Journal of Marine Research, 62, 787–814, 2004

Passive tracer reconstruction as a least-squares problem with a semi-Lagrangian constraint: An application to fish eggs and larvae

by G. G. Panteleev[1,2], B. deYoung[1,3], C. S. Reiss[4] and C. T. Taggart[4]

ABSTRACT

A variational, data assimilation, algorithm was developed for reconstruction of a two-dimensional, nonstationary, passive tracer field in the ocean with open boundaries and a known velocity field. The observations, spatial smoothing terms and passive tracer conservation equation were included as weak constraints. The algorithm was tested with simulated, nonstationary, pseudo-oceanographic data integrated for a 7-day period. Simulations were run to determine the robustness of the algorithm and the effect of theoretical, simulated 'sampling events,' mimicking the standard oceanographic survey. We explored the sensitivity of the reconstructed tracer fields to the distribution of the pseudo-oceanographic sampling strategy, essentially an antenna problem, and to errors in the velocity field and the observations. The algorithm was applied to observations of silver hake (*Merluccius bilinearis*) eggs and larvae obtained in August 1998 on the Scotian Shelf. Finally, the evolution of fish eggs and larvae concentration was found. The corresponding mortality rate of fish eggs and larvae was determined to be 0.28 day^{-1}, with errors of 0.03 day^{-1}. The approach, which is quite general and could be applied to many different problems requiring minimization subject to constraints, allows for error analysis of the results.

1. Introduction

The volume and variety of data collected in the ocean has substantially increased during the last decade. Besides traditional measurements of temperature and salinity, new observational and sampling techniques have enabled the collection of large quantities of biological and hydrochemical data that can be treated as passive tracers as they are not strongly coupled with the ocean circulation. The collection of these new data raises the problem of how to transfer these data, usually sampled on irregular grids, onto a regular grid, necessary for numerical analysis. One of the key interpretive questions is how to relate the observed pattern of the scalar or tracer fields to the circulation field which may be determined from different observations or derived from a numerical model. Some of these

1. Department of Physics and Physical Oceanography, Memorial University of Newfoundland, St. John's, A1B 3X7, Canada.
2. Present address: Shirshov Institute of Oceanology, Nakhimovsky Prospect 36, Moscow 117218, Russia.
3. Corresponding author: ***email: bdeyoung@physics.mun.ca***
4. Department of Oceanography, Dalhousie University, Halifax, Nova Scotia B3H 4J1, Canada.

problems are simply numerical issues of spatial analysis (e.g. Daley, 1991) but others have more to do with understanding the limitations of the tracer and circulation data. That understanding may be formulated via mathematical models which can be used for the analysis.

One illustrative problem for which these questions became apparent is in the reconstruction of the abundance of fish eggs and small larvae when samples are taken in a limited region of the ocean with open boundaries. Obviously (for example Hjort, 1926) the physical environment strongly influences the distribution of eggs and larvae and has a substantial effect on growth and survival, effecting the concentration of organisms observed at different times.

The simplest model equation, which describes the evolution of the concentration of fish eggs and larvae, is the advection-reaction equation. This equation includes the direct effect of circulation on the distribution and the influence of natural mortality on the concentration. Behavior (growth, vertical migration or horizontal swimming) of the different organisms may be rather complicated and unknown, and it seems natural to simplify the problem and use the advection-reaction equation, with mortality rate (m) alone, as a first step in predicting the abundance and distribution of animals in time. Such a model description may be relevant to the more general problem of determining the influence of the circulation on any passive tracers, such as hydro-chemical water properties, water pollution etc. In this paper, we concentrate our attention on biological passive tracers (e.g. plankton, fish eggs and small larvae).

On the continental shelf, biological sampling of eggs and larvae taken as part of an oceanographic survey can require cruises that last days to weeks. If a mathematical model can describe the evolution of the passive tracer field during the period of sampling, then the reconstruction of passive tracer fields from the observations can be viewed as a standard inverse problem with constraints implied by the model (Marchuk, 1975; Wunsch, 1978; Thacker and Long, 1988). It can be formulated as either a "weak" or "strong" constraint problem depending on the accuracy of the chosen model (Sasaki, 1970). In the absence of any interaction between the different passive tracers and with a known velocity distribution, then the advection equation is linear. It is helpful that we can linearize the problem, particularly since the biological data are usually sampled with substantial errors and nonlinear effects may not be resolved by these data.

Coupling the physical circulation to the passive tracer equation, even for the simple linear transport problem, can be difficult if the goal is a reliable simulation at long time scales. Of course, long is relative, but in the coastal environment even a few days can present a substantial challenge (Helbig and Pepin, 1998). The solution of the inverse problem (Wunsch, 1996) implies additional difficulties that are derived from the minimization of the cost function over the range of the initial/boundary conditions and other independent parameters of the model. These problems require careful choice and preliminary analysis of the numerical formulation of the model and the corresponding inverse problem. In this paper, we present a data assimilation scheme that incorporates a

variational algorithm for the reconstruction of the distribution of passive tracers. Our algorithm uses a Lagrangian approach that helps to minimize the number of independent variables. Reducing the number of variables helps to simplify the minimization problem and permits possible estimation of *a posteriori* errors through the direct inversion of the dense Hessian matrix (Thacker and Long, 1988).

We tested our approach with simulated data and explored the sensitivity of the variational approach to the distribution of "stations" in our simulated data and to the variance of the error in the observations and the velocity field. We show that the approach is fairly robust and can provide estimates of mortality that exhibit small error rates relative to the variability in abundance estimates. Finally, we applied the technique to data for silver hake eggs and larvae collected during two surveys (01 August 1998 and 07 August 1998) from Western Bank on the Scotian Shelf (Fig. 1, Reiss *et al.,* 2000).

Although we present the technique with an application for fish eggs and larvae, with minimal modifications it could be applied to many other tracer problems, e.g. oil spill dispersal, fluorocarbon dispersal, and others.

The paper is organized as follows. In the next section we outline the assumptions that underlie the proposed algorithm. In Section 3, we develop the model to be applied here and present the variational algorithm. In Section 4 we explore the sensitivity of the proposed method to the selection of different parameters. Then (Section 5), we apply the result to fish eggs and larvae from the Scotian Shelf to determine the natural rate of mortality. We end with conclusions and some discussion.

2. Assumptions

We begin with a discussion of the assumptions and develop a hypothesis that we believe to be pertinent to the passive tracer problem. Proper specification of the problem simplifies our task and limits the subjectivity of the algorithm as we apply it to real data. Naturally, these particular two-dimensional assumptions are primarily relevant to our application on the Scotian Shelf.

The first biological assumption is that the distribution of the passive tracer (e.g. fish eggs or larvae) is essentially two dimensional in space. That is, the passive tracer distribution is restricted to the same vertical layer, and periodic or cyclic vertical movement of the tracer is not an important determinant of advection at short (from days to weeks) time scales. On Western Bank (Fig. 1) this assumption has been validated for a variety of fish species including cod, silver hake and capelin (Fortier and Villeneuve 1996; Taggart *et al.,* 1996; Lochmann *et al.,* 1997; Reiss *et al.,* 2000). In particular the eggs and small (smaller than 5 mm length) larvae of silver hake are most abundant in the upper 20 to 40 m, above the pycnocline (Reiss *et al.,* 2000). Any loss from this layer will appear as a mortality in our calculation.

In this study, fish eggs and larvae were collected by tows of a standard 61 cm bongo net (333 μ mesh) that was fished from the surface to 50 m or the pycnocline and then back to the surface (oblique). The velocity field was derived from measurements conducted at

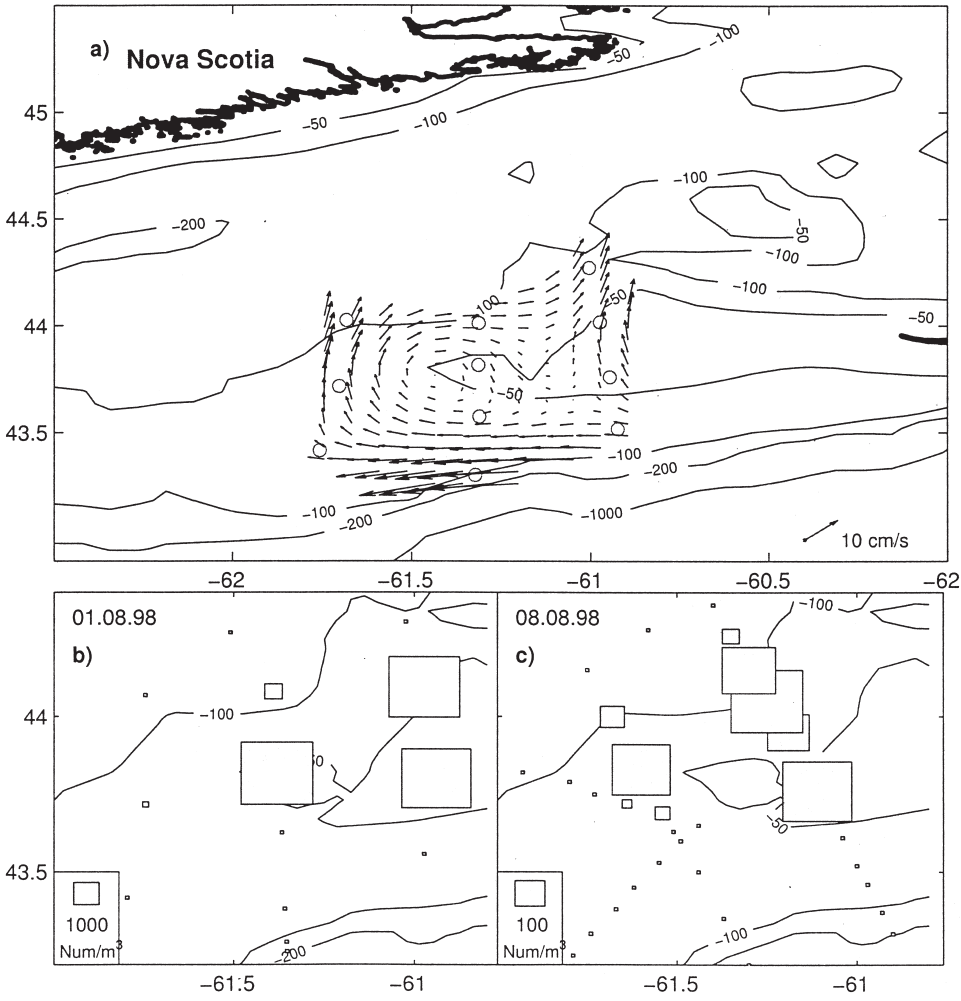

Figure 1. Upper panel: (a) Local geography and mean interpolated velocity field for the period 01–08 August 1998. Lower panel: Expanding symbols of the concentration (numbers/m^3) and distribution of silver hake (*Merluccius bilinearis*); (a) eggs and recently hatched (smaller than 3 mm TL) larvae on Western Bank on 1 August 1998 (first cruise); (b) from 3 to 4 mm larvae collected during the second cruise on August 8th (second cruise).

30 m (Panteleev *et al.,* 2001) that roughly correspond to the average depth of the oblique tows. Therefore, we will assume that the passive tracers (fish eggs and larvae) are distributed in a two-dimensional horizontal plane of water above the thermocline.

The second assumption is that the mortality of the eggs and larvae do not depend on location within the survey or the time between the two surveys (roughly seven days) since we have no way to determine whether such variability exists. In general, previous estimates

of larval mortality range from 0.1 to 0.4 day^{-1} (Houde, 1996). On the Scotian Shelf, Taggart *et al.* (1996) tracked a patch of larvae for 25 days and estimated the mortality rate for cod and silver hake as 0.2 day^{-1} using an exponential mortality model. This estimate provides a useful first guess for our analysis.

Many of the processes of larval transport and evolution are difficult to define or are poorly known, e.g. the natural mortality and diffusion. It is, therefore, difficult to generate meaningful solutions for this problem, even if we know one or another of the terms, e.g. the velocity field, quite well. We will assume that our model describes the evolution of the passive tracer with some accuracy. This third assumption allows us to formulate the data assimilation algorithm through the "weak" constraint (Sasaki, 1970) rather than applying it in a "strong," fixed, deterministic fashion.

The fourth assumption requires us to define the typical scale of the passive tracer field. We need to define that scale in order to introduce "bogus data" (Thacker and Long, 1988) into the assimilation algorithm. The reason for adding bogus data is the lack of observations and the necessary regularizing of the data assimilation problem through the prior hypothesis that defines the spatial smoothness of the larval distribution. Otherwise the problem will be underdetermined. According to Lochmann *et al.* (1997) and Taggart *et al.* (1996), the typical spatial scale of the larval distribution on Western Bank is about 10 km. Our measurements of the larval concentration made on the northern flank of Western Bank (08 August 1998, Fig. 1c) show that this spatial scale may be as much as two times greater. Thus, we defined the length scale $L \approx 20$ km, taking into account the relatively sparse sample grid during the first survey (01 August 1998, Fig. 1b).

The fifth assumption is that the tracer (eggs and larvae) concentration can be treated as a stochastic δ-correlated function with unknown means. This allows us to simplify the inverse algorithm described below and formulate it in terms of the Gaussian probability distribution (Wunsch, 1996). In general, we can use information about the spatial cross-correlation function, but its estimate demands large amounts of data, and given the generally limited data available, this function is often known rather imperfectly. It is clear that we need to develop approaches that will work in the absence of 'perfect' knowledge of the spatial fields for which we have data.

3. Methodology

a. The model

In accord with the assumptions discussed above for the description of the passive tracer field (concentration of fish eggs and larvae), we used the following two-dimensional equation:

$$\partial C/\partial t + \nabla \cdot (\mathbf{u}C) + mC - D_C \Delta C = 0 \tag{1}$$

where C is the larval concentration, D_C the diffusion coefficient and m the mortality coefficient. With a known velocity field $\mathbf{u}$, and given initial $C(x, y, t = 0)$ and boundary

$C(b, t)$ conditions, the solution $C(x, y, t)$ of this equation can be determined with traditional finite-difference methods (Roache, 1998).

We know that the solution $C(x, y, t)$ can be approximated with a set of some functions C_k advected by the flow field **u** as a Lagrangian particle, moving with the fluid elements. This approximation may be written as

$$C(x, y, t) = \sum_K C_k. \tag{2}$$

We will call C_k a function-particle. In case of zero mortality ($m = 0$) and diffusion $D_C = 0$, the simplest choice of C_k will be a function defined as a constant of some radius scale and zero elsewhere. But if mortality and diffusion are nonzero, as is generally the case, then we must account for these processes in our determination of the form of the function-particles. So, we define C_k as the analytical solutions to the diffusion equation in cylindrical coordinates:

$$\partial C/\partial t = D_C(\partial^2 C/\partial r^2 + r^{-1}\partial C/\partial r) - mC \tag{3}$$

$$C_k = C_k^0(1/(t + t_0)) \exp(-r^2/4D_C(t + t_0)) \exp(-mt) \tag{4}$$

where r is a radius and t_0 is some period of time introduced to avoid infinite values of the concentration at the initial time $t = 0$. That period can be also interpreted as the time needed for the initial growth and diffusion of eggs spanning the initial point. For our runs, which are seven days long, we used $t_0 = 8$ days and $D_C = 5 \times 10^5$ cm^2 s^{-1}. The horizontal diffusion coefficient was estimated from a number of trajectories of the particles started and integrated in the local nonstationary velocity field using the approach utilized by Sanderson and Booth (1991) and Bauer *et al.* (1998). Our results revealed an error of approximately 20–30% in the definition of the horizontal diffusion coefficient. According to (4), this inaccuracy leads to an increase of 2–4 km in the spatial scale $(2 \times D_C \times (t_0 + t))^{0.5}$ of the function C_k, while the typical advection over one week gives a scale of 30–40 km. Therefore, the influence of possible errors in the horizontal diffusion coefficient is relatively unimportant. Our solution should not be strongly sensitive to errors in the diffusion coefficient over a period of seven days.

With a known velocity field **u** and a specified set of function-particles with initial amplitudes C_k^0, Eqs. (2–4) allow us to determine an approximate solution to the passive tracer evolution described by Eq. (1). The solution will be a function of the initial amplitudes C_k^0 and the mortality coefficient m, which will form a vector of independent parameters for our model (C_k^0, m).

The accuracy of the proposed approximation will depend on the initial number of function-particles and the length of the integration time. We do not provide a theoretical analysis of the relation between the number of function-particles and quality of the approximation (i.e. Eqs. 2–4), but do run simulations to show that for real-world application, we can start with an appropriate number of function-particles and obtain

satisfactory results. Mathematically, this procedure is similar to finding the solution to Eq. (1) by separating the physical processes, as in operator splitting (Press *et al.,* 1992).

b. Inverse algorithm

The solution to Eq. (1) by the superposition of the individual solutions to Eqs. (2–4) allows us to formulate the inverse problem in terms of the corresponding vector of independent parameters for our model (C_k^0, m). Since the solution to Eqs. (2–4) is approximate, we formulate the inverse problem in terms of 'weak' constraints (Sasaki, 1970). We minimize the following cost function

$$\mathbf{F} = \sum_{i=1}^{4} \mathbf{f_i}, \tag{5}$$

$$\mathbf{f_1} = \sum_{n=1}^{N} g_1(x_n, y_n)(C - C_n^*)^2, \tag{6}$$

$$\mathbf{f_2} = \int_{S,t=0,t=T} g_{21}(x, y, t)(C_{xx}^2 + C_{yy}^2)ds + \int_{B,T} g_{22}(b, t)C_{bb}^2 + g_{23}(B, t)C_{tt}^2 dbdt, \tag{7}$$

$$\begin{aligned}\mathbf{f_3} = &\int_{S,t=0} g_3(x, y)[\partial C/\partial t + \nabla(\mathbf{u}C) + mC - D_C\Delta C]^2 ds \\ &+ \int_{b,t} g_3(x, y)[\partial C/\partial t + \nabla(\mathbf{u}C) + mC - D_C\Delta C]^2 dbdt,\end{aligned} \tag{8}$$

$$\mathbf{f_4} = \sum_{k=1}^{K} g_4 \exp(-\gamma C_k^0), \tag{9}$$

where C is the concentration, K is the number of particle-functions (Eqs. 2–4), C_n^* are the larval measurements at the locations with coordinates (x_n, y_n), and N the number of measurements. S, B and T indicate the two-dimensional space, one-dimensional boundary and time domain respectively. Each function $\mathbf{f_i}$, $i = 1, 2, 3$ plays its own role and respectively provides the necessary agreement to the data, smoothness and dynamical correspondence of the minimization function. The function $\mathbf{f_4}$ helps to avoid nonphysical negative values of larval concentration. It is possible to exclude the function $\mathbf{f_2}$ from the cost function $\mathbf{F}$ if the number of data N is equal to the number of particle-functions K. Otherwise, penalizing the smoothness plays the role of the 'bogus' data (Thacker and Long, 1988). g_4 and γ are constant coefficients. g_1, g_{21}, g_{22}, g_{23}, g_3 are weight functions

that according to control theory (Wunsch, 1996) are the inverse covariance of the corresponding physical data.

The normal way to specify these covariance functions is through preliminary analysis of the observations and incorporation of prior knowledge of the relevant dynamical processes. For the applications presented below, we used the simplest ideas about the quality of the observations and typical features of the larval distribution. We specified the relative measurement error ξ_{data}, the time scale T_C, the space scale L_C, the time-dependent mean concentration $\overline{C^*_{t_n}}$ and the maximum initial and final concentrations C_{in}, C_{fi}. Using these parameters we defined

$$g_1(x_n, y_n) = (\overline{C^*_{t_n}}\xi_{data})^{-2}, \tag{10}$$

$$g_{21}(x, y, 0) = (L_C^2/C_{in})^2,\ g_{21}(x, y, T) = (L_C^2/C_{fi})^2, \tag{11}$$

$$g_{22}(b, t) = (2L_C^2/(C_{in} + C_{fi}))^2,\ g_{23}(b, t) = (2T_C^2/(C_{in} + C_{fi}))^2. \tag{12}$$

The introduction of initial C_{in} and final C_{fi} concentrations and the time dependent mean concentration $\overline{C^*_{t_n}}$ were necessary because high mortality rates influence the absolute values and as a consequence the spatial smoothness of the larval concentration. The sensitivity of the different variational approaches to errors in the spatial scale L_C has been investigated in numerous papers (e.g. McIntosh, 1990; Panteleev *et al.,* 2002). These analyses show that reasonable errors in the specification of the spatial scale L_C do not influence the results. As discussed by McIntosh (1990), this is one of the more attractive feature of the variational data assimilation approach.

The definition of g_3 is rather subjective especially as we do not know how well Eq. (1) describes the relevant processes. Taking into account typical errors of $\xi_{eq} = 0.2$ for the approximate solution of (1) with (2–4), we defined

$$g_3 = (\xi_{eq} * 0.5 * (C_{in} + C_{fi})/T_C)^{-2}. \tag{13}$$

g_4 and γ have no physical meaning and were defined empirically based upon the results of sensitivity testing.

With known mortality m, the cost function $\mathbf{F}$ depends only upon the control vector $q = (C_k^0)^T$ consisting of the concentration C_k^0 of initial and inflow particles. The mortality coefficient may also be treated as an unknown variable and in that case $q = (C_k^0, m)^T$. Mortality will be estimated from measurements of concentration directly. If we do have estimates of mortality it seems natural to add the additional functions

$$\mathbf{f_5} = g_5(m - m^*)^2, \tag{14}$$

$$\mathbf{f_6} = g_6 \exp(-\beta m), \tag{15}$$

whose goal is to attract m to the prior estimated value and to avoid nonphysical negative values of mortality.

This constraint essentially limits our search to regions where we expect m to lie. g_5, g_6

and β can be defined as g_1, g_4 and γ, respectively. As will be shown below, specification of the mortality as a part of the control vector offers the potential for a carefully constrained determination of natural mortality.

c. Testing

The Lagrangian and semi-Lagrangian approaches have received much attention during the past three decades (Sawyer, 1963; Robert, 1981; Pellerin *et al.,* 1995). The most attractive feature of the Lagrangian approach is the relatively large time step of the integration of the model in comparison with numerical schemes realized in Eulerian coordinates. However, there are two problems associated with the Lagrangian approach. The first is the redistribution (divergence or convergence) in time of the initially uniformly distributed particles. The second is the difficulty of remapping the Lagrangian solution onto the Eulerian grid. Both of these problems demand complicated numerical algorithms. If these numerical algorithms are not computationally efficient, then there is little to be gained.

The efficiency of the data assimilation algorithm strongly depends on the successful minimization of the cost function, and therefore results in additional requirements for the model formulation. The two key issues are the dimension of the independent variables (control vector) and the level of numerical noise in the model solution. The first follows from the fact that the computational effort required for the minimization of the cost function is proportional to N^2, where N is the dimension of the control vector. The second is explained by the complicated form of the cost function and the difficulty in finding the global minimum in the presence of numerical noise and/or a high frequency component in the model solution. The dimension of the minimization problem becomes especially important if we need to calculate errors in the solution. Determining the errors is accomplished by inverting the N-by-N dimensional Hessian matrix associated with the cost function (Tziperman and Thacker, 1989), and may be extremely complicated for large N.

We illustrate the application of the semi-Lagrangian model (Eqs. 2–4) by comparing its results with those of Eq. (1) solved on an Eulerian grid. For the comparison, we used a stationary nondivergent velocity field similar in spatial scale to that observed for Western Bank (Fig. 1) and defined by the superposition of three harmonics:

$$\psi = \sum_{k=1}^{3} A_k \sin(\varphi_{xk} x + \varphi_{yk} y) \tag{16}$$

$$u = -\partial\psi/\partial y \tag{17}$$

$$v = \partial\psi/\partial x \tag{18}$$

where A_k is the amplitude and the sine functions represent the spatial harmonics on the regular grid with $\delta x = 3$ km and $\delta y = 3.3$ km. This velocity field was also applied to other

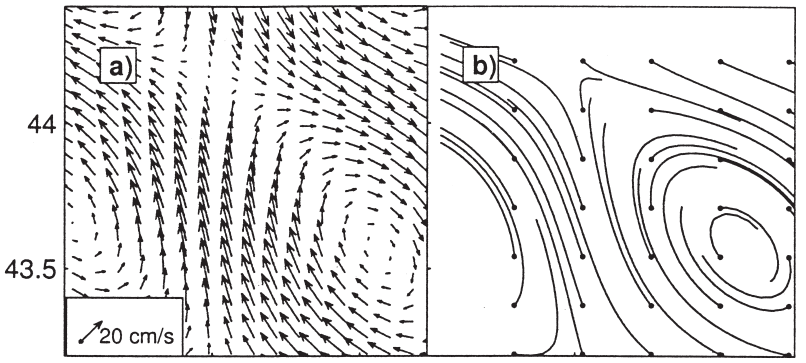

Figure 2. (a) The velocity field defined by Eqs. (16–18); (b) Typical trajectories of the Lagrangian particles in the velocity field shown in (a).

numerical experiments described below. Figures 2a and 2b show the spatial structure and corresponding trajectories of the Lagrangian particles in the velocity field defined by (16–18). Note that the proposed variational technique does not assume stationarity hypothesis and nonstationary velocity fields can be easily applied.

The Eulerian finite-difference solution of Eq. (1) was obtained by using a conservative second-order in space and leap-frog in time ($\delta t = 0.7$ hours) finite-difference scheme (Roache, 1998). The dimension of the Eulerian grid was $32 \times 42 = 1344$. We set the initial concentration of the passive tracer to be zero outside of the study region (Fig. 3a). Therefore, zero boundary conditions were specified along boundary points where particles flow into the domain. Boundary conditions for the outflow portions of the domain were extrapolated from the internal points. Figure 3a illustrates the evolution of the Eulerian solution over seven days.

The corresponding semi-Lagrangian solution was obtained via the following procedure:

a) Each grid point from the initial Eulerian distribution (Fig. 3a, 0 days) was treated as exact 'data.' This gives 1344 'measurements' but a large number of these measurements were equal to zero, and so only 1000 measurements were used.

b) The trajectory of 1000 particles initialized in points corresponding to the data were calculated using a standard Runge-Kutta fourth order accuracy algorithm (Roache, 1998). Some of these trajectories are shown in Figure 2b.

c) The set of optimal initial amplitudes C_k^0 was calculated as a function that minimized the sum of the functions (6), (8) and (9). Note, that Eq. (7) may be ignored because the number of particle-functions is equal to the number of data extracted from the Eulerian solution.

The final semi-Lagrangian solution (Eqs. 2–4) is shown in Figure 3b. We note that the semi-Lagrangian solution does not have the noise observed in the finite-difference solution because of the weak divergence of the velocity field and imperfect reflection from open boundaries.

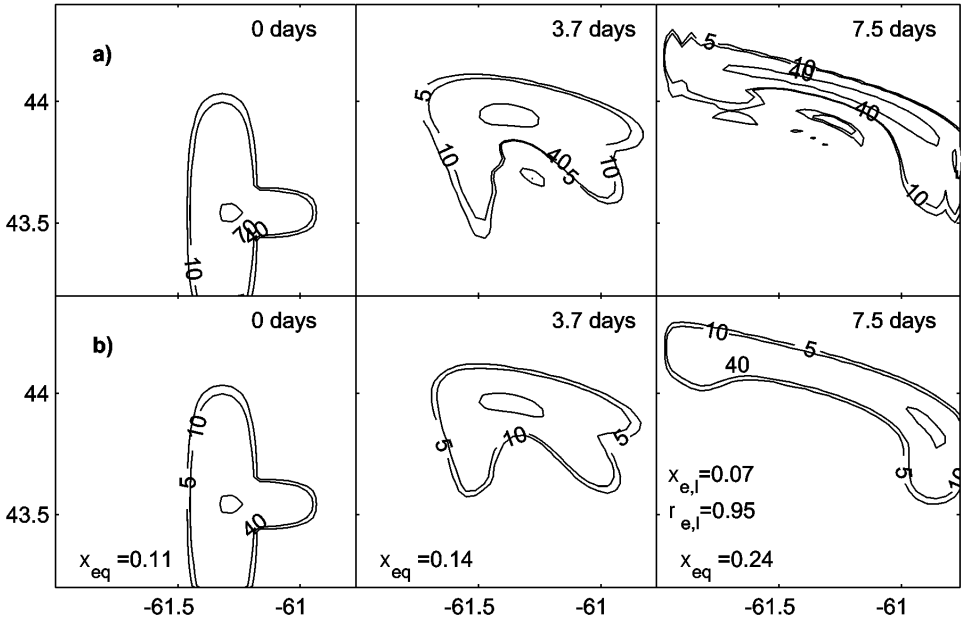

Figure 3. (a) Evolution of the finite-difference Eulerian solution of Eq. (1); (b) Evolution of the semi-Lagrangian solution defined by Eqs. (2–4). Time-dependent error in the dynamical equation ξ_{eq}, averaged correlation $\varrho_{e,l}$ and the error $\xi_{e,l}$ between the Eulerian and Lagrangian solutions.

There are some disadvantages with this approach. The principal drawback is the practical difficulty of specifying an infinite number of particles in order to obtain the desired accuracy for the solution to Eq. (1) over a long time period, when the spatial scale of the function-particle $2 \times (D_C(t + t_0))^{0.5}$ becomes comparable with the typical velocity spatial scale. This can be particularly important if the flow field has strong spatial gradients. Another problem is the error that develops when mapping the analytical solution onto the Eulerian grid. Such errors can lead to violation of the mass conservation property of the analytical solution (Eqs. 2–4). However, with careful calculation of the integral $\int C_k dx dy$ over the Eulerian grid these problems can be minimized. Furthermore, the imprecise specification of the initial and boundary conditions can be problematic, although for the short integration times, when the number of specified functions (Eq. 4) can be limited without much loss of accuracy, this drawback is not critical.

For quantitative analysis of the semi-Lagrangian algorithm we need to define the 'true' solution of Eq. (1). The semi-Lagrangian solution is smoother, but violates mass conservation properties being mapped to the Eulerian grid (see above), while the Eulerian finite-difference conservation scheme is free of this drawback. The limitations of the finite-difference scheme (dispersion, banding with negative values) are well known and will be amplified with an increase in the time integration period. For our example, however, where the spatial extent of the grid is about 200 km and the velocity scale is about

10 cm/sec, the advective lifetime is $T = L/U \approx 20$ days, so for the 8-day modeled period the difference between the finite-difference and true solution is small and despite its imperfections we can consider the Eulerian solution as the 'true' solution. Meanwhile, comparison of the plots (Fig. 3) shows that both the Eulerian and semi-Lagrangian solutions are highly correlated.

We can also consider a measure of similarity between the Eulerian and Lagrangian solutions in terms of their correlation $\varrho_{e,l}$ and error $\xi_{e,l}$ defined respectively as

$$\varrho_{e,l} = \int_{t,x,y} (C_e - \overline{C_e})(C_l - \overline{C_l})dxdydt \Big/ \times \left(\int_{t,x,y} (C_e - \overline{C_e})^2 dxdydt \int_{t,x,y} (C_l - \overline{C_l})^2 dxdydt \right)^{0.5} \tag{19}$$

$$\xi_{e,l} = \left(\int_{t,x,y} (C_e - C_l)^2 dtdxdy \Big/ \int_{t,x,y} C_e^2 dtdxdy \right)^{0.5} \tag{20}$$

where C_e and C_l are the Eulerian and semi-Lagrangian solutions. For the case illustrated in Figure 3, the introduced correlation and error were about 0.95 and 0.07, respectively.

As a quantitative measure of the character of the semi-Lagrangian solution, we introduced an error function into the solution of Eqs. (2–4) as an approximation of the finite-difference solution to Eq. (1):

$$\xi_{eq} = \left(\int_{t,x,y} [\partial C/\partial t + \nabla(\mathbf{u}C) + mC - D_C \Delta C]^2 dtdxdy \Big/ \int_{t,x,y} (\partial C/\partial t)^2 dtdxdy \right)^{0.5}. \tag{21}$$

The error is approximately 0.1–0.15 at the beginning and increases to 0.25 by the seventh day (Fig. 3).

4. Sensitivity studies

a. Antenna sensitivity

We carried out many test studies in order to determine the characteristics of our approach. Our primary goal was to reconstruct the nonstationary passive tracer field with data analogous to observations derived from a shelf-scale oceanographic survey in a region with open boundaries and a known velocity field. To begin, we generated the Eulerian solution to Eq. (1) with a stationary velocity field (Eqs. 16–18) and initial conditions as shown in Figure 3a. Additionally, we simulated the constant inflow of a passive tracer along the eastern boundary. The final result with zero mortality ($m = 0$) is shown in Figure 4a.

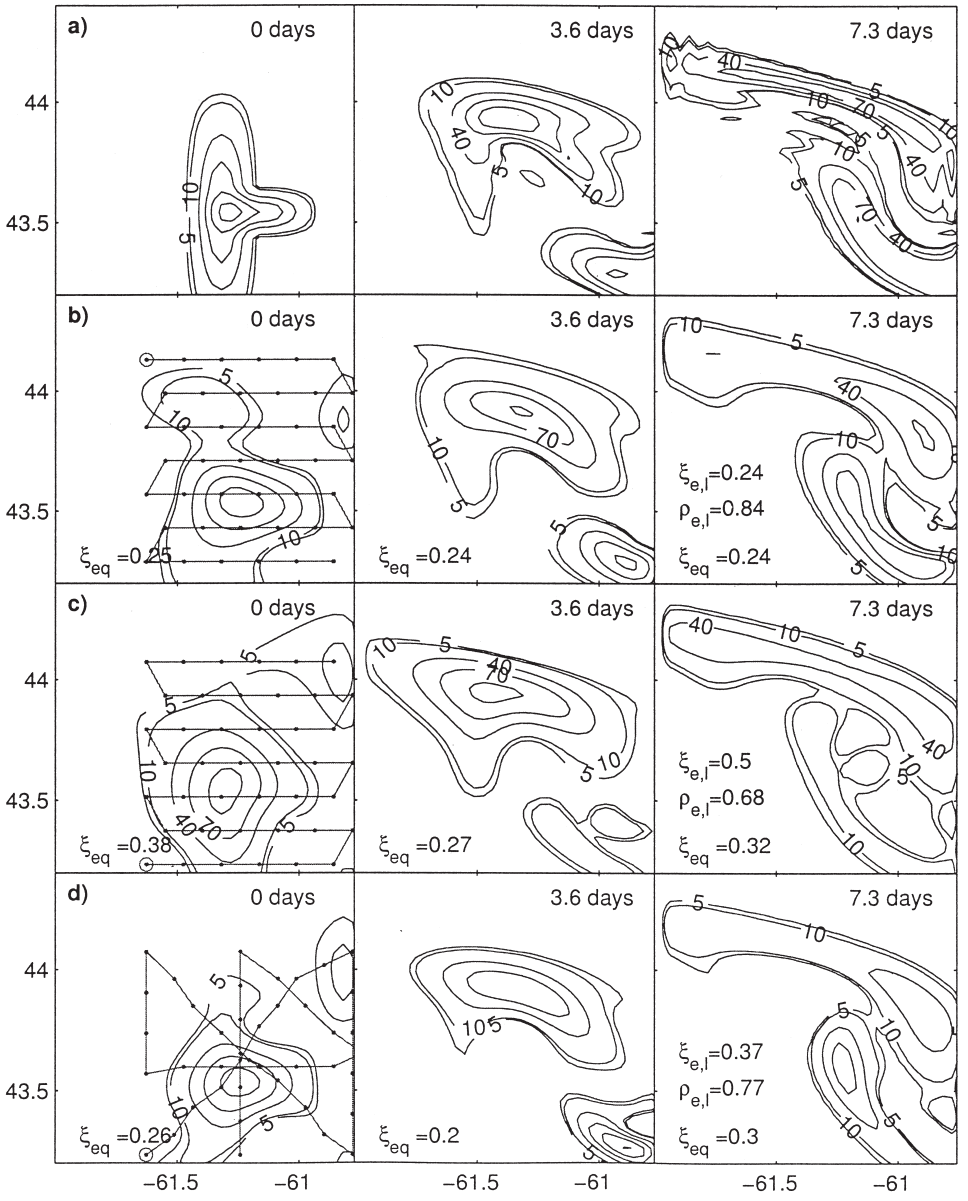

Figure 4. Results showing the sensitivity of the reconstruction to the sampling approach, testing the influence of the survey design. (a) —the original finite-difference solution. (b), (c) and (d)—reconstruction of the passive tracer field based on data extracted according to zigzag line on the left panels. The empty circles denote the starting point of the sampling line. Time-dependent error in the dynamical equation ξ_{eq}, the averaged correlation $\varrho_{e,l}$ and error $\xi_{e,l}$ between the Eulerian and Lagrangian solutions are shown.

Following different zigzag lines, corresponding to cruise tracks, we extracted 42 'observations' from the passive tracer field and imposed the relative measurement error $\xi_{data} = 0.1$. The distance between 'stations' was chosen to enable resolution of the spatial structure of the observed phenomena. Following Eq. (10), each observation has a prescribed error covariance of $(\overline{C^*_{t_n}}\xi_{data})^2$. We applied three different sampling grids to explore the sensitivity to 'cruise' path: (i) with samples taken from north to south (opposed to the mean current), (ii) with samples taken from south to north (following the mean current), and (iii) with samples made following a star pattern.

The sampling lines and the results of the passive tracer reconstruction (Figs. 4b, 4c and 4d) show that there is substantial sensitivity to the survey path and structure. The typical size of the control vector (the number of initialized initial and inflow particles) was about 3000. The best reconstruction was achieved in case (i) where the correlation $\varrho_{e,l}$ and error $\xi_{e,l}$ between the Eulerian and semi-Lagrangian solution were 0.84 and 0.24, respectively. The worst reconstruction ($\varrho_{e,l} = 0.68$, $\xi_{e,l} = 0.5$) was obtained in case (iii), perhaps because of difficulty in obtaining reliable information about the inflow passive tracer with measurements made in the direction of the mean current (case iii). The reconstruction of the inflow tracer in the southeastern part of our region (Fig. 3c) supports this interpretation. Similar results were obtained by Panteleev and Semenov (1988) who found that sampling along the Gulf Stream was much improved if the survey moved in a direction opposed to the current direction.

The efficiency of the star pattern scheme (correlation 0.77, Fig. 4d) is also reduced when sampling took place in a direction opposed to the prevailing current. Its advantage, in comparison with the current-following scheme, is a consequence of the observations taken in the southeastern corner halfway through the survey. In general, these results confirm other work that suggests that sampling should be conducted in a direction opposed to the prevailing circulation. Taking into account these results, we suggest that prior analysis of the mean circulation should be used to construct more efficient sampling grids for biological and hydro-physical studies. Such an analysis provides an opportunity to optimize the sampling strategy when there is information on the circulation.

An interesting result of this preliminary analysis is that errors in the transport equation, represented by ξ_{eq}, do not depend upon the sampling scheme and lie within the range 0.25–0.30. Similar values were obtained from all of our numerical experiments, and so we will not discuss this parameter further.

b. Errors in the passive tracer data

The data simulated in the previous subsection were extracted from the finite-difference solution and can be treated as perfect, or exact, data. Of course, all observations have some errors which can strongly influence the output. Therefore, it is important to investigate the sensitivity of the proposed algorithm to errors in the observations.

For this analysis we contaminate the known, 'perfect,' observations with 10, 20, 30 and 40% noise and apply our algorithm with an appropriately modified weight function g_1. The

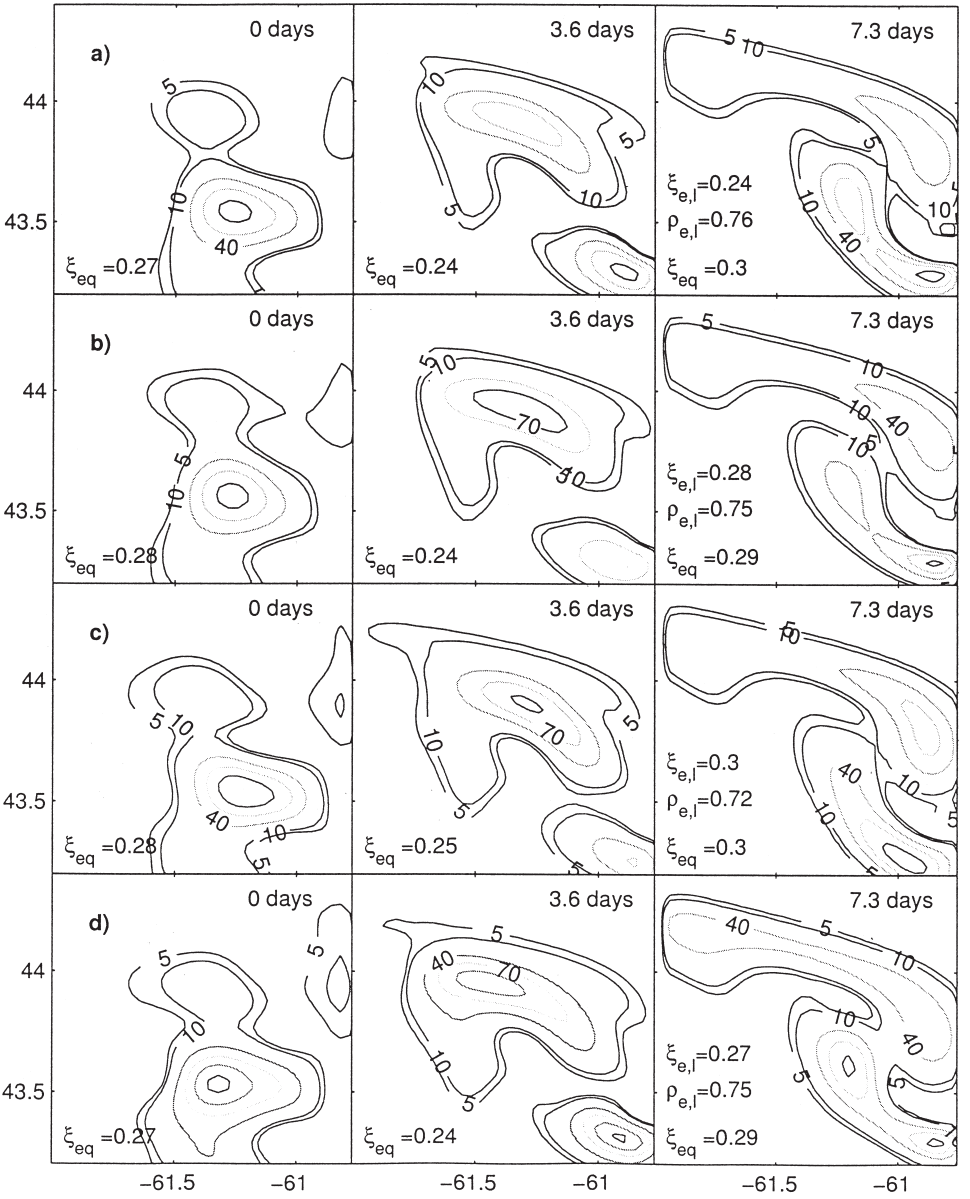

Figure 5. Results showing the sensitivity to errors in the observations of the passive tracer data: reconstruction of the passive tracer field based on data contaminated by error at (a) 10%, (b) 20%, (c) 30% and (d) 40% amplitudes. Time-dependent error in the dynamical equation ξ_{eq}, the averaged correlation $\varrho_{e,l}$ and error $\xi_{e,l}$ between the Eulerian and Lagrangian solutions are shown.

results of this experiment (Fig. 5) show that the proposed algorithm has a filtering property such that the output has only a weak dependence upon the level of the observational error. Figure 5 shows that the $\xi_{e,l}$ and the correlation $\varrho_{e,l}$ between the reconstructed and the

Eulerian solutions remain nearly constant and equal 0.24–0.3 and 0.72–0.76, correspondingly.

These experiments show that the algorithm is quite robust in the face of errors in the observations. This property of the algorithm allows us to use it even in cases where the errors in the data are not well defined or the errors are suspected to be large. Errors in oceanographic data are often large and meaningful; estimation of the true errors of the observations is difficult at best (Taggart and Frank, 1991).

c. Errors in the velocity field

For the application developed here, we assume that the velocity field is known, so error analysis of the velocity field is not crucial for our discussion. Nonetheless, it is instructive to consider the potential influence of errors in the velocity field on the final distribution of the passive tracers.

The velocity field is a superposition of three harmonics (Eqs. 16–18). Therefore, the sources of possible errors in the flow field can be split into errors of the phase and amplitude of these harmonics. Figures 6a, 6b, 6c and 6d represent some examples of the reconstructed passive tracer field based on imperfect velocity fields. The modifications include changes in phase (Figs. 6a and 6b) and amplitude (Figs. 6c and 6d) of the harmonics which make up the velocity field, with deviations of about 5% (Figs. 6a and 6c) and 20% (Figs. 6b and 6d), respectively. As expected, the algorithm is rather sensitive to errors in the flow field (Fig. 6). This result is a logical consequence of the primary role of the advective terms in Eq. (1). At the same time, even if the velocity field applied differs from the true field by up to 20%, many of the features are conserved in the reconstructed passive tracer field.

d. Mortality estimates

The mortality rate coefficient m can be included as part of the control vector. Therefore, mortality can be defined in an optimal way using the available passive tracer, current data and the actual estimated decrease in the passive tracer as described by the third term of Eq. (1).

Errors in the observations will influence the final mortality estimates, so we investigated the relationship between a pre-defined mortality rate and errors in the passive tracer observations. As before, we began with the finite-difference solution to Eq. (1) with $m = 0.1$ (Fig. 7a). Then we extracted 42 pseudo-measurements and contaminated them with noise at levels corresponding to 10, 20, and 30% of their amplitude. No information about the size of the mortality coefficient was included, but we did require that the mortality rate should be positive. We neglected the function $\mathbf{f_5}$ ($g_5 = 0$) but used the $\mathbf{f_6}$ function to avoid any negative values of mortality. The reconstructed passive tracer fields and the optimized mortality rates (Figs. 7b, 7c and 7d) $m = 0.083$ were very close to the original

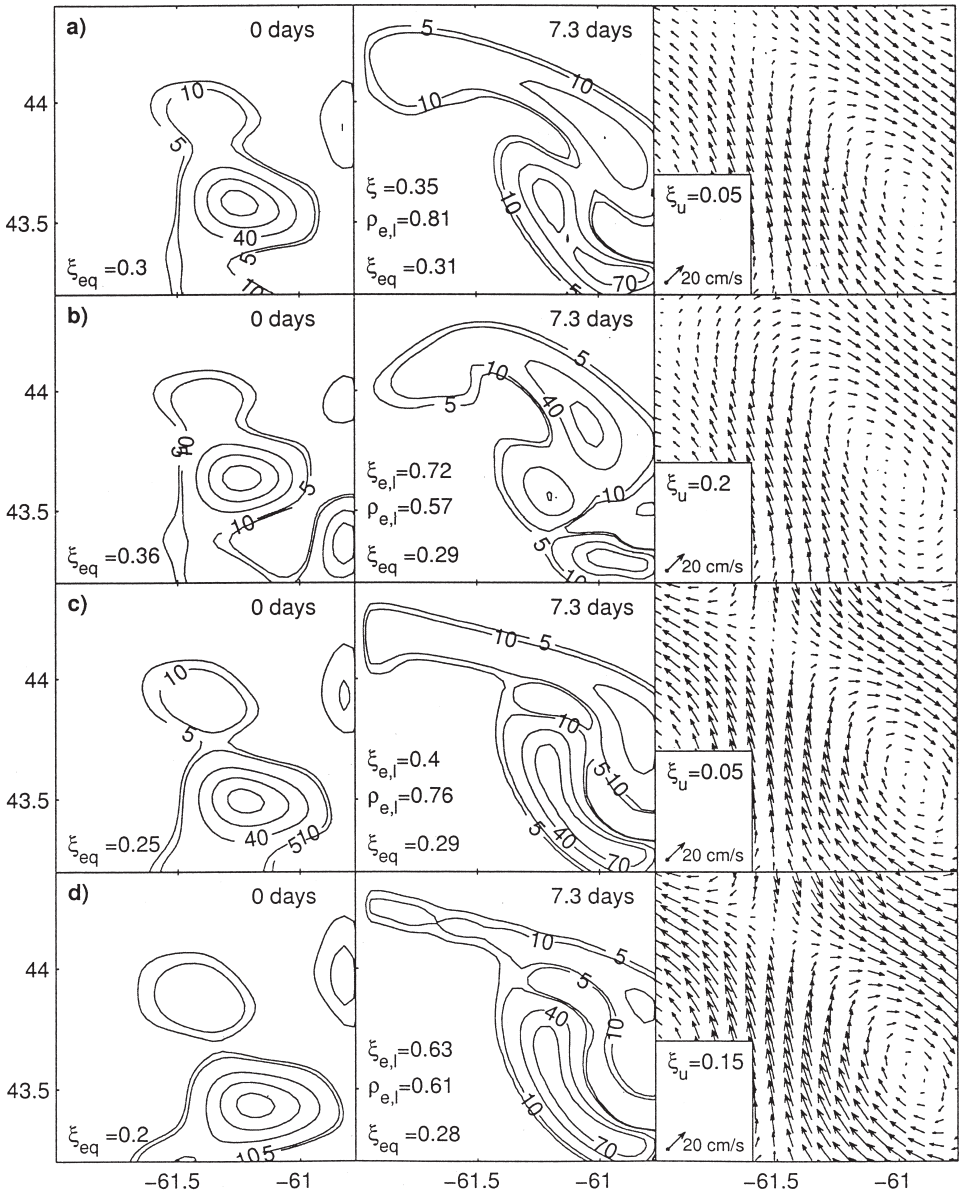

Figure 6. Results showing sensitivity to errors in the velocity field: (a), (b), (c) and (d)—reconstruction of the passive tracer field with exact data and incorrect velocity field. The error in the velocity field is shown on the right side. Time-dependent error in the dynamical equation ξ_{eq}, the average correlation $\varrho_{e,l}$ and error $\xi_{e,l}$ between the Eulerian and Lagrangian solutions are shown on the left and central panels.

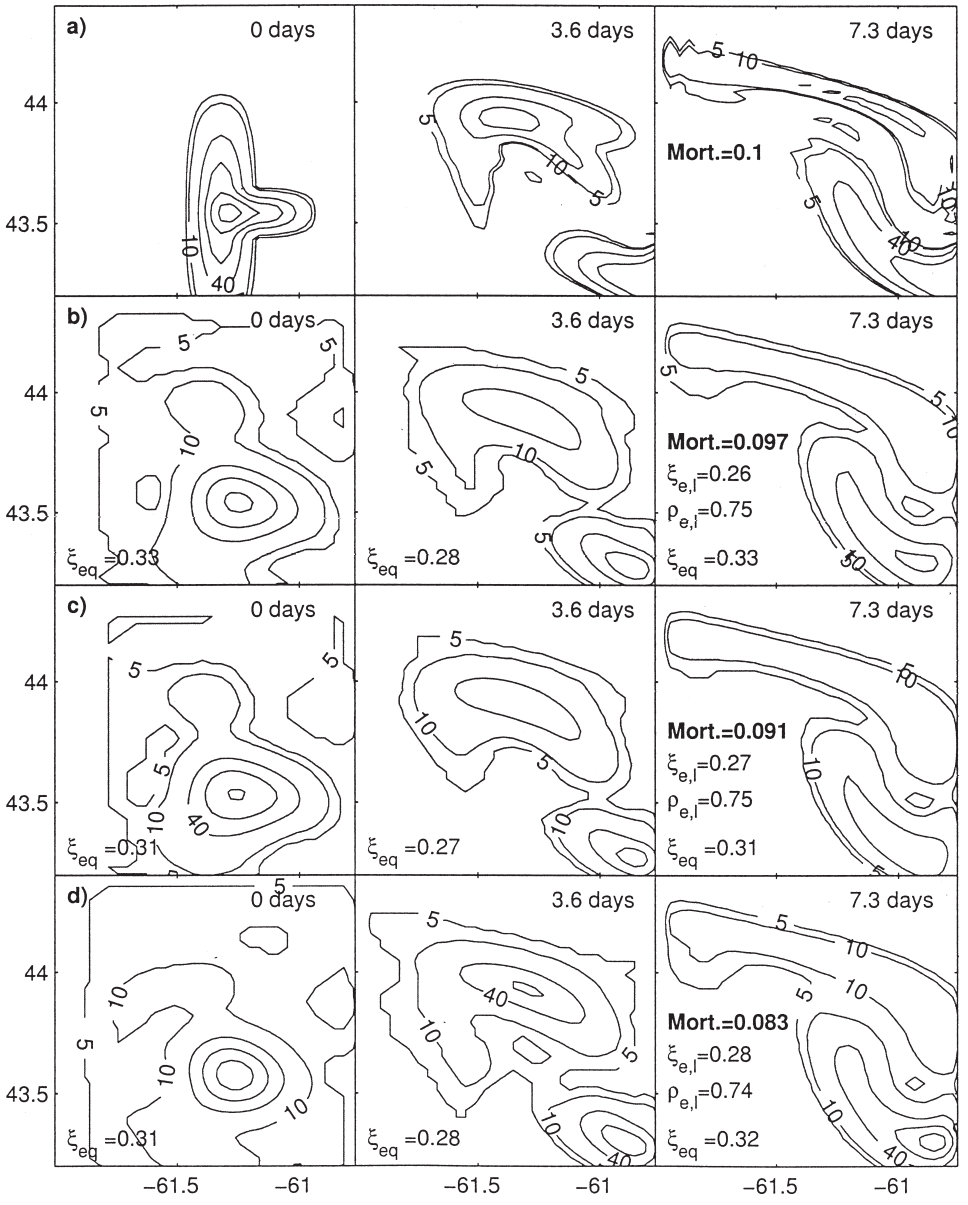

Figure 7. Determination of mortality: (a) finite-difference solution to Eq. (1) with mortality $m = 0.1$. Reconstruction of the passive tracer field and mortality coefficient by using data contaminated with errors with amplitudes of (b) 10%, (c) 20% and (d) 40%. Time-dependent error in the dynamical equation ξ_{eq}, the averaged correlation $\varrho_{e,l}$ and error $\xi_{e,l}$ between the Eulerian and Lagrangian solutions are shown.

observations and the prescribed value of the mortality $m = 0.1$ even with 30% noise in the data (Fig. 7d).

As in the previous simulations, the influence of errors in the observations on the resulting mortality rate is rather weak. We speculate that the filtering properties of the algorithm result from the incorporation of a semi-Lagrangian approach that strictly (analytically) prescribes the movement, mortality and diffusion of each particle. Such prescriptions guarantee the smoothness of the solution, thus helping to avoid numerical noise that can result from any finite-difference scheme.

We note that the resulting mortality coefficients ($m = 0.097, 0.091, 0.083$) are smaller than the prescribed value $m = 0.1$. This results from the inclusion of the smoothing function $\mathbf{f_2}$ into $\mathbf{F}$. This function, through the minimization of the cost function, causes the solution to underfit somewhat the initially larger amplitude, which therefore leads to a smaller value of the natural mortality. It may be that a different choice of the smoothing function $\mathbf{f_2}$ would reduce this effect. For real applications we should take this bias into account. Nonetheless, the bias appears to be quite small; it is only 17% when the error in the observations is as great as 30%.

The strategy of data sampling may be different from the utilized upstream sampling if the principal goal of the survey is to define the larval mortality. Therefore we did two additional experiments with downstream sampling and two 'simultaneous' survey in the beginning (0 day) and at the end (7.3 days) of the integration period.

The optimized solution and mortality coefficient obtained with the downstream sampling data, contaminated by 10% noise, are shown in Figure 8a. The downstream strategy gives worse results both in the reconstruction of the passive tracer and in mortality coefficient (Figs. 7 and 8). The estimated mortality is approximately 0.07 day^{-1} e.g. much worse than estimated with 30% noise incorporated into the data (Fig. 7d). This makes sense since the variational algorithm tries to reconstruct the 'true' solution over the whole domain while the 'true' solution can be achieved only with 'true' mortality. Therefore, the downstream strategy, at least as defined in Figure 8a, has no advantages. Obviously, the greatest accuracy in the definition of the mortality will be achieved if the ship is drifting and periodically sampling the larvae in the ocean. However, in that case the total number of larvae and their spatial distribution cannot be estimated.

The results from two simulations with no additional noise added (Fig. 8b) to the data show remarkable accuracy (0.103 day^{-1}), even though the spatial distribution of the larvae still contains moderate errors. We believe the small error in the mortality estimate is because of the good estimate of total larvae at the beginning and end of the survey. The example presented in Figure 8b is obviously unrealistic but it gives guidance as to how to plan the 'best' experiment. It is interesting to note that the definition of the mortality through the fitting of the initial C_0 and final C_0 average concentration to the function $C_0 \exp(-mt)$ gives $m = 0.015$ day^{-1}, a very poor result, a consequence of neglecting inflow and outflow of larvae through the open boundary.

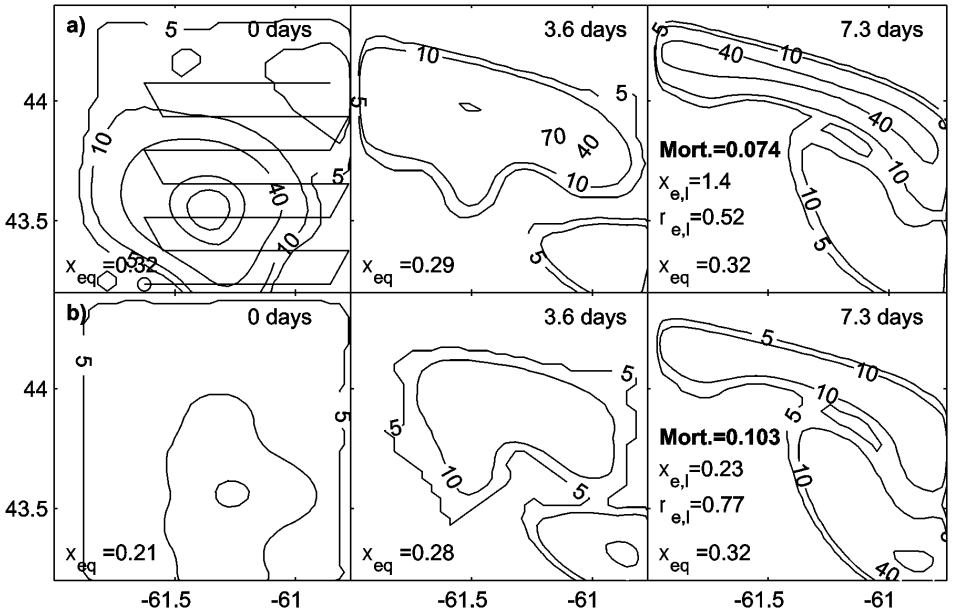

Figure 8. Determination of mortality with (a)—down-stream sampling strategy with 20% noise added to the data, and (b)—two simultaneous surveys conducted at the beginning and at the end of the integrated period.

5. Application to silver hake eggs and larvae on Western Bank

a. Eggs and larvae distributions on the Western Bank

The sensitivity studies presented above indicate that the proposed algorithm can be applied to the reconstruction of larval fish distributions and to the estimation of the mortality rate of those fish eggs and larvae over short periods of time (days to weeks). In this section, we apply the algorithm to observations of fish eggs and larvae collected on and around Western Bank, on the Scotian Shelf, during two cruises in early August 1998. During the first cruise (1–3 August) fish eggs and larvae were collected at 17 stations, in an approximate grid (Fig. 1a, closed squares). On the second cruise (6–8 August), larval fish were collected along four fine-scale transects, designed to intersect fronts or other hydrographic features around the bank (Fig. 1c). Plankton tows were made with a 63 cm bongo net that provided tow depth and temperature monitored in real time (Reiss *et al.,* 2000). Tows were made obliquely to 50 m, or to 5 m from the bottom where the water depth was less than 55 m (Hazen and Reiss, 1999). Aboard ship, samples from both sides of the bongo nets were rinsed through a 333 μm mesh screen and the port side was stored in 5% buffered formalin seawater. In the laboratory, eggs and larvae of silver hake (*Merluccius bilinearis*) were removed, enumerated and measured to the nearest 0.1 mm total length. The typical concentration of eggs and young larvae (smaller than 3 mm) for

the first cruise and 3–4 mm larvae for the second cruise are shown in Figures 1b and 1c. We note that the larval concentration is very patchy, one of the reasons why direct calculations of mortality, or calculations using strong constraints, are so difficult.

b. Features of Western Bank

A mooring array consisting of 11 moorings with a total of 25 current meters was deployed on Western Bank from 1 August to 15–16 October 1998 (Fig. 1a). Distances between the moorings were below the observed decorrelation length scale for this Bank (Thompson and Griffin, 1998), 25 km. The data from the mooring array allow us to use the time-dependent velocity measurements to define the subtidal flow fields (cf. Panteleev *et al.,* 2001) for the period from 1–8 August 1998.

Analysis of the current data showed that both tidal and inertial residuals for periods longer than two days were relatively unimportant, so we included only the subinertial currents (Panteleev *et al.,* 2001), while taking into account the tidal and inertial variability in our estimates of the horizontal diffusivity. The current observations were interpolated onto a regular grid to generate a nonstationary flow field that was in general rather stable in time (the average variance over the period was less than 8×10^{-4} m^2 s^{-1} over the seven day period). The seven-day mean circulation (Fig. 1a) revealed a pattern that was typical of the circulation over Western Bank (Sanderson, 1995; Griffin and Thompson, 1996; Reiss *et al.,* 2000; Panteleev *et al.,* 2001) dominated by an anticyclonic gyre with a maximum velocity U of approximately 10 cm s^{-1}.

The stability of the flow field is primarily the result of rather calm summer wind conditions. The mean wind speed was about 3 m/sec. The stability of the circulation field, and the agreement with earlier observations, supports the reliability of these observations. As noted above, errors in the velocity field can strongly influence the results. These current observations, made at distances shorter than the decorrelation scale of the subtidal currents, provide a very robust circulation field for both the reconstruction of the passive tracer (eggs and larvae) field and for the determination of the associated mortality rate that we seek to recover.

At summer water temperatures on the Scotian Shelf, silver hake eggs hatch in 24 to 36 hrs. The growth rate of silver hake larvae is based on age data obtained from the otoliths (ear stones) and is best described by a linear function for larvae less than 8 mm total length (Jeffrey, 2000). We applied the simple growth rate formula and assumed that eggs hatched 24 hrs after release. The size at hatch is estimated at 1.7 mm total length, and the growth rate is fixed at 0.18 mm day^{-1} (Jeffrey, 2000). For this analysis, we used eggs and larvae smaller than 3 mm observed during the first cruise and compared the distributions to larvae sized between 3 and 4 mm collected during the second cruise.

Some of the egg and larval collections were made before the deployment of the entire current mooring array. Therefore, at those locations eggs and larvae must have experienced some mortality. So, at those stations we recalculated the egg and larval concentrations by applying a mortality rate of 0.2 day^{-1} with an exponential decay. This mortality is a

commonly accepted value for natural mortality (Heath, 1992; Houde, 1996). Unfortunately, these adjustments do, of course, lead to some error in the data used by the algorithm. So, we specified a somewhat high error rate ($\xi_{data} = 0.4$) for these data. We set $\xi_{data} = 0.2$ for the data sampled within the investigated period and region. The data were sampled at the beginning and at the end of the study period, so it is sufficient to define just the initial and final mean concentrations $\overline{C^*_{t_{in}}} = 1000$ and $\overline{C^*_{t_{in}}} = 80$ per 100 m^3 and to define inverse error covariance of the measurements according to Eq. (10). Because our knowledge of mortality is rather limited, we begin with a mortality estimate of 0.2 day^{-1} with a standard deviation of 0.15 day^{-1}. Taking into account the exponential influence of the mortality on the concentration, this standard deviation is rather large and should overlap all possible realistic values of the mortality coefficient.

The station distributions for the two surveys were rather irregular and sparse relative to a truly desirable sampling scheme. Obviously available ship time, and other practical constraints, limited and determined the survey design. These limitations make the data analysis and definition of physical scales quite difficult. As mentioned above, we specified a typical spatial scale of $L_C = 20$ km as a minimum scale resolved by the measurements during the first survey. The typical initial and final scales for the egg and larval concentrations were defined as the maximum concentration measured during the first and second survey, respectively. The time scale was chosen to be 2.5 days based upon a scaling of $T_C = L_C/U$.

c. Egg and larval distributions

Following the proposed algorithm (Eqs. 2–15), we defined the corresponding weight functions g_i and reconstructed the evolution of eggs and larvae in space and time. The small velocity amplitude observed during the study period (see Fig. 1a) allowed us to substantially decrease the number of particle-functions needed for the approximation (Eqs. 2–4). Combined analysis of data and particle trajectories from several numerical experiments showed that just 898 particle-functions would be required. So, the dimension of the control vector q (initial concentration C^o_k and mortality m) was 899.

The initial (1 August) and final (8 August) distributions are shown in Figures 9a and 9b. Comparison with the optimally interpolated (Gandin, 1964) larval concentration presented in Figures 9c and 9d reveals that the variational algorithm preserves the sharp gradients which disappear with optimal interpolation. This sharpness is a result of the nonstationarity and the dynamic constraints utilized here. We note the difference in the locations of the concentration maxima are in better agreement with the flow field (Fig. 1a) in Figures 9a and 9b (maxima A1, B1 and A2, B2) than in Figures 9c and 9d. (maxima C1, C2).

The determination of the mortality rate between the two sampling periods is another result of the proposed algorithm. Our recovered mortality rate was 0.27 day^{-1}, somewhat greater than the preliminary estimated value of 0.2 day^{-1}. Note that the first-guess mortality of 0.2 day^{-1} was made from data observed for a 25-day period (Taggart *et al.,*

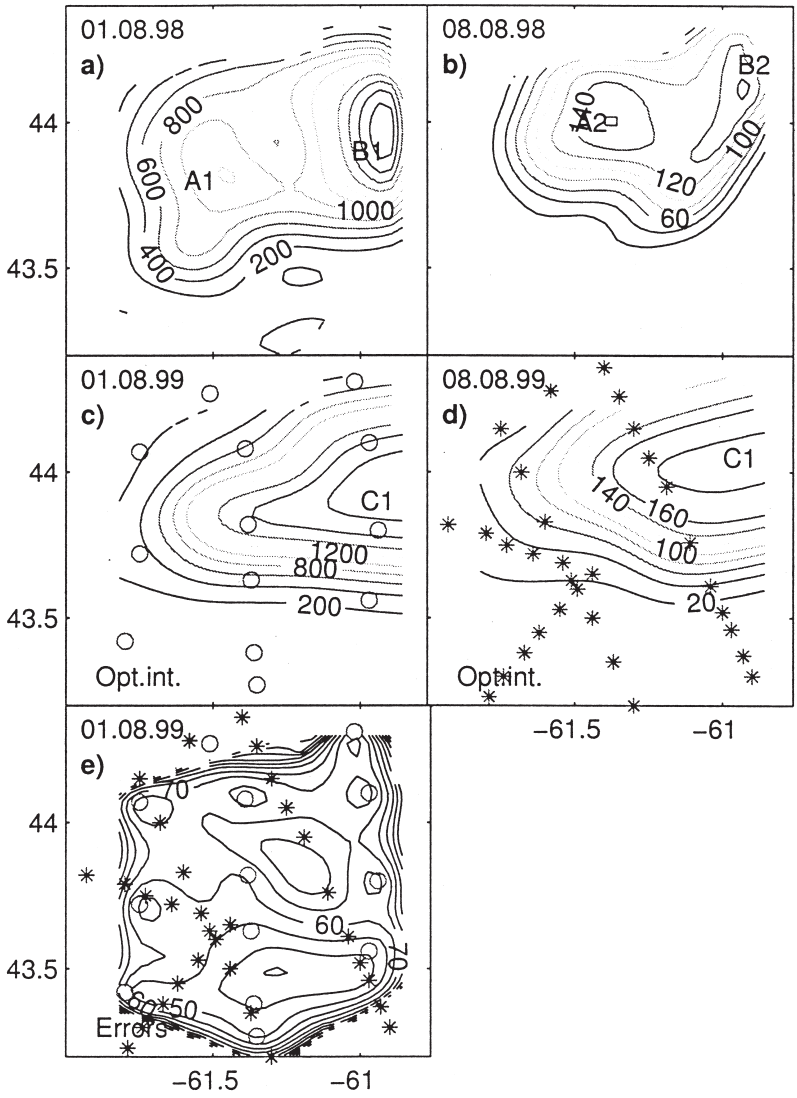

Figure 9. Reconstruction of the evolution of the concentration of (a) eggs and (b) young larvae of silver hake on Western Bank. Optimal interpolation method was used to generate the contours of the concentration of eggs and young larvae for the (c) first and (d) second surveys. (e) Standard deviation of the initial larvae concentration. The station locations are shown with asterisks and circles. The letters A1, A2, B1, B2, C1, C2 denote the initial and final position of local maximums discussed in the text.

1996) and could take into account mortality of both young and relatively old larvae. Our results are based on a shorter period and reflect the mortality of early stages which should be larger.

d. Errors

The treatment of the elements of the control vector as a stochastic δ-correlated function creates a floor for the definition of the errors of the estimated physical values. According to Wunsch (1996), the error covariance of the corresponding elements of the control vector can be calculated as diagonal elements of the inverse Hessian matrix $\mathbf{H}^{-1} = (\partial^2\mathbf{F}/\partial q_i \partial q_j)^{-1}$. Several approaches have been suggested for the computation of $\mathbf{H}$. The most straightforward and simple method of calculation is to obtain its finite-different approximation (Tziperman and Thacker, 1989). Following this approach, we individually perturb the components of the control vector $q = (C_k^o, m)^T$ in the vicinity of their optimal values. After that the finite-different approximation of the Hessian can be defined as:

$$H_{i,j} = (\nabla_{q+\delta q_i}\mathbf{F} - \nabla_{q-\delta q_i}\mathbf{F})_j/2\delta q_i. \tag{22}$$

The inverse Hessian matrix was calculated through the sequential solution of N matrix equations

$$\mathbf{H}X = I_i,\ i = 1 \cdots N, \tag{23}$$

where $I_i = (0, 0, 0 \ldots 1, \ldots 0)^T$ is an n-dimensional zero-vector whose i-th element is equal to one. The solutions of these equations were obtained with the generalized minimum residuals algorithm proposed by Saad (1992). Each solution gives us one column of the inverse Hessian matrix. The square root of the first $N - 1$ diagonal elements of the inverse Hessian matrix are mapped in Figure 9e. Their physical meaning is a standard deviation or standard error δC_k^o of the corresponding physical values e.g. estimated initial larvae and eggs concentration C_k^o. As can be seen from Figures 1 and 9e, the minimal errors in general correspond to data points in the southern part of the study area where there are the most measurements and the strongest currents. This current effectively spreads information from the point of measurement to the neighboring points according to the velocities (Fig. 1). The error peaks are concentrated in the central part of the region where there are the fewest measurements and where currents are weak. The weakness of the circulation creates a situation in which information from the nearest measurements cannot 'spread' into other parts of the grid during the seven-day observation period. Far from the data, the errors naturally increase but still do not exceed the prior defined expected values discussed above (Wunsch, 1996).

The standard deviation of the mortality estimate is 0.02 day^{-1}. Obviously, with these two measurements of the concentration made at the start and the end of the survey, only one estimate can be made of the mortality, so in our case 60 measurements of the concentration should give an estimate for the standard deviation of $0.2/\sqrt{30} \approx 0.036$ day^{-1}, which is in rough agreement with the obtained error estimate.

We also explored the influence of possible errors in the weight function specification upon our estimates of mortality by carrying out numerical experiments with other possible, realistic time and space scales. For all of the experiments that we conducted, we found that the mortality ranged from 0.25 to 0.3 day^{-1}. Even though this algorithm exhibited a small

bias (a persistent tendency to underestimate the natural mortality) due to smoothing terms incorporated into the cost function, the best estimate for the natural mortality of larval fish during this experiment was 0.28 day^{-1} with an error of just 0.03 day^{-1}.

Our error estimates do not take into account possible errors in the velocity field. Theoretically the influence of these errors may be estimated the same way, but the time-dependent velocity field could be treated as part of the control vector. Even with a stationary velocity field, the dimension of the control vector will be roughly three times greater and the procedure for inverting the Hessian will demand 27 times more computational effort.

As we show above, the influence of the diffusivity coefficient is not crucial relative to the importance of advection. In order to explore the influence of velocity errors on the mortality coefficient, we carried out several experiments with the velocity field. These experiments revealed that even a 20% increase or decrease of the velocity amplitude causes a difference of only 0.01 day^{-1} in the optimized mortality coefficient. This relatively small error results from the sampling strategy: two quasi-simultaneous surveys were conducted at the beginning and at the end of the study period. These two fields play the most important role with relatively little influence from errors in the velocity field and diffusion.

6. Discussion and conclusions

We have presented a variational algorithm for the reconstruction of a nonstationary two-dimensional passive tracer field. This algorithm should provide the best linear estimator for the passive tracer problem (Wunsch, 1996). Our algorithm was formulated as a semi-Lagrangian, weakly constrained problem. The semi-Lagrangian formulation allowed us to decrease the number of control variables (899 in the case of the Western Bank study) and generated useful filtering of the tracer. The latter property allows the use of rather noisy data, a common situation in oceanography in general, and biological oceanography in particular. Although this algorithm has only been applied to the problem of a single passive tracer, additional tracers could be added. The definition of the weight functions g_i, although subjective, is straightforward. For the Western Bank example, we applied a rather simple approach to estimate these functions from the analysis of the spatial scale and amplitude of the larval distribution. Other approaches (for example, Jameson and Waseda (2000)) could be applied for different physical situations. At the same time, additional information (for example about spatial anisotropy) can be taken into account and in that case modification of the weight function g_i will be necessary. Such adjustments are likely to be particularly useful for larger spatial domains.

From the mathematical and practical point of view, the proposed method is rather simple but requires substantial computer memory. For example, when started with 3000 particles, our algorithm required about 200 megabytes (MB) of memory. This requirement could be substantially decreased (by about 20–30 MB) by excluding the constraint function (8) from (5). Our numerical experiment revealed that solutions with $g_3 \neq 0$ and with $g_3 = 0$ are similar, but minimization of (5) without (8) was less computationally demanding (by

about 5 times). At the same time, application of the additional dynamical constraints defined by (8) produced smaller errors ξ_{eq} and better agreement between the reconstructions and the observations.

Our proposed semi-Lagrangian approach revealed several attractive features in comparison with finite-different Eulerian techniques. Obviously, the comparison with the finite-different Eulerian approximation that we selected is incomplete and other more sophisticated numerical schemes (Boris and Book, 1973; Smolarkiewicz, 1983) might be utilized in order to decrease dispersion and other negative features seen in Figure 2c. For our problem, the development of a passive tracer data assimilation algorithm that could recover the mortality estimates from the concentration data measured with substantial errors, the comparisons are legitimate. A more thorough comparison between our approach and data assimilation algorithms built on sophisticated Eulerian schemes, while outside the scope of this paper, would clearly be interesting.

We designed our algorithm for the rather simple problem of one passive tracer with natural mortality based upon general biological principles. It should be possible, however, to add interactions between two or more tracers, or to add more sophisticated biological parameterization. Of course, for those more demanding problems, we might require more data and the application of the algorithm would be more challenging. One could also apply this algorithm in the backward sense; for example, use passive tracer information, something like temperature and chlorophyll patterns, to estimate the surface circulation field.

The ability to estimate the error is an attractive feature of our algorithm. Although the velocity errors were not taken into account, our error estimates are physically reasonable and so may be used in practice. For silver hake eggs and larvae on the Scotian Shelf in August 1998, we also performed sensitivity tests that revealed that natural mortality lies between 0.25 to 0.3 day^{-1}. Given the persistent tendency of our algorithm to underestimate this function, we suggest that the best estimate for the natural mortality for these data is 0.28 ± 0.03 day^{-1}. Houde (1996) showed that changes in larval mortality rates in the second decimal place could have large impacts on the numbers of surviving larvae in as short a time as 45 days. The generation of mortality rate estimates with an accuracy greater than 0.05 day^{-1} would be a significant advance for biological oceanographic applications.

Acknowledgments. This study was part of the GLOBEC Canada program supported by the Natural Sciences and Engineering Research Council (NSERC) of Canada. We thank the crew of the CCGS *Parizeau* for their assistance in the collection of the data and our colleagues in the Western Bank program. We are grateful to Dmitri Nechaev for interesting discussions and useful advice. This manuscript was prepared with the AGU LATEX macros v3.0.

REFERENCES

Bauer, S., M. Swenson, A. Griffa, A. Mariano and K. Owens. 1998. Eddy mean flow decomposition and eddy-diffusivity estimates in the tropical Pacific Ocean 1: Methodology. J. Geophys Res., *103,* 30855–30871.

Boris, J. and D. L. Book. 1973. Flux-corrected transport. I:SHASTA, a fluid transport algorithm that works. J. Comput. Phys., *18,* 248–283.

Daley, R. 1991. Atmospheric Data Analysis, Cambridge University Press, 457 pp.

Fortier, L. and A. Villeneuve. 1996. Cannibalism and predation on fish larvae by larvae of Atlantic mackerel, *Scomber scombrus:* trophodynamics and potential impact on recruitment. U.S. Fish. Bull., *94,* 268–281.

Gandin, L. S. 1964. On the optimal interpolation of vector fields. Proc. GGP, *165,* 47–59.

Griffin, D. A. and K. R. Thompson. 1996. The adjoint method of data assimilation used operationally for shelf circulation. J. Geophys. Res., *101,* 3457–3478.

Hazen, D. G. and C. S. Reiss. 1999. Cruise Summary: CCGS-PARIZEAU Cruise (98-058) to Western Bank, September 28 October 22, 1998. GLOBEC-CANADA Western Bank Report 1997/7. Dalhousie University, Department of Oceanography.

Heath, M. R. 1992. Field investigations of the early life stages of marine fish. Adv. Mar. Biol., *28,* 1–173.

Helbig, J. and P. Pepin. 1998. Partitioning the influence of physical properties on the estimation of ichthyoplankton mortality rates. I. Theory. Can. J. Fish. Aquat. Sci., *55,* 2189–2205.

Hjort, J. 1926. Fluctuations in the year classes of important food fishes. J. Cons. Int. Explor. Mer, *1,* 5–38.

Houde, E. 1996. Evaluating stage-specific survival during the early life of fish, *in* Survival Strategies in Early Life Stages of Marine Resources, Y. Watanabe, Y. Yamashita and Y. Oozeki, eds., Proc. Inter. Workshop, Yokohama. Balkema Publishers, Rotterdam, 51–66.

Jameson, L. and T. Waseda. 2000. Error estimation using the wavelet analysis for data assimilation: EEWADAi. J. Atmos. Oceanic. Technol., *17,* 1235–1246.

Jeffrey, J. 2000. Growth variation of silver hake (*Merluccius bilinearis*) larvae in relation to oceanographic conditions on Western Bank, Scotian Shelf. MS Thesis, Oceanography Department, Dalhousie University, Halifax, NS, Canada.

Lochmann, S. E., C. T. Taggart, D. A. Griffin, K. R. Thompson and G. A. Maillet. 1997. Abundance and condition of larval cod (*Gadus morhua*) at a convergent front on Western Bank, Scotian Shelf. Can. J. Fish. Aquat. Sci., *54,* 1461–1479.

McIntosh, P. C. 1990. Oceanographic data interpolation: objective analysis and splines. J. Geophys. Res., *35,* 13529–13541.

Marchuk, G. I. 1975. Methods of Numerical Mathematics, Springer-Verlag, New York, Heidelberg, Berlin. 316 pp.

Panteleev, G., B. deYoung and M. Yaremchuk. 2001. The mean summer and fall circulation of Western Bank on the Scotian Shelf. Atmos. Oceans, *39,* 127–144.

Panteleev, G. G., N. A. Maximenko, B. deYoung, C. Reiss and T. Yamagato. 2002. Variational interpolation of circulation with nonlinear, advective smoothing. J. Ocean Atmos. Tech., *19,* 1442–1450.

Panteleev, G. G. and E. V. Semenov. 1988. On the strategy of hydrological array measurements. Oceanologia *6,* 1032–1034.

Pellerin, P., R. Laprise and I. Zawadski. 1995. The performance of a semi-Lagrangian transport scheme for the advection-condensation problem. Mon. Weath. Rev., *123,* 3318–3330.

Press, W. H., S. A. Teulkolsky, W. T. Vetterling and B. P. Flannery. 1992. Numerical Recipes in Fortran, 2nd ed., Cambridge University Press, 963 pp.

Reiss, C. S., F. Panteleev, C. T. Taggart, J. Sheng and B. deYoung. 2000. Observations on larval fish transport and retention on the Scotian Shelf in relation to geostrophic circulation. Fish. Oceanogr., *9,* 195–213.

Roache, P. J. 1998. Fundamentals of Computational Fluid Dynamics, Hermosa Publishers, Albuquerque, New Mexico, 648 pp.

Robert, A. 1981. A stable numerical integration scheme for the primitive meteorological equations. Atmos. Oceans, *19,* 35–46.

Saad, Y. 1992. Numerical Methods for Large Eigenvalue Problems, Manchester University Press, Manchester, UK, 346 pp.

Sanderson, B. 1995. Structure of an eddy measured with drifters. J. Geophys. Res., *100,* 6761–6776.

Sanderson, B. and D. A. Booth. 1991. The fractal dimension of drifter trajectories and estimates of horizontal eddy-diffusivity. Tellus, *43A,* 334–349.

Sasaki, Y. 1970. Some basic formalism in numerical variational analysis. Mon. Weath. Rev., *98,* 738–742.

Sawyer, J. S. 1963. A semi-Lagrangian method of solving the vorticity advection equation. Tellus, *15,* 336–342.

Smolarkiewicz, P. K. 1983. A simple positive definite advection scheme with small implicit diffusion. Mon. Weath. Rev., *111,* 479–486.

Taggart, C. T. and K. T. Frank. 1991. Perspectives on larval fish ecology and recruitment processes: Probing the Scales of Relationships, *in* Large Marine Ecosystems: Patterns, Processes and Yields (AAAS Selected Symposium Series), K. Sherman and L. M. Alexander, eds., Washington, DC, Am. Assoc. Advancement Science, 151–164.

Taggart, C. T., S. E. Lochmann, D. A. Griffin, K. R. Thompson and G. L. Maillet. 1996. Abundance distribution of larval cod (*Gadus morhua*) and zooplankton in a gyre-like water mass on the Scotian Shelf, *in* Survival Strategies in Early Life Stages of Marine Resources. Y. Watanabe, Y. Yamashita and Y. Oozeki, eds., Proc. Internat. Workshop, Yokohama. Balkema Press, Rotterdam, 367 pp.

Thacker, W. C. and R. Long. 1988. Fitting dynamics to data. J. Geophys. Res., *93,* 10655–10665.

Thompson, K. R. and D. A. Griffin. 1998. A model of the circulation on the outer Scotian Shelf with open boundary conditions inferred by data assimilation. J. Geophys. Res., *103,* 30641–30660.

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

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

——1996. The Ocean Circulation Inverse Problem, Cambridge University Press, Cambridge, 442 pp.

Received: *12 November, 2001;* revised: *22 July, 2004.*