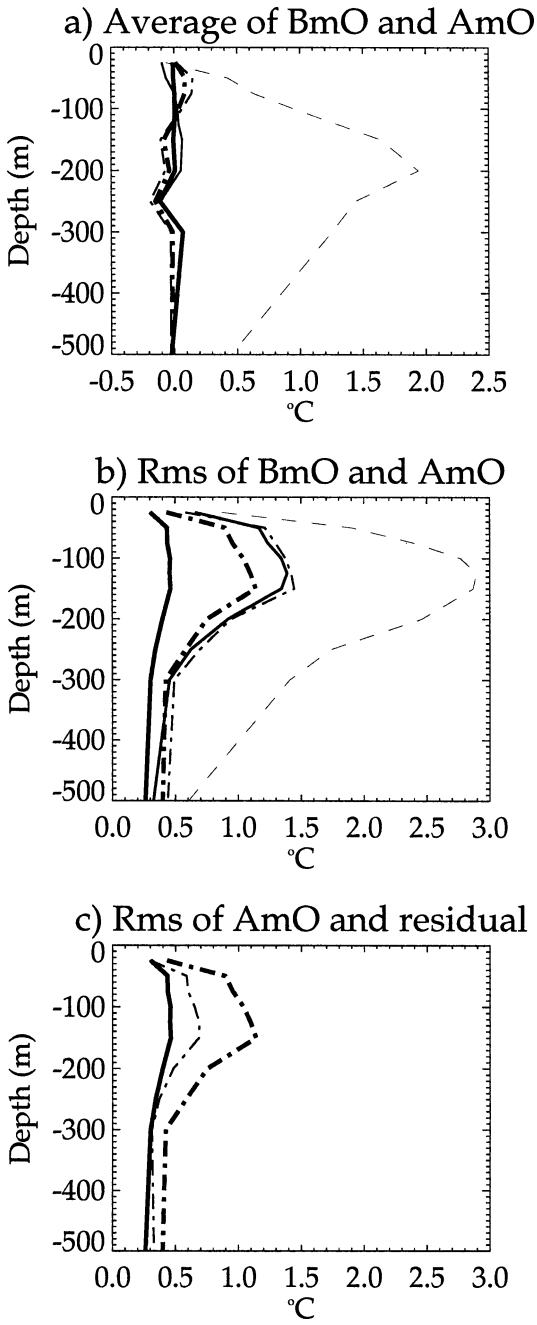

FIG. 9. Time-averaged statistics, plotted as a function of depth, of BmO and AmO for TAO data during the 1993–98 period. (a) The average of BmO (thin curves) and AmO (thick curves), (b) the rms of BmO (thin curves) and AmO (thick curves), and (c) the rms of AmO (thick curves) and of the J_o residual (thin curves). In (a) and (b), the dashed curve corresponds to EXCL; in all panels, the dashed–dotted curves correspond to EX3D, and the solid curves to EX4D. Note that the AmO and J_o residual of EX4D [solid curves in (c)] are indistinguishable.

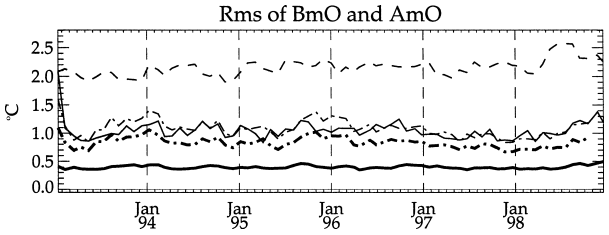

FIG. 10. Spatially averaged rms of BmO (thin curves) and AmO (thick curves) for TAO data, plotted as a function of the assimilation cycle during the 1993–98 period. The dashed curve corresponds to EXCL, the dashed–dotted curves to EX3D, and the solid curves to EX4D. The statistics for EX3D are displayed as an average over three cycles (30 days) in order to be compared with those from EX4D.

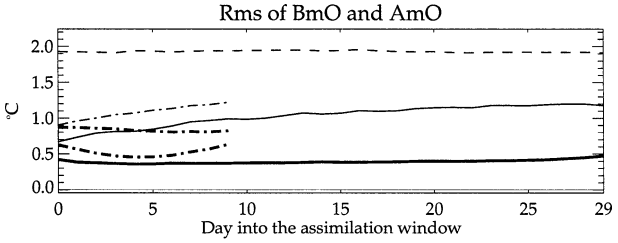

FIG. 11. Rms of the 1993–98 cycle average of BmO (thin curves), AmO (thick curves), and J_o residual (thick curves) for TAO data plotted as a function of the day within the assimilation window. The dashed curve corresponds to EXCL, the dashed–dotted curves to EX3D, and the solid curves to EX4D. The upper (lower) of the two thick dashed–dotted curves corresponds to the AmO (J_o residual) of EX3D. The AmO and J_o residual (thick solid curves) of EX4D are indistinguishable.

in which the (temperature) analysis increment was added directly to the background initial conditions as in 4DVAR. The statistics of the AmO and J_o residual for that experiment were comparable to those of EX3D but the overall impact of the assimilation on the model fields was significantly worse and therefore justifies our approach of applying the analysis increment gradually. For EX4D, the residual and AmO are indistinguishable in Fig. 9c. This result is consistent with Fig. 6 of section 4c, which showed that, after several outer iterations, the full and incremental cost functions converged to similar values at the end of minimization. That result was demonstrated for a particular cycle (the second); Fig. 9c just confirms that this feature is consistent for all cycles of EX4D.

Figure 9 provided a time-averaged view of the BmO and AmO statistics for the assimilated TAO data. Figure 10 now provides a view of how these statistics change with time during the 1993–98 analysis period. On the first assimilation cycle, the background is provided by the control and as a result the BmO is very large (Fig. 10). After the first assimilation cycle, however, the properties observed in Fig. 9 begin to emerge: the average of AmO is small in both EX3D and EX4D, and the rms of AmO is equal to about 1°C in EX3D and 0.5°C in EX4D. It was found that, in both EX3D and EX4D, the statistics of BmO and AmO are approximately stationary, and in particular do not seem to exhibit any obvious dependence on either the number of assimilated observations or interannual variability.

The cycle average of the temporal evolution of BmO and AmO *within* the assimilation window is shown in Fig. 11. For the background, which is not constrained by observations, the fit to the data degrades with time. After 10 days in EX3D and 30 days in EX4D, the rms of BmO increases by 0.2° and 0.5°C from initial values of 1.0° and 0.6°C, respectively. The rms of BmO at the end of the window is 1.2°C in both experiments, compared to 1.9°C in EXCL. In both EX3D and EX4D, the fit of the analysis to the data is uniform over their respective assimilation windows, apart from a very slight increase of the AmO at the window boundaries in EX4D. Note that, since the analysis at the end of a given cycle is used as the background at the beginning of the next cycle, the BmO at day 0 should be equal to the AmO at day 30, except for the first and last cycle for which there is no corresponding AmO (day 30) and BmO (day 0), respectively. This latter point explains the slight difference between these quantities in Fig. 11. When model error is important, and not explicitly accounted for in the assimilation method, it often manifests itself as a U shape in the fit to the data (Ménard and Daley 1996). Any signal of model error, however, is more likely to be visible in the J_o residual than in the AmO since the residual is the quantity that is minimized objectively. The residual will contain information about linearization errors introduced by the approximation (3) as well as errors in the nonlinear model itself.

The cycle average of the J_o residuals in EX3D and EX4D is also shown in Fig. 11. For EX3D (lower thick dashed–dotted curve), the U shape is evident, though quite weak, with a difference of about 0.2°C between the highest point at the window boundaries and the lowest point in the middle of the window. If this is indeed a sign of model error, then it would be consistent with the results of section 4b, which illustrated that the persistence model did have some limitations near the equator. In contrast, in EX4D, the residuals are nearly flat over the whole window width (and are very similar to the AmO). This result then suggests that model error is not a significant problem over the 30-day window used in 4DVAR.

5. Summary

Three- and four-dimensional variational assimilation systems have been developed for the rigid-lid version of the OPA OGCM (Madec et al. 1998). The assimilation problem is defined by a cost function that penalizes departures from the data and from a background estimate that is the result of a previous assimilation. The control variable of the cost function is the initial condition at the start of each assimilation window. The cost function is minimized using an incremental approach

by which the full minimization problem, involving the nonlinear OGCM as a constraint, is approximated by a sequence of quadratic minimization problems involving linear constraints. The control variable of each quadratic cost function is an increment to the initial conditions. In the incremental 3DVAR, the increment is transported forward in time by a persistence model. In the incremental 4DVAR, a linearized version of the full OGCM is used to propagate the increment. The adjoint of the linear model propagator is used to compute the gradient of the cost function with respect to the initial condition increment thus allowing the solution to be found iteratively using a gradient descent method. Once the analysis increment is found, it is applied at the beginning of the assimilation window to derive the 4DVAR analysis trajectory. In the case of 3DVAR, it is applied as a constant 3D forcing to the model equations over the assimilation window as a way of minimizing spurious adjustments.

The systems have been applied to assimilate in situ temperature data in the tropical Pacific Ocean. Three experiments were performed for the 1993–98 period: a control experiment without data assimilation, an experiment using 3DVAR cycled with a 10-day window, and an experiment using 4DVAR cycled with a 30-day window. The validity of the persistence model (3DVAR) and of the TL model (4DVAR) was investigated. It was shown that persistence was a reasonable assumption over 10 days, at least outside the equatorial waveguide, and that the TL model provides a good description of the large-scale oceanic state over at least 30 days. However, because of the nonlinear nature of convective and vertical mixing processes, the validity of the TL model could be degraded at small vertical scales. Tropical instability waves (TIWs) are nonlinear oscillations with a timescale of about 1 month. At their horizontal scale, the TL model is also less accurate. Nevertheless, in 4DVAR a very good fit to the observed TIWs was achieved (see Part II). This result points to the important role of the outer iterations in the incremental 4DVAR formulation in providing a feedback mechanism between the TL and nonlinear models so that the model can eventually achieve a very close fit to the data. With outer iterations, the final values of the incremental and nonincremental cost functions were shown to be very close, which provides a good measure of the consistency of the incremental approach for solving the original nonlinear minimization problem. Without outer iterations, the performance of the 4DVAR system is seriously degraded.

Single-observation experiments have been performed to illustrate the effect of the TL dynamics in modifying the prior estimates of the background-error variances. The TL dynamics were shown to modify the variances in a physically sensible way. The variances were diminished in the mixed layer, and the maximum value of the variance in the profile could be increased and displaced to the level of the background thermocline, where thermal variability (and background error) is greatest.

A detailed examination of the fit of the different analyses to the assimilated data has been made. The control experiment displays a large bias below the thermocline, which is strongly reduced in the 3DVAR analyses and almost entirely absent from the 4DVAR analyses. The rms difference between the analyses and observations is also very much reduced in the thermocline region. For example, for TAO data, whereas the rms difference is about 2.8°C in the control, it falls to 1.4°C in 3DVAR and to below 0.5°C in 4DVAR, which is less than the specified standard deviation of the observation error. Over the TAO region, the spatially averaged rms difference between the observations and the 3DVAR and 4DVAR analyses is stationary in time (equal to 1°C in 3DVAR and 0.4° in 4DVAR) and, in particular, does not exhibit any dependence on the number of observations or interannual variability.

In this paper, the 3DVAR and 4DVAR systems have been described and evaluated in terms of certain algorithmic and statistical diagnostics. In Part II, the analyses produced by the two systems are examined from a physical perspective. A discussion of the results of the two papers and possible avenues for future development are given at the end of Part II.

Acknowledgments. This work was initiated at LODYC with funding from an EC Human Capital and Mobility fellowship for the first author. Further funding was provided by the French MERCATOR and EC-FP5 ENACT projects. Discussions with Philippe Courtier, Pascale Delecluse, Eric Greiner, Gurvan Madec, and François Vandenberghe in the early stages of development of the assimilation system are gratefully acknowledged. The first author would like to thank ECMWF for hosting him for a year when much of this work was carried out. Erik Andersson, Magdalena Balmaseda, Mike Cullen, Mike Fisher, Andy Moore, Andrea Piacentini, and Philippe Rogel provided many useful suggestions for improving the manuscript. We are also grateful to three anonymous reviewers for their constructive and insightful remarks.

REFERENCES

Alves, J. O., M. A. Balmaseda, D. L. T. Anderson, and T. N. Stockdale, 2002: Sensitivity of dynamical seasonal forecasts to ocean initial conditions. ECMWF Tech. Memo. 369, 24 pp. [Available online at http://www.ecmwf.int/publications/.]

Andersson, E., M. Fisher, R. Munro, and A. McNally, 2000: Diagnosis of background errors for radiances and other observable quantities in a variational data assimilation system, and the explanation of a case of poor convergence. *Quart. J. Roy. Meteor. Soc.*, **126,** 1455–1472.

Balmaseda, M., D. L. T. Anderson, and M. Davey, 1994: ENSO prediction using a dynamical ocean model coupled to statistical atmospheres. *Tellus*, **46A,** 497–511.

Barnett, T. P., M. Latif, N. Graham, M. Flugel, S. Pazan, and W. White, 1993: ENSO and ENSO-related predictability. Part I:

Prediction of equatorial Pacific sea surface temperature with a hybrid coupled ocean–atmosphere model *J. Climate,* **6,** 1545–1566.

Behringer, D., M. Ji, and A. Leetma, 1998: An improved coupled model for ENSO prediction and implications for ocean initialization. Part I: The ocean data assimilation system. *Mon. Wea. Rev.,* **126,** 1013–1021.

Bennett, A. F., B. S. Chua, D. E. Harrison, and M. J. McPhaden, 2000: Generalized inversion of Tropical Atmosphere–Ocean (TAO) data using a coupled model of the tropical Pacific. *J. Climate,* **13,** 2770–2785.

Bloom, S. C., L. L. Takacs, A. M. Da Silva, and D. Ledvina, 1996: Data assimilation using incremental analysis updates. *Mon. Wea. Rev.,* **124,** 1256–1271.

Bonekamp, H., G. J. van Oldenborgh, and G. Burgers, 2001: Variational assimilation of TAO and XBT data in the HOPE OGCM, adjusting the surface fluxes in the tropical ocean. *J. Geophys. Res.,* **106,** 16 693–16 709.

Bryan, K., 1969: A numerical method for the study of the circulation of the world ocean. *J. Comput. Phys.,* **4,** 347–376.

Cane, M. A., S. E. Zebiak, and S. C. Dolan, 1986: Experimental forecasts of El Niño. *Nature,* **321,** 827–832.

Courtier, P., J.-N. Thépaut, and A. Hollingsworth, 1994: A strategy for operational implementation of 4DVAR, using an incremental approach. *Quart. J. Roy. Meteor. Soc.,* **120,** 1367–1388.

——, and Coauthors, 1998: The ECMWF implementation of three dimensional variational assimilation (3DVAR). Part I: Formulation. *Quart. J. Roy. Meteor. Soc.,* **124,** 1783–1808.

Derber, J. C., and A. Rosati, 1989: A global oceanic data assimilation system. *J. Phys. Oceanogr.,* **19,** 1333–1347.

Desroziers, G., and S. Ivanov, 2001: Diagnosis and adaptive tuning of observation error parameters in a variational assimilation. *Quart. J. Roy. Meteor. Soc.,* **127,** 1433–1452.

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

Fisher, M., and E. Andersson, 2001: Developments in 4DVAR and Kalman filtering. ECMWF Tech. Memo. 347, 36 pp. [Available online at http://www.ecmwf.int/publications/.]

Gandin, L. S., 1965: *Objective Analysis of Meteorological Fields.* Israeli Program for Scientific Translations, 242 pp.

Gauthier, P., C. Charette, L. Fillion, P. Koclas, and S. Laroche, 1999: Implementation of a 3D variational data assimilation system at the Canadian Meteorological Centre. Part I: The global analysis. *Atmos.–Ocean,* **37,** 103–156.

Gibson, J., P. Kållberg, S. Uppala, A. Noumura, A. Hernandez, and E. Serrano, 1997: ERA description. ECMWF Re-Analysis Project Report Series, No. 1, 77 pp. [Available online at http://www.ecmwf.int/research/era/ERA-15/Report_Series/.]

Giering, R., and T. Kaminski, 1998: Recipes for adjoint code construction. *ACM Trans. Math. Software,* **24,** 437–474.

Gilbert, J.-C., and C. Lemaréchal, 1989: Some numerical experiments with variable-storage quasi-Newton algorithms. *Math. Program.,* **45,** 407–435.

Gill, A. E., 1982: *Atmosphere–Ocean Dynamics.* Academic Press, 662 pp.

Greiner, E., and S. Arnault, 2000: Comparing the results of a 4D-variational assimilation of satellite and in situ data with WOCE CITHER hydrographic measurements in the tropical Atlantic. *Progress in Oceanography,* Vol. 47, Pergamon Press, 1–68.

——, ——, and A. Morlière, 1998a: Twelve monthly experiments of 4D-variational assimilation in the tropical Atlantic during 1987: Part I: Method and statistical results. *Progress in Oceanography,* Vol. 41, Pergamon Press, 141–202.

——, ——, and ——, 1998b: Twelve monthly experiments of 4D-variational assimilation in the tropical Atlantic during 1987: Part II: Oceanographic interpretation. *Progress in Oceanography,* Vol. 41, Pergamon Press, 203–247.

Grima, N., A. Bentamy, K. Katsaros, Y. Quilfen, P. Delecluse, and C. Lévy, 1999: Sensitivity study of an oceanic general circulation model forced by satellite wind-stress fields. *J. Geophys. Res.,* **104,** 7967–7989.

Hollingsworth, A., and P. Lönnberg, 1989: The verification of objective analyses: Diagnostics of analysis system performance. *Meteor. Atmos.,* **40,** 3–27.

Ide, K., P. Courtier, M. Ghil, and A. C. Lorenc, 1997: Unified notation for data assimilation: Operational, sequential and variational. *J. Meteor. Soc. Japan,* **75,** 181–189.

Janiskova, M., J.-N. Thépaut, and J.-F. Geleyn, 1999: Simplified and regular physical parametrizations for incremental four-dimensional variational assimilation. *Mon. Wea. Rev.,* **127,** 26–45.

Ji, M., and A. Leetma, 1997: Impact of data assimilation on ocean initialization and El Niño prediction. *Mon. Wea. Rev.,* **125,** 742–753.

Kessler, W. S., M. C. Spillane, M. J. McPhaden, and D. E. Harrison, 1996: Scales of variability in the equatorial Pacific inferred from the Tropical Atmosphere–Ocean buoy array. *J. Climate,* **9,** 2999–3024.

Laroche, S., and P. Gauthier, 1998: A validation of the incremental formulation of 4D variational data assimilation in a nonlinear barotropic flow. *Tellus,* **50A,** 557–572.

Latif, M., and M. Flügel, 1991: An investigation of short-range climate predictability in the tropical Pacific. *J. Geophys. Res.,* **96,** 2661–2673.

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

Legeckis, R., 1977: Long waves in the eastern equatorial Pacific; a view from a geostationary satellite. *Science,* **197,** 1177–1181.

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

Lorenc, A. C., 1981: A global three-dimensional multivariate statistical interpolation scheme. *Mon. Wea. Rev.,* **109,** 701–721.

——, and Coauthors, 2000: The Met. Office global three-dimensional variational data assimilation scheme. *Quart. J. Roy. Meteor. Soc.,* **126,** 2991–3012.

Madec, G., P. Delecluse, M. Imbard, and C. Lévy, 1998: OPA 8.1 Ocean General Circulation Model reference manual. LODYC/IPSL Technical Note 11, Paris, France, 91 pp. [Available online at http://www.lodyc.jussieu.fr/opa/.]

Mahfouf, J.-F., 1999: Influence of physical processes on the tangent-linear approximation. *Tellus,* **51A,** 147–166.

McCarty, M. E., L. J. Mangum, and M. J. McPhaden, 1997: Temperature errors in TAO data induced by mooring motion. NOAA Tech. Memo. ERL PMEL-108, 14 pp.

McPhaden, M. J., 1993: TOGA-TAO and the 1991-93 El Niño–Southern Oscillation event. *Oceanography,* **6,** 36–44.

Ménard, R., and R. Daley, 1996: The application of Kalman smoother theory to the estimation of 4DVAR error statistics. *Tellus,* **48A,** 221–237.

Meyers, G., H. Phillips, N. Smith, and J. Sprintall, 1991: Space and time scales for optimal interpolation of temperature—Tropical Pacific Ocean. *Progress in Oceanography,* Vol. 28, Pergamon Press, 189–218.

Palmer, T. N., and D. L. T. Anderson, 1994: Prospects for seasonal forecasting. *Quart. J. Roy. Meteor. Soc.,* **120,** 755–794.

Parrish, D. F., and J. C. Derber, 1992: The National Meteorological Center's spectral statistical interpolation analysis system. *Mon. Wea. Rev.,* **120,** 1747–1763.

Philander, S. G., 1989: *El Niño, La Niña, and the Southern Oscillation.* Academic Press, 293 pp.

Rabier, F., A. McNally, E. Andersson, P. Courtier, P. Undén, J. Eyre, A. Hollingsworth, and F. Bouttier, 1998: The ECMWF implementation of three-dimensional variational assimilation (3DVAR). Part II: Structure functions. *Quart. J. Roy. Meteor. Soc.,* **124,** 1809–1829.

——, H. Järvinen, E. Klinker, J. F. Mahfouf, and A. Simmons, 2000: The ECMWF operational implementation of four-dimensional variational assimilation. Part I: Experimental results with simplified physics. *Quart. J. Roy. Meteor. Soc.,* **126,** 1143–1170.

Reynolds, R. W., and T. M. Smith, 1994: Improved global sea surface temperature analyses using optimal interpolation. *J. Climate,* **7,** 929–948.

Rosati, A., K. Miyakoda, and R. Gudgel, 1997: The impact of ocean initial conditions on ENSO forecasting with a coupled model. *Mon. Wea. Rev.,* **125,** 754–772.

Roullet, G., and G. Madec, 2000: Salt conservation, free surface, and varying levels: A new formulation for ocean general circulation models. *J. Geophys. Res.,* **105,** 23 927–23 942.

Rutherford, I., 1972: Data assimilation by statistical interpolation of forecast error fields. *J. Atmos. Sci.,* **29,** 809–815.

Segschneider, J., D. L. T. Anderson, and T. N. Stockdale, 2000: Towards the use of altimetry for operational seasonal forecasting. *J. Climate,* **13,** 3116–3138.

——, ——, J. Vialard, M. A. Balmaseda, and T. N. Stockdale, 2001: Initialization of seasonal forecasts assimilating sea level and temperature observations. *J. Climate,* **14,** 4292–4307.

Smith, N. R., J. E. Blomley, and G. Meyers, 1991: A univariate statistical interpolation scheme for subsurface thermal analyses in the tropical oceans. *Progress in Oceanography,* Vol. 28, Pergamon Press, 219–256.

Talagrand, O., 1991: The use of adjoint equations in numerical modeling of the atmospheric circulation. *Automatic Differentiation of Algorithms: Theory, Implementation, and Application,* A. Griewank and G. F. Corliss, Eds., SIAM, 169–180.

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

Tarantola, A., 1987: *Inverse Problem Theory: Methods for Data Fitting and Model Parameter Estimation.* Elsevier, 613 pp.

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

Thépaut, J.-N., R. N. Hoffman, and P. Courtier, 1993: Interactions of dynamics and observations in four-dimensional variational assimilation. *Mon. Wea. Rev.,* **121,** 3393–3414.

——, P. Courtier, G. Belaud, and G. LeMaître, 1996: Dynamical structure functions in a four-dimensional variational assimilation: A case study. *Quart. J. Roy. Meteor. Soc.,* **122,** 535–561.

Tzipermann, 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.

——, ——, ——, ——, 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.

Vialard, J., C. Menkes, J.-P. Boulanger, P. Delecluse, E. Guilyardi, M. J. McPhaden, and G. Madec, 2001: A model study of oceanic mechanisms affecting equatorial Pacific sea surface temperature during the 1997–98 El Niño. *J. Phys. Oceanogr.,* **31,** 1649–1675.

——, A. T. Weaver, D. L. T. Anderson, and P. Delecluse, 2003: Three- and four-dimensional variational assimilation with a general circulation model of the tropical Pacific Ocean. Part II: Physical validation. *Mon. Wea. Rev.,* **131,** 1379–1395.

Weaver, A. T., and P. Courtier, 2001: Correlation modelling on the sphere using a generalized diffusion equation. *Quart. J. Roy. Meteor. Soc.,* **127,** 1815–1846.

——, J. Vialard, D. L. T. Anderson, and P. Delecluse, 2002: Three- and four-dimensional variational assimilation with a general circulation model of the tropical Pacific Ocean. ECMWF Tech. Memo. 365, 74 pp. [Available online at http://www.ecmwf.int/publications/.]

Xu, Q., 1996: Generalized adjoint for physical processes with parametrized discontinuities. Part I: Basic issues and heuristic examples. *J. Atmos. Sci.,* **53,** 1123–1142.

Zhu, J., and M. Kamachi, 2000: The role of time step size in numerical stability of tangent linear models. *Mon. Wea. Rev.,* **128,** 1562–1572.

Zou, X., 1997: Tangent-linear and adjoint of 'on–off' processes and their feasibility for use in 4-dimensional variational data assimilation. *Tellus,* **49A,** 3–31.