JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 102, NO. C6, PAGES 12,479–12,491, JUNE 15, 1997

Parameter estimation solving a weak constraint variational formulation for an Ekman model

Mette Eknes and Geir Evensen

Nansen Environmental and Remote Sensing Center, Bergen, Norway

Abstract. A weak constraint variational formulation is used for inverse calculations and parameter estimation in a one-dimensional Ekman model. When parameters in the model are allowed to contain errors, the inverse problem becomes nonlinear even if the model itself is linear. It is shown that a convergent iteration can be defined for the nonlinear system of Euler–Lagrange equations and that improved estimates of the poorly known parameters can be calculated by solving the inverse problem for each of the linear iterates using the representer method. The formulation of the variational problem and the solution methods are illustrated using a simple example. The use of a simple dynamical model makes it possible to give an instructive presentation of the representer method. The method is finally used in an example using real current meter data. It is shown that the weak constraint formulation results in smooth solutions in good agreement with the data all through the water column and that it is superior to the traditional strong constraint inverse estimate.

Introduction

Data assimilation and inverse methods are normally used for generating estimates of dynamical variables, taking into account both the information about the dynamics from a dynamical model and the information about the true state which is contained in a set of measurements. Such techniques have also been proposed as a tool for parameter estimation in dynamical models, although only a few works have so far considered the parameter estimation problem in this context. Examples are *Smedstad and O'Brien* [1991], who estimated the phase speeds in a reduced-gravity model of equatorial waves, using a strong constraint variational formulation, and *Yu and O'Brien* [1991, 1992], who used a modified one-dimensional Ekman layer model in combination with variational optimal control techniques to simultaneously estimate the surface wind drag coefficient and the vertical profile of the eddy viscosity from observed data.

Yu and O'Brien [1991] defined a cost function where the first guesses of the diffusion coefficient and the wind drag coefficient were penalized in addition to the residual between the model results and the observations. However, there was no penalty on the first-guess initial conditions. Without such a penalty, every choice of initial conditions can be used, and there may exist many initial conditions resulting in a solution which interpolates the data and gives a penalty function equal to zero [*Bennett and Miller*, 1990]. *Yu and O'Brien* [1992] included such a term although the resulting equation for the initial condition was incorrect. The cost function was minimized using the adjoint technique to calculate the gradient of the penalty function with respect to the control parameters, i.e., the initial condition, the wind drag, and the eddy viscosity with the dynamical model acting as a strong constraint.

Here the parameter estimation problem is reexamined using a weak constraint variational inverse formulation. Thus the model, the initial and boundary conditions, the measurements, the diffusion parameter, and the wind drag coefficient are all allowed to contain errors. The strong constraint inverse may be obtained as a limiting case where the model errors approach zero.

If the wind drag C_d and the vertical diffusion parameter A are considered known, the inverse formulation for the Ekman model is linear. However, allowing them to contain errors, the problem becomes nonlinear and some sort of iteration procedure must be used to solve the Euler–Lagrange equations. Here it is shown that by defining an iteration for the unknown parameters, each of the linear iterates for the Euler–Lagrange equations can be solved exactly using the representer method [*Bennett*, 1992]. Such an approach is similar to previous applications of the representer method with nonlinear dynamics. For instance, *Bennett and Thorburn* [1992] and *Bennett et al.* [1993] solved for the weak constraint inverse of a nonlinear barotropic quasi-geostrophic model by defining a convergent sequence of linear iterates of the Euler–Lagrange equations, where each iterate could be solved using the representer technique.

In the following sections, the inverse formulation is discussed and the Euler–Lagrange equations are derived. Then, the representer method is applied to decouple the Euler–Lagrange equations for the linear problem resulting when the diffusion parameter and the wind drag coefficient are assumed known. Next we define iterations of the Euler–Lagrange equations which are used to solve for the diffusion parameter and the wind drag coefficient, and an example that illustrates the method is presented. Finally, the LOTUS-3 data set [*Tarbell et al.*, 1984; *Bowers et al.*, 1986] has been used in a real data assimilation experiment to compare the

Paper number 96JC03454.
0148-0227/97/96JC-03454$09.00

results of a weak constraint formulation with the results of *Yu and O'Brien* [1991, 1992].

Inverse Formulation

The Ekman model can be written in nondimensional form by defining characteristic scales for the dependent and independent variables. The timescale is $\tau = f^{-1}$, where f is the Coriolis parameter, and the depth scale is the Ekman layer thickness $\delta_e = \sqrt{2A(0)f^{-1}}$, where $A(0)$ is the diffusion coefficient at the surface. A velocity scale becomes $\delta_e\tau^{-1} = \sqrt{2A(0)f}$, and the diffusion coefficient is scaled by $2A(0)$. The atmospheric wind speed is scaled by $\sqrt{2A(0)f\rho_w\rho_a^{-1}}$, where ρ_a and ρ_w are the air and water densities. The nondimensional Ekman layer model then becomes

$$\frac{\partial \mathbf{u}}{\partial t} + \mathbf{k} \times \mathbf{u} = \frac{\partial}{\partial z}\left(A\frac{\partial \mathbf{u}}{\partial z}\right) + \mathbf{q}, \tag{1}$$

where $\mathbf{u}(z,t)$ is the horizontal velocity vector, $A = A(z)$ is the diffusion coefficient, and $\mathbf{q}(z,t)$ is the unknown model error. The initial conditions are given as

$$\mathbf{u}(z,0) = \mathbf{U}_0 + \mathbf{a}, \tag{2}$$

where $\mathbf{a}$ contains the errors in the first-guess initial condition $\mathbf{U}_0$. The boundary conditions for the model are

$$A\frac{\partial \mathbf{u}}{\partial z}\bigg|_{z=0} = \left(C_d\sqrt{u_a^2 + v_a^2}\right)\mathbf{u}_a + \mathbf{b}_0, \tag{3}$$

$$A\frac{\partial \mathbf{u}}{\partial z}\bigg|_{z=-H} = \mathbf{o} + \mathbf{b}_H, \tag{4}$$

where the position $z = 0$ is at the ocean surface and the lower boundary is at $z = -H$, C_d is the wind drag coefficient, $\mathbf{u}_a$ is the atmospheric wind speed, and $\mathbf{b}_0$ and $\mathbf{b}_H$ are the unknown errors in the boundary conditions.

Now a set of measurements, $\mathbf{d}$, of the true solution are assumed given and linearly related to the model variables by the measurement equation

$$\mathbf{d} = \mathcal{L}[\mathbf{u}^t] + \epsilon. \tag{5}$$

Here $\mathcal{L}$ is a vector of linear measurement functionals, $\mathbf{u}^t$ is the true state, and ϵ is the measurement error.

If all the error terms are zero, the problem becomes overdetermined and no solution can be found in general. However, by allowing the model dynamics, the initial and boundary conditions, and the measurements to contain errors, a solution can be found which minimizes these error terms in a weighted least squares sense.

Here we also allow for the first guesses of the wind drag and the diffusion coefficient, C_{d_0} and $A_0(z)$, to contain errors, i.e.,

$$C_d = C_{d_0} + p_{C_d}, \tag{6}$$

$$A(z) = A_0(z) + p_A(z), \tag{7}$$

where p_{C_d} and $p_A(z)$ are the unknown error terms. Thus a combined state estimation and parameter estimation problem is formulated. All the error terms are assumed to be normal distributed with zero mean and known covariances.

A convenient variational formulation is

$$\begin{aligned}
&\mathcal{J}[\mathbf{u}, C_d, A] \\
&= \int_0^T dt_1 \int_0^T dt_2 \int_{-H}^0 dz_1 \int_{-H}^0 dz_2 \\
&\qquad \mathbf{q}^{\mathrm{T}}(z_1,t_1)\,\mathbf{W}_{qq}(z_1,t_1,z_2,t_2)\,\mathbf{q}(z_2,t_2) \\
&+ \int_{-H}^0 dz_1 \int_{-H}^0 dz_2\; \mathbf{a}^{\mathrm{T}}(z_1)\,\mathbf{W}_{aa}(z_1,z_2)\,\mathbf{a}(z_2) \\
&+ \int_0^T dt_1 \int_0^T dt_2\; \mathbf{b}_0^{\mathrm{T}}(t_1)\,\mathbf{W}_{b_0b_0}(t_1,t_2)\,\mathbf{b}_0(t_2) \\
&+ \int_0^T dt_1 \int_0^T dt_2\; \mathbf{b}_H^{\mathrm{T}}(t_1)\,\mathbf{W}_{b_Hb_H}(t_1,t_2)\,\mathbf{b}_H(t_2) \\
&+ \int_{-H}^0 dz_1 \int_{-H}^0 dz_2\; p_A(z_1) W_{AA}(z_1,z_2)\,p_A(z_2) \\
&+ p_{C_d} W_{C_dC_d} p_{C_d} \\
&+ \epsilon^T \mathbf{w} \epsilon.
\end{aligned} \tag{8}$$

A simpler way of writing this may be

$$\begin{aligned}
\mathcal{J}[\mathbf{u}, C_d, A] &= \mathbf{q}^T \bullet \mathbf{W}_{qq} \bullet \mathbf{q} + \mathbf{a}^T \circ \mathbf{W}_{aa} \circ \mathbf{a} \\
&\quad + \mathbf{b}_0^T * \mathbf{W}_{b_0b_0} * \mathbf{b}_0 + \mathbf{b}_H^T * \mathbf{W}_{b_Hb_H} * \mathbf{b}_H \\
&\quad + p_A \circ W_{AA} \circ p_A + p_{C_d} W_{C_dC_d} p_{C_d} + \epsilon^T \mathbf{w} \epsilon,
\end{aligned} \tag{9}$$

where the solid circles mean integration both in space and time, the open circles mean integration in space, and the asterisks mean integration in time.

The weights $\mathbf{W}_{??}$, are functional inverses of the respective covariances $\mathbf{Q}_{??}$, e.g., for the model weight, $\mathbf{Q}_{qq} \bullet \mathbf{W}_{qq} = \delta(z_1 - z_3)\delta(t_1 - t_3)\mathbf{I}$, or written out,

$$\begin{aligned}
&\int_0^T dt_2 \int_{-H}^0 dz_2\; \mathbf{Q}_{qq}(z_1,t_1,z_2,t_2)\mathbf{W}_{qq}(z_2,t_2,z_3,t_3) \\
&= \delta(z_1 - z_3)\delta(t_1 - t_3)\mathbf{I},
\end{aligned} \tag{10}$$

where $\mathbf{w}$ is the inverse of the measurement error covariance matrix $\mathbf{w}^{-1}$ and the weight $W_{C_dC_d}$ is 1 over the error variance of p_{C_d}. These weights determine the spatial and temporal scales for the physical problem and ensure smooth influences from the measurements. Note that the first guesses of all unknown quantities are penalized. This is required to ensure a unique solution of the inverse as shown by *Bennett and Miller* [1990]. Even if no measurements are available, the inverse will still have a unique solution corresponding to the first-guess solution. This implies that the role of each of the measurements is to add a specific contribution (influence function) to the first- guess solution.

Note that estimators other than least squares could be defined. However, the least squares formulation is attractive for several reasons. If the unknown errors are Gaussian, i.e., completely explained by the two first statistical moments, mean and covariance, then minimizing (8) is equivalent to finding the maximum likelihood estimate. When working with methods that involve the Euler–Lagrange equations,

these are readily derived and the derivatives of the penalty function exist everywhere.

Some important differences between the formulation used in this paper and the previous works by *Yu and O'Brien* [1991, 1992] should be mentioned. Here, in addition to allowing the model and boundary conditions to contain errors, nondiagonal weights have been used in the penalty function (8) to ensure smooth results. In the works by *Yu and O'Brien* [1991, 1992] the weights were diagonal, which means that there was no regularization imposed on the control variables; that is, they would accept noisy estimates as solutions.

By substituting for the error terms from (1) to (7) in the penalty function $\mathcal{J}[\mathbf{u}, C_d, A]$ and taking the variation with respect to $\mathbf{u}$, the system of Euler-Lagrange equations becomes

$$\frac{\partial \mathbf{u}}{\partial t} + \mathbf{k} \times \mathbf{u} = \frac{\partial}{\partial z}(A\frac{\partial \mathbf{u}}{\partial z}) + \mathbf{Q}_{qq} \bullet \boldsymbol{\lambda}, \tag{11}$$

with initial conditions

$$\mathbf{u}|_{t=0} = \mathbf{U}_0 + \mathbf{Q}_{aa} \circ \boldsymbol{\lambda}, \tag{12}$$

and boundary conditions

$$A\frac{\partial \mathbf{u}}{\partial z}\bigg|_{z=0} = C_d\sqrt{u_a^2 + v_a^2}\,\mathbf{u}_a + \mathbf{Q}_{b_0 b_0} * \boldsymbol{\lambda}, \tag{13}$$

$$A\frac{\partial \mathbf{u}}{\partial z}\bigg|_{z=-H} = -\mathbf{Q}_{b_H b_H} * \boldsymbol{\lambda}. \tag{14}$$

The so-called adjoint equation becomes

$$\begin{aligned} -\frac{\partial \boldsymbol{\lambda}}{\partial t} - \mathbf{k} \times \boldsymbol{\lambda} &= \frac{\partial}{\partial z}(A\frac{\partial \boldsymbol{\lambda}}{\partial z}) \\ &+ \mathcal{L}^{\mathrm{T}}[\delta(z - z_2)\delta(t - t_2)]\mathbf{w}[\mathbf{d} - \mathcal{L}[\mathbf{u}]], \end{aligned} \tag{15}$$

subject to the "final" condition

$$\boldsymbol{\lambda}|_{t=T} = \mathbf{o}, \tag{16}$$

and the boundary condition

$$\frac{\partial \boldsymbol{\lambda}}{\partial z}\bigg|_{z=0, z=-H} = \mathbf{o}. \tag{17}$$

The system (11) to (17) is the Euler-Lagrange equations which here comprise a two-point boundary value problem in space and time, and since they are coupled, they must be solved simultaneously. Equation (11) is the dynamical model forced by a term that estimates the model errors and contains the adjoint variable. A similar coupling to the adjoint variable is also contained in the initial conditions (12) and boundary conditions (13) and (14). The "backward" or adjoint equation (15), which strictly speaking is the Euler-Lagrange equation, contains a weak coupling to the "forward" variable $\mathbf{u}$ at measurement locations. Note that the strong constraint assumption removes the coupling of (11) to the adjoint variables, but the Euler-Lagrange equations are still coupled through the initial and boundary conditions (12) to (14). In the adjoint technique this coupling is iterated using a gradient descent procedure.

Since the drag coefficient and the diffusion are allowed to contain errors, the variation of the penalty function with respect to these parameters must also be taken. This results in the additional equations

$$C_d = C_{d_0} + Q_{C_d C_d} \int_0^T \boldsymbol{\lambda}^T(0, t)\,\mathbf{u}_a dt, \tag{18}$$

$$A = A_0 - Q_{AA} \bullet \frac{\partial \boldsymbol{\lambda}^{\mathrm{T}}}{\partial z}\frac{\partial \mathbf{u}}{\partial z}, \tag{19}$$

for the wind drag coefficient and the diffusion parameter. The addition of the two equations (18) and (19) makes the system of Euler–Lagrange equations nonlinear.

Representer Solution

The nonlinearity appearing in the Euler-Lagrange equations for parameter estimation problems suggests that some kind of iterative procedure should be used. If the parameters A and C_d are given, the remaining Euler–Lagrange equations are linear and can be solved exactly using a representer expansion. It is therefore proposed to iterate (18) and (19) and to solve for each of the linear iterates using the representer method.

Assume now that the forward and backward variables can be expressed as

$$\mathbf{u} = \mathbf{u}_F + \sum_{m=1}^{M} b_m \mathbf{r}_m, \tag{20}$$

$$\boldsymbol{\lambda} = \boldsymbol{\lambda}_F + \sum_{m=1}^{M} b_m \boldsymbol{\alpha}_m. \tag{21}$$

The coefficients b_m are amplitudes for the M influence functions or representers, $\mathbf{r}_m(z, t)$. There is one coefficient for each of the measurements. Here the forward variable contains both a u and a v component and so will each of the representers.

If expressions (20) and (21) are substituted into the Euler-Lagrange equations, one first observes that the first-guess solution, $\boldsymbol{\lambda}_F = \mathbf{o}$, is the solution of the homogeneous backward model, while $\mathbf{u}_F$ is a first-guess model solution with $\boldsymbol{\lambda}_F = \mathbf{o}$, i.e., the model solution found when no information from the measurements is used.

The M representers are found by solving the initial value problems

$$\frac{\partial \mathbf{r}_m}{\partial t} + \mathbf{k} \times \mathbf{r}_m = \frac{\partial}{\partial z}\left(A\frac{\partial \mathbf{r}_m}{\partial z}\right) + \mathbf{Q}_{qq} \bullet \boldsymbol{\alpha}_m, \tag{22}$$

with initial condition

$$\mathbf{r}_m|_{t=0} = \mathbf{Q}_{aa} \circ \boldsymbol{\alpha}_m, \tag{23}$$

and boundary conditions

$$A\frac{\partial \mathbf{r}_m}{\partial z}\bigg|_{z=0} = \mathbf{Q}_{b_0 b_0} * \boldsymbol{\alpha}_m, \tag{24}$$

$$A\frac{\partial \mathbf{r}_m}{\partial z}\bigg|_{z=-H} = -\mathbf{Q}_{b_H b_H} * \boldsymbol{\alpha}_m. \tag{25}$$

These equations are coupled to the adjoints of the representers, $\boldsymbol{\alpha}_m$, which satisfy the "final" value problems

$$-\frac{\partial \boldsymbol{\alpha}_m}{\partial t} - \mathbf{k} \times \boldsymbol{\alpha}_m = \frac{\partial}{\partial z}\left(A\frac{\partial \boldsymbol{\alpha}_m}{\partial z}\right) + \mathcal{L}_m[\delta(z-z_2)\delta(t-t_2)], \tag{26}$$

with "final" conditions

$$\boldsymbol{\alpha}_m|_{t=T} = \mathbf{o}, \tag{27}$$

and boundary conditions

$$\left.\frac{\partial \boldsymbol{\alpha}_m}{\partial z}\right|_{z=0,\, z=-H} = \mathbf{o}. \tag{28}$$

The equations for the representers and their adjoints are now decoupled since the dependence to the forward variable has been removed in (26). This has been done by choosing $\mathbf{b}$ to satisfy the linear system

$$(\mathbf{R} + \mathbf{w}^{-1})\mathbf{b} = \mathbf{d} - \mathcal{L}[\mathbf{u}_F], \tag{29}$$

where the representer matrix $\mathbf{R}$ is defined as

$$\mathbf{R}(:, m) = \mathcal{L}[\mathbf{r}_m]; \tag{30}$$

that is, $\mathbf{R}$ is constructed by measuring the representers. By rearranging (29), one can write

$$\mathbf{b} = \mathbf{w}[\mathbf{d} - \mathcal{L}(\mathbf{u}_F + \sum_{m=1}^{M} b_m \mathbf{r}_m)] = \mathbf{w}[\mathbf{d} - \mathcal{L}(\mathbf{u})]. \tag{31}$$

This expression can be used on the right-hand side of (15) to decouple the Euler–Lagrange equations.

The solution procedure for the representer method can be summerized as follows. First, each of the representers is calculated by a backward integration of (26)–(28) to get $\boldsymbol{\alpha}_m$, followed by a forward integration of (22)–(25) to get $\mathbf{r}_m$. Note that only the representer matrix is required to find $\mathbf{b}$, so only the "measurements" of the representers need to be stored. The calculations of the M representers are entirely independent of each other and may therefore be computed very efficiently on a multiprocessor computer [*Bennett and Baugh*, 1992]. When the representer matrix has been generated, the system (29) can be solved after a forward integration for $\mathbf{u}_F$ to calculate the residual between the measurements and the first-guess solution appearing on the right-hand side of (29). The coefficient vector $\mathbf{b}$ is then used in (15) to find $\boldsymbol{\lambda}$ by a backward integration of (15)–(17), followed by a forward integration of (11)–(14) to find the inverse solution. The total cost of the solution algorithm is the integration of $2M+3$ initial value problems, and only one field as function of (z, t) needs to be stored simultaneously.

It should be noted that the expression (20) for $\mathbf{u}$ does not represent all arbitrary functions of x and t. However, all observable fields can be represented by (20), and only the unobservable fields have been rejected [see *Bennett*, 1992]. Thus the problem has been reduced to searching for the solution in an M-dimensional space spanned by the representers.

A posterior error covariance estimate can be calculated for the solution by using

$$\mathbf{C_{uu}}(z_1, t_1, z_2, t_2) = \boldsymbol{\Gamma}(z_1, t_1, z_2, t_2) - \mathbf{r}(z_1, t_1)^{\mathrm{T}}\left(\mathbf{R} + \mathbf{w}^{-1}\right)^{-1}\mathbf{r}(z_2, t_2), \tag{32}$$

where $\boldsymbol{\Gamma}(z_1, t_1, z_2, t_2)$ is the representer or prior space-time covariance function for the first-guess solution and $\mathbf{r}^{\mathrm{T}}\left(\mathbf{R} + \mathbf{w}^{-1}\right)^{-1}\mathbf{r}$ is the explained error covariance. Normally, only the variances are needed and $\boldsymbol{\Gamma}(z, t, z, t)$ may be estimated using statistical simulations. It should also be mentioned that each of the representers can be expressed as

$$\mathbf{r}_m = \mathcal{L}_m[\boldsymbol{\Gamma}]; \tag{33}$$

thus the representer method is equivalent to Gauss-Markov smoothing in space and time [*Bennett*, 1992]. A comparison of the representer method and the Kalman filter has been given by *Evensen* [1994a] .

Parameter Estimation

In the previous section it was illustrated how the Euler–Lagrange equations for the weak constraint inverse formulation could be solved exactly when $A(z)$ and C_d were known. When the parameters are allowed to contain errors, the inverse problem becomes nonlinear and therefore an iteration will be used for (18) and (19) for $A(z)$ and C_d. In each iteration, the representer technique will be used to solve for the corresponding inverse estimate.

The equations (18) and (19) were here iterated using a gradient descent method, i.e.,

$$C_d^{n+1} = C_d^n - \gamma\left(C_d^n - C_{d_0} - Q_{C_dC_d}\int_0^T \boldsymbol{\lambda}_n^{\mathrm{T}}\sqrt{u_a^2 + v_a^2}\,\mathbf{u}_a\,dt\right), \tag{34}$$

$$A^{n+1}(z) = A^n(z) - \beta\left[A^n(z) - A_0(z) + Q_{pp} \bullet \frac{\partial \mathbf{u}_n^{\mathrm{T}}}{\partial z}\frac{\partial \boldsymbol{\lambda}_n}{\partial z}\right], \tag{35}$$

Note that the expressions inside the parantheses and brackets are the actual gradients used in the gradient descent algorithm. The constants γ and β determine length of the steps in the direction of the gradient in the parameter space and have an important impact on the convergence. If chosen too large, the new values for the parameters will be unphysical and the method will probably diverge. If chosen too small, the method will converge but slowly. Thus there is an optimal value that provides the fastest convergence. The Newton method was also examined and gave similar convergence.

On the other hand, the problem studied here is rather simple and the number of iterations is not crucial since only a few CPU seconds are needed for each iteration. If the convergence rate is important, as it will be for a larger problem having a huge state space, then a more sophisticated method should be used.

The equations (34) and (35) are now iterated to generate new guesses C_{d}^{n+1} and A^{n+1}, which are used to solve for $\mathbf{u}_{n+1}$ and $\boldsymbol{\lambda}_{n+1}$ using the representer technique described in the previous section.

As pointed out by *Bennett* [1992], *Yu and O'Brien* [1991] did not impose a smoothing constraint on the diffusion co-

Table 1. Physical Parameters Used in the Twin Experiment

Symbol	Description	Value
T	integration time	50 hours
Δt	time step	6 min
H	depth	40 m
Δz	grid size	1.4 m
f	Coriolis parameter	1.3×10^{-6} s^{-1}
δ_E	Ekman layer thickness	6.3 m
ρ_a	density of air	1.2 kg m^{-3}
ρ_w	density of water	1.025×10^3 kg m^{-3}

efficient $A(z)$. It was therefore not clear if there was any difference in varying $A(0)$ or C_d in the surface condition (13), since $A(z)$ may be discontinuous. However, here the nondiagonal weights will ensure a smooth $A(z)$. It is therefore expected that a vertical profile of the solution for $\mathbf{u}$, which is consistent with the measurements, will determine the profile for $A(z)$, while C_d will adjust to provide the correct surface forcing. Here we also included an error term in the actual boundary condition to account for errors in the atmospheric wind data. Clearly, these error sources give rise to a highly nonlinear problem, where multiple minima may exist and a unique solution is not guaranteed.

Example

Here a simple example will be used to illustrate the method which is proposed for parameter estimation. A constant wind with $\mathbf{u}_a = (10 \text{ m s}^{-1}, 10 \text{ m s}^{-1})$ has been used to spin up the vertical velocity structure in the first-guess solution, starting with an initial condition $\mathbf{u}(z, 0) = \mathbf{o}$ and then performing 50 hours of integration. The reference case, from which velocity data are extracted, is generated by continuing the integration for another 50 hours. Values for some of the physical parameters are given in Table 1. Note that the values for variables in the text and the tables are all given in dimensional units.

By measuring the reference case and adding Gaussian noise, eight simulated measurements of $\mathbf{u}$ were generated; that is, a total of 16 measurements of u and v components were used. The locations of the measurements are shown in Figure 1.

All error terms are assumed to be unbiased, and the error covariances were specified as follows:

$$\mathbf{Q}_{aa}(z_1, z_2) = \sigma_a^2 \exp\left\{-\left(\frac{z_1 - z_2}{l_a}\right)^2\right\}\mathbf{I}, \quad (36)$$

$$\mathbf{Q}_{b_0 b_0}(t_1, t_2) = \sigma_{b_0}^2 \delta(t_1 - t_2)\mathbf{I}, \quad (37)$$

$$\mathbf{Q}_{b_H b_H}(t_1, t_2) = \sigma_{b_H}^2 \delta(t_1 - t_2)\mathbf{I}, \quad (38)$$

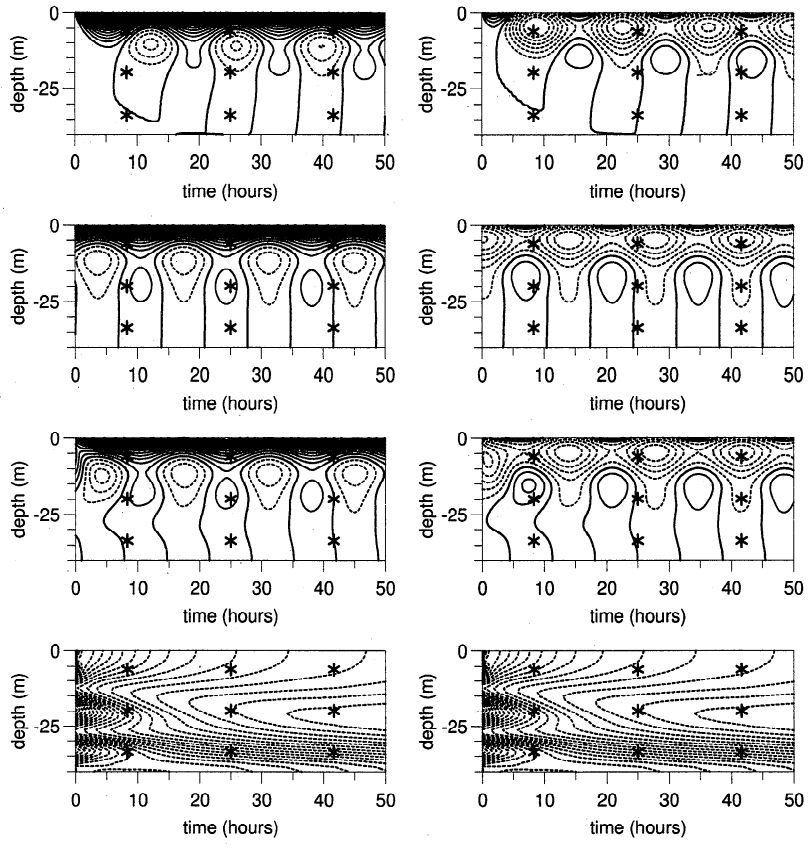

Figure 1. Solution from the identical twin experiment. The left and right columns contain the u and v components, respectively, of (from top to bottom) the first-guess estimate $\mathbf{u}_F$, the reference case $\mathbf{u}$, the inverse estimate $\hat{\mathbf{u}}$ and the error variance reduction. The contour intervals are 0.05 m s^{-1} for the velocity plots and 0.00006 m^2 s^{-2} for the error variance reduction. For all the plots the thick lines are the zero contour, dotted lines denote negative values, and solid lines denote positive values. The measurement locations are marked with an asterisk.

Table 2. Dimensional Values of the Variances and the Decorrelation Lengths Used in the Twin Experiment

Symbol	Description	Value
σ_q^2	model error variance	1.0×10^{-13} m^2 s^{-4}
σ_a^2	initial error variance	0.0025 m^2 s^{-2}
$\sigma_{b_0}^2$	boundary error variance	3.0×10^{-10} m^4 s^{-4}
$\sigma_{b_H}^2$	boundary error variance	3.0×10^{-10} m^4 s^{-4}
σ_o^2	measurement error variance	2.5×10^{-5} m^2 s^{-2}
σ_A^2	diffusion error variance	6.25×10^{-8} m^4 s^{-2}
$\sigma_{C_d}^2$	wind drag error variance	1.7×10^{-8}
l_q	e-folding scale for model error covariance	$1.0\,\delta_e$
l_a	e-folding scale for initial error covariance	$1.0\,\delta_e$
l_A	e-folding scale for diffusion error covariance	$2.0\,\delta_e$

$$\mathbf{Q}_{qq}(z_1, t_1, z_2, t_2) = \sigma_q^2 \exp\left\{-\left(\frac{z_1 - z_2}{l_q}\right)^2\right\} \delta(t_1 - t_2)\mathbf{I}, \quad (39)$$

$$Q_{AA}(z_1, z_2) = \sigma_A^2 \exp\left\{-\left(\frac{z_1 - z_2}{l_A}\right)^2\right\}, \quad (40)$$

$$Q_{C_d C_d} = \sigma_{C_d}^2, \quad (41)$$

$$\mathbf{w}^{-1} = \sigma_o^2 \mathbf{I}. \quad (42)$$

Here it has been assumed that the model and the boundary errors are uncorrelated in time. This is convenient for computational reasons, but for more realistic applications, such a correlation should probably be included. Values for the variances and the decorrelation lengths are given in Table 2. The error variances all correspond to a 5–10% standard deviation of the variables or terms they represent errors in. This means that all first guesses and the model dynamics are assumed to be reasonably accurate and they have all about the same impact on the inverse solution. Small perturbations in the weights give only small perturbations in the inverse estimate. However, large perturbation may cause problems; for example, with zero weights on some of the first guesses, the inverse problem may become illposed. The decorrelation lengths are similar to the characteristic length scales of the dynamical system. This ensures that the representers also become smooth with similar length scales as the dynamical solution.

To illustrate the solution procedure using the representer method in more detail, the variables $\boldsymbol{\alpha}_5$, $\mathbf{r}_5$, and $\boldsymbol{\lambda}$ and the right-hand sides, $\mathbf{Q}_{qq} \bullet \boldsymbol{\alpha}_5$ and $\mathbf{Q}_{qq} \bullet \boldsymbol{\lambda}$, are given in Figure 2 with u and v components in the left and the right column, respectively. These plots demonstrate how the information from the measurements is taken into account and influences the solution. Measurement number five corresponds to the u component at the location $(Z_5, T_5) = (-20.0, 25.0)$.

The top row shows the components of $\boldsymbol{\alpha}_5$, and it is clear from (26) that the u component of $\boldsymbol{\alpha}_5$ is forced by the δ function at the measurement location. This information is then propagated backward in time, while the u and v components interact during the integration.

The $\boldsymbol{\alpha}_m$ are then used on the right-hand side of the forward equation for the representer and also in the initial and boundary conditions. The convolution $\mathbf{Q}_{qq} \bullet \boldsymbol{\alpha}_5$ tends to smooth the $\boldsymbol{\alpha}_5$ field according to the covariance functions contained in $\mathbf{Q}_{qq}$, as can be observed from the second row in Figure 2.

The representer $\mathbf{r}_5$ is smooth and is oscillating in time with a period reflecting the inertial oscillations described by the dynamical model. Note that the representers will have a discontinuous time derivative at the measurement location since the right-hand side $\mathbf{Q}_{qq} \bullet \boldsymbol{\alpha}_5$ is discontinuous there. However, if a correlation in time was allowed in $\mathbf{Q}_{qq}$, then $\mathbf{Q}_{qq} \bullet \boldsymbol{\alpha}_5$ would be continuous and the representer $\mathbf{r}_5$ would be smooth.

After all the representers have been calculated and measured to generate the representer matrix, the coefficient $\mathbf{b}$ is calculated and used in (15) to decouple the Euler–Lagrange equations. The u and v components of $\boldsymbol{\lambda}$ (Figure 2) illustrate how the various measurements have a different impact determined by values of the coefficients in $\mathbf{b}$, which again are determined by the quality of the first-guess solution versus the quality of the measurements and the residual between the measurements and the first-guess solution. After $\boldsymbol{\lambda}$ is found, the right-hand side in the forward model equation can be constructed through the convolution $\mathbf{Q}_{qq} \bullet \boldsymbol{\lambda}$, and this field is given at the bottom of Figure 2. Clearly, the role of this term is to force the solution to smooth the measurements.

The first-guess, the reference solution, and the inverse estimate are given in Figure 1. The reference solution is regenerated quite well, even though the first-guess solution is out of phase with the reference case and the measurements do not resolve the time period of the oscillation. In fact, a single measurement may suffice for reconstructing the correct phase since the corresponding representer will carry the information both forward and backward in time, although the errors will be larger with less measurements. Note that the quality of the inverse estimate is poorest near the initial time. This is probably caused by a poor choice of weights for the initial conditions relative to the initial condition that was actually used.

The estimation of the diffusion parameter $A(z)$ is illustrated in Figure 3 , where the first-guess $A_0(z)$ and the reference $A(z)$ are shown together with the estimate $\hat{A}(z)$. The weak signal below the Ekman layer makes it difficult to correct an erroneous first-guess of the diffusion parameter in the deep ocean. Note also that the estimate for A does not coincide with the reference diffusion parameter but is located

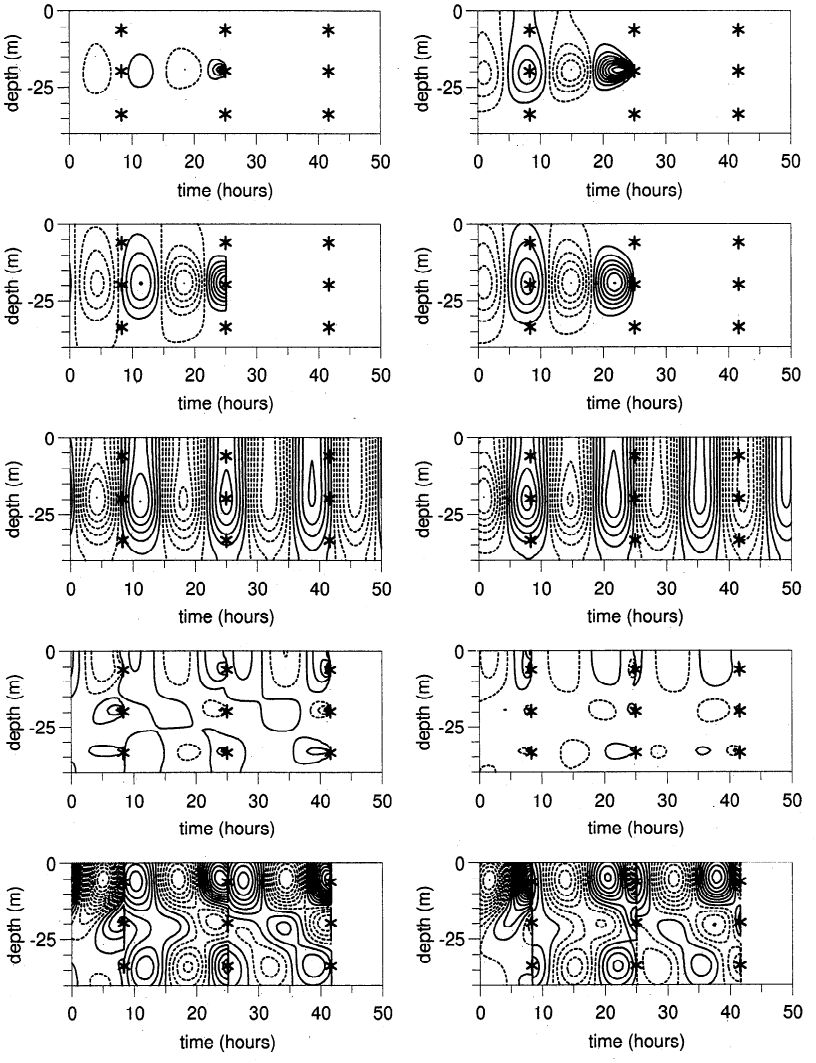

Figure 2. The (left) u and (right) v components of (top to bottom) α_5, $Q_{qq} \bullet \alpha_5$, r_5, the adjoint λ, and $Q_{qq} \bullet \lambda$ from the identical twin experiment. The measurement locations are marked with an asterisk.

somewhere in between the first-guess A_0 and the exact A at most of the depths. In some places, however, the estimate is located to the left of both the first guess and the reference diffusion. This is not unexpected for this nonlinear problem where the minimum of the penalty function does determine a both the inverse solution and estimated parameters simultaneously. Note that these are mutually dependent. The minimizing solution is determined by the prescribed weights and should be located within the corresponding error variances.

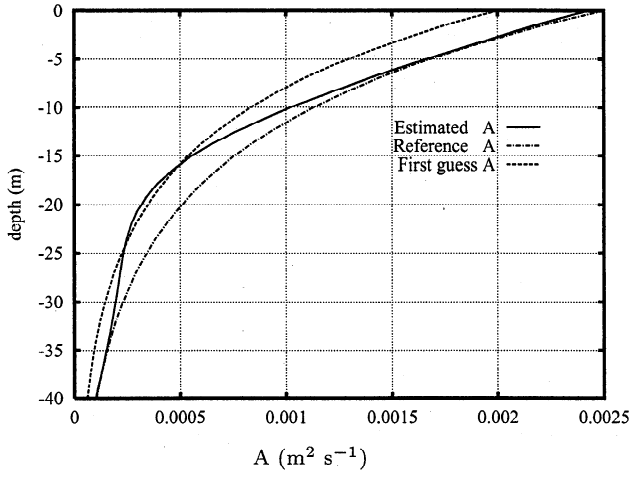

Figure 3. The estimation of the eddy viscosity profile A from the identical twin experiment.

The estimation of the the wind drag coefficient C_d is shown in Figure 4 , and the estimate is a value somewhere in between the first-guess and the reference value. It should at this time be commented on that the estimated values for the unknown parameters found by *Yu and O'Brien* [1991, 1992] did not show any effect from the penalty of the first-guess values. This indicates that zero weights were used for the first guesses. What they actually did was to replace the first-guess value with the current estimate in each iteration. A different variational problem was therefore implicitly assumed in each iteration, and they had in reality no penalty on the first-guess estimates. They therefore allowed for every arbitrary and nonsmooth function to be a solution for $A(z)$. Unless there is enough independent information contained

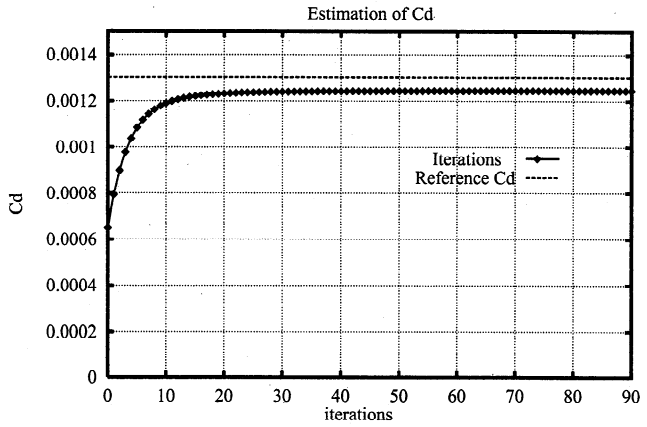

Figure 4. The estimation of the wind-drag coefficient C_d from the identical twin experiment. The number of iterations is given along the x axis.

in the measurements to close this problem, many solutions may be found which interpolate the measurements and give the value zero to the penalty functions.

Assimilation of LOTUS-3 Time Series

The representer implementation will now be examined using the LOTUS-3 data set [*Bowers et al.*, 1986] in a similar setup to the one used by *Yu and O'Brien* [1991, 1992] . The LOTUS-3 measurements were collected in the northwestern Sargasso Sea (34° N, 70° W) during the summer of 1982. Current m fixed at depths 5, 10, 15, 20, 25, 35, 50, 65, 75 and 100 meters measured the in situ currents, and a wind recorder mounted on top of the LOTUS-3 tower measured the wind speeds. The sampling interval was 15 min, and the data used by *Yu and O'Brien* [1991, 1992] were collected in the period from June 30 to July 9, 1982. Here data from the same time period are used. However, while *Yu and O'Brien* [1991, 1992] used all data collected during the 10 days, we have used a subsampled data set consisting of measurements collected at a 5-hour time interval at the depths 5, 25, 35, 50, and 75 m. The reason for not using all the measurements is to reduce the size of the representer matrix $\mathbf{R}$ and thus the computational cost. The inertial period and the vertical length scale are still resolved, and it is expected that mainly small-scale noise is rejected by subsampling the measurements.

The model was initialized by the first measurements collected on June 30, 1982. The standard deviation of the small-scale variability of the velocity observations was estimated to be close to 0.025 m s^{-1}, and this value was used to determine the error variances for the observations and the initial conditions. A similar approach was also used for the surface boundary conditions by looking at small-scale variability of the wind data. The model error variance and the error variances of the two parameters were specified after a few test runs to give a relatively smooth inverse estimate which seemed to be nearly consistent with the model dynamics and at the same time was close to the observations without over fitting them. Some of the physical parameters used are given in Table 3, while values for the variances and the decorrelation lengths are given in Table 4.

The Ekman model describes wind driven currents and inertial oscillations only, while the measurements may also contain contributions from, e.g., pressure-driven currents. Therefore some drift in the measurements has been removed from the deepest moorings as was also done by *Yu and O'Brien* [1991, 1992] .

The results from the inverse calculation are shown in Figures 5 and 6 as time series of the u and v component of the velocity at various depths. The inverse estimate is plotted together with the full time series of the measurements. The measurements which were used in the inversion are shown as diamonds.

It is first of all evident that the inverse estimate is close to the measurements at all times and depths and also at 10 m where no measurements were assimilated. Both the amplitude and phase are in good agreement with the measurements at all depths. Note also that the inverse estimate is smooth and does not interpolate the measurements; however, it is close to the measurements at all depths. By a closer examination of the inverse estimate, it is possible to see that the time derivative of the inverse estimate is discontinuous at measurement locations. This is caused by neglecting the time correlation in the model error covariances.

The posterior error variances $\mathbf{C_{uu}}(z,t,z,t)$ can be calculated from (32), and normally, the prior error variances $\boldsymbol{\Gamma}(z,t,z,t)$ are estimated from a statistical simulation, [e.g., *Evensen*, 1994b] . However, note that from (30) and (33),

$$\mathbf{R} = \mathcal{L}\mathcal{L}^T[\boldsymbol{\Gamma}]. \tag{43}$$

The use of direct measurements implies that the diagonal of the representer matrix $\mathbf{R}$ will contain the error variances at the measurement locations. Having a high density of measurements and assuming a smooth prior error variance field in space and time makes it possible to interpolate the prior error variances at measurement locations to the full space-time grid. Thus the posterior error variances are readily calculated without much computational effort.

In Figure 7, the prior and posterior error variances for the u component of the horizontal velocity $\mathbf{u}$ at 5 m and 50 m are shown. What may be observed is that the prior errors are growing in time as a result of incorporating model errors. The posterior error variances have the characteristic structure one would expect from a smoother solution for linear dynamics, with minimum values at the measurement locations [e.g., *Bennett and Budgell*, 1989] . The error estimates are given for the final iteration in the estimation of the diffusion A and the wind drag C_d and are thus those corresponding to the best estimates of the parameters.

Table 3. Physical Parameters Used in the LOTUS-3 Data Assimilation Experiment

Symbol	Description	Value
T	integration time	240 hours
Δt	time step	30 min
H	depth	100 m
Δz	grid size	0.5 m
f	Coriolis parameter	0.8×10^{-4} s^{-1}
δ_E	Ekman layer thickness	2.71 m
ρ_a	density of air	1.2 kg m^{-3}
ρ_w	density of water	1.025×10^3 kg m^{-3}

Table 4. Dimensional Values for the Variances and the Decorrelation Lengths Used in the LOTUS-3 Data Assimilation Experiment

Symbol	Description	Value
σ_q^2	model error variance	5.0×10^{-12} m^2 s^{-4}
σ_a^2	initial error variance	6.25×10^{-4} m^2 s^{-2}
$\sigma_{b_0}^2$	boundary error variance	3.0×10^{-10} m^4 s^{-4}
$\sigma_{b_H}^2$	boundary error variance	3.0×10^{-10} m^4 s^{-4}
σ_o^2	measurement error variance	6.25×10^{-4} m^2 s^{-2}
σ_A^2	diffusion error variance	0.5×10^{-8} m^4 s^{-4}
$\sigma_{C_d}^2$	wind drag error variance	1.7×10^{-8}
l_q	e-folding scale for model error covariance	$2.2\delta_e$
l_a	e-folding scale for initial error covariance	$2.2\delta_e$
l_A	e-folding scale for diffusion error covariance	$5.0\delta_e$

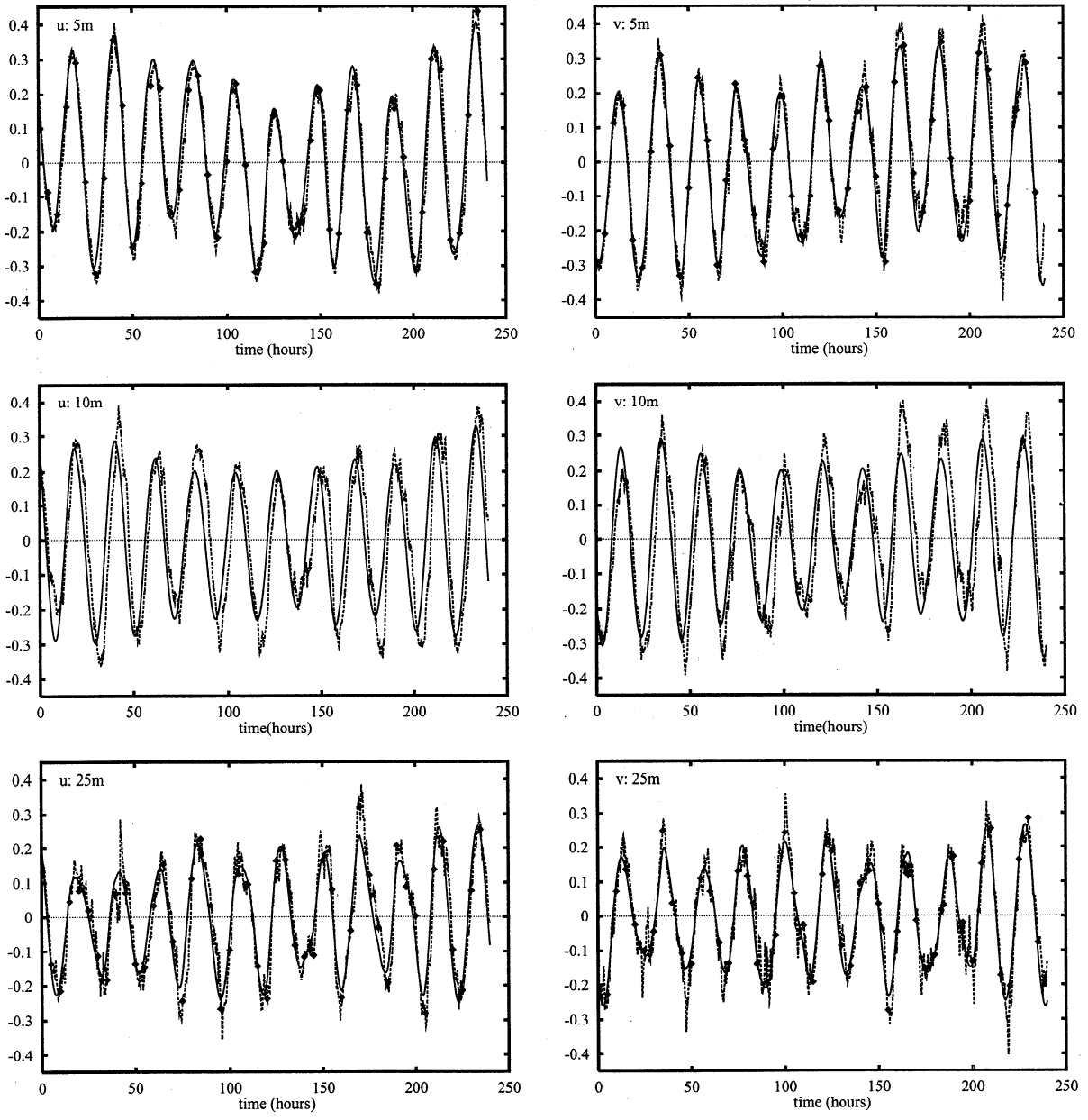

Figure 5. Weak constraint inverse estimate (solid line), the time series of measurements (dashed lines), and the subsampled measurements (diamonds) at 5, 10, and 25 m from the LOTUS-3 data assimilation experiment. The u and v components (meters per second) are shown in the left and right columns, respectively.

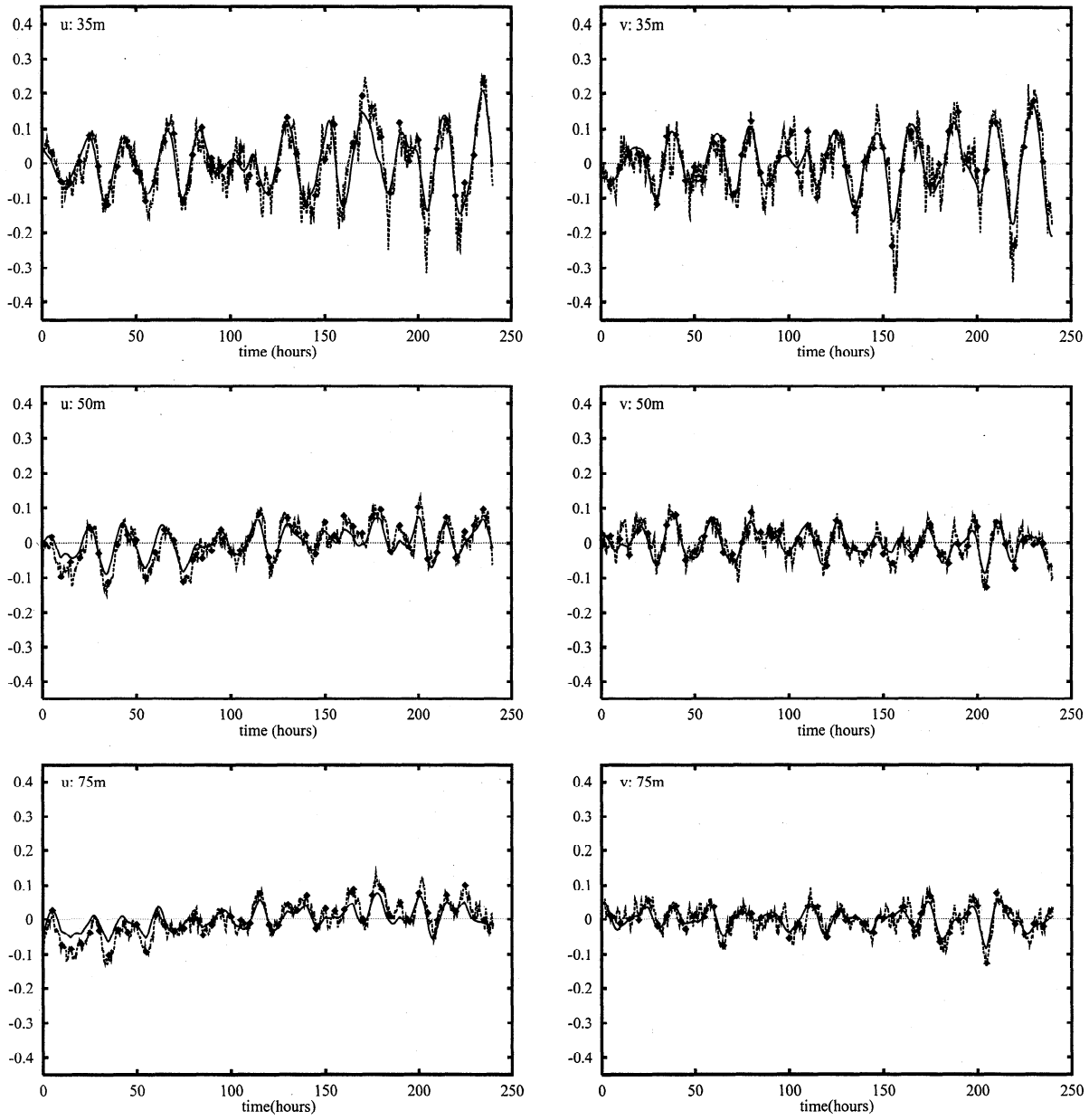

Figure 6. Same as Figure 5, but for the depths 35, 50, and 75 m.

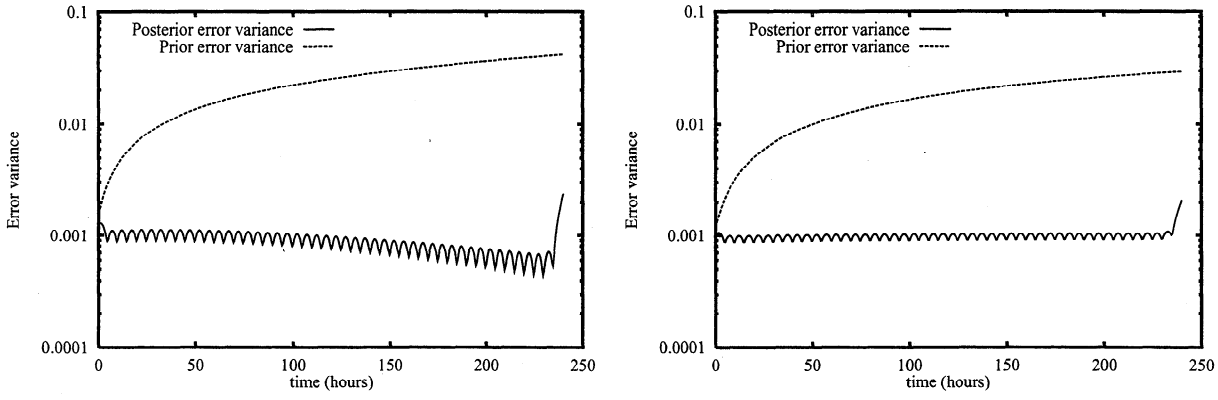

Figure 7. The error variance for the u component of the inverse estimate at (left) 5 m and (right) 50 m in the weak constraint case.

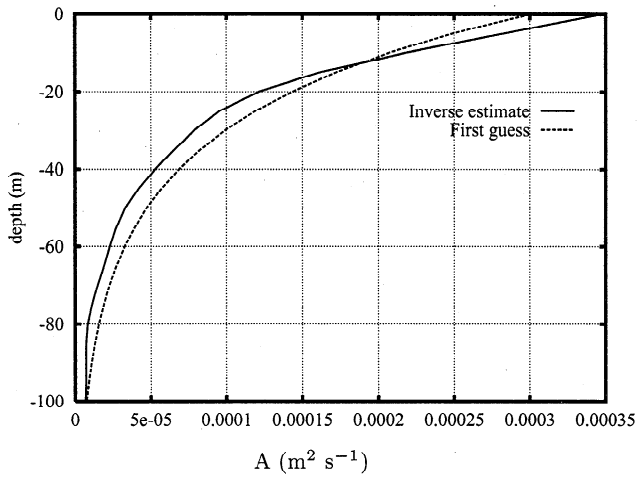

Figure 8. Results of the estimation of the eddy viscosity profile A from the LOTUS-3 data assimilation experiment.

The results of the estimation of the vertical diffusion parameter are shown in Figure 8. It generally decreases at all depths from the first-guess profile, e.g., at the surface from the first-guess value 0.0003 to 0.00035 m^2 s^{-1}. The corresponding value found by *Yu and O'Brien* [1991] starting from a first-guess of 0.00001 m^2 s^{-1} was 0.0029 m^2 s^{-1}.

The wind drag coefficient C_d converges rapidly from the first-guess value 0.001 to the value 0.00068 as shown in Figure 9. The value found by *Yu and O'Brien* [1991] was 0.00126, starting from 0.00134.

There is a substantial difference in the values for the diffusion coefficient and the wind drag obtained in the two works. It should be noted that while *Yu and O'Brien* [1991] replaced the first-guess values, C_{d_0} and $A_0(z)$, with the current estimate in each iteration of (18) and (19), these were kept constant in our calculations. Clearly, *Yu and O'Brien* [1991] solved a different inverse problem in each iteration. Actually, it is not clear from their figures that their iterations did converge.

By comparing our results for the diffusion parameter and wind drag coefficient with results found by, e.g., *Price et al.* [1987], who inferred an effective viscosity $A = 0.006$ m^2 s^{-1} by separating the wind-driven current from the measured LOTUS-3 current and averaging over the whole period, it may seem as if our final diffusion and wind- drag are too small. Note that the inverse estimate for the model state is in good agreement with the observations also in the deep ocean, even if the estimated diffusion may be to small to produce the correct velocities in a model simulation without assimilation. Thus the observations in the deep ocean may account for the use of a too low vertical diffusion coefficient. On the other hand, the diffusion parameter is within reasonable values given from simple turbulence closure schemes. A shear profile estimated from the observations results in a typical shear of 0.003–0.004 s^{-1} from the $z = 0$ m to $z = -25$ m; then the shear increases to about 0.007 s^{-1} from $z = -25$ m to $z = -35$ m, before it decreases to a much lower value at greater depths. Using a shear $\mathcal{S} \simeq 0.005$ s^{-1} and a diffusion coefficient $A \simeq 0.00025$ m^2 s^{-1}, in the expression $A = \ell^2 \mathcal{S}$ which relates shear and diffusion through the Prandtl turbulence mixing length [*Speziale*, 1991], results in a value of the Prandtl mixing length of $\ell \simeq 4.5$ m, which is of the same order as the characteristic Ekman layer thickness (see Table 3).

For comparison a strong constraint inversion was performed and the results are shown in Figures 10 and 11. The final estimates of the diffusion A and the wind drag C_d from the weak constraint inversion were used. Note that the strong constraint inverse for a linear model is easily solved for without any iterations simply by calculating the representer solution with the model error covariance set to zero.

It is clear from comparisons that the strong constraint solution in the upper part of the ocean is in reasonable phase with the measurements, as determined by the initial conditions, while the amplitudes are not as good as in the weak constraint inverse. The only way the amplitudes can change when the model is assumed to be perfect is by vertical transfer of momentum from the surface. This is seen to work reasonably well near the surface, while in the deeper ocean, there is hardly any effect from the wind stress and the strong constraint inverse solution is also far from the measurements. The solution is actually rather close to a sine curve representing the pure inertial oscillations. The strong constraint results from *Yu and O'Brien* [1992] are similar to ours and also have the same problems with amplitude and phase. These results indicate that model deficiencies, such as neglected physics, should be accounted for through a weak constraint variational formulation to ensure an inverse solution in agreement with the measurements.

Summary

A combined state estimation and parameter estimation problem has been formulated for a one-dimensional Ekman model. The formulation of the inverse problem and the solution method which is based on the representer method has been outlined and discussed in some detail. The inclusion of unknown physical parameters as control variables yields a nonlinear inverse problem even if the model itself is linear. It has been illustrated how an iterative technique in combination with the representer method can be used to estimate poorly known parameters in the model. This was done by defining an iteration for the unknown parameters in

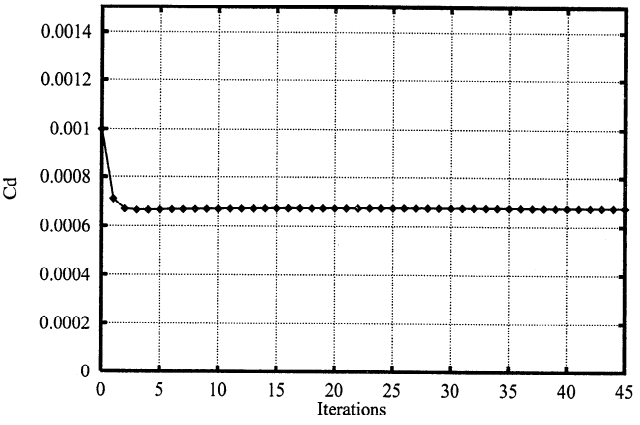

Figure 9. Results of the estimation of the wind-drag C_d from the LOTUS-3 data assimilation experiment.

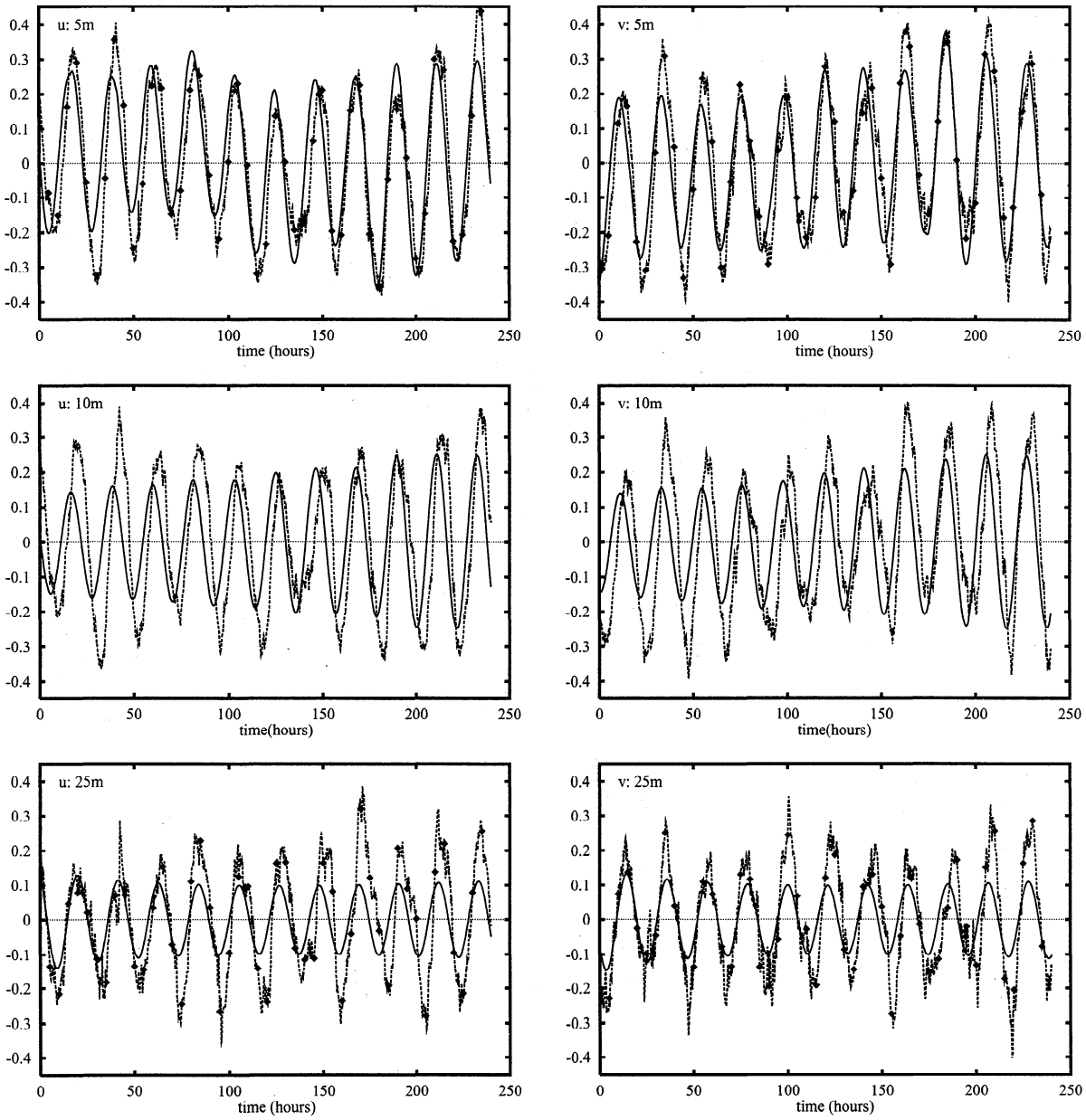

Figure 10. Strong constraint inverse estimate (solid line), the time series of measurements (dashed lines), and the subsampled measurements (diamonds) at 5, 10, and 25 m from the LOTUS-3 assimilation experiment. The u and v components (meters per second) are shown in the left and right columns, respectively.

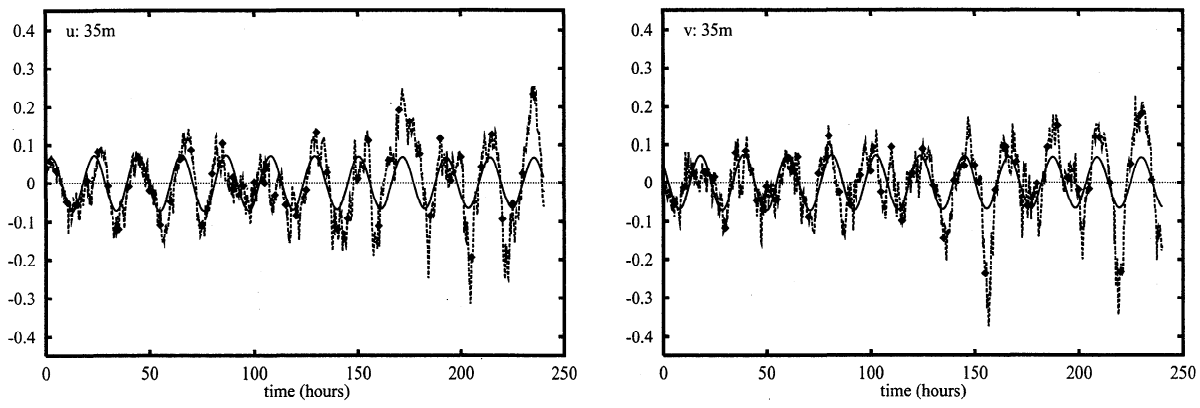

Figure 11. Same as Figure 10 but for the depths 35, 50, and 75 m.

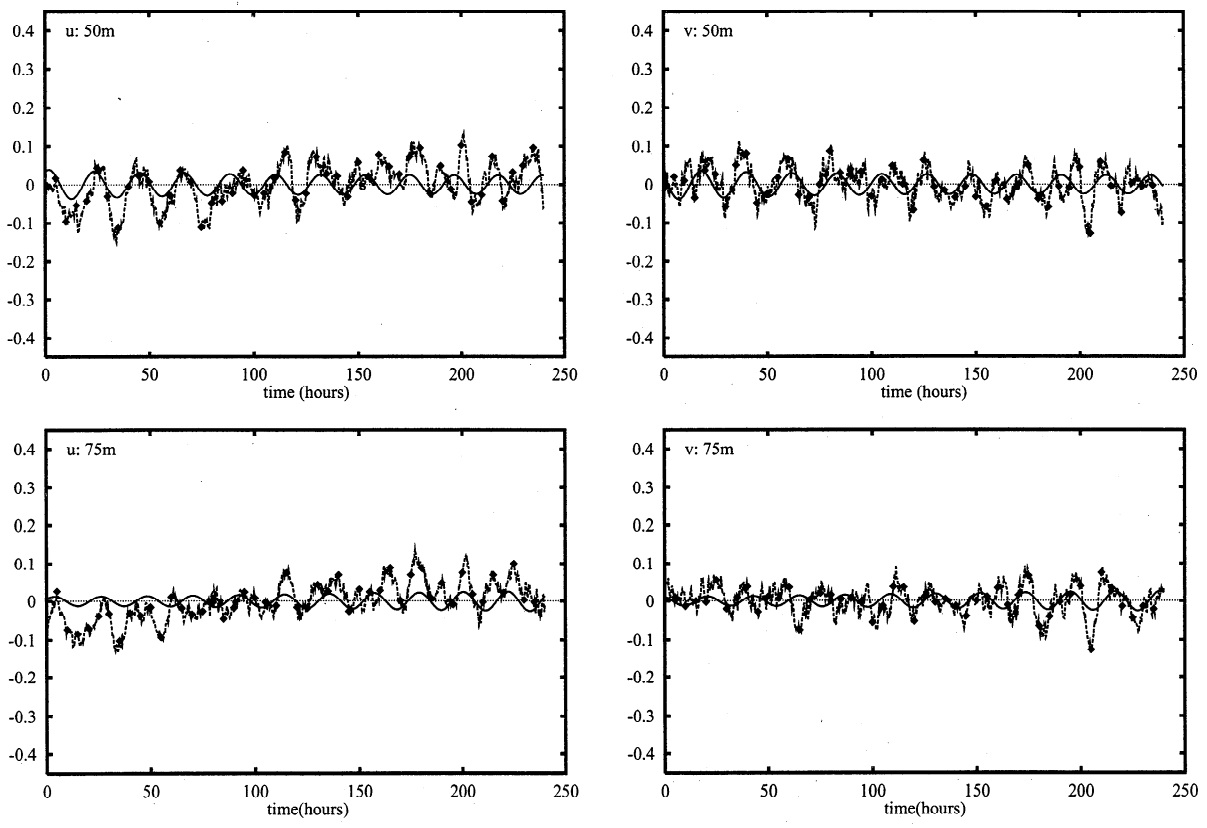

Figure 11. (continued)

the Euler–Lagrange equations and then solving each linear iterate exactly using the representer method. In addition to illustrating the method on a simple twin experiment, a comparison was also made with the strong constraint solution found by *Yu and O'Brien* [1991, 1992], where measurements from the LOTUS-3 data set were assimilated. It was shown that the weak constraint inverse solution was in good agreement with the observations and that it was superior to the strong constraint inverse.

Acknowledgments. Mette Eknes was supported by a grant from the Nansen Center's fellowship and research fund, and Geir Evensen was supported by the Norwegian Research Council.

References

Bennett, A. F., *Inverse Methods in Physical Oceanography*, Cambridge Univ. Press, New York, 1992.

Bennett, A. F., and J. R. Baugh, A parallel algorithm for variational assimilation in oceanography and meteorology, *J. Atmos. Oceanic Technolo.*, *9*, 426–433, 1992.

Bennett, A. F., and W. P. Budgell, The Kalman smoother for a linear quasigeostrophic model of ocean circulation, *Dyn. Atmos. Oceans*, *13*(3-4), 219–267, 1989.

Bennett, A. F., and R. N. Miller, Weighting initial conditions in variational assimilation schemes, *Mon. Weather Rev.*, *119*, 1098–1102, 1990.

Bennett, A. F., and M. Thorburn, The generalized inverse of a nonlinear quasigeostrophic ocean circulation model, *J. Phys. Oceanogr.*, *22*, 213–230, 1992.

Bennett, A. F., L. M. Leslie, C. R. Hagelberg, and P. E. Powers, Tropical cyclone prediction using a barotropic model initialized by a generalized inverse method, *Mon. Weather Rev.*, *121*, 1714–1729, 1993.

Bowers, C. M., J. F. Price, R. A. Weller, and M. G. Briscoe, Data tabulations and analysis of diurnal sea surface temperature variability observed at LOTUS, *Tech. Rep.* 5, Woods Hole Oceanogr. Inst. Woods Hole, Mass., 1986.

Evensen, G., Inverse methods and data assimilation in nonlinear ocean models, *Physica D*, *77*, 108–129, 1994a.

Evensen, G., Sequential data assimilation with a nonlinear quasi-geostrophic model using Monte Carlo methods to forecast error statistics, *J. Geophys. Res.*, *99*(C5), 10,143–10,162, 1994b.

Price, J. F., R. A. Weller, and R. Schudlich, Wind-driven ocean currents and Ekman transport, *Science*, *238*, 1534–1538, 1987.

Smedstad, O. M., and J. J. O'Brien, Variational data assimilation and parameter estimation in an equatorial Pacific Ocean model, *Prog. Oceanogr.*, *26*, 179–241, 1991.

Speziale, C. G., Analytical methods for the developments of Reynolds-stress closures in turbulence, *Annu. Rev. Fluid. Mech.*, *23*, 107–157, 1991.

Tarbell, S. A., N. J. Pennington, and M. G. Briscoe, A compilation of moored current meter and wind recorder data: Long-term upper ocean study (LOTUS), May 1982–April 1983., *Tech. Rep.* 36, Woods Hole Oceanogr. Inst. Woods Hole, Mass., 1984.

Yu, L., and J. J. O'Brien, Variational estimation of the wind stress drag coefficient and the oceanic eddy viscosity profile, *J. Phys. Oceanogr.*, *21*, 709–719, 1991.

Yu, L., and J. J. O'Brien, On the initial condition in parameter estimation, *J. Phys. Oceanogr.*, *22*, 1361–1364, 1992.

M. Eknes and G. Evensen, Nansen Environmental and Remote Sensing Center, Edvard Griegsvei 3A, N-5037 Solheimsviken, Bergen, Norway. (e-mail: Mette.Eknes@nrsc.no)

(Received January 17, 1996; revised August 23, 1996; accepted October 7, 1996.)