Tellus (2004), 56A, 141–153
Printed in UK. All rights reserved

TELLUS

Initial-boundary compatibility for inverse regional models

By JULIA C. MUCCINO[1]* and ANDREW F. BENNETT[2], [1]*Department of Civil and Environmental Engineering, PO Box 875306, Ira A. Fulton School of Engineering, Arizona State University, Tempe, AZ 85287-5306, USA;* [2]*College of Oceanic and Atmospheric Sciences, 104 Oceanography Administration Building, Corvallis, OR 97331-5503, USA*

(Manuscript received 5 August 2003; in final form 23 October 2003)

ABSTRACT

In this paper we investigate the importance of compatibility conditions in inverse regional models. In forward models, the initial condition and boundary condition must be continuous at the space–time corner to ensure that the solution is continuous at the space–time corner and along the characteristic emanating from it. Furthermore, the initial condition, boundary condition and forcing must satisfy the partial differential equation at the space–time corner to ensure that the solution is continuously differentiable at the space–time corner and along the characteristic emanating from it.

Assimilation of data into a model requires a null hypothesis regarding residuals in the dynamics (including, for example, forcing, boundary conditions and initial conditions) and data. In most published work, these errors are assumed to be uncorrelated, although Bogden does investigate the importance of correlation between boundary conditions and forcing, and shows that by generalizing the penalty functional to include this cross-correlation the open boundaries behave more like an interface with a true ocean. In this paper, the more general importance of cross-correlations and their impact on the smoothness of the optimal solution are discussed. It is shown that in the absence of physical diffusion, cross-correlations are necessary to obtain a smooth solution; if errors are assumed to be uncorrelated, discontinuities propagate along the characteristic emanating from the space–time corner, even when the initial condition, boundary condition and forcing are compatible, as described above.

Although diffusion damps such discontinuities, it does not reproduce the inherently smooth solution achieved when cross-correlation of residuals are accounted for. Rather, the discontinuity is smoothed into a spurious front.

1. Introduction

Regional models are the critical tools for short-range, high-resolution numerical weather prediction (e.g. Hodur, 1997) and for ocean forecasting (e.g. Burnett et al., 2002). These immensely complex codes combine pseudo-laminar fluid dynamics variously with empirical parametrization schemes for unresolved internal processes and for imposed surface fluxes, significant orography, complex bathymetry including basin shape, radiative transfer and a host of other important features. The complexity will only become greater, as advanced data assimilation systems become incorporated into the models. On the other hand, it will always remain essential that the most elemental mathematical considerations not be overlooked. Tompkins, writing in Bernstein (1950), stresses the need to strive to retain the best balance between practical imperatives on the one hand, and on the other seeming abstraction such as existence theorems for the partial differential equations (PDEs) that are at the heart of environmental models. One especially simple consideration, although spelled out in intermediate-level math texts, seems never to rate discussion. The requirement is for compatibility between initial and boundary data, where they overlap in space–time. If the requirement is ignored, non-differentiable or even discontinuous behavior will emanate from the space–time 'corner', with deleterious effect upon the solution in the model region. Satisfaction of the requirement not only for forecasts, but also for fixed-interval assimilation of data, is the focus of this investigation.

As an instructive example, consider the simple, non-dimensional, non-homogeneous, linear convection problem over the domain $0 \le x \le L$, $0 \le t \le T$:

$$\frac{\partial u_F}{\partial t} + \frac{\partial u_F}{\partial x} = F(x,t), \quad u_F(x,0) = I(x), \quad u_F(0,t) = B(t) \tag{1}$$

where $F(x, t)$, $I(x)$ and $B(t)$ are the dimensionless forcing, initial condition and boundary condition, respectively, and x

*Corresponding author.
e-mail: jmuccino@asu.edu

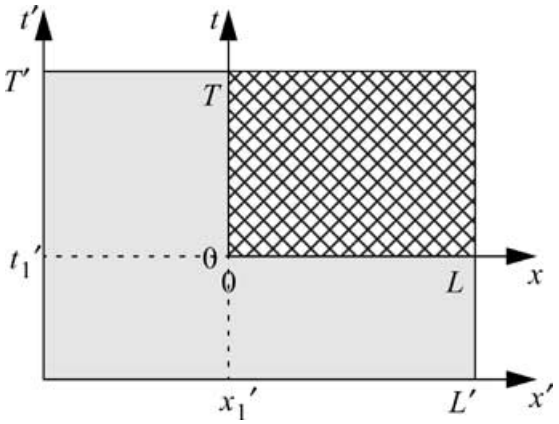

Fig. 1. Domain of interest (cross-hatched) nested within a larger domain (shaded). Space–time corner (where $(x, t) = (0, 0)$) in the smaller domain is an interior point in the larger domain.

and t are dimensionless space and time, respectively. For the solution to be continuous along the characteristic emanating from the space–time corner, that is, where $(x, t) = (0, 0)$ (hereafter, simply 'corner'), the initial condition and boundary condition must be be consistent there:

$$I(0) = B(0). \tag{2}$$

Furthermore, for the solution to be continuously differentiable across the characteristic emanating from the corner (hereafter, 'corner characteristic'), the initial condition, boundary condition and forcing must satisfy the PDE at the corner:

$$\left.\frac{\partial B}{\partial t}\right|_{t=0} + \left.\frac{\partial I}{\partial x}\right|_{x=0} = F(0, 0). \tag{3}$$

Both requirements (2) and (3) may be achieved by appropriate specification of $B(t)$ and $I(x)$. While mathematically necessary, these requirements are also physically reasonable within the context of a regional model nested within a larger model, as shown in Fig. 1. Consider, for example, a typical regional model, with forcing $F(x, t)$, initial condition $I(x)$, boundary condition $B(t)$ and solution $u_F(x, t)$, nested within a larger unbounded or periodic model with forcing $F'(x', t')$, initial condition $I'(x')$ and solution $u'_F(x', t')$ that satisfies the same PDE as the smaller model. Both $I(x)$ and $B(t)$ serve to link the smaller model with the larger model. If $I(x)$ takes on the corresponding values from the larger domain (i.e. $I(x) = u'_F(x', t'_1)$, $0 \le x \le L$, $x'_1 \le L'$) and $B(t)$ takes on the corresponding values from the larger domain (i.e. $B(t) = u'_F(x'_1, t')$, $0 \le t \le T$, $t'_1 \le t' \le T'$), then eq. (2) will be satisfied because $u'_F(x'_1, t'_1)$ is uniquely and smoothly defined (assuming that F' and I' are smoothly defined). Furthermore, because the PDE is satisfied at all interior points of the larger model, including the space/time boundaries of the smaller domain, the PDE will be satisfied at the corner of the smaller domain (0, 0) as long as $F(0, 0) = F'(0, 0)$ and $I(x)$ and $B(t)$ are specified as stipulated above.

Suppose that, in addition to eq. (1), there are also M pointwise measurements of the solution available:

$$u_F(x_m, t_m) = d_m, \quad m = 1, M. \tag{4}$$

Equation (4) overdetermines the dynamics (1) and thus the problem becomes ill-posed; the ill-posedness may be resolved with a generalized solution that satisfies the dynamics and the data in a least-squares sense. That is, residuals $f(x, t)$, $i(x)$ and $b(t)$ are admitted in the forcing, initial condition and boundary condition, respectively, and eq. (1) becomes

$$\begin{aligned} &\frac{\partial u}{\partial t} + \frac{\partial u}{\partial x} = F(x, t) + f(x, t), \quad u(x, 0) = I(x) + i(x), \\ &u(0, t) \quad = B(t) + b(t). \end{aligned} \tag{5a}$$

Likewise, a residual ε_m is admitted in the measurements:

$$u(x_m, t_m) = d_m, +\varepsilon_m, \quad m = 1, M. \tag{5b}$$

A probabilistic hypothesis is made about the residuals, such as H_0:

$$\langle f(x, t)\rangle = \langle i(x)\rangle = \langle b(t)\rangle = 0, \quad \langle \mathbf{e}\rangle = 0 \tag{6a}$$

$$\begin{aligned} &\langle f(x, t) f(y, s)\rangle = C_{ff}(x, t, y, s), \quad \langle i(x) i(y)\rangle = C_{ii}(x, y), \\ &\langle b(t) b(s)\rangle = C_{bb}(t, s), \quad \langle \mathbf{e}\mathbf{e}^{\mathrm{T}}\rangle = \mathbf{C_e} \end{aligned} \tag{6b}$$

$$\begin{aligned} &\langle f(x, t) i(y)\rangle = \langle f(x, t) b(s)\rangle = \langle i(x) b(s)\rangle = 0, \\ &\langle f(x, t)\mathbf{e}\rangle \quad = \langle i(x)\mathbf{e}\rangle = \langle b(t)\mathbf{e}\rangle = \mathbf{0} \end{aligned} \tag{6c}$$

where $\langle\ \rangle$ indicates the expected value and $\mathbf{e}$ is a vector of ε_m. Note that eq. (6c) indicates that all cross-correlations between $f(x, t)$, $i(x)$, $b(t)$, and $\mathbf{e}$ are neglected. These first and second moments completely characterize a Gaussian random process. The associated estimator of maximum likelihood is the quadratic form or 'penalty functional':

$$J = f \bullet C_{ff}^{-1} \bullet f + i \circ C_{ff}^{-1} \circ i + b * C_{bb}^{-1} * b + \mathbf{e}\mathbf{C_e}^{-1}\mathbf{e}^{\mathrm{T}} \tag{7}$$

where $q \bullet p = \int \mathrm{d}y \int \mathrm{d}s q(x, t, y, s) p(y, s)$ indicates integration over all space and time, $q \circ p = \int \mathrm{d}y q(x, y) p(y)$ indicates integration over all space, and $q * p = \int \mathrm{d}s q(t, s) p(s)$ indicates integration over all time. The non-diagonal, symmetric, positive-definite weight functions C_{ff}^{-1}, C_{ii}^{-1}, and C_{bb}^{-1} are the functional inverses, with respect to $\bullet$, $\circ$ and $*$, of the covariances in eq. (6b) and the symmetric, positive-definite weight matrix $\mathbf{C_e}^{-1}$ is the matrix inverse of $\mathbf{C_e}$. The Euler–Lagrange (EL) equations for the extrema of eq. (7) are (see Bennett, 2002):

$$-\frac{\partial \lambda}{\partial t} - \frac{\partial \lambda}{\partial x} = \delta^{\mathrm{T}}\mathbf{C_e}^{-1}(\mathbf{d} - \hat{\mathbf{u}}), \quad \lambda(x, T) = 0, \quad \lambda(0, L) = 0 \tag{8}$$

$$\begin{aligned} &\frac{\partial \hat{u}}{\partial t} + \frac{\partial \hat{u}}{\partial x} = F(x, t) + \hat{f}(x, t), \quad \hat{u}(x, 0) = I(x) + \hat{i}(x), \\ &\hat{u}(0, t) = B(t) + \hat{b}(t) \end{aligned} \tag{9}$$

where the impulse vector δ has, as its mth component, the scalar impulse $\delta(x-x_m)\,\delta(t-t_m)$, $\mathbf{d}$ is the vector of the data d_m and $\hat{\mathbf{u}}$ is the vector of point measurements of the optimal solution, $\hat{u}(x,t)$ at the space/time locations of the data. The optimal residuals in eq. (9) are

$$\begin{aligned}&\hat{f}(x,t) = C_{ff} \bullet \lambda(y,s), \quad \hat{i}(x) = C_{ii} \circ \lambda(y,0),\\ &\hat{b}(t) = C_{bb} * \lambda(0,s)\end{aligned} \tag{10}$$

where the integration is over the (y, s) arguments. The coupling between eqs. (8) and (9) is resolved by the representer method (Bennett, 2002).

One might expect that $\hat{u}(x,t)$ should be smooth throughout the spatial-temporal domain if $I(x)$, $B(t)$ and $F(x,t)$ satisfy eqs. (2) and (3). However, by comparing eq. (9) with eq. (1), it is apparent that to ensure $\hat{u}(x,t)$ is everywhere smooth and differentiable, $\hat{f}(x,t)$, $\hat{i}(x)$ and $\hat{b}(t)$ must satisfy requirements analogous to eqs. (2) and (3)

$$\hat{i}(0) = \hat{b}(0) \tag{11}$$

$$\left.\frac{\partial \hat{b}}{\partial t}\right|_{t=0} + \left.\frac{\partial \hat{i}}{\partial x}\right|_{x=0} = \hat{f}(0,0), \tag{12}$$

where eq. (11) ensures continuity of the residuals at the space–time corner and eq. (12) ensures continuous differentiability of the residuals at the space–time corner. However, the data assimilation problem described thus far does not explicitly enforce these constraints on the residuals. As a result, $\hat{i}(x)$, $\hat{b}(t)$ and $\hat{f}(x,t)$ will not, in general, satisfy eqs. (11) and (12) even when $I(x)$, $B(t)$ and $F(x,t)$ satisfy eqs. (2) and (3). The optimal solution will therefore, in general, be discontinuous at the corner and along the corner characteristic.

2. Cross-correlations and implications at the space–time corner

The obvious question then is how should eqs. (11) and (12) be enforced within the confines of a data assimilation algorithm? Consider a more general hypothesis, H_{0g} that includes cross-correlations between $i(x)$, $b(t)$ and $f(x,t)$:

$$\langle f(x,t)\rangle = \langle i(x)\rangle = \langle b(t)\rangle = 0, \quad \langle \mathbf{e}\rangle = 0 \tag{13a}$$

$$\begin{aligned}&\langle f(x,t)f(y,s)\rangle = C_{ff}(x,t,y,s), \quad \langle i(x)i(y)\rangle = C_{ii}(x,y),\\ &\langle b(t)b(s)\rangle \qquad = C_{bb}(t,s), \quad \langle \mathbf{e}\mathbf{e}^{\mathrm{T}}\rangle = \mathbf{C_e}\end{aligned} \tag{13b}$$

$$\begin{aligned}&\langle f(x,t)i(y)\rangle = C_{fi}(x,t,y), \quad \langle f(x,t)b(s)\rangle = C_{fb}(x,t,s),\\ &\langle i(x)f(y,s)\rangle = C_{if}(x,y,s), \quad \langle i(x)b(s)\rangle = C_{ib}(x,s),\\ &\langle b(t)f(y,s)\rangle = C_{bf}(t,y,s), \langle b(t)i(y)\rangle = C_{bi}(t,y)\\ &\langle f(x,t)\mathbf{e}\rangle = \langle i(x)\mathbf{e}\rangle = \langle b(t)\mathbf{e}\rangle = \mathbf{0}.\end{aligned} \tag{13c}$$

Note that this hypothesis is a generalization of Bogden (2001) that includes cross-covariances only between f and b. A penalty functional associated with this more general null hypothesis is

$$\begin{aligned}I = {} & f \bullet C_{ff}^{-1} \bullet f + f \bullet C_{fi}^{-1} \circ i + f \bullet C_{fb}^{-1} * b\\ & + i \circ C_{if}^{-1} \bullet f + i \circ C_{ii}^{-1} \circ i + i \circ C_{ib}^{-1} * b\\ & + b * C_{bf}^{-1} \bullet f + b * C_{bi}^{-1} \circ i + b * C_{bb}^{-1} * b + \mathbf{e}\mathbf{C_e}^{-1}\mathbf{e}^{\mathrm{T}}.\end{aligned} \tag{14}$$

The EL equations associated with eq. (14) are just as in eqs. (8) and (9), but the residuals on the right-hand side of eq. (9) now include convolutions of cross-covariances:

$$\hat{f} = C_{ff} \bullet \lambda(y,s) + C_{fi} \circ \lambda(y,0) + C_{fb} * \lambda(0,s) \tag{15a}$$

$$\hat{i} = C_{if} \bullet \lambda(y,s) + C_{ii} \circ \lambda(y,0) + C_{ib} * \lambda(0,s) \tag{15b}$$

$$\hat{b} = C_{bf} \bullet \lambda(y,s) + C_{bi} \circ \lambda(y,0) + C_{bb} * \lambda(0,s), \tag{15c}$$

where again, integration is over the (y, s) arguments and where the covariances are defined[1] to satisfy:

$$\begin{aligned}&\begin{bmatrix} C_{ff}\bullet & C_{fi}\circ & C_{fb}* \\ C_{if}\bullet & C_{ii}\circ & C_{ib}* \\ C_{bf}\bullet & C_{bi}\circ & C_{bb}* \end{bmatrix} \begin{bmatrix} C_{ff}^{-1} & C_{fi}^{-1} & C_{fb}^{-1} \\ C_{if}^{-1} & C_{ii}^{-1} & C_{ib}^{-1} \\ C_{bf}^{-1} & C_{bi}^{-1} & C_{bb}^{-1} \end{bmatrix}\\ &\quad = \begin{bmatrix} \delta(x-y)\delta(t-s) & 0 & 0 \\ 0 & \delta(x-y) & 0 \\ 0 & 0 & \delta(t-s) \end{bmatrix}.\end{aligned} \tag{16}$$

Substituting the generalized residual expressions (15a)–(15c) into the first corner condition (11) yields the following constraints on the covariances:

$$C_{bf}(0,y,s) - C_{if}(0,y,s) = 0 \tag{17a}$$

$$C_{bi}(0,y) - C_{ii}(0,y) = 0 \tag{17b}$$

$$C_{bb}(0,s) - C_{ib}(0,s) = 0. \tag{17c}$$

These are necessary and sufficient constraints for requirement (11) to be met for arbitrary λ. Next, substituting these generalized residuals expressions into the second corner condition (12) and grouping by types of integration yields

$$\begin{aligned}&\left\{\frac{\partial C_{bf}}{\partial t} + \frac{\partial C_{if}}{\partial x} - C_{ff}\right\} \bullet \lambda(\,,\,) + \left\{\frac{\partial C_{bi}}{\partial t} + \frac{\partial C_{ii}}{\partial x} - C_{fi}\right\} \circ \lambda(\,,0)\\ &\quad + \left\{\frac{\partial C_{bb}}{\partial t} + \frac{\partial C_{ib}}{\partial x} - C_{fb}\right\} * \lambda(0,\,) = 0\end{aligned}$$

where differentiation is over the first argument and all terms are evaluated at the corner. Thus, eq. (12) is satisfied for all λ if and only if

[1] The representer solution algorithm never requires the inverses of these covariances, only the covariances themselves.

$$\frac{\partial C_{bf}}{\partial t} + \frac{\partial C_{if}}{\partial x} - C_{ff} = 0 \tag{18a}$$

$$\frac{\partial C_{bi}}{\partial t} + \frac{\partial C_{ii}}{\partial x} - C_{fi} = 0 \tag{18b}$$

$$\frac{\partial C_{bb}}{\partial t} + \frac{\partial C_{ib}}{\partial x} - C_{fb} = 0, \tag{18c}$$

where these constraints hold at the space–time corner. Thus, eqs. (11) and (12) can be enforced by augmenting H_{0g} with conditions (17) and (18). Reference to H_{0g} hereafter includes these augmenting constraints on the covariances and cross-covariances.

It is easy to show that these mathematical conditions have physical as well as mathematical relevance. For example, multiply the second corner condition (12), which should apply also to any f, i, and b, by $f(\,,\,)$

$$\left.\frac{\partial b(t) f(\,,\,)}{\partial t}\right|_{t=0} + \left.\frac{\partial i(x) f(\,,\,)}{\partial x}\right|_{x=0} = f(0,0) f(\,,\,) \tag{19}$$

and take the expectation

$$\left.\frac{\partial \langle b f(\,,\,)\rangle}{\partial t}\right|_{t=0} + \left.\frac{\partial \langle i f(\,,\,)\rangle}{\partial x}\right|_{x=0} = \langle f(\,,\,) f(0,0)\rangle. \tag{20}$$

Thus, we expect eq. (20) to be true in a natural system, while eq. (18a) is required for mathematical consistency. Because these are the same, the mathematical requirement is physically reasonable. Similar manipulations show that eqs. (18b) and (18c) are physically reasonable as well.

3. Determining cross-covariances consistent with mathematical requirements

Statistical simulation is used to construct the nine covariances in eq. (15) such that they satisfy the constraints in eqs. (17) and (18) in the limit of a large number of samples.

3.1. Realistic simplification of the covariance constraints

The construction of the error covariances is simplified by assuming that there is no correlation between the forcing and the initial condition; this is justified by recognizing that the initial state should not be influenced by the forcing after $t = 0$. Although the temporally correlated forcing indicates correlation between the forcing at very early time and forcing before $t = 0$, which in turn would influence the initial condition, this effect is neglected here. (The impact of this simplification is examined in Section 3.3.) It is therefore assumed that $C_{if}(x, y, s) = C_{fi}(x, t, y) = 0$, and the covariance constraints in eqs. (17) and (18) become

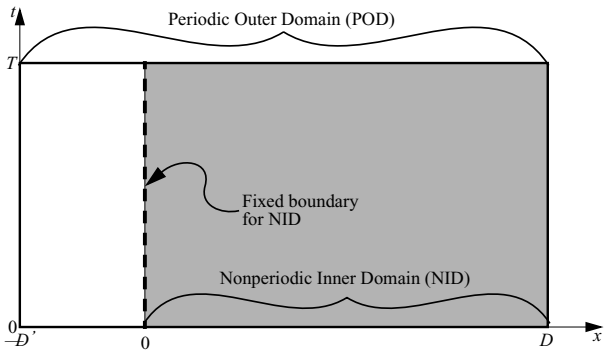

Fig. 2. Relationship between the POD and NID. The NID is nested within the POD and the boundary condition for the NID is determined by the solution at $x = 0$ in the POD.

$$C_{bf}(0, y, s) = 0 \tag{21a}$$

$$C_{bi}(0, y) - C_{ii}(0, y) = 0 \tag{21b}$$

$$C_{bb}(0, s) - C_{ib}(0, s) = 0 \tag{21c}$$

$$\frac{\partial C_{bf}}{\partial t} - C_{ff} = 0 \tag{22a}$$

$$\frac{\partial C_{bi}}{\partial t} + \frac{\partial C_{ii}}{\partial x} = 0 \tag{22b}$$

$$\frac{\partial C_{bb}}{\partial t} + \frac{\partial C_{ib}}{\partial x} - C_{fb} = 0 \tag{22c}$$

where eq. (22) holds at the corner.

3.2. Statistical simulation of the covariances

Statistical simulation of the necessary covariances begins with the solution of weak dynamics on a so-called 'periodic outer domain' (POD, as shown in Fig. 2) with no forcing $F(x, t)$ or initial condition $I(x)$:

$$\begin{aligned} &\frac{\partial u_{\mathrm{POD}}}{\partial t} + \frac{\partial u_{\mathrm{POD}}}{\partial x} = f(x, t), \\ &u_{\mathrm{POD}}(x, 0) = i(x), \quad u_{\mathrm{POD}}(-D', t) = u_{\mathrm{POD}}(D, t) \\ &\quad \text{for} \quad -D' \le x \le D \quad \text{and} \quad 0 \le t \le T. \end{aligned} \tag{23}$$

Here, samples of $f(x, t)$ and $i(x)$ are periodic and satisfy a periodic null hypothesis, H_{0P}:

$$\langle f(x, t)\rangle = \langle i(x)\rangle = \langle f(x, t) i(x)\rangle = 0, \tag{24a}$$

$$\begin{aligned} &\langle f(x, t) f(y, s)\rangle = C_{ff}(x, t, y, s), \\ &\langle i(x) i(y)\rangle = C_{ii}(x, y), \quad \langle \mathbf{e}\mathbf{e}^{\mathrm{T}}\rangle = \mathbf{C}_{\mathbf{e}} \end{aligned} \tag{24b}$$

$$\langle f(x, t)\mathbf{e}\rangle = \langle i(x)\mathbf{e}\rangle = \langle b(t)\mathbf{e}\rangle = 0. \tag{24c}$$

Note that H_{0P} follows directly from H_{0g} (that is, all cross terms vanish) because:

(i) the spatial domain is periodic, and thus there is no $b(t)$;

(ii) $C_{if}(x, y, s) = C_{fi}(x, t, y) = 0$ by the simplification described in Section 3.1.

A forcing residual is generated such that its covariance is

$$C_{ff}(x, t, y, s) = C_{0f} \exp\left(-\frac{|x-y|^2}{L_f^2}\right) \exp\left(-\frac{|t-s|}{\tau}\right) \quad (25)$$

where C_{0f}, L_f, and τ are the variance, decorrelation length-scale and decorrelation time-scale of the forcing residual, respectively. An initial condition residual is generated such that its covariance is

$$C_{ii}(x, y) = C_{0i} \exp\left(-\frac{|x-y|^2}{L_i}\right) \quad (26)$$

where C_{0i} and L_i are the variance and decorrelation length-scale of the initial condition residual, respectively.

Samples of $f(x, t)$ and $i(x)$ that are consistent with H_{0P} and with the covariance forms in eqs. (25) and (26) are synthesized by convolving white noise with the square roots of the covariances, as in Muccino and Bennett (2002). The dynamics in eq. (23) are then solved on the POD to yield a sample of $u_{POD}(x, t)$ for the residual pair $f(x, t)$ and $i(x)$.

A non-periodic inner domain (NID), the actual domain of interest, is nested within the POD, as shown in Fig. 2; the NID has the same temporal extent as the POD, but the spatial extent of the NID is $0\ x \leq D$. The upstream boundary condition required for the NID by the convection equation is obtained from the value of $u_{POD}(x, t)$ at $x = 0$; thus $b(t) = u_{POD}(0, t)$. In this way, a trio of compatible residuals $f(x, t)$, $i(x)$, and $b(t)$ is obtained for the NID that satisfies H_{0P} and therefore H_{0g}.

This procedure is repeated N times to generate N residual trios. Then, the seven required covariances are generated according to the definitions in eqs. (13b) and (13c).

3.3. Numerical details

The convection equation is discretized using first-order, explicit upwinding

$$u_{Fi}^{k+1} = u_{Fi}^{k} - C\left(u_{Fi}^{k} - u_{Fi-1}^{k}\right) + \Delta t\, F_i^k \quad (27)$$

where $C = \Delta x/\Delta t$. We let $C = 1$ ($\Delta x = \Delta t = 0.1$); this scheme has perfect amplitude and phase propagation for free solutions (i.e. $F_i^k = 0$) (see, for example, Celia and Gray, 1992).

As a demonstration of the impact of incompatibilities at the corner on the solution, consider a case where

$$F(x, t) = 0.1 \quad (28a)$$

$$I(x) = \exp\left(-\left|\frac{x}{2}\right|^2\right) \quad (28b)$$

$$B(t) = -\sin\left(\frac{\pi t}{2.5}\right) \quad (28c)$$

for $0 < x < 6$ and $0 < t < 6$, where $I(0) \neq B(0)$ and $\partial B/\partial t + \partial I/\partial x \neq F$ at the corner. The solution (with no data), shown in Fig. 3, displays a discontinuity propagating along the corner characteristic, seen as a vertical 'wall of discontinuity' along the line $x = t$. Note that the slope is also discontinuous along that characteristic.

With appropriate care, the modeler can specify F, B and I to be compatible at the corner, and thus avoid the propagation of discontinuities in u_F like those demonstrated in Fig. 3. Here, we are interested only in the behavior of the residuals $\hat{f}, \hat{i}$ and $\hat{b}$; these are not explicitly under the modeler's control. The model here is linear, so we shall, without loss of generality, restrict the remainder of the discussion to the special case where $F(x, t) = I(x) = B(t) = 0$, and thus $u_F\ (x, t) = 0$. Therefore, the optimal solution, $\hat{u}(x, t)$, is the solution to eq. (5a) arising when forced only with the optimal residuals $\hat{f}, \hat{i}$ and $\hat{b}$.

The following forcing and initial condition covariance parameters were used on the POD to generate the full set of covariances by statistical simulation. The standard error of the forcing residual is assumed to be 10% of the magnitude of the convection term, such that $C_{0f} = 0.063$. The decorrelation length- and time-scales for the forcing are assumed be one wavelength such that $L_f = \tau = 2.5$. Likewise, the standard error of the initial condition is assumed to be 20% of the amplitude of the initial condition, such that $C_{0i} = 0.04$. The decorrelation length-scale for the initial condition is also assumed to be one wavelength, such that $L_i = 2.5$. The data are uncorrelated with a variance of $V_\varepsilon = 0.04$. The spatial extent of the POD is $-2.5 \leq x \leq 10$, so the extent of the NID is $0 \leq x \leq 10$. The temporal extent of both the POD and NID is $0 \leq t \leq 6$.

Ten thousand samples ($N = 10\,000$) of residual trios (f, i and b) were generated to construct estimates of the covariances. The statistically simulated covariances satisfy the general covariance constraints (17) and (18) to within roundoff error, when the derivatives of the covariances are differenced using the same scheme used to generate the synthetic residuals.

However, as stated earlier in Section 3.1, the covariance constraints have been simplified by neglecting cross-correlation between the initial condition and the forcing, although in a strict sense there is some correlation between the initial condition and the forcing at very early time owing to the temporal correlation of the forcing. Thus, it is appropriate to examine the error introduced by this simplification. Consider, for example, general constraint (17a)

$$C_{bf}(0, y, s) - C_{if}(0, y, s) = 0$$

and its simplified counterpart (21a)

$$C_{bf}(0, y, s) = 0.$$

In terms of expected values, the former is written:

$$\langle b(0) f(y, s)\rangle - \langle i(0) f(y, s)\rangle = 0. \quad (29)$$

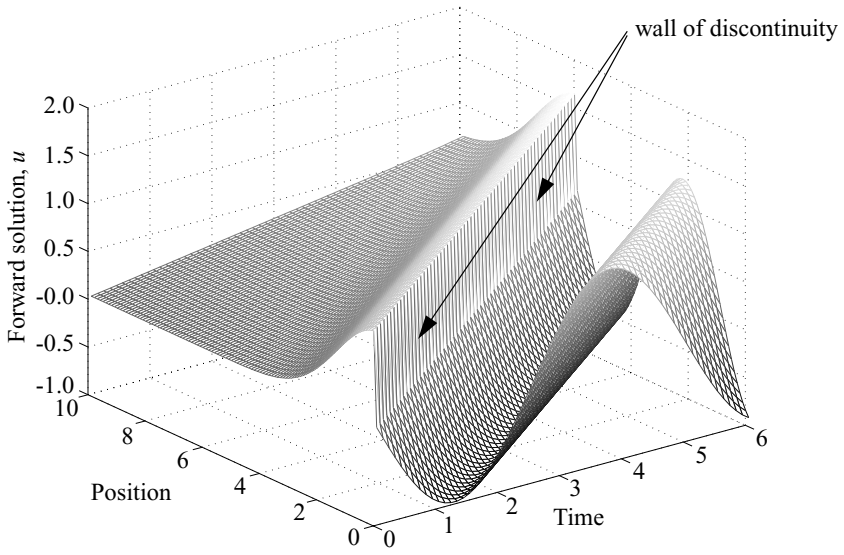

Fig. 3. Example of the impact of incompatibilities at the corner on the forward solution, u. Note the 'wall' of discontinuity along the corner characteristic. Not only is the solution discontinuous, but there is also a discontinuity in the slope along that characteristic.

Although $\langle i(0)\, f(y, s)\rangle = 0$ in the limit of large N, it does not equal zero for small N. As an extreme example, consider $N = 1$ where $\langle i(0)\, f(y, s)\rangle = i(0) f(y, s)$. This quantity equals zero only if either the initial condition vanishes at the corner or the forcing vanishes for all space and time (or some combination), which is clearly not the case here. These observations are borne out by computation. The mean absolute error[2] introduced by neglecting the correlation between the forcing and the initial condition in eq. (21a), as a function of the number of samples, is indicated by the solid line in Fig. 4a. Indeed, the error decreases with increasing N. Similar arguments may be made regarding the differentiability conditions (22a) and (22b) as shown by the solid and dash-dotted lines, respectively, in Fig. 4b. [Note that this simplification does not impact covariance conditions (21b), (21c) and (22c), so these are satisfied to within roundoff error regardless of the number of samples; see the dashed and dash-dotted lines in Fig. 4a and the dashed line in Fig. 4b.] Thus, even when correlation between the forcing and the initial condition is neglected, the general constraints (17) and (18) are satisfied for large N.

Consider now the errors introduced into the general covariance constraints when *all* cross-correlations are neglected, as in H_0 as described by eq. (6). In this case, the covariance constraints become

$$0 = 0 \tag{30a}$$

$$C_{ii}(0, y) = 0 \tag{30b}$$

$$C_{bb}(0, s) = 0 \tag{30c}$$

[2]Mean absolute error is defined as the mean absolute value of the error in the constraint. Thus, in these discrete calculations, the absolute error is summed over the free arguments, and divided by the number of free arguments.

$$C_{ff} = 0 \tag{31a}$$

$$\frac{\partial C_{ii}}{\partial x} = 0 \tag{31b}$$

$$\frac{\partial C_{bb}}{\partial t} = 0. \tag{31c}$$

In Fig. 4c, the dash-dotted line shows the (absolute) error in eq. (30b) and the dashed line shows the error in eq. (30c). Neither shows a tendency to decrease with increasing sample size. [Note that there is no line in Fig. 4c for eq. (30a) because all terms in that equation are cross terms.] Similar observations may be made regarding Fig. 4d, which illustrates the errors in eq. (31a) by the solid line, eq. (31b) by the dash-dotted line, and eq. (31c) by the dashed line. Thus, increasing the sample size used in constructing the covariances does not compensate for neglecting all cross terms.

4. The data

The data are synthesized such that they are consistent with the general null hypothesis, H_{0g}. First, a single sample residual trio $(f(x, t), i(x), b(t))$ consistent with H_{0g} is obtained as described above. Using these residuals, the associated sample solution, $u(x, t)$, is determined by solving the weak dynamics in eq. (5a). This field represents one realization of a 'true' ocean that is consistent with the statistical statements regarding the dynamic residuals; it is then 'sampled' at desired space–time locations, and a data residual consistent with H_{0g} (i.e. a random number from a population with zero mean and variance V_ε) is added to each sampled value. These synthesized data are consistent with the statistical statements regarding both the dynamics and data residuals in H_{0g}.

In these experiments, the data are gathered from a synthetic mooring at $x = 3.0$. That is, the spatial location is fixed and a

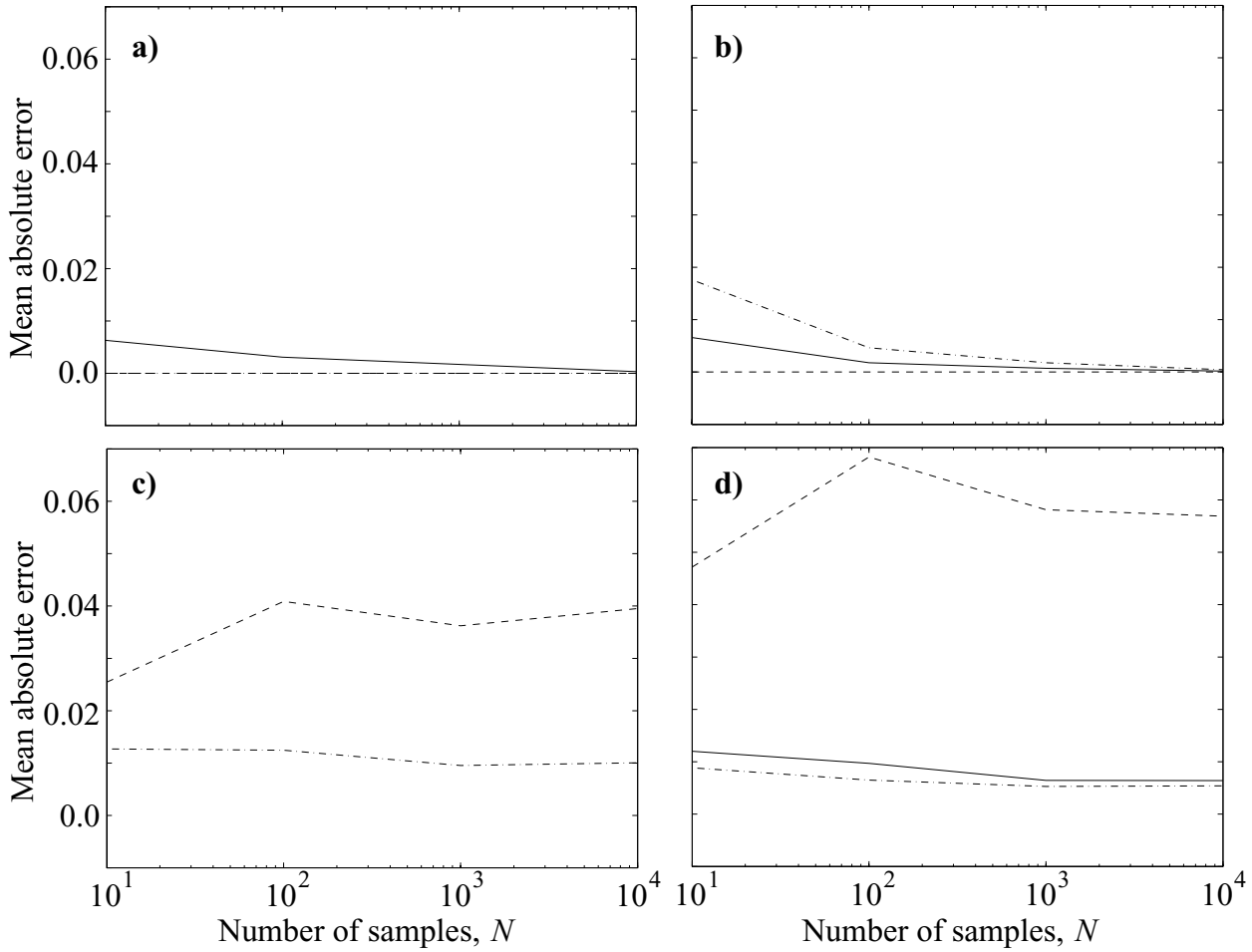

Fig. 4. Mean absolute error in covariance constraints: (a) eq. (21a) as the solid line, eq. (21b) as the dash-dotted line and eq. (21c) as the dashed line (note that the dashed and dash-dotted lines are coincident with a mean absolute error of zero); (b) eq. (22a) as the solid line, eq. (22b) as the dash-dotted line and eq. (22c) as the dashed line; (c) eq. (30b) as the dash-dotted line and eq. (30c) as the dashed line; (d) eq. (31a) as the solid line, eq. (31b) as the dash-dotted line and eq. (31c) as the dashed line.

datum is gathered at each time-step for $0.1 \leq t \leq 5.0$ except at $t = 3.0$ where the space–time location of the datum is coincident with the corner characteristic. As discussed earlier, the optimal solution can support discontinuities along that characteristic, and placing a datum there results in difficulties with the representer matrix as well as a failed consistency check on the EL solver (see the Appendix for a more detailed explanation). To avoid these inconveniences, the synthetic mooring does not take a measurement when the corner characteristic passes through it. The data set consists, therefore, of 49 point measurements of $u(x, t)$.

5. Results

The objective of this paper is to demonstrate the importance of cross-covariance terms in the data assimilation formulation. Thus, the comparison we wish to make is between

(i) the 'Standard Approach' (SA), which minimizes penalty functional J (eq. 7) *with no cross terms*, and

(ii) the 'Cross Term Approach' (CTA), which minimizes the penalty functional I (eq. 14) *with cross terms*.

Note that both computations use multidimensional numerical integration to evaluate all necessary convolutions.

Three-dimensional plots of the optimal solution $\hat{u}(x, t)$ for both approaches are shown in Fig. 5. The wall of discontinuity propagating along the characteristic in the SA solution arises entirely from discontinuous residuals, as the user specified F, I and B are compatible and cause no discontinuities. Figure 6a shows a comparison between the SA (red) and CTA (blue) optimal solutions and the true solution used to generate the data (black) at $x = 3.0$, the location of the synthetic mooring. Both data assimilation approaches are relatively successful in reproducing the true residual away from the characteristic (seen here at $t = 3.0$). However, only the CTA is continuous at the corner

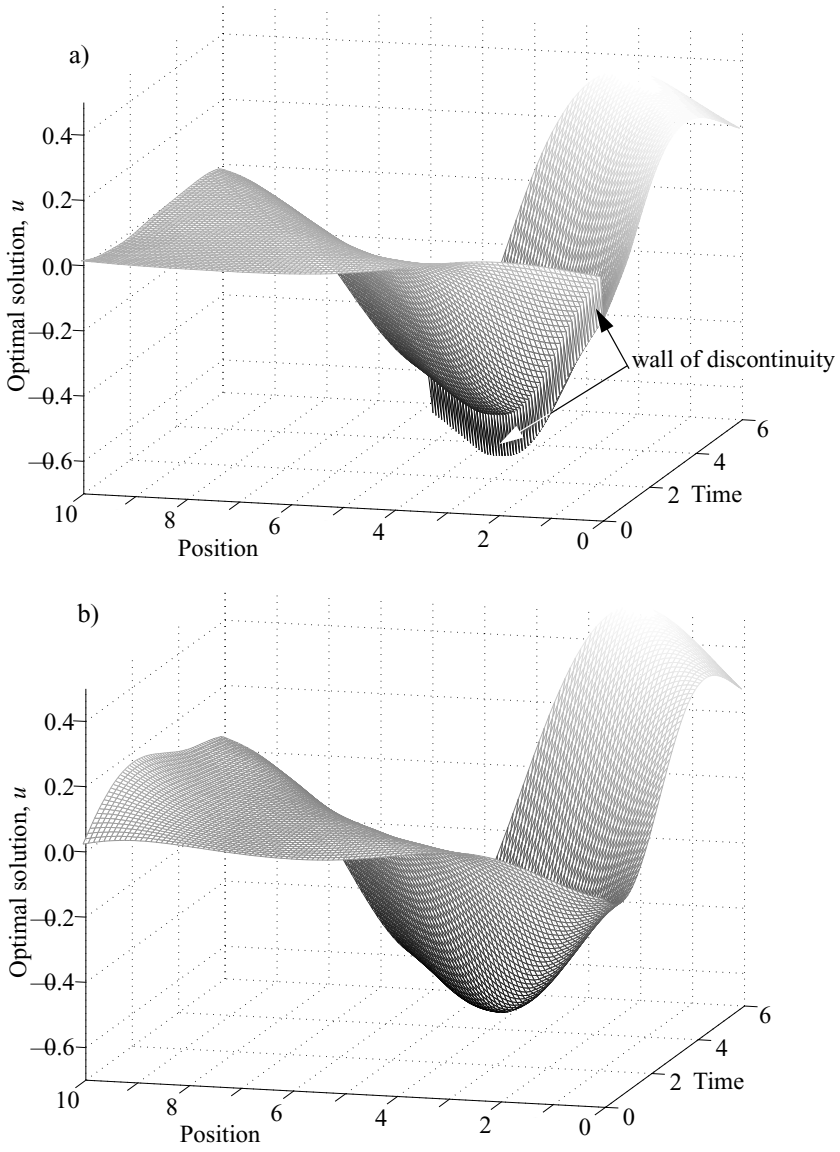

Fig. 5. Optimal solutions computed by (a) the SA and (b) the CTA. There is a wall of discontinuity along the corner characteristic of the SA solution, but the CTA solution is everywhere smooth. In the former case, the discontinuity arises entirely from incompatible residuals, as user-specified F, B and I (here $F = 0$, $B = 0$, and $I = 0$) are compatible at the corner and generate no discontinuities.

characteristic; the SA shows a discontinuity of approximately 15% of the wave amplitude. [As an interesting aside, notice that both assimilation approaches begin to fail as soon as the mooring is 'turned off' at $t = 5.0$.] The optimal solutions obtained at another fixed position, between the boundary and the mooring ($x = 0.8$) are shown in Fig. 6b along with the true solution. Again, the SA solution is discontinuous at the characteristic (seen here at $t = 0.8$), but the CTA solution is smooth there. [Note also that at this location, away from the mooring, the assimilation begins to fail much earlier than at the mooring.]

The results in Fig. 6 are representative of the results throughout the space–time domain; the SA solution has a persistent discontinuity that propagates from the corner along the corner characteristic while the CTA solution is everywhere smooth. These results demonstrate that if the cross terms are included in the penalty functional, and additionally if covariances satisfy the covariance constraints (21) and (22), then the residuals $\hat{f}$, $\hat{i}$, and $\hat{b}$ do indeed satisfy eqs. (11) and (12), resulting in a smooth and differentiable optimal solution across the corner characteristic.

It is particularly interesting to note that the hypothesis test is insensitive to the discontinuity. The minimal value of $\hat{J}$ for SA is 41.1 and the minimal value of $\hat{I}$ for CTA is 42.5. If the hypothesis is valid, as it is here, these are χ^2_M variables (where M is the number of data) and thus have means of 49 and standard deviations of $\sqrt{2 \times 49} \approx 10$; for more details, see Bennett (2002) and Muccino et al. (2003). Both values fall within 15% of one standard deviation of each other, and within one standard deviation of the mean. Thus, the hypothesis test does not justify rejection of the optimal solution (or any other data assimilation products) associated with either of the approaches. This 'failure'

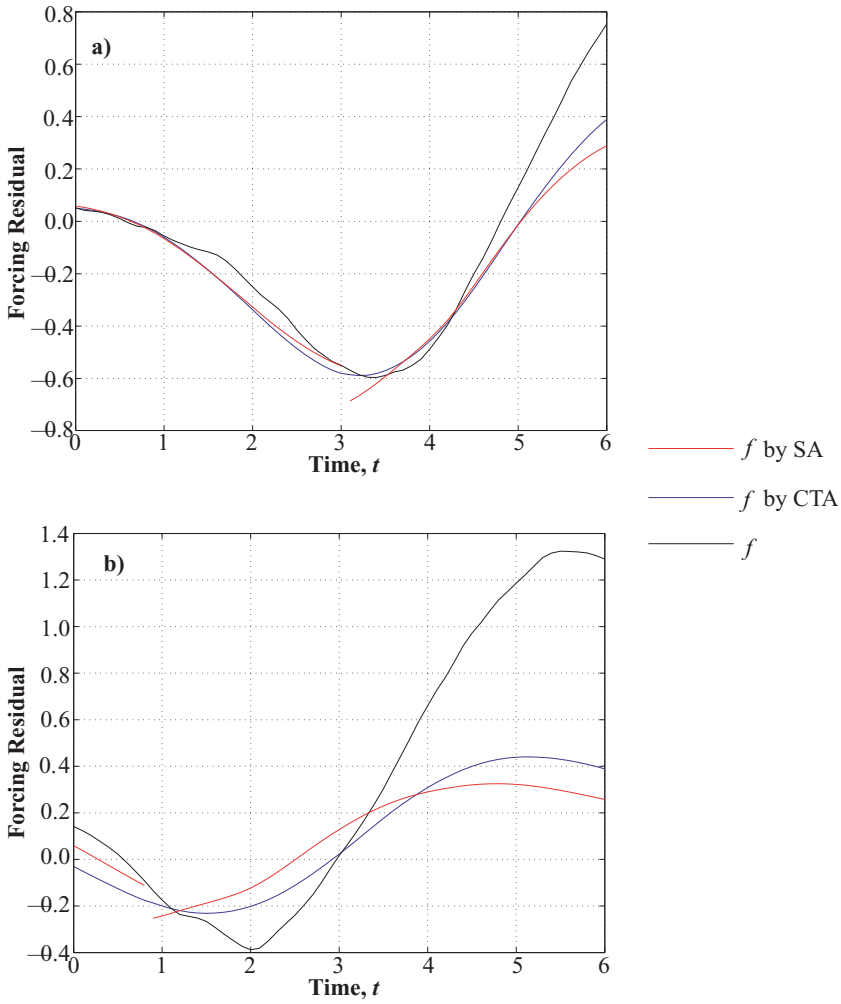

Fig. 6. (a) Comparison of the SA and CTA solutions and the true solution used to generate the data at $x = 3.0$, the location of the mooring. Both data assimilation approaches are relatively successful in reproducing the true solution away from the characteristic, but only the CTA solution is continuous at the characteristic ($t = 3.0$). (b) Same as (a), but at $x = 0.8$. Away from the mooring, neither data assimilation approach is as successful in reproducing the true solution, but again the SA solution is discontinuous at the corner characteristic ($t = 0.8$) while the CTA solution is everywhere smooth.

of the hypothesis test can be explained by considering that it can be shown that

$$\hat{J} = \mathbf{h}^{\mathrm{T}}\mathbf{P}^{-1}\mathbf{h} \tag{32}$$

where $\mathbf{h} = \mathbf{d} - \mathbf{u}_F$ is the misfit between the data and the dataless solution, and $\mathbf{P} = \mathbf{R} + \mathbf{C}_\varepsilon$ is the sum of the representer matrix and the data covariance matrix, $\mathbf{C_e}$. Thus, $\hat{J}$ depends upon the representers and the forward solution only at the data points and, in this problem, data along the discontinuity were disqualified, as justified in the Appendix. More fundamentally, although eq. (32) relates $\hat{J}$ to a data space scalar, the penalty functional is an integral with smooth kernels (i.e. inverse covariances) and therefore tends to smear the discontinuity.

6. Smoothing of the discontinuity with diffusion

One might assume that such discontinuities are interesting for their own sake, but really not of concern for realistic modeling because diffusion (both numerical and physical) might serve to smooth the discontinuity into a differentiable solution. However, will diffusion reproduce the inherently smooth solution free of bogus scales not in the hypothesis regarding the residuals nor in the internal dynamics of the model? To gain perspective on this issue, consider the same non-dimensional convection equation as before

$$\frac{\partial u}{\partial t} + \frac{\partial u}{\partial x} = F(x,t), \quad u(x,0) = I(x), \quad u(0,t) = B_1(t) \tag{33}$$

where (33) is discretized as in Section 3. Consider also a similar non-dimensional, unforced convection–diffusion equation:

$$\begin{aligned} &\frac{\partial u}{\partial t} + \frac{\partial u}{\partial x} = D\frac{\partial^2 u}{\partial x^2} + F(x,t), \quad u(x,0) = I(x), \\ &u(0,t) = B_1(t), \quad u(L,t) = B_2(t). \end{aligned} \tag{34}$$

Here, the first term of eq. (34) is discretized as in Section 3, with the diffusion term treated with an explicit centered difference

$$u_i^{k+1} = u_i^k - C\left(u_i^k - u_{i-1}^k\right) + \rho\left(u_{i-1}^k - 2u_i^k + u_{i+1}^k\right) \tag{35}$$

where $\rho = D\,\Delta t\,/\Delta x^2$ and $D = 0.05$. For stability, the time-step is reduced to 0.05, but the space-step remains 0.1. Of course,

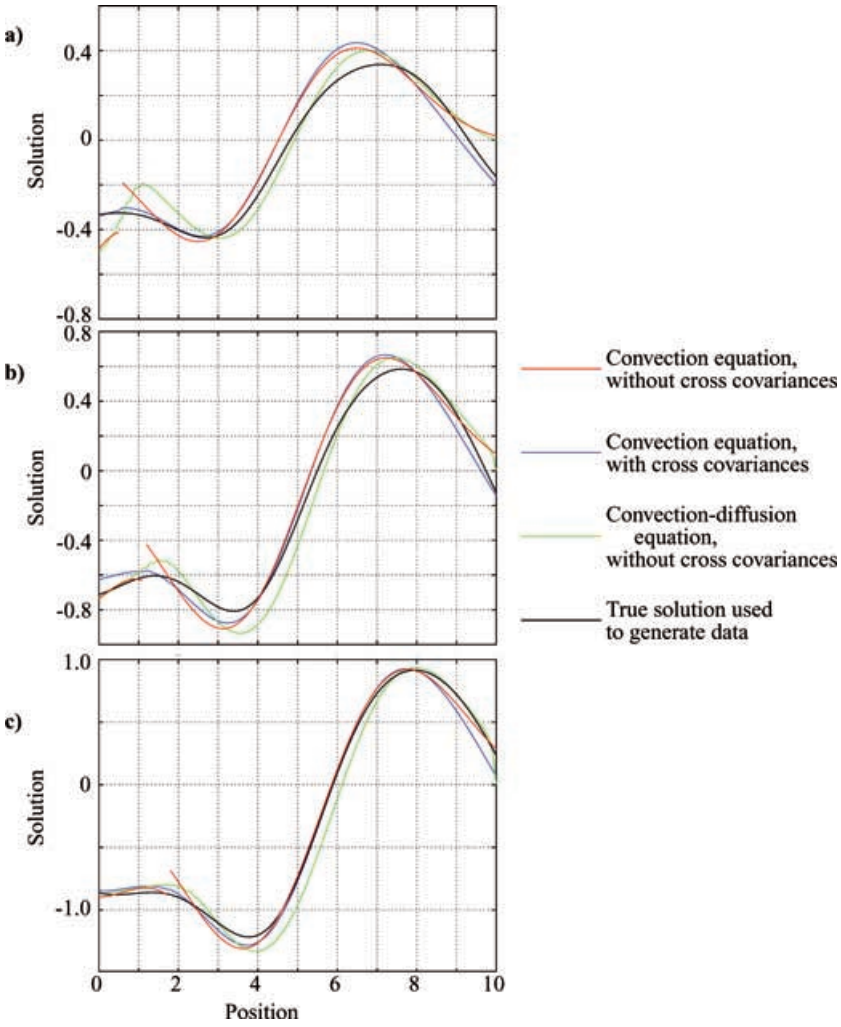

Fig. 7. Optimal solutions for the convection equation without and with cross-covariances and the convection–diffusion equation without cross-covariances as well as true solution used to generate the data at (a) $t = 0.6$, (b) $t = 1.2$, and (c) $t = 1.8$. Although the diffusion smoothes the discontinuity substantially, it does not reproduce the inherently smooth solution to the convection equation with cross-covariance terms; rather, it results in the development of a spurious boundary layer along the corner characteristic. The boundary layer broadens with time.

whereas the solution to eq. (33) may not be continuous or differentiable along the corner characteristic (depending upon the compatibility of F, I and B), one would expect the solution to eq. (34) to be differentiable for arbitrary F, I, B_1 and B_2 for $t > 0$. However, we again assume that the forcing, initial condition and boundary condition(s) are compatible and, without loss of generality, assume these are all equal to zero as we focus attention on variational assimilation with eqs. (33) and (34), respectively, and the impact of $\hat{f}$, $\hat{i}$ and $\hat{b}$ on the optimal solution.

The spatial extent of the POD used to generate the covariances is $-2.5 \leq x \leq 10$, so the extent of the NID is $0 \leq x \leq 10$. The temporal extent of both the POD and NID is $0 \leq t \leq 3$. The standard errors and scales (and therefore covariances) are the same as before. Note the continuity conditions on the covariances are independent of the dynamics. Also, the diffusion term in the conditions for continuously differentiability is small away from real physical boundaries, so it suffices to use the 'no diffusion' covariances with the small amount of diffusion in eq. (35).

For comparison, the solution to eq. (33) is computed without and with cross-covariances, and the solution to eq. (34) is computed without cross-covariances. Each of the three assimilations uses the same set of data which consists of synthetic measurements described in Section 4 gathered at $x = 1$, 2, 4, and 8 for each time-step for $0.1 \leq t \leq 2.4$, with the exception of (1, 1) and (2, 2), as justified in the Appendix. The optimal solutions for these three assimilations are shown at three different times in Fig. 7 along with the true solution used to generate the data. The diffusion does smooth over the discontinuity, but does not reproduce the inherently smooth solution obtained by including the cross-covariance terms in the non-diffusive convection equation. In fact, the diffusive process smoothes the discontinuity into a spurious internal boundary layer that broadens as a function of time. The width of the boundary scales is $\sqrt{DT}$, where T is the time elapsed since the characteristic exited the corner. This is not an internal scale such as D/c, where c is the advecting phase velocity (here, $c = 1$).

These results demonstrate that it is misleading to rely on diffusion to reproduce the smoothing impact of the cross-covariance terms, especially at early time. Also, as before, the hypothesis test fails to identify poor performance of the optimal solution when the cross-covariance terms are neglected, either without or with diffusion.

7. Conclusions

When solving hyperbolic PDEs, mathematical texts such as Courant and Hilbert (1962) have made it clear that the forcing, initial condition and boundary condition must be compatible at the space–time corner (i.e. $(x, t) = (0, 0)$ in two-dimensional problems). While often ignored, a modeler can ensure such compatibility by carefully specifying (1) initial and boundary conditions that are continuous at the corner, and (2) forcing, initial conditions and boundary conditions that satisfy the PDE at the corner.

When assimilating data, the optimal solution does not exactly satisfy the governing equation, initial condition and boundary condition; rather, it minimizes a penalty functional that penalizes residuals in the model and in the data. Therefore, it is not immediately clear whether appropriate compatibility conditions are satisfied at the corner.

Indeed, it has been shown here, both theoretically and numerically, that the 'standard' penalty functional, which neglects cross-correlations between residuals, does not impose the necessary constraints to ensure the necessary compatibility. The result is a solution that is not continuous, nor continuously differentiable, along the characteristic emanating from the space–time corner. Furthermore, it has also been demonstrated that, although diffusion does indeed smooth the resulting discontinuities, it does not necessarily reproduce the inherently smooth solution obtained by correctly admitting cross-correlations between residuals. Rather, in addition to the scales imposed by the hypothesis and the internal dynamics of the model, diffusion can introduce a spurious internal boundary layer, which broadens in time, along the corner characteristic.

Theory has been presented to demonstrate that in order to ensure a continuous and continuously differentiable optimal solution, cross-correlations satisfying necessary constraints must be included in the penalty functional. Experiments have demonstrated that this, indeed, is the case.

However, inclusion of the cross-covariance terms is computationally burdensome in two respects.

(i) Compatible closed-form covariances are not obvious; the covariances have been statistically simulated to ensure that the covariance constraints (15) and (17) are satisfied in the limit of large number of samples. Even if this task is simplified by assuming $C_{if} = C_{fi} = 0$, the statistical simulation requires a substantial number of samples to effectively convey the length- and time-scales in the null hypothesis.

(ii) Because the covariances are not represented by typical bell and Markovian forms, no efficient convolution algorithms like those described in, for example, Egbert et al. (1994) and Bennett (2002) are known and the convolutions must be computed by much more laborious numerical integration.

Of these tasks, the most computationally intense are the statistical simulation of C_{ff} and the convolution associated with it. This is a four-dimensional array ($NT \times NX \times NT \times NX$), requiring roughly five operations per array element per sample for statistical simulation and convolution (of the order of 10^{12} operations for these calculations). However, C_{ff} can be constructed and convolved according to the efficient algorithms in Egbert et al. (1994) and Bennett (2002) (requiring only 10^6 operations) without compromise. The computational savings are even more significant in higher dimensions. Efficient algorithms may also be used for C_{ii}, although its operation cost is negligible in comparison to that associated with C_{ff}.

Statistical simulation of C_{fb} and C_{bf} is also computationally burdensome (of the order of 10^{10} operations). One might reasonably argue that these cross-covariances are significant only in ensuring the differentiability of residuals, and that, from a practical perspective, diffusion (either numerical or physical) or finite element smearing would effectively smooth out kinks introduced by incompatibility in the differentiability constraints (18). One might therefore choose to ignore the differentiability constraints altogether, satisfying only the continuity constraints (17):

$$C_{bf}(0, ,) = 0 \tag{36a}$$

$$C_{bf}(0,) - C_{ii}(0,) = 0 \tag{36b}$$

$$C_{bb}(0,) - C_{ib}(0,) = 0. \tag{36c}$$

The only cross-covariances required here are C_{bi} and C_{ib}, which are easily computed.

Characteristics are also present in quasi-linear hyperbolic problems (Courant and Hilbert, 1962), so the same issues, discussed here within the framework of a linear problem, must be addressed. The representer algorithm requires linear EL equations. Nonlinearities are effectively handled within this framework by iterating on a sequence of linear EL equations, whose solutions converge to the solution of the original nonlinear problem (it is hoped). Thus, the importance of the cross-covariance terms in linear data assimilation problems directly attests to their importance in nonlinear problems as well.

Finally, we note that all calculations in this paper have used the representer algorithm to solve the fixed interval smoothing problem defined by eq. (14). Attempts to construct a sequential Kalman filter from the EL eqs. (8), (9), and (15) using the Gelfand and Fomin sweep algorithm (see, for example, Bennett, 2002) are as yet unsuccessful, owing to the assumption of

time-correlated boundary residuals in particular. Investigations are continuing.

8. Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant Nos. OCE-0121315 and OCE-0121542.

9. Appendix A: Impact of Incompatible Covariances on Representers

A very simplified version of the problem described in the body of the paper is used here to demonstrate the impact of neglecting cross-covariances on the representer matrix. Consider the non-dimensional, homogeneous convection equation over the domain $0 \le x \le 5$, $0 \le t \le 5$

$$\frac{\partial u}{\partial t} + \frac{\partial u}{\partial x} = 0, \quad u(x,0) = 0, \quad u(0,t) = 0 \tag{A1}$$

and two data

$$u(3,3) = d_1 \quad \text{and} \quad u(3,4) = d_2, \tag{A2}$$

as shown in Fig. 8. The null hypothesis neglects the cross terms, as in eq. (6), and the penalty functional is eq. (7), with $C_{ff} = 1$, $C_{ii} = 1$, and C_{bb} as some positive function of (t, s). The adjoint-representer and representer equations (see Bennett, 2002 for details) follow from eqs. (8) and (9)

$$\begin{aligned} &\frac{\partial \gamma_m}{\partial t} - \frac{\partial \gamma_m}{\partial x} = \delta(x - x_m)\,\delta(t - t_m), \\ &\gamma_m(x,T) = 0, \quad \gamma_m(0,L) = 0 \end{aligned} \tag{A3}$$

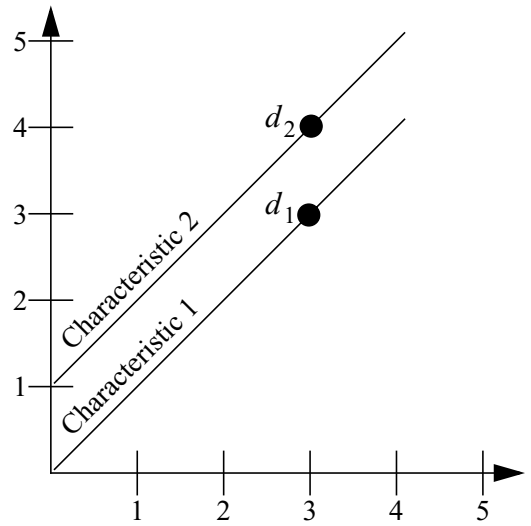

Fig. 8. Domain and data for demonstration of the impact of incompatible covariances on representers.

$$\begin{aligned} &\frac{\partial r_m}{\partial t} + \frac{\partial r_m}{\partial x} = \gamma_m(x,t), \\ &r_m(x,0) = \gamma_m(x,0), \quad r_m(0,t) = C_{bb} \bullet \gamma_m(0,t) \end{aligned} \tag{A4}$$

where $m = 1, 2$.

Consider first d_1. The function $\gamma_1(x, t)$ is a singularity that begins at d_1 and traces back to the origin along characteristic 1 and is zero everywhere else. Thus, $\gamma_m(x, 0) = r_m(x, 0)$ is singular at $x = 0$ and zero everywhere else. Likewise, $\gamma_m(0, t)$ is singular at $t = 0$ and zero elsewhere, but after convolving according to the third of eq. (A4), $r_m(0, t)$ is no longer singular. Here lies the crux of the issue – there is an incompatibility at the space–time corner. The initial condition is singular at the corner, but the boundary condition is not. For computation, we must make a choice and, in these calculations, we consistently choose the initial condition value at the corner. Thus, $r_1(x, t)$ is zero below, non-zero on and non-zero above Characteristic 1. Evaluating $r_1(x, t)$ at the data points yields $r_{11} = r_1(x_1, t_1) \neq 0$ and $r_{12}=r_1(x_2, t_2) \neq 0$.

Proceeding similarly for d_2, $\gamma_2(x, t)$ is a singularity that begins at d_2 and traces back to (0, 1) along characteristic 2 and is zero everywhere else. Thus, $\gamma_m(x, 0) = r_m(x, 0)$ is zero for all x. Furthermore, $\gamma_m(0, t)$ is singular at $t = 1$ and zero everywhere else, but after convolving according to the second term of eq. (A4), $r_m(0, t)$ is no longer singular, and is non-zero at the corner. Again, there is an incompatibility at the corner because the initial condition is zero, but the boundary condition is non-zero. Again, we choose the initial condition value over the boundary condition, and thus $r_2(x, t)$ is zero below, zero on and non-zero above characteristic 2. Evaluating $r_2(x, t)$ at the data points yields $r_{21} = r_2(x_1, t_1) = 0$ and $r_{22} = r_2(x_2, t_2) \neq 0$.

Thus, the representer matrix is

$$\mathbf{R} = \begin{bmatrix} r_{11} & r_{12} \\ r_{21} & r_{22} \end{bmatrix} = \begin{bmatrix} nz & nz \\ 0 & nz \end{bmatrix} \tag{A5}$$

where nz indicates a non-zero value; clearly $\mathbf{R}$ is non-symmetric. Thus, the incompatibility at the corner causes the symmetry requirement of the representer matrix to be violated, but only when the space–time location of a datum coincides with the corner characteristic.

It is easy to generalize this simple example to the problem in the body of the paper. Convolution of $\gamma_m(x, 0)$ to yield the initial condition and convolution of $\gamma_m(0, t)$ to yield the boundary condition for the representer calculation will, in general, yield different values at (0, 0). If we consistently choose the initial condition value for the computations, data above the characteristic will break the symmetry of the representer matrix. Of course, this broken symmetry could be restored by opting for the boundary condition value at the corner, but then data below the characteristic would break the symmetry.

It is important to recognize that when cross-covariances are neglected, even data located away from the corner

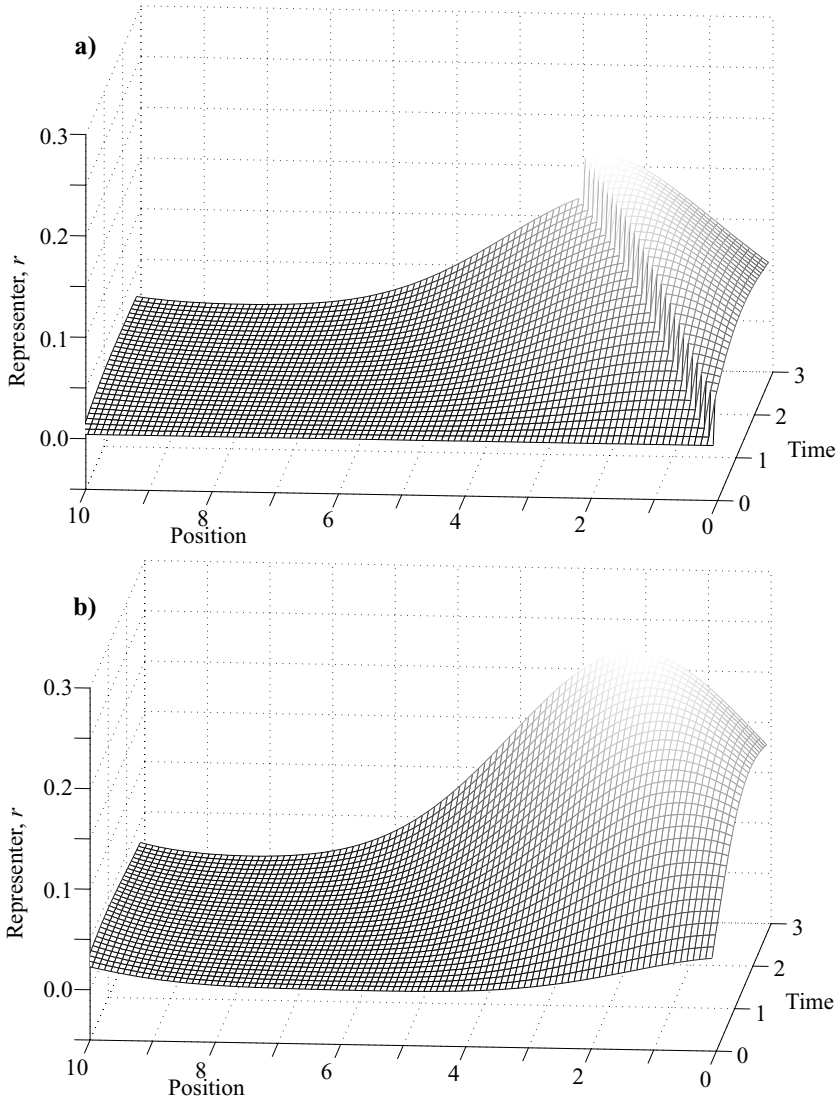

Fig. 9. Representer for data point at (1, 1.5) without cross-covariances and (b) representer for the same data point with cross-covariances. The cross-covariances are required to avoid incompatibility at the space–time corner, which in turn leads to discontinuities which propagate along the corner characteristic.

characteristic yield representers that are discontinuous along that characteristic, as demonstrated in Fig. 9. Because the optimal solution is the sum of the forward solution (which is continuous) and a linear combination of representers (each of which is discontinuous along the corner characteristic), one should expect discontinuities in the optimal solution along the corner characteristic.

References

Bennett, A. F. 2002. *Inverse Modeling of the Ocean and Atmosphere*. Cambridge Univ. Press, Cambridge.

Bernstein, D. L. 1950. *Existence Theorems in Partial Differential Equations*. Princeton Univ. Press, Princeton, NJ.

Bogden, P. S. 2001. The impact of model-error correlation on regional data assimilative models and their observational arrays. *J. Marine Res.* **59**, 831–857.

Burnett, W., Harding, J. and Heburn, G. 2002. Overview of operational forecasting in the U.S. Navy: Past, present and future. *Oceanography* **15**, 4–12.

Celia, M. A. and Gray, W. G. 1992. *Numerical Methods for Differential Equations*. Prentice Hall, Englewood Cliffs, NJ.

Courant, R. and Hilbert, D. 1962. *Methods of Mathematical Physics*, **Vol. 2**. Wiley Intersciences, New York.

Egbert, G. D., Bennett, A. F. and Foreman, M. G. G. 1994. TOPEX/POSEIDON tides estimated using a global inverse model. *J. Geophys. Res.* **99**C12, 24 821–24 852.

Hodur, R., 1997. The Naval Research Laboratory Coupled Ocean/Atmosphere Mesoscale Prediction System (COAMPS). *Mon. Wea. Rev.* **125**, 1414–1430.

Muccino, J. C. and Bennett, A. F. 2002. Generalized inversion of the Korteweg-de Vries equation. *Dyn. Atmos. Oceans* **35**, 227–263.

Muccino, J. C., Hubele, N. F. and Bennett, A. F. 2003. Significance testing for variational assimilation. *Q. J. R. Meteorol. Soc.*, in press.