

## Adaptive Tuning of Numerical Weather Prediction Models: Simultaneous Estimation of Weighting, Smoothing, and Physical Parameters

JIANJIAN GONG AND GRACE WAHBA

*Department of Statistics, University of Wisconsin—Madison, Madison, Wisconsin*

DONALD R. JOHNSON

*Space Science and Engineering Center, University of Wisconsin—Madison, Madison, Wisconsin*

JOSEPH TRIBBIA

*National Center for Atmospheric Research, Boulder, Colorado*

(Manuscript received 24 July 1996, in final form 23 January 1997)

### ABSTRACT

In Wahba et al. it was shown how the randomized trace method could be used to adaptively tune numerical weather prediction models via generalized cross validation (GCV) and related methods. In this paper a “toy” four-dimensional data assimilation model is developed (actually one space and one time variable), consisting of an equivalent barotropic vorticity equation on a latitude circle, and used to demonstrate how this technique may be used to simultaneously tune weighting, smoothing, and physical parameters. Analyses both with the model as a strong constraint (corresponding to the usual 4D-Var approach) and as a weak constraint (corresponding theoretically to a fixed-interval Kalman smoother) are carried out. The conclusions are limited to the particular toy problem considered, but it can be seen how more elaborate experiments could be carried out, as well as how the method might be applied in practice. The authors have considered five adjustable parameters, two related to a distributed coefficient in the equivalent barotropic vorticity equation (“physical” parameters), one governing the relative weight given to observations versus forecast, one governing the relative weight given to observations versus goodness of fit to the model (in the weak constraint case), and one governing the damping of high-frequency oscillations in the analysis at the final time point (“smoothing” parameter). The weighting parameters and the smoothing parameter can, if desired, be interpreted as ratios of parameters in prior covariances. Analyses are made with a low-resolution model of the dynamics of the equivalent barotropic vorticity equation given noisy forecast (initial conditions) and noisy wind observations, and compared with nature evolved from exact initial conditions using a high-resolution forward integration. The authors found that these five (carefully chosen) parameters are simultaneously tunable on line, that is, simultaneously with the analysis, and 1) that the analysis is *equally* and strongly sensitive to both the choice of the observed versus forecast weighting parameter and the choice of the smoothing parameter; 2) that the analysis with the model as a weak constraint, based on the tuned estimate of the parameter governing how close the analysis satisfies the model, is somewhat better than the analysis using the model as a strong constraint, although estimation of this tuning parameter varies much more than the other parameters with replications of the experiment; and 3) good estimates of the physical parameters are obtained; however, these estimates are closer to those that make the model integrated forward with perfect initial conditions best fit nature, and these are not exactly the “true” parameters.

### 1. Introduction

We consider the adaptive tuning of a “toy” four-dimensional variational data assimilation problem, which contains terms that are toy versions of terms included in operational numerical weather prediction problems and reanalysis problems. The purpose of this

paper is to demonstrate the feasibility and effectiveness of quasi-on-line simultaneous adaptive tuning of multiple weighting, smoothing, and physical parameters in such a variational problem, via the use of randomized trace versions of generalized cross validation (GCV) and unbiased risk (UBR) methods for the tuning. The toy variational problem we study is of the form of the “standard” variational numerical weather prediction problem (Lorenz 1986; Parrish and Derber 1992; Wahba 1982a) as extended to time as well as space with the model as a weak constraint, and we also examine the variational problem with the model as a strong constraint as a toy version of 4D-Var as discussed in, for example, Rabier

---

*Corresponding author address:* Prof. Grace Wahba, Department of Statistics, University of Wisconsin, 1210 W. Dayton St., Madison, WI 53706.  
E-mail: wahba@stat.wisc.edu

et al. (1993) and Zou et al. (1993). Under some assumptions the solution of the 4D-Var problem with the model as a weak or a strong constraint can be shown to be equivalent to a Kalman smoother with model error or without model error; see Thépaut et al. (1993) and Cohn et al. (1994). We include terms that govern the fit of the analysis to the data and to the forecast, and a “smoothness” (read “balance”) penalty term. In the weak constraint case we also include a penalty term that governs the closeness of the analysis to the model. The strong constraint case here forces the analysis to fit some model trajectory and can be considered as a limiting case of the model as a weak constraint. In practice these terms include numerous weighting and smoothing parameters, as well as coefficients in the model parameterizations, calibration coefficients, scale lengths, spatially varying terms in covariances, and so forth.

In the usual 3D-Var or OI analysis, the weight given to the data is governed by an assumed observation error covariance matrix, which may include errors of representativeness and errors in any forward model involved, as well as instrumental errors. The components of this matrix are frequently obtained from known instrument characteristics and other system knowledge. The weight given to the forecast is governed by a forecast error covariance, which is in practice generally obtained from historical information about the system (see, e.g., Hollingsworth and Lonnberg 1986; Bartello and Mitchell 1992; Parrish and Derber 1992). Kalman filter (KF) theory tells how in theory the forecast error covariance is propagated along with the forecast, but in practice propagating a complete KF forecast error covariance is not practical. Recently several authors have proposed simplified versions of the forecast error covariance prescribed by KF theory and then have derived or estimated a small number of coefficients in this covariance; see, for example, Cohn (1993) and Todling and Cohn (1994) and references therein. Model error covariances are in theory part of the KF equations but are not easy to specify realistically. Recent work in the development of model error covariances includes Daley (1992) and Mitchell and Daley (1997a,b). In particular Mitchell and Daley note that correlation of the model error and the signal can be important, a problem that specifically exists but is ignored in the present work. Other discussions of model errors include Boer (1984), Dalcher and Kalnay (1987), Tribbia and Baumhefner (1988), and Dee (1995). In addition, model errors may well be correlated from time to time, in violation of the usual KF assumptions.

In the present work, and its predecessor Wahba et al. (1995, hereafter WJGG), it is our thesis that, under certain conditions, it may be both feasible and beneficial to attempt adaptive on-line (or quasi-on-line) tuning or fine-tuning of a *relatively small number of* those parameters, *to which the analyses are especially sensitive*. By quasi-on-line we mean, as part of the variational problem for obtaining the analysis. This work is parallel to

the pioneering work of Dee (1995), who uses maximum likelihood methods for this purpose. In addition, the analysis can be sensitive to inadequate specification of physical parameters in the model, and if this is the case, there is also the potential for quasi-on-line tuning of some of them. See Wergen (1992), O’Sullivan (1991), Wahba (1990a), and Kravaris and Seinfeld (1985). This has led us to examine the possibility of simultaneously tuning or fine-tuning physical parameters on line.

In this paper we selected an equivalent barotropic vorticity equation for the streamfunction on a latitude circle and evolved “nature” on a fine time and space grid via a leapfrog scheme for integrating the vorticity equation, while we defined the model dynamics on a coarse subgrid via a first-order finite-difference scheme. The barotropic vorticity equation we use has two physical parameters, and we set these to be unknowns to be tuned in the model dynamics. We simulated a forecast for the streamfunction, by adding (correlated) random errors to the nature streamfunction at the start, and wind observations by adding (white) random error to the nature wind at selected points in time and space. We attempted to scale the experiment to realistic values of the various error variances. Given the simulated forecast and observations, we then tune and solve a variational problem in time and space, using the model dynamics both as a weak and then a strong constraint. We compute the GCV and UBR functions that are minimized to tune the model, via the randomized trace technique described in detail in WJGG. As discussed in WJGG, the randomized trace method may be implemented on any size problem that has a working code for solving the variational problem, at the cost of solving the variational problem several times. Thus it is potentially feasible with operational-sized variational problems, including mildly nonquadratic ones solved in a limited number of iterations.

The parameters we chose to tune were selected carefully in that we had reason to believe a priori that the estimates of the state vector and then the predicted observations were sensitive to them. This is certainly a minimum requirement for the parameters to be tuned this way. Using the fact that we know nature in this experiment, we can quantify the sensitivity fairly precisely. We find that for these five sensitive parameters the methods can tune the variational problem very well. Let  $\theta$  be the vector of parameters to be tuned and  $R(\theta)$  be the predictive mean square error, that is, the mean square difference between nature and the analysis in observation space, when  $\theta$  is used, and let  $\hat{\theta}$  be the estimate of the optimal  $\theta$  via GCV or UBR. We find that the inefficiency, which is defined as  $R^{1/2}(\hat{\theta})/\min_{\theta} R^{1/2}(\theta)$  (see Craven and Wahba 1979) is mostly in the range of 1.05–1.2. (Note that  $R$  can be computed only in simulation studies where nature is known.) In the case of the model as a weak constraint, we have formulated the weak model constraint in the simplest possible terms, namely as penalty on the ordinary sum

of squares differences, even though the model error is surely correlated in space and time. That correlation would in theory dictate a more sophisticated penalty. In this experiment the analysis is less sensitive to the parameter that controls the relative fit of the analysis to the model than to the other comparable parameters, at least in the region of parameter space near the optimum parameters. Nevertheless some tuning appears to be beneficial. The relative lack of sensitivity to this parameter may be related to the fact that the form of the lack of fit to model penalty is appropriate to model errors independent in time and space, while model errors surely are not. At the scale of our toy problem, we found that the model as a (tuned) weak constraint is slightly better than as a strong constraint. The parameter controlling the relative weight given to forecast versus observations can be interpreted as a variance ratio, and the tuned value was a reasonable estimate of the ratio of the variances used in the simulation of forecast error and observation error. A smoothness penalty (read “balance”) at the last time step was definitely tunable and beneficial. In fact the sensitivity to this parameter was roughly similar to that of the parameter controlling the relative fit to observations versus forecast. We do not know to what extent numerical results in our toy model scale up to more realistic settings, but we think that the experimental design demonstrates an approach that has the potential for answering the scale-up question and also suggests directions for how the methods might be implemented in practice. Theoretical results concerning sensitivity will appear separately.

In section 2 we review 4D-Var and KF theory that is behind the variational problem that we use. In section 3 we describe the experimental design and the calculations. Section 4 describes the experimental results, section 5 discusses some related issues, and section 6 gives a summary, conclusions, and some directions for future research.

## 2. Tuning the 4D-Var problem

We briefly review the (usual) assumptions from Kalman filtering and 4D-Var, organized in a form corresponding to our experiment. Let  $t = 1, \dots, T$  denote discrete time and let  $\Psi_t, t = 1, \dots, T$  be a sequence of state vectors representing (some part of) nature that evolves according to

$$\Psi_{t+1} = \mathbf{M}_t \Psi_t + \mathbf{N}_t + \xi_t, \quad t = 1, \dots, T-1, \quad (2.1)$$

where  $\mathbf{M}_t$  is the model evolution operator,  $\mathbf{N}_t$  is a forcing function, and the  $\xi_t$  are assumed to be independent, zero mean Gaussian random vectors with covariances  $\sigma_m^2 \mathbf{Q}_t$ ,  $t = 1, \dots, T$ . Here,  $\Psi_*$  is the forecast for  $t = 1$ , assumed to satisfy

$$\Psi_* = \Psi_1 + \epsilon_* \quad (2.2)$$

with  $\epsilon_* \sim \mathcal{N}(0, \sigma_f^2 \mathbf{Q}_*)$ .<sup>1</sup> The usual linear Kalman filter theory (in the present notation) would evolve the forecast error covariance as  $\sigma_f^2 \mathbf{Q}_* = \mathbf{M}_0 \mathbf{P}_0^a \mathbf{M}_0^t + \sigma_m^2 \mathbf{Q}_0$ , where  $\mathbf{M}_0$  and  $\sigma_m^2 \mathbf{Q}_0$  are the model evolution operator and model error covariance for the time step preceding  $t = 1$  and  $\mathbf{P}_0^a$  is the analysis error covariance at the preceding time step. However, in this work we will deal with  $\sigma_f^2 \mathbf{Q}_*$  directly, although other options based on an evolved covariance are possible; see Cohn (1993) and Dee (1995) and references cited therein. The observations  $\mathbf{y}_t$  are assumed to satisfy

$$\mathbf{y}_t = \mathbf{K}_t \Psi_t + \epsilon_t, \quad t \in \Lambda, \quad (2.3)$$

where  $\mathbf{K}_t$  is a map from state vector space to observation space at time  $t \in \Lambda$  and  $\Lambda$  is the subset of  $\{1, \dots, T\}$  where there are observations. The  $\epsilon_t$  are assumed to be independent in time,  $\epsilon_t \sim \mathcal{N}(0, \sigma_o^2 \mathbf{S}_t)$ ,  $t \in \Lambda \subset \{1, \dots, T\}$ .

The joint density of  $\{\mathbf{y}_t, t \in \Lambda, \Psi_*, \Psi_1, \dots, \Psi_T\}$  may then be written

$$\begin{aligned} p(\{\mathbf{y}_t, t \in \Lambda\}, \Psi_*, \Psi_1, \dots, \Psi_T) \\ = p(\{\mathbf{y}_t, t \in \Lambda\}, \Psi_* | \Psi_1, \dots, \Psi_T) \\ \times p(\Psi_1, \dots, \Psi_T). \end{aligned} \quad (2.4)$$

Provided the model operator  $\mathbf{M}_t$  is invertible, we may write

$$\begin{aligned} p(\Psi_1, \dots, \Psi_T) \\ = p(\Psi_1) p(\Psi_2 | \Psi_1) p(\Psi_3 | \Psi_2) \cdots p(\Psi_T | \Psi_{T-1}) \\ = p(\Psi_T) (\Psi_{T-1} | \Psi_T) p(\Psi_{T-2} | \Psi_{T-1}) \cdots p(\Psi_1 | \Psi_2). \end{aligned} \quad (2.5)$$

If we specify  $p(\Psi_T)$ , the joint density on the left of (2.4) is completely specified. We will allow an improper Gaussian prior for  $\Psi_T$ , which will allow the imposition of a penalty term on  $\Psi_T$  that is not necessarily of full rank. Practically speaking this amounts to treating  $\Psi_T$  as though it had the prior covariance  $b\mathbf{\Sigma}$  with  $\mathbf{\Sigma}^{-1} = \mathbf{J}$  where  $\mathbf{J}$  is not of full rank, and this entails penalizing only part of  $\Psi_T$ . See Wahba (1990b, section 1.5) for further discussion of this point.

Letting  $\|\mathbf{y}\|_b^2 = \mathbf{y}' \mathbf{B} \mathbf{y}$ , then the conditional expectation of  $\Psi_1, \dots, \Psi_T$  given  $\{\mathbf{y}_t, t \in \Lambda, \Psi_*\}$  is the minimizer, call it  $\Psi = (\Psi'_1, \dots, \Psi'_T)$ , of

$$\begin{aligned} \frac{1}{\sigma_o^2} \sum_{t \in \Lambda} \|\mathbf{y}_t - \mathbf{K}_t \Psi_t\|_{\mathbf{S}_t}^2 \\ + \frac{1}{\sigma_m^2} \sum_{t=1}^{T-1} \|\Psi_{t+1} - \mathbf{M}_t \Psi_t - \mathbf{N}_t\|_{\mathbf{Q}_t}^2 \\ + \frac{1}{\sigma_f^2} \|\Psi_* - \Psi_1\|_{\mathbf{Q}_*}^2 + \frac{1}{b} \|\Psi_T\|_{\mathbf{J}}^2. \end{aligned} \quad (2.6)$$

<sup>1</sup> This notation means that  $\epsilon_*$  is a Gaussian random vector with zero mean and covariance matrix  $\sigma_f^2 \mathbf{Q}_*$ .

Note that  $\Psi_t$  for  $t = T$  has been singled out for special treatment in the last term. In theory, we could have used any  $t = 1, 2, \dots, T$ . See also Bennett and Miller (1991), who discuss the necessity of a term like  $b^{-1}\|\Psi_T\|_3^2$ . Letting  $\gamma = \sigma_o^2/\sigma_f^2$ ,  $\alpha = \sigma_o^2/\sigma_m^2$ , and  $\eta = \sigma_o^2/b$ , then the minimizer of (2.6) is the same as the minimizer of

$$\sum_{t \in \Lambda} \|\mathbf{y}_t - \mathbf{K}_t \Psi_t\|_{\hat{\mathbf{S}}_t}^2 + \alpha \sum_{t=1}^{T-1} \|\Psi_{t+1} - \mathbf{M}_t \Psi_t - \mathbf{N}_t\|_{\hat{\mathbf{Q}}_t}^2 + \gamma \|\Psi_* - \Psi_1\|_{\hat{\mathbf{Q}}_*}^2 + \eta \|\Psi_T\|_3^2. \quad (2.7)$$

Letting  $\alpha \rightarrow \infty$  is equivalent to the ‘‘perfect model’’ assumption (or constraint). In that case (2.7) becomes the following. Find  $\Psi_1$  to minimize

$$\sum_{t \in \Lambda} \|\mathbf{y}_t - \mathbf{K}_t \Psi_t\|_{\hat{\mathbf{S}}_t}^2 + \gamma \|\Psi_* - \Psi_1\|_{\hat{\mathbf{Q}}_*}^2 + \eta \|\Psi_T\|_3^2, \quad (2.8)$$

where the  $\Psi_t$ ,  $t = 2, \dots, T$  are constrained to satisfy

$$\Psi_t = \mathbf{M}_{t-1} \Psi_{t-1} + \mathbf{N}_{t-1}. \quad (2.9)$$

Except (possibly) for the inclusion of the penalty term  $\eta \|\Psi_T\|_3^2$ , (2.7), (2.8), and (2.9) are the basis for the ‘‘standard’’ three- and four-dimensional variational data assimilation methods and also correspond to the fixed-lag Kalman smoother. See Lorenc (1986), Zou et al. (1992), Bennett (1992), Rabier et al. (1993), Thepaut et al. (1993), Parrish and Derber (1992), and Cohn et al. (1994). Terms that penalize large gravity waves, lack of balance, or other nonmeteorological phenomena are in fact frequently included in operational data assimilation models.

In what follows, we will assume that  $\mathbf{S}_t$  is known reasonably well, and, hence, the variational problem can be rescaled so that we can set  $\mathbf{S}_t = \mathbf{I}$ . In order to use the generalized cross-validation and unbiased risk methods here, it is necessary to have a large dataset<sup>2</sup> whose error structure is close to white or at least well known. In practice this assumption is reasonably well satisfied when the data includes a large subset from, for example, radiosondes, and correlation in the so-called errors of representativeness can be neglected. Highly positively correlated errors that are not accounted for may cause the cross validation to confuse noise with signal.

We will be considering the adaptive estimation of the weighting parameters  $\gamma$  and (in the weak constraint case)  $\alpha$ , the smoothing parameter  $\eta$ , as well as two ‘‘physical’’ parameters  $U_0$  and  $\delta$ , which are part of a distributed parameter in  $\mathbf{M}_t$ , to be described later. Let  $\theta$  stand collectively for  $\theta = (U_0, \delta, \eta, \alpha, \gamma)$  or, if  $\alpha = \infty$  (the strong constraint case),  $\theta = (U_0, \delta, \eta, \gamma)$ . Letting  $\hat{\mathbf{y}}_t = \mathbf{K}_t \Psi_t$ ,  $\hat{\mathbf{y}} = (\hat{\mathbf{y}}_1, \dots, \hat{\mathbf{y}}_T)'$ , then, either in the weak constraint case (2.7) or the strong constraint case (2.8) and (2.9), there exists a matrix  $\mathbf{A}(\theta)$ , known as the influence matrix, such that

$$\hat{\mathbf{y}} = \mathbf{A}(\theta)\mathbf{y} + \text{quantities independent of } \mathbf{y}. \quad (2.10)$$

The GCV estimate of  $\theta$  is given by the minimizer of  $V(\theta)$ , where

$$V(\theta) = \frac{n_{\text{dat}}^{-1} \text{RSS}(\theta)}{\{n_{\text{dat}}^{-1} \text{Tr}[\mathbf{I} - \mathbf{A}(\theta)]\}^2}, \quad (2.11)$$

where  $n_{\text{dat}}$  is the number of data points (dimension of  $\mathbf{y}$ ) and  $\text{RSS}(\theta) = \|\mathbf{y} - \hat{\mathbf{y}}\|^2$ . Here and elsewhere, if there is no subscript on  $\|\cdot\|$ , then the Euclidean norm is intended. The UBR estimate of  $\theta$ , which may be used when  $\sigma_o^2$  is known, is given by the minimizer of  $U(\theta)$  defined by

$$U(\theta) = \frac{1}{n_{\text{dat}}} \text{RSS}(\theta) + 2\sigma_o^2 \frac{1}{n_{\text{dat}}} \text{Tr}\mathbf{A}(\theta). \quad (2.12)$$

Letting  $\Psi_t^{\text{TRUE}}$  be the ‘‘true’’ but unknown  $\Psi_t$ , both  $U(\theta)$  and  $V(\theta)$  are, under suitable assumptions, known to be proxies for the predictive mean square error (PMSE), given by  $R(\theta)$  where

$$R(\theta) = \frac{1}{n_{\text{dat}}} \sum_{t \in \Lambda} \|\mathbf{K}_t \Psi_t^{\text{TRUE}} - \mathbf{K}_t \hat{\Psi}_t(\theta)\|^2. \quad (2.13)$$

By proxies for  $R(\theta)$ , here it is meant that the GCV and UBR estimates are good estimates of the minimizer of  $R(\theta)$  in the sense that their inefficiencies, as measured by  $R^{1/2}(\hat{\theta})/\min_{\theta} R^{1/2}(\theta)$ , where  $\hat{\theta}$  is either the minimizer of  $U$  or of  $V$ , tend to 1 as the sample size becomes large, under fairly general assumptions. WJGG has several examples that demonstrate the properties of these estimates, and appendices A and B there give a brief discussion of the mathematical conditions sufficient to ensure that these estimates are good proxies for  $R$ . The GCV estimate, proposed in Craven and Wahba (1979) and Golub et al. (1979), was first used in an atmospheric sciences context in Wahba and Wendelberger (1980) and is discussed in the books of Golub and van Loan (1989), Eubank (1988), Wahba (1990b), Hastie and Tibshirani (1990), Green and Silverman (1994), and elsewhere. Rigorous theoretical mathematical justification of these estimates has been provided by Li (1986).

### 3. The experiment

#### a. The equivalent barotropic vorticity equation on a latitude circle

The equivalent barotropic vorticity equation that we used is

$$\frac{\partial}{\partial t}(\Psi_{xx} - \lambda^2 \Psi) + U(x)\Psi_{xxx} + \beta \Psi_x = -f_0 U_x, \quad (3.1)$$

where  $x$  is the space variable, on the latitude circle at  $\Phi = 45^\circ$ ;  $\Psi$  is the streamfunction;  $f_0 = 2\Omega \sin\Phi$  is the Coriolis parameter;  $\beta = 2\Omega \cos\Phi/a$ , the meridional derivative, and  $a$  is the radius of the earth;  $\lambda^2 = f_0^2/gH$ ; and  $U(x) = U_0[1 + \delta g(x)]$ , where  $U_0$  and  $\delta$  are ‘‘distributed’’ parameters to be estimated, and  $g(x)$  is a fixed

<sup>2</sup> This may be a subset of the entire dataset; see Wahba et al. (1994).

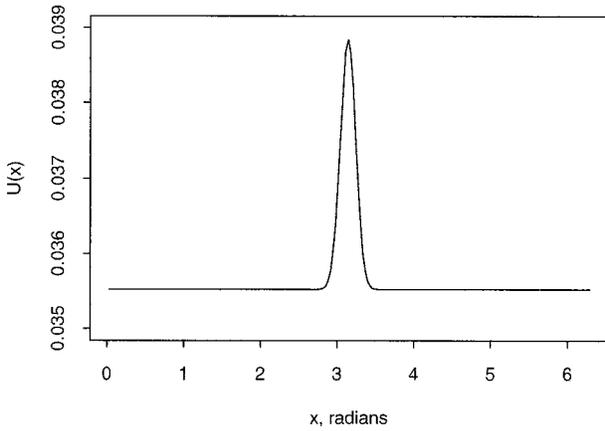


FIG. 1. The (true) distributed parameter  $U(x)$ , in dimensionless form.

perturbation function. Taking  $H \approx 10^3$  m,  $g \approx 10$  m s<sup>-2</sup> results in  $f_0 \approx 10^{-4}$  s<sup>-1</sup>,  $\beta \approx 1.57 \times 10^{-11}$  s<sup>-1</sup> m<sup>-1</sup>, and  $\lambda^2 \approx 10^{-12}$  m<sup>-2</sup>. For numerical calculations a dimensionless form of (3.1) was used via the transformations

$$\left\{ \begin{array}{l} t \rightarrow f_0 t \\ x \rightarrow L^{-1} x \\ \Psi(x, t) \rightarrow \frac{1}{L^2 f_0} \Psi(x, t) \\ U(x) \rightarrow \frac{1}{L f_0} U(x) \\ \beta \rightarrow \frac{L}{f_0} \beta \\ \lambda^2 \rightarrow L^2 \lambda^2 \\ \frac{\partial \Psi}{\partial x} \rightarrow \frac{1}{L f_0} \frac{\partial \Psi}{\partial x} \end{array} \right\}, \quad (3.2)$$

where  $L = 4496$  km (radius of the earth times  $\cos 45^\circ$ ). The result is

$$\frac{\partial}{\partial t} (\Psi_{xx} - \lambda^2 \Psi) + U(x) \Psi_{xxx} + \beta \Psi_x = -U_x, \quad (3.3)$$

where the parameters above become  $\beta = \cos 45^\circ \approx 0.707$ ;  $\lambda^2 \approx 20$ . Here,  $U(x)$  is given in dimensionless form in Fig. 1,  $U_0 = 0.0355$ ,  $\delta = 0.10$ .

b. “Nature” and “the model”

In reality the atmosphere (“nature”) exists in continuous space and time, evolving according to laws of physics at all scales, and on which observations, consisting of (noise-contaminated) values of functionals, which may involve derivatives, point values, or weighted integrals over various-sized regions, are available. The computer model, discrete in time and space, is only an approximation to nature at the cruder scales, and the model observation operator  $\mathbf{K}_t$ , which is a map from

state space to observation space, is only an approximation to the real observation operator, which is a map from nature space to observation space. In the present work, we will avoid the (important) issues of the best ways to go from continuous time and space to discrete time and space<sup>3</sup> by letting our proxy for nature be an atmosphere and an evolution operator defined on a fine time and space grid (the “nature” grid), while the model evolution operator  $\mathbf{M}_t$  and state vector  $\Psi_t$  are defined on a crude time and space grid (the “model” grid), which is a subgrid of the nature grid. Similarly, the nature observation functionals will be defined by an observation operator from the nature grid to observation space, while the model observation operator  $\mathbf{K}_t$  will be defined from the model grid to observation space. The model error in atmospheric and oceanic dynamical systems models has to do with the fact that the objects that in reality play the role of the nature grid and nature evolution operator are not the same as those that play the role of the model grid and the model evolution operator. Thus the assumption that the  $\xi_t$  in the Kalman filter model described above behave as though they are independent from time step to time step (either in the present work or in reality) is probably the least tenable assumption made here (and elsewhere). Nevertheless, the description has been found to be useful in some contexts.

The nature grid for our experiment is a fine grid with space points approximately 21 km apart and time grid points 8.7 s apart. The nature streamfunction at the start (0 h or model time  $t = 1$ ) for our primary experiment is defined by discretizing  $\Psi_{\text{start}}(x)$  given by

$$\Psi_{\text{start}}(x) = 0.002(\sin x + 0.5 \cos 2x + 0.6 \sin 3x), \quad x \in [0, 2\pi] \quad (3.4)$$

on the nature space grid. The nature streamfunction at 0 h appears in the upper left-hand corner of Fig. 3. A secondary experiment used a stronger “signal” obtained by multiplying  $\Psi_{\text{start}}$  in (3.4) by 6.28. The nature dynamics are defined via a leapfrog scheme on the nature grid for the numerical solution of (3.3), initiated by a first-order difference scheme, and using the initial state determined by (3.4) or Eq. (3.4)  $\times$  6.28. Nature was evolved for 48 h. The nature streamfunction is then known on the nature grid, and the nature wind field is defined as the central divided difference of the nature streamfunction on the nature space grid. The model uses as its state vector values of the streamfunction on the coarse model grid of 194 points approximately 146 km apart in space (every seventh nature space grid point), with a time step of 4 h (=1667 nature time steps). The model wind vector is defined as the central divided difference of the model streamfunction on the model space grid. The model dynamics was represented by a first-

<sup>3</sup> This issue is discussed in Wahba (1990b) and Bennett (1992).

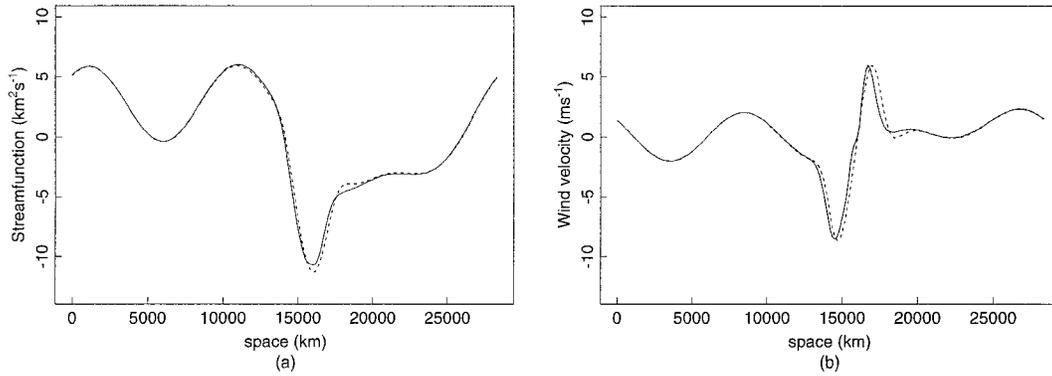


FIG. 2. Nature and model streamfunction (a) and nature and model wind (b) at 48 h (model time  $t = 13$ ), based on discretizing  $\Psi_{\text{start}}$  on the nature and model grids, respectively, and integrating forward via the nature and model evolution operators, respectively. Dotted lines are nature and solid lines are model.

order forward difference scheme to integrate (3.3) on the model grid. Let  $\tilde{\Delta}_1$ ,  $\tilde{\Delta}_2$ , and  $\tilde{\Delta}_3$  stand for the first-, second-, and third-order circulant difference matrices given, respectively, by

$$\begin{pmatrix} -1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & -1 & 1 \\ 1 & 0 & 0 & \cdots & 0 & -1 \end{pmatrix}, \begin{pmatrix} -2 & 1 & 0 & \cdots & 1 \\ 1 & -2 & 1 & \cdots & 0 \\ 0 & 1 & -2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 1 & 0 & \cdots & 1 & -2 \end{pmatrix},$$

$$\begin{pmatrix} -3 & 1 & 0 & 0 & \cdots & -1 & 3 \\ 3 & -3 & 1 & 0 & \cdots & 0 & -1 \\ -1 & 3 & -3 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots & \vdots \\ 1 & 0 & 0 & \cdots & -1 & 3 & -3 \end{pmatrix},$$

and let  $\Delta_k = \tilde{\Delta}_k(\Delta x)^{-k}$ ,  $k = 1, 2, 3$ , where  $\Delta x$  is the distance between two model space grid points. Let  $x_i$  be the  $i$ th model grid point and  $\mathbf{G}$  be the diagonal matrix with  $g(x_i)$  in the  $i$ th position. The vector form of this forward difference scheme is then

$$\Psi_{t+1} = [\mathbf{I} - \mathbf{B}^{-1}(\mathbf{C} + \mathbf{D})]\Psi_t - \mathbf{B}^{-1}\mathbf{F}_t, \quad (3.5)$$

where

$$\mathbf{B} = (\Delta_2 + \lambda^2\mathbf{I})\Delta x^2$$

$$\mathbf{C} = U_0(\mathbf{I} + \delta\mathbf{G})\Delta_3\Delta x^2\Delta t$$

$$\mathbf{D} = \beta\Delta_1\Delta x^2\Delta t,$$

$$\mathbf{F}_t = \begin{pmatrix} \frac{\partial U(x_1)}{\partial x} \\ \frac{\partial U(x_2)}{\partial x} \\ \vdots \\ \frac{\partial U(x_n)}{\partial x} \end{pmatrix} \Delta x^2\Delta t \doteq U_0\delta \begin{pmatrix} \frac{g(x_2) - g(x_1)}{\Delta x} \\ \frac{g(x_3) - g(x_2)}{\Delta x} \\ \cdots \\ \frac{g(x_1) - g(x_n)}{\Delta x} \end{pmatrix} \Delta x^2\Delta t,$$

$$t = 1, \dots, T - 1,$$

where  $\Delta t$  is the time between two model time steps. We attempted to choose the nature and model grids and dynamics so that the differences between the evolution of nature and evolution defined by the model were relatively small but not negligible compared to observation errors. The dotted line in Fig. 2a gives the nature streamfunction at 48 h (model time  $t = T = 13$ ) as was determined by evolving  $\Psi_{\text{start}}(x)$  of (3.4) via the nature dynamics on the nature grid, and the solid line is the streamfunction at the same time as obtained by evolving  $\Psi_{\text{start}}$  via the model dynamics on the model grid. Figure 2b gives the corresponding wind fields computed as the central divided difference of the nature and model streamfunctions, respectively, on the nature and model grids. The root-mean-square difference between the nature and model streamfunctions, averaged over the model grid points at the last time step, was about  $0.2 \text{ km}^2 \text{ s}^{-1}$ , and the root-mean-square difference in the wind fields on the model grid points was about  $0.37 \text{ m s}^{-1}$ .

*c. The observations*

The wind observations were generated at model times  $t \in \Lambda = \{1, 4, 7, 10, 13\}$  (corresponding to 0, 12, 24, 36, and 48 h) on a regular space grid with a hole in it, specifically on the 164 model grid points numbered 1–30 and 61–194. This was done by first computing the nature wind at these points in space and time, then adding observation errors to the nature wind as independent zero mean Gaussian random variables with a common standard deviation. Figure 3 gives the nature streamfunction (evolved from  $\Psi_{\text{start}}$ ) and wind field at the observation times, with observed winds superimposed on the nature wind plot. These wind observations were generated with a standard deviation of  $\sigma_o^2 = 2 \text{ m s}^{-1}$  in the wind observation error.

*d. The forecast*

The forecast used by the model in our experiment was obtained by adding a 194-dimensional forecast

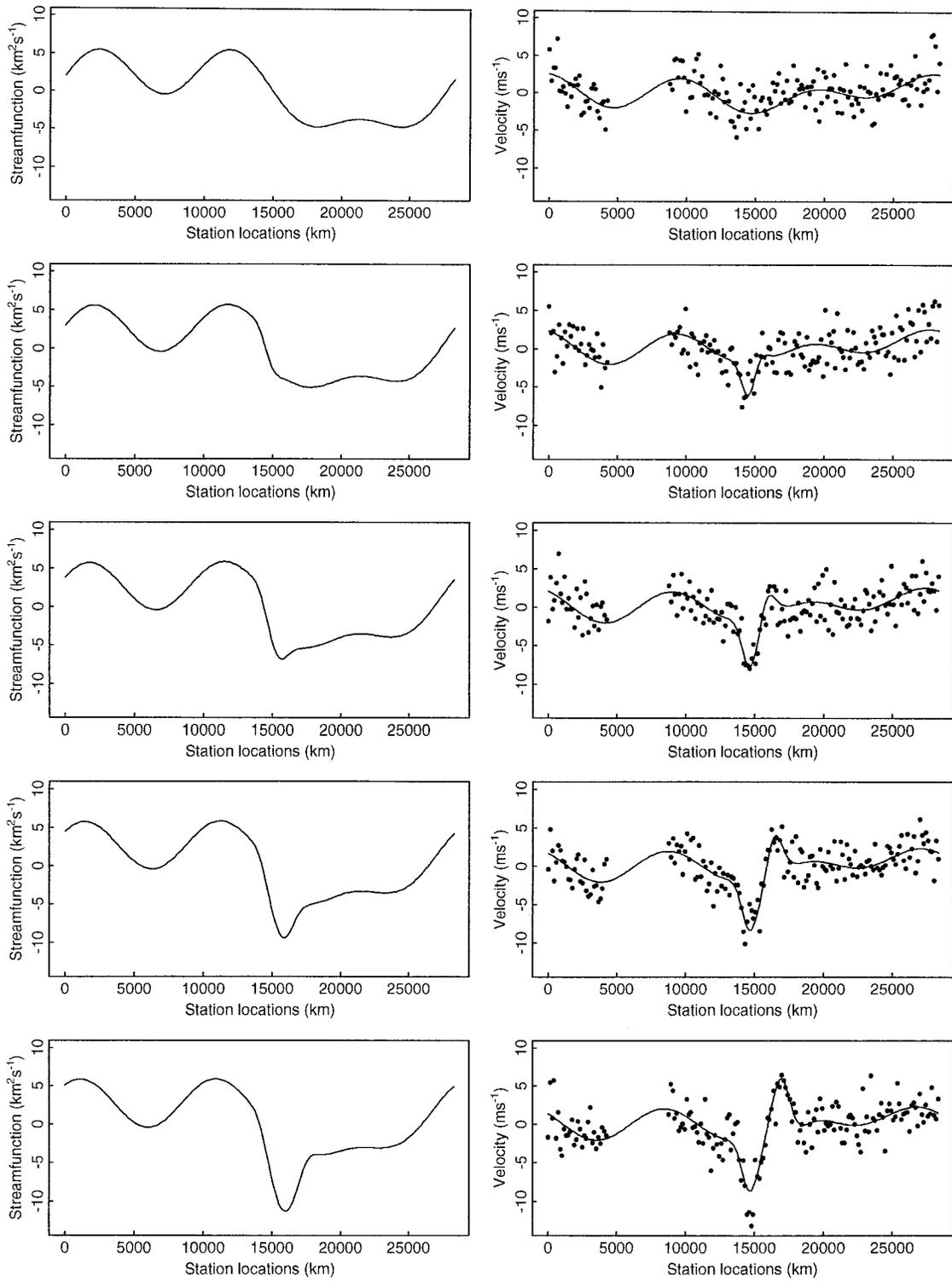


FIG. 3. Nature streamfunction at the observation times  $t = 1, 4, 7, 10, 13$  (left column) and corresponding nature wind and wind observations (right column).

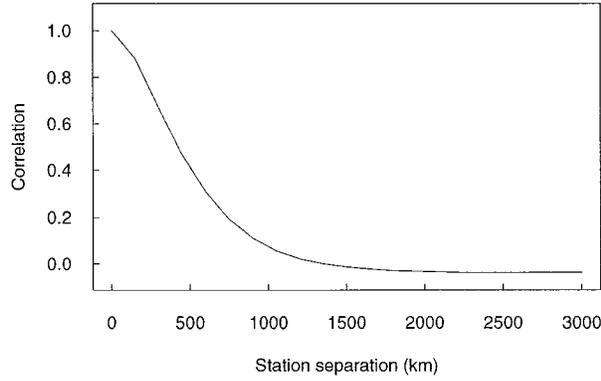


FIG. 4. Wind field forecast error correlation function  $\rho(r)$ , chosen to resemble Hollingsworth and Lonnberg (1986, Fig. 5).

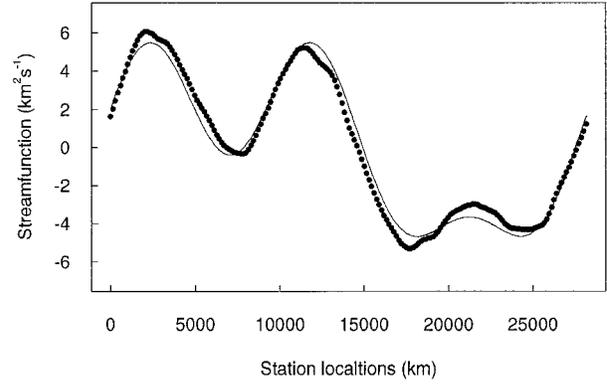


FIG. 5. Nature streamfunction (solid line) and noisy streamfunction forecast (dots) at  $t = 1$ .

error vector  $\boldsymbol{\epsilon}_{\Psi_f}$  to the vector of true streamfunctions at the initial time obtained by discretizing  $\Psi_{\text{start}}(x)$  on the model space grid. Here,  $\boldsymbol{\epsilon}_{\Psi_f}$  was generated as a zero mean Gaussian random vector with covariance  $\sigma_f^2 \mathbf{Q}_*$ , where  $\mathbf{Q}_*$  is a circulant (isotropic) correlation matrix to be described. Since the wind field is defined as the central divided difference  $\Delta_c$  of the streamfunction, then  $\boldsymbol{\epsilon}_{\Psi_f}$  and  $\boldsymbol{\epsilon}_{V_f}$ , the wind field forecast error, are related by  $\boldsymbol{\epsilon}_{V_f} = \Delta_c \boldsymbol{\epsilon}_{\Psi_f}$ . Figure 4 gives the wind field forecast error correlation function  $\rho(r)$ , which corresponds to the streamfunction correlation that we used. Here,  $\rho$  was chosen to be visually indistinguishable from the longitudinal wind field forecast error correlation implied in Fig. 5 of Hollingsworth and Lonnberg (1986). The wind field forecast error covariance is then  $\sigma_{V_f}^2 \mathbf{C}_{V_f}$ , where  $\sigma_{V_f}$  is to be specified and the  $j, k$ th entry of  $\mathbf{C}_{V_f}$  is  $\rho(|j - k| \times 150 \text{ km})$ . The  $\sigma_f^2$  and  $\mathbf{Q}_*$  that we used satisfy  $\Delta_c(\sigma_f^2 \mathbf{Q}_*)\Delta_c' = \sigma_{V_f}^2 \mathbf{C}_{V_f}$ . Since  $\Delta_c$  is not of full rank, this equation does not always have a solution, and when it does, it is not unique. We did a small amount of tinkering [in Fourier transform space; see Wahba (1968)] with both  $\mathbf{Q}_*$  and  $\rho$  so that both the equation is satisfied and  $\mathbf{Q}_*$  is of full rank for numerical stability. Figure 5 gives a plot of the  $\Psi_{\text{start}}(x)$  along with a noisy streamfunction forecast obtained by adding a random  $\boldsymbol{\epsilon}_{\Psi_f}$ . For this figure,  $\sigma_{V_f}$  was taken as  $0.485 \text{ m s}^{-1}$ , which resulted in  $\sigma_f = 0.726 \text{ km}^2 \text{ s}^{-1}$ . Note the evidence of spatial correlation.

#### e. Calculations for the experiment

We now give details for the calculation of the minimizers of (2.7) and (2.8), subject to (2.9):  $\Psi_t$  is the 194-dimensional model state vector of streamfunction values at time  $t = 1, \dots, 13$ , and  $\mathbf{y}_t$  is the 164-dimensional vector of wind observations at  $t \in \Lambda = 1, 4, 7, 10, 13$ . Here,  $\mathbf{M}_t = \mathbf{M}(U_0, \delta) = \mathbf{M}$  independent of  $t$  is obtained from (3.5) as  $\mathbf{M} = [\mathbf{I} - \mathbf{B}^{-1}(\mathbf{C} + \mathbf{D})]$ , and  $\mathbf{N}_t = -\mathbf{B}^{-1}\mathbf{F}_t = \mathbf{N}(U_0, \delta) = \mathbf{N}$ , independent of  $t$ , where  $\mathbf{F}_t$  is given in (3.5). In this experiment we set  $\mathbf{Q}_t = \mathbf{I}$ . This

is partly because methods for modeling the model error are beyond the scope of this paper. But also, we are particularly interested in seeing what might happen when a misspecified “model error covariance” is used, something that is likely to occur in practice. Extension of the formulas below to the general  $\mathbf{Q}_t$  case is straightforward, however, and they could be extended to the case where model error is, for example, described as a low-order autoregressive or moving average scheme on the model time grid. See Bennett et al. (1996) and Bennett et al. (1997), who employ an autoregressive scheme for model error. We set  $\mathbf{J}$  to correspond to a penalty on the sum of squares of the second differences of the streamfunction values; that is,

$$\mathbf{J} = \Delta_2' \Delta_2. \quad (3.6)$$

Let  $\mathbf{y} = (\mathbf{y}'_1, \mathbf{y}'_4, \mathbf{y}'_7, \mathbf{y}'_{10}, \mathbf{y}'_{13})'$ ,  $\Psi = (\Psi'_1, \dots, \Psi'_{13})'$ , and let  $\mathbf{P}$  be the  $164 \times 194$  submatrix of the identity matrix obtained by deleting columns 31–60 (corresponding to no observations). Then  $\mathbf{K}_t = \mathbf{P}\Delta_c$ . Let  $\mathbf{K}$  be the  $(5 \times 164) \times (13 \times 194)$  matrix defined by  $\sum_{t \in \Lambda} \|\mathbf{y}_t - \mathbf{P}\Delta_c \Psi_t\|^2 = \|\mathbf{y} - \mathbf{K}\Psi\|^2$ . The functional to be minimized becomes

$$\|\mathbf{y} - \mathbf{K}\Psi\|^2 + \alpha \sum_{t=1}^{12} \|\Psi_{t+1} - \mathbf{M}\Psi_t + \mathbf{N}\|^2 + \gamma(\Psi_1 - \Psi_*)' \mathbf{Q}_*^{-1} (\Psi_1 - \Psi_*) + \eta \Psi_T' \mathbf{J} \Psi_T. \quad (3.7)$$

The minimizer  $\hat{\Psi}$  then depends on the choice of the parameters  $\theta = (U_0, \delta, \eta, \alpha, \gamma)$ . The normal equations for  $\hat{\Psi} = \hat{\Psi}(\theta)$  are

$$\mathbf{E}_{\text{weak}} \hat{\Psi} = \mathbf{b}_{\text{weak}}, \quad (3.8)$$

where

$$\mathbf{E}_{\text{weak}} = \mathbf{K}'\mathbf{K} + \alpha \mathbf{C}_M + \gamma \begin{pmatrix} \mathbf{Q}_*^{-1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 0 \end{pmatrix} + \eta \begin{pmatrix} 0 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \mathbf{J} \end{pmatrix}, \quad (3.9)$$



the fact that the model integration has larger discretization errors due to its coarser finite-difference mesh. This leads to errors in the phase and group velocities of Rossby waves that can be compensated for statistically by using parameter values that are not precisely those of the nature simulation. Below we will see that the fitted  $(U_0, \delta)$  are closer to the minimizer of  $\text{MSE}_V$  than to the nature value, which is not surprising considering the optimality criteria  $R(\theta) = n_{\text{dat}}^{-1} \sum_{t \in \Lambda} \|\mathbf{K}_t \Psi_t^{\text{TRUE}} - \mathbf{K}_t \Psi_t(\theta)\|^2$  of (2.13) associated with the estimation procedure.

#### 4. Results

##### a. The weak constraint case

We used this model in several different experiments. In the first experiment, we generated a nature run with the streamfunction as in (3.4) and a set of observations and a forecast as shown in Figs. 3 and 5. Using the nature run and this data, we examined in detail the sensitivity of the target criteria  $R^{1/2}(\theta)$  to the parameters  $\theta = (U_0, \delta, \eta, \alpha, \gamma)$ . Then we examined the efficacy of  $V(\theta)$  and  $U^{1/2}(\theta)$  as proxies for the in practice unobservable  $R^{1/2}(\theta)$ . We evaluated  $R$ ,  $V$ , and  $U$  on a grid in five dimensions. For  $(U_0, \delta)$  we used a fine-grid scheme,  $41 \times 41$  for  $U_0 \in [0.031, 0.063]$  and  $\delta \in [0.07, 0.11]$ , which covers the true values  $(0.0355, 0.100)$ , while for  $(\eta, \alpha, \gamma)$  we used a coarse grid of  $3 \times 5 \times 4$  values;  $\log_{10}\eta = 3, 4, 5$ ;  $\log_{10}\alpha = 3.04, 3.91, 4.78, 5.65, 6.51$ ; and  $\log_{10}\gamma = 1.94, 2.24, 2.54, 2.84$ . These values actually were chosen after some preliminary experimentation. Only plots for the three values  $3.04, 4.78, 6.51$  for  $\log_{10}\alpha$  will be shown here. Figure 6 gives 36 contour plots for  $R^{1/2}(U_0, \delta, \eta, \alpha, \gamma)$  for the  $36 = 3 \times 3 \times 4$  values of  $\eta, \alpha, \gamma$  noted. The global minimum over this grid occurs at  $\log_{10}\eta = 4, \log_{10}\alpha = 4.78, \log_{10}\gamma = 2.54$  (middle plot in the lower left block), and  $(U_0, \delta) = (0.041, 0.102)$ , and is indicated by an asterisk. This value for  $(U_0, \delta)$  may be compared with the minimizer  $(U_0, \delta) = (0.038, 0.095)$  of  $\text{MSE}_V^{1/2}(U_0, \delta)$  of section 3f, as well as with the true values  $(0.0355, 0.100)$ . From Fig. 6 the sensitivity of the target function  $R^{1/2}$  to changes in  $(U_0, \delta)$  for each set of values of  $(\eta, \alpha, \gamma)$  may be observed. Letting  $\theta_R$  be the value of  $\theta$  noted above where the global minimum occurs, the sensitivity of  $R(\theta)$  near  $\theta_R$  to changes in the other parameters may be seen by examining  $R^{1/2}(\theta)/R^{1/2}(\theta_R)$  while varying one parameter at a time and setting the other parameters at their minimizing values. Using the data files that were used to generate Fig. 6, the results are plotted schematically in Fig. 7. The horizontal axis in Fig. 7 is  $\log_{10}\eta, \log_{10}\alpha$ , or  $\log_{10}\gamma$ . The dotted line gives  $\min_{\eta} R^{1/2}(\theta)/R^{1/2}(\theta_R)$ , where  $\min_{\eta}$  means that the minimum is taken over all the parameters except  $\eta$ , similarly for  $\alpha$  (solid line), and  $\gamma$  (dashed line). Here,  $R^{1/2}$  is clearly sensitive to changes in  $\eta$  and  $\gamma$ , while

it is sensitive to decreases in  $\alpha$  below the minimizer but is quite flat as  $\alpha$  increases. Meanwhile,  $R^{1/2}(\theta)$  increases very slowly as  $\alpha$  becomes large.

Figures 8 and 9 give the same 36 contour plots, except for  $V(\theta)$  and  $U^{1/2}(\theta)$ , respectively. Within the resolution of the plots, the minimizers of  $V(\theta)$  and  $U^{1/2}(\theta)$  are the same and are  $\log_{10}\eta = 4, \log_{10}\alpha = 4.78, \log_{10}\gamma = 2.24$  (middle plot in the upper-right block), and  $(U_0, \delta) = (0.0397, 0.100)$ . These minima are indicated by an asterisk on the plots. It can be seen that they both follow  $R^{1/2}$  fairly well. (In theory,  $V \approx R + \sigma_v^2$  in the neighborhood of the minimizer, while  $U \approx R$ ; see WJGG). Letting  $\theta_R, \theta_V$ , and  $\theta_U$  be the minimizers of  $R(\theta), V(\theta)$ , and  $U(\theta)$ , respectively, we can measure the inefficiencies  $I_V$  and  $I_U$  of the estimates provided by  $\theta_V$  and  $\theta_U$  by  $I_V = R^{1/2}(\theta_V)/R^{1/2}(\theta_R)$  and  $I_U = R^{1/2}(\theta_U)/R^{1/2}(\theta_R)$ . In this case the inefficiencies, which measure the root-mean-square error when the estimate is used divided by the minimum root-mean-square error obtainable if nature were known, are both given by  $0.2991/0.2764 = 1.082$ , to within the resolution of the crude search employed. Theoretical discussions of the properties of these estimates may be found in Li (1986) and Wahba (1990b). We conclude that (for this example) the minimizers  $\theta_V$  and  $\theta_U$  of  $V(\theta)$  and  $U(\theta)$  do very well from the point of view that  $R(\theta_V)$  and  $R(\theta_U)$  are not much larger than  $R(\theta_R)$ . We can conclude also that since  $R^{1/2}$  is quite sensitive to all of the parameters except  $\alpha$ , we should be able to obtain relatively stable estimates of these parameters for similar signals. Notice that in each of Figs. 6, 8, and 9 the location of the minimizing value of  $(U_0, \delta)$  is nearly the same in the four lower-right blocks in each  $3 \times 3$  contour plot in  $(U_0, \delta)$ , but not elsewhere. This suggests that good estimates of  $(U_0, \delta)$  may be obtained independently of  $\eta, \alpha, \gamma$  for a range of values of these parameters about the minimizer but, if these parameters are fixed at poor values, then a search made to find the minimum over  $(U_0, \delta)$  may give quite different values.

##### b. The strong constraint case

Using the same data as in section 4a, Fig. 10 gives nine contour plots each for  $R^{1/2}(\theta), V(\theta)$ , and  $U^{1/2}(\theta)$ , where now  $\theta = (U_0, \delta, \eta, \gamma)$ . The grid scheme is similar to the one for the weak constraint:  $41 \times 41$  for  $U_0 \in [0.0335, 0.0446]$ , and  $\delta \in [0.075, 0.115]$ , with three values for  $\log_{10}\eta, 3, 4, 5$ , and three values for  $\log_{10}\gamma, 1.54, 2.04, 2.54$ . These values were chosen after some trial and error. The minimum of  $R^{1/2}$  over this fine grid in  $(U_0, \delta)$  and crude grid in  $\eta, \gamma$  occurred at  $\log_{10}\gamma = 2.04, \log_{10}\eta = 4$ , and  $(U_0, \delta) = (0.041, 0.106)$ , and is marked with an asterisk, as are the other minima discussed next. The minimum of the  $V(\theta)$  occurs for the same  $\eta$  and  $\gamma$ , at  $(U_0, \delta) = (0.039, 0.099)$ . Here,  $R^{1/2}(\theta_V)$ , where  $\theta_V$  is the minimizer of  $V$  given above, was  $0.330$ . Thus  $I_V = 0.330/0.319 = 1.034$ . As before,  $U(\theta)$  behaves similar to  $V(\theta)$  and gives the

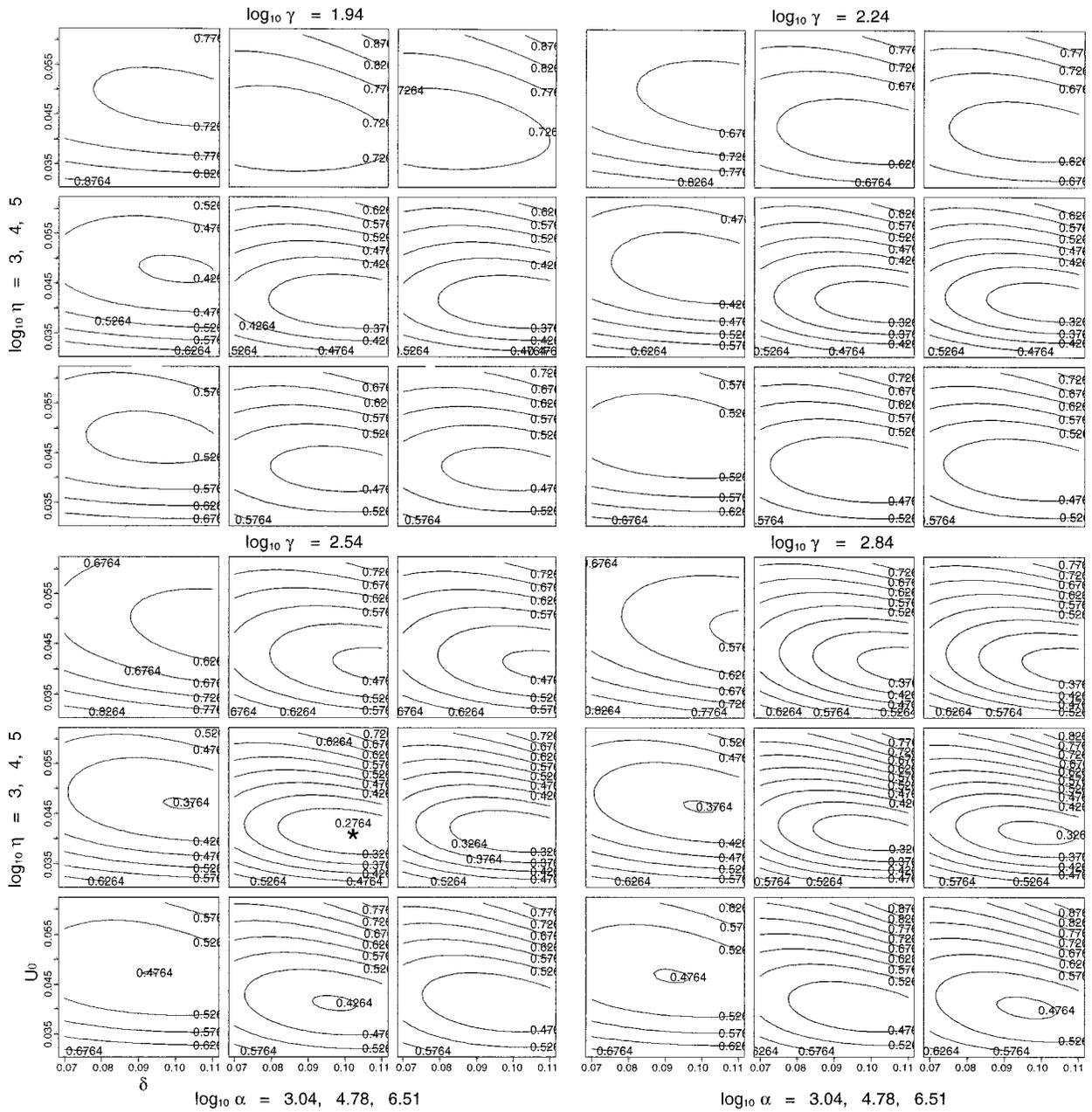


FIG. 6.  $R^{1/2}(\theta)$ .

same minimizer as  $V(\theta)$ , to the resolution of the crude grid search, and the minimizer of  $U(\theta)$  or  $V(\theta)$  provides a good estimate of the minimizer of  $R(\theta)$ . In theory the strong constraint case is a special case of the weak constraint case as  $\alpha \rightarrow \infty$ . However, we cannot compute the strong constraint case as a limit of the weak constraint case, and the strong constraint case has been computed via an independent algorithm. We did try a set of weak constraint cases with  $U_0, \delta,$

$\eta,$  and  $\gamma$  set to their values above which minimized  $R^{1/2}$  in the strong constraint case, and then let  $\alpha$  increase. We found that  $R^{1/2}(\theta)$  appeared to approach its strong constraint value of 0.319. This may be compared with the global minimum of  $R^{1/2}(\theta)$  found in section 4a of 0.2764 (although the search grid in  $\log_{10} \gamma$  was not exactly the same).

The extremely wiggly curves in some parts of the contour plots in Fig. 9 are caused by the instability

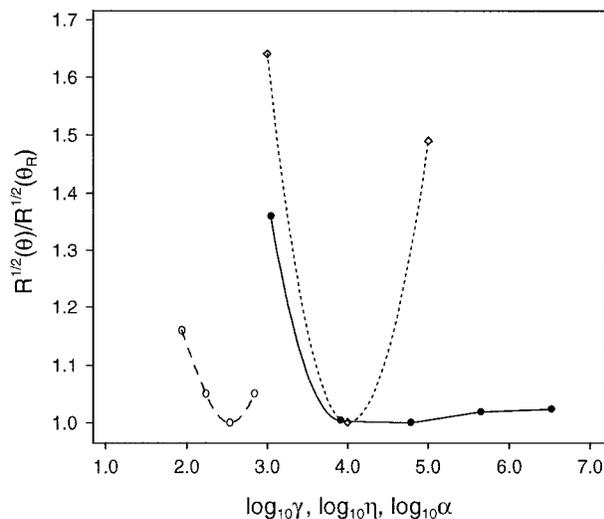


FIG. 7. Relative sensitivities of  $R^{1/2}(\theta)$  to changes in  $\log_{10}\eta$ ,  $\log_{10}\alpha$ , and  $\log_{10}\gamma$ . Dotted line,  $\eta$ ; solid line,  $\alpha$ ; dashed line,  $\gamma$ .

of the computation.<sup>5</sup> When  $(U_0, \delta)$  is in those wiggly regions of the contour plots, the condition number of the matrix  $\mathbf{E}_{\text{strong}}$  in the normal equations of (3.13) is very large. Physically, there is an admissible range for the distributed parameters in the barotropic vorticity equation. The discretization scheme we used to form the dynamical model is a conditionally stable scheme. Beyond the admissible range of the distributed parameters, the instability of the discretization scheme will show up. The wiggly curves may correspond to the boundary or near boundary of this range. This instability was not evident in the weak constraint case for the values of  $\alpha$  that we tried.

*c. Replications*

It is known that variability in the random errors in the data that go into adaptive estimation of smoothing parameters in the moderate sample size case can have an observable effect on smoothing parameter estimation. We have therefore replicated this experiment 10 additional times with the same nature but with new random numbers drawn for the forecast errors and the observation errors. Instead of a global search, we have used a slightly modified Powell’s direction set method (Press et al. 1990, section 10.5; Avriél 1976) to search for the minimizers of  $R(\theta)$ ,  $V(\theta)$ , and  $U(\theta)$ . Since all three functions are expected to have their minima with all parameters positive, Powell’s algorithm was modified so that when the search approaches or crosses the boundary between positive and negative values, a large func-

tion value will be returned so that the next search direction chosen by Powell’s algorithm will be changed to move toward the interior of the positive orthant. Letting  $F(\theta_k)$  stand for  $R(\theta_k)$ ,  $V(\theta_k)$ , or  $U(\theta_k)$ , where  $\theta_k$  is the  $k$ th value of  $\theta$ , the modified Powell’s algorithm was set to stop when  $|F(\theta_{k+1}) - F(\theta_k)| \leq 10^{-4}[F(\theta_{k+1}) + F(\theta_k)]$ .

The results may be seen in Table 1. Replicate (11) used the same observations and forecast as went into Figs. 6, 8, 9, and 10, but the Powell’s algorithm was used for the search. The columns headed by  $(U_0, \delta)$ ,  $\log_{10}\alpha$ ,  $\log_{10}\gamma$ , and  $\log_{10}\eta$  give, for replicates (1)–(11), the values of the components of  $\theta$  that were found by the Powell’s algorithm to minimize the PMSE, UBR, and GCV functions, that is, to minimize  $R$ ,  $U$ , and  $V$ . The column marked  $R^{1/2}(\theta)$  gives the value of  $R^{1/2}$  at the  $\theta$  minimizing  $R$ ,  $U$ , and  $V$ , and the column marked “Ineff” gives  $R^{1/2}(\hat{\theta})/R^{1/2}(\theta_R)$ , where  $\hat{\theta}$  is the minimizer that was found for  $U$  or  $V$  and  $\theta_R$  is the minimizer of  $R$ . The replicates marked (11A) and (11B) give, respectively, the same information previously obtained with the crude grid search behind Figs. 6, 8, and 9 (11A) and Fig. 10 (11B).

It can be seen that  $R^{1/2}(\theta)$  at the minimizers found in (11) and (11A) are quite similar, which are shown to indicate how close minimizers obtained by the grid search and the Powell’s algorithm are. Several noteworthy observations are to be made from this table.

- 1) The values of  $U_0$  that minimize  $R$  are systematically larger than the “true”  $U_0 = 0.0355$ , and the values of  $U_0$  that minimize  $V$  and  $U$  appear to be scattered about the minimizer of  $R$ , not about 0.0355. This is suggestive of the perfectly plausible idea that the “best” value of a distributed parameter for fitting purposes in a computer model that is an imperfect representation of the truth is not necessarily the true value.
- 2) An examination of the Ineff column indicates that 8 out of these 11 cases had both the UBR and the GCV inefficiencies less than 1.20. There were three GCV and two UBR inefficiencies between 1.2 and 1.38 (replicates 1, 2, and 8). In each of these cases the difficulty appears to be the apparent inability of the minimizer of  $V$  or  $U$  to find an  $\alpha$  close to the minimizer of the  $\alpha$  that minimizes  $R$ . This is certainly related to the fact that  $R$  appears to be insensitive to  $\alpha$  over a wide range of  $\alpha$ , and this insensitivity may be related to the fact that we are using a least squares penalty on model error (i.e.,  $\mathbf{Q}_t$  was set to  $\mathbf{I}$ ), whereas the true model error is certainly correlated in space, as well as time.
- 3) Putting the model in as a strong constraint avoids the problem of choosing  $\alpha$ , but this involves an assumption that the model error is sufficiently small over the time period considered. In the present example the weak constraint appears to be beneficial, as can be seen by comparing replicates (11A) and

<sup>5</sup> Some chaotic appearing contours above the extremely wiggly contours have been removed. Future experiments must, of course, be concerned with stable calculations.

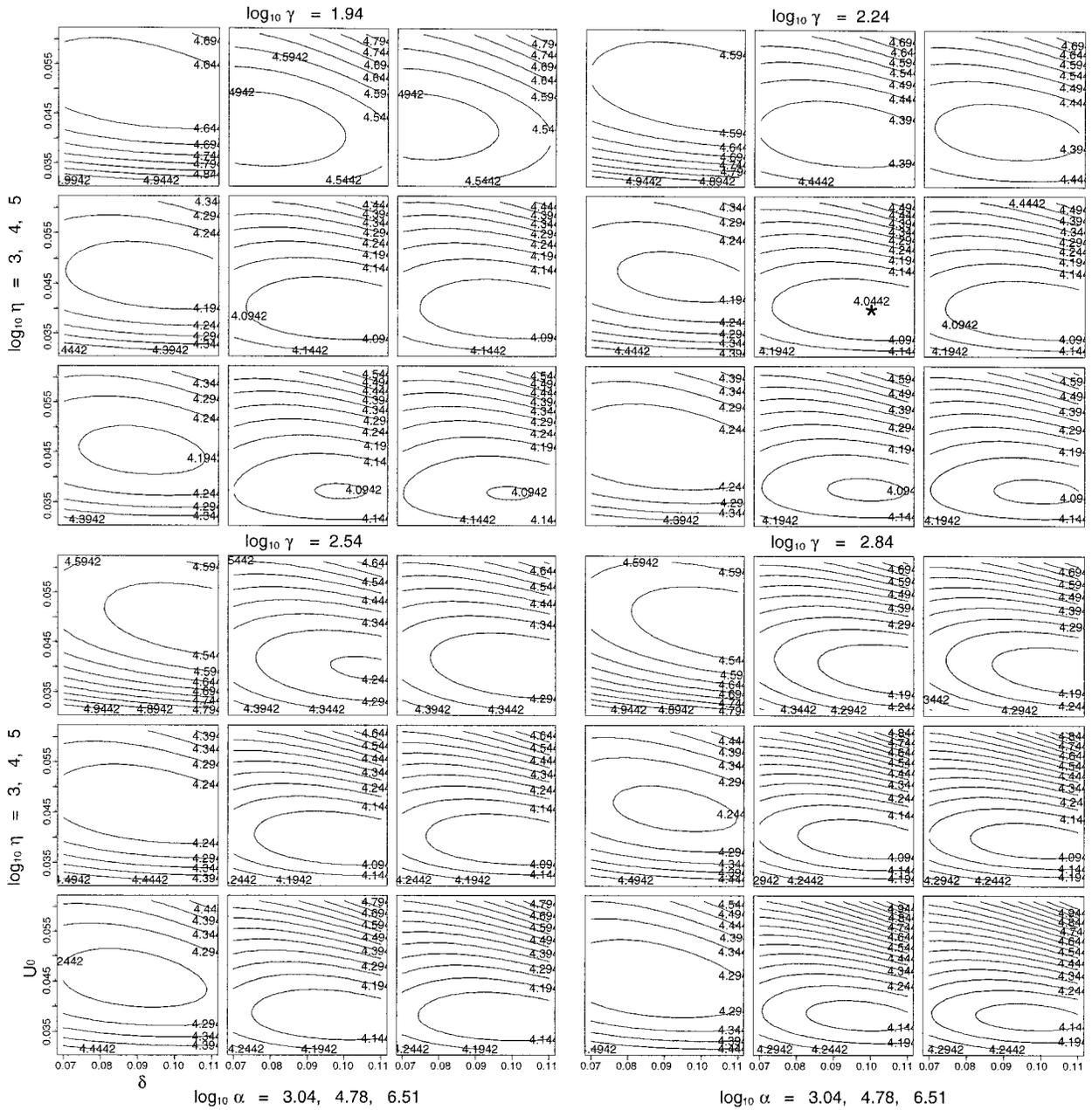


FIG. 8.  $V(\theta)$ .

(11B), but only experimentation with more realistic systems can answer this question in practice.

In Tables 2 and 3 we report results of repeating the experiment of Table 1 with smaller noise and stronger signal, respectively. For Table 2 the observation error standard deviation has been reduced from 2 to 1 m s<sup>-1</sup>, and the forecast error standard deviation in the wind has been reduced from 0.485 to 0.16 m s<sup>-1</sup>. For Table 3 the observation and wind forecast error standard deviations have been set back to the same values as in Table 1, but the signal has been made stronger by replacing  $\Psi_{\text{start}}(x)$  of (3.4)

by  $6.28\Psi_{\text{start}}(x)$ . Thus, the signal-to-noise ratio has gone up in both of these cases. We note the following.

- 4) As the signal-to-noise ratio behind both Tables 2 and 3 has gone up, the inefficiencies overall are less than those in Table 1, falling between a low of 1.01 and a high of 1.16.
- 5) Some of the replicates in Tables 2 and 3 demonstrate an insensitivity to  $\alpha$  in that the UBR or GCV estimate of  $\alpha$  is quite different than the PMSE value while the inefficiency is still quite small.
- 6) As in Table 1, the values of  $U_0$  that minimize  $U$

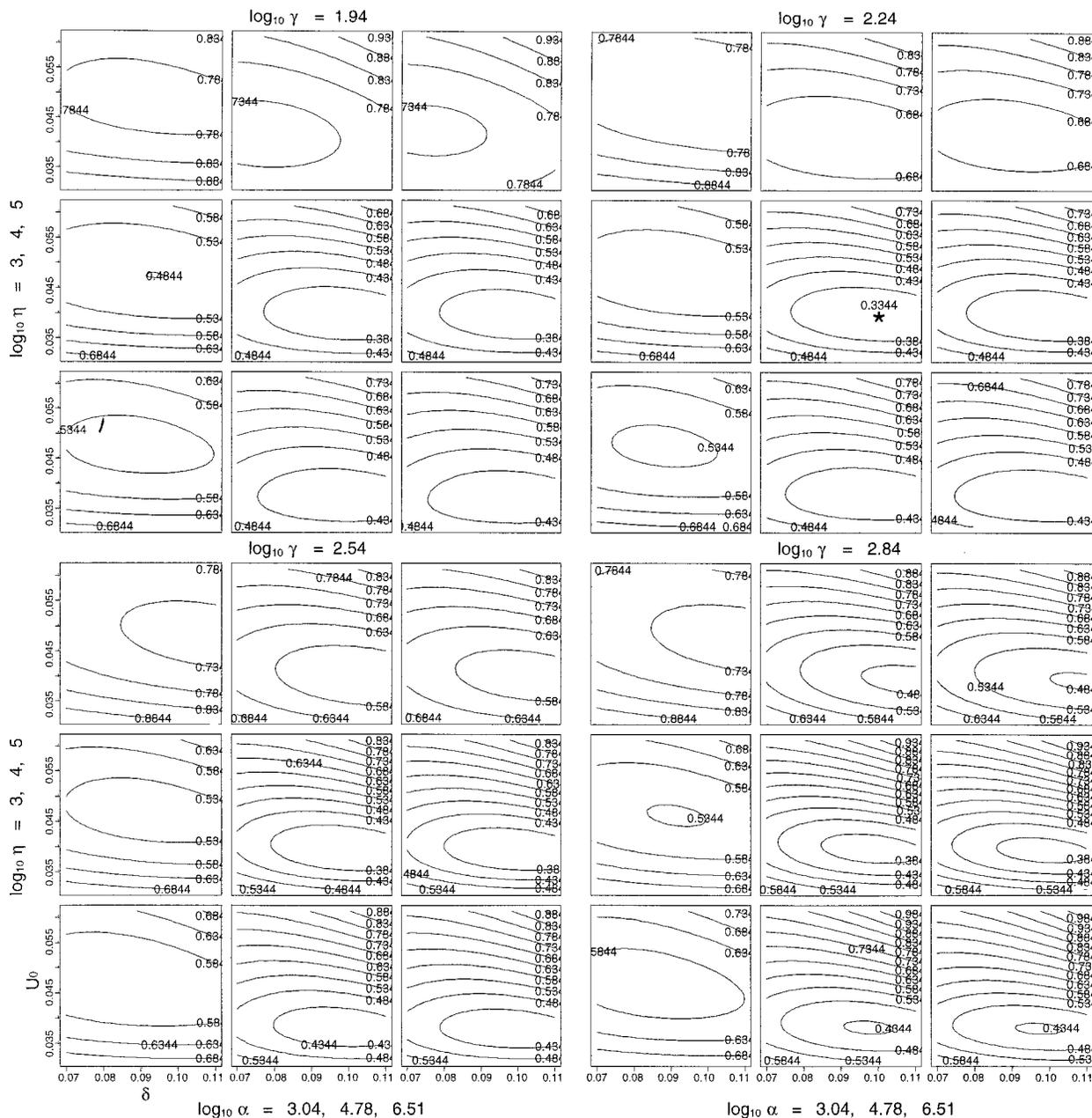


FIG. 9.  $U^{1/2}(\theta)$ .

and  $V$  in Tables 2 and 3 appear to be scattered about the  $U_0$ , which minimizes  $R$ , not about 0.0355. The value of  $U_0$ , which minimized  $R$ , is about the same in Tables 1 and 2, which have the same signal, while it is less in Table 3, where the signal is more intense.

- 7) The variability in the estimates in Tables 1, 2, and 3 suggest the variability that might be encountered in practice due to random errors in the observation and forecast, with similar sample size and data distribution. This component of the variability can be expected to go down as the sample size goes

up; conversely, changes in the synoptic situation can be expected to cause changes in the tuning parameter estimates.

*d. Imputed and estimated weighting and smoothing parameters*

As noted and described in section 2, the statistical formalism behind the Kalman filter leads to specific meanings for  $\alpha$ ,  $\gamma$ , and  $\eta$  as ratios of variances. Suppose observation error, model error, and forecast error

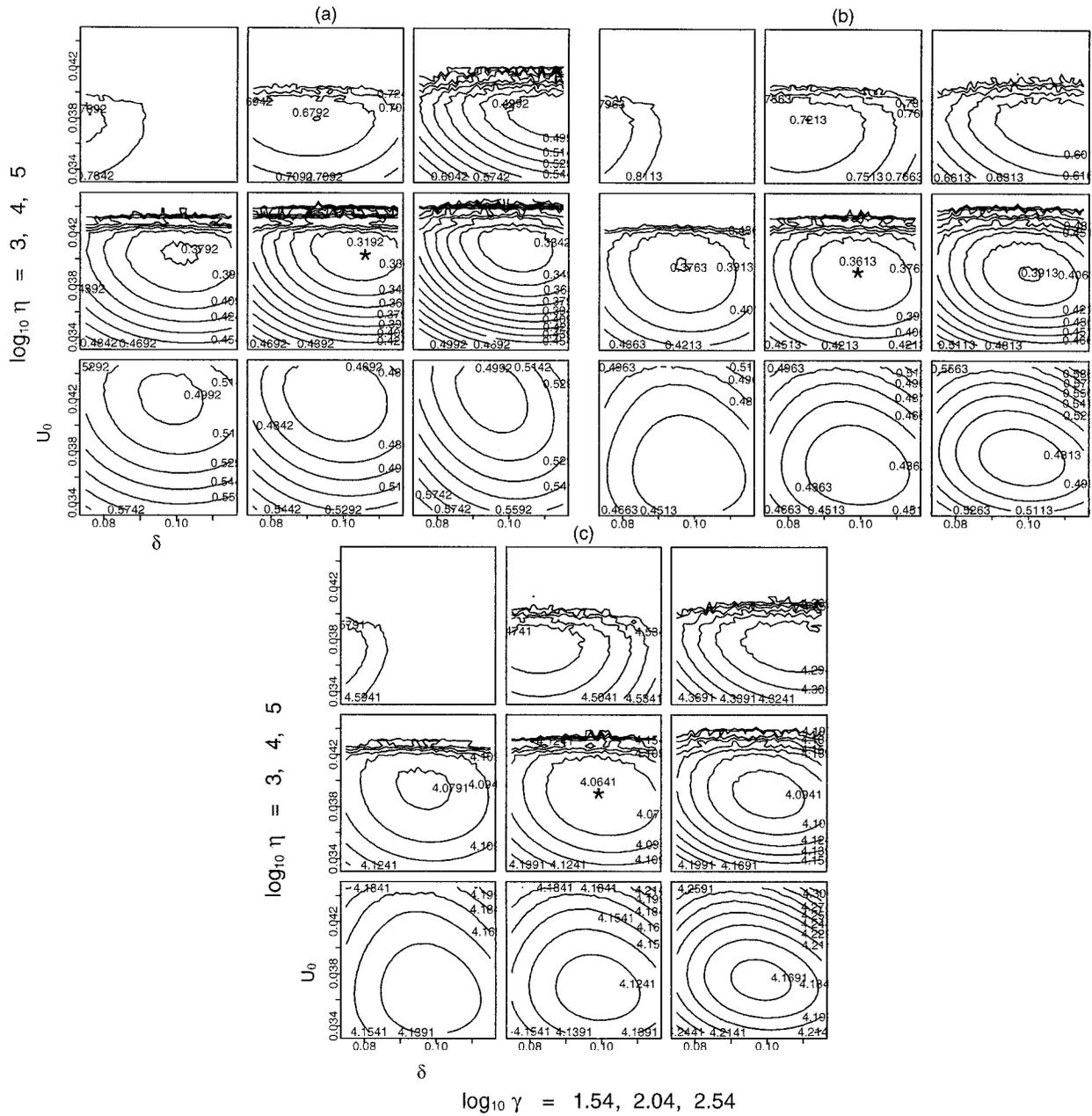


FIG. 10. (a)  $R^2(\theta)$ , (b)  $U^{1/2}(\theta)$ , (c)  $V(\theta)$ , strong constraint.

were all random vectors with the covariance matrices  $\sigma_o^2 \mathbf{S}_t$ ,  $\sigma_m^2 \mathbf{Q}_t$ , and  $\sigma_f^2 \mathbf{Q}_*$ , which are implied by the variational problem (2.6), which is being solved. Suppose we pretended that  $\Psi_T$  were a random vector with the covariance  $b\mathbf{J}^{-1}$ . Then  $\alpha = \sigma_o^2/\sigma_m^2$ ,  $\gamma = \sigma_o^2/\sigma_f^2$ , and  $\eta = \sigma_o^2/b$ . In this section we compare the tuned values of  $\alpha$ ,  $\gamma$ , and  $\eta$  that were obtained, with theoretical values for these variance ratios that we derived or imputed as though the assumptions behind the Kalman filter were true. To the extent that the reasoning described here can be applied in practice, it can provide a method for getting starting values for the tuning

parameters and may also provide a reasonableness check on the values obtained and/or the model being fitted. It does have to be kept in mind, however, that, as in the estimation of  $(U_o, \delta)$ , that the optimum “theoretical” statistical parameters may not be the best for prediction if the assumptions behind the theoretical model [which consists here of (2.1), (2.2), and (2.3), including misspecification of  $\mathbf{S}_t$ ,  $\mathbf{Q}_t$ , and  $\mathbf{Q}_*$ ] are not satisfied.

In our experiment, the observation error and the forecast error were simulated according to the given random assumptions, and so we know the true  $\sigma_o^2$  and

TABLE 1. Case 1: 2 m s<sup>-1</sup> observation error standard deviation and 0.485 m s<sup>-1</sup> wind forecast error standard deviation.

Replicate		$R^{1/2}(\theta)$	Ineff	$\theta: \{(U_0, \delta)\}$	$\log_{10}\alpha$	$\log_{10}\gamma$	$\log_{10}\eta$
1	PMSE	0.3147		(0.0381, 0.1018)	6.66	2.31	4.25
	UBR	0.4337	1.3781	(0.0416, 0.0903)	9.48	2.38	3.32
	GCV	0.4175	1.3267	(0.0417, 0.0904)	9.48	2.45	3.47
2	PMSE	0.2994		(0.0413, 0.1035)	5.08	2.37	4.15
	UBR	0.3464	1.1570	(0.0451, 0.0940)	3.67	2.33	3.95
	GCV	0.3755	1.2542	(0.0444, 0.0926)	3.40	2.15	3.80
3	PMSE	0.3278		(0.0408, 0.1027)	4.85	2.08	4.00
	UBR	0.3319	1.0125	(0.0413, 0.1015)	9.47	2.24	4.04
	GCV	0.3349	1.0217	(0.0406, 0.1032)	9.13	2.31	4.17
4	PMSE	0.3345		(0.0404, 0.1055)	5.80	2.35	4.06
	UBR	0.3567	1.0663	(0.0422, 0.1121)	9.75	2.58	4.42
	GCV	0.3527	1.0544	(0.0410, 0.1143)	9.04	2.56	4.43
5	PMSE	0.2986		(0.0426, 0.1056)	4.54	2.74	4.33
	UBR	0.3475	1.1637	(0.0391, 0.1229)	9.51	2.49	4.25
	GCV	0.3505	1.1738	(0.0393, 0.1224)	9.48	2.37	4.14
6	PMSE	0.3488		(0.0396, 0.0999)	6.58	2.24	4.11
	UBR	0.3636	1.0424	(0.0406, 0.0933)	6.49	2.55	4.41
	GCV	0.3601	1.0324	(0.0401, 0.0926)	6.57	2.46	4.33
7	PMSE	0.3694		(0.0411, 0.1034)	5.82	2.17	4.00
	UBR	0.3925	1.0625	(0.0436, 0.0973)	4.50	2.23	4.29
	GCV	0.3926	1.0628	(0.0437, 0.0972)	4.55	2.24	4.30
8	PMSE	0.3665		(0.0407, 0.1078)	4.89	2.54	4.20
	UBR	0.4581	1.2499	(0.0389, 0.0971)	6.41	2.46	3.29
	GCV	0.4595	1.2537	(0.0390, 0.0970)	6.16	2.46	3.28
9	PMSE	0.3715		(0.0415, 0.1057)	5.37	2.33	4.10
	UBR	0.3824	1.0293	(0.0415, 0.1059)	4.04	2.25	3.93
	GCV	0.3915	1.0538	(0.0416, 0.1054)	3.85	2.19	3.86
10	PMSE	0.3047		(0.0401, 0.1008)	6.67	1.98	4.07
	UBR	0.3363	1.1037	(0.0429, 0.1006)	6.67	2.21	3.98
	GCV	0.3370	1.1060	(0.0428, 0.1008)	6.67	2.25	4.03
11	PMSE	0.2750		(0.0411, 0.1041)	4.58	2.68	4.33
	UBR	0.3023	1.0993	(0.0397, 0.0995)	4.27	2.21	3.84
	GCV	0.3087	1.1225	(0.0397, 0.0993)	4.17	2.17	3.82
11A	PMSE	0.2764		(0.0407, 0.1020)	4.78	2.54	4.00
	UBR	0.2991	1.0821	(0.0397, 0.1000)	4.78	2.24	4.00
	GCV	0.2991	1.0821	(0.0397, 0.1000)	4.78	2.24	4.00
11B	PMSE	0.3192		(0.0411, 0.1060)	$\infty$	2.04	4.00
	UBR	0.3305	1.0354	(0.0397, 0.0993)	$\infty$	2.04	4.00
	GCV	0.3305	1.0354	(0.0397, 0.0993)	$\infty$	2.04	4.00

$\sigma_f^2$  and hence the theoretical  $\gamma$ . Assumptions like these are very natural in practice. We are of course oversimplifying things here by assuming that  $\mathbf{S}_i$  and  $\mathbf{Q}^*$  are known. However, it is not unreasonable to identify  $\sigma_o^2$  and  $\sigma_f^2$  and hence  $\gamma$  with quantities whose values are at least approximately known in practice.

Recalling the definition of  $\mathbf{J}$  given in (3.6), we imputed a value for  $b^{1/2}$  as  $(b^{put})^{1/2} = \{(1/194) \sum_{i=1}^{194} [\Psi_T^{TRUE}(x_{i+2}) - 2\Psi_T^{TRUE}(x_{i+1}) + \Psi_T^{TRUE}(x_i)]^2\}^{1/2}$ . Note that  $b^{put}$  would be an unbiased estimate for  $b$  under the (artificial) assumption that  $\Psi_T^{TRUE}$  were a random vector whose second differences on the model space grid were independent  $\mathcal{N}(0, b)$ . This assumption corresponds to the smoothness penalty that has been imposed on  $\Psi_T$  in (3.7) and should be considered as a standin for more realistic penalty functionals that could be developed based on, for example, damping out waves inversely proportional to the energy that they are

a priori believed to have in the real atmosphere. A simple example may be found in Wahba (1982b).

Of all the variances, it is probably least realistic to attempt to impute a theoretical value to  $\sigma_m^2$ , even given everything we know in this experiment, because the “model” for the model error is unrealistic. Here, nature was generated by evolving an assumed (smooth) streamfunction at the start, and the model error was generated by differences between the nature and model dynamics, but the model for model error is zero mean, independent from time to time with covariance [from (2.6)]  $\sigma_m^2 \mathbf{Q}_i$  with  $\mathbf{Q}_i = \mathbf{I}$ . However, since we “know everything” about nature and the estimates here, we will attempt to impute values to  $\sigma_m^2$  from this knowledge. We imputed the ad hoc value  $\sigma_m^{put}$  to  $\sigma_m$  as  $\sigma_m^{put} = [(13 \times 194)^{-1} \sum_{t=1}^{13} \|\Psi_t^{TRUE} - \Psi_t^{MODEL}(U_0, \delta)\|^2]^{1/2}$ . Under the (artificial) assumption that

TABLE 2. Case 2: 1 m s<sup>-1</sup> observation error standard deviation and 0.16 m s<sup>-1</sup> wind forecast error standard deviation.

Replicate		$R^{1/2}(\theta)$	Ineff	$\theta: \{(U_0, \delta)\}$	$\log_{10}\alpha$	$\log_{10}\gamma$	$\log_{10}\eta$
1	PMSE	0.2376		(0.0399, 0.0996)	5.07	2.609	4.04
	UBR	0.2713	1.1418	(0.0416, 0.0952)	3.50	2.625	3.38
	GCV	0.2601	1.0947	(0.0415, 0.0950)	3.83	2.809	3.54
2	PMSE	0.2320		(0.0410, 0.1010)	4.20	2.527	3.90
	UBR	0.2489	1.0728	(0.0429, 0.0958)	3.44	2.540	3.57
	GCV	0.2581	1.1125	(0.0428, 0.0955)	3.27	2.493	3.46
3	PMSE	0.2388		(0.0409, 0.1011)	3.74	2.263	3.76
	UBR	0.2458	1.0293	(0.0408, 0.0998)	4.92	2.386	3.78
	GCV	0.2509	1.0507	(0.0409, 0.0996)	9.87	2.618	4.02
4	PMSE	0.2445		(0.0408, 0.1014)	4.49	2.576	3.89
	UBR	0.2586	1.0577	(0.0419, 0.1044)	8.77	2.871	4.25
	GCV	0.2604	1.0650	(0.0418, 0.1050)	8.55	2.950	4.27
5	PMSE	0.2397		(0.0414, 0.1015)	8.77	2.829	4.28
	UBR	0.2534	1.0572	(0.0399, 0.1093)	8.78	2.653	4.14
	GCV	0.2542	1.0605	(0.0399, 0.1094)	8.78	2.565	4.07
6	PMSE	0.2350		(0.0401, 0.1008)	4.07	2.182	3.85
	UBR	0.2659	1.1315	(0.0412, 0.0984)	3.23	2.525	3.62
	GCV	0.2730	1.1617	(0.0412, 0.0982)	3.15	2.524	3.59
7	PMSE	0.2442		(0.0407, 0.0988)	4.42	2.527	3.88
	UBR	0.2485	1.0176	(0.0407, 0.0962)	7.04	2.696	4.17
	GCV	0.2490	1.0196	(0.0407, 0.0961)	7.11	2.713	4.20
8	PMSE	0.2484		(0.0407, 0.1013)	3.97	2.472	3.72
	UBR	0.2600	1.0467	(0.0422, 0.0978)	4.14	2.431	4.02
	GCV	0.2612	1.0515	(0.0423, 0.0976)	4.38	2.501	4.11
9	PMSE	0.2451		(0.0413, 0.1020)	3.90	2.843	3.98
	UBR	0.2909	1.1868	(0.0403, 0.0959)	6.89	2.718	3.28
	GCV	0.2877	1.1738	(0.0404, 0.0957)	7.02	2.783	3.36
10	PMSE	0.2478		(0.0411, 0.1023)	4.25	2.617	3.99
	UBR	0.2506	1.0112	(0.0415, 0.1015)	3.99	2.457	3.78
	GCV	0.2517	1.0157	(0.0415, 0.1015)	3.91	2.427	3.75
11	PMSE	0.2251		(0.0411, 0.1016)	4.03	2.560	3.98
	UBR	0.2421	1.0755	(0.0406, 0.0989)	3.88	2.107	3.61
	GCV	0.2403	1.0675	(0.0406, 0.0990)	3.94	2.140	3.63

$\Psi_i^{\text{MODEL}}(U_0, \delta) \sim \mathcal{N}(\Psi_i^{\text{TRUE}}, \sigma_m^2 \mathbf{I})$ , independent for different  $t$ , then  $(\sigma_m^{\text{put}})^2$  would be an unbiased estimate of  $\sigma_m^2$ .

The results for  $(b^{\text{put}})^{1/2}$  were  $(b^{\text{put}})^{1/2} = 0.055 \text{ km}^2 \text{ s}^{-1}$  for cases 1 and 2 and  $0.345 \text{ km}^2 \text{ s}^{-1} = 6.28 \times 0.055 \text{ km}^2 \text{ s}^{-1}$  for case 3.

To obtain  $\sigma_m^{\text{put}}$ , for cases 1 and 2, which have the same signal, we used the values of  $(U_0, \delta)$ , which minimized  $\sigma_m^{\text{put}}$ , the result was  $\sigma_m^{\text{put}} = 0.158 \text{ km}^2 \text{ s}^{-1}$ . For case 3,  $\sigma_m^{\text{put}}$  was obtained in the same way and was  $0.279 \text{ km}^2 \text{ s}^{-1}$ .

The third column of Table 4 gives the imputed values of  $\kappa^{1/2}$  for  $\kappa = \alpha, \gamma, \eta$ , as the ratios of the actual or imputed standard deviations, in physical units, and the fourth column gives  $\log_{10}$  of the square of the values in the third column, converted to dimensionless form for comparison with the estimated  $\log_{10}\alpha, \log_{10}\gamma$ , and  $\log_{10}\eta$ . The dimensionless form of the ratios in column 3 are obtained by multiplying them by  $(4496 \text{ km}) \times 10^{-3} \text{ km} \times \text{m}^{-1}$ . The fifth column gives the mean and standard deviation for the 11 values of  $\log_{10}\alpha, \log_{10}\gamma$ , and  $\log_{10}\eta$ , which minimize  $R(\theta)$  from the eleven replicates of cases 1, 2, and 3 from Tables 1, 2, and 3, and

the sixth and seventh columns give the same information for  $U(\theta)$  and  $V(\theta)$ . It can be seen in all three cases that the minimizers of  $R, V$ , and  $U$  in  $\log_{10}\gamma$  are generally close to the imputed value, it appears that these minimizers provide an estimate of the observed versus forecast variance ratio  $\sigma_o^2/\sigma_f^2$ . In cases 1 and 2 the minimizers in  $\eta$  are close to their imputed values, while in case 3 (stronger signal) the minimizer is larger. The minimizers in  $\alpha$  are larger by a factor of between  $10^{1/2}$  and  $10^3$  than their ad hoc imputed values. An examination of the evolution of  $\Psi_i^{\text{TRUE}} - \Psi_i^{\text{MODEL}}(U_0, \delta)$  (not shown here) indicates that this difference is fairly systematic in time, concentrated in the neighborhood of the maximum of  $U(x)$ . Thus the “model” for model error as well as the procedure for imputing  $\sigma_m^2$  are probably both unrealistic and may provide an explanation for the difference between the imputed and estimated  $\alpha$ . However, it appears that the fitting is drawing strongly to the model. Further work in accounting for model errors is warranted.

The imputed values allow common physical units to be attached to this particular set of smoothing parameters, which allows a direct comparison of sensitivity.

TABLE 3. Case 3: signal times 6.28, 2 m s<sup>-1</sup> observation error, and 0.485 m s<sup>-1</sup> wind forecast error.

Replicate		$R^{1/2}(\theta)$	Ineff	$\theta: \{(U_0, \delta)\}$	$\log_{10}\alpha$	$\log_{10}\gamma$	$\log_{10}\eta$
1	PMSE	0.2751		(0.0373, 0.1036)	3.719	2.205	3.578
	UBR	0.2879	1.0465	(0.0373, 0.1007)	3.310	2.247	3.252
	GCV	0.2799	1.0174	(0.0374, 0.1005)	3.557	2.226	3.348
2	PMSE	0.2656		(0.0373, 0.1046)	3.591	2.295	3.607
	UBR	0.2783	1.0478	(0.0379, 0.0983)	3.354	2.068	3.291
	GCV	0.2869	1.0802	(0.0378, 0.0982)	3.198	2.069	3.212
3	PMSE	0.2651		(0.0372, 0.1042)	3.408	2.196	3.494
	UBR	0.2691	1.0151	(0.0378, 0.1032)	3.697	2.123	3.439
	GCV	0.2856	1.0773	(0.0372, 0.1031)	5.051	2.304	3.586
4	PMSE	0.2847		(0.0373, 0.1035)	3.655	2.257	3.404
	UBR	0.2993	1.0577	(0.0373, 0.1116)	3.413	2.769	3.918
	GCV	0.2994	1.0650	(0.0374, 0.1109)	3.578	2.730	3.963
5	PMSE	0.2882		(0.0370, 0.1048)	3.332	2.629	3.584
	UBR	0.3176	1.0572	(0.0378, 0.1103)	7.524	2.566	3.892
	GCV	0.3155	1.0605	(0.0378, 0.1105)	7.520	2.473	3.825
6	PMSE	0.2655		(0.0370, 0.1033)	3.671	2.181	3.515
	UBR	0.2971	1.1190	(0.0367, 0.1052)	3.339	2.665	3.545
	GCV	0.3060	1.1525	(0.0366, 0.1051)	3.335	2.684	3.478
7	PMSE	0.2698		(0.0375, 0.1031)	3.601	1.972	3.499
	UBR	0.2835	1.0508	(0.0374, 0.0953)	5.524	1.994	3.519
	GCV	0.2834	1.0504	(0.0374, 0.0954)	5.525	1.999	3.526
8	PMSE	0.2706		(0.0372, 0.1057)	3.512	2.449	3.426
	UBR	0.2758	1.0192	(0.0372, 0.1021)	3.435	2.345	3.583
	GCV	0.2760	1.0199	(0.0373, 0.1020)	3.520	2.330	3.613
9	PMSE	0.2856		(0.0370, 0.1081)	3.402	2.841	3.762
	UBR	0.3150	1.1029	(0.0379, 0.0976)	6.826	2.098	3.128
	GCV	0.3138	1.0987	(0.0380, 0.0977)	7.730	2.135	3.167
10	PMSE	0.2797		(0.0370, 0.1062)	3.442	2.559	3.683
	UBR	0.2867	1.0250	(0.0379, 0.1038)	3.810	2.168	3.609
	GCV	0.2870	1.0261	(0.0378, 0.1039)	3.847	2.176	3.621
11	PMSE	0.2755		(0.0374, 0.1058)	3.418	2.605	3.649
	UBR	0.2890	1.0490	(0.0377, 0.1009)	3.755	1.905	3.382
	GCV	0.2877	1.0443	(0.0376, 0.1012)	3.842	1.963	3.409

In the case of variance ratios not involving the model errors, it appears that imputed values may well serve as starting points for the tuning.

### 5. Further discussion

#### a. Search algorithms

With regard to Powell’s search algorithm, we set a particularly stringent convergence criterion, and in each of the iterative searches in  $\theta$  behind the results of Tables 1, 2, and 3, there were of the order of several hundred iterations. This is clearly not acceptable in practice. Functions like  $R$ ,  $V$ , and  $U$  here tend to be smooth functions of their arguments. It is very expensive to compute a single value, and derivatives are not readily available. Without derivatives, effective descent algorithms like the conjugate gradient algorithm are not available. In these circumstances we think an appropriate way to proceed in the future is via what is called the “design” approach in the statistics literature. It is to 1) carefully select a set of points (the so-called design points), where the function is to be evaluated; 2) evaluate the function

at these design points; 3) interpolate the function at the design points using a smoothly differentiable, “legitimate” interpolation scheme; 4) find the minimum of the interpolant; 5) if desired, lay out another design (i.e., set of design points) around this minimum and evaluate the function at the new design points; and 6) interpolate to the design points lying within some region surrounding the minimum so far and find the minimum of the new interpolant. By a legitimate interpolation scheme, we mean one based on a positive definite or conditionally positive definite function; see Bates et al. (1993, section 5). There is extensive literature utilizing this so called design approach to efficiently find minima of computer-generated functions of several variables that are very expensive to evaluate. Choice of the design points in higher dimensions is an important issue. This approach may also be used to explore sensitivity to various parameters of interest. Bowman et al. (1993) describe the use of this approach to study the sensitivity of a global equivalent-barotropic model to certain parameters in the model, for example, viscosity. They use Latin hypercube designs and positive definite functions

TABLE 4. Comparison of imputed and estimated parameters.

Case	$\kappa$	Imputed $\kappa^{1/2}$	Dimensionless $\log_{10} \kappa$	Minimizer $R$ (mean $\pm$ sd)	Minimizer $U$ (mean $\pm$ sd)	Minimizer $V$ (mean $\pm$ sd)
1	$\alpha$	$\frac{\sigma_o}{\sigma_m^{\text{put}}} = \frac{2 \text{ m s}^{-1}}{0.158 \text{ km}^2 \text{ s}^{-1}}$	3.514	$5.531 \pm 0.826$	$6.745 \pm 2.450$	$6.591 \pm 2.389$
	$\gamma$	$\frac{\sigma_o}{\sigma_f} = \frac{2 \text{ m s}^{-1}}{0.726 \text{ km}^2 \text{ s}^{-1}}$	2.186	$2.345 \pm 0.235$	$2.357 \pm 0.142$	$2.328 \pm 0.140$
	$\eta$	$\frac{\sigma_o}{b^{\text{put}1/2}} = \frac{2 \text{ m s}^{-1}}{0.055 \text{ km}^2 \text{ s}^{-1}}$	4.428	$4.145 \pm 0.119$	$3.975 \pm 0.385$	$3.966 \pm 0.362$
2	$\alpha$	$\frac{\sigma_o}{\sigma_m^{\text{put}}} = \frac{1 \text{ m s}^{-1}}{0.158 \text{ km}^2 \text{ s}^{-1}}$	2.912	$4.628 \pm 1.420$	$5.325 \pm 2.143$	$5.801 \pm 2.501$
	$\gamma$	$\frac{\sigma_o}{\sigma_f} = \frac{1 \text{ m s}^{-1}}{0.242 \text{ km}^2 \text{ s}^{-1}}$	2.538	$2.546 \pm 0.199$	$2.546 \pm 0.204$	$2.593 \pm 0.220$
	$\eta$	$\frac{\sigma_o}{b^{\text{put}1/2}} = \frac{1 \text{ m s}^{-1}}{0.055 \text{ km}^2 \text{ s}^{-1}}$	3.826	$3.934 \pm 0.151$	$3.782 \pm 0.327$	$3.818 \pm 0.324$
3	$\alpha$	$\frac{\sigma_o}{\sigma_m^{\text{put}}} = \frac{2 \text{ m s}^{-1}}{0.2790 \text{ km}^2 \text{ s}^{-1}}$	3.018	$3.523 \pm 0.131$	$4.362 \pm 1.532$	$4.609 \pm 1.655$
	$\gamma$	$\frac{\sigma_o}{\sigma_f} = \frac{2 \text{ m s}^{-1}}{0.726 \text{ km}^2 \text{ s}^{-1}}$	2.186	$2.381 \pm 0.256$	$2.268 \pm 0.285$	$2.281 \pm 0.258$
	$\eta$	$\frac{\sigma_o}{b^{\text{put}1/2}} = \frac{2 \text{ m s}^{-1}}{0.3455 \text{ km}^2 \text{ s}^{-1}}$	2.832	$3.564 \pm 0.109$	$3.505 \pm 0.247$	$3.523 \pm 0.240$

from a convenient family, which includes covariances of Gaussian form. Other design possibilities include the I-, D-, and A-optimal designs (see Hardin and Sloane 1993), blending function designs also known as hyperbolic cross points (see Wahba 1978), and other interpolants including higher-order spline-like interpolants.

It may be of benefit to attempt to divide the various tuning parameters into groups that could be studied independently of each other, at least near the minimizing values. See the remarks at the end of section 4a concerning minimization with respect to  $(U_o, \delta)$ .

It is possible to estimate derivatives of  $V$  or  $U$  by difference quotients or, in some cases, by differentiating analytically and evaluating the resulting expressions by the randomized trace technique and then using a conjugate gradient or other descent algorithm using derivatives. Whether or not this will be more efficient than one of the design approaches above is a question for future research.

#### b. Other tuning criteria

Two other tuning criteria are in wide use, namely ordinary cross validation (leaving out one, or leaving out several) and various forms of maximum likelihood. Let  $\hat{\Psi}^{(k)}(\theta)$  be the estimate of  $\Psi$  obtained by solving the relevant variational problem with the  $k$ th data point left out; here  $k$  would run over the  $(5 \times 164)$  components of  $\mathbf{y}$ . Letting  $\mathbf{y}^k$  be the  $k$ th component of  $\mathbf{y}$  that was left out, the leaving-out-one estimate of  $\theta$  is the minimizer of  $V_o(\theta)$  defined by  $V_o(\theta) = n_{\text{dat}}^{-1} \sum_k [\mathbf{y}^k - \mathbf{K}^k \hat{\Psi}^{(k)}(\theta)]^2$ , where  $\mathbf{K}^k \hat{\Psi}^{(k)}(\theta)$  is the predicted value of  $\mathbf{y}^k$  based on

$\hat{\Psi}^{(k)}(\theta)$ ;  $V(\theta)$  and  $V_o(\theta)$  are closely related (see Wahba 1990b), although their computational costs are vastly different. A leaving-out-one estimate of  $\theta$  may be based on cross-validating only the data at the last time step; one minimizes  $V_{o,T}(\theta) = [n_{\text{dat}}(T)]^{-1} \sum_{k:\text{y}^k \text{ observed at } T} [\mathbf{y}^k - \mathbf{K}^k \hat{\Psi}^{(k)}(\theta)]^2$ . The corresponding (partial) GCV estimate is the minimizer of  $V_T(\theta)$  given by

$$V_T(\theta) = \frac{1}{n_{\text{dat}}(T)} \frac{\text{RSS}_T(\theta)}{\left\{ \frac{1}{n_{\text{dat}}(T)} \text{Tr}[\mathbf{I} - \mathbf{A}_{TT}(\theta)] \right\}^2},$$

where  $\mathbf{A}_{TT}$  is the  $TT$ th block of  $\mathbf{A}(\theta)$  and  $\text{RSS}_T(\theta)$  is the residual sum of squares of the observations minus the analysis carried forward to observation space  $(o - a)$  at time  $T$ . The partial UBR estimate is defined analogously. The partial GCV and partial UBR estimates are candidates for future study; however, the number of observations at time  $T$  in the present experiment (164) is probably not enough to estimate all five parameters in a reproducible fashion. The GML (generalized maximum likelihood) estimate for the parametrization employed here, that is, based on factoring out  $\sigma_o^2$ , may be derived following Wahba (1990b); see also Wahba et al. (1994) and Wahba (1985). Calculations with the GML in the present context appear to be more costly than the randomized trace version of the GCV, although this may change as more advanced numerical methods become available. Other parametrizations for maximum likelihood estimates are also available; see, for example, Dee (1995).

## 6. Summary and conclusions

In this paper we have demonstrated an approach to quasi-on-line tuning of multiple weighting, smoothing, and physical parameters in variational data assimilation procedures involving model constraints, via the generalized cross-validation and unbiased risk methods. In the examples of our experiment we have found that physical parameters to which the analysis is sensitive can be tuned along with one or two weighting parameters and a smoothing parameter. We found in our example that the smoothing parameter is equally important as a parameter controlling the trade-off between fit to the data and fit to the forecast (a result that may surprise some) and that a parameter that controls the relative strength of the model as a weak constraint is somewhat tunable, although not as tunable as the other parameters. There are a number of oversimplifications that could be removed in future experiments, including the fact that the forecast error was generated from a known correlation function that was then used in the analysis, the linearity of nature and the model, the simplified penalty when using the model as a weak constraint, and the regularity of the observation points. We think the approach is worth pursuing in more sophisticated experiments and raises a number of issues for further work. The first requirement for tunability is that the analysis be sensitive to the parameters being tuned. Sensitivity analysis in general and for parameter estimation in particular is an important area of ongoing research. A referee has noted that the use of adjoint sensitivity analysis may allow the ranking of the importance of parameters with respect to a given forecast criteria. Adjoint sensitivity analysis is discussed in Rabier et al. (1992), LeDimet et al. (1995), LeDimet et al. (1997), Sun and Yeh (1990a), and elsewhere. See also O'Sullivan (1991). Since sensitivity analyses for certain kinds of parameters require evaluations of the model for each setting of the parameters of interest, further work is needed along the lines of the "design" approach to efficiently study sensitivity. Developments in the papers just mentioned may be useful. In particular, it would be helpful to be able to ascertain which groups of parameters, if any, may be tuned independently of other parameters. We noted that the minimizers in  $(U, \delta)$  are relatively insensitive to changes of the other parameters *within a certain range near the minimum*, but not everywhere. Sensitivity to some other parameters, specifically certain parameters inside covariances, can be partly studied outside of a model and some results along these lines are under development. Of interest are scale lengths and other parameters governing spatial variability inside forecast error or model error covariances.

One of the referees has asked us to discuss the concept of identifiability of physical parameters. "Tunability," as used in this paper, is not exactly the same as the classic notion of identifiability, where the goal is to estimate the "true" value of the unknown (distributed)

parameter, although the two are certainly closely related. In this paper we consider a parameter in the model as tunable, whether it is a physical, smoothing, or weighting parameter, if the analysis, obtained in conjunction with a particular model and particular pattern and error structure of observations, is sensitive to the choice of the parameter. An optimally tuned physical parameter is not necessarily an optimal estimate of the true parameter according to some criteria other than providing an optimal analysis. The best tuned parameter may be a biased estimate of the true parameter while compensating well for model error in the analysis, as discussed in section 3f. Sun and Yeh (1990b) discuss the concept of " $\delta$ -prediction equivalence identifiability," which quantifies parameter "identifiability" in terms of the effect of the parameter on prediction. Although their definition is not exactly the same as tunability, the idea of defining a kind of parameter identifiability in terms of its effect on prediction is relevant here. In Sun and Yeh (1990b) as here the answer generally depends on the model as well as the pattern and error structure of the observations. Other relevant references include Kravaris and Seinfeld (1986), Wolfenbarger and Seinfeld (1991), and McLaughlin and Townley (1966). Note that in this work we have only considered "carefully chosen" parameters that we had previously determined were tunable (via some pilot experimentation) since we were concerned with the development of the tuning method, given tunable parameters.

The question of the accuracy of an analysis using the model as a strong constraint relative to an analysis based on the model as a tuned weak constraint has been raised here, but answered only within the very limited confines of part of this particular experiment, and has to be raised anew within the confines of any particular model, signal or families of signals, and time interval. Other approaches to weakening the strong model constraint in 4D-Var are also available—see Derber (1989), Zupanski (1993), Griffith and Nichols (1996)—and certain of their parameters have the potential of being tuned this way. Another important open question involves indirect data such as satellite radiance data. Radiance data can (and should) be incorporated into the variational problem, even possibly including some unknown parameters in the forward problem, which in principle could be included in the set of parameters to be tuned. However, the error structure of such data may have large biases and be hard to model correctly, and so care must be taken if it is to be included in the cross validation. According to D. Dee (1996, personal communication) "bias estimation is the most underrated problem in data assimilation." Further research along all of these lines is needed.

*Acknowledgments.* We thank Dick Dee and David Houghton for helpful discussions, and Mike Navon, Roger Daley, and Andrew Lorenc for early versions of their manuscripts. This work was supported by NASA

under Grant NAGW 2961, by NSF under Grant DMS9121003, and the NCAR Geophysical Statistics Project under NSF Grant DMS9312686.

## REFERENCES

- Avriel, M., 1976: *Nonlinear Programming: Analysis and Methods*. Prentice Hall, 512 pp.
- Bartello, P., and H. Mitchell, 1992: A continuous three-dimensional model of short-range forecast error covariances. *Tellus*, **44A**, 217–235.
- Bates, D., F. Reames, and G. Wahba, 1993: Getting better contour plots with S and GCVPACK. *Comput. Stat. Data Anal.*, **15**, 329–342.
- Bennett, A., 1992: *Inverse Methods in Physical Oceanography*. Cambridge University Press, 346 pp.
- Bennett, A. F., and R. N. Miller, 1991: Weighting initial conditions in variational assimilation schemes. *Mon. Wea. Rev.*, **119**, 1098–1102.
- , B. Chua, and L. Leslie, 1996: Generalized inversion of a global weather prediction model. *Meteor. Atmos. Phys.*, **60**, 165–178.
- , —, and —, 1997: Generalized inversion of a global weather prediction model, II, analysis and implementation. *Meteor. Atmos. Phys.*, in press.
- Boer, G., 1984: A spectral analysis of predictability and error in an operational forecast system. *Mon. Wea. Rev.*, **112**, 1183–1197.
- Bowman, K., J. Sacks, and Y. Chang, 1993: Design and analysis of numerical experiments. *J. Atmos. Sci.*, **50**, 1267–1278.
- Cohn, S., 1993: Dynamics of short-term univariate forecast error covariances. *Mon. Wea. Rev.*, **121**, 3123–3149.
- , N. Sivakumaran, and R. Todling, 1994: A fixed-lag Kalman smoother for retrospective data assimilation. *Mon. Wea. Rev.*, **122**, 2838–2867.
- Craven, P., and G. Wahba, 1979: Smoothing noisy data with spline functions: Estimating the correct degree of smoothing by the method of generalized cross-validation. *Numer. Math.*, **31**, 377–403.
- Dalcher, A., and E. Kalnay, 1987: Error growth and predictability in operational ECMWF forecasts. *Tellus*, **39A**, 474–491.
- Daley, R., 1992: Estimating model-error covariances for application to atmospheric data assimilation. *Mon. Wea. Rev.*, **120**, 1735–1746.
- Dee, D., 1995: On-line estimation of error covariance parameters for atmospheric data assimilation. *Mon. Wea. Rev.*, **123**, 1128–1145.
- Derber, J., 1989: A variational continuous assimilation technique. *Mon. Wea. Rev.*, **117**, 2437–2466.
- Dongarra, J., J. Bunch, C. Moler, and G. Stewart, 1979: *Linpack Users' Guide*. SIAM, 320 pp.
- Eubank, R., 1988: *Spline Smoothing and Nonparametric Regression*. Marcel Dekker, 438 pp.
- Girard, D., 1989: A fast 'Monte-Carlo cross-validation' procedure for large least squares problems with noisy data. *Numer. Math.*, **56**, 1–23.
- , 1991: Asymptotic optimality of the fast randomized versions of GCV and  $C_L$  in ridge regression and regularization. *Ann. Stat.*, **19**, 1950–1963.
- Golub, G., and C. van Loan, 1989: *Matrix Computations*. 2d ed. The Johns Hopkins University Press, 642 pp.
- , M. Heath, and G. Wahba, 1979: Generalized cross validation as a method for choosing a good ridge parameter. *Technometrics*, **21**, 215–224.
- Green, P., and B. Silverman, 1994: *Nonparametric Regression and Generalized Linear Models*. Chapman and Hall, 182 pp.
- Griffith, A., and N. Nichols, 1996: Accounting for model error in data assimilation using adjoint methods. *Computational Differentiation: Techniques, Applications and Tools, Proc. Second Int. SIAM Workshop on Computational Differentiation*, Santa Fe, NM, SIAM, 105–204.
- Hardin, R., and N. Sloane, 1993: A new approach to the construction of optimal designs. *J. Stat. Planning Inference*, **37**, 339–369.
- Hastie, T., and R. Tibshirani, 1990: *Generalized Additive Models*. Chapman and Hall, 335 pp.
- Hollingsworth, A., and P. Lönnberg, 1986: The statistical structure of short-range forecast errors as determined from radiosonde data. Part I: The wind field. *Tellus*, **38A**, 111–136.
- Hutchinson, M., 1989: A stochastic estimator of the trace of the influence matrix for Laplacian smoothing splines. *Commun. Stat. Part B-Simulation and Computation*, **18**, 1059–1076; Corrigendum, **19**, 433–450.
- Kravaris, C., and J. Seinfeld, 1985: Identification of parameters in distributed parameter systems by regularization. *SIAM J. Control Opt.*, **23**, 217–241.
- , and —, 1986: Identification of spatially varying parameters in distributed parameter systems by discrete regularization. *J. Math. Anal. Appl.*, **119**, 128–152.
- LeDimet, F., H. Ngodock, and I. Navon, 1995: Sensitivity analysis in variational data assimilation. Supercomputer Computations Research Institute Tech. Rep. FSU-SCRI-95T-103, The Florida State University, Tallahassee, FL.
- , H. Ngodock, B. Luong, and J. Verron, 1997: Sensitivity analysis in variational data assimilation. *Data Assimilation in Meteorology and Oceanography: Theory and Practice*, M. Ghil et al., Eds., Universal Academy Press, 135–145.
- Li, K. C., 1986: Asymptotic optimality of  $C_L$  and generalized cross validation in ridge regression with application to spline smoothing. *Ann. Stat.*, **14**, 1101–1112.
- Lorenc, A., 1986: Analysis methods for numerical weather prediction. *Quart. J. Roy. Meteor. Soc.*, **112**, 1177–1194.
- McLaughlin, D., and L. Townley, 1996: A reassessment of the groundwater inverse problem. *Water Resour. Res.*, **32**, 1131–1161.
- Mitchell, H., and R. Daley, 1997a: Discretization error and signal/error correlation in atmospheric data assimilation. Part I: All scales resolved. *Tellus*, **49**, 32–53.
- , and —, 1997b: Discretization error and signal/error correlation in atmospheric data assimilation. Part II: The effect of unresolved scales. *Tellus*, **49**, 54–73.
- O'Sullivan, F., 1991: Sensitivity analysis for regularized estimation in some system identification problems. *SIAM J. Sci. Start. Comput.*, **12**, 1266–1283.
- Parrish, D., and J. Derber, 1992: The National Meteorological Center's spectral statistical interpolation analysis system. *Mon. Wea. Rev.*, **120**, 1747–1763.
- Press, W., B. Flannery, S. Teukolsky, and W. Vetterling, 1990: *Numerical Recipes (Fortran Version)*. Cambridge University Press, 702 pp.
- Rabier, F., P. Courtier, and O. Talagrand, 1992: An application of adjoint models to sensitivity analysis. *Beitr. Phys. Atmos.*, **65**, 177–192.
- , —, J. Pailleux, O. Talagrand, J. Thépaut, and D. Vasiljevic, 1993: A comparison between four-dimensional variational assimilation and simplified sequential assimilation relying on three-dimensional analysis. *Quart. J. Roy. Meteor. Soc.*, **119**, 845–880.
- Sun, N., and W. Yeh, 1990a: Coupled inverse problems in groundwater modeling. 1. Sensitivity analysis and parameter identification. *Water Resour. Res.*, **26**, 2507–2525.
- , and —, 1990b: Coupled inverse problems in groundwater modeling. 2. Identifiability and experimental design. *Water Resour. Res.*, **26**, 2527–2540.
- Thépaut, J., R. Hoffman, and P. Courtier, 1993: Interactions of dynamics and observations in a four-dimensional variational assimilation. *Mon. Wea. Rev.*, **121**, 3393–3414.
- Todling, R., and S. Cohn, 1994: Suboptimal schemes for atmospheric data assimilation based on the Kalman filter. *Mon. Wea. Rev.*, **122**, 2530–2557.
- Tribbia, J., and D. Baumhefner, 1988: The reliability of improvements in deterministic short-range forecasts in the presence of initial

- state and modeling deficiencies. *Mon. Wea. Rev.*, **116**, 2276–2288.
- Wahba, G., 1968: On the distribution of some statistics useful in the analysis of jointly stationary time series. *Ann. Math. Stat.*, **39**, 1849–1862.
- , 1978: Interpolating surfaces: High order convergence rates and their associated designs, with applications to x-ray image reconstruction, Tech. Rep. 523, Dept. of Statistics, University of Wisconsin, Madison, WI.
- , 1982a: Variational methods in simultaneous optimum interpolation and initialization. *The Interaction Between Objective Analysis and Initialization*, D. Williamson, Ed., Atmospheric Analysis and Prediction Division, NCAR, 178–185.
- , 1982b: Vector splines on the sphere, with application to the estimation of vorticity and divergence from discrete, noisy data. *Multivariate Approximation Theory*, W. Schempp and K. Zeller, Eds., Vol. 2, Birkhauser Verlag, 407–429.
- , 1985: A comparison of GCV and GML for choosing the smoothing parameter in the generalized spline smoothing problem. *Ann. Stat.*, **13**, 1378–1402.
- , 1990a: Regularization and cross validation methods for nonlinear, implicit, ill-posed inverse problems. *Geophysical Data Inversion Methods and Applications*, A. Vogel, C. Ofoegbu, R. Gorenflo, and B. Ursin, Eds., Vieweg, Weisbaden-Braunschweig, 3–13.
- , 1990b: *Spline Models for Observational Data*. CBMS-NSF Regional Conference Series in Applied Mathematics, Vol. 59, SIAM, vii and 169 pp.
- , and J. Wendelberger, 1980: Some new mathematical methods for variational objective analysis using splines and cross-validation. *Mon. Wea. Rev.*, **108**, 1122–1145.
- , D. Johnson, F. Gao, and J. Gong, 1994: Adaptive tuning of numerical weather prediction models: Part I, randomized GCV and related methods in three and four dimensional data assimilation. Tech. Rep. 920, Department of Statistics, University of Wisconsin, Madison, WI.
- , —, —, and —, 1995: Adaptive tuning of numerical weather prediction models: Randomized GCV in three and four dimensional data assimilation. *Mon. Wea. Rev.*, **123**, 3358–3369.
- Wergen, W., 1992: The effect of model errors in variational assimilation. *Tellus*, **44A**, 297–313.
- Wolfenbarger, K., and J. Seinfeld, 1991: Regularized solutions to the aerosol data inversion problem. *SIAM J. Sci. Stat. Comput.*, **12**, 342–361.
- Zou, X., I. Navon, and F-X. LeDimet, 1992: Incomplete observations and control of gravity waves in variational data assimilation. Part II: Applications and numerical results. *Tellus*, **44A**, 297–313.
- , —, and J. Sela, 1993: Control of gravitational oscillations in variational data assimilation. *Mon. Wea. Rev.*, **121**, 272–289.
- Zupanski, M., 1993: Regional four-dimensional variational data assimilation in a quasi-operational forecasting environment. *Mon. Wea. Rev.*, **121**, 2396–2408.