Weighting Initial Conditions in Variational Assimilation Schemes

ANDREW F. BENNETT AND ROBERT N. MILLER

College of Oceanography, Oregon State University, Corvallis, Oregon

9 May 1990, and 6 September 1990

ABSTRACT

An analysis of variational data assimilation schemes for linear dynamical forecast models shows that the penalty functional must include an explicit contribution from the initial conditions in order to ensure a unique, low-noise forecast. The noise level is related to the effective number of data being assimilated.

1. Introduction

The use of data to guide the choice of initial conditions for a forecast model was considered in Talagrand and Courtier (1987) and in Courtier and Talagrand (1987). The approach used in these widely influential papers (hereafter referred to as TC and CT, respectively) was to choose initial conditions such that the corresponding forecast was as close as possible to data at later times, in a least squares sense. It was reported in CT that the forecast providing the best fit possessed considerable small-scale noise, but the noise level could be greatly reduced by minimizing a weighted sum of squares of data misfits, plus squares of spatial derivatives of the initial conditions.

This note offers a partial explanation for the reported relationship between forecast noise and the penalizing of initial roughness. The essential point is that the relationship goes beyond simple low-pass filtering, and involves the well conditioning of the underlying inverse problem. Our explanation is only partial since we analyze a linear forecast model, whereas TC and CT considered a nonlinear model. Otherwise our model is virtually the same as theirs; it is single-layer quasi-geostrophic, with periodic boundary conditions. For simplicity alone we replace their hemispheric geometry and spherical harmonic expansion with β-plane Cartesian coordinates and a trigonometric Fourier series expansion.

2. The model initial conditions and data

The linear single-level quasi-geostrophic model may be expressed in terms of complex Fourier coefficients,

$$\frac{d}{dt}\psi_{\mathbf{n}} = L_{\mathbf{n}}\psi_{\mathbf{n}} + F_{\mathbf{n}} + q_{\mathbf{n}}, \qquad (2.1)$$

where $\psi_{\mathbf{n}}(t)$ is a Fourier coefficient for the streamfunction:

$$\psi(\mathbf{x}, t) = \sum_{\mathbf{n}} \psi_{\mathbf{n}}(t)e^{i\mathbf{n}\cdot\mathbf{x}}, \qquad (2.2)$$

while $\mathbf{x} = (x_1, x_2)$, $\mathbf{n} = (n_1, n_2)$, $L_{\mathbf{n}}$ is a complex number with negative real part expressing the model physics (see Bennett and Budgell 1989), $F_{\mathbf{n}}(t)$ is a first guess for the model forcing, and $q_{\mathbf{n}}(t)$ is the error in the first guess for the forcing; that is, $F_{\mathbf{n}} + q_{\mathbf{n}}$ is the "true" forcing and $\psi_{\mathbf{n}}$ is the "true" streamfunction. Suitable initial conditions for (2.1) are provided by

$$\psi_{\mathbf{n}}(0) = \Psi_{\mathbf{n}} + a_{\mathbf{n}} \qquad (2.3)$$

where $\Psi_{\mathbf{n}}$ is the first-guess initial condition and $a_{\mathbf{n}}$ is the first-guess error.

The most general forms of the measurements that we consider are linear functionals acting on the streamfunction field:

$$\psi(\mathbf{x}, t) \rightarrow \mathcal{L}[\psi] \qquad (2.4)$$

where $\mathcal{L}[\psi]$ is a single real number, such as the value of ψ at a single point and time; an average over some region in space or time; or the local value of a geostrophic velocity component. For example, in terms of the Fourier coefficients, streamfunction measurement at position $\mathbf{z}$ and time s has the functional form

$$\mathcal{L}[\psi] = \sum_{\mathbf{n}} \psi_{\mathbf{n}}(s)e^{i\mathbf{n}\cdot\mathbf{z}}. \qquad (2.5)$$

In practice there will be a finite number M of measurements, expressed in the form

$$d_m = \mathcal{L}_m[\psi] + \epsilon_m, \quad (1 \le m \le M) \qquad (2.6)$$

where d_m is the mth datum, $\mathcal{L}_m$ is the mth measurement functional, ψ is again the true streamfunction, while ϵ_m is the error in the mth datum. The data may be conveniently defined in the compact form

Corresponding author address: Dr. Andrew F. Bennett, College of Oceanography, Oregon State University, Corvallis, OR 97331-5503.

$$\mathbf{d} = \mathscr{L}[\psi] + \epsilon \tag{2.7}$$

where $\mathbf{d}$, $\mathscr{L}$, and ϵ are vectors of length M.

The first-guess solution $\psi_{\mathbf{n}_F}$ satisfies (2.1) and (2.3) with $q_\mathbf{n}$ and $a_\mathbf{n}$ set to zero. The first-guess data misfit is

$$\mathbf{h} = \mathbf{d} - \mathscr{L}[\psi_F], \tag{2.8}$$

and in general is not zero.

3. Weighted least squares

We seek the streamfunction estimate $\hat{\psi}_\mathbf{n}$ such that the corresponding model residual $\hat{q}_\mathbf{n}$, the corresponding initial residual $\hat{a}_\mathbf{n}$, and the corresponding data residual $\hat{\epsilon}$ are minimal in a weighted least-squares sense during some fixed time interval $0 \leqslant t < T$. The approach taken by TC is a special case in which the model residual has infinite weight, while the initial residual has zero weight. In terms of Fourier coefficients a suitable quadratic penalty functional is

$$\mathscr{I}[\psi] = \int_0^T \sum_\mathbf{n} W_\mathbf{n} |q_\mathbf{n}(t)|^2 dt + \sum_\mathbf{n} V_\mathbf{n} |a_\mathbf{n}|^2 + w|\epsilon|^2 \tag{3.1}$$

where $q_\mathbf{n}$, $a_\mathbf{n}$, and ϵ are residuals related to ψ by (2.1), (2.3), and (2.7), while $W_\mathbf{n}$, $V_\mathbf{n}$, and w are real constant weights. It is assumed that the data were collected during the time interval $0 \leqslant t \leqslant T$. In a more general formulation, the quadratic functional $\mathscr{I}$ would involve products of Fourier coefficients with different indices and at different times, and products of data residuals for different measurements, but the simple "diagonal" form (3.1) suffices for this analysis. The inclusion of a dynamical residual such as $q_\mathbf{n}$ is standard in the general theory of state estimation. It is routinely included in Kalman filtering and smoothing (Gelb 1974). Bennett and McIntosh (1982) and Derber (1989) provide oceanographic and meteorological examples, respectively, of the explicit inclusion of dynamical residuals in variational assimilation.

There is substantial literature on least-squares inverse theory, such as the monographs by Menke (1984) and Tarantola (1987) in the context of solid earth geophysics. The general principles that we shall discuss may be deduced from these monographs, but are developed here in the context of a time-dependent dynamical model.

It is a textbook exercise in the calculus of variations to show that $\mathscr{I}$ has an extremum for $\psi_\mathbf{n} = \hat{\psi}_\mathbf{n}$ satisfying

$$\frac{d}{dt}\hat{\psi}_\mathbf{n} = L_\mathbf{n}\hat{\psi}_\mathbf{n} + F_\mathbf{n} + W_\mathbf{n}^{-1}\lambda_\mathbf{n}, \tag{3.2}$$

subject to

$$\hat{\psi}_\mathbf{n}(0) = \Psi_\mathbf{n} + V_\mathbf{n}^{-1}\lambda_\mathbf{n}(0) \tag{3.3}$$

where the "adjoint variable" $\lambda_\mathbf{n}(t)$ satisfies the "adjoint" or Euler–Lagrange equation

$$-\frac{d}{dt}\lambda_\mathbf{n} = L_\mathbf{n}^*\lambda_\mathbf{n} + w\mathscr{L}'[e^{i\mathbf{n}\cdot\mathbf{x}'}\delta(t-t')]^*(\mathbf{d} - \mathscr{L}[\hat{\psi}]) \tag{3.4}$$

and

$$\lambda_\mathbf{n}(T) = 0. \tag{3.5}$$

In (3.4) the superscript asterisk denotes the complex conjugate of a scalar and the conjugate transpose of a vector. The prime on the measurement functional indicates that it acts on $\mathbf{x}'$ and t'. For example, if $\mathcal{L}_m$ is evaluation at $\mathbf{x}_m$ and t_m, then, referring to (2.5),

$$\mathscr{L}'_m[e^{i\mathbf{n}\cdot\mathbf{x}'}\delta(t-t')] = e^{i\mathbf{n}\cdot\mathbf{x}_m}\delta(t-t_m). \tag{3.6}$$

The optimal estimates for the residuals may be seen to be

$$\hat{q}_\mathbf{n}(t) = W_\mathbf{n}^{-1}\lambda_\mathbf{n}(t),$$
$$\hat{a}_\mathbf{n} = V_\mathbf{n}^{-1}\lambda_\mathbf{n}(0), \quad \hat{\epsilon} = \mathbf{d} - \mathscr{L}[\hat{\psi}]. \tag{3.7}$$

The system (3.2)–(3.5) defines a linear, second-order two-point boundary value problem for the time interval $0 \leqslant t \leqslant T$. It may be solved using the Gelfand and Fomin sweep algorithm (see for example Bennett and Budgell 1989) or using a finite sum of the Riesz representers for the M measurement functionals $\mathcal{L}_m$ (see for example, Bennett 1985). We will not be concerned with solution methods, but rather the uniqueness of solutions. Indeed, uniqueness of the solution must be established before choosing a particular solution method.

It is appropriate at this point to identify the special case considered in TC and CT. In effect they chose w finite but

$$W_\mathbf{n} \to \infty, \quad V_\mathbf{n} \to 0. \tag{3.8}$$

Consequently, they made a null estimate for the model residual: $\hat{q}_\mathbf{n}(t) \equiv 0$, which is consistent with assuming a perfect model, having perfect dynamics $L_\mathbf{n}$, and a perfect first-guess $F_\mathbf{n}$ for the forcing. Interpretation of (3.3) as $V_\mathbf{n} \to 0$ requires care. The actual condition that $\mathscr{I}$ have an extremum with respect to variations of $\psi_\mathbf{n}(0)$ is that

$$\lambda_\mathbf{n}(0) = V_\mathbf{n}\{\hat{\psi}_\mathbf{n}(0) - \Psi_\mathbf{n}\}, \tag{3.9}$$

which reduces simply to

$$\lambda_\mathbf{n}(0) = 0 \tag{3.10}$$

when $V_\mathbf{n} \to 0$.

4. Uniqueness of the extremum

Suppose that, for fixed choices of $F_\mathbf{n}$, $\Psi_\mathbf{n}$, and $\mathbf{d}$, the system (3.2)–(3.5) has two solutions $\hat{\psi}_\mathbf{n}^{(1)}$ and $\hat{\psi}_\mathbf{n}^{(2)}$, with corresponding adjoint variables $\lambda_\mathbf{n}^{(1)}$ and $\lambda_\mathbf{n}^{(2)}$.

Then it is readily shown that their differences $\hat{\phi}_{\mathbf{n}} = \hat{\psi}_{\mathbf{n}}^{(1)} - \hat{\psi}_{\mathbf{n}}^{(2)}$, $\mu_{\mathbf{n}} = \lambda_{\mathbf{n}}^{(1)} - \lambda_{\mathbf{n}}^{(2)}$ satisfy the homogeneous form of (3.2)–(3.5); that is, $F_{\mathbf{n}}$, $\Psi_{\mathbf{n}}$, and $\mathbf{d}$ all vanishing, and consequently

$$\int_0^T \sum_{\mathbf{n}} W_{\mathbf{n}}^{-1}|\mu_{\mathbf{n}}(t)|^2 dt + \sum_{\mathbf{n}} V_{\mathbf{n}}|\hat{\phi}_{\mathbf{n}}(0)|^2 + w|\mathscr{L}[\hat{\phi}]|^2 = 0. \quad (4.1)$$

This quadratic functional is similar but not identical to the penalty functional (3.1). A series of conclusions C1–C7 may be drawn from (4.1).

(C1)

$$w > 0: \quad \text{then} \quad \mathscr{L}[\hat{\phi}] = 0 \quad \text{so} \quad \mathscr{L}[\hat{\psi}^{(1)}] = \mathscr{L}[\hat{\psi}^{(2)}]; \quad (4.2)$$

that is, the difference between the solutions is unobservable.

(C2)

$$W_{\mathbf{n}} > 0, \quad (\text{all } \mathbf{n}): \quad \text{then} \quad \mu_{\mathbf{n}}(t) \equiv 0 \quad \text{so} \quad \hat{q}_{\mathbf{n}}^{(1)} = W_{\mathbf{n}}^{-1}\lambda_{\mathbf{n}}^{(1)} = W_{\mathbf{n}}^{-1}\lambda_{\mathbf{n}}^{(2)} = \hat{q}_{\mathbf{n}}^{(2)}; \quad (4.3)$$

that is, the solutions yield the same estimate for the model residual.

(C3)

$$V_{\mathbf{n}} > 0, \quad (\text{all } \mathbf{n}): \quad \text{then} \quad \hat{\phi}_{\mathbf{n}}(0) \equiv 0 \quad \text{so} \quad \hat{\psi}_{\mathbf{n}}^{(1)}(0) = \hat{\psi}_{\mathbf{n}}^{(2)}(0); \quad (4.4)$$

that is, the solutions yield the same estimate for the initial noise.

(C4)

$$W_{\mathbf{n}} > 0, V_{\mathbf{n}} > 0 \text{ (all } \mathbf{n}\text{): then } \mu_{\mathbf{n}}(t) \equiv 0 \text{ and } \hat{\phi}(0) = 0, \text{ so } \hat{\phi}_{\mathbf{n}}(t) \equiv 0 \text{ and thus } \hat{\psi}_{\mathbf{n}}^{(1)}(t) \equiv \hat{\psi}_{\mathbf{n}}^{(2)}(t); \quad (4.5)$$

that is, the solution $\hat{\psi}_{\mathbf{n}}(t)$ is unique.

Conclusion (C4) is the uniqueness theorem for (3.2)–(3.5). Note that it holds even if $w = 0$; that is, there are no data or only worthless data. In such circumstances the optimal estimate $\hat{\psi}_{\mathbf{n}}$ would coincide with the first-guess $\psi_{\mathbf{n}_F}$, and the extreme value of $\mathscr{I}$ would be zero (and hence the minimum value).

It may be shown (see Bennett 1985) that the unique solution $\hat{\psi}_{\mathbf{n}}$ may be expressed as the sum of the first-guess solution $\psi_{\mathbf{n}_F}$, plus a linear combination of M "representers" $r_{\mathbf{n}}^{m}$:

$$\hat{\psi}_{\mathbf{n}}(t) = \psi_{\mathbf{n}_F}(t) + \sum_{m=1}^{M} b_m r_{\mathbf{n}}^{m}(t). \quad (4.6)$$

The coefficients $(b_1, \cdots, b_M) = \mathbf{b}^*$ in (4.6) are solutions of the finite-dimensional linear system:

$$(\mathbf{R} + w^{-1}\mathbf{I})\mathbf{b} = \mathbf{h} \quad (4.7)$$

where $\mathbf{h}$ is given by (2.8), $\mathbf{I}$ is the $M \times M$ unit matrix, and $\mathbf{R}$ is a positive definite Hermitian matrix. The representer $r_{\mathbf{n}}^{m}(t)$ and representer matrix $\mathbf{R}$ are determined by the dynamics $L_{\mathbf{n}}$, the model weight $W_{\mathbf{n}}$, the initial weight $V_{\mathbf{n}}$, and the measurement functional $\mathcal{L}^{m}$.

(C5) $W_n \to \infty$ (all $\mathbf{n}$): the perfect model. Then the integral over time in (4.1) vanishes identically, as does the model residual $\hat{q}_{\mathbf{n}}(t) = W_{\mathbf{n}}^{-1}\lambda_{\mathbf{n}}(t)$ in (3.2). It may be shown that the adjoint variable $\lambda_{\mathbf{n}}(t)$ is the Lagrange multiplier for the perfect model, the latter being a "strong constraint" (Sasaki 1970) for the minimization of $\mathscr{I}$. Note that $W_{\mathbf{n}}|q_{\mathbf{n}}|^2 = \lambda_{\mathbf{n}}^*(t)q_{\mathbf{n}}(t) = W_{\mathbf{n}}^{-1}|\lambda_{\mathbf{n}}|^2 \to 0$ as $W_{\mathbf{n}} \to \infty$, so $\mathscr{I}$ includes only the initial and data residuals in this limit. Note also that C5 is a special case of C2 and C4.

(C6) It is clearly of *no* interest to consider $W_{\mathbf{n}} = 0$ (all $\mathbf{n}$); that is, a worthless model due either to worthless dynamics of a worthless first guess for the forcing. Obviously there is a large class of functions that vanish initially and that interpolate the data $\mathbf{d}$.

(C7) $w > 0$, $W_{\mathbf{n}} > 0$, $V_{\mathbf{n}} = 0$ (all $\mathbf{n}$): then $\mathscr{L}[\hat{\phi}] = 0$ and $\hat{\mu}_{\mathbf{n}}(t) \equiv 0$ so the difference between two solutions is unobservable and the solutions have the same model residual. However, since $V_{\mathbf{n}} = 0$ (all $\mathbf{n}$) there is no constraint on the initial values of the solution; it is only known that the model residual vanishes at time $t = 0$ [see (3.10)]. Uniqueness of the optimal estimate $\hat{\psi}_n$ cannot be inferred. Indeed, let $\hat{\theta}_{\mathbf{n}}$ be any (set of) Fourier coefficients satisfying

$$\frac{d}{dt}\hat{\theta}_{\mathbf{n}} = L_{\mathbf{n}}\hat{\theta}_{\mathbf{n}} + F_{\mathbf{n}} \quad (4.8)$$

and

$$\mathscr{L}[\hat{\theta}] = \mathbf{d}. \quad (4.9)$$

Since $V_{\mathbf{n}} = 0$ (all $\mathbf{n}$), it follows that $\mathscr{I}[\hat{\theta}] = 0$, for any values of $\hat{\theta}_{\mathbf{n}}(0)$. Note that there is no residual in (4.8), even if $W_{\mathbf{n}} < \infty$ for some $\mathbf{n}$ (imperfect dynamics). Further, $\mathscr{I}[\hat{\theta} + \hat{\xi}] = 0$ for any $\hat{\xi}_{\mathbf{n}}$ satisfying

$$\frac{d}{dt}\hat{\xi}_{\mathbf{n}} = L_{\mathbf{n}}\hat{\xi}_{\mathbf{n}}, \quad (4.10)$$

for any values of $\hat{\xi}_{\mathbf{n}}(0)$ such that

$$\mathscr{L}[\hat{\xi}] = 0. \quad (4.11)$$

That is, the streamfunction that minimizes $\mathscr{I}$ is undefined up to any unobservable addition to the initial streamfunction. Moreover, the model weight and data weight have no influence on the many possible minima. This last remark may be shown to hold even if these weights are nondiagonal matrices.

Finally, it is remarked that in case C7, minima of $\mathscr{I}$ are of the general form

$$\hat{\psi}_{\mathbf{n}} = \psi_{F_{\mathbf{n}}} + \eta_{\mathbf{n}} \quad (4.12)$$

where

$$\frac{d}{dt}\eta_{\mathbf{n}} = L_{\mathbf{n}}\eta_{\mathbf{n}} \tag{4.13}$$

and

$$\mathscr{L}[\eta_{\mathbf{n}}] = \mathbf{h} = \mathbf{d} - \mathscr{L}[\psi_F], \tag{4.14}$$

but $\eta_{\mathbf{n}}(0)$ is otherwise arbitrary.

If the observing system was dense in space or time, then any such "correction" field $\eta_{\mathbf{n}}(t)$ would have fine spatial or temporal structure. We shall argue in section 5 that nontrivial $\eta_{\mathbf{n}}$ evidently exist.

5. Practical considerations

In practice the number of Fourier coefficients $\psi_{\mathbf{n}}(t)$ is some finite number, say N^2. If $N^2 \leqslant M$, where again M is the number of measurements, then there is a unique extremum for $\mathscr{J}$, and the extreme value of $\mathscr{J}$ is zero in general only if $N^2 = M$. Usually $N^2 \gg M$, and if $V_{\mathbf{n}} = 0$ (all $\mathbf{n}$), then there are $N^2 - M$ ways of choosing the N^2 initial values $\hat{\psi}_{\mathbf{n}}(0)$ such that the solution of

$$\frac{d}{dt}\hat{\psi}_{\mathbf{n}} = L_{\mathbf{n}}\hat{\psi}_{\mathbf{n}} + F_{\mathbf{n}} \tag{5.1}$$

satisfies

$$\mathbf{d} = \mathscr{L}[\hat{\psi}]; \tag{5.2}$$

that is, $\mathscr{J} = 0$.

In TC and CT the choices $V_{\mathbf{n}} = 0$ (all $\mathbf{n}$), $W_{\mathbf{n}} \to \infty$ (all $\mathbf{n}$) were made, even though for the linear model at least the second choice is irrelevant once the first choice is made. Minimization of $\mathscr{J}$ was achieved by gradient descent; that is, $\hat{\psi}_{\mathbf{n}}(0)$ was calculated by moving down the gradient with component

$$\mathscr{J}_{\mathbf{n}} = \frac{\partial \mathscr{J}}{\partial \psi_{\mathbf{n}}(0)}. \tag{5.3}$$

It is readily seen that the adjoint or Euler–Lagrange initial condition (3.10) is actually the condition

$$\mathscr{J}_{\mathbf{n}} = 0. \tag{5.4}$$

Computation of $\mathscr{J}_{\mathbf{n}}$ was elegantly performed in CT by integrating (3.4) subject to (3.5), the value of $\mathcal{L}[\hat{\psi}]$ on the right-hand side of (3.4) being given by integration of (3.2) from a previous estimate of $\hat{\psi}_{\mathbf{n}}(0)$. Recall that $W_{\mathbf{n}}^{-1} = 0$ in TC and CT. Regardless of how minimization is achieved, the analysis of the linear model in section 4 indicates that any minimizing field $\hat{\psi}_{\mathbf{n}}$ must possess fine structure in space or time, as reported in CT.

The number of initial spectral coefficients in CT was equivalent to $N^2 = 231$, while there were $M = 5479$ observations of either geopotential height or velocity components on the 500-mb surface. Indeed, the quadratic penalty function in CT consisted of a diagonal sum over all 5479 measurement residuals, similar to the last term in (3.1). Consequently the inverse problem would appear to be substantially overdetermined, and hence possess a unique solution. However, the effective number of observations is not easily estimated. In terms of our analysis, the scalar weight w in (3.1) should, in general, be replaced by a symmetric positive definite matrix. At issue is the number of independent measurement errors. If the errors were instrumental in origin or were simply mesoscale variability, then all 5479 residuals could be regarded independent and a diagonal sum would be appropriate. On the other hand, one might expect that a significant proportion of the error variance would be correlated on synoptic scales. For example, velocity errors of synoptic scale could be due to divergent motions of synoptic scale while geopotential errors of synoptic scale could be due to relative errors in measuring synoptic-scale fields. Such variance would possess relatively few degrees of freedom. Indeed, 81% of the variance in geopotential height at 700-mb may be represented with just 16 EOFs (Tracton et al. 1989). Moreover, these 700-mb EOFs were computed from hemispheric analyses having a finer spatial truncation (rhomboidal 40) than that used in CT. In conclusion, the effective value of M for CT could be anywhere in the range 16–5479, depending upon the proportions of synoptic-scale and subsynoptic-scale measurement error.

In order to reduce the fine structure in their solution, CT introduced nonzero weighting for the initial streamfunction of the form

$$V_{\mathbf{n}} = V_0(|\mathbf{n}| + 1)^p \tag{5.5}$$

where V_0 is a positive scale factor and p a positive power. Consequently, the optimal solution $\hat{\psi}_{\mathbf{n}}$, at least for the linear model, was rendered unique. As discussed in section 4, the unique solution, at least for the linear model, may be expressed in the form (4.6) and (4.7) where $\mathbf{R}$ is inversely proportional to the scale factor V_0. It follows that the condition of the system (4.7) is determined by the ratio $\rho = w/V_0$, with $\rho \gg 1$ yielding poor conditioning in general. Thus, while introducing nonzero initial weights leads to a unique solution, the solution may still be very sensitive to the data. The sensitivity would also manifest itself as fine spatial or temporal structure, since the smallest eigenvalues of $\mathbf{R} + \mathrm{w}^{-1}\mathbf{I}$ would be associated with eigenvectors having components of comparable magnitudes but rapidly fluctuating signs. The solution for the representer coefficient vector $\mathbf{b}$ would also possess this fine component structure and would induce corresponding fine spatial and temporal structure on $\hat{\psi}_{\mathbf{n}}(t)$ and hence $\hat{\psi}(\mathbf{x}, t)$. This fine structure is reduced as p is increased, since the representers $r_{\mathbf{n}}{}^m$ are $O(|\mathbf{n}|^{-s})$ as $|\mathbf{n}| \to \infty$, where $s = O(p)$ as $p \to \infty$.

6. Conclusions

Analysis of a linear model indicates that fine structure should be expected in the results of variational data assimilation schemes unless the initial conditions are included in the penalty functional, with the smaller scales incurring higher penalty than larger scales. We offer this analysis as a partial explanation for the results reported in TC and CT, which involved a nonlinear model.

Acknowledgments. Conversations with Leonard Walstad have been helpful. Andrew F. Bennett is supported by National Science Foundation Grant OCE-8800004, and Robert N. Miller by the Office of Naval Research Grant N00014-90-J-1125. Florence Beyer typed the manuscript.

REFERENCES

Bennett, A. F., 1985: Array design by inverse methods. *Progr. Oceanogr.,* **15,** 129–156.

——, and W. P. Budgell, 1989: The Kalman smoother for a linear quasi-geostrophic model of ocean circulation. *Dyn. Atmos. Oceans,* **13,** 219–267.

——, and P. C. McIntosh, 1982: Open ocean modeling as an inverse problem: Tidal theory. *J. Phys. Oceanogr.,* **12,** 1004–1018.

Courtier, P., and O. Talagrand, 1987: Variational assimilation of meteorological observations with the adjoint vorticity equation. II: Numerical results. *Quart. J. Roy. Meteorol. Soc.,* **113,** 1329–1347.

Derber, J. C., 1989: A variational continuous assimilation scheme. *Mon. Wea. Rev.,* **117,** 2437–2446.

Gelb, A., 1974: *Applied Optimal Estimation.* MIT Press, 374 pp.

Menke, W., 1984: *Geophysical Data Analysis: Discrete Inverse Theory.* Academic Press, New York, 260 pp.

Sasaki, Y., 1970; Some basic formalism in numerical variational analysis. *Mon. Wea. Rev.,* **98,** 43–60.

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

Tarantola, A., 1987: *Inverse Problem Theory.* Elsevier, New York, 613 pp.

Tracton, M. S., K. Mo, W. Chen, E. Kalnay, R. Kistler and G. White, 1989: Dynamical extended range forecasting (DERF) at the National Meteorological Center. *Mon. Wea. Rev.,* **117,** 1604–1635.