Journal of Marine Research, 59, 831–857, 2001

Journal of
MARINE RESEARCH

Volume 59, Number 6

The impact of model-error correlation on regional data assimilative models and their observational arrays

by Philip S. Bogden[1]

ABSTRACT

Data assimilative models often minimize a penalty functional that measures model adjustment and model-data misfit. The penalty functional builds assumptions about model error into the analysis. Usually, errors from different parts of the model (e.g., dynamics and boundary conditions) are presumed to be uncorrelated. This is clearly not a valid assumption in regional models where uncertain large-scale forcing affects open-ocean boundary conditions. In this study, calculations with a regional wind-driven inverse model provide a specific example where model error from uncertain wind stress is correlated with model error from uncertain open boundary conditions. This physically realistic scenario motivates development of a more general penalty functional that includes model-error correlation. In fact, model-error correlations must be included in order to meet the objective of making the open-ocean boundaries behave like the open ocean. Statistical issues for the generalized inverse model are described in the context of objective analysis. Implications for array design are addressed. For data assimilative models that incorrectly neglect model-error correlation, data should not come from open-ocean boundary regions. Rather, data should come from the interior of the regional domain. There is no such restriction on data placement for the assimilative model that correctly accounts for model-error correlation.

1. Introduction

Suppose a limited number of current meter moorings will be used to estimate coastal wind-driven currents. Suppose also that the data will be used in a regional data assimilative model to estimate the entire current field. As with many regional models, problematic open-ocean boundary conditions will be a major source of model error. Therefore, should

1. The Gulf of Maine Ocean Observation System (GoMOOS), P.O. Box 4919, Portland, Maine, 04112, U.S.A.
email: bogden@gomoos.org

the moorings be placed near the open boundaries of the regional model so as to directly influence the open boundary conditions, or should they be placed in the interior of the domain and allowed to influence the open boundary conditions from afar? Results from this study indicate that optimal mooring placement is highly sensitive to statistical assumptions incorporated into the assimilative model.

Statistical assumptions are generally used to define a penalty functional that provides a quadratic measure of model adjustment and model-data misfit; the optimal solution minimizes the penalty functional. Statistical issues are especially subtle with regional coastal models because of the open boundary conditions (OBCs). Uncertain OBCs provide the central focus for this study. The scenario presented in Section 2 is based on a regional model of central Long Island Sound that has open-ocean boundaries on two sides. Dynamical errors arise from uncertain local wind forcing, and OBC errors arise from uncertain remotely driven flows. We presume that large-scale winds provide the ultimate source of both types of model error. Other potential sources of model error are neglected in order to focus on the treatment of OBCs.

This study is based on simplified dynamics of a single shallow-water layer. Nevertheless, the open-boundary effects encountered here are characteristic of regional models with more complex dynamics. The seasonal Gulf of Maine calculation by Xue *et al.* (2000) provides an example. Their simulation exhibits unrealistic flows near the open boundary, although flows in the interior seem unaffected by the open-boundary artifacts. They use a radiation condition for barotropic flow that is similar to an OBC employed here.

Bogden *et al.* (1996) use a generalized inverse to estimate OBC forcing for a shallow-water model of the subtidal wind-driven circulation in Massachusetts Bays. Their inverse uses moored current meter data from the interior of the regional model domain. Wind is prescribed from observations, and a radiation OBC is adjusted in the inverse. Estimates of the interior flow are strongly influenced by OBC adjustments. The resulting flow estimates are robust near interior moorings. However, flows near the open boundary exhibit unrealistic features that are relatively sensitive to assumptions about OBC-error statistics.

One goal of the study described here is to determine model-error statistics that produce accurate velocity estimates throughout the regional domain. That is, model-error covariances must make the open boundary behave like the open ocean. A problem occurs because the model errors have large horizontal scale so that local winds are correlated with remote winds. Remote winds drive currents outside the model domain which subsequently affect the OBCs. Therefore, local winds are correlated with remotely driven flows at the open boundary. The implication for the data assimilative model is that errors in the local forcing are correlated with errors in the OBCs. If the two sources of model error are correlated, then data-based adjustment for one type of model error must correlate with adjustment for the other.

This kind of cross-correlation between different model errors can arise in a variety of contexts. Nevertheless, model-error correlations are rarely included in data assimilative

models. Model errors are often presumed to be autocorrelated over large length and time scales, but cross-correlation is another issue. In analogy with maximum likelihood methods, Section 3 shows how model-error correlations are incorporated into the penalty functional. Bennett (1992) states that such terms "greatly clutter the inversion and are neglected in most applications, corresponding to the assumption of uncorrelated errors. Indeed, there is little prospect of providing credible estimates for such covariances, although it is plausible that they are small." This study provides a specific counter-example to Bennett's statement. The example is particularly relevant for regional data-assimilative models of the coastal ocean.

We show that the neglect of model-error correlation has a detrimental impact on the analysis by considering a case where the exact model-error statistics are known. Exact statistics for the regional inverse model are computed from a second dynamical model whose domain includes all of Long Island Sound. Statistics are computed with OBCs for the regional model that have been extracted from the large-domain model. Such OBCs are "correct" in that they make the regional model exactly reproduce the circulation of the large-domain model. Moreover, OBC statistics computed from the large-domain model allow the regional inverse model to produce OBCs that are statistically consistent with the large-domain model. That is, with the correct model-error statistics, the open boundaries in the regional inverse model behave like the open ocean.

Results of this study, which include implications for array design, are described in the context of objective analysis (Bretherton *et al.,* 1976). We exploit the fact that the well-known technique of objective analysis is equivalent to the generalized inverse of a linear dynamical model (Bennett, 1992; Egbert and Bennett, 1996). The two seemingly different procedures for estimating velocity fields from velocity data yield the same results if they are based on the same statistics. The equivalence is not obvious, largely because of different problem formulations and statistical prescriptions. With objective analysis, one generally specifies velocity statistics. Such velocity statistics are implicit in the inverse, although they can be determined explicitly from the model and model-error statistics employed in the penalty functional. Bennett (1992) establishes the equivalence between the generalized inverse and objective analysis by showing that the penalty functional and the data distribution define a set of functions, called representers, which correspond to the velocity statistics used in standard objective analysis. The representers in this study exhibit complex structure associated with the variable coastal bathymetry. This complexity would make it very difficult to specify the velocity statistics without a numerical model.

Section 2 describes examples of the wind-driven velocity response, and its statistics, for the basic model. Regional velocities based on OBC forcing from the large-domain model are compared with those based on a standard radiation condition. The comparison sets the stage for statistical analysis of the data assimilative models.

Section 3 presents a general penalty functional that correctly accounts for model-error correlation. This general form is contrasted with a more standard penalty functional that

neglects such terms. Objective analysis motivates the comparison of the two inverse models.

Implications for array design and dynamical hypothesis testing are addressed in Section 4. Again, the implications of different penalty functionals are assessed. Results with a small array show that tests based on expected error variance may not detect the neglect of model-error correlation. Nevertheless, flow structures clearly show that optimal mooring placement depends on the accuracy of statistical assumptions used to formulate the penalty functional.

2. Basic wind-driven model

The basic model uses linear shallow-water dynamics for a single layer. In standard notation, the equations for momentum and mass conservation are

$$\begin{aligned} \frac{\partial u}{\partial t} - fv &= -g\frac{\partial \eta}{\partial x} - \frac{r}{H}u + \varepsilon_x, \\ \frac{\partial v}{\partial t} + fu &= -g\frac{\partial \eta}{\partial y} - \frac{r}{H}v + \varepsilon_y, \\ \frac{\partial \eta}{\partial t} &= -\frac{\partial (Hu)}{\partial x} - \frac{\partial (Hv)}{\partial y}. \end{aligned} \tag{1}$$

The friction parameter is $r = 0.0015$ m/s, and $H(x, y)$ is ocean depth. Here and in (3), the ε terms represent model errors whose statistics are prescribed and whose specific values are estimated with the generalized inverse. In principle, model errors can have a variety of sources including neglected physics, inaccurate parameterizations, numerical truncation, and uncertain forcing. The analysis in this study is simplified by presuming that model errors arise only from uncertain wind stress forcing $\tau = (\tau_x, \tau_y)^T$. The associated dynamical errors are $\boldsymbol{\varepsilon} = (\varepsilon_x, \varepsilon_y)^T = \tau/H$.

Two types of boundary conditions apply:

$$un_x + vn_y = 0, \tag{2}$$

$$un_x + vn_y = \sqrt{\frac{g}{H}}\,\eta + \varepsilon_o. \tag{3}$$

The first specifies no normal flow at coastal boundaries; (n_x, n_y) is the unit normal vector directed out of the domain. The second is a well-known radiation condition for open-ocean boundaries, and ε_o is the OBC error. In one dimension, with $\varepsilon_o = 0$ and constant H, Eq. (3) radiates outward-propagating gravity waves traveling at $\sqrt{gH}$. The OBC is discussed at length by Bennett (1992).

The data assimilative model in Section 3 uses velocity data from a few moorings to estimate the entire wind-driven velocity field. Since the dynamical model that relates wind

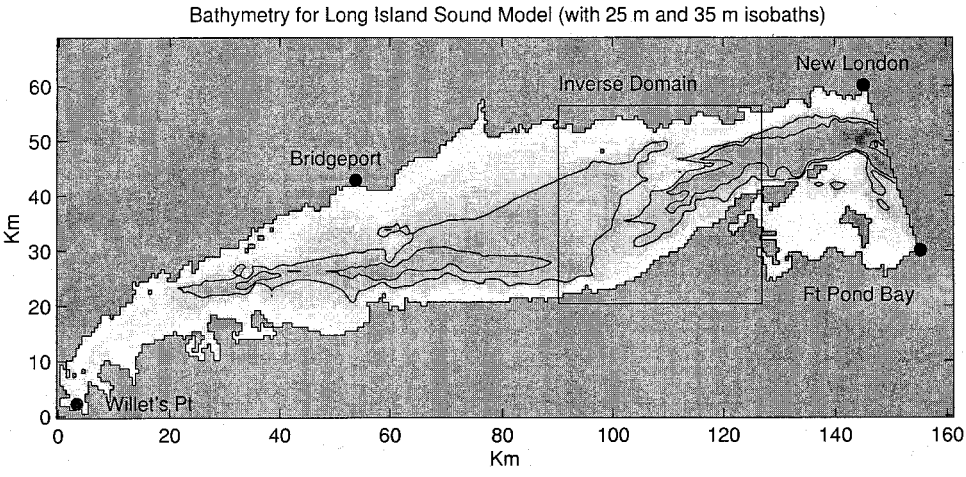

Figure 1. Model bathymetry for Long Island Sound. Gray shades darken with depth, and contours show 25 and 35 meter isobaths. The line between New London and Fort Pond Bay is an open-ocean boundary for the large domain model. The smaller square, with open ocean boundaries on the east and west, outlines the regional domain for the generalized inverse.

and velocity is linear, the problem is equivalent to estimating $\tau(\mathbf{x})$ from measurements of $\mathbf{u} = (u, v)^T$. In practice, one might have prior estimates of τ from coastal stations, buoys, or an atmospheric model. These priors would yield prior estimates of wind-driven velocity, and $\boldsymbol{\varepsilon}$ would then represent an error in the prior estimate of τ. While it is straightforward to include a nonzero prior in the analysis, we presume that the prior is zero in order to clarify the relationship between uncertain wind forcing and uncertain OBCs.

This study considers only steady-state solutions; the model is allowed enough spin-up time ($H/r \approx 1$ day) so that results do not depend on initial conditions. The steady-state simplification is not necessary. In fact, the intention is to apply the methodology to realistic time-dependent problems for which the OBC can be especially troublesome. For time-dependent regional models, some sort of radiative or absorbing condition is necessary at the open boundary in order to suppress unrealistic basin oscillations. The radiation OBC obtained with $\varepsilon_o = 0$ (Eq. (3)) has no such oscillations and produces an accurate response in the interior of the domain, as shown below. Near the open boundaries, however, the radiation OBC produces unrealistic artifacts.

Eqs. (1) are discretized on a standard C grid. Figure 1 shows bathymetry for the large-domain model. The line between New London and Fort Pond Bay is an open-ocean boundary for the large domain. The smaller square outlines the small domain used for the regional calculations and the generalized inverse. The smaller domain has complex bathymetry with open-ocean boundaries on the east and west. A relatively shallow sill (evident in the 25 meter isobath) extends from coast to coast in the center of the domain. Grid spacing is 500 meters and 750 meters in the east and north directions, respectively. The small domain is 66 by 41 grid points and the large domain is 285 by 92.

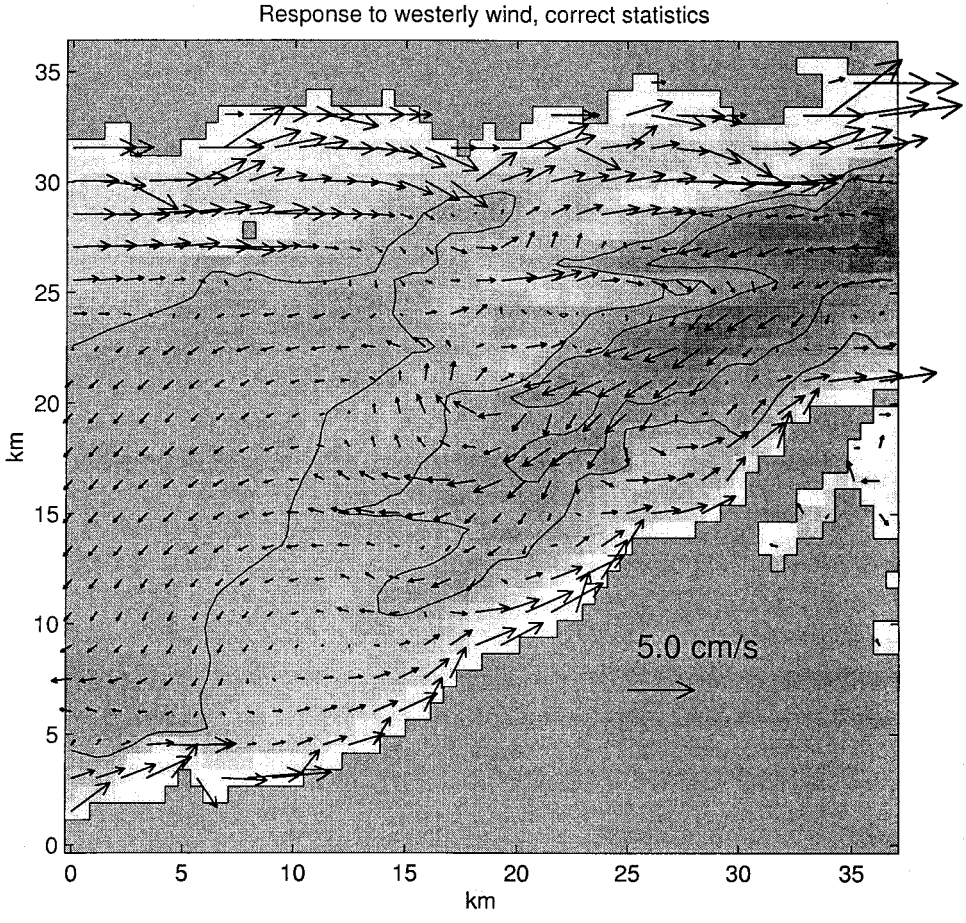

Figure 2. Large-domain response to westerly wind for the regional domain outlined in Figure 1. Velocity vectors for every other grid point are superposed on the bathymetry.

a. Response to large-scale wind stress

Figure 2 shows the large-domain response to a 1 dyne/cm^2 westerly wind. Velocity vectors for every other grid point are superposed on the bathymetry. Currents are presented only for the small domain. The response is strongest near the coasts, where currents flow parallel to shore in the downwind direction. In deeper water, large-scale pressure gradients drive upwind currents.

The maximum speeds are 7 cm/s. These are weak relative to tidal currents but comparable to residual flows associated with tidal rectification in a fully nonlinear shallow-water model. Realistic tides exhibit weak nonlinearity in this area. The linear model (1) produces tidal flow, $\mathbf{u}_{tide}$, that compares well with the observed depth-averaged tidal currents and with tidal flows in the nonlinear model. Furthermore, the linear model

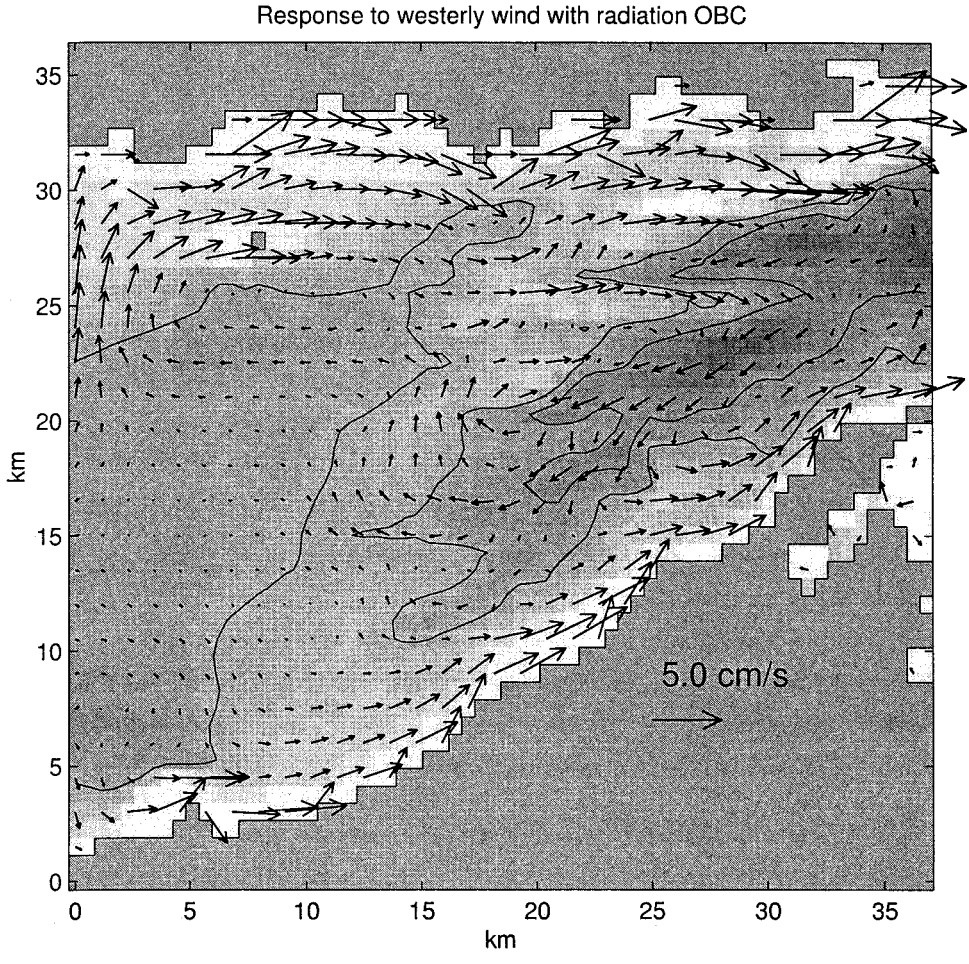

Figure 3. Response to westerly wind using the radiation condition (3) with $\varepsilon_o = 0$ on the open boundaries of this small domain. Except within roughly 5 km of the open boundaries, the response here is similar to the large-domain response in Figure 2.

forced by the time-average of $\mathbf{u}_{tide} \cdot \nabla \mathbf{u}_{tide}$ produces the same tidal residual as the fully nonlinear model (unpublished results). Thus, the linear dynamics (1) provides a plausible working hypothesis for subtidal circulation. With weak nonlinearity, the tidal residual could be included as part of the ensemble-averaged circulation. So we neglect the tidal residual without loss of generality.

Figure 3 shows the response to westerly wind for the small-domain model. This case uses the homogeneous radiation condition, $\varepsilon_o = 0$ in (3), along the open boundaries of the small domain. Away from the open boundaries, the small-domain response is nearly identical to the large-domain response in Figure 2. But the radiation condition produces

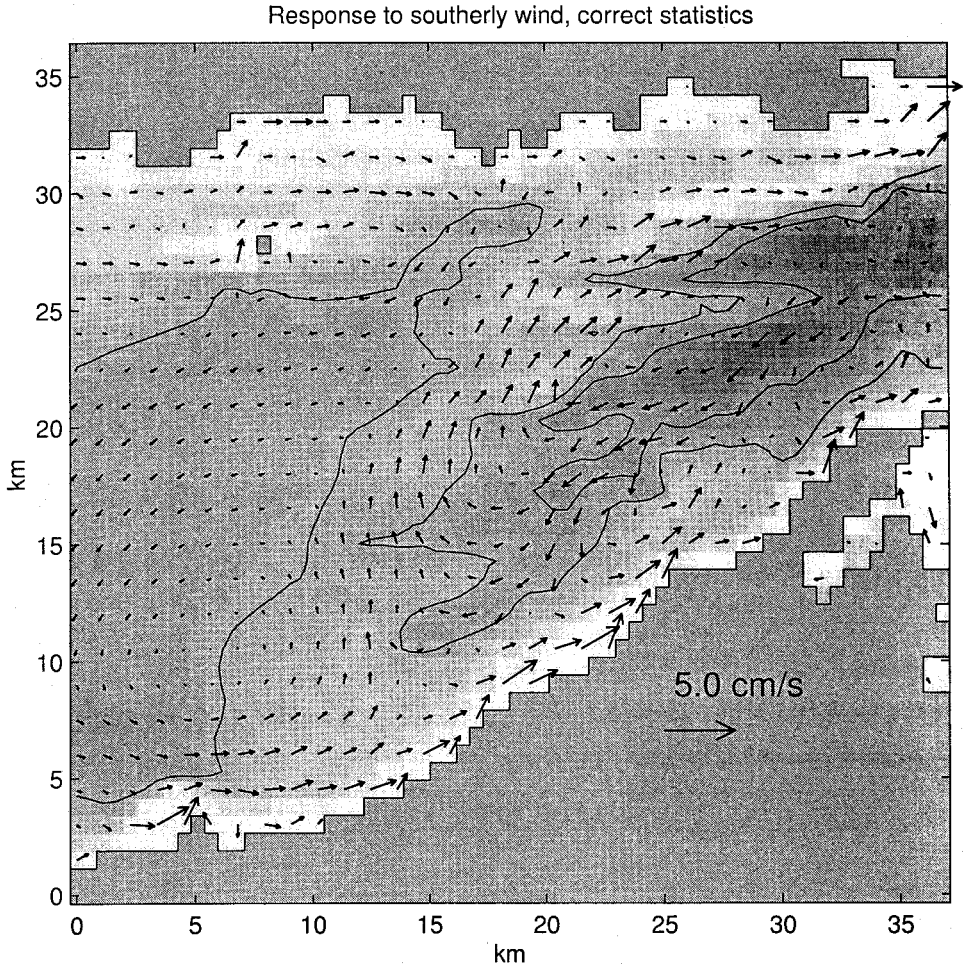

Figure 4. Large-domain response to southerly wind plotted for the small domain outlined in Figure 1.

unrealistic flow parallel to the open boundaries and suppresses some of the cross-boundary flow observed in the large-domain response. In the regional model with $\varepsilon_o = 0$, strong coastal flows come from unrealistic currents that flow parallel to the open boundary. In contrast, the strong coastal currents in the large-domain model flow across the open boundary of the regional domain.

Figure 4 shows the large-domain response to southerly wind. In most areas, wind stress is approximately balanced by a large-scale opposing pressure gradient. Consequently, the velocity response is generally smaller in magnitude than that in Figure 2. The pressure gradient drives small southwestward flow in some areas with deep water. But the largest currents are found in shallow areas where bottom drag approximately balances wind stress.

In the regional model with $\varepsilon_o = 0$, the response to southerly winds (not shown) is nearly

identical to that in Figure 4 for the interior of the small domain. As with westerly wind forcing, the biggest differences between the small- and large-domain calculations occur along the open boundary where the radiation OBC produces unrealistic artifacts.

Using the large-domain calculation as reference, ε_o can be computed so that the regional model response is identical to the large-domain response. This is done by extracting $\mathbf{u}$ and η from the large-domain calculation and then computing $\varepsilon_o(x, y)$ along the open boundary of the regional model with (3). We refer to OBCs computed this way as the "correct" OBCs. The "correct" values of ε_o are used to compute the statistics for the inverse.

b. Statistical description of the wind-driven model

The mean and covariance of velocity provide a concise description of the wind-driven model. These velocity statistics depend on the statistics of $\boldsymbol{\varepsilon} = \tau/H$. For simplicity, wind stress $\tau = (\tau_x, \tau_y)^T$ is presumed to have zero mean,

$$\langle \tau \rangle = \mathbf{0},$$

where angle brackets denote an ensemble average.

The zero-mean assumption is not necessary. Indeed, as mentioned above, one could use a nonzero time-dependent prior estimate of wind stress $\langle \tau \rangle$. The corresponding prior velocity $\langle \boldsymbol{u} \rangle$ is easily obtained by integrating the dynamical model forced by $\langle \tau \rangle$. With linear dynamics, the inverse formalism doesn't change. An accurate nonzero prior would reduce the variance of model error $\boldsymbol{\varepsilon} = \tau - \langle \tau \rangle$. Nevertheless, as long as the model errors have long time and length scales, which is usually the case, the results of this study continue to apply. Thus, we consider the zero-mean case without loss of generality.

The wind stress is presumed to be homogeneous and isotropic with covariance

$$\langle \tau(\mathbf{x}_1)\tau(\mathbf{x}_2)^T \rangle = \sigma^2 \mathbf{I},$$

where $\sigma^2 = 1\ (\text{dyne/cm}^2)^2$ is wind stress variance, and $\mathbf{I}$ is the 2-by-2 identity matrix. We consider steady-state forcing, so all time-lagged autocorrelations are unity. Length scales for wind stress are presumed to be larger than the model domain, so τ^x and τ^y are perfectly autocorrelated at all space lags, but uncorrelated with each other. Because of the large length and time scales, wind stress can be expressed with two parameters, and any covariance matrix based on multiple wind-stress measurements has rank 2.

The dynamical errors, $\boldsymbol{\varepsilon} = \tau/H(\mathbf{x})$, have correspondingly simple statistics, with zero mean and an isotropic and diagonal covariance,

$$\langle \boldsymbol{\varepsilon}(\mathbf{x}_1)\boldsymbol{\varepsilon}(\mathbf{x}_2)^T \rangle = \frac{\sigma^2}{H(\mathbf{x}_1)H(\mathbf{x}_2)} \mathbf{I}.$$

As with wind stress, model errors are perfectly autocorrelated in time and space. Dependence on $H(\mathbf{x})$ makes the variance inhomogeneous.

The dynamical model and wind stress statistics can be used to compute covariances $\langle \boldsymbol{u}(\mathbf{x}_1)\boldsymbol{u}(\mathbf{x}_2)^T \rangle$ for velocity. This could be done with a pseudo-ensemble average of velocity

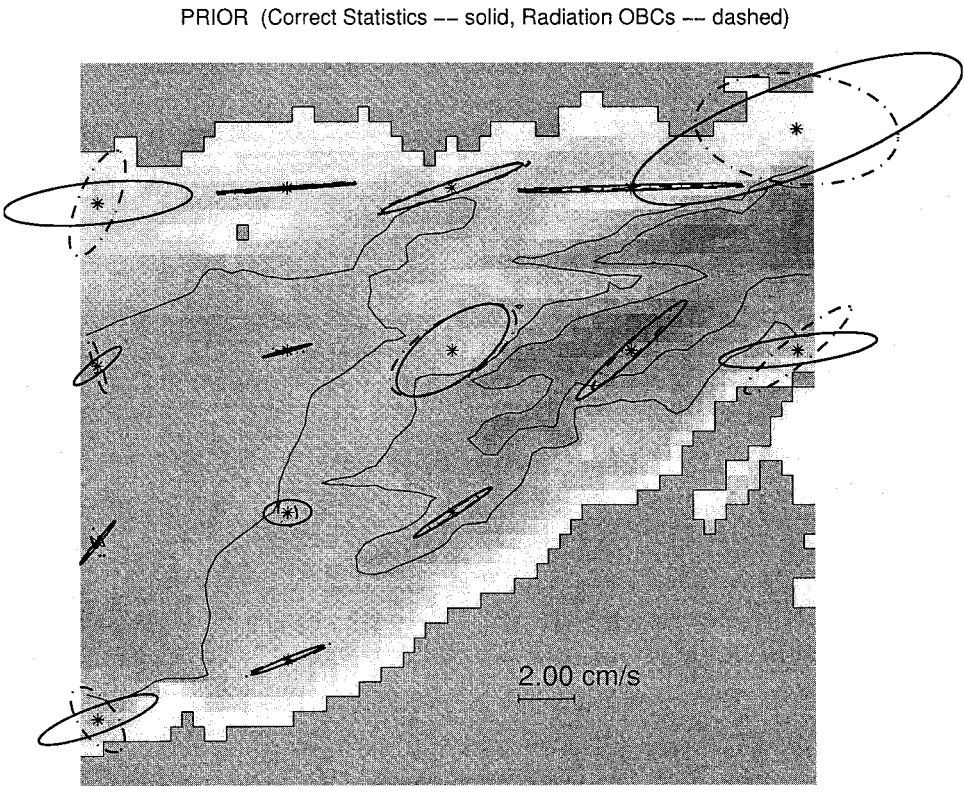

Figure 5. Velocity ellipses for response to homogeneous and isotropic large-scale wind stress. Ellipse radii are proportional to the standard deviation in the response. The scale is indicated in the lower right of the figure for a wind stress with standard deviation of 1 dyne/cm^2. The solid ellipses are associated with the large domain response. The dashed ellipses are the small domain response using a radiation condition, $\varepsilon_o = 0$ in (3), at the open boundary of this small domain.

responses. The alternative method used here is based on representers, as described in the Appendix (see also, Bennett, 1992). Representers have an important statistical interpretation that is described in the next section.

The full velocity covariance is difficult to visualize because it is a complex function of two space variables. However, most of the features in the wind-driven velocity are captured by plotting the variance $\langle \boldsymbol{u}(\mathbf{x})\boldsymbol{u}(\mathbf{x})^T \rangle$ at a few isolated locations. Variances for the large-domain and regional models are plotted in Figure 5. Ellipse radii are proportional to one standard deviation of the velocity, and the velocity scale is included in the lower right of the figure.

The solid ellipses in Figure 5 characterize the "correct" wind-driven currents for the large-domain model in Figures 2 and 4. The largest ellipses occur in the shallow water near the coast where the current response is strongest. The response would be isotropic and

velocity ellipses would be circles if wind stress were exactly balanced by bottom drag. But such a balance only approximately describes shallow regions in the middle of the domain. Near the coast and in deep water, velocity ellipses are highly anisotropic. These anisotropies arise because the topography supports large-scale pressure gradients that can dominate wind stress. This complex velocity variance contrasts with the simple homogeneous isotropic winds.

The dashed ellipses in Figure 5 represent velocity statistics for the regional wind-driven model using the radiation OBC obtained when $\varepsilon_o = 0$ in (3). The large dashed and solid ellipses are nearly identical away from the open boundaries. This is because the velocity response in the interior of the regional model is only weakly affected by the OBCs. In contrast, along the open boundary velocity ellipses are substantially different. Unlike the large-domain response, the radiation OBC allows only weak flow across the open boundary and produces relatively strong unrealistic flow parallel to the open boundary.

Statistics of the small-domain model are "correct" when they agree exactly with those of the large-domain model. "Correct" values of ε_o are computed from the large-domain model. The "correct" statistics involve correlations between OBC errors ε_o and forcing errors ε_x and ε_y. These statistics are computed with an ensemble average in which a westerly wind is associated with ε_o from the large-domain response to westerly wind, and likewise for wind in any other direction. These model-error statistics are easy to compute because of the simplicity of the wind statistics. The relatively complex velocity covariances are easily computed with the representer calculation described in the Appendix. When the "correct" cross terms $\langle \varepsilon_o \varepsilon_x \rangle$ and $\langle \varepsilon_o \varepsilon_y \rangle$ are used, velocity ellipses for the regional model are identical to the solid ellipses in Figure 5.

Figure 6 emphasizes the importance of the cross terms $\langle \varepsilon_o \varepsilon_x \rangle$ and $\langle \varepsilon_o \varepsilon_y \rangle$. The solid ellipses in Figures 5 and 6 are identical and "correct." The dashed ellipses in Figure 6 differ only because they involve the incorrect assumption of uncorrelated model errors, $\langle \varepsilon_o \varepsilon_x \rangle = \langle \varepsilon_o \varepsilon_y \rangle = 0$. This case is equivalent to a regional model that is independently forced by local wind stress and OBC inhomogeneities. Velocity ellipses in the center of the domain are not sensitive to the OBC influence. But near the open boundary, the neglect of cross terms results in qualitatively different velocity statistics. In some places, flows near the open boundary are more isotropic and more energetic than they are with the correct statistics. In other places, open-boundary artifacts are highly anisotropic and oriented parallel to the open boundary.

3. Data assimilation with the regional wind-driven model

Model-error statistics are incorporated into the generalized inverse through the penalty functional. Each penalty functional in this study has a unique and easily obtained minimum (see the Appendix). The objective here is to construct a penalty functional for the regional model so that the OBCs drive the "correct" response observed in the large-domain model.

Most data assimilative models use a penalty functional based on quadratic measures of

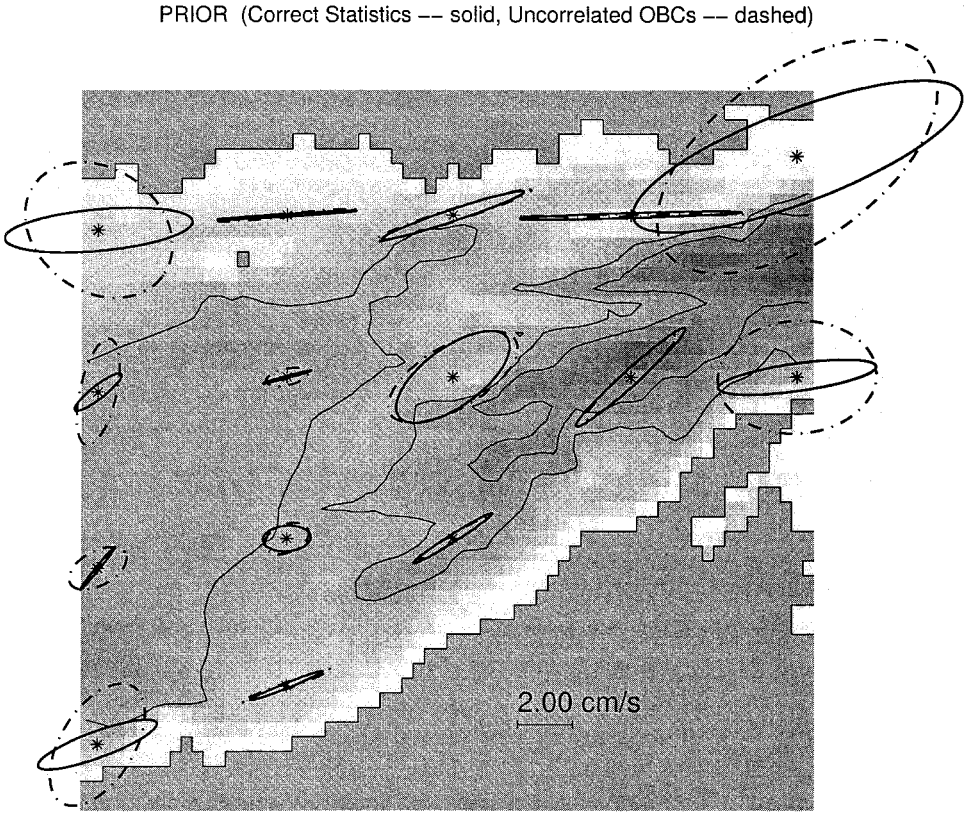

Figure 6. The velocity ellipses for correct statistics (solid) are identical to those in Figure 5. The dashed ellipses are based on the same prior statistics, except the OBC errors are presumed to be uncorrelated with model errors.

model-data misfit and model-adjustment. Using a notational shorthand, a standard form for the wind-driven model is

$$J_{standard}[u, v] = \varepsilon_x \bullet C_{xx}^{-1} \bullet \varepsilon_x + \varepsilon_y \bullet C_{yy}^{-1} \bullet \varepsilon_y + \varepsilon_o \circ C_{oo}^{-1} \circ \varepsilon_o + \mathbf{e} \cdot \mathbf{C}_{ee}^{-1}\mathbf{e}. \qquad (4)$$

The solid and open circles represent multiple integrals. For arbitrary functions $a(\mathbf{x}, t)$ and $b(\mathbf{x}, t)$, the solid circle is an integral over all model space $\mathbf{x}$ and time t,

$$a \bullet b \equiv \int d\mathbf{x} \int_0^T dt\, a(\mathbf{x}, t) b(\mathbf{x}, t).$$

The open circle represents an integral over time and along the open boundary,

$$a_o \circ b_o \equiv \int_o ds \int_0^T dt\, a(s, t) b(s, t).$$

For covariance functions such as $C_{xx}(\mathbf{x}_1, t_1, \mathbf{x}_2, t_2) = \langle \varepsilon_x(\mathbf{x}_1, t_1)\varepsilon_x(\mathbf{x}_2, t_2)\rangle$ the integral operators can be applied twice,

$$a \bullet C \bullet b \equiv \int d\mathbf{x}_1 \int dt_1 \int d\mathbf{x}_2 \int dt_2 a(\mathbf{x}_1, t_1) C(\mathbf{x}_1, t_1, \mathbf{x}_2, t_2) b(\mathbf{x}_2, t_2).$$

The superscripts in 4 represent functional inverses defined as follows,

$$C_{xx} \bullet C_{xx}^{-1} = \delta(x_1 - x_2)\delta(y_1 - y_2)\delta(t_1 - t_2).$$

As shown in the Appendix, C_{xx} must be prescribed but it is never necessary to evaluate C_{xx}^{-1}.

The first term on the right-hand side of (4) is an integral of squared error in the x-momentum equation weighted by its inverse covariance $C_{xx}^{-1}(\mathbf{x}_1, t_1, \mathbf{x}_2, t_2)$. The second and third terms are contributions from ε_y and ε_o. The last term is a weighted sum of the squared model-data misfit, where

$$\mathbf{d} = \mathbf{L}[u, v] + \mathbf{e}$$

is an N-vector of data values. Each element of $\mathbf{L}[u, v]$ is a component of model velocity evaluated at the location and time of the corresponding datum, and $\mathbf{e}$ is the corresponding measurement-error vector. The model-data misfit is presumed to come from uncorrelated, homogeneous and isotropic measurement errors, so

$$\mathbf{C}_{ee} \equiv \langle \mathbf{e}\mathbf{e}^T \rangle = \sigma_{ee}^2 \mathbf{I},$$

with constant variance σ_{ee}^2.

The model-error covariances for ε_x and ε_y are defined in the previous section. The covariance for OBC error ε_o in the regional model is computed with values of $\varepsilon_o = un^x + vn^y - \sqrt{g/H}\eta$ extracted from the large-domain wind-driven model.

Up to this point, the development closely follows that given by Bennett (1992), where $\mathbf{L}[u, v]$ would be referred to as a vector of linear evaluation functionals. The penalty functional (4) is "standard" because it neglects model-error cross products between ε_o, ε_x, and ε_y. Neglect of such cross terms corresponds to the assumption that model errors are uncorrelated, $\langle \varepsilon_x \varepsilon_y \rangle = \langle \varepsilon_o \varepsilon_x \rangle = \langle \varepsilon_o \varepsilon_y \rangle = 0$. The "correct" penalty functional includes the cross terms,

$$J_{correct} = \begin{bmatrix} \varepsilon_x \\ \varepsilon_y \\ \varepsilon_o \end{bmatrix}^T \begin{bmatrix} \bullet C_{xx} \bullet & 0 & \bullet C_{xo} \circ \\ 0 & \bullet C_{yy} \bullet & \bullet C_{yo} \circ \\ \bullet C_{xo} \circ & \bullet C_{yo} \circ & \circ C_{oo} \circ \end{bmatrix}^{-1} \begin{bmatrix} \varepsilon_x \\ \varepsilon_y \\ \varepsilon_o \end{bmatrix} + \mathbf{e} \cdot \mathbf{C}_{ee}^{-1}\mathbf{e}. \tag{5}$$

In this more general form, individual model errors are elements of a model-error vector, and the inverse of the model-error covariance matrix involves a combination of vector and integral operations. The inverse operator is symbolic in this equation. As shown in the Appendix, $J_{correct}$ can be minimized without evaluating the matrix inverse or inverse covariances in (5). We use the so-called representer method (Bennett, 1992). The representer method is elegant, and it can be extremely efficient. With linear dynamics there is no iteration. See the Appendix for details.

a. Generalized inverse and objective analysis

The focus here is on the statistical interpretation of representers and their analogs in objective analysis. We begin by restating the data assimilation problem as an example of objective analysis.

Objective analysis provides the linear minimum-variance unbiased estimate $\hat{\mathbf{u}}(\mathbf{x}, t)$ of $\mathbf{u}(\mathbf{x}, t)$ based on the available data $\mathbf{d}$. That is, $\hat{\mathbf{u}}(\mathbf{x}, t)$ minimizes $E = \langle(\hat{\mathbf{u}} - \mathbf{u})^2\rangle$. The Gauss-Markov theorem states that

$$\hat{\mathbf{u}}(\mathbf{x}, t) = \langle \mathbf{u}(\mathbf{x}, t)\mathbf{d}^T\rangle \mathbf{C}_{dd}^{-1}\mathbf{d}, \tag{6}$$

where $\mathbf{C}_{dd} = \langle \mathbf{d}\mathbf{d}^T\rangle$ is the data-data covariance. $\mathbf{C}_{dd}$ is the sum of two components,

$$\mathbf{C}_{dd} = \mathbf{R} + \mathbf{C}_{ee}.$$

$\mathbf{R}$ is associated with measurements of the velocity field that is being mapped. That is,

$$\mathbf{R} = \langle \mathbf{L}[u, v]\mathbf{L}[u, v]^T\rangle$$

is the covariance of noise-free data.

This optimal estimate $\hat{\mathbf{u}}$ in (6) is generally derived by first writing $\hat{\mathbf{u}}(\mathbf{x}, t)$ as a linear combination of the data, $\hat{\mathbf{u}}(\mathbf{x}, t) = \sum_{i=1}^{N} \alpha_i d_i$. The optimal data weights $\alpha_i(\mathbf{x}, t)$ are then obtained by minimizing E. But it is instructive to consider an equivalent expression for the optimal estimate,

$$\hat{\mathbf{u}}(\mathbf{x}, t) = \sum_{i=1}^{N} \beta_i \langle d_i \mathbf{u}(\mathbf{x}, t)\rangle, \tag{7}$$

where $\beta_i = (\mathbf{C}_{dd}^{-1}\mathbf{d})_i$ are data-dependent weights for the covariance functions $\langle d_i\mathbf{u}(\mathbf{x}, t)\rangle$. When measurement errors are uncorrelated with model errors, as presumed here, then $\langle d_i\mathbf{u}(\mathbf{x}, t)\rangle = \langle u(\mathbf{x}_i, t_i)\mathbf{u}(\mathbf{x}, t)\rangle$ when d_i is a measurement of u, and $\langle d_i\mathbf{u}(\mathbf{x}, t)\rangle = \langle v(\mathbf{x}_i, t_i)\mathbf{u}(\mathbf{x}, t)\rangle$ when d_i is a measurement of v.

In the generalized inverse, the covariances between the velocity field and the velocity measurements, $\langle u(\mathbf{x}_i, t_i)\mathbf{u}(\mathbf{x}, t)\rangle$ and $\langle v(\mathbf{x}_i, t_i)\mathbf{u}(\mathbf{x}, t)\rangle$, are referred to as "representers" (Bennett, 1992). There is one for each scalar datum. The measurement covariance $\mathbf{R}$ can be computed by measuring the representers,

$$\mathbf{R}_{ij} \equiv \mathbf{L}_j[\langle d_i \mathbf{u}(\mathbf{x})\rangle].$$

In the generalized inverse, **R** is that part of the data-data covariance associated with the model adjustments. **R** is referred to as the representer matrix.

The so-called "representer method" uses representers to minimize J elegantly and efficiently. Egbert and Bennett (1996) describe an extremely efficient algorithm. But it is the statistical interpretation of representers that is important here. In particular, objective analysis and the generalized inverse are equivalent when $\langle d_i \mathbf{u}(\mathbf{x}, t)\rangle$ and $\mathbf{C}_{dd}$ are consistent with the model and the model-error statistics that define J. In this case, the optimal estimate $\hat{\mathbf{u}}(\mathbf{x}, t)$ is the velocity field that minimizes J.

With objective analysis, the standard approach is to prescribe the covariances $\langle d_i \mathbf{u}(\mathbf{x}, t)\rangle$ and $\mathbf{C}_{dd}$, compute β_i based on the data, and then compute the sum in (7). Bretherton *et al.* (1976) could estimate velocity covariances with data because statistics of the eddy field are fairly simple. However, $\langle d_i \mathbf{u}(\mathbf{x}, t)\rangle$ and $\mathbf{C}_{dd}$ for the coastal wind-driven problem are complex functions of **x**. These statistics are difficult to estimate with the sparse data sets available to most oceanographers.

With the generalized inverse, however, one prescribes the relatively simple statistics of $\boldsymbol{\varepsilon}$ and ε_o and then easily computes $\langle d_i \mathbf{u}(\mathbf{x}, t)\rangle$ and **R** as described in the Appendix. It is worth emphasizing that each penalty functional is associated with its own set of representers. That is, each penalty functional implies a unique set of velocity covariances and data-data covariances. These covariances form the basis of the following analysis.

b. Representers

Figure 7 shows a representer based on the incorrect assumption of uncorrelated model errors, $C_{xo} = C_{yo} = 0$, used in $J_{standard}$. The datum, d_i, for this representer is a measurement of u near the open boundary in the northwest corner of the regional domain; the datum location is indicated by a solid circle. The largest amplitude flows occur near the western open boundary, with relatively weak flow in the interior. The representer is dominated by unrealistic flow near the western open boundary that is qualitatively different from any of the correct responses in the wind-driven model.

In contrast, when the correct model-error covariances C_{xo} and C_{yo} are used with $J_{correct}$, the representer structure (not shown) is barely distinguishable from large-domain response to a westerly wind shown in Figure 2. Consequently, a data assimilative model using $J_{standard}$ and this single datum would produce reasonable flows throughout the regional domain, including the cross-boundary flows on the opposite side of the domain.

This seemingly remarkable result is actually quite reasonable. The correct values of C_{xo} and C_{yo} make the regional model consistent with the large-domain model. In this case, any model adjustment must give rise to a linear combination of the two large-scale fields in Figures 2 and 4. The response to westerly winds dominates the structure in the representer for $J_{correct}$ because it is the strongest response to the isotropic winds.

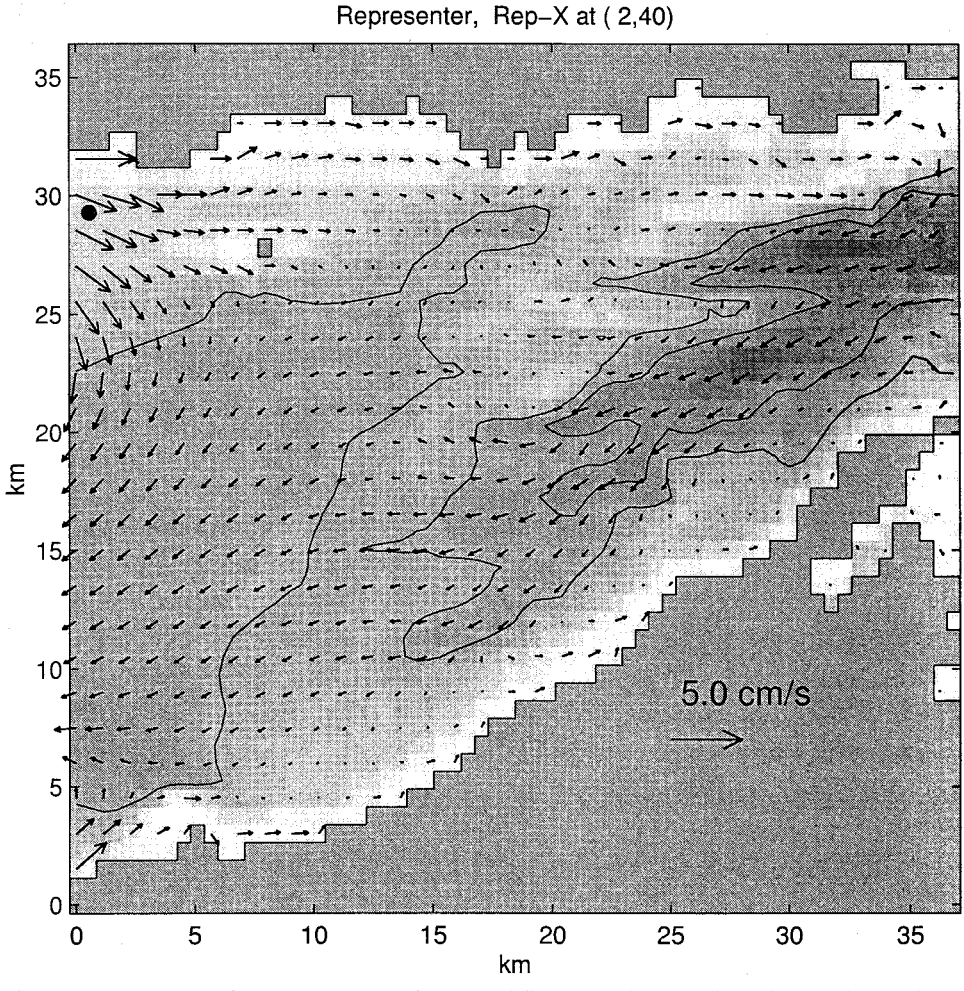

Figure 7. Representer for measurement of eastward flow near the open boundary at the northeast corner of the regional model domain. The measurement location is indicated by a solid circle. The representer has been normalized by one standard deviation of the model measurement.

When the datum is located near the center of the regional domain, away from the open boundaries, the representer for $J_{correct}$ again looks like the large-domain response to westerly wind forcing. However, the incorrect assumption $C_{xo} = C_{yo} = 0$ in $J_{standard}$ produces a representer that looks like the response to wind forcing with a homogeneous OBC, $\varepsilon_o = 0$, as shown in Figure 3. This makes sense because the interior of the domain is weakly affected by the OBC forcing, and the representer includes only the flows that are correlated with the datum. Therefore, when $C_{xo} = C_{yo} = 0$, a datum in the interior of the domain will be strongly correlated with flows driven by local wind, but uncorrelated with OBC forcing.

To summarize, unrealistic results with one datum are entirely attributable to setting $C_{xo} = C_{yo} = 0$. For $J_{standard}$, all datum locations produce open boundary artifacts. Datum placement along the open boundary will produce unrealistic flows even in close proximity to the datum. With the datum in the interior, solutions with $J_{standard}$ will resemble those using the radiation condition $\varepsilon_o = 0$; interior flows near the datum will be accurate and boundary regions will have artifacts. Thus, with $J_{standard}$, the preferred datum location is in the interior of the domain, away from the effects of problematic OBCs.

With $J_{correct}$ and the correct values of C_{xo} and C_{yo}, solutions are much less sensitive to datum placement. The assimilative model will have realistic large-scale flows throughout the domain, even across the open boundaries. In this case, it makes sense to place the datum in a region where the wind-driven flows are large, so as to maximize the signal to noise ratio.

c. Analysis of mooring distributions

In order to generalize these results to arrays with multiple moorings, consider the data-data covariance matrix, $\mathbf{C}_{dd} = \mathbf{R} + \mathbf{C}_{ee}$ for the special noise-free data, $\mathbf{C}_{ee} = 0$, so that $\mathbf{C}_{dd} = \mathbf{R}$. The eigenvector-eigenvalue decomposition of $\mathbf{R}$ is

$$\mathbf{R} = \mathbf{U}\Lambda\mathbf{U}^T,$$

and the unitary eigenvector matrix $\mathbf{U}$ can be normalized,

$$\mathbf{Z} = \Lambda^{1/2}\mathbf{U}^T,$$

so that $\mathbf{Z}$ has units of velocity and $\mathbf{R} = \mathbf{Z}\mathbf{Z}^T$. The columns of $\mathbf{Z}$ are the empirical orthogonal functions (EOFs) of noise-free data. If $\mathbf{C}_{ee}$ is nonzero and diagonal, as often presumed, then $\mathbf{C}_{dd}$ and $\mathbf{R}$ have the same eigenvectors.

Let $\mathbf{z}$ denote the dominant EOF associated with the largest eigenvalue λ. Since $\mathbf{z} = \lambda^{-1}\mathbf{R}\mathbf{z}$, the representers can be used to relate $\mathbf{z}$ to the entire velocity field as follows:

$$\mathbf{z} = \lambda^{-1} \sum_i \mathbf{L}[\langle u(\mathbf{x}_i, t_i)\mathbf{u}(\mathbf{x}, t)\rangle] z_i = \mathbf{L}\left[\lambda^{-1} \sum_i z_i \langle u(\mathbf{x}_i, t_i)\mathbf{u}(\mathbf{x}, t)\rangle\right]. \tag{8}$$

Bennett (1992) refers to the linear combination of representers $\lambda^{-1} \Sigma_i z_i \langle u_i \mathbf{u}\rangle$ as a "solution antenna mode." Eq. (8) shows that $\mathbf{z}$. is recovered by measuring the solution antenna mode with $\mathbf{L}$. Thus, an antenna mode is the velocity field associated with an EOF, and the antenna mode equals the EOF at the data locations.

Figure 8 shows the dominant antenna mode for $J_{standard}$ and an array of 4 moorings located near the open boundaries of the regional model. The data locations are indicated by solid circles. There is strong cross-boundary flow at the eastern end of the domain and strong flow in the interior. Both features are qualitatively consistent with westerly wind forcing. However, flows at the western open boundary are unrealistically oriented along the

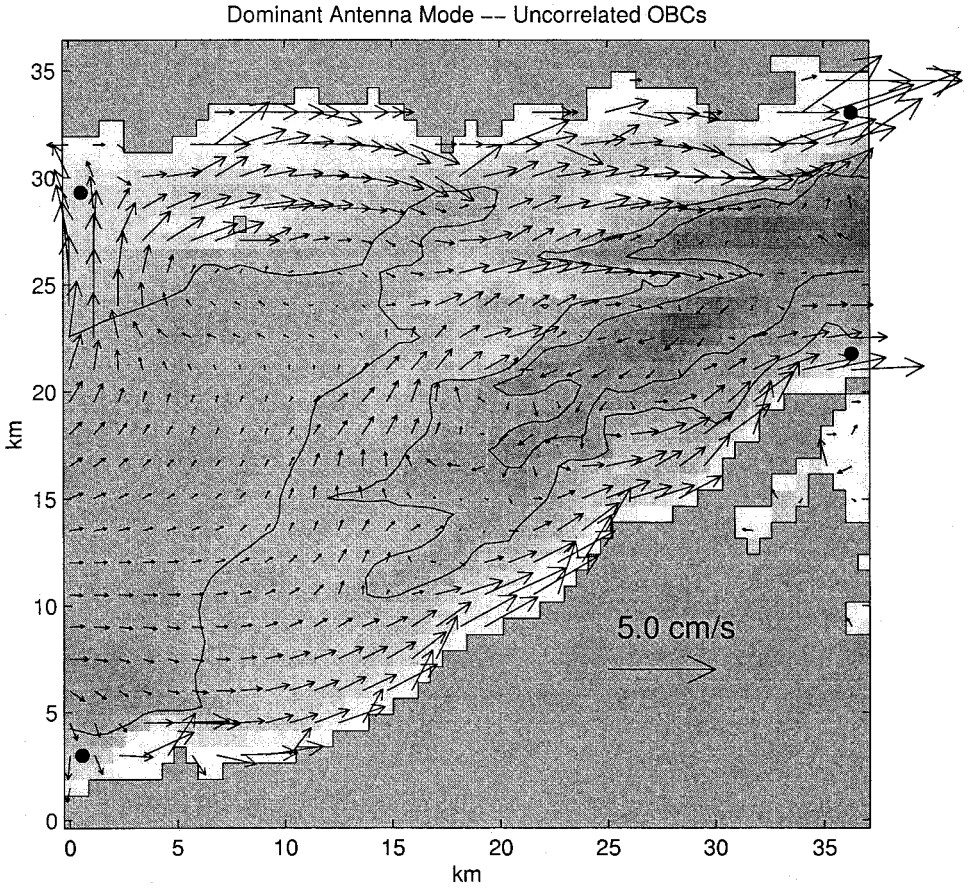

Figure 8. Dominant antenna mode for $J_{standard}$ and an array of moorings near the open boundaries. Data locations are indicated by solid circles.

boundary. These western open-boundary flows are characteristic of the homogeneous OBC with $\varepsilon_o = 0$.

Figure 9 shows the dominant antenna mode again based on $J_{standard}$ but with the moorings located in the interior of the domain. The structure is similar to that with moorings near the open boundary, except this antenna mode has relatively strong flow in the interior and near the solid boundary.

Figure 10 shows the dominant antenna mode based on $J_{correct}$ and the interior mooring distribution. There are no open-boundary artifacts in this case. With correct model-error statistics, the dominant antenna mode depends only weakly on the mooring locations.

In summary, the antenna mode structure in the interior of the domain is not sensitive to neglect of model-error correlations when the moorings are located in the interior. $J_{standard}$ and $J_{correct}$ will yield similar solutions near the interior data locations, but $J_{standard}$ will exhibit open-boundary artifacts. This contrasts the sensitivity that occurs with moorings

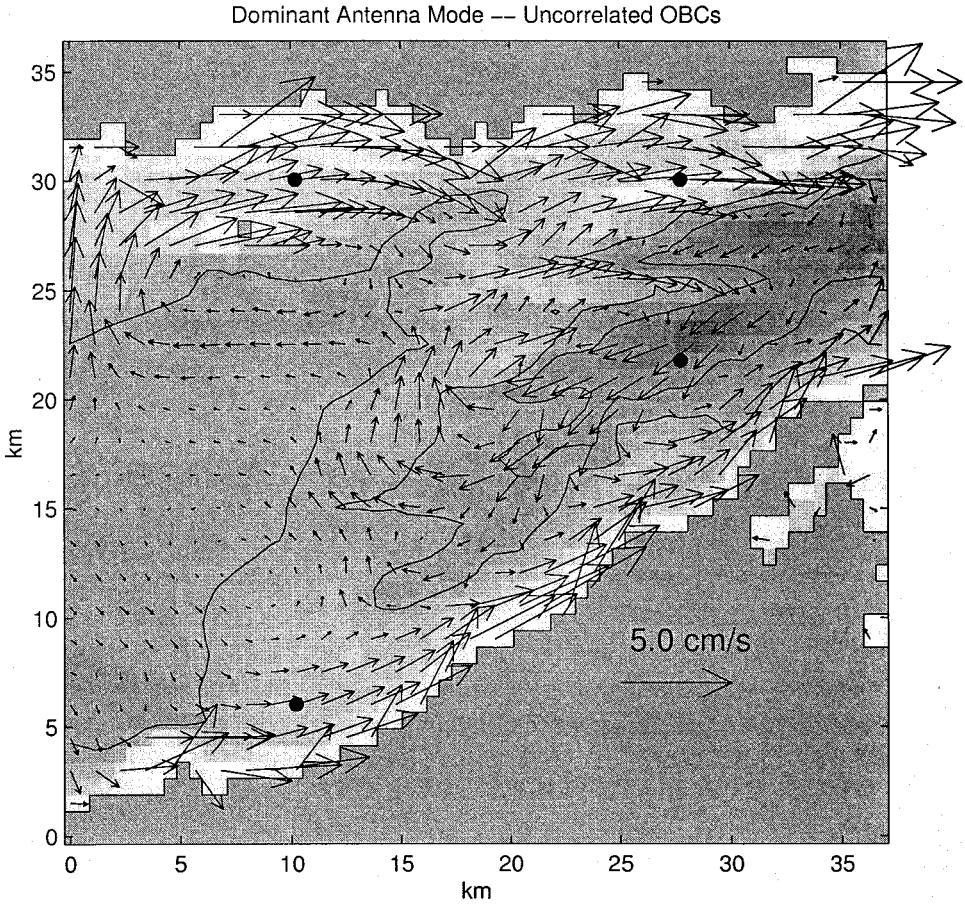

Figure 9. Dominant antenna mode for $J_{standard}$ and a 4-mooring array in the interior of the domain. As in Figure 8, the model-error statistics are based on the incorrect assumption $C_{uo} = C_{vo} = 0$.

near the open boundaries. With an open-boundary array and $C_{uo} = C_{vo} = 0$, the dominant antenna mode has weak interior flow, and strong unrealistic flows near the moorings. With correct statistics, antenna modes for $J_{correct}$ are not sensitive to mooring placement.

In practice, antenna modes should have accurate structure in the vicinity of the data. Otherwise, measurement errors that project onto these modes will produce unrealistic flow fields. That is, unrealistic antenna-mode structure makes the assimilative model sensitive to noise in the data. Therefore, with a data assimilative model based on $J_{standard}$, moorings should not be placed near the open boundary.

d. Testing dynamical hypotheses

The statistical assumptions used to define the penalty functional comprise a null hypothesis about the model and its errors. It is desirable to have an objective test of such

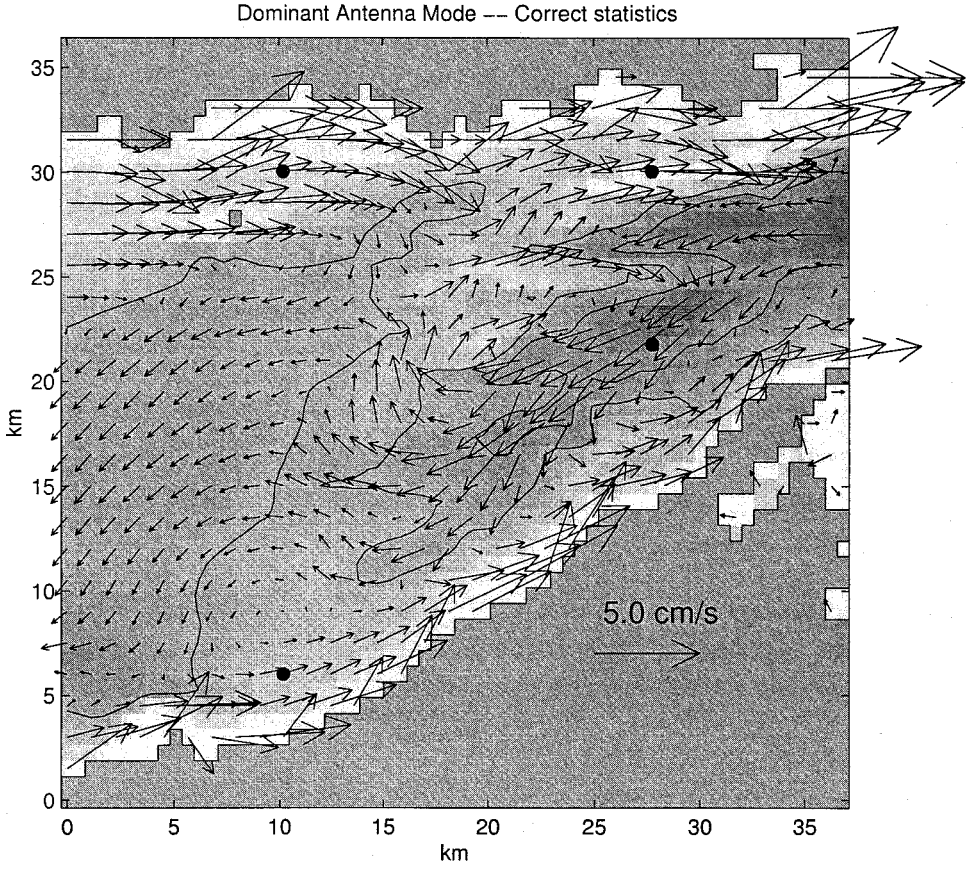

Figure 10. Dominant antenna mode for correct statistics, $J_{correct}$, and moorings in the interior of the domain. The antenna mode for moorings near the open boundaries is nearly identical to the one shown here.

hypotheses. Even the "correct" statistics in our scenario comprise a hypothesis worthy of testing because the large-domain model may have significant errors.

The data-dependent posterior penalty functional $\hat{J}$ provides a test statistic for the statistical assumptions that are built into J. Bennett (1992) shows that $\hat{J} = \mathbf{d}\mathbf{C}_{dd}^{-1}\mathbf{d}$ for zero-mean data, as presumed here. Summation notation can be used to rewrite this as $\hat{J} = Tr[\mathbf{C}_{dd}^{-1/2}\mathbf{d}^T\mathbf{d}\mathbf{C}_{dd}^{-1/2}]$. When the null hypothesis is correct, that is, when the model-error covariances from the large-domain model are accurate (as presumed here), then $\mathbf{C}_{dd} \equiv \langle \mathbf{d}\mathbf{d}^T \rangle = \mathbf{R} + \mathbf{C}_{ee}$, so

$$\langle \hat{J}_{correct} \rangle = Tr[\mathbf{C}_{dd}^{-1/2} \langle \mathbf{d}^T\mathbf{d} \rangle \mathbf{C}_{dd}^{-1/2}] = N,$$

where N equals the number of observations. Furthermore, if the model errors satisfy a normal distribution (or if one appeals to the central limit theorem) then $\hat{J}$ is a χ^2 random

Table 1. Expected values of the posterior penalty functional for four different null hypotheses, and two different moorings arrays. The first four cases are based on a 4-mooring array near the open boundary, and cases 5–6 are based on moorings in the interior of the regional domain. $J_{correct}$ uses the correct statistics, $J_{standard}$ neglects correlation of different model errors, $J_{interior}$ allows adjustment of the wind forcing but keeps the OBCs fixed, and J_{OBC} allows adjustment of the OBC but makes interior dynamics a strong constraint. The right-most column shows the velocity variance at the moorings based on each null hypothesis.

		$\langle \hat{J} \rangle$				
Case	Null hypothesis	$\sigma_{ee}^2 = 10$	$\sigma_{ee}^2 = 1$	$\sigma_{ee}^2 = 0.1$	$\sigma_{ee}^2 = 0.01$	$(1/N)Tr[\mathbf{R}]$ (cm/s)^2
1	$\langle \hat{J}_{correct} \rangle$	8.0	8.0	8.0	8.0	8.8
2	$\langle \hat{J}_{standard} \rangle$, $C_{uo} = C_{vo} = 0$	7.9	8.6	14.6	65.3	9.8
3	$\langle \hat{J}_{interior} \rangle$, $\varepsilon_o = 0$	11.1	25.7	240.4	2663.7	3.9
4	$\langle \hat{J}_{OBC} \rangle$, $\varepsilon_u = \varepsilon_v = 0$	10.3	19.9	175.0	1873.8	5.8
5	$\langle \hat{J}_{correct} \rangle$	8.0	8.0	8.0	8.0	4.1
6	$\langle \hat{J}_{standard} \rangle$, $C_{uo} = C_{vo} = 0$	8.0	7.7	7.1	6.5	4.0
7	$\langle \hat{J}_{interior} \rangle$, $\varepsilon_o = 0$	8.1	8.6	13.3	60.5	3.8
8	$\langle \hat{J}_{OBC} \rangle$, $\varepsilon_u = \varepsilon_v = 0$	11.1	38.9	306.3	2445.6	0.2

variable with N degrees of freedom. Thus, standard statistical tests can be used to test the null hypothesis that defines J.

Let $\mathbf{C}_{standard}$ denote the data-data covariance matrix based on the incorrect null hypothesis associated with $J_{standard}$. The expected value of the posterior penalty functional is

$$\langle \hat{J}_{standard} \rangle = Tr[\mathbf{C}_{standard}^{-1/2} \mathbf{C}_{dd} \mathbf{C}_{standard}^{-1/2}] \neq N.$$

If the difference between $\langle \hat{J}_{standard} \rangle$ and $\langle \hat{J}_{correct} \rangle$ is large compared to $\sqrt{N}$ (the standard deviation of χ_N^2), then $\hat{J}$ should provide a sensitive statistic to test for erroneous neglect of correlated errors.

Table 1 shows $\langle \hat{J} \rangle$ computed for two different 4-mooring arrays, and four different null hypotheses. Cases 1 through 4 are based on moorings near the open boundaries, as in Figure 8. Case 1 uses the correct statistics, so $\langle \hat{J}_{correct} \rangle = 8$ because four 2-component vectors provide $N = 8$ scalar measurements. Note also that $\langle \hat{J}_{correct} \rangle = 8$ regardless of the measurement-error variance σ_{ee}^2.

Case 2 is based on the data-data covariance $\mathbf{C}_{standard}$ associated with $J_{standard}$, so $C_{uo} = C_{vo} = 0$. In this case, $\langle \hat{J}_{standard} \rangle$ and $\langle \hat{J}_{correct} \rangle$ are not substantially different for reasonable values of measurement error. They only become distinguishable for rms measurement errors much less than 1 cm/s, which is unreasonably small for most situations.

Case 3 is based on an assimilative model, $J_{interior}$, that allows dynamical adjustments ε_x and ε_y, but does not allow adjustment of the homogeneous OBC, so $\varepsilon_o = 0$. Without OBC adjustment, the average velocity variance at the moorings, $(1/N)Tr[\mathbf{R}]$, is less than half the correct value. For this case, $\langle \hat{J}_{interior} \rangle$ substantially exceeds $N = 8$ for even the largest value of σ_{ee}^2.

In case 4, J_{OBC}, only the OBCs are allowed to vary, so the unforced momentum equations provide strong constraints on the inverse. As with case 3, $(1/N)Tr[\mathbf{R}]$ is much smaller than the correct value of 8.8 $(cm/s)^2$, and $\langle \hat{J}_{OBC} \rangle$ exceeds $N = 8$ for largest value of σ_{ee}^2.

Cases 5 through 8 in Table 1 show results for the same four null hypotheses, but for a 4-mooring array located in the interior of the regional model domain (as in Fig. 9). The data variances for cases 5–7 are virtually identical because OBC adjustments have a small effect on the interior of the regional model. For these three cases, $\langle \hat{J} \rangle$ is not significantly different from $\langle \hat{J}_{correct} \rangle$. In contrast, $\langle \hat{J}_{OBC} \rangle$ provides a sensitive test for Case 8. So, for interior moorings, J_{OBC} is the only case for which $(1/N)Tr[\mathbf{R}]$ differs substantially from the correct value.

In summary, results show that $\hat{J}$ provides a sensitive test statistic only when the velocity variance based on incorrect statistics differs substantially from the correct variance. For moorings near the boundary, we should easily detect erroneous neglect of OBC errors with J_{OBC}, or forcing errors with $J_{interior}$, since the expected data variance changes up to 50%. For moorings in the interior, only J_{OBC} has significantly different variance.

The neglect of model-error correlation with $J_{standard}$ has relatively little effect on average variance regardless of mooring location. Model-error correlation changes orientation and ellipticity of the velocity ellipses near the open boundary and these account for the observed artifacts in the flow. However, the posterior penalty functional $\hat{J}_{standard}$ is not sensitive to these effects. Fortunately, this seemingly pessimistic conclusion cannot be generalized because fractional errors in $\hat{J}$ become more significant as N increases. N is quite small in this study. In general, $\hat{J}$ provides a test statistic for the hypotheses used to define J and we advocate using it as such.

4. Discussion

The regional model in this study has two types of model error: forcing error from uncertain local winds and OBC error from uncertain remotely driven flows. The open-boundary flows are driven by remote winds which have infinite length scales. Therefore, model errors have only two degrees of freedom corresponding to the two independent components of wind stress. To account for this fact, the statistically consistent penalty functional, $J_{correct}$, includes the cross terms, C_{xo} and C_{yo}, that relate winds to OBCs. The more standard form, $J_{standard}$, neglects such off-diagonal terms. Their neglect allows OBC adjustments, ε_o, to be independent of interior adjustments, ε_x and ε_y, thereby increasing the degrees of freedom from two to six; there are two from ε_x and ε_y, two from associated values of ε_o on the western open boundary, and two more from ε_o on the east.

Idealized scenarios can be constructed where the cross terms don't matter. For example, if the data are perfect and the number of independent data, N, exceeds the degrees of freedom in the model, then the model can interpolate the data. However, this idealized case is not realistic. In reality, data errors make it undesirable to have the model and data in exact agreement. For example, with $J_{standard}$ the model can develop unrealistic artifacts

near the open boundary in order to minimize model-data misfit. With $J_{correct}$, the flows are reasonable because adjustments are statistically consistent with the "correct" model. On average, assimilation with $J_{standard}$ will yield better agreement with the data than $J_{correct}$. But, the smaller model-data misfit will come at the expense of unrealistic boundary artifacts.

Representers illustrate how the problems arise. The one-datum scenario is particularly informative because there is only one representer for each penalty functional, and optimal velocity fields are proportional to their respective representers. With $J_{correct}$, the representer exhibits realistic and complex flow structure characteristic of the correct response to wind. In contrast, the representer for $J_{standard}$ exhibits unrealistic artifacts along the open boundary. In both cases, spatially uniform winds generate complex flows because of the dynamics and the highly variable bathymetry. But realistic flows require accurate model-error statistics. The antenna-mode analysis shows that these conclusions generalize to arrays of moorings.

Sensitivity to mooring location develops because of the inaccurate model-error statistics in $J_{standard}$. Representers for the single-datum scenario are again informative. For a single datum located in the interior of the regional domain, representers for $J_{standard}$ and $J_{correct}$ have similar structure near the datum and substantial differences near the open boundary. Thus, with one interior datum, both models will produce reasonable velocity estimates in the interior. Near the open boundary, $J_{correct}$ still produces realistic flows but $J_{standard}$ produces artifacts.

The results for $J_{standard}$ degrade further for a single datum placed near the open boundary, in which case the flows are unrealistic even in the vicinity of the datum. In fact, $J_{standard}$ will produce unrealistic flows regardless of data accuracy because representer structure does not depend on the actual data value. As before, antenna modes generalize these results for arrays of moorings. In general for the data assimilative model, $J_{standard}$, which neglects cross terms, moorings should be placed in the interior, away from the open boundaries. There is no such restriction on mooring location with $J_{correct}$.

Two simplifications used in this study are worth discussion. First, the situation described here is extreme because the correct inverse has only two degrees of freedom. This simplification arises because the model errors are correlated with infinite length scales. Many things, including finite length scales, can contribute to make the errors uncorrelated and thereby increase the degrees of freedom. Second, the focus on neglected cross terms is really a special case of a more general sensitivity to inaccurate model-error statistics. The related sensitivity to mooring placement encountered here can also arise in very different scenarios.

The study by Bogden *et al.* (1996) provides an example. Prescribed winds for their regional inverse model were presumed correct so the only model adjustments were associated with OBC errors. With error-free interior forcing, the "correct" cross terms are identically zero. In addition, their inverse calculation had many degrees of freedom because the OBC errors had finite length scales comparable to grid resolution. They found

that velocities along the open boundary were highly sensitive to assumptions about OBC-error covariance. In contrast, they found that interior velocity estimates were relatively robust. This is fortunate because the OBC-error statistics were poorly understood. They concluded that only the large-scale interior flows were reliable. Their conclusions are consistent with this study because the moorings were located in the interior of their domain. In retrospect, the interior mooring placement was fortunate. Had the moorings been located near the open boundary, we infer from results presented here that velocities everywhere in the domain would have been sensitive to uncertain OBC-error statistics.

This example addresses one reviewer's suggestion that it may be better to invert in the large domain and expect results to be sensible only in the smaller area near the moorings. And if details of flow at the boundary are ignored or masked, then one might replace unknown model-error statistics in the large domain with ad hoc smoothness criteria that "regularize" the data assimilative model. We do not advocate such an approach. Increasing the size of the model domain increases computational expense and complicates accurate specification of "interior" forcing. Furthermore, unless the larger domain is a closed basin, one must still verify that the open boundaries are sufficiently far away to be unimportant. This is not a trivial task.

We prefer to use a large-domain model to generate model-error statistics for the small domain inverse, as was done here. The prescribed model-error statistics represent hypotheses that should be tested with data. The two penalty functionals in this study correspond to two different hypotheses. The statistics employed with $J_{standard}$ are clearly wrong, and $J_{correct}$ is "correct" because it is consistent with the large-domain model. But the large-domain model is itself a hypothesis. Thus, a regional inverse based on $J_{correct}$ can be used to test whether the large-domain model is statistically consistent with the data.

This approach can be extended to test models of varying dynamical complexity. For example, the neglected terms in the momentum balance of a simple model represent a type of model error. The statistics of such errors can be estimated from a more complex model which includes those terms. This approach has advantages because assimilation schemes based on the simplified dynamics can be much easier to implement and understand. In contrast, generalized inversion of complex stratified nonlinear models is neither simple nor straightforward. For example, Bennett (1992) discusses the problems that arise because OBCs of primitive equation models generally lead to ill-posed forward problems.

While issues such as these continue to present formidable challenges, the transition to more complex dynamics need not be so abrupt. Some oceanographers such as Thompson *et al.* (2000) and Lynch *et al.* (1998) employ a so-called incremental approach to data assimilation. They use a barotropic inverse model, similar to the one used here, to provide depth-averaged OBCs for a stratified model that has more complex dynamics. Their approach is incremental in its treatment of dynamical complexity and may benefit from the issues discussed here.

Acknowledgments. This study was supported by the National Science Foundation (Grant OCE-9796225) and the Office of Naval Research (N000140010226).

APPENDIX

The procedures for minimizing $J_{standard}$ and $J_{correct}$ are nearly identical. In both cases, standard application of the calculus of variations leads to an Euler-Lagrange system of equations. Results are presented below, and the reader is referred to Bennett (1992) for the theoretical development. The only difference introduced by model-error correlation occurs when relating the adjoint variables $\mu_x(\mathbf{x}, t)$, $\mu_y(\mathbf{x}, t)$, and $\mu_\eta(\mathbf{x}, t)$, to the model adjustments. This involves the model-error covariance matrix operating on the adjoint variables,

$$\begin{bmatrix} \varepsilon_x \\ \varepsilon_y \\ \varepsilon_o \end{bmatrix} = \begin{bmatrix} C_{xx}\bullet & 0 & C_{xo}\circ \\ 0 & C_{yy}\bullet & C_{yo}\circ \\ C_{xo}\circ & C_{yo}\circ & C_{oo}\circ \end{bmatrix} \begin{bmatrix} \mu_x \\ \mu_y \\ \mu_o \end{bmatrix}.$$

These relations also hold for $J_{standard}$, in which case $C_{xo} = C_{yo} = 0$.

The Euler-Lagrange system includes the basic model equations (1) and boundary conditions (2) and (3) along with a similar system for the adjoint variables,

$$\begin{aligned} \frac{\partial \mu_x}{\partial t} - f\mu_y &= -H\frac{\partial \mu_\eta}{\partial x} + \frac{r}{H}\mu_x - (\mathbf{d} - \mathbf{L}[u, v]) \cdot \mathbf{C}_{ee}^{-1}\boldsymbol{\Delta}_x, \\ \frac{\partial \mu_y}{\partial t} + f\mu_x &= -H\frac{\partial \mu_\eta}{\partial y} + \frac{r}{H}\mu_y - (\mathbf{d} - \mathbf{L}[u, v]) \cdot \mathbf{C}_{ee}^{-1}\boldsymbol{\Delta}_y, \\ \frac{\partial \mu_\eta}{\partial t} &= -g\left(\frac{\partial \mu_x}{\partial x} + \frac{\partial \mu_y}{\partial y}\right), \end{aligned} \tag{9}$$

with coastal boundary condition

$$\mu_x n_x + \mu_y n_y = 0,$$

and OBCs

$$\mu_x n_x + \mu_y n_y = -\sqrt{\frac{\overline{H}}{g}}\,\mu_\eta,$$

$$\mu_o = -H\mu_\eta.$$

The vectors $\boldsymbol{\Delta}_x$ and $\boldsymbol{\Delta}_y$ represent delta functions at the locations of u and v measurements, respectively, so $\mathbf{L}[u, v] = \boldsymbol{\Delta}_x \bullet u + \boldsymbol{\Delta}_y \bullet v$.

Adjoint variables satisfy the terminal condition

$$\mu_x(\mathbf{x}, T) = \mu_y(\mathbf{x}, t) = \mu_\eta(\mathbf{x}, T) = 0.$$

In this study, setting T equal to one day after the last measurement makes results independent of the terminal conditions.

The Euler-Lagrange system represents a coupled two-point boundary value problem because the solution variables u and v appear as forcing terms for μ_x and μ_y. The problem is efficiently solved with representers, and Bennett (1992) describes the methodology in detail.

For completeness, we present the representer solution. The notation is cumbersome but the manipulations are straightforward. Let $\mathbf{r}_x(\mathbf{x}, t)$, $\mathbf{r}_y(\mathbf{x}, t)$, and $\mathbf{r}_\eta(\mathbf{x}, t)$ denote column vectors of length N. If the scalar variable r_x denotes component i of $\mathbf{r}_x$, then

$$(r_x, r_y) = \langle d_i \mathbf{u}(\mathbf{x}, t) \rangle.$$

That is, the three functions from row i of the N-vectors $\mathbf{r}_x(\mathbf{x}, t)$, $\mathbf{r}_y(\mathbf{x}, t)$, and $\mathbf{r}_\eta(\mathbf{x}, t)$ comprise representer for d_i. The representer matrix $\mathbf{R}$ is obtained by evaluating $\mathbf{r}_x$ and $\mathbf{r}_y$ at the data locations with $\mathbf{L}$.

The representers are obtained by integrating the system

$$\frac{\partial \mathbf{r}_x}{\partial t} - f\mathbf{r}_y = -g\frac{\partial \mathbf{r}_\eta}{\partial x} - \frac{r}{H}\mathbf{r}_x + C_{xx} \bullet \boldsymbol{\alpha}_x + C_{xo} \circ \boldsymbol{\alpha}_o,$$

$$\frac{\partial \mathbf{r}_y}{\partial t} + f\mathbf{r}_x = -g\frac{\partial \mathbf{r}_\eta}{\partial y} - \frac{r}{H}\mathbf{r}_y + C_{yy} \bullet \boldsymbol{\alpha}_y + C_{yo} \circ \boldsymbol{\alpha}_o,$$

$$\frac{\partial \mathbf{r}_\eta}{\partial t} = -\frac{\partial (H\mathbf{r}_x)}{\partial x} - \frac{\partial (H\mathbf{r}_y)}{\partial y},$$

with coastal boundary condition,

$$\mathbf{r}_x n_x + \mathbf{r}_x n_x = 0,$$

and OBC,

$$\mathbf{r}_x n_x + \mathbf{r}_y n_y = \sqrt{\frac{g}{H}}\mathbf{r}_\eta - HC_{oo} \circ \boldsymbol{\alpha}_\eta - C_{ox} \bullet \boldsymbol{\alpha}_x - C_{oy} \bullet \boldsymbol{\alpha}_y.$$

The adjoint representer vectors, denoted by $\boldsymbol{\alpha}$, are obtained by integrating the uncoupled system,

$$\frac{\partial \boldsymbol{\alpha}_x}{\partial t} - f\boldsymbol{\alpha}_y = -H\frac{\partial \boldsymbol{\alpha}_\eta}{\partial x} - \frac{r}{H}\boldsymbol{\alpha}_x - \boldsymbol{\Delta}_x,$$

$$\frac{\partial \boldsymbol{\alpha}_y}{\partial t} + f\boldsymbol{\alpha}_x = -H\frac{\partial \boldsymbol{\alpha}_\eta}{\partial y} - \frac{r}{H}\boldsymbol{\alpha}_y - \boldsymbol{\Delta}_y,$$

$$\frac{\partial \boldsymbol{\alpha}_\eta}{\partial t} = -g\left(\frac{\partial \boldsymbol{\alpha}_x}{\partial x} + \frac{\partial \boldsymbol{\alpha}_y}{\partial y}\right).$$

with coastal boundary condition

$$\alpha_x n_x + \alpha_y n_y = 0,$$

open boundary condition

$$\alpha_x n_x + \alpha_y n_y = -\sqrt{\frac{\overline{H}}{g}}\,\alpha_\eta,$$

and terminal conditions

$$\alpha_\eta(\mathbf{x}, T) = \alpha_x(\mathbf{x}, T) = \alpha_y(\mathbf{x}, T) = 0.$$

This system is integrated backward in time. Note that the forcing terms are delta functions, so each adjoint representer is a Greens function for the wind-driven model.

The optimal solution is obtained by first defining the N-vector $\mathbf{b}$ of representer coefficients

$$\mathbf{b} = (\mathbf{d} - \mathbf{L}[u, v]) \cdot \mathbf{C}_{ee}^{-1} \tag{10}$$

so that the adjoint variables are $\mu_x = \mathbf{b} \cdot \alpha_x$, $\mu_y = \mathbf{b} \cdot \alpha_y$, $\mu_\eta = \mathbf{b} \cdot \alpha_\eta$. The optimal flow fields are $\hat{u} = \mathbf{b} \cdot \mathbf{r}_x$ and $\hat{v} = \mathbf{b} \cdot \mathbf{r}_y$. It can be shown that

$$\mathbf{b} = (\mathbf{R} + \mathbf{C}_{ee})^{-1}\mathbf{d},$$

so that, once $\mathbf{R}$ is known, $\mathbf{b}$ can be computed with the data and substituted into (9) using (10). This uncouples the system so that the adjoint equations can be integrated backward in time to obtain the optimal model adjustments. The optimal solution is then obtained by integrating the model equations. There is no iteration.

REFERENCES

Bennett, A. F. 1992. Inverse Methods in Physical Oceanography, Cambridge University Press, 346 pp.

Bogden, P. S., P. Malanotte-Rizzoli and R. Signell. 1996. Open-ocean boundary conditions from interior data: Local and remote forcing of Massachusetts Bay. J. Geophys. Res., *101,* 6487–6500.

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

Egbert, G. D. and A. F. Bennett. 1996. Data assimilation for ocean tides, *in* Modern Approaches to Data Assimilation in Ocean Modeling, P. Malanotte-Rizzoli, ed., Elsevier Sciences, 468 pp.

Lynch, D. R., C. E. Naimie and C. G. Hannah. 1998. Hindcasting the Georges Bank circulation, Part I: Detiding. Cont. Shelf Res., *18,* 607–639.

Thompson, K. R., M. Dowd, Y. Lu and B. Smith. 2000. Oceanographic data assimilation and regression analysis. Econometrics, *11,* 183–196.

Xue, H., F. Chai and N. R. Pettigrew. 2000. A model study of the seasonal circulation in the Gulf of Maine. J. Phys. Oceanogr., *30,* 1111–1135.

Received: *27 July, 2000;* revised: *2 October, 2001.*