Tellus (1996), 48A, 221–237
Printed in Belgium – all rights reserved

TELLUS
ISSN 0280–6495

The application of Kalman smoother theory to the estimation of 4DVAR error statistics

By RICHARD MÉNARD*, *Department of Atmospheric and Oceanic Sciences, McGill Univeristy, Montréal, Québec, Canada* and ROGER DALEY[1], *Atmospheric Environment Service, Toronto, Ontario, Canada*

(Manuscript received 16 September 1994; in final form 10 July 1995)

ABSTRACT

Modern atmospheric data assimilation theory is dominated by the four-dimensional variational (4DVAR) and Kalman filter/smoother approaches. Both generate analysis weights (explicitly or implicitly) which are dynamically determined by the assimilation model. A Kalman smoother is basically a generalization of the Kalman filter which can process future observations. In control theory, a generalization of 4DVAR called Pontryagin optimization can account for an imperfect assimilating model. Pontryagin optimization and the fixed-interval Kalman smoother are equivalent when both methods use the same statistical information. We use the equivalence between Pontryagin optimization and the Kalman smoother to examine the effect of the perfect model assumption on the error statistics and analysis weights of the 4DVAR algorithm. This is done by developing the Kalman smoother equations for a a very simple assimilating model. A procedure for diagnosing the effect of model error, based on the observational cost function, is also developed.

1. Introduction

Atmospheric data assimilation requires an atmospheric forecast model for advancing in time, forward (observation) models for each of the instruments and detailed knowledge of forecast and observation error statistics to optimally weight forecast and observations. Inaccurate atmospheric and forward models as well as mis-specified error statistics all lead to analysis degradation, but it is the role of the error statistics which has been least studied and remains the most problematic.

We distinguish 3 types of error statistics: *observation errors* associated with each instrument, *model errors* which are due to discretization and/or parameterization errors in the atmospheric model and *forecast* errors which are a complex function of model and observation errors as well as the spatial/temporal configuration of the observing system.

In the optimal interpolation (OI) and 3-dimensional variational (3DVAR) algorithms, the forecast error statistics are usually specified to be stationary, homogeneous and (perhaps) isotropic. Estimations of the forecast and observation errors have traditionally been obtained by the *innovation technique* (Hollingsworth and Lönnberg, 1986; Daley, 1991, chs. 4 and 5). This procedure used the North American radiosonde network and a few simple assumptions to disentangle observation and forecast error and was quite adequate for optimum interpolation (OI) algorithms. However, this method was not adequate for the more advanced three dimensional variational (3DVAR) algorithm (which required global statistics), because the radiosonde network is badly distributed with respect to the global scales and effective

[1] Corresponding author: Roger Daley, Naval Research Laboratory, 7 Grace Hopper Avenue, Monterey, CA 93943-5502, USA.

* Present affiliation: Data Assimilation Office, Code 910.3, NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA.

use of satellite innovation statistics remains elusive. Consequently, the present operational procedures for obtaining 3DVAR forecast error statistics are based on an unproven relation between forecast errors and forecast differences (Parrish and Derber, 1992; Bouttier, 1994).

In the more advanced assimilation techniques (four-dimensional variational, 4DVAR, and the extended Kalman filter, EKF), the analysis weights are determined dynamically and are flow dependent. These weights depend explicitly on the forecast error covariances for the EKF and implicitly for the 4DVAR algorithm. Propagation of forecast errors for these algorithms has been studied by Monte Carlo methods (Molteni and Palmer, 1993) and through the use of the tangent linear (TLM) approximation to the forecast model (Gauthier et al., 1992; Cohn, 1993; Ménard, 1994). However, as noted above, forecast error depends on model error, and while forecast error is not very well understood or estimated; the situation with respect to model error is much worse. In fact, with a few exceptions (Dee, 1995), very little work has been done on model error statistics. In the absence of such statistical information, data assimilation procedures such as 4DVAR, simply avoid the problem by assuming that the forecast model is perfect.

Little is known about the impact of the perfect model assumption. For advanced sequential data assimilation methods such as the Kalman filter, the perfect model assumption may lead to filter divergence (Jazwinski, 1970; Ménard et al., 1995), where valuable data are given progressively smaller weights, eventually leading to data rejection. For fixed-interval data assimilation algorithms, such as 4DVAR, the effect of the perfect model assumption has received relatively little attention, apart from the studies of Wergen (1993) and Zupanski (1993).

The purpose of this paper is to study the behaviour of fixed-interval data assimilation systems with respect to specifications or assumptions about model error statistics. This will be done by exploiting the equivalence between the *Kalman smoother* (a generalization of the Kalman filter which can assimilate future observations, as well as past and present observations) and its dual control problem, the *Pontryagin optimization* or generalized inverse algorithm (which can be considered as a generalization of 4DVAR, to allow for an imperfect model). Since our emphasis is on the impact of model error statistics, rather than on atmospheric dynamics, and because second moment statistical calculations with very large state space dimensions are presently impractical, a simplified atmospheric model has been used. With this approach, exact covariance calculations can sometimes be carried out analytically, enhancing understanding of the problem.

The organization of the paper is as follows. In Section 2, the Kalman smoother is derived by optimally combining the standard Kalman filter with a reverse-time information filter. Pontryagin optimization (and its equivalence with the fixed-interval Kalman smoother) is introduced in Section 4. In Section 3, we describe a very simple assimilating model and derive the corresponding Kalman smoother equations. Analytic results can be obtained with this model (following Daley and Ménard (1993)) and are particularly useful in understanding the role of future observations in atmospheric data assimilation. Finally, in Section 5, we will use the fixed-interval Kalman smoother to examine the effect of statistical misspecification (in particular, an inappropriate perfect model assumption), and practical performance diagnostics for the 4DVAR algorithm.

2. Obtaining the Kalman smoother equations by information filter

The Kalman smoother is derived in several texts on estimation theory such as Gelb (1974). It has also been discussed in the oceanographic literature by Wunsch (1988), Gaspar and Wunsch (1989), Bennett and Budgell (1989) and Moiseenko and Saenko (1992). A derivation of the *fixed-lag* Kalman smoother by Cohn et al. (1994) has recently appeared in the meteorological literature. In this section, we will introduce a quite different derivation of the Kalman smoother by simply extending familiar Kalman filter theory (Daley, 1991, Subsection 13.3) to the assimilation of future observations. The Kalman smoother is mathematically complex and for reasons of economy, the present derivation is condensed.

This derivation of the Kalman smoother, based on Fraser and Potter (1969), has 3 steps: a Kalman filter for the forward integration, a reverse-time information filter for the backward integration

from future data and a procedure for optimally combining the 2 estimates. The result is a minimum variance estimate based on past, present and future observations. This derivation is not intended to lead to an efficient algorithm. Other Kalman smoother algorithms are more efficient, the RTS algorithm (Rauch et al., 1965), the representer algorithm (Bennett, 1992) or the retrospective algorithm (Cohn et al., 1994).

2.1. *The forward (Kalman) filter*

The forward part of the filter consists of the standard Kalman filter equations described in Daley (1992a). We define a regular time-independent analysis grid r_j, $1 \leqslant j \leqslant J$ and an irregular time-dependent observation network r_i, $1 \leqslant i \leqslant I_n$ (where n indicates time t_n). The 2nd moment error statistics (on the analysis grid) are governed by the error covariance propagation equation and the analysis error covariance equation,

$$\boldsymbol{P}^{\mathrm{f}}_{n+1} = \boldsymbol{M}\boldsymbol{P}^{\mathrm{a}}_{n}\boldsymbol{M}^{\mathrm{T}} + \boldsymbol{Q}_n, \tag{2.1}$$

$$\begin{aligned}\boldsymbol{P}^{\mathrm{a}}_{n} &= [\boldsymbol{I} - \boldsymbol{K}_n\boldsymbol{H}_n]\,\boldsymbol{P}^{\mathrm{f}}_{n} \\ &= [[\boldsymbol{P}^{\mathrm{f}}_{n}]^{-1} + \boldsymbol{H}^{\mathrm{T}}_{n}\boldsymbol{R}^{-1}_{n}\boldsymbol{H}_n]^{-1},\end{aligned} \tag{2.2}$$

where

$$\begin{aligned}\boldsymbol{K}_n &= \boldsymbol{P}^{\mathrm{f}}_{n}\boldsymbol{H}^{\mathrm{T}}_{n}[\boldsymbol{H}_n\boldsymbol{P}^{\mathrm{f}}_{n}\boldsymbol{H}^{\mathrm{T}}_{n} + \boldsymbol{R}_n]^{-1} \\ &= \boldsymbol{P}^{\mathrm{a}}_{n}\boldsymbol{H}^{\mathrm{T}}_{n}\boldsymbol{R}^{-1}_{n},\end{aligned} \tag{2.3}$$

is the optimal (Kalman) gain. $\boldsymbol{M}_n$ is the (linear) assimilation model at time t_n and $\boldsymbol{H}_n$ is the (linear) forward interpolation operator from the analysis grid to the observation network. $\boldsymbol{P}^{\mathrm{f}}_{n}$ and $\boldsymbol{P}^{\mathrm{a}}_{n}$ are the forecast and analysis error covariances, $\boldsymbol{R}_n$ is the observation error covariance and $\boldsymbol{Q}_n$ is the model error covariance between t_n and t_{n+1} which must be specified. The corresponding first moment equations for the forecast and analyzed state variables are given by,

$$\boldsymbol{x}^{\mathrm{f}}_{n+1} = \boldsymbol{M}_n\boldsymbol{x}^{\mathrm{a}}_{n}, \tag{2.4}$$

$$\boldsymbol{x}^{\mathrm{a}}_{n} - \boldsymbol{x}^{\mathrm{f}}_{n} = \boldsymbol{K}_n[\boldsymbol{z}_n - \boldsymbol{H}_n\boldsymbol{x}^{\mathrm{f}}_{n}], \tag{2.5}$$

where $\boldsymbol{x}^{\mathrm{a}}_{n}$, $\boldsymbol{x}^{\mathrm{f}}_{n}$ are the vectors of analysis and forecast values at t_n on the analysis grid. $\boldsymbol{z}_n$ is the vector of observations.

2.2. *The reverse-time information filter*

The reverse-time information filter can be thought of as a Kalman filter which runs backward in time. It is derived in the same way as the Kalman filter in Daley (1992a, Section 2). Thus, the true state vector on the analysis grid at time t_n can be obtained from the corresponding vector at t_{n-1} by,

$$\boldsymbol{x}_n = \boldsymbol{M}_{n-1}\boldsymbol{x}_{n-1} + \boldsymbol{e}^{\mathrm{q}}_{n-1}, \tag{2.6}$$

where $\boldsymbol{e}^{\mathrm{q}}_{n}$ is the model error growth vector between t_{n-1} and t_n and $\boldsymbol{x}_n$ is the "true" state vector at t_n. Suppose for the moment that the model can be run backward in time, that is we can define $\boldsymbol{M}^{-1}_{n}$. (The Kalman smoother equations can be derived without this assumption). Then,

$$\boldsymbol{x}_{n-1} = \boldsymbol{M}^{-1}_{n-1}[\boldsymbol{x}_n - \boldsymbol{e}^{\mathrm{q}}_{n-1}]. \tag{2.7}$$

Now suppose that we are integrating the model backwards in this fashion from some future analyzed state. We introduce the superscript "β" to indicate a *hindcast* (i.e., backward forecast) from a future time. We use the superscript "α" to indicate an analysis which is produced from *present* and *future data only*. Thus, corresponding to the forecast eq. (2.4), there is a hindcast equation,

$$\boldsymbol{x}^{\beta}_{n-1} = \boldsymbol{M}^{-1}_{n-1}\boldsymbol{x}^{\alpha}_{n}. \tag{2.8}$$

Substracting (2.7) from (2.6) gives,

$$\boldsymbol{e}^{\beta}_{n-1} = \boldsymbol{M}^{-1}_{n-1}[\boldsymbol{e}^{\alpha}_{n} + \boldsymbol{e}^{\mathrm{q}}_{n-1}], \tag{2.9}$$

where $\boldsymbol{e}^{\beta}_{n} = \boldsymbol{x}^{\beta}_{n} - \boldsymbol{x}_n$ and $\boldsymbol{e}^{\alpha}_{n} = \boldsymbol{x}^{\alpha}_{n} - \boldsymbol{x}_n$ are the hindcast and analysis errors. Right multiplication by $(\boldsymbol{e}^{\beta}_{n-1})^{\mathrm{T}}$ and application of the expectation operator $\langle\ \rangle$ gives,

$$\boldsymbol{P}^{\beta}_{n-1} = \boldsymbol{M}^{-1}_{n-1}[\boldsymbol{P}^{\alpha}_{n} + \boldsymbol{Q}_{n-1}][\boldsymbol{M}^{-1}_{n-1}]^{\mathrm{T}}, \tag{2.10}$$

where $\boldsymbol{P}^{\beta}_{n} = \langle \boldsymbol{e}^{\beta}_{n}(\boldsymbol{e}^{\beta}_{n})^{\mathrm{T}} \rangle$, $\boldsymbol{P}^{\alpha}_{n} = \langle \boldsymbol{e}^{\alpha}_{n}(\boldsymbol{e}^{\alpha}_{n})^{\mathrm{T}} \rangle$ are the hindcast and reverse time analysis error covariances. $\boldsymbol{Q}_n = \langle \boldsymbol{e}^{\mathrm{q}}_{n}(\boldsymbol{e}^{\mathrm{q}}_{n})^{\mathrm{T}} \rangle$ is the model error covariance (2.1). We have assumed (as in the derivation of eqs. 2.1–5) that the model error is unbiased, not serially correlated and not mutually correlated with the observation error.

Eqs. (2.8 and 2.10) are not in a very desirable form because of the appearance of $\boldsymbol{M}^{-1}_{n-1}$.

However, these equations can be written in *information form*, which involves $\boldsymbol{M}_n^{\mathrm{T}}$ instead of $\boldsymbol{M}_n^{-1}$. For any error covariance matrix $\boldsymbol{P}$, we define a corresponding *information matrix* $\boldsymbol{N}=\boldsymbol{P}^{-1}$. We also introduce a new state variable $\boldsymbol{y}=\boldsymbol{N}\boldsymbol{x}=\boldsymbol{P}^{-1}\boldsymbol{x}$. When the error becomes large, then $\boldsymbol{y}$ and $\boldsymbol{N}$ approach zero and there is said to be no information.

Substitution of $\boldsymbol{N}$ into eqs. (2.8 and 2.10) gives,

$$\begin{aligned}\boldsymbol{y}_{n-1}^{\beta}&=\boldsymbol{N}_{n-1}^{\beta}\boldsymbol{M}_{n-1}^{-1}[\boldsymbol{N}_n^{\alpha}]^{-1}\boldsymbol{y}_n^{\alpha}\\&=\boldsymbol{M}_{n-1}^{\mathrm{T}}[\boldsymbol{I}+\boldsymbol{N}_n^{\alpha}\boldsymbol{Q}_{n-1}]^{-1}\boldsymbol{y}_n^{\alpha},\\\boldsymbol{N}_{n-1}^{\beta}&=\boldsymbol{M}_{n-1}^{\mathrm{T}}[\boldsymbol{I}+\boldsymbol{N}_n^{\alpha}\boldsymbol{Q}_{n-1}]^{-1}\boldsymbol{N}_n^{\alpha}\boldsymbol{M}_{n-1}.\end{aligned}\tag{2.11}$$

Note that eqs. (2.11) involve $\boldsymbol{M}$ and $\boldsymbol{M}^{\mathrm{T}}$ but not $\boldsymbol{M}^{-1}$.

We also have to derive the information form of the analysis equations for the reverse-time filter corresponding to equations (2.2 and 2.5). Thus, the appropriate equations for $\boldsymbol{N}_n^{\alpha}$ and $\boldsymbol{y}_n^{\alpha}$ are,

$$\begin{aligned}\boldsymbol{N}_n^{\alpha}&=\boldsymbol{N}_n^{\beta}+\boldsymbol{H}_n^{\mathrm{T}}\boldsymbol{R}_n^{-1}\boldsymbol{H}_n,\\\boldsymbol{y}_n^{\alpha}&=\boldsymbol{y}_n^{\beta}+\boldsymbol{H}_n^{\mathrm{T}}\boldsymbol{R}_n^{-1}\boldsymbol{z}_n.\end{aligned}\tag{2.12}$$

The backward filter is started at some time t_N, by setting $\boldsymbol{y}_N^{\beta}$ and $\boldsymbol{N}_N^{\beta}$ to zero, implying that there is no information for $t>t_N$. It might be noted, that unlike the Kalman filter, the reverse-time information filter has a relatively simple analysis step (2.12); but the hindcast step (2.11) is more complex.

2.3. *The smoothing step*

At any time t_n, there are now two estimates of the state vector, one obtained from past information using the Kalman filter and one obtained from future information using the reverse-time information filter. The smoothing step combines these two estimates optimally by minimum variance estimation. There are three ways to do this. In this treatment, we combine $\boldsymbol{x}_n^{\mathrm{a}}$ (which uses the observations at t_n) with $\boldsymbol{y}_n^{\beta}$ (which does not). Alternatively, we could combine $\boldsymbol{x}_n^{\mathrm{f}}$ with $\boldsymbol{y}_n^{\alpha}$, or $\boldsymbol{x}_n^{\mathrm{f}}$ with $\boldsymbol{y}_n^{\beta}$ and $\boldsymbol{z}_n$. In all these procedures, the observations are only used once. We will denote the smoothed state at time t_n as $\boldsymbol{x}_n^{\mathrm{s}}$ and the corresponding error covariance as $\boldsymbol{P}_n^{\mathrm{s}}$. The optimal combination (see Daley, 1991, chapter 2) is given by,

$$\begin{aligned}&[\boldsymbol{P}_n^{\mathrm{s}}]^{-1}=[\boldsymbol{P}_n^{\mathrm{a}}]^{-1}+\boldsymbol{N}_n^{\beta},\\&\text{or}\quad\boldsymbol{P}_n^{\mathrm{s}}=[\boldsymbol{I}+\boldsymbol{P}_n^{\mathrm{a}}\boldsymbol{N}_n^{\beta}]^{-1}\boldsymbol{P}_n^{\mathrm{a}},\\&\boldsymbol{x}_n^{\mathrm{s}}=[\boldsymbol{I}+\boldsymbol{P}_n^{\mathrm{a}}\boldsymbol{N}_n^{\beta}]^{-1}[\boldsymbol{x}_n^{\mathrm{a}}+\boldsymbol{P}_n^{\mathrm{a}}\boldsymbol{y}_n^{\beta}].\end{aligned}\tag{2.13}$$

The complete Kalman smoother equations then consist of the forward Kalman filter (2.1–5), the reverse-time information filter (2.11–12) and the optimal combination of the two estimates (2.13).

3. Application to a simple assimilating model

Before introducing Pontryagin optimization and its relation to the Kalman smoother derived in Section 2, we will illustrate the Kalman smoother mechanics by application to a very simple assimilating system. The assimilating model is an exact semi-Lagrangian discretization of the one-dimensional linear advective/diffusive equation. Daley and Ménard (1993), herafter referred to as DM93, examined a number of aspects of the Kalman filter with this system using a *maximum network* in which there was an observation station collocated with every analysis gridpoint. With this system, we can obtain analytic solutions which are used to examine the role of future observations and model error in data assimilation. While this model is much simpler than present day forecast models, it is ideal for the study of error statistics.

3.1. *The assimilation system*

The assimilating model is based on the equation,

$$\frac{\partial h}{\partial t}+U\frac{\partial h}{\partial x}-\nu\frac{\partial^2 h}{\partial x^2}=0,\tag{3.1}$$

where U is a constant advection velocity, ν is a dissipation coefficient and h is (say) the geopotential height. Equation (3.1) is applied over the one dimensional periodic domain $-\pi a\leqslant x\leqslant\pi a$, and $h(-\pi a,t)=h(\pi a,t)$. The model and analysis gridpoints are defined as $x_j=-\pi a+(j-1)\,\Delta x$, $1\leqslant j\leqslant J$, where J is the number of gridpoints and $\Delta x=2\pi a/J$. Here $a=2500$ km and $J=49$, corresponding to $\Delta x=320$ km. We assume that the analysis grid and the observation network coincide and the forecasts and observations are of the same variable ($\boldsymbol{H}_n$ is identity matrix). We also assume that $\boldsymbol{R}_n$, $\boldsymbol{Q}_n$ are stationary and homogeneous. Equation (3.1) is discretized as $\boldsymbol{M}=\boldsymbol{F}\hat{\boldsymbol{M}}\boldsymbol{F}^{\mathrm{T}}$, where

$\boldsymbol{F}$ is a Fourier transform matrix and $\hat{\boldsymbol{M}}$ is a block diagonal matrix with 2×2 blocks $\hat{\boldsymbol{m}}(p)$ for Fourier wavenumber p, $1 \leqslant p \leqslant (J-1)/2$, given by DM93 (equation 2.8). Defining $m^2(p) = \exp(-2\nu\,\Delta t p^2/a^2)$ (Δt is the insertion interval), then $\hat{m}^{-1} = m^{-2}(p)\,\hat{m}^{\mathrm{T}}$. Following DM93, the Kalman smoother equations (Section 2) can then be diagonalized. Define scalars q^2, r^2, a_n^2, f_n^2, β_n^{-2}, α_n^{-2} and s_n^2 as spectral forms for wavenumber p corresponding to the covariances Q, R, P_n^{a}, $\boldsymbol{P}_n^{\mathrm{f}}$, $\boldsymbol{N}_n^{\beta}$, $\boldsymbol{N}_n^{\alpha}$ and $\boldsymbol{P}_n^{\mathrm{s}}$. Define $\hat{\boldsymbol{x}}_n^{\mathrm{f}}$, $\hat{\boldsymbol{x}}_n^{\mathrm{a}}$, $\hat{\boldsymbol{z}}_n$, $\hat{\boldsymbol{y}}_n^{\beta}$, $\hat{\boldsymbol{y}}_n^{\alpha}$ and $\hat{\boldsymbol{x}}_n^{\mathrm{s}}$ as the vectors of length 2 (the cosine and sine components of wavenumber p) corresponding to $\boldsymbol{x}_n^{\mathrm{f}}$, $\boldsymbol{x}_n^{\mathrm{a}}$, $\boldsymbol{z}_n$, $\boldsymbol{y}_n^{\beta}$, $\boldsymbol{y}_n^{\alpha}$ and $\boldsymbol{x}_n^{\mathrm{s}}$. Also define scalar $\hat{k}_n$ corresponding to Kalman gain $\boldsymbol{K}_n$. Then, the spectral form of the Kalman filter equations (2.1–5) is,

$$
\begin{aligned}
&f_{n+1}^2 = m^2 a_n^2 + q^2, \qquad a_n^2 = r^2 f_n^2/[r^2 + f_n^2],\\
&\hat{k}_n = f_n^2/[f_n^2 + r^2],\\
&\boldsymbol{x}_{n+1}^{\mathrm{f}} = \hat{\boldsymbol{m}}\hat{\boldsymbol{x}}_n^{\mathrm{a}}, \qquad \hat{\boldsymbol{x}}_n^{\mathrm{a}} = \hat{\boldsymbol{x}}_n^{\mathrm{f}} + \hat{k}_n[\hat{\boldsymbol{z}}_n - \hat{\boldsymbol{x}}_n^{\mathrm{f}}].
\end{aligned}
\tag{3.2}
$$

The spectral form of the reverse-time filter (2.11–12) is,

$$
\begin{aligned}
&\hat{\boldsymbol{y}}_{n-1}^{\beta} = \hat{\boldsymbol{m}}^{\mathrm{T}}[1 + \alpha_n^{-2} q^2]^{-1}\,\hat{\boldsymbol{y}}_n^{\alpha},\\
&\beta_{n-1}^{-2} = m^2[1 + \alpha_n^{-2} q^2]^{-1}\,\alpha_n^{-2},\\
&\alpha_n^{-2} = \beta_n^{-2} + r^{-2}, \qquad \hat{\boldsymbol{y}}_n^{\alpha} = \hat{\boldsymbol{y}}_n^{\beta} + r^{-2}\hat{\boldsymbol{z}}_n.
\end{aligned}
\tag{3.3}
$$

The spectral form of the smoothing equations (2.13) is,

$$
\begin{aligned}
&s_n^2 = a_n^2/[1 + a_n^2\beta_n^{-2}],\\
&\hat{\boldsymbol{x}}_n^{\mathrm{s}} = [\hat{\boldsymbol{x}}_n^{\mathrm{a}} + a_n^2\hat{\boldsymbol{y}}_n^{\beta}]/[1 + a_n^2\beta_n^{-2}].
\end{aligned}
\tag{3.4}
$$

Each wavenumber is decoupled and there is no propagation of covariances in this system. Following DM93, it is possible to obtain analytic solutions for the second moment equations of (3.2–4), thus obtaining the smoother error covariance at any time from knowledge of the observation, model and initial ($t = 0$) forecast error covariances.

3.2. *Analysis error variance for the fixed-interval smoother*

The second moment equations from (3.2–4) can be used to analyze the error variances to be expected from the Kalman smoother. We consider the case of fixed time interval $t_0 \leqslant t \leqslant t_N = t_0 + N\Delta t$, where $N = 7$. There is no information after t_N, so that $\boldsymbol{N}_N^{\beta} = 0$ and thus $\beta_N^{-2}(p)$ is also zero for all p. The model and observation error covariances are defined the same way as in DM93. Thus, when the model is imperfect, the model error covariance $\boldsymbol{Q}$ is homogeneous and stationary and its spectrum takes the form $q^2(p) = c(1 + p^2 l_q^2/a^2)^{-2}$, where $l_q = a/6 = 417$ km is a correlation length scale and c is a constant required to make the total model error variance $(E^q)^2 = 100\ \mathrm{m}^2/6\ \mathrm{h}$. The observation error is assumed to be uncorrelated with variance $(E^{\mathrm{r}})^2 = 100\ \mathrm{m}^2$. The red model and white observation error spectra are shown in DM93 (Fig. 2a). The Courant number is specified as $U\,\Delta t/\Delta x = 0.5$, and the settings for the dissipation parameter ν will be discussed subsequently. The second moment statistics are independent of the Courant number.

In Fig. 1a–c, we plot the rms height error (m) as a function of insertion time for 3 cases: (a) perfect, inviscid, (b) imperfect, inviscid and (c) perfect, viscous ($\nu\,\Delta t/a^2 = 0.001$). The dashed curve is the rms observation error (10 m). Curve "a" is the Kalman filter analysis error, curve "b" is the hindcast (not the analysis) error of the reverse-time information filter and curve "s" is the smoother estimate. In these 3 panels, there was no information prior to t_0, that is $\boldsymbol{P}_0^{\mathrm{f}}$ is infinite or $f_0^{-2}(p) = 0$.

Referring first to panel (a), it can be seen that the smoother error and Kalman filter error are the same at t_7. At t_0, the smoother error is much smaller than the Kalman filter error and slightly smaller than the hindcast error, because the hindcast does not include observations at t_0. The smoother error is constant over the interval and is equal to $(N+1)^{-1/2}\,E^{\mathrm{r}}$ (see Daley, 1992, ch. 13, problem 13.3). It is always smaller than either the hindcast error or the Kalman filter error. In panel (b) the effect of the imperfect model is to increase the smoother error and produce a flat minimum in the center of the interval. The effect of diffusion in panel (c) is to produce an asymmetry in the smoother with a maximum error at $t = 0$.

Experiments were also performed in which *there was information prior to* t_0. We assumed $\boldsymbol{P}_0^{\mathrm{f}} = \boldsymbol{Q}$. In Fig. 1d is shown the smoother curves only for the inviscid perfect model ($\boldsymbol{Q} = 0$) and the inviscid, imperfect case ($\boldsymbol{Q} \neq 0$), which can be compared with the corresponding "s" curves of panels (a, b). It is clear that the proper use of information prior to the beginning of the interval decreases the

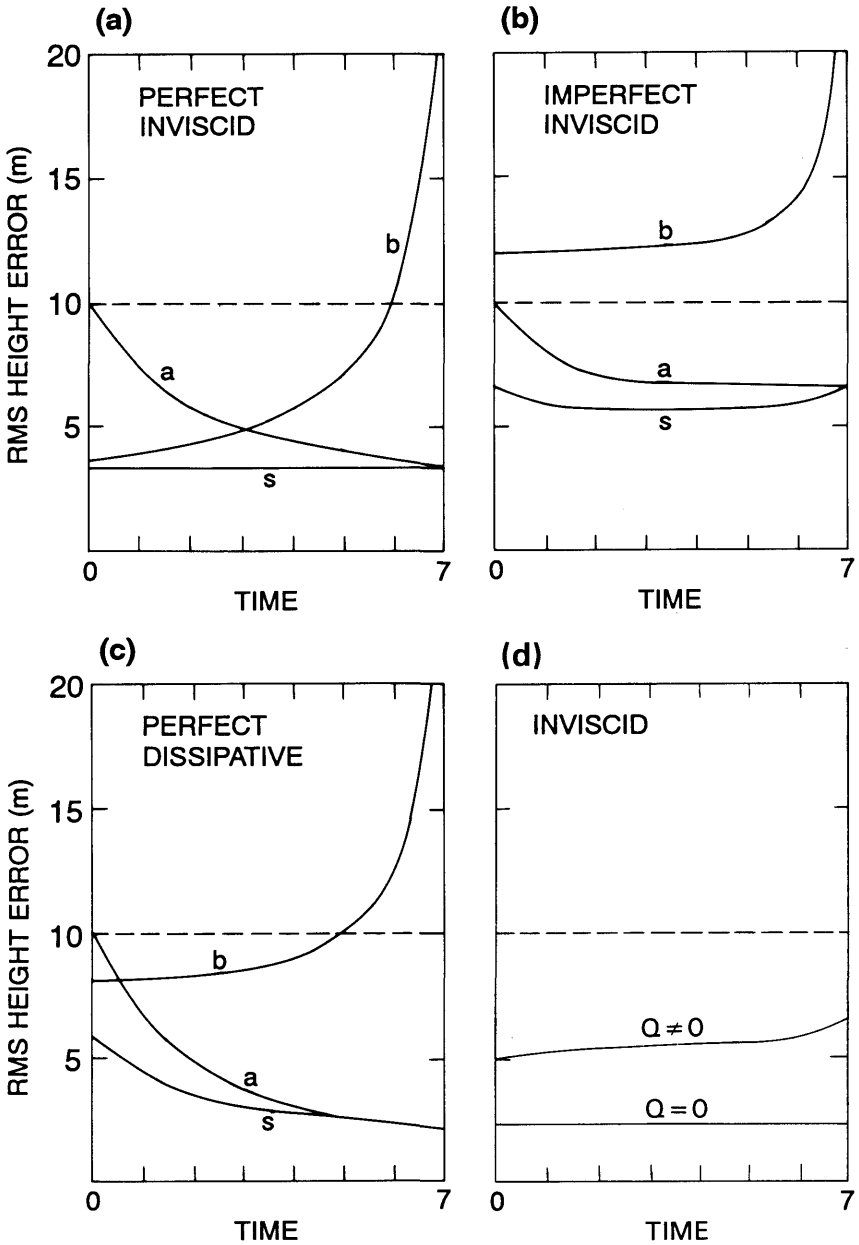

Fig. 1. Plots of the Kalman filter analysis error (curve a), the hindcast error (curve b) and the smoother error (curve s) as a function of time during the fixed interval $t_0 \leqslant t \leqslant t_N$ with no observations prior to t_0. (a) Perfect ($Q = 0$) and inviscid ($\nu = 0$), (b) imperfect and inviscid and (c) perfect and dissipative. Panel (d) shows the smoother error variance only *when there is information prior to* t_0 corresponding to panel (a) $Q = 0$ and (b) $Q \neq 0$.

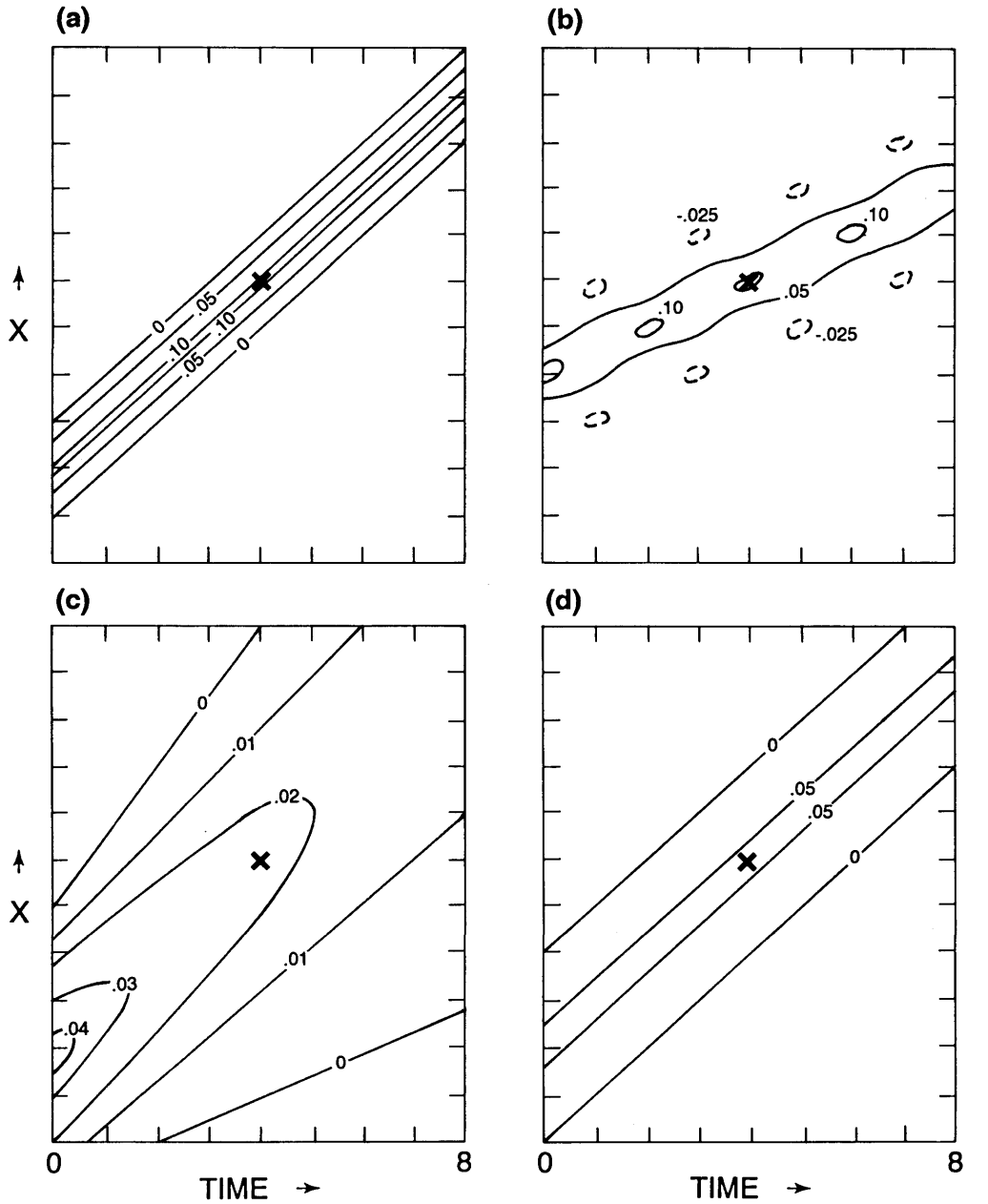

Fig. 2. Weights given to observations in the analysis of the point marked with an "x" for a perfect ($Q=0$) model. Observations are located at different values of space (ordinate) and time (abscissa). Panels (a, b and d) are inviscid, panel (c) is viscous. Panel (d) has information prior to ($t=0$). The Courant number equals 1 except in panel (b) where it equals $\frac{1}{2}$.

smoother error over the whole time interval for the perfect model, but affects only the beginning of the time interval for the imperfect model.

The effects of model error, dissipation and future observations which are apparent in Fig. 1, can be clearly understood from the following simple analysis. Suppose at time t_n, we have produced an analysis using observations for time $t \leqslant t_n$ with analysis error variance $a_n^2(p)$ for wavenumber p. This could have been achieved by any assimilating procedure which used only past and present observations. Now suppose that there are observations at time t_{n+1} with error $r^2(p)$ and integrate the reverse-time filter from t_{n+1} to t_n assuming no information for $t > t_{n+1}$ (i.e., $\beta_{n+1}^{-2}(p) = 0$). Then, using the second moment spectral equations from (3.2–4), it is easily shown that the smoother error at t_n is,

$$s_n^2(p) = a_n^2(p) \frac{r^2(p) + q^2(p)}{r^2(p) + q^2(p) + m^2(p)\, a_n^2(p)}. \quad (3.5)$$

From (3.5), it can be seen that future observations will always improve the analysis ($s_n^2(p) \leqslant a_n^2(p)$) in this optimal system. The improvement will be greatest for a perfect model ($q^2(p) = 0$), or for perfect future observations ($r^2(p) = 0$), or if the model is inviscid or unstable ($m^2(p) \geqslant 1$). Note that $m^2(p)$ becomes small for large dissipation ν or small spatial scales (p large). Thus, future observations provide little information about present states for small scales in dissipative systems.

3.3. *Smoother weights for perfect and imperfect models*

The effect of the perfect model assumption can be looked at in another way by examining the observation weights using equations (3.2–4). We first consider a very simple case, where the solution can be obtained analytically. Assume that there are only 3 observation times t_{n-1}, t_n, t_{n+1} with observation vectors z_{n-1}, z_n, z_{n+1} (with no information prior to t_{n-1} or after t_{n+1}). Then, application of both 1st and 2nd moment eqs. (3.2–4) yields (after some manipulation),

$$\hat{x}_n^s = \frac{\begin{pmatrix} (r^2+q^2)\, r^2 \hat{\boldsymbol{m}} \hat{z}_{n-1} + (r^2+q^2)(m^2r^2+q^2)\, \hat{z}_n \\ + r^2(m^2r^2+q^2)\, \hat{\boldsymbol{m}}^{\mathrm{T}} \hat{z}_{n+1} \end{pmatrix}}{\begin{pmatrix} r^2(r^2+q^2) + (r^2+q^2)(m^2r^2+q^2) \\ + m^2r^2(m^2r^2+q^2) \end{pmatrix}}. \quad (3.6)$$

Eq. (3.6) describes how the 3 observation vectors $\hat{z}_{n-1}$, $\hat{z}_n$ and $\hat{z}_{n+1}$ are to be weighted in the analysis of $\hat{x}_n^s$. We consider three special cases.

$$\hat{x}_n^s = \frac{[\hat{\boldsymbol{m}} \hat{z}_{n-1} + m^2 \hat{z}_n + m^2 \hat{\boldsymbol{m}}^{\mathrm{T}} \hat{z}_{n+1}]}{[1 + m^2 + m^4]},$$

$$(q^2 = 0)$$

$$\hat{x}_n^s = \frac{[r^2 \hat{\boldsymbol{m}} \hat{z}_{n-1} + (r^2 + q^2)\, \hat{z}_n + r^2 \hat{\boldsymbol{m}}_{n+1}^{\mathrm{T}} \hat{z}_{n+1}]}{[3r^2 + q^2]},$$

$$(m^2 = 1)$$

$$\hat{x}_n^s = \frac{[\hat{\boldsymbol{m}} \hat{z}_{n-1} + \hat{z}_n + \hat{\boldsymbol{m}}^{\mathrm{T}} \hat{z}_{n+1}]}{3},$$

$$(m^2 = 1,\ q^2 = 0) \quad (3.7)$$

$\hat{\boldsymbol{m}} \hat{z}_{n-1}$, $\hat{z}_n$ and $\hat{\boldsymbol{m}}^{\mathrm{T}} \hat{z}_{n+1}$ all lie along the same model characteristic. For the perfect, inviscid case ($q^2 = 0$, $m^2 = 1$) the analyzed values are simply obtained by averaging along the characteristics (see Daley, 1992, chapter 13, problem 13.3). This type of averaging effectively filters the (random) observation error. For perfect, damped models, more weight is given to earlier observations than later ones, consistent with (3.5). When the model is imperfect, most weight is given to observations which occur at the same time as the analysis, earlier or later observations are given less weight.

These ideas are illustrated more graphically in Figs. 2 and 3. We assume that there are observations $t_0 \leqslant t_n \leqslant t_N$, where $N = 8$. The experimental set up is as described earlier in this section. Panels of Figs. 2 and 3 correspond to each other, except that $\boldsymbol{Q} = 0$ in Fig. 2. In each panel of Figs. 2 and 3 we plot the weights for the (smoother) analysis, as a function of time (abscissa) and space (ordinate), given to an observation at the point marked with an "X". This point is an arbitrary analysis grid-point at $t = 4$. There is no information prior to $t = 0$, except in Figs. 2d and 3d. Courant number $U\,\Delta t/\Delta x = 1$ for (a, c, d) and $U\,\Delta t/\Delta x = 0.5$ for (b). All cases are inviscid except for (c), where $\nu\,\Delta t/a^2 = 0.1$.

Consider first, the perfect ($\boldsymbol{Q} = 0$) model case of Fig. 2, corresponding to the 4DVAR algorithm. In Fig. 2a, the weights given each of the observations along the characteristic which passes through the point "X" are all equal to $1/(N+1) = 0.111$. The weights given to all observations which do not lie

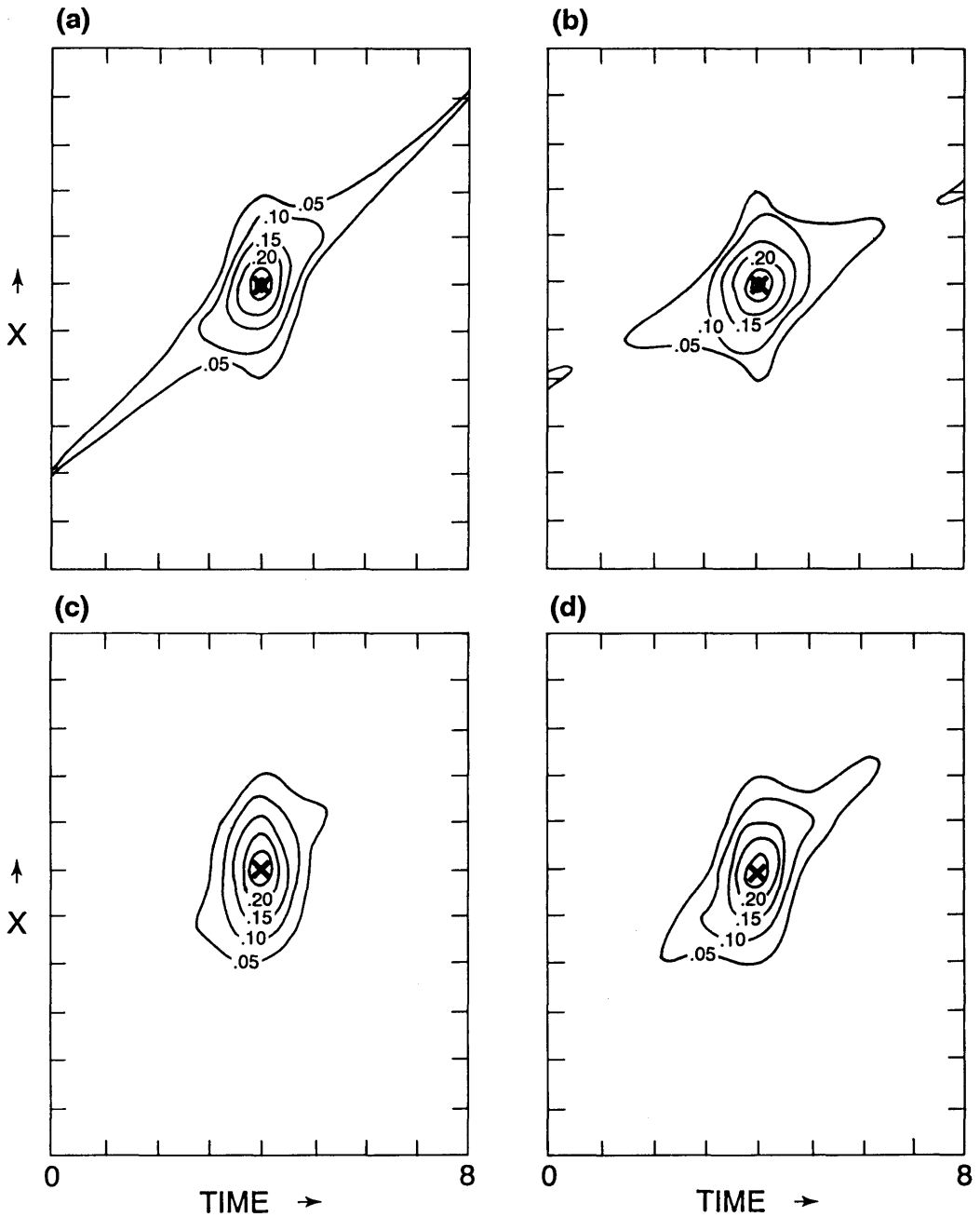

Fig. 3. Same as Fig. 2, except for an imperfect model.

along this characteristic are equal to zero in this case. The weights in Fig. 2b, are more complex, because the characteristic which passes through the point "X" does not necessarily pass through an observation point at other times. For the inviscid, perfect cases (panels a, b), the observational noise is filtered by appropriate smoothing along characteristics and there is no spatial filtering. For the perfect, damped case (panel c), the observations at t_0 are given the most weight as expected. The large scales are obtained by averaging the observations along the characteristics, but the smaller scales are obtained from observations at the beginning of the period. The primary effect of prior information (panel d) is to introduce spatial averaging. ($\boldsymbol{P}_0^f = \boldsymbol{Q}$ has a red spectrum.)

The introduction of model error into the algorithm (Fig. 3) changes the weights substantially. In particular, much more weight is given to observations which are concurrent with the analysis, and much less weight is given to observations which are more distant in time. Prior information (Fig. 3d) has relatively little effect on the weights when the model is imperfect. The dissipation causes some asymmetry in Fig. 3c, but it is not detectable in the panel.

Figs. 2 and 3 dramatically illustrate that the weights given to observations are very different for perfect and imperfect models. In particular, a perfect model may give much more weight to observations which are distant in time and/or space. If the model truly is perfect, then these weights are appropriate, but if the model is not perfect then the resulting analyses will be substantially in error because distant observations have been given too much weight.

3.4. *The fixed-lag smoother*

The Kalman smoother equations (2.1–5, 2.11–12 and 2.13), together with the simple assimilating model (3.1), were also used to investigate another potentially valuable assimilating algorithm, the so-called *fixed-lag* smoother. This smoother has potential value for analyzing climatological information which does not have to be processed in strict real time. The fixed-lag smoother is a Kalman filter with a look-ahead, which for the analysis at t_n, processes observations from $t_n < t \leqslant t_n + k\,\Delta t$ (k is a lag) as well as observations prior to t_n. ($k = 0$ is the Kalman filter). Cohn et al. (1994) have designed a form of fixed-lag smoother referred to as the *retrospective algorithm* for this purpose.

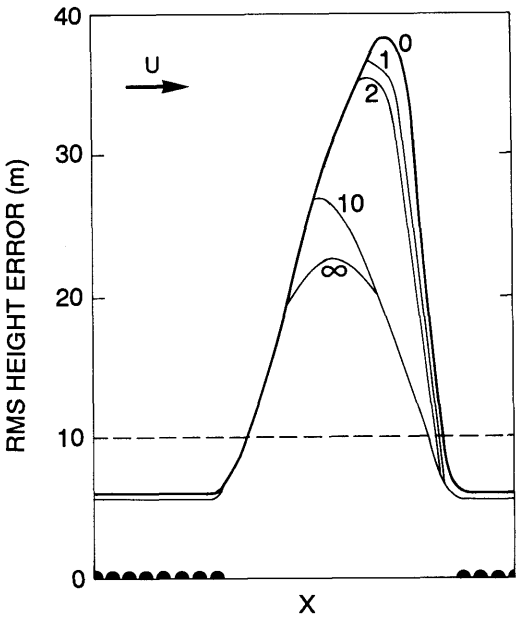

Fig. 4. Stationary rms smoother error variance (in the inviscid case) in the vicinity of the data void for lags of $k = 0, 1, 2, 10, \infty$. Flow is from the left. Observations positions are indicated by solid half-circles on lower margin. Dashed line is rms observation error.

The fixed-lag smoother experiments used an observation network which coincided with the analysis grid except for a *single large data void* (13 observations missing). Covariances propagate in these experiments and the fixed-lag smoother was integrated to stationarity for lags of $k = 0$, 1, 10 and ∞. The resulting asymptotic analysis error variance is shown for different lags as a function of x in the vicinity of the data void in Fig. 4. This figure shows that while the Kalman filter ($k = 0$) propagates information downstream, the Kalman smoother can also propagate future information upstream, thus substantially improving analyses near the downstream edge of data voids. Other experiments, not shown, suggest that this effect is maximized for the inviscid case, consistent with (3.5).

4. Pontryagin optimization

In Section 2 we derived the Kalman smoother from estimation theory by deriving the minimum

variance estimate given past, present and future observations. For the fixed-interval problem, the smoothing equations can also be obtained from optimal control principles using the initial state $\boldsymbol{x}_0$ and the model error trajectory $\{\boldsymbol{e}_n^{\mathrm{q}}, 0 \leqslant n \leqslant N\}$ as control variables. The purpose of this section is to demonstrate the equivalence between the estimation theory and optimal control formulations (under certain linearity conditions) and introduce a modified adjoint equation that accounts for model error. The demonstration of this equivalence is actually classical and is due to Rauch, Tung and Striebel (1965). It has been presented in the context of data assimilation for the scalar case by Bennett and Budgell (1989) and less powerful forms of the equivalence can be seen in Thacker (1986) by neglecting future observations or Rabier et al. (1993) by invoking the perfect model assumption.

The equivalence stated above has led to efficient methods of finding solutions to the weak constrained control problem, such as the representer method or simulated annealing (Bennett and Thorburn, 1992 and Evensen, 1994). However, this section has a different focus, using the sweep method to decouple the Euler-Lagrange equations and obtain a modified adjoint equation. This equation is a generalized form of the adjoint equation that occurs in the 4DVAR algorithm, which accounts for the model error covariance. The 4DVAR algorithm, where the model is introduced as a strong constraint, can then be derived as a special case. In Section 5, this equivalence will be used to assess the estimation error in 4DVAR analyses.

We define a non-linear model $\mathcal{M}$, which operates on the true state vector $\boldsymbol{x}_n$. Then, the equivalent of eq. (2.6) at time t_n is,

$$\boldsymbol{x}_{n+1} = \mathcal{M}(\boldsymbol{x}_n) + \boldsymbol{e}_n^{\mathrm{q}}, \tag{4.1}$$

where $\boldsymbol{e}_n^{\mathrm{q}}$ is the model error at time t_n. Now define a cost function J, over the interval $0 \leqslant n \leqslant N$, as,

$$J = J_0 + \sum_{n=0}^{N} J_{\mathrm{r}}(n) + \sum_{n=0}^{N-1} J_{\mathrm{q}}(n), \tag{4.2}$$

where

$$J_0 = \tfrac{1}{2}[\boldsymbol{x}_0^{\mathrm{f}} - \boldsymbol{x}_0]^{\mathrm{T}} [\boldsymbol{P}_0^{\mathrm{f}}]^{-1} [\boldsymbol{x}_0^{\mathrm{f}} - \boldsymbol{x}_0],$$

$$J_{\mathrm{r}}(n) = \tfrac{1}{2}[\boldsymbol{z}_n - \boldsymbol{H}_n \boldsymbol{x}_n]^{\mathrm{T}} \boldsymbol{R}_n^{-1} [\boldsymbol{z}_n - \boldsymbol{H}_n \boldsymbol{x}_n],$$

$$J_{\mathrm{q}}(n) = \tfrac{1}{2}[\boldsymbol{x}_{n+1} - \mathcal{M}(\boldsymbol{x}_n)]^{\mathrm{T}} \boldsymbol{Q}_n^{-1} [\boldsymbol{x}_{n+1} - \mathcal{M}(\boldsymbol{x}_n)],$$

and $\boldsymbol{x}_0^{\mathrm{f}}$ is the forecast vector of the state variable(s) obtained from observations prior to t_0 and the remaining variables are defined in Section 2. The three cost functions J_0, J_{r} and J_{q} will be referred to as the initial, observational and model cost functions respectively. (The symbol "J" for cost function should not be confused with the symbol for number of gridpoints defined in Section 3). If there is no information prior to t_0, then $(\boldsymbol{P}_0^{\mathrm{f}})^{-1} = 0$, and the initial cost function J_0 is ignored. We have assumed a linear forward interpolation operator $\boldsymbol{H}_n$ as in Section 2, although the derivation can be generalized to include a nonlinear operator.

An optimal control of the initial state $\boldsymbol{x}_0$ and the model error trajectory $\{\boldsymbol{e}_n^{\mathrm{q}}, 0 \leqslant n \leqslant N\}$ can be obtained by minimizing the cost function (4.2) under the (weak) constraint (4.1) yielding a least square estimate. An equivalent control method is derived from Pontryagin's maximum principle (Pontryagin et al., 1962). The optimal control is obtained by maximizing the Hamiltonian of the optimization problem $\mathcal{H}(\boldsymbol{x}_n, \lambda_{n+1}, \boldsymbol{e}_n^{\mathrm{q}}) = \lambda_{n+1}^{\mathrm{T}}[\mathcal{M}(\boldsymbol{x}_n) + \boldsymbol{e}_n^{\mathrm{q}}] - J$, where J is the cost function (4.2). The control method developed by Lev Semenovich Pontryagin and co-workers is actually more general than least-square estimation or variational methods, but for quadratic cost functions (i.e., L_2 norms) and unrestricted control, it makes no difference. In order to distinguish this procedure from the 4DVAR procedure (the model used as strong constraint), we will refer to the above control method as *Pontryagin optimization.* Other authors has used different terminology, such as *generalized inverse* (Bennett, 1992).

It should be noted that if the initial error $\boldsymbol{e}_0^{\mathrm{f}} = \boldsymbol{x}_0^{\mathrm{f}} - \boldsymbol{x}_0$, model error $\boldsymbol{e}_n^{\mathrm{q}}$ and observation error $\boldsymbol{e}_n^{\mathrm{r}} = \boldsymbol{z}_n - \boldsymbol{H}_n \boldsymbol{x}_n$ are independent random Gaussian vectors (with covariances $\boldsymbol{P}_0^{\mathrm{f}}$, $\boldsymbol{Q}_n$ and $\boldsymbol{R}_n$ respectively) and $\boldsymbol{e}_n^{\mathrm{r}}$ and $\boldsymbol{e}_n^{\mathrm{q}}$ are white-noise processes; then the Pontryagin estimate $\boldsymbol{x}_n^*$, which minimizes (4.2) is also a maximum likelihood estimate, since the joint probability distribution of $\boldsymbol{e}_0^{\mathrm{f}}$, $\boldsymbol{e}_n^{\mathrm{q}}$, $\boldsymbol{e}_n^{\mathrm{r}}$, $0 \leqslant n \leqslant N$ is proportional to $\exp(-J)$.

To derive the Pontryagin optimal control, we can minimize J subject to the constraint (4.1). Thus, we introduce Lagrange multipliers and define a modified functional as,

$$\bar{J} = J - \sum_{n=0}^{N-1} \lambda_{n+1}^{\mathrm{T}} [\boldsymbol{x}_{n+1} - \mathcal{M}(\boldsymbol{x}_n) - \boldsymbol{e}_n^{\mathrm{q}}], \tag{4.3}$$

where $\{\lambda_n\}$ is a set of N undetermined Lagrange multiplier vectors. We introduce the tangent linear approximation,

$$\mathcal{M}(\boldsymbol{x}_n+\delta\boldsymbol{x}_n)\approx\mathcal{M}(\boldsymbol{x}_n)+\boldsymbol{M}(\boldsymbol{x}_n)\,\delta\boldsymbol{x}_n, \tag{4.4}$$

where $\boldsymbol{M}(\boldsymbol{x}_n)=\partial\mathcal{M}(\boldsymbol{x})/\partial\boldsymbol{x}|_{\boldsymbol{x}=\boldsymbol{x}_n}$ is the tangent linear model of $\mathcal{M}$ evaluated at $\boldsymbol{x}_n$ and $\delta\boldsymbol{x}_n$ is the variation of $\boldsymbol{x}_n$. If the model $\mathcal{M}$ is linear then the tangent linear model $\boldsymbol{M}(\boldsymbol{x}_n)$ is independent of the state and becomes equal to the model $\boldsymbol{M}_n$ of Section 2.

We denote the optimum state and model error (which minimize 4.2) as $\boldsymbol{x}_n=\boldsymbol{x}_n^*$ and $\boldsymbol{e}_n^{\mathrm{q}}=\boldsymbol{e}_n^{\mathrm{q}*}$, respectively. For eq. (4.3), variation on λ_n yields the *adjoint equation*,

$$\lambda_n=\boldsymbol{M}^{\mathrm{T}}(\boldsymbol{x}_n^*)\,\lambda_{n+1}-\boldsymbol{H}_n^{\mathrm{T}}\boldsymbol{R}_n^{-1}[\boldsymbol{z}_n-\boldsymbol{H}_n\boldsymbol{x}_n^*],\quad 1\leqslant n\leqslant N-1, \tag{4.5}$$

subject to the adjoint final condition,

$$\lambda_N=-\boldsymbol{H}_N^{\mathrm{T}}\boldsymbol{R}_N^{-1}[\boldsymbol{z}_N-\boldsymbol{H}_N\boldsymbol{x}_N^*]. \tag{4.6}$$

While (4.5–6) have the same mathematical form as the 4DVAR adjoint equations, they involve $\boldsymbol{x}_n^*$ which has not been determined. Variation on $\boldsymbol{e}_n^{\mathrm{q}}$, assuming $\boldsymbol{Q}_n^{-1}$ is non-singular and applying (4.1) yields,

$$\boldsymbol{x}_{n+1}^*=\mathcal{M}(\boldsymbol{x}_n^*)-\boldsymbol{Q}_n\lambda_{n+1},\quad 0\leqslant n\leqslant N-1, \tag{4.7}$$

subject to the *initial condition*,

$$\boldsymbol{x}_0^*=\boldsymbol{x}_0^{\mathrm{a}}-\boldsymbol{P}_0^{\mathrm{a}}\boldsymbol{M}^{\mathrm{T}}(\boldsymbol{x}_0^*)\,\lambda_1, \tag{4.8}$$

where $\boldsymbol{x}_0^{\mathrm{a}}$ is the analysis produced by observations for $t\leqslant t_0$. The system (4.5–8) forms the *discrete-time Euler-Lagrange* equations of the variational problem (4.2). This system of equations is coupled and *cannot* be solved with a simple direct integration, as it can for the 4DVAR algorithm.

We note that in the absence of statistical information about the model error ($\boldsymbol{Q}_n^{-1}=0$), the cost function J_q in (4.2) vanishes and J becomes identical (or similar) to the cost function used in 4DVAR. Pontryagin optimization is ill-defined in this case, because the variation on $\boldsymbol{e}_n^{\mathrm{q}}$ leads to the condition $\lambda_n^{\mathrm{T}}=0$, $0\leqslant n\leqslant N$, and there is no control on the model error. Controllability can be regained if the control over model error is removed, that is minimize J with respect to initial conditions only, or equivalently assume a perfect model as in the 4DVAR algorithm.

Several methods have been proposed to decouple the Euler-Lagrange equations (see Bennett, 1992), but to show the relation between minimization of (4.2) and the Kalman smoother, we shall use the sweep method (Gel'fand and Fomin, 1963). The application of the sweep method to this problem can be described briefly as follows (see also Bennett (1992)). The initial condition (4.8) suggests a solution of the form,

$$\boldsymbol{x}_n^*=\boldsymbol{x}_n^{\mathrm{a}}-\boldsymbol{P}_n^{\mathrm{a}}\boldsymbol{M}^{\mathrm{T}}(\boldsymbol{x}_n^{\mathrm{a}})\,\lambda_{n+1}, \tag{4.9}$$

where $\boldsymbol{x}_n^{\mathrm{a}}$ and $\boldsymbol{P}_n^{\mathrm{a}}$ are obtained from past and present information only. Application of the non-linear model $\mathcal{M}$ to (4.9), using (4.5–8), dropping higher order terms in λ (in the non-linear case) and some manipulation, yields,

$$\begin{aligned}&[\boldsymbol{P}_{n+1}^{\mathrm{a}}]^{-1}\,[\boldsymbol{x}_{n+1}^{\mathrm{a}}-\mathcal{M}(\boldsymbol{x}_n^{\mathrm{a}})]\\&\quad-\boldsymbol{H}_{n+1}^{\mathrm{T}}\boldsymbol{R}_{n+1}^{-1}[\boldsymbol{z}_{n+1}-\boldsymbol{H}_{n+1}\mathcal{M}(\boldsymbol{x}_n^{\mathrm{a}})]\\&\quad=\{I+[\boldsymbol{H}_{n+1}^{\mathrm{T}}\boldsymbol{R}_{n+1}^{-1}\boldsymbol{H}_{n+1}-[\boldsymbol{P}_{n+1}^{\mathrm{a}}]^{-1}]\\&\quad\times[\boldsymbol{M}(\boldsymbol{x}_n^{\mathrm{a}})\,\boldsymbol{P}_n^{\mathrm{a}}\boldsymbol{M}^{\mathrm{T}}(\boldsymbol{x}_n^{\mathrm{a}})+\boldsymbol{Q}_n]\}\,\lambda_{n+1}.\end{aligned} \tag{4.10}$$

This equation is exact for the linear case and shows that if the forward Kalman filter equations (2.1–5) are satisfied, then equation (4.10) will be true for any values of λ_n. Thus, for the linear case, the forward Kalman filter equations can be derived by minimizing (4.2). For the non-linear case (replacing 2.4 by $\boldsymbol{x}_{n+1}^{\mathrm{f}}=\mathcal{M}(\boldsymbol{x}_n^{\mathrm{a}})$ for example), eq. (4.10) is still valid to first order in λ, and thus, the equivalence is also true for the *extended* Kalman filter equations. It remains to be shown that the reverse-time information filter (2.11–12) and the smoothing step (2.13) are consistent with the minimization of (4.2). This is basically done by developing evolution equations for the covariance of λ_n and error covariances of the sweep estimate $\boldsymbol{x}_n^*$ and then showing that the resulting algorithm is equivalent to (2.11–13), see Bennett and Budgell (1989).

From (4.10), the sweep method has effectively decoupled the Euler-Lagrange equations (4.5–8). Thus, insertion of the sweep solution (4.9) into the adjoint equations (4.5–6) results in a *modified adjoint equation*,

$$\lambda_n = [\boldsymbol{I} - \boldsymbol{H}_n^{\mathrm{T}} \boldsymbol{R}_n^{-1} \boldsymbol{H}_n \boldsymbol{P}_n^{\mathrm{a}}]\, \boldsymbol{M}^{\mathrm{T}}(\boldsymbol{x}_n^{\mathrm{a}})\, \lambda_{n+1} - \boldsymbol{H}_n^{\mathrm{T}} \boldsymbol{R}_n^{-1} [z_n - \boldsymbol{H}_n \boldsymbol{x}_n^{\mathrm{a}}], \tag{4.11}$$

subject to the final condition,

$$\lambda_N = -\boldsymbol{H}_N^{\mathrm{T}} \boldsymbol{R}_N^{-1} [z_N - \boldsymbol{H}_N \boldsymbol{x}_N^{\mathrm{a}}]. \tag{4.12}$$

Unlike equations (4.5–6), the modified adjoint equations depend on $\boldsymbol{x}_n^{\mathrm{a}}$ and $\boldsymbol{P}_n^{\mathrm{a}}$ which can be obtained from a forward integration (the forward sweep) of the Kalman filter eqs. (2.1–5 or their non-linear extensions). The modified adjoint equations (4.11–12) are then integrated backwards in time (the backward sweep) using the stored values $\boldsymbol{x}_n^{\mathrm{a}}$ and $\boldsymbol{P}_n^{\mathrm{a}}$. From the values of $\boldsymbol{x}_n^{\mathrm{a}}$, $\boldsymbol{P}_n^{\mathrm{a}}$ and λ_n equation (4.9) is applied to obtain $\boldsymbol{x}_n^*$, which minimizes (4.2). Note that the observations are processed twice, once in the forward sweep and once in the backward sweep.

As noted earlier, the standard adjoint equation (similar to 4.5–6) used in the 4DVAR algorithm, cannot be used when the model is imperfect. While the sweep algorithm itself is expensive, a practical weak constraint variational scheme can perhaps be formulated, by inserting sub-optimal filtering estimates for $\boldsymbol{x}_n^{\mathrm{a}}$ and $\boldsymbol{K}_n = \boldsymbol{P}_n^{\mathrm{a}} \boldsymbol{H}^{\mathrm{T}} \boldsymbol{R}_n^{-1}$ (e.g., Todling and Cohn, 1994) into equations (4.11–12). We note that the value of the Lagrange multiplier at initial time is related to the gradient of the cost function, which could then be used in a descent algorithm. Thus, the addition of a sub-optimal Kalman filter to a 4DVAR algorithm, and the use of the modified adjoint equation might partially account for an imperfect model or model error covariance effects, although the algorithm would still be expensive compared to 4DVAR.

In Section 5, we will assume the equivalence between the minimization of (4.2) and the Kalman smoother equations of Section 2, and use this equivalence to examine the effect of making an inappropriate perfect model error assumption in the 4DVAR algorithm.

5. Mis-specification and the observational cost function

We can now re-interpret the results of Section 3, using the equivalence between the Kalman smoother and the minimization of the cost function shown in Section 4. In particular, it can be seen that Figs. 1a, 1b and 2 actually refer to the 4DVAR algorithm. In Figs. 1d, 2d, 3d, the J_0 term in the cost function (4.2) is non-zero, but in all other panels of Figs. 1–3, J_0 is zero.

5.1. *The mis-specification equations*

In the 4DVAR algorithm, it is assumed that the model is perfect. In reality, of course, no model is perfect. The effect of the perfect model assumption in the 4DVAR algorithm was examined recently by Wergen (1993). Another completely different way of examining this assumption, is to use the Kalman smoother equations developed in Sections 2 and 3. This is done by deriving mis-specification equations as in Daley (1991, Subsection 4.9) and Daley (1992a). These equations can be used to determine the expected smoother error statistics if the model is specified to be perfect when it is actually imperfect. In this treatment, we will derive the spectral form of the mis-specification equations assuming homogeneity and a collocated observation network and analysis grid as in Section 3.

The correct model and observation error variances for wavenumber p are denoted $q^2(p)$, $r^2(p)$. Suppose the model and observation error variances are *incorrectly specified* as $\tilde{q}^2(p)$ and $\tilde{r}^2(p)$. If we assume the model is perfect (i.e., ignore the J_q term in 4.2), then $\tilde{q}^2(p) = 0$. From (3.2–3), the incorrect analysis, forecast and smoother error variances $\tilde{a}_n^2(p), \tilde{f}_n^2(p), \tilde{\alpha}_n^2(p), \tilde{\beta}_n^2(p), \tilde{s}_n^2(p)$ for each wavenumber p are given by,

$$\begin{aligned}
\tilde{f}_{n+1}^2 &= m^2 \tilde{a}_n^2, \\
\tilde{a}_n^2 &= \tilde{r}^2 \tilde{f}_n^2 / [\tilde{r}^2 + \tilde{f}_n^2], \\
\tilde{\beta}_{n-1}^{-2} &= m^2 \tilde{\alpha}_n^{-2}, \\
\tilde{\alpha}_n^{-2} &= \tilde{\beta}_n^{-2} + \tilde{r}^{-2}, \\
\tilde{s}_n^2 &= \tilde{a}_n^2 [1 + \tilde{a}_n^2 \tilde{\beta}_n^{-2}],
\end{aligned} \tag{5.1}$$

while the corresponding correct error variances satisfy,

$$\begin{aligned}
f_{n+1}^2 &= m^2 a_n^2 + q^2, \\
a_n^2 &= [\tilde{r}^4 f_n^2 + \tilde{f}_n^4 r^2] / [\tilde{r}^2 + \tilde{f}_n^2]^2, \\
\beta_{n-1}^{-2} &= m^2 \alpha_n^{-2} / [1 + q^2 \alpha_n^{-2}], \\
\alpha_n^{-2} &= [\tilde{r}^2 + \tilde{\beta}_n^2]^2 / [\tilde{r}^4 \beta_n^2 + r^2 \tilde{\beta}_n^4], \\
s_n^2 &= [\tilde{\beta}_n^4 a_n^2 + \tilde{a}_n^4 \beta_n^2] / [\tilde{\beta}_n^2 + \tilde{a}_n^2]^2.
\end{aligned} \tag{5.2}$$

If $\tilde{r}^2(p) = r^2(p)$, $\tilde{q}^2(p) = q^2(p)$, then (5.1) and (5.2) are the same. Mis-specification of $\boldsymbol{Q}$ and $\boldsymbol{R}$ introduces errors in the forward integration, the reverse integration and the smoother step. In these experiments we were concerned with the mis-specification of the model error $\boldsymbol{Q}$, so the observation error covariance $\boldsymbol{R}$ was assumed to be correctly specified ($\tilde{r}^2(p) = r^2(p)$).

Experiments were conducted with exactly the same experimental set-up as in Section 3. In the solid curves of Fig. 5, we show the error variances corresponding to the forward analysis (a), the hindcast (b) and smoother error (s) over time-interval $t_0 \leqslant t_0 + n\,\Delta t \leqslant t_N$, with $N = 7$. In this case, the model is inviscid and imperfect, with the spectrum of $\boldsymbol{Q}$ given in Section 3. There is no information prior to t_0 ($\boldsymbol{P}_0^{\mathrm{f}} \to \infty$). However, we apply the 4DVAR algorithm, which assumes that the model is perfect. Fig. 5 can be compared with the optimal 4DVAR results in Fig. 1a (perfect, inviscid) and 1b (imperfect, inviscid). It is obvious that the error of 4DVAR is much larger here than in either of the optimal cases, particularly at t_0 and t_N. Only at $t_{N/2}$ is the 4DVAR error variance smaller than the observation error variance. This is the consequence of imposing a strong, but erroneous constraint.

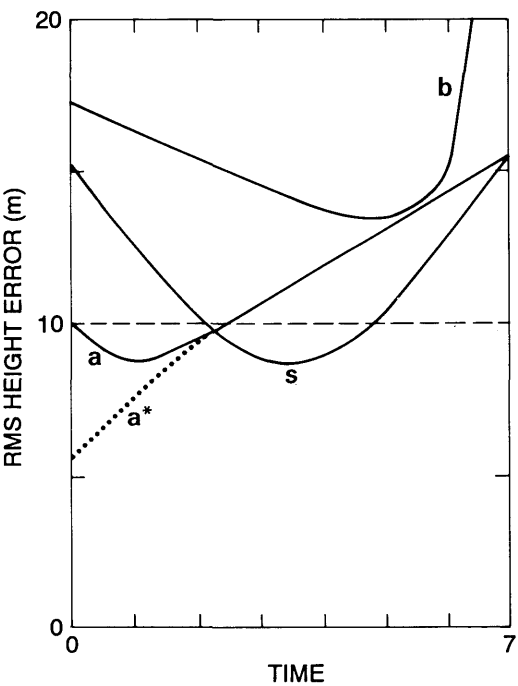

Fig. 5. Error variances for the mis-specification experiment in the same format as Fig. 1. The solid curves indicate no information prior to t_0, while the dotted curve (a*) indicates the effect of prior information on the forward analysis curve.

The effect of information prior to t_0 ($\boldsymbol{P}_0^{\mathrm{f}}$ finite) was also tested and is illustrated by dotted curve "a*" of Fig. 5. Information prior to $t = 0$ reduced the error below the "a" curve of Fig. 5, but had no effect on the hindcast ("b" curve). The perfect model assumption gives too much weight to the incorrect hindcast at $t = 0$; and the resulting smoother error variance is not reduced.

5.2. *The observational cost function as a performance diagnostic*

The mis-specification equations can only be used to determine the true analysis or smoother error statistics in idealized experiments because the "true" observation and model error statistics are required. Under more realistic (operational) conditions, the "true" model and observation error statistics are not available, and can only be estimated. However, performance diagnostics (Daley, 1992a) can be used to assess the optimality of an assimilation system under operational conditions. We will now discuss such a performance diagnostic, based on the temporal evolution of $J_r(n)$, the observational cost function in eq. (4.2).

This is demonstrated by using the system discussed in Section 3. Then, for this system, the observational cost function (4.2) is,

$$J_r(n) = [\boldsymbol{z}_n - \boldsymbol{H}_n \boldsymbol{x}_n^{\mathrm{s}}]^{\mathrm{T}}\, \boldsymbol{R}_n^{-1} [\boldsymbol{z}_n - \boldsymbol{H}_n \boldsymbol{x}_n^{\mathrm{s}}]. \tag{5.3}$$

Taking the expected value of (5.3) and using (3.2–4, 5.1–2) gives,

$$\langle J_r(n) \rangle = \sum_{p=0}^{(J-1)/2} \delta_p [1 + s_n^2(p)\, r^{-2}(p) - 2\tilde{s}_n^2(p)\, r^{-2}(p)], \tag{5.4}$$

where $s_n^2(p)$ and $\tilde{s}_n^2(p)$ are the correct and incorrect smoother error variances defined in (5.1–2) and $\delta_p = 1$ for $p = 0$ and $\delta_p = 2$ otherwise. The optimal case is obtained from (5.4) by setting $\tilde{s}_n^2 = s_n^2$. Finally, defining $I(n)$ as the number of observations at t_n, we calculate $\langle J_r(n) \rangle / I(n)$ and plot it as a function of n. This quantity, which measures the

fit to the observations, is dimensionless and O(1), and should be relatively insensitive both to the type and number of observations. This quantity will be referred to as *the normalized fit*, and equals 1 when the smoothed estimate is the truth.

We plot in Fig. 6 $\langle J_r(n)\rangle / I(n)$ as a function of time, $0 \leqslant n \leqslant 7$, for the inviscid case (with no prior information). For this simple spectral model, $I(n) = J = 49$ as before. The model is inviscid. The case marked "PONTRYAGIN" refers to the imperfect model with model error spectrum of Section 3, and is the same case shown in Fig. 1b. The case marked "4DVAR $Q = 0$" corresponds to Fig. 1a. The curve marked "4DVAR $Q \neq 0$" corresponds to Fig. 5 (solid curves). As expected, when the model is perfect and the 4DVAR algorithm is used, the normalized fit is independent of time. Pontryagin optimization draws more closely to the observations at the beginning and end of the interval because there is less information available from the model.

However, when the 4DVAR algorithm is applied with an imperfect model, the algorithm is not able to fit observations at the beginning and end of the interval. The 4DVAR algorithm (see Fig. 2a) gives considerable weight to temporally distant observations. Thus, when the model is imperfect, the 4DVAR algorithm gives insufficient weight to temporally local observations. This problem will obviously be minimized in the center of the assimilating interval, where no observation is more temporally distant than half the interval. What is particularly interesting is that there is a change of curvature for the normalized fit depending upon whether the system is optimal or not. In particular, for the 4DVAR algorithm, the more positive the curvature, the less justifiable is the perfect model assumption.

For Pontryagin optimization or the Kalman smoother algorithm, it might be possible to operationally tune a formulation for the model error Q, based on examination of the normalized fit. For example, if it were supposed that the model error covariance was homogeneous and stationary with a spectrum given in Section 3, then examination of the normalized fit could be used to choose the optimal values of l_q and E_q. Thus, the optimal values of these parameters would be those that occurred when the curvature of the normalized fit changed sign.

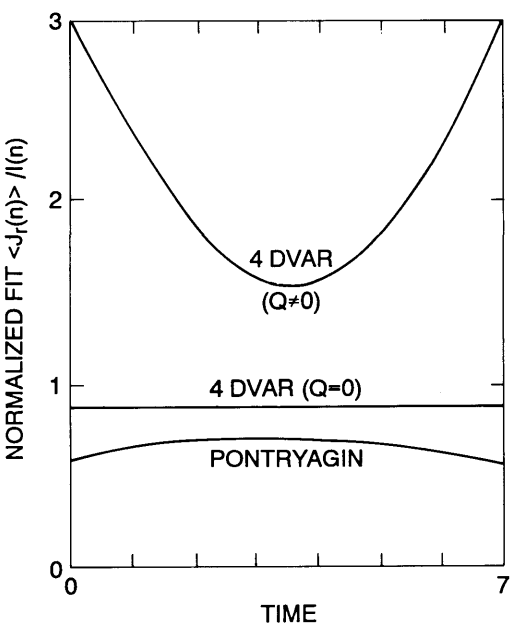

Fig. 6. Normalized fit for Pontryagin optimization and the 4DVAR algorithm for perfect and imperfect models.

6. Summary and conclusions

The Kalman smoother is a generalization of the Kalman filter algorithm, which permits the processing of future observations. The fixed-interval Kalman smoother is equivalent to Pontryagin optimization, a generalization of the four dimensional variational (4DVAR) algorithm which does not require the perfect model assumption.

The Kalman smoother can be derived as the optimal combination of a forward Kalman filter and a reverse-time information filter which processes future observations. Future observations improve the analysis, particularly when they are accurate, the flow is inviscid or unstable and the model is perfect. Future information can propagate upstream and improves the analysis near the downstream edge of data voids. For dissipative systems, future observations provide little information about the smaller spatial scales.

The fixed-interval Kalman smoother was shown to minimize a four-dimensional cost function involving observation errors, initial errors and model errors (Pontryagin optimization). The

4DVAR algorithm assumes a perfect model and thus minimizes a simpler cost function which contains no model error term. In this way, the 4DVAR algorithm can be thought of as a special case of Pontryagin optimization, or alternatively as a sub-optimal Kalman smoother. The 4DVAR algorithm does not produce error statistics; but this connection between the 4DVAR algorithm and the Kalman smoother made it possible to examine the error characteristics of the 4DVAR algorithm by examining the second moment error statistics produced by the Kalman smoother.

It was shown that the perfect model assumption (made in the 4DVAR algorithm) may have a significant effect on the forecast/analysis statistics and the weights given to observations. In particular, the perfect model assumption required by the 4DVAR algorithm gives substantial weight to observations which are distant in time and/or space. It was found that applying the 4DVAR algorithm, when the model is actually incorrect, may cause large errors, particularly at the beginning and end of the assimilation period. A procedure, based on the observational cost function (fit to observations) was developed to diagnose the effect of model error on the 4DVAR algorithm.

Modification of the 4DVAR algorithm to account for the model error covariance is a desirable goal. The use of the modified adjoint equation (4.11–12) in a 4DVAR algorithm, combined with a sub-optimal Kalman filter to derive the analysis and analysis error statistics from past information, might prove to be a feasible way to account for model error.

In this paper, we have demonstrated the consequences of an inappropriate perfect model assumption when using the 4DVAR algorithm. However, it is intrinsically difficult to separate model error from forecast error, and consequently, it is not clear how "perfect" todays' assimilating models really are. For example, recent 4DVAR experiments by Thépaut et al. (1993) suggest that the assimilating models might be sufficiently accurate to make the perfect model assumption a less serious drawback than other limitations of the 4DVAR algorithm. On the other hand, short-range weather forecasts by these models frequently "bust" and it is known from climate modelling studies that there are serious model biases and inadequacies in the physical parameterizations. Consequently, at this time, it would still seem wise to regard the perfect model assumption with suspicion.

7. Acknowledgements

The authors would like to acknowledge S. E. Cohn, N. S. Suvakumaran and R. Todling of NASA Goddard Space Flight Center for their careful review of the manuscript and for many helpful discussions. The graphics were done by Tom Chivers of AES. Support for this work was provided by AES both in the form of a scholarship and a contract.

REFERENCES

Bennett, A. 1992. *Inverse methods in physical oceanography.* Cambridge University Press, New York, 346 pp.

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

Bennett, A. and Thorburn, M. 1992. The generalized inverse of a nonlinear quasigeostrophic ocean circulation model. *J. Phys. Oceanogr.* **22**, 213–230.

Bouttier, F. 1994. *Sur la prévision de la qualité des prévision météorologiques.* PhD Thesis, Université Paul Sabatier à Toulouse.

Cohn, S. E. 1993. Dynamics of short-range univariate forecast error covariances. *Mon. Wea. Rev.* **121**, 3123–3149.

Cohn, S. E., Sivakumaran, N. S. and Todling, R. 1994. A fixed-lag Kalman smoother for retrospective atmospheric data assimilation. *Mon. Wea. Rev.* **122**, 2838–2867.

Courtier, P. and Talagrand, O. 1990. Variational assimilation of meteorological observations with the direct and adjoint shallow water equations. *Tellus* **42A**, 531–549.

Daley, R. 1991. *Atmospheric data analysis.* Cambridge University Press. 457 pp.

Daley, R. 1992a. The lagged innovation covariance: a performance diagnostic for atmospheric data assimilation. *Mon. Wea. Rev.* **120**, 178–196.

Daley, R. and Ménard, R. 1993. Spectral characteristics of Kalman filter systems for atmospheric data assimilation. *Mon. Wea. Rev.* **121**, 1554–1565.

Dee, D. 1995. On-line estimation of error covariance parameters for atmospheric data assimilation. *Mon. Wea. Rev.* **123**, 1182–1196.

Evensen, G. 1994. Inverse methods and data assimilation in non-linear ocean models. *Physica D* **77**, 108–129.

Fraser, D. and Potter, J. 1969. The optimum linear smoother as a combination of two optimum linear filters. *IEEE Trans. Automat. Contr.* **AC-14**, 387–390.

Gauthier, P., Courtier, P. and Moll, P. 1993. Assimilation of simulated wind lidar data with a Kalman filter. *Mon. Wea. Rev.* **121**, 1803–1820.

Gaspar, P. and Wunsch, C. 1989. Estimates from altimeter data of barotropic Rossby waves in the Northwestern Atlantic Ocean. *J. Phys. Oceanogr.* **19**, 1821–1844.

Gelb, A. 1974. *Applied optimal estimation.* Cambridge University Press. 374 pp.

Gel'fand, I. and Fomin, S. 1963. *Calculus of variations.* Prentice-Hall, Englewood Cliffs, 582 pp.

Hollingsworth, A. and Lönnberg, P. 1986. The statistical structure of short-range forecast errors as determined from radiosonde data. Part I: the wind field. *Tellus* **38A**, 111–136.

Jazwinski, A. 1970. *Stochastic processes and filtering theory.* Academic Press, New York, 343 pp.

Ménard, R. 1994. *Kalman filtering of Burgers' equation and its application to atmospheric data assimilation.* PhD Thesis, McGill University, Montréal.

Ménard, R., Lyster, P., Chang, L.-P. and Cohn, S. E. 1995. Middle atmosphere assimilation of UARS constituent data using a Kalman filter. *Proceedings of the 2nd International Symposium on Assimilation of observations in meteorology and oceanography.* Tokyo, Japan. World Meteorological Organization, pp. 235–238.

Moiseenko, A. and Saenko, O. 1992. Using the Kalman smoother to analyze and process oceanographic data. *Russ. J. Numer. Anal. Math. Modelling* 7, 241–255.

Molteni, F. and Palmer, T. 1993. Predictability and finite-time instability of the northern winter circulation. *Quart. J. Roy. Meteor. Soc.* **119**, 269–298.

Parrish, D. and Derber, J. 1992. The National Meteorological Center's spectral statistical interpolation analysis system. *Mon. Wea. Rev.* **120**, 1747–1763.

Pontryagin, L., Boltyanskii, V., Gamkrelidge, R. and Mishchenko, E. 1962. *The mathematical theory of optimal processes.* Translation by K. Trogonoff, edited by L. Neustadt, Wiley-Interscience, New York, 357 pp.

Rabier, F., Courtier, P., Pailleux, J., Talagrand, O. and Vasiljevic, D. 1993. A comparison between four-dimensional variational assimilation and simplified sequential assimilation, relying on three-dimensional variational analysis. *Quart. Journ. Roy. Met. Soc.* **119**, 845–880.

Rauch, H., Tung, F. and Striebel, C. 1965. Maximum likelihood estimates of linear dynamic systems. *AIAA Journal* **3**, 1445–1450.

Thacker, W. 1986. Relationship between statistical and deterministic methods of data assimilation. *Variational methods in the geosciences* (ed. Y. Sasaki). Elsevier, Amsterdam, 173–179.

Thépaut, J.-N., Vasiljevic, D., Courtier, P. and Pailleux, J. 1993. Variational assimilation of conventional observations with a multi-level primitive equations model. *Quart. Journ. Roy. Met. Soc.* **119**, 153–186.

Todling, R. and Cohn, S. E. 1994. Sub-optimal schemes for atmospheric data assimilation based on the Kalman filter. *Mon. Wea. Rev.* **122**, 2530–2557.

Wergen, W. 1992. The effect of model errors in variational assimilation. *Tellus* **44A**, 297–313.

Wunsch, C. 1988. Transient tracers as a problem in control theory. *J. Geophys. Res.* **93**, 8099–8110.

Zupanski, M. 1993. Regional four-dimensional variational data assimilation in a quasi-operational forecasting environment. *Mon. Wea. Rev.* **121**, 2396–2408.