Assimilation of Stratospheric Chemical Tracer Observations Using a Kalman Filter. Part I: Formulation

RICHARD MÉNARD

Data Assimilation Office, NASA Goddard Space Flight Center, Greenbelt, Maryland, and Joint Center for Earth System Technology, University of Maryland, Baltimore County, Catonsville, Maryland

STEPHEN E. COHN, LANG-PING CHANG,* AND PETER M. LYSTER+

Data Assimilation Office, NASA Goddard Space Flight Center, Greenbelt, Maryland

(Manuscript received 12 January 1999, in final form 13 September 1999)

ABSTRACT

The first part of this two-part article describes the formulation of a Kalman filter system for assimilating limb-sounding observations of stratospheric chemical constituents into a tracer transport model. The system is based on a two-dimensional isentropic approximation, permitting a full Kalman filter implementation and a thorough study of its behavior in a real-data environment. Datasets from two instruments on the *Upper Atmosphere Research Satellite* with very different viewing geometries are used in the assimilation experiments. A robust chi-squared diagnostic, which compares statistics of the observed-minus-forecast residuals with those calculated by the filter algorithm, is used to help formulate the statistical inputs to the filter, as well as to tune covariance parameters and to validate the assimilation results.

Two significant departures from the standard (discrete) Kalman filter formulation were found to be important in this study. First, it was discovered that the standard Kalman filter covariance propagation is highly inaccurate for this problem. Spurious and rapid loss of variance and increase of correlation length scales occur as a result of diffusion of the small-scale structures inherent in tracer error covariance fields. A new formulation based on well-understood properties of the continuum error covariance propagation was therefore introduced. Second, validation diagnostics suggested that the initial error, model error, and representativeness error are all more appropriately assumed to be relative than absolute in this problem. A filter formulation for relative errors was therefore devised. With these two modifications, this Kalman filter assimilation system has only three tunable variance parameters and one tunable correlation length-scale parameter.

1. Introduction

This article introduces the formulation of a Kalman filter system designed for assimilating chemical tracer observations obtained by sounding the earth limb. The system has been implemented on distributed-memory parallel computers (Lyster et al. 1997) and has been used to assimilate observations from instruments on board the *Upper Atmosphere Research Satellite* (*UARS*). It also serves the purpose of providing a benchmark for simplified chemical tracer assimilation schemes, such as schemes analogous to those described by Todling and Cohn (1994) and Cohn and Todling (1996) for meteorological problems. In a companion paper (Ménard and Chang 2000, referred to hereafter as Part II) we present results of *UARS* data assimilation experiments and discuss the relative importance of different components of the Kalman filter assimilation system.

The Kalman filter is often used as a paradigm to help understand sequential data assimilation schemes (e.g., Daley and Ménard 1993). Its extension to smoothing is also used to explain and help simplify other four-dimensional schemes such as 4D-Var (Courtier et al. 1994; Ménard and Daley 1996). However, due to its computational complexity, the Kalman filter in full extent can rarely be applied to realistic atmospheric data assimilation problems involving actual observations. As a result, little is known to date about how to formulate the Kalman filter appropriately for realistic problems and, in particular, about how to formulate its required sta-

* Additional affiliation: General Sciences Corporation (a subsidiary of Science Applications International Corporation), Beltsville, Maryland.

+ Additional affiliation: Earth System Science Interdisciplinary Center, University of Maryland at College Park, College Park, Maryland.

Corresponding author address: Dr. Richard Ménard, NASA GSFC, Data Assimilation Office, Code 910.3, Greenbelt, MD 20771.
E-mail: menard@dao.gsfc.nasa.gov

tistical inputs, such as the model error covariance. Consequently, knowledge about its actual behavior in practice is lacking.

In this article we present a formulation and application of the Kalman filter tailored specifically to our assimilation problem of chemical tracer observations taken from limb-sounding instruments. The assimilation system is based on a two-dimensional approximation, described in (A) below, but otherwise contains no computational simplification or approximation of the horizontal dynamics of the forecast error covariance. The two-dimensional approximation makes the Kalman filter implementation feasible with present-day computers. Our assimilation problem also has two important characteristics (C1 and C2):

(A) The assimilation of chemical tracer data obtained from limb-sounding instruments can be accurately approximated as a two-dimensional problem. Short-term prediction of long-lived chemical constituents is well approximated as a two-dimensional transport on individual isentropic surfaces. Soundings of the earth limb provide no information below and little information above the tangent point in the retrieved quantities; that is, vertical observational error correlation can be neglected.

(C1) The propagation of the forecast/analysis error covariance is linear and thus linear filtering theory applies. Both the transport model and the forward observation operator are linear operators. No moment closure assumption or tangent linear assumption is required.

(C2) The continuum dynamics of the forecast error covariance function and its main properties have been established by Cohn (1993). The forecast error variance satisfies transport dynamics, and the covariance function is a Lagrangian-conserved quantity.

One important finding of this article is that the standard (discrete) Kalman filter forecast error covariance evolution equation for tracer dynamics, which approximates the continuum forecast error covariance dynamics, can entail significant and completely spurious discretization error. This error turns out to be so large for our assimilation problem that it is comparable to all other combined sources of model error, which include the error of the driving winds as well as all physical and chemical effects not accounted for in pure two-dimensional transport (A), as we shall conclude in Part II. Part of our Kalman filter formulation is, therefore, a modification of the standard one to conform more closely to the continuum properties (C2).

Our formulation of the statistical inputs for the Kalman filter—the model and observational error covariances, and the initial state and its error covariance—is based partly upon aircraft measurements of spectral properties of tracer fields and partly upon the principle of minimizing *data shocks.* Free variance parameters are estimated using a *robust* χ^2 validation diagnostic (Bennett and Thornburn 1992; Dee 1995a). This diagnostic is robust in the sense that it is independent of the probability distribution of the observed-minus-forecast (OmF) residuals. Although the χ^2 diagnostic is most often used as a validation tool, our filter formulation has only one free variance parameter for the observational error, one for the model error, and one for the initial error, and we found that the χ^2 diagnostic could also be used for tuning these few parameters. This tuning has allowed us to conclude in particular, knowing the measurement error, that accounting for the representativeness error is important in our assimilation problem.

Assimilation experiments and χ^2 validation results have also suggested that the observational, model, and initial errors are all state dependent, and we have formulated a generalized Kalman filter algorithm accordingly. The standard algorithm assumes that these errors are not correlated with the signal (or true state). This lack of state dependence has been criticized as being inappropriate for atmospheric data assimilation. Errors arising from numerical discretization of the continuum dynamics and forward observation operators, for example, have been clearly related to the signal (Cohn and Dee 1988; Cohn 1997). This signal–error correlation yields, however, a larger set of Kalman filtering equations (Mitchell and Daley 1997a,b). Instead, we take a pragmatic approach of modeling state dependence directly with state estimates, either the forecast or the analysis. Our generalized Kalman filter algorithm then retains the standard form, as we show in the appendix.

The organization of this paper is as follows. In section 2 we present the theoretical background, including the stochastic hypotheses and the assumptions on state-dependent errors. We also introduce the χ^2 validation tool. In section 3 we describe the transport model, assess the numerical accuracy of the standard Kalman filter error covariance computation, and modify this computation to obey more closely the continuum forecast error covariance dynamics. To motivate the use of appropriate covariance models and to describe the *UARS* data, we present in section 4 the result of assimilation experiments using a naive Kalman filter configuration, accounting for neither the model error nor the representativeness error and using a simple initial error covariance. We identify appropriate covariance models for the initial, observational, and model errors in section 5, through a series of assimilation experiments. We summarize our Kalman filter formulation in section 6.

2. Theory of estimation of chemical tracer fields

a. Kalman filter assimilation method

In our assimilation problem, the *true state* $\boldsymbol{\mu}_k^t$ is a discrete representation of the continuum consistent with the transport model (Cohn 1997). It is a random vector

of mixing ratios, of dimension n, to be estimated on the basis of observations. The minimum variance estimate is a conditional expectation given the available observations (e.g., Jazwinski 1970; Cohn 1997). The minimum variance *analysis* estimate $\boldsymbol{\mu}_k^a$ at time t_k is defined as the n vector

$$\boldsymbol{\mu}_k^a \equiv \langle \boldsymbol{\mu}_k^t \mid \mathbf{m}_k^o \rangle, \tag{2.1}$$

where

$$\mathbf{m}_k^o \equiv \{\boldsymbol{\mu}_1^o, \boldsymbol{\mu}_2^o, \ldots, \boldsymbol{\mu}_k^o\} \tag{2.2}$$

denotes the set of all *observed values* of mixing ratio up to and including time t_k. The *observation vectors* $\boldsymbol{\mu}_i^o$ have dimension p_i. To extend the Kalman filter theory to state-dependent errors it is important to note that the conditional expectation $\langle \ \mid \mathbf{m}_k^o \rangle$ is made with respect to a *realization* of the observations, so that the analysis estimate $\boldsymbol{\mu}_k^a$ is *not* a random variable and is a function of the observed values. The minimum variance *forecast* estimate $\boldsymbol{\mu}_{k+1}^f$ is similarly defined as the conditional expectation of the true state at the next model time step, t_{k+1}, given all observed values up to and including time t_k,

$$\boldsymbol{\mu}_{k+1}^f \equiv \langle \boldsymbol{\mu}_{k+1}^t \mid \mathbf{m}_k^o \rangle. \tag{2.3}$$

Note again that the forecast estimate is *not* a random variable and is a function of the observed values.

The observations are related to the true state through an observation equation, taken to have the form

$$\boldsymbol{\mu}_k^o = \mathbf{H}_k \boldsymbol{\mu}_k^t + \boldsymbol{\epsilon}_k^o. \tag{2.4}$$

The observations are mixing ratios retrieved from limb-sounding instruments and, therefore, are of the same physical quantity as the prognostic variable of the transport model. The discrete forward observation operator, $\mathbf{H}_k$, is taken to be a simple bilinear interpolation on the latitude–longitude grid of the transport model. The observational error,

$$\boldsymbol{\epsilon}_k^o = \boldsymbol{\epsilon}_k^m + \boldsymbol{\epsilon}_k^r, \tag{2.5}$$

is composed of the measurement (retrieval) error, $\boldsymbol{\epsilon}_k^m$, and the representativeness error, $\boldsymbol{\epsilon}_k^r$. The measurement error includes effects due to detector noise, as well as forward and inverse modeling errors in the retrieval process. The measurement error variance can be estimated through error analysis of the retrieval scheme (Rodgers 1990) and is reported by the instrument teams. The representativeness error arises from subgrid-scale variability (Cohn 1997), numerical discretization error associated with the operator $\mathbf{H}_k$ (Daley 1993), and misspecification of the discrete forward operator, all of which depend on the state. The representativeness error covariance will be estimated through the OmF statistics using the χ^2 diagnostic, as described in section 5 and Part II.

To express state dependence of the representativeness error, we assume the following form:

$$\boldsymbol{\epsilon}_k^r = \mathbf{g}(\mathbf{H}_k \boldsymbol{\mu}_k^f) \circ \tilde{\boldsymbol{\epsilon}}_k^r, \tag{2.6}$$

where $\mathbf{g}$ denotes some vector function of the interpolated forecast state $\mathbf{H}_k \boldsymbol{\mu}_k^f$. The symbol $\circ$ denotes the component-wise (Hadamard) product; $\mathbf{g}$, $\tilde{\boldsymbol{\epsilon}}_k^r$, and $\boldsymbol{\epsilon}_k^r$ all have the dimension of the observation vector, $\boldsymbol{p}_k$. Here $\tilde{\boldsymbol{\epsilon}}_k^r$ is assumed to be state independent, white in time, normally distributed with zero mean and unit covariance, and uncorrelated with the measurement error. In the Kalman filter, the observational error covariance is actually defined as a conditional expectation (see the appendix). Under the stated assumptions, the observational error covariance is given by

$$\mathbf{R}_k^o = \mathbf{R}_k^m + \mathbf{R}_k^r, \tag{2.7}$$

where $\mathbf{R}_k^m$ is the measurement error covariance;

$$\begin{aligned} \mathbf{R}_k^r &\equiv \langle \boldsymbol{\epsilon}_k^r (\boldsymbol{\epsilon}_k^r)^{\mathrm{T}} \mid \boldsymbol{\mu}_k^t, \mathbf{m}_{k-1}^o \rangle \\ &= [\mathbf{g}(\mathbf{H}_k \boldsymbol{\mu}_k^f) \mathbf{g}(\mathbf{H}_k \boldsymbol{\mu}_k^f)^{\mathrm{T}}] \circ \mathbf{I}_{m_k}, \end{aligned} \tag{2.8}$$

is the representativeness error covariance, a diagonal matrix; and $\mathbf{I}_{m_k}$ is the $p_k \times p_k$ identity matrix. The form of $\mathbf{g}$ will be established later, in section 5. We assume $\mathbf{R}_k^m$ to be diagonal, and use the values of the measurement error variance reported by the instrument teams as the diagonal elements. The observational error covariance matrix $\mathbf{R}_k^o$ is therefore diagonal: we have assumed no spatial error correlation. We remark that although such an assumption may be unrealistic, establishing the error correlation length scale with existing methodology would require additional observations having spatially *uncorrelated* observational errors (e.g., Dee and da Silva 1999; Dee et al. 1999), which are unavailable.

To evolve the true state we discretize the (stochastic) dynamics of a conservative tracer on isentropic surfaces with random forcing,

$$\frac{\partial \mu}{\partial t} + \mathbf{V} \cdot \boldsymbol{\nabla} \mu = f, \tag{2.9}$$

where μ is the tracer mixing ratio, $\mathbf{V}$ is the horizontal wind field along an isentropic surface, and $\boldsymbol{\nabla}$ is the horizontal gradient on the isentropic surface. The random forcing term f represents, in part, the physical and chemical effects that are neglected in the isentropic advection dynamics but are present in the actual transport of long-lived species. Vertical and horizontal diffusion and chemical sources and sinks are among these effects. Wind error effects are also part of the random forcing f, since we use an analyzed wind field $\mathbf{V}$ instead of the true winds.

Discretizing (2.9) gives an evolution equation for $\boldsymbol{\mu}_k^t$. An additional forcing term arises from numerical discretization error (Cohn and Dee 1988). The resulting equation is written in the form

$$\boldsymbol{\mu}_{k+1}^t = \mathbf{M}_k \boldsymbol{\mu}_k^t + \boldsymbol{\epsilon}_k^q, \tag{2.10}$$

where the *model error* $\boldsymbol{\epsilon}_k^q$ represents collectively the numerical discretization error and the effect of f integrated from time t_k to t_{k+1}. We use the Lin and Rood

(1996) transport model based on the second-order-accurate van Leer scheme with the monotonic and positive constraints removed, yielding a linear transport model $\mathbf{M}_k$. This linearity is important to avoid moment closure problems in the filter algorithm.

Numerical experiments to be described in section 5 suggest that the model error $\boldsymbol{\epsilon}_k^q$ is state dependent. Similarly to the representativeness error, we assume a model error of the form

$$\boldsymbol{\epsilon}_k^q = \mathbf{f}(\boldsymbol{\mu}_k^q) \circ \tilde{\boldsymbol{\epsilon}}_k^q, \tag{2.11}$$

where $\mathbf{f}$ depends on the analysis state. Here $\tilde{\boldsymbol{\epsilon}}_k^q$ is assumed to be state independent, white in time, normally distributed with zero mean and unit variance, and uncorrelated with the measurement error and initial error. The model error covariance is given by

$$\mathbf{Q}_k \equiv \langle \boldsymbol{\epsilon}_k^q (\boldsymbol{\epsilon}_k^q)^{\mathrm{T}} \mid \mathbf{m}_k^o \rangle = [\mathbf{f}(\boldsymbol{\mu}_k^a)\mathbf{f}(\boldsymbol{\mu}_k^a)^{\mathrm{T}}] \circ \mathbf{C}_k^q, \tag{2.12}$$

where $\mathbf{C}_k^q$ is a correlation matrix. The initial error covariance, $\mathbf{P}_0$, is also taken to have a similar form, and will be described in section 5.

The Kalman filter algorithm that results from these state-dependent error assumptions is shown in the appendix to have the standard (Kalman filter) form. In particular it is shown that the Bayesian update leads to the usual analysis equation, given by (2.15) below. The algorithm is composed of two steps, as usual. In the analysis step, the analysis estimate (2.1) is calculated along with the analysis error covariance, defined by

$$\mathbf{P}_k^a \equiv \langle (\boldsymbol{\mu}_k^t - \boldsymbol{\mu}_k^a)(\boldsymbol{\mu}_k^t - \boldsymbol{\mu}_k^a)^{\mathrm{T}} \mid \mathbf{m}_k^o \rangle. \tag{2.13}$$

Similarly, the forecast estimate (2.3) and corresponding forecast error covariance,

$$\mathbf{P}_{k+1}^f \equiv \langle (\boldsymbol{\mu}_{k+1}^t - \boldsymbol{\mu}_{k+1}^f)(\boldsymbol{\mu}_{k+1}^t - \boldsymbol{\mu}_{k+1}^f)^{\mathrm{T}} \mid \mathbf{m}_k^o \rangle, \tag{2.14}$$

are calculated in the forecast step.

The analysis step consists of the equations

$$\boldsymbol{\mu}_k^a = \boldsymbol{\mu}_k^f + \mathbf{K}_k(\boldsymbol{\mu}_k^o - \mathbf{H}_k\boldsymbol{\mu}_k^f) \quad \text{and} \tag{2.15}$$

$$\mathbf{P}_k^a = (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)[(\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)\mathbf{P}_k^f]^{\mathrm{T}} + \mathbf{K}_k\mathbf{R}_k^o\mathbf{K}_k^{\mathrm{T}}, \tag{2.16}$$

where

$$\mathbf{K}_k = (\mathbf{H}_k\mathbf{P}_k^f)^{\mathrm{T}}[\mathbf{H}_k(\mathbf{H}_k\mathbf{P}_k^f)^{\mathrm{T}} + \mathbf{R}_k^o]^{-1} \tag{2.17}$$

is the *Kalman gain matrix*. The forecast step consists of the equations

$$\boldsymbol{\mu}_{k+1}^f = \mathbf{M}_k\boldsymbol{\mu}_k^a \quad \text{and} \tag{2.18}$$

$$\mathbf{P}_{k+1}^f = \mathbf{M}_k(\mathbf{M}_k\mathbf{P}_k^a)^{\mathrm{T}} + \mathbf{Q}_k. \tag{2.19}$$

The expression $\mathbf{M}_k(\mathbf{M}_k\mathbf{P}_k^a)^{\mathrm{T}}$ in (2.19) is often written as $\mathbf{M}_k\mathbf{P}_k^a\mathbf{M}_k^{\mathrm{T}}$. Actually, the transpose operator only appears as a result of how the covariance is stored, that is, as a matrix. In fact, in the expression $\mathbf{M}_k\mathbf{P}_k^a\mathbf{M}_k^{\mathrm{T}}$, the transposed model matrix $\mathbf{M}_k^{\mathrm{T}}$ or adjoint model is not used as an operator: $\mathbf{M}_k^{\mathrm{T}}$ is not used to left-multiply a vector or a matrix. Rather, it is $\mathbf{M}_k$ itself that is used to left-multiply the matrices $\mathbf{P}_k^a$ and $(\mathbf{M}_k\mathbf{P}_k^a)^{\mathrm{T}}$. It is a common mistake to believe that the Kalman filter algorithm uses an adjoint model.

Numerical implementation of the Kalman filter algorithm was made on distributed-memory parallel computers (Lyster et al. 1997) to obtain efficient throughput times. The code uses the covariance decomposition method of Lyster et al. (1997), and runs on the National Aeronautics and Space Administration (NASA) Goddard Space Flight Center Cray T3E using the Message Passing Interface. Currently, a throughput rate of about 2 min (wall clock time) per day of assimilation at $4^\circ \times 5^\circ$ resolution for a single isentropic level is achieved with 128 processors.

b. χ^2 validation

Optimality of the Kalman filter depends strongly on correct specification of the input error covariances, $\mathbf{R}_k^o$, $\mathbf{Q}_k$, and $\mathbf{P}_0$, which determine how accurately the covariances $\mathbf{P}_k^a$ and $\mathbf{P}_k^f$ computed according to the Kalman filter algorithm reflect the actual estimation error (Bennett and Thorburn 1992; Daley 1992; Dee 1995a,b). The χ^2 validation diagnostic presented here compares the sample covariance of the OmF residual,

$$\begin{aligned}\boldsymbol{\nu}_k &\equiv \boldsymbol{\mu}_k^o - \mathbf{H}_k\boldsymbol{\mu}_k^f = (\boldsymbol{\mu}_k^o - \mathbf{H}_k\boldsymbol{\mu}_k^t) - \mathbf{H}_k(\boldsymbol{\mu}_k^f - \boldsymbol{\mu}_k^t) \\ &= \boldsymbol{\epsilon}_k^o - \mathbf{H}_k\boldsymbol{\epsilon}_k^f,\end{aligned} \tag{2.20}$$

which is a combination of observational error $\boldsymbol{\epsilon}_k^o$ and forecast error $\boldsymbol{\epsilon}_k^f$, with the Kalman filter covariance counterpart,

$$\mathbf{S}_k \equiv \langle \boldsymbol{\nu}_k\boldsymbol{\nu}_k^{\mathrm{T}} \mid \mathbf{m}_{k-1}^o \rangle = \mathbf{H}_k(\mathbf{H}_k\mathbf{P}_k^f)^{\mathrm{T}} + \mathbf{R}_k^o. \tag{2.21}$$

Note that from the identity

$$\boldsymbol{\nu}_k = \boldsymbol{\mu}_k^o - \langle \boldsymbol{\mu}_k^o \mid \mathbf{m}_{k-1}^o \rangle, \tag{2.22}$$

the OmF residual can be interpreted as *that part of the current observation not already explained by past observations*, known also as the *innovation.*

The χ^2 quantities can be introduced in a variety of ways (see, e.g., Dee 1995a). For our purpose, it is useful to define χ^2 in the observation space as

$$\begin{aligned}\chi_k^2 &\equiv \boldsymbol{\nu}_k^{\mathrm{T}}\mathbf{S}_k^{-1}\boldsymbol{\nu}_k \\ &= (\boldsymbol{\mu}_k^o - \mathbf{H}_k\boldsymbol{\mu}_k^f)^{\mathrm{T}}[\mathbf{H}_k(\mathbf{H}_k\mathbf{P}_k^f)^{\mathrm{T}} + \mathbf{R}_k^o]^{-1}(\boldsymbol{\mu}_k^o - \mathbf{H}_k\boldsymbol{\mu}_k^f).\end{aligned} \tag{2.23}$$

Taking the conditional expectation of χ_k^2 yields

$$\begin{aligned}\langle \chi_k^2 \mid \mathbf{m}_{k-1}^o \rangle &= \langle \boldsymbol{\nu}_k^{\mathrm{T}}\mathbf{S}_k^{-1}\boldsymbol{\nu}_k \mid \mathbf{m}_{k-1}^o \rangle = \langle \mathrm{trace}[\mathbf{S}_k^{-1}\boldsymbol{\nu}_k\boldsymbol{\nu}_k^{\mathrm{T}}] \mid \mathbf{m}_{k-1}^o \rangle \\ &= \mathrm{trace}[\mathbf{S}_k^{-1}\langle \boldsymbol{\nu}_k\boldsymbol{\nu}_k^{\mathrm{T}} \mid \mathbf{m}_{k-1}^o \rangle] = \mathrm{trace}[\mathbf{S}_k^{-1}\mathbf{S}_k] \\ &= p_k.\end{aligned} \tag{2.24}$$

That is, the conditional mean of χ_k^2 is equal to the number of observations (Bryson and Ho 1975, section 12.2), provided that $\mathbf{S}_k$ is nonsingular. The importance of this result owes to the fact that (2.24) is a *robust* relationship: it neither assumes nor requires knowledge about the

probability distribution of the innovations. However, under an assumption that the innovations are Gaussian distributed, χ_k^2 is chi-square distributed with p_k degrees of freedom.

The relationships (2.23)–(2.24) can be used to verify the correctness of the innovation covariance matrix that uses the $\mathbf{P}_k^f$ computed by the Kalman filter algorithm and the $\mathbf{R}_k^o$ specified in the algorithm. As described in Part II, we normalize χ_k^2 by the number of observations p_k, replace the conditional mean by a time mean, and compare the result with unity. In fact, in section 5 we introduce variance parameters into the specification of the input error covariances $\mathbf{R}_k^o$, $\mathbf{Q}_k$, and $\mathbf{P}_0$, and in Part II we use the normalized χ_k^2 to tune these parameters. We remark that since temporal (serial) correlations do not alter time averages, the time mean χ_k^2 diagnostic is independent of any temporal correlation that might exist in the innovations.

The χ_k^2 condition (2.24) is a general diagnostic that can be used in a variety of assimilation schemes. This diagnostic generalizes the ''fit to the observations'' introduced by Ménard and Daley (1996) to include the forecast error covariance in addition to the observational error covariance. The χ_k^2 diagnostic has been applied to 4D-Var by Dee (1995a). In 3D-Var (Courtier et al. 1998), χ_k^2 in (2.23) corresponds to twice the cost function

$$J = \frac{1}{2}(\boldsymbol{\mu}_k^o - \mathbf{H}_k\boldsymbol{\mu}_k)^{\mathrm{T}}(\mathbf{R}_k^o)^{-1}(\boldsymbol{\mu}_k^o - \mathbf{H}_k\boldsymbol{\mu}_k) + \frac{1}{2}(\boldsymbol{\mu}_k - \boldsymbol{\mu}_k^f)^{\mathrm{T}}(\mathbf{P}_k^f)^{-1}(\boldsymbol{\mu}_k - \boldsymbol{\mu}_k^f), \quad (2.25)$$

evaluated at the minimum (Tarantola 1987, section 4.3.6). The evaluation of χ_k^2 in the Physical-space Statistical Analysis System (Cohn et al. 1998) is particularly inexpensive: using the solution $\mathbf{y}_k$ of the equation

$$(\mathbf{H}_k\mathbf{P}_k^f\mathbf{H}_k^{\mathrm{T}} + \mathbf{R}_k^o)\mathbf{y}_k = \boldsymbol{\nu}_k, \quad (2.26)$$

χ_k^2 is just the inner product

$$\chi_k^2 = \boldsymbol{\nu}_k^{\mathrm{T}}\mathbf{y}_k. \quad (2.27)$$

3. Transport of the error covariance

a. Transport model

The transport model is an adaptation of the model of Lin and Rood (1996). The Lin and Rood model solves the tracer equation in conservation form to evolve density. To apply this model to the advection dynamics of mixing ratio μ, we write the advection equation in the form

$$\frac{\partial\mu}{\partial t} + \boldsymbol{\nabla}\cdot(\mu\mathbf{V}) = \mu\boldsymbol{\nabla}\cdot\mathbf{V}, \quad (3.1)$$

which is solved by operator splitting (Strang 1968). At each time step, we solve

$$\frac{\partial\mu}{\partial t} + \boldsymbol{\nabla}\cdot(\mu\mathbf{V}) = 0, \quad (3.2)$$

using the second-order van Leer scheme of Lin and Rood (1996), and then solve

$$\frac{\partial\mu}{\partial t} = \mu\boldsymbol{\nabla}\cdot\mathbf{V} \quad (3.3)$$

with the backward Euler scheme, using the result of (3.2) as the initial condition. Note that time discretization of (3.3) is actually not necessary as (3.3) admits an analytical resolvent of the form $\exp(\Delta t\boldsymbol{\nabla}\cdot\mathbf{V})$. Use of the analytical resolvent showed, however, little difference in transport results when compared with the backward Euler scheme. Although the original transport problem formulated for the continuum is linear in μ, Eq. (2.9), the discrete transport model of Lin and Rood (1996) is nonlinear due to monotonic and positive constraints. These constraints are, however, nondifferentiable and render ill-defined the linearization of the model (Polavarapu et al. 1996; Xu 1996). To avoid this problem, we removed the monotonic and positive constraints from the transport model.

The transport model is driven off-line by wind analyses obtained from the Goddard Earth Observing System Data Assimilation System (GEOS DAS) version 1 (Pfaendtner et al. 1995) in a tropospheric–stratospheric configuration. This configuration has a horizontal resolution of $4^\circ \times 5^\circ$ with 46 levels (18 analysis levels) and the uppermost level at 0.4 hPa. In GEOS DAS an Incremental Analysis Update (IAU; Bloom et al. 1996) is used, which acts as an effective time smoother. Time smoothing has been recommended for off-line transport models driven by analyzed winds (Waugh et al. 1997). The IAU analysis winds, stored at 6-hourly intervals, are linearly interpolated in time to the transport model time step. In the vertical, the winds are linearly interpolated in the logarithm of potential temperature to the isentropic surfaces. Above 10 hPa, a sixth-order Shapiro (1970) filter has been applied in the zonal direction to the winds after all interpolations have been performed, to reduce small-scale noise. The transport model runs at the same horizontal resolution as the wind analyses.

b. Covariance dynamics

Although technically correct for the discrete Kalman filter, we have found that the $\mathbf{M}_k(\mathbf{M}_k\mathbf{P}_k^a)^{\mathrm{T}}$ computation in (2.19) gives rise to significantly large discretization error. It has been argued from theoretical considerations by Cohn and Dee (1988) and then by Cohn (1993) and Mitchell and Daley (1997a,b) that this computation could give rise to significant discretization error, the size of which has remained unknown in practice. We first sized the importance of this error when we began validating the Kalman filter analyses in unobserved regions. The analysis error variance computed by the Kalman filter indicated that the analysis had much more

skill (i.e., lower analysis error variance) in jet regions, for instance, than what was actually supported by comparison with a control dataset. Spurious results of this type will be shown later in this section.

To understand this problem, let us first consider the evolution of the covariance *function* corresponding to the continuum dynamics (2.9). In the absense of random forcing f, the covariance function obeys generalized advection dynamics in a four-dimensional space (Cohn 1993), twice that of the spatial dimension of the transport problem (2.9):

$$\frac{\partial P}{\partial t} + \mathbf{V}_1 \cdot \boldsymbol{\nabla}_1 P + \mathbf{V}_2 \cdot \boldsymbol{\nabla}_2 P = 0, \qquad (3.4)$$

where $P = P(\mathbf{x}_1, \mathbf{x}_2, t)$ is the covariance function of two points, $\mathbf{x}_1 = (\lambda_1, \phi_1)$ and $\mathbf{x}_2 = (\lambda_2, \phi_2)$. Here $\mathbf{V}_1 = [u(\mathbf{x}_1, t), v(\mathbf{x}_1, t)]$ is the wind vector at the point $\mathbf{x}_1$, and similarly for $\mathbf{V}_2$, and $\boldsymbol{\nabla}_1$, $\boldsymbol{\nabla}_2$ are the horizontal gradient operators with respect to the points $\mathbf{x}_1$ and $\mathbf{x}_2$. Since the operators $\mathbf{V}_1 \cdot \boldsymbol{\nabla}_1$ and $\mathbf{V}_2 \cdot \boldsymbol{\nabla}_2$ commute, the resolvent operator of (3.4) is the product of two resolvent operators, one solving the advection problem with respect to $\mathbf{x}_1$ and the other with respect to $\mathbf{x}_2$ (Cohn 1993, appendix B), which is in fact operator splitting. If we were to discretize each of these resolvents as a discrete operator $\mathbf{M}_k$, we would get $\mathbf{P}^f_{k+1} = \mathbf{M}_k(\mathbf{M}_k\mathbf{P}^a_k)^{\mathrm{T}}$ with $\mathbf{P}^a_k$ and $\mathbf{P}^f_{k+1}$ stored as matrices, which is precisely the discrete Kalman filter covariance evolution equation. If, on the other hand, we were to discretize (3.4) directly, we would in general obtain a different and possibly more accurate result. For example, solving a two-dimensional advection equation by using operator splitting with one-dimensional advection schemes is known to have adverse consequences in general (e.g., Lin and Rood 1996, section 2). Solving the two-dimensional equation in a manner that obeys more directly the physics generally leads to superior results.

To assess the numerical accuracy of the covariance computed from $\mathbf{M}_k(\mathbf{M}_k\mathbf{P}^a_k)^{\mathrm{T}}$, we use some properties of the analytical solution of the continuum covariance evolution equation (3.4). Equation (3.4) can be solved by the method of characteristics. Its solution has the property that the covariance between a pair of material particles is conserved during transport (Cohn 1993); that is, the covariance function is a Lagrangian-conserved quantity. Consequently, the variance $V(\mathbf{x}, t) \equiv P(\mathbf{x}, \mathbf{x}, t)$, which is the covariance of a material particle with itself, is conserved along the particle trajectory. That is to say, the variance obeys the advection dynamics:

$$\frac{\partial V}{\partial t} + \mathbf{V} \cdot \boldsymbol{\nabla} V = 0. \qquad (3.5)$$

Let us compare the variance obtained from $\mathbf{M}_k(\mathbf{M}_k\mathbf{P}^a_k)^{\mathrm{T}}$ with the variance obtained using the advection dynamics (3.5) directly. Using a spatially uniform initial variance field $V \equiv 1$, an integration on the 1100-K isentropic surface was performed in pure transport mode, that is, with no data assimilation, using the GEOS DAS winds from 0000 UTC 6 September to 0000 UTC 10 September 1992. A first-order autoregressive (FOAR) correlation model [see Eq. (5.2)] with a length scale of 3600 km was used as the initial correlation. This length scale is typical and a discussion on how it was obtained will be given in Part II. According to the Lagrangian conservative property, the variance should remain uniform and of unit value regardless of the advecting wind field. The results are displayed in Fig. 1. In Fig. 1a the variance obtained from $\mathbf{M}_k(\mathbf{M}_k\mathbf{P}^a_k)^{\mathrm{T}}$ is depicted. We observe a significant decrease in magnitude from unity, especially in the southern midlatitudes with values below 0.5. Figure 1b displays the variance obtained by applying the discrete transport model directly to the advection equation (3.5). A departure from unity of at most 3% is seen. These results show that the standard covariance evolution equation (2.19) used in discrete Kalman filtering can give rise to a significant and spurious *loss of variance.*

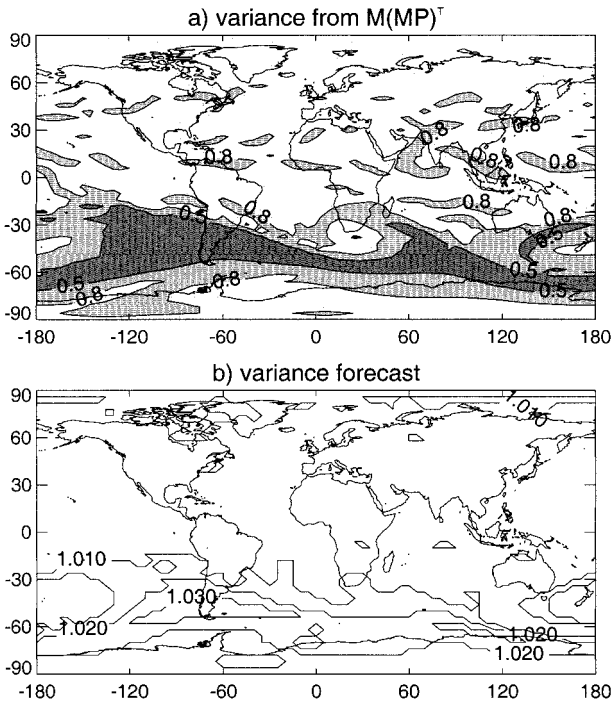

FIG. 1. Forecast error variance after 4 days, starting from unit variance. (a) The variance obtained from using $\mathbf{M}(\mathbf{MP})^{\mathrm{T}}$, and (b) the variance obtained by using the advection equation for variance.

To shed some light on this loss of variance, let us take a closer look at the $\mathbf{M}_k(\mathbf{M}_k\mathbf{P}^a_k)^{\mathrm{T}}$ computation, in which the covariance evolution is obtained by applying the transport model on each column of $\mathbf{P}$ and then on each row. Each column or row of $\mathbf{P}$ represents the covariance between a grid point and all the other grid points of the domain. These columns and rows have generally a near-zero value away from the diagonal, which rises rapidly over a relatively short distance to-

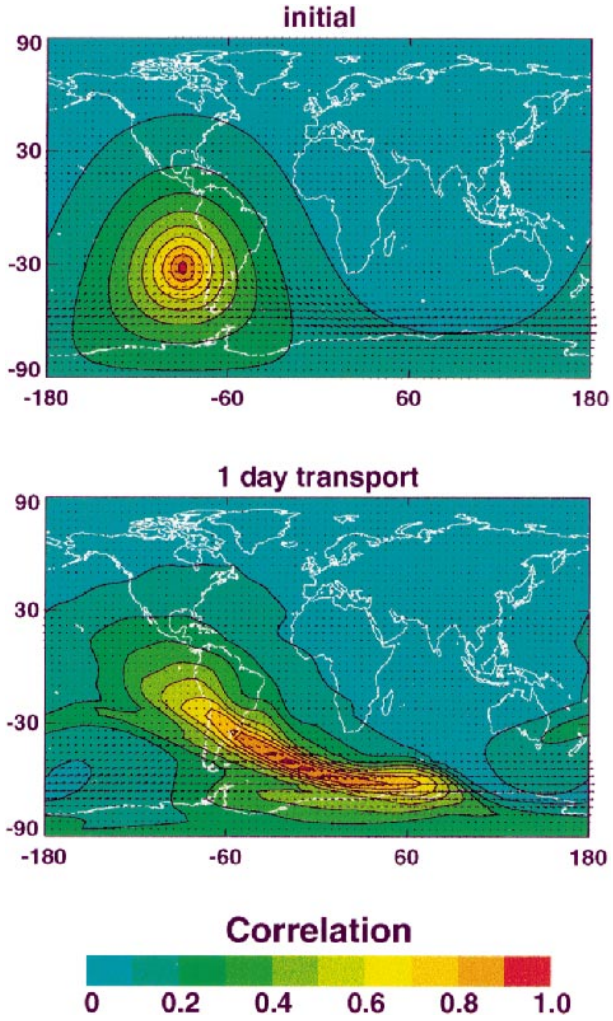

FIG. 2. Error correlation with respect to the reference point located at (32°S, 90°W). (a) The initial FOAR correlation, and (b) the forecast error correlation after 1 day using $\mathbf{M}(\mathbf{MP})^{\mathrm{T}}$.

ward the diagonal of the $\mathbf{P}$ matrix, as shown in Fig. 2 with respect to the reference point located at (32°S, 90°W). Figure 2a shows the FOAR initial condition, and Fig. 2b shows the correlation after one day of covariance evolution; the covariance has been normalized by the computed variance to yield the correlation. Shearing of the wind field, also depicted in Fig. 2, results in contraction and stretching of the correlation pattern in different directions. The correlation field in Fig. 2b has a much finer structure than constituent fields for which the model was designed to provide accurate simulations. The operation $\mathbf{M}_k(\mathbf{M}_k\mathbf{P}_k^a)^{\mathrm{T}}$ involves, in particular, differencing across the diagonal of $\mathbf{P}$, giving rise to diffusion of the peak of the covariance structure, manifested as diffusion of the variance. The maximum loss of variance seen in Fig. 1a corresponds to the region of strongest wind shear and, thus, of largest gradients near the diagonal of $\mathbf{P}$. Additional experiments (results not shown) have demonstrated that this loss of variance also increases with decreasing initial correlation length, as expected.

To remedy the loss of variance arising in the computation $\mathbf{M}_k(\mathbf{M}_k\mathbf{P}_k^a)^{\mathrm{T}}$, a corrected-*variance forecast* error covariance matrix, denoted by $\mathbf{P}_k^{fv}$, is obtained by evolving the variance according to (3.5) using the transport model, with the *correlation* obtained from $\mathbf{M}_k(\mathbf{M}_k\mathbf{P}_k^a)^{\mathrm{T}}$. That is, neglecting the model error for the moment, we compute first

$$\tilde{\mathbf{P}}_{k+1}^f = \mathbf{M}_k(\mathbf{M}_k\mathbf{P}_k^a)^{\mathrm{T}} \quad \text{and} \tag{3.6}$$

$$\tilde{\mathbf{V}}_{k+1}^f = \mathbf{M}_k\mathbf{V}_k^a, \tag{3.7}$$

where $\mathbf{V}_k^a(i) = \mathbf{P}_k^a(i, i)$ is the analysis error variance at the grid point $\mathbf{x}_i$. Because small negative variance values can result from (3.7) when no monotonic and positive constraints are used in the transport model, advection of the logarithm of the variance was used instead, with the appropriate transformations at the beginning and the end of each time step. The corrected-variance forecast error covariance matrix is then given by

$$\mathbf{P}_{k+1}^{fv}(i, j) = \sqrt{\tilde{\mathbf{V}}_{k+1}^f(i)\tilde{\mathbf{V}}_{k+1}^f(j)}\,\tilde{\mathbf{C}}_{k+1}^f(i, j), \tag{3.8}$$

where

$$\tilde{\mathbf{C}}_{k+1}^f(i, j) = \frac{\tilde{\mathbf{P}}_{k+1}^f(i, j)}{\sqrt{\tilde{\mathbf{P}}_{k+1}^f(i, i)\tilde{\mathbf{P}}_{k+1}^f(j, j)}}. \tag{3.9}$$

Another important property of the covariance evolution follows from the continuum dynamics (3.4). In the absence of model error covariance, and for *nondivergent winds*, the *total covariance* $[P]$ defined as the integral over the four-dimensional covariance space,

$$[P] \equiv \int_{S_1}\int_{S_2} P(\mathbf{x}_1, \mathbf{x}_2)\, dS_1\, dS_2, \tag{3.10}$$

is a conservative quantity,

$$\frac{\partial[P]}{\partial t} = 0. \tag{3.11}$$

Here dS_i is the element of surface area on the sphere S_i. In Fig. 3 we show the evolution of the total covariance starting at the same time and on the same isentrope as in the previous experiment, still with no model error, but with the divergent part of the wind removed. The solid curve corresponds to the standard form of the Kalman filter, that is, using $\mathbf{M}_k(\mathbf{M}_k\mathbf{P}_k)^{\mathrm{T}}$. The total covariance is constant because the transport model of Lin and Rood (1996) is mass conserving. The dashed curve corresponds to the corrected-variance evolution scheme (3.6)–(3.9). The steady *increase* of the total covariance represented by the dashed curve results from the fact that this scheme compensates for the loss of variance in the denominator of (3.9). We thus conclude that the spurious numerical diffusion in the computation of $\mathbf{M}_k(\mathbf{M}_k\mathbf{P}_k)^{\mathrm{T}}$ produces not only a loss of variance but also an increase of the correlation length scales. To correct for this increase of the correlation length scales in the

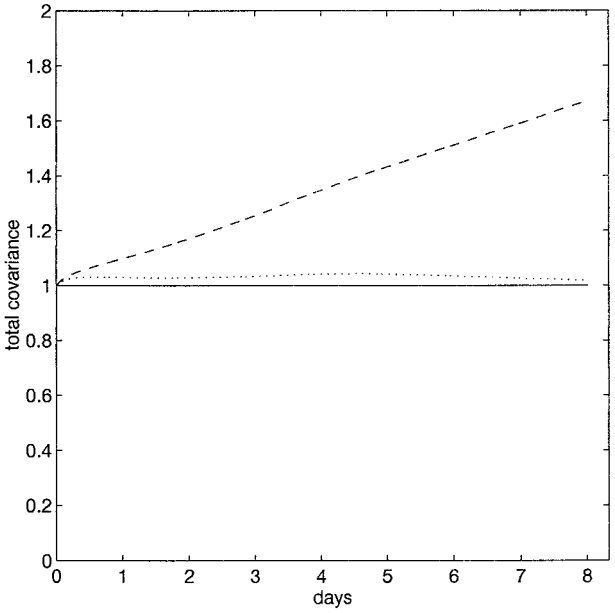

FIG. 3. Evolution of the total covariance [P]. The solid curve corresponds to the standard discrete Kalman filter. The dashed curve corresponds to the corrected-variance evolution scheme (3.6)–(3.9). The dotted curve is obtained by using further the Hadamard product (3.12).

corrected-variance evolution scheme (3.6)–(3.9), we take the Hadamard product of $\mathbf{P}^{fv}$ and a ''prescribed correlation model'' $\mathbf{C}$, such as an isotropic FOAR model, yielding a *corrected* forecast error covariance matrix,

$$\mathbf{P}^{c}_{k+1}(i, j) = \mathbf{P}^{fv}_{k+1}(i, j)\mathbf{C}(i, j), \tag{3.12}$$

that has smaller length scales than either of the two covariance matrices being multiplied (Gaspari and Cohn 1999). The dotted curve in Fig. 3 depicts the evolution of the total covariance using an isotropic FOAR correlation model $\mathbf{C}$ with a very large length scale, $L = 8 \times 10^9$ m. This length scale, obtained by trial and error, was chosen to yield a nearly constant total covariance. The multiplication (3.12) represents only a small correction *per time step* since this L is more than a thousand times the radius of the earth.

Summarizing, we replace the standard Kalman filter covariance evolution on equation (2.19) by

$$\mathbf{P}^{f}_{k+1} = \mathbf{P}^{c}_{k+1} + \mathbf{Q}_k, \tag{3.13}$$

with $\mathbf{C}$ and L as described above. This scheme propagates variance directly, according to the continuum variance dynamics, and very nearly conserves the total covariance. It will be applied throughout this study in data assimilation mode, and its effect will be demonstrated in Part II. The corrected scheme is admittedly crude, and in particular it cannot be validated readily in data assimilation mode. Numerical schemes discretizing the continuum covariance dynamics (3.4) more directly could be developed in the future and may address the issues of covariance dynamics more properly.

4. A simple assimilation experiment

The Cryogenic Limb Array Etalon Spectrometer (CLAES) and the Halogen Occultation Experiment (HALOE) are two *UARS* limb-sounding instruments that provide very different observation patterns, and will be used throughout this paper to compare and validate results of the Kalman filter. The CLAES instrument measures thermal emission from the earth limb, retrieving about 1300 vertical profiles per day for several chemical constituents (Roche et al. 1993). To minimize exposure to the sun, thus preventing heating of the infrared instruments, the spacecraft changes its orientation with respect to its orbit each 36 days, from a fly-forward position to a fly-backward position and vice versa (Reber 1993). This results for CLAES in two different regions of latitudinal coverage, 80°S–34°N (south limb view) and 34°S–80°N (north limb view), alternating each 36 days. The HALOE instrument, on the other hand, provides more limited coverage as it operates with a solar occultation technique (Russell et al. 1993). The observing pattern of HALOE consists of about 15 profiles per day along each of two latitude circles, usually in opposite hemispheres. With orbital precession and revolution, nearly complete latitudinal coverage is obtained in roughly a month and a half. The positions of the CLAES and HALOE observations for a single day, on 8 September 1992, are depicted in Fig. 4.

In this paper the retrieved mixing ratios for HALOE CH_4 (version 18) and CLAES CH_4 (version 8) are used as the observations. These limb-sounding retrievals are sensitive to the gas mixing ratio close to the tangent point over a distance typically of about 400 km along the ray path, which is about the size of our transport model grid. There is also averaging in the direction of satellite motion, due to the duration of the measurement, that does not exceed 500 km. Since the aperture function is therefore barely resolvable by the transport model grid, we have neglected its effect and treated the retrieved information as point observations, using a simple bilinear interpolation for the discrete forward observation operator $\mathbf{H}_k$. We should note that the retrievals assume spherical symmetry along the ray path, so that the retrieved data have little dependence on the prior (Rodgers 1990; Marks and Rodgers 1993), and are thus suitable for data assimilation (Joiner and da Silva 1998). We have not attempted to account for this assumption nor to assimilate the radiances.

The instrument error is generally obtained by perturbing the input radiance with random test error values, and its standard deviation has been reported as ''quality'' in the level 3AT dataset. Detailed analyses of errors in HALOE CH_4 and CLAES CH_4 retrievals are given by Park et al. (1996) and Roche et al. (1996), respectively. The relative error, or quality/mixing ratio, is a nearly constant quantity characterizing the retrieval error for a given instrument and gas.

In this first assimilation experiment, we have made

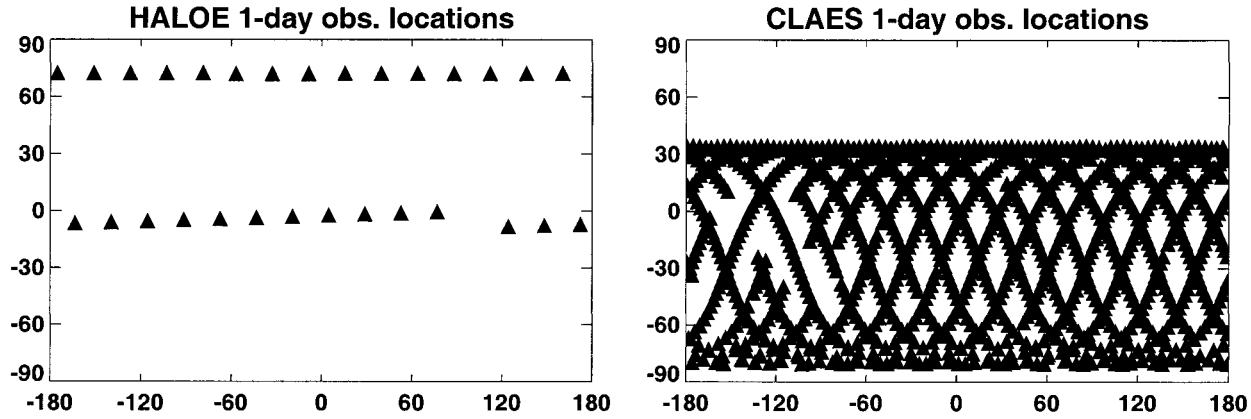

FIG. 4. Observational coverage for (a) HALOE and (b) CLAES for 8 Sep 1992.

no attempt to specify any input covariance information based on the innovation statistics. Instead, we have taken the instrument error to be the sole type of observational error and have assumed no model error. The measurement error is about 0.4% for HALOE CH_4 at 1100 K, and about 7.3% for CLAES CH_4. For the initial condition, we used the zonal, monthly mean of the CLAES observational data (binned into 4° latitude bands) over the preceding month, and the variance about this mean. Values near the poles were obtained by extrapolation with a cubic spline with zero slope at the poles. As in the experiment of section 3, the initial error correlation was taken to be the FOAR model (5.2) with a length scale of 3600 km. This length scale was obtained by using the maximum-likelihood method (Dee 1995b), as discussed in Part II.

Experiments assimilating either CLAES or HALOE CH_4 mixing ratio data were carried out on the 1100-K isentropic surface starting again at 0000 UTC 6 September 1992. The results after 4 and 5 days of assimilation of CLAES data only are depicted in Figs. 5a and 5b, respectively. The CLAES instrument is in south limb view during this period, providing coverage between 80°S and 34°N (Fig. 4) and, thus, sees directly the ''tongue'' of tropical air breaking out into the southern midlatitudes. The experiment was repeated with the HALOE observations that, throughout the period, remained limited to the tropical region and to a narrow latitude band near 72°N (Fig. 4). Results are depicted in Figs. 5c and 5d. The CLAES and HALOE observations differ systematically by a factor of about 1.4, and the color bars have been adjusted to account for this observational bias between the two instruments. The initial condition for the HALOE assimilation was also modified by dividing the mean and variance by 1.4 and $(1.4)^2$, respectively.

Since there are no HALOE observations in the southern extratropics and northern midlatitudes, comparison of the CLAES and HALOE analyses can serve as a basis to assess the ability of the Kalman filter to infer information in these unobserved regions. Unrealistically large mixing ratio values are obtained with the HALOE assimilation on day 5 in these regions, particularly throughout most of the Southern Hemisphere, which contrasts with a relatively successful analysis on day 4. We conclude that the simple error covariances for $\mathbf{P}_0$, $\mathbf{R}_k$, and $\mathbf{Q}_k$ (set to zero) produce Kalman filter assimilation results that are not reliable for sparse observations.

5. Modeling of the input error covariances

As noted in section 2b and illustrated in the previous section, successful implementation of a Kalman filter depends on correct specification of the input error covariances, $\mathbf{R}_k^o$, $\mathbf{Q}_k$, and $\mathbf{P}_0$. In this section we present the rationale for and the description of our choice of covariance models, along with the choice of initial state $\boldsymbol{\mu}_0$.

a. Initial conditions

The initial condition for a Kalman filter (2.15)–(2.19) consists of the unconditional mean estimate $\boldsymbol{\mu}_0 = \langle \boldsymbol{\mu}_0^t \rangle$ of the initial state, and the corresponding unconditional error covariance $\mathbf{P}_0 = \langle (\boldsymbol{\mu}_0^t - \langle \boldsymbol{\mu}_0^t \rangle)(\boldsymbol{\mu}_0^t - \langle \boldsymbol{\mu}_0^t \rangle)^{\mathrm{T}} \rangle$. The unconditional mean estimate is the minimum variance estimate obtained *without observations.* Such an estimate can be thought of as being provided by a long-term integration of a three-dimensional chemical transport model with sources and sinks of CH_4. This simulation would not depend directly on observations except possibly through an initial condition whose effect should become negligible after a long period of time. For practical reasons, we used instead our two-dimensional tracer transport of CH_4 over a period of time, starting from a zonal state described in section 4. The integration period was chosen to be those 16 days prior to the initial time of the assimilation. We found this period to be long enough to yield a ''spunup'' tracer field that is consistent with the winds, but not so long

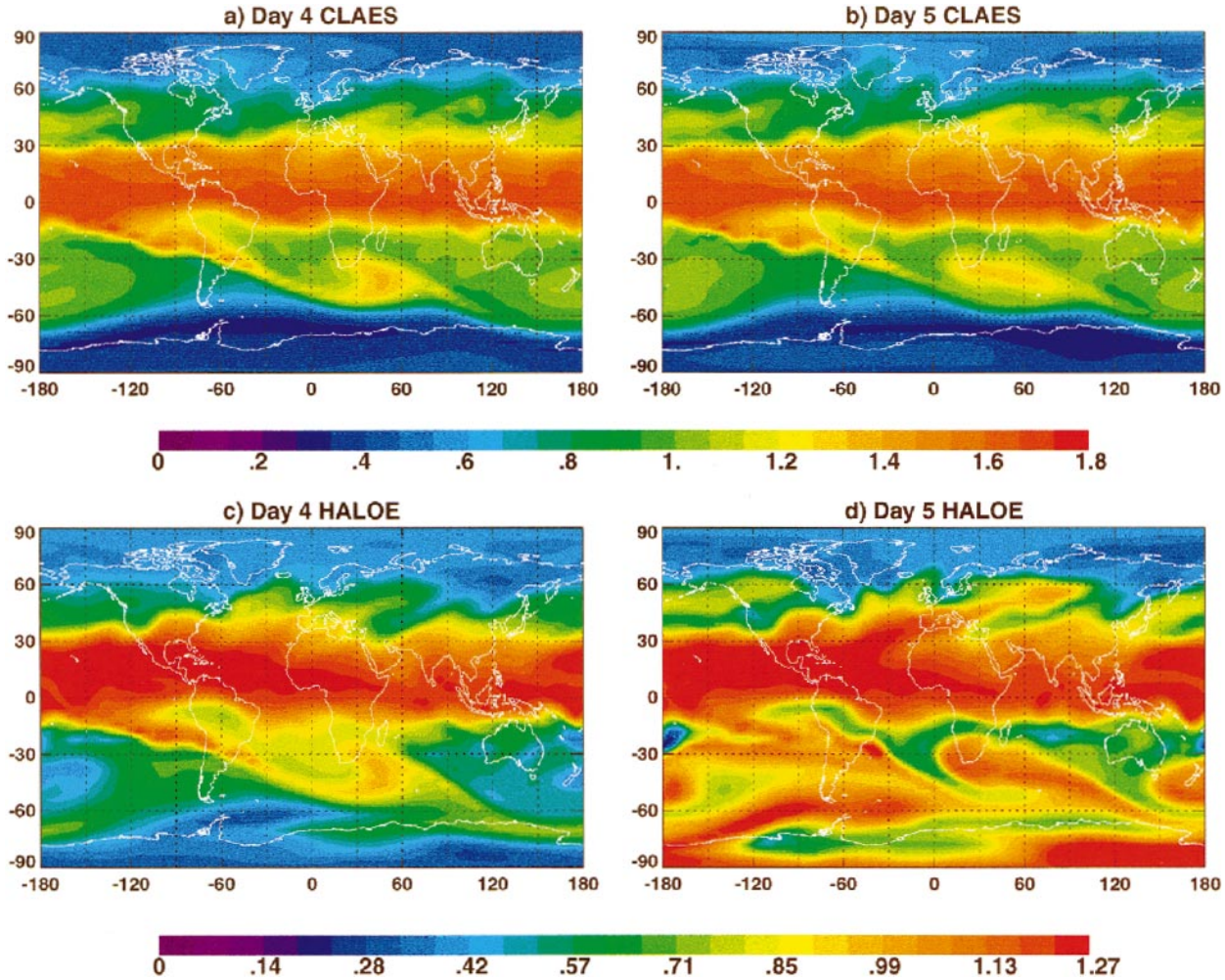

FIG. 5. Assimilation of methane observations using a zonal initial condition, no model error, and instrument error only as the observational error. (a) and (b) The CLAES assimilation, and (c) and (d) the HALOE assimilation.

that numerical diffusion degrades the large-scale structure.

The initial error covariance is decomposed into a variance and a correlation, which we specify separately. In the absence of model error, the variance obeys the advection dynamics (3.5), and so does its square root, the standard deviation. Consequently, they both behave as tracers. Different tracer fields are often significantly correlated, differing mainly by a constant of proportionality (e.g., Schoeberl and Lait 1992; Riishøjgaard and Källén 1997). We used this property as an assumption to prescribe the initial error standard deviation, by taking it to be proportional to the initial CH_4 mixing ratio. The constant of proportionality, γ, was assigned a value of 12% based on the global root-mean-square deviation about the zonal state described in section 4.

The initial error correlation was chosen to be isotropic on the sphere. The correlation model was chosen on the basis of its spectral characteristics, which in turn dictate the spectral characteristics of the analysis increments (Daley 1991, section 3.3) and, therefore, of the mixing ratio analyses. One-dimensional spectra of long-lived chemicals obtained from stratospheric aircraft measurements at altitudes of about 25 km typically show spectral slopes of about -2 at scales of tens to a few thousand kilometers (Nastrom et al. 1986; Ngan and Shepherd 1997). The FOAR correlation model (5.2), whose spectrum $\mathcal{R}^1$ decays asymptotically as -2, was chosen for the initial error correlation model, although our assimilation is carried out at 1100 K (about 38 km of altitude). Contrary to most correlation models used in meteorological data assimilation, such as the second-order autoregressive and Gaussian models (e.g., Daley 1991), the FOAR correlation model is nondifferentiable, having a cusp at the origin. Cusplike features have, in fact, been noted in autocorrelation functions of stratospheric aircraft measurements of chemical tracers (Bacmeister et al. 1997).

The initial error covariance matrix thus takes the form

$$\mathbf{P}_0(i, j) = \gamma^2 \boldsymbol{\mu}_0(i)\boldsymbol{\mu}_0(j)\mathbf{C}(i, j), \qquad (5.1)$$

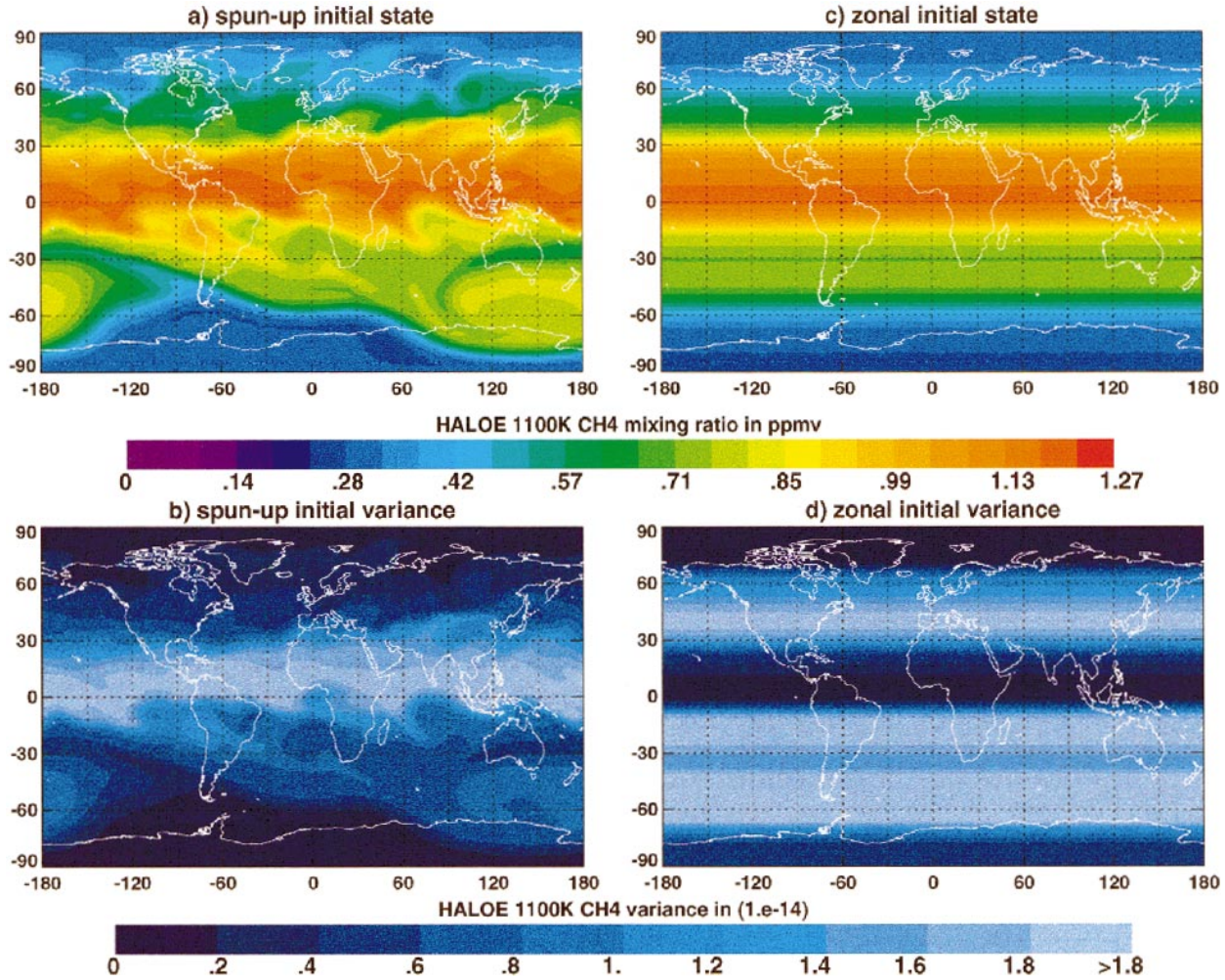

FIG. 6. Initial mixing ratio and error variance for HALOE assimilations. (a) and (b) The spunup initial state and corresponding error variance described in section 5a, respectively. (c) and (d) The zonal initial state and error variance described in section 4. The maximum values in (d) are 5.5×10^{-14} for the band around 38°N, 4.0×10^{-14} for the band around 20°S, and 6.8×10^{-14} for the band around 60°S. These bands of high variance are related to the wave activity along the subtropical edges and in the surf zone.

where the correlation matrix $\mathbf{C}$ is given by the FOAR model

$$\mathbf{C}(i, j) = \exp\left(-\frac{|\mathbf{r}_i - \mathbf{r}_j|}{L}\right). \tag{5.2}$$

Here $\mathbf{r}_i$ and $\mathbf{r}_j$ are the position vectors of two points $\mathbf{x}_i$ and $\mathbf{x}_j$ on the sphere, $|\cdot|$ denotes Euclidean distance in $\mathcal{R}^3$ (chordal distance on the sphere), and L is the correlation length scale. Use of position vectors restricted to the sphere guarantees that the correlation matrix is positive definite and periodic on the sphere (Gaspari and Cohn 1999).

Using the initial state described above and the initial error covariance (5.1)–(5.2), depicted in Figs. 6a and 6b, we carried out an 8-day assimilation of HALOE CH_4 data on the 1100-K isentrope starting again at 0000 UTC 6 September 1992. This experiment is identical to the previous HALOE experiment, described in section 4, except for the initial conditions, shown for comparison in Figs. 6c and 6d. The new experiment resulted in a smooth time evolution of the mixing ratio analysis, the zonal mean of which has been plotted for conciseness in Fig. 7a. This contrasts with the evolution obtained in the previous HALOE experiment, plotted in Fig. 7b, which shows considerable noise beyond day 4 associated with the Southern Hemisphere shear zone; compare with Fig. 5d.

These experiments were conducted with no model error and an observational error variance that is small relative to the initial error variance, that is, about 0.004^2 versus 0.12^2. In addition, the total number of observations up to eight days, about 240, is far less than the number of state variables $n = 3312$ for our transport model at $4° \times 5°$ resolution. Therefore, the observations are insufficient to determine the state after eight days, and the analysis still depends on the initial conditions

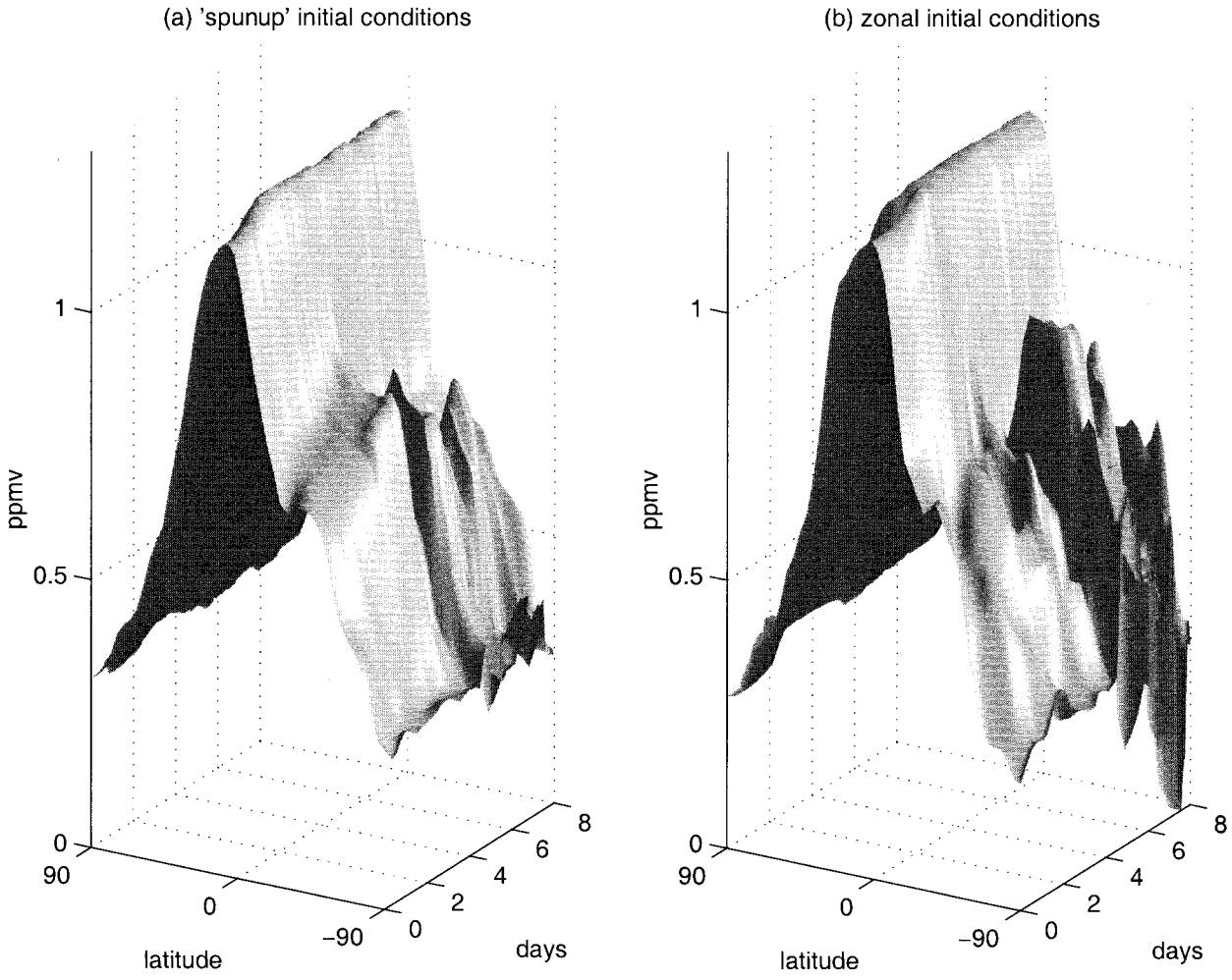

FIG. 7. Zonal mean time evolution of HALOE CH_4 analyses. (a) The result starting from the spunup initial conditions described in section 5a. (b) The result of starting from the zonal mean initial conditions of section 4.

(e.g., Cohn and Dee 1988; Ménard 1994; Ménard et al. 1995; Daley 1997). Furthermore, because the observational error variance is small, the analysis increments can be large, and *data shocks*, as evident in Fig. 5d, can occur if the initial conditions are inappropriate. Our experimental results lead us to the conclusion that appropriate initial conditions under these stringent circumstances, and therefore under most circumstances, can be obtained by using a spunup tracer field, an error standard deviation proportional to the mixing ratio, and the FOAR correlation model.

b. Representativeness error

Two covariance models were considered for the representativeness error, an *absolute error model* and a *relative error model.* The choice of model will be made on the basis of a regional χ^2_k diagnostic in HALOE assimilation experiments.

In the absolute error model, the function $\mathbf{g}$ in (2.6) is simply the product of a dimensionless constant α and a typical global mean mixing ratio $\overline{\mu}$. The representativeness error covariance is then given by

$$\mathbf{R}_k^r = \alpha^2 \overline{\mu}^2 \mathbf{I}_{m_k}, \tag{5.3}$$

according to (2.8). In the relative error model we choose the function $\mathbf{g}$ to be linear, so that $\mathbf{g}(\mathbf{H}_k \boldsymbol{\mu}_k^f) = \beta(\mathbf{H}_k \boldsymbol{\mu}_k^f)$, where the dimensionless constant β is the relative error standard deviation. The representativeness error covariance is then given by

$$\mathbf{R}_k^r = \boldsymbol{\beta}^2 \mathbf{R}_k, \tag{5.4}$$

where

$$\begin{aligned} \mathbf{R}_k &= [(\mathbf{H}_k \boldsymbol{\mu}_k^f)(\mathbf{H}_k \boldsymbol{\mu}_k^f)^{\mathrm{T}}] \circ \mathbf{I}_{m_k} \\ &= \mathrm{diag}[(\mathbf{H}_k \boldsymbol{\mu}_k^f)_1^2, \ldots, (\mathbf{H}_k \boldsymbol{\mu}_k^f)_{m_k}^2]. \end{aligned} \tag{5.5}$$

Here, $\mathrm{diag}(a_1, \ldots, a_n)$ is a diagonal matrix whose diagonal elements are $a_1, \ldots, a_n$.

We have conducted two series of 8-day HALOE CH_4

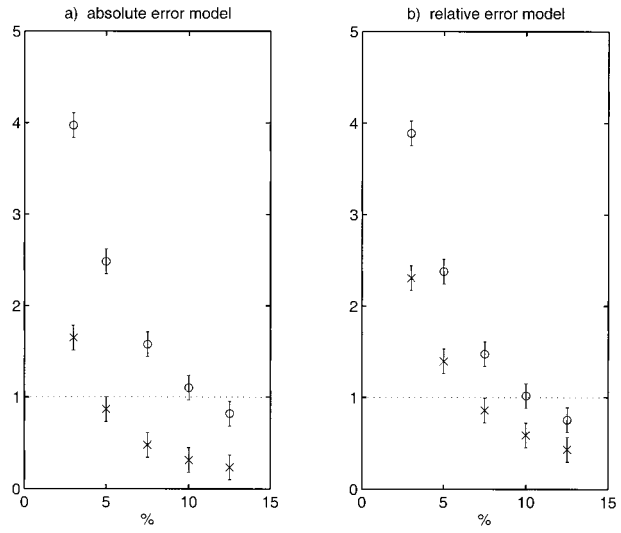

FIG. 8. Normalized mean χ_k^2 from HALOE CH_4 assimilation experiments. The circles depict the normalized mean χ_k^2 using northern latitude observations, and crosses depict the normalized mean χ_k^2 using the tropical observations. Estimation error of one σ is represented by error bars. (a) The results using the absolute error model, where the abscissa is α (in %). (b) The results using the relative error model, where the abscissa is β (in %).

assimilation experiments, each experiment being identical to that of the previous subsection except for the inclusion of the representativeness error covariance model. In the first series of experiments, the absolute error model was used with values $\alpha = 0.03$, 0.05, 0.075, 0.1, and 0.125. The χ_k^2 diagnostic (2.23) was carried out separately for two groups of data, the northern latitude and the tropical innovations (see Fig. 4a), with the mean χ_k^2 normalized by the number of observations in each group. The χ_k^2 results are presented in Fig. 8a. Circles (crosses) indicate the normalized χ_k^2 for the northern (tropical) innovations; the estimation error bar of 1σ assumes Gaussian-distributed innovations. We note large discrepancies between the results for northern latitude and tropical innovations. In the second series of experiments, we used the relative error model formulation with $\beta = 0.03$, 0.05, 0.075, 0.10, and 0.125. The results depicted in Fig. 8b indicate a better agreement between the χ_k^2 values for the northern latitude and tropical innovations. In fact, for $\beta \geq 0.075$, the χ_k^2 values agree within less than 2σ. The normalized χ_k^2 is closest to one for $\beta \cong 0.086$.

These results suggest that the relative representative error formulation is more appropriate than the absolute formulation. Another finding is that the relative representativeness error for HALOE CH_4 of about 8.6% is much larger than the relative measurement error, which is about 0.4%. This value of the relative representativeness error is also in rough agreement with the ''data repeatability'' obtained for HALOE CH_4 at the same altitude, near 4 hPa (see Park et al. 1996). We choose the relative error model for all subsequent assimilation experiments, with $\beta = 0.086$ in the case of HALOE CH_4.

We remark that the remaining discrepancy between the northern latitude and tropical χ^2 results could possibly be ameliorated by choosing the function **g** of (2.6) to depend on derivatives of the mixing ratio field, to reflect interpolation errors in the observation operators $\mathbf{H}_k$. While our formulation of the Kalman filter accommodates such dependence, we have not tried such a model of the representativeness error.

c. Model error

It is important in general to include effects of model error in a Kalman filter. First and foremost, when model error is neglected, the computed forecast error variance becomes unrealistically small, and therefore the filter ultimately stops ''paying attention'' to new data—a problem known as filter divergence (e.g., Daley and Ménard 1993). Second, accounting for model error prevents the analysis from always having sensitivity to the initial error covariance specification.

Although little is known about the model error covariance in general, we know that it will impact the forecast error covariance at each time step. Rather than attempting to account for each individual source of error (e.g., Houtekamer et al. 1996), a potentially cumbersome task, we decided to adopt the relative formulation of the initial error covariance with the purpose of minimizing data shocks. Consequently the model error covariance matrix has the form

$$\mathbf{Q}_k(i, j) = \delta^2 \boldsymbol{\mu}_k^q(i) \boldsymbol{\mu}_k^q(j) \mathbf{C}^q(i, j), \tag{5.6}$$

where δ is a *relative error* parameter and $\mathbf{C}^q$ is the model error correlation matrix, held constant in time. In accordance with the general formulation, Eq. (2.12), the model error covariance matrix prescribed here depends on the current analysis, and in particular assumes the model error std dev to be proportional to the analysis field. The relative error parameter δ will be determined by χ^2 tuning in Part II. The model error correlation model $\mathbf{C}^q$ has been chosen to be identical to the initial error correlation model (5.2), with the same value of the length-scale parameter.

In lieu of prescribing a model error covariance explicitly, we also implemented a fading memory filter (Sorenson and Sacks 1971), in which the computed forecast error covariance is simply multiplied by a scalar slightly greater than one. This is equivalent to giving less weight to old observations and is a well-known technique in engineering to prevent filter divergence (e.g., Jazwinski 1970). The application of this scheme to our data assimilation problem, however, gave rise to exponential growth of variance away from the observations and therefore created large and unrealistic analysis increments away from the observations (results not shown). This approach was thus abandoned in favor of adding a model error covariance matrix (5.6).

6. Conclusions

We have introduced the formulation of a Kalman filter system for the assimilation of chemical tracer observations obtained from limb-sounding instruments. The system is tailored to this specific application. A two-dimensional approximation on isentropic surfaces has been made for the whole assimilation problem, permitting full computation of the forecast and analysis error covariances. We thus neglect the vertical error correlation in the observations and in the dynamics, diabatic effects, vertical mixing, and also chemistry as well as sources and sinks of the constituents. The approximation in the chemical transport, including wind errors, is accommodated by a model error covariance in the Kalman filter algorithm.

The Kalman filter requires specification of the initial, observational, and model error covariances. The design of these covariances followed from a series of assimilation experiments. To validate and tune covariance parameters, we introduced a robust χ^2 diagnostic that compares the sample covariance of the innovations with the innovation covariance obtained from the Kalman filter algorithm.

To formulate the initial error covariance, we set up an experimental configuration acutely sensitive to the initial covariance specification by neglecting model error and the representativeness error component of the observational error, and by assimilating only HALOE observations, which are accurate but sparse. Smooth evolution of the CH_4 mixing ratio field, with little data shock, was obtained by taking the initial error std dev to be proportional to a spunup initial mixing ratio field. This initial field was obtained from an integration of the transport model for the 16 days prior to the initial time of the assimilation period. The initial error correlation was chosen on the grounds that the spectral response to a single observation should be in agreement with the spectrum of the state itself, which is known from aircraft measurements to have a spectral slope near -2 at the relevant spatial scales. This argument suggested the use of a FOAR correlation.

The model error covariance was designed on the basis of the same considerations. The std dev was taken to be proportional to the analysis state, and the FOAR correlation was used.

The observational error, assumed to be horizontally uncorrelated, was decomposed into a measurement error and a representativeness error. The measurement error variance was provided by the *UARS* instrument teams. By monitoring the χ^2 diagnostic in different geographical regions in the assimilation experiments, we deduced that taking the representativeness error std dev to be proportional to the forecast state is more appropriate than taking it to be constant for all observations.

These findings about error characterization suggested state-dependent initial, observational, and model errors, which dictated a revision of the original Kalman filter theory. We found, however, that the Kalman filter retains its usual form as long as the state dependence is expressed in terms of the conditional expectations, that is, given the forecast or the analysis states, which is a pragmatic approach.

The numerical accuracy of the standard (discrete) Kalman filter forecast error covariance computation was also assessed against known properties of the continuum forecast error covariance dynamics. We found a strong loss of variance and an increase of correlation length scales arising from spurious dissipation of small-scale covariance structures by the transport model. This was especially severe in regions of strong wind shear. An alternative computational scheme without these defects was developed on the basis of the theory of Cohn (1993) and was applied to all the experiments in this study and in Part II.

Our Kalman filter formulation reduces the number of unknowns to one correlation length-scale parameter and three variance parameters: an FOAR length scale, a relative initial error standard deviation, a relative representativeness error standard deviation, and a relative model error standard deviation. This Kalman filter will be used for assimilation experiments in Part II.

Acknowledgments. The authors wish to thank Ricky Rood for his encouragement and support for this project. We are especially grateful to Roger Daley for his insightful comments, enthusiasm, and support during the whole course of this research. Lars Peter Riishøjgaard, Lynn Sparling, Susan Strahan, and Roger Daley, as well as three anonymous referees, are also thanked for their careful reviews of the manuscript. This research was supported by NASA Grants 578-41-52-20 and 579-21-32-20, the Canadian Atmospheric Environment Service (RM), the EOS Interdisciplinary Investigation on Data Assimilation (SEC), and the NASA High Performance Computing and Communications Program Earth and Space Sciences project (PML).

APPENDIX

A Derivation of the Kalman Filter for State-Dependent Errors

a. Analysis step

The analysis equation of the Kalman filter is based on an update of the *prior* conditional probability density function $p(\boldsymbol{\mu}_k^t \mid \mathcal{M}_{k-1}^o = \mathbf{m}_{k-1}^o)$ to yield the *posterior* probability density function $p(\boldsymbol{\mu}_k^t \mid \mathcal{M}_k^o = \mathbf{m}_k^o)$, for a given *realization* $\mathbf{m}_k^o$ of the set $\mathcal{M}_k^o$ of all possible values of the observation vectors from time t_1 to t_k. We distinguish between the set $\mathcal{M}_k^o$ of *random* observation vectors, that is, $\mathcal{M}_k^o \equiv \{\mathbf{M}_1^o, \ldots, \mathbf{M}_k^o\}$, where $\mathbf{M}_k^o$ is the random observation vector at time t_k, and its realization $\mathbf{m}_k^o \equiv \{\boldsymbol{\mu}_1^o, \ldots, \boldsymbol{\mu}_k^o\}$, where $\boldsymbol{\mu}_k^o$ is the vector of *observed* values or realization, $\mathbf{M}_k^o = \boldsymbol{\mu}_k^o$, of the observation vector $\mathbf{M}_k^o$. To derive this update of conditional probabilities

we first use Bayes' theorem for the random observation vectors. Using the definition of conditional probability density function (pdf), we get

$$p(\boldsymbol{\mu}_k^t \,|\, \mathcal{M}_k^o) = p(\boldsymbol{\mu}_k^t \,|\, \mathbf{M}_k^o, \mathcal{M}_{k-1}^o) = \frac{p(\boldsymbol{\mu}_k^t, \mathbf{M}_k^o, \mathcal{M}_{k-1}^o)}{p(\mathbf{M}_k^o, \mathcal{M}_{k-1}^o)}$$
$$= \frac{p(\mathbf{M}_k^o \,|\, \boldsymbol{\mu}_k^t, \mathcal{M}_{k-1}^o) p(\boldsymbol{\mu}_k^t, \mathcal{M}_{k-1}^o)}{p(\mathbf{M}_k^o, \mathcal{M}_{k-1}^o)}$$
$$= \frac{p(\mathbf{M}_k^o \,|\, \boldsymbol{\mu}_k^t, \mathcal{M}_{k-1}^o) p(\boldsymbol{\mu}_k^t, \mathcal{M}_{k-1}^o) p(\mathcal{M}_{k-1}^o)}{p(\mathbf{M}_k^o \,|\, \mathcal{M}_{k-1}^o) p(\mathcal{M}_{k-1}^o)}$$
$$= \frac{p(\mathbf{M}_k^o \,|\, \boldsymbol{\mu}_k^t, \mathcal{M}_{k-1}^o) p(\boldsymbol{\mu}_k^t \,|\, \mathcal{M}_{k-1}^o)}{p(\mathbf{M}_k^o \,|\, \mathcal{M}_{k-1}^o)}. \tag{A.1}$$

In the identity (A.1) the conditional pdf's can always be restricted to a realization $\mathcal{M}_k^o = \mathbf{m}_k^o$ of the observations. Thus we have

$$p(\boldsymbol{\mu}_k^t \,|\, \mathcal{M}_k^o = \mathbf{m}_k^o) = \frac{p(\boldsymbol{\mu}_k^o \,|\, \boldsymbol{\mu}_k^t, \mathcal{M}_{k-1}^o = \mathbf{m}_{k-1}^o) p(\boldsymbol{\mu}_k^t \,|\, \mathcal{M}_{k-1}^o = \mathbf{m}_{k-1}^o)}{p(\boldsymbol{\mu}_k^o \,|\, \mathcal{M}_{k-1}^o = \mathbf{m}_{k-1}^o)}. \tag{A.2}$$

Assuming, as usual in Kalman filtering, that the (relative) observational error $\boldsymbol{\epsilon}_k$ is Gaussian distributed, we will show that a Gaussian-distributed prior pdf yields a Gaussian-distributed posterior pdf. The evaluation of each term in (A.2) follows.

The prior density $p(\boldsymbol{\mu}_k^t \,|\, \mathcal{M}_{k-1}^o = \mathbf{m}_{k-1}^o)$, with conditional mean $\boldsymbol{\mu}_k^f \equiv \langle \boldsymbol{\mu}_k^t \,|\, \mathcal{M}_{k-1}^o = \mathbf{m}_{k-1}^o \rangle$ and conditional covariance $\mathbf{P}_k^f \equiv \langle (\boldsymbol{\mu}_k^t - \boldsymbol{\mu}_k^f)(\boldsymbol{\mu}_k^t - \boldsymbol{\mu}_k^f)^{\mathrm{T}} \,|\, \mathcal{M}_{k-1}^o = \mathbf{m}_{k-1}^o \rangle$, is written as

$$p(\boldsymbol{\mu}_k^t \,|\, \mathcal{M}_{k-1}^o = \mathbf{m}_{k-1}^o) = \frac{1}{\sqrt{(2\pi)^n |\mathbf{P}_k^f|}} \exp\left[-\frac{1}{2} (\boldsymbol{\mu}_k^t - \boldsymbol{\mu}_k^f)^{\mathrm{T}} (\mathbf{P}_k^f)^{-1} (\boldsymbol{\mu}_k^t - \boldsymbol{\mu}_k^f) \right]. \tag{A.3}$$

Conditional expectations such as $\langle \boldsymbol{\mu}_k^t \,|\, \mathcal{M}_{k-1}^o = \mathbf{m}_{k-1}^o \rangle$ are abbreviated in the text as $\langle \boldsymbol{\mu}_k^t \,|\, \mathbf{m}_{k-1}^o \rangle$.

To evaluate $p(\boldsymbol{\mu}_k^o \,|\, \boldsymbol{\mu}_k^t, \mathcal{M}_{k-1}^o = \mathbf{m}_{k-1}^o)$, let us first refer to the observation equation,

$$\boldsymbol{\mu}_k^o = \mathbf{H}_k \boldsymbol{\mu}_k^t + \mathbf{g}_k(\mathbf{H}_k \boldsymbol{\mu}_k^f) \circ \boldsymbol{\epsilon}_k. \tag{A.4}$$

We remark first that, by definition, the forecast $\boldsymbol{\mu}_k^f \equiv \langle \boldsymbol{\mu}_k^t \,|\, \mathcal{M}_{k-1}^o = \mathbf{m}_{k-1}^o \rangle$ is not a random variable. Thus in (A.4) for a given $\boldsymbol{\mu}_k^t$ and realization $\mathcal{M}_{k-1}^o = \mathbf{m}_{k-1}^o$, $\boldsymbol{\mu}_k^o$ depends only on the single random variable $\boldsymbol{\epsilon}_k$, which is Gaussian distributed. The conditional pdf $p(\boldsymbol{\mu}_k^o \,|\, \boldsymbol{\mu}_k^t, \mathcal{M}_{k-1}^o = \mathbf{m}_{k-1}^o)$ is thus Gaussian and its mean and covariance is obtained as follows:

$$\langle \boldsymbol{\mu}_k^o \,|\, \boldsymbol{\mu}_k^t, \mathcal{M}_{k-1}^o = \mathbf{m}_{k-1}^o \rangle$$
$$= \langle \mathbf{H}_k \boldsymbol{\mu}_k^t \,|\, \boldsymbol{\mu}_k^t, \mathcal{M}_{k-1}^o = \mathbf{m}_{k-1}^o \rangle + \langle \mathbf{g}_k(\mathbf{H}_k \langle \boldsymbol{\mu}_k^t \,|\, \mathcal{M}_{k-1}^o = \mathbf{m}_{k-1}^o \rangle) \circ \boldsymbol{\epsilon}_k \,|\, \boldsymbol{\mu}_k^t, \mathcal{M}_{k-1}^o = \mathbf{m}_{k-1}^o \rangle$$
$$= \mathbf{H}_k \boldsymbol{\mu}_k^t + \mathbf{g}_k(\mathbf{H}_k \langle \boldsymbol{\mu}_k^t \,|\, \mathcal{M}_{k-1}^o = \mathbf{m}_{k-1}^o \rangle) \circ \langle \boldsymbol{\epsilon}_k \,|\, \boldsymbol{\mu}_k^t, \mathcal{M}_{k-1}^o = \mathbf{m}_{k-1}^o \rangle = \mathbf{H}_k \boldsymbol{\mu}_k^t \tag{A.5}$$

and

$$\langle (\boldsymbol{\mu}_k^o - \mathbf{H}_k \boldsymbol{\mu}_k^t)(\boldsymbol{\mu}_k^o - \mathbf{H}_k \boldsymbol{\mu}_k^t)^{\mathrm{T}} \,|\, \boldsymbol{\mu}_k^t, \mathcal{M}_{k-1}^o = \mathbf{m}_{k-1}^o \rangle$$
$$= \langle \mathbf{g}_k(\mathbf{H}_k \boldsymbol{\mu}_k^f) \circ \boldsymbol{\epsilon}_k \boldsymbol{\epsilon}_k^{\mathrm{T}} \circ \mathbf{g}_k^{\mathrm{T}}(\mathbf{H}_k \boldsymbol{\mu}_k^f) \,|\, \boldsymbol{\mu}_k^t, \mathcal{M}_{k-1}^o = \mathbf{m}_{k-1}^o \rangle$$
$$= \mathbf{R}_k \circ [\mathbf{g}_k(\mathbf{H}_k \boldsymbol{\mu}_k^f) \mathbf{g}_k^{\mathrm{T}}(\mathbf{H}_k \boldsymbol{\mu}_k^f)] \equiv \mathbf{R}_k^o. \tag{A.6}$$

Note that in (A.5) we assumed $\boldsymbol{\epsilon}_k$ to be of zero mean, and independent of the true state and of the observations, so that $\langle \boldsymbol{\epsilon}_k \,|\, \boldsymbol{\mu}_k^t, \mathcal{M}_{k-1}^o = \mathbf{m}_{k-1}^o \rangle = \langle \boldsymbol{\epsilon}_k \rangle = 0$. Also in (A6) we used $\langle \boldsymbol{\epsilon}_k \boldsymbol{\epsilon}_k^{\mathrm{T}} \rangle = \mathbf{R}_k$. The conditional pdf $p(\boldsymbol{\mu}_k^o \,|\, \boldsymbol{\mu}_k^t, \mathcal{M}_{k-1}^o = \mathbf{m}_{k-1}^o)$ can thus be written as

$$p(\boldsymbol{\mu}_k^o \,|\, \boldsymbol{\mu}_k^t, \mathcal{M}_{k-1}^o = \mathbf{m}_{k-1}^o) = \frac{1}{\sqrt{(2\pi)^{p_k} |\mathbf{R}_k^o|}} \times \exp\left[-\frac{1}{2} (\boldsymbol{\mu}_k^o - \mathbf{H}_k \boldsymbol{\mu}_k^t)^{\mathrm{T}} (\mathbf{R}_k^o)^{-1} (\boldsymbol{\mu}_k^o - \mathbf{H}_k \boldsymbol{\mu}_k^t) \right]. \tag{A.7}$$

To evaluate the conditional pdf of $\boldsymbol{\mu}_k^o$ given $\mathcal{M}_{k-1}^o = \mathbf{m}_{k-1}^o$, that is, $p(\boldsymbol{\mu}_k^o \,|\, \mathcal{M}_{k-1}^o = \mathbf{m}_{k-1}^o)$, we note that from (A.4)

$$\langle \boldsymbol{\mu}_k^o \,|\, \mathcal{M}_{k-1}^o = \mathbf{m}_{k-1}^o \rangle$$
$$= \mathbf{H}_k \langle \boldsymbol{\mu}_k^t \,|\, \mathcal{M}_{k-1}^o = \mathbf{m}_{k-1}^o \rangle + \langle \mathbf{g}_k(\mathbf{H}_k \langle \boldsymbol{\mu}_k^t \,|\, \mathcal{M}_{k-1}^o = \mathbf{m}_{k-1}^o \rangle) \circ \boldsymbol{\epsilon}_k \,|\, \mathcal{M}_{k-1}^o = \mathbf{m}_{k-1}^o \rangle$$
$$= \mathbf{H}_k \boldsymbol{\mu}_k^f + \mathbf{g}_k(\mathbf{H}_k \langle \boldsymbol{\mu}_k^t \,|\, \mathcal{M}_{k-1}^o = \mathbf{m}_{k-1}^o \rangle) \circ \langle \boldsymbol{\epsilon}_k \,|\, \mathcal{M}_{k-1}^o = \mathbf{m}_{k-1}^o \rangle = \mathbf{H}_k \boldsymbol{\mu}_k^f. \tag{A.8}$$

Also, the conditional covariance is given by

$$\langle (\boldsymbol{\mu}_k^o - \mathbf{H}_k \boldsymbol{\mu}_k^f)(\boldsymbol{\mu}_k^o - \mathbf{H}_k \boldsymbol{\mu}_k^f)^{\mathrm{T}} \,|\, \mathcal{M}_{k-1}^o = \mathbf{m}_{k-1}^o \rangle$$
$$= \langle (\mathbf{H}_k \boldsymbol{\epsilon}_k^f + \mathbf{g}_k(\mathbf{H}_k \boldsymbol{\mu}_k^f) \circ \boldsymbol{\epsilon}_k) \times (\mathbf{H}_k \boldsymbol{\epsilon}_k^f + \mathbf{g}_k(\mathbf{H}_k \boldsymbol{\mu}_k^f) \circ \boldsymbol{\epsilon}_k)^{\mathrm{T}} \,|\, \mathcal{M}_{k-1}^o = \mathbf{m}_{k-1}^o \rangle$$
$$= \mathbf{H}_k (\mathbf{H}_k \mathbf{P}_k^f)^{\mathrm{T}} + \mathbf{R}_k^o, \tag{A.9}$$

where we assumed that the forecast error $\boldsymbol{\epsilon}_k^f \equiv \boldsymbol{\mu}_k^t - \boldsymbol{\mu}_k^f$ is uncorrelated with $\boldsymbol{\epsilon}_k$. From (A.4) we establish that the conditional pdf $p(\boldsymbol{\mu}_k^o \,|\, \mathcal{M}_{k-1}^o = \mathbf{m}_{k-1}^o)$ is Gaussian since $\boldsymbol{\mu}_k^o$ is the sum of two Gaussian-distributed random vectors, $\mathbf{H}_k \boldsymbol{\mu}_k^t$ and $\mathbf{g}_k(\mathbf{H}_k \boldsymbol{\mu}_k^f) \circ \boldsymbol{\epsilon}_k$. Thus we have

$$p(\boldsymbol{\mu}_k^o \mid \mathcal{M}_{k-1}^o = \mathbf{m}_{k-1}^o) = \frac{1}{\sqrt{(2\pi)^{p_k}|\mathbf{H}_k(\mathbf{H}_k\mathbf{P}_k^f)^{\mathrm{T}} + \mathbf{R}_k^o|}} \times \exp\left\{-\frac{1}{2}(\boldsymbol{\mu}_k^o - \mathbf{H}_k\boldsymbol{\mu}_k^f)^{\mathrm{T}}[\mathbf{H}_k(\mathbf{H}_k\mathbf{P}_k^f)^{\mathrm{T}} + \mathbf{R}_k^o]^{-1}(\boldsymbol{\mu}_k^o - \mathbf{H}_k\boldsymbol{\mu}_k^f)\right\}. \tag{A.10}$$

Having derived the three conditional pdf's in (A.2), namely (A.3), (A.7), and (A.10), the posterior pdf can now be calculated using a standard derivation of the analysis step for Gaussian-distributed random vectors found in many references on estimation theory (e.g., Jazwinski 1970, chapter 7; Cohn 1997). The posterioir pdf $p(\boldsymbol{\mu}_k^t \mid \mathcal{M}_k^o = \mathbf{m}_k^o)$ is Gaussian since the pdf's (A.3), (A.7), and (A.10) are Gaussian. The analysis, or conditional mean $\boldsymbol{\mu}_k^a \equiv \langle \boldsymbol{\mu}_k^t \mid \mathcal{M}_k^o = \mathbf{m}_k^o \rangle$, is given by

$$\boldsymbol{\mu}_k^a = \boldsymbol{\mu}_k^f + \mathbf{K}_k(\boldsymbol{\mu}_k^o - \mathbf{H}_k\boldsymbol{\mu}_k^f), \tag{A.11}$$

where

$$\mathbf{K}_k = (\mathbf{H}_k\mathbf{P}_k^f)^{\mathrm{T}}[\mathbf{H}_k(\mathbf{H}_k\mathbf{P}_k^f)^{\mathrm{T}} + \mathbf{R}_k^o]^{-1}. \tag{A.12}$$

The analysis error covariance $\mathbf{P}_k^a \equiv \langle (\boldsymbol{\mu}_k^t - \boldsymbol{\mu}_k^a)(\boldsymbol{\mu}_k^t - \boldsymbol{\mu}_k^a)^{\mathrm{T}} \mid \mathcal{M}_k^o = \mathbf{m}_k^o \rangle$ is given by

$$\mathbf{P}_k^a = (\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)[(\mathbf{I} - \mathbf{K}_k\mathbf{H}_k)\mathbf{P}_k^f]^{\mathrm{T}} + \mathbf{K}_k\mathbf{R}_k^o\mathbf{K}_k^{\mathrm{T}}. \tag{A.13}$$

b. Forecast step

Let us consider a stochastic-dynamic equation of the form

$$\boldsymbol{\mu}_{k+1}^t = \mathbf{M}_k\boldsymbol{\mu}_k^t + \mathbf{f}_k(\boldsymbol{\mu}_k^a) \circ \boldsymbol{\epsilon}_k^q, \tag{A.14}$$

where $\langle \boldsymbol{\epsilon}_k^q(\boldsymbol{\epsilon}_k^q)^{\mathrm{T}} \rangle = \mathbf{C}_k^q$. Here again we note that, by definition, $\boldsymbol{\mu}_k^a \equiv \langle \boldsymbol{\mu}_k^t \mid \mathcal{M}_k^o = \mathbf{m}_k^o \rangle$ is not a random vector. Taking the conditional expectation of (A.14) thus gives

$$\begin{aligned}
\boldsymbol{\mu}_{k+1}^f &\equiv \langle \boldsymbol{\mu}_{k+1}^t \mid \mathcal{M}_k^o = \mathbf{m}_k^o \rangle \\
&= \mathbf{M}_k\langle \boldsymbol{\mu}_k^t \mid \mathcal{M}_k^o = \mathbf{m}_k^o \rangle \\
&\quad + \langle \mathbf{f}_k(\langle \boldsymbol{\mu}_k^t \mid \mathcal{M}_k^o = \mathbf{m}_k^o \rangle) \circ \boldsymbol{\epsilon}_k^q \mid \mathcal{M}_k^o = \mathbf{m}_k^o \rangle \\
&= \mathbf{M}_k\langle \boldsymbol{\mu}_k^t \mid \mathcal{M}_k^o = \mathbf{m}_k^o \rangle \\
&\quad + \mathbf{f}_k(\langle \boldsymbol{\mu}_k^t \mid \mathcal{M}_k^o = \mathbf{m}_k^o \rangle) \circ \langle \boldsymbol{\epsilon}_k^q \mid \mathcal{M}_k^o = \mathbf{m}_k^o \rangle \\
&= \mathbf{M}_k\langle \boldsymbol{\mu}_k^t \mid \mathcal{M}_k^o = \mathbf{m}_k^o \rangle + \mathbf{f}_k(\langle \boldsymbol{\mu}_k^t \mid \mathcal{M}_k^o = \mathbf{m}_k^o \rangle) \circ \langle \boldsymbol{\epsilon}_k^q \rangle,
\end{aligned} \tag{A.15}$$

where we have used the fact that $\boldsymbol{\epsilon}_k^q$ is independent of the observations. Assuming an unbiased dynamical model, that is, $\langle \boldsymbol{\epsilon}_k^q \rangle = 0$, yields

$$\boldsymbol{\mu}_{k+1}^f = \mathbf{M}_k\boldsymbol{\mu}_k^a. \tag{A.16}$$

To obtain the forecast error covariance $\mathbf{P}_{k+1}^f \equiv \langle (\boldsymbol{\mu}_{k+1}^t - \boldsymbol{\mu}_{k+1}^f)(\boldsymbol{\mu}_{k+1}^t - \boldsymbol{\mu}_{k+1}^f)^{\mathrm{T}} \mid \mathcal{M}_k^o = \mathbf{m}_k^o \rangle$, we need to consider the conditional expectation of the following four terms of the outer product:

$$\begin{aligned}
&(\boldsymbol{\mu}_{k+1}^t - \boldsymbol{\mu}_{k+1}^f)(\boldsymbol{\mu}_{k+1}^t - \boldsymbol{\mu}_{k+1}^f)^{\mathrm{T}} \\
&\quad = \mathbf{M}_k(\boldsymbol{\mu}_k^t - \boldsymbol{\mu}_k^a)(\boldsymbol{\mu}_k^t - \boldsymbol{\mu}_k^a)^{\mathrm{T}}\mathbf{M}_k^{\mathrm{T}} \\
&\qquad + \mathbf{M}_k(\boldsymbol{\mu}_k^t - \boldsymbol{\mu}_k^a)[\mathbf{f}_k(\boldsymbol{\mu}_k^a) \circ \boldsymbol{\epsilon}_k^q]^{\mathrm{T}} \\
&\qquad + [\mathbf{f}_k(\boldsymbol{\mu}_k^a) \circ \boldsymbol{\epsilon}_k^q](\boldsymbol{\mu}_k^t - \boldsymbol{\mu}_k^a)^{\mathrm{T}}\mathbf{M}_k^{\mathrm{T}} \\
&\qquad + [\mathbf{f}_k(\boldsymbol{\mu}_k^a) \circ \boldsymbol{\epsilon}_k^q][\mathbf{f}_k(\boldsymbol{\mu}_k^a) \circ \boldsymbol{\epsilon}_k^q]^{\mathrm{T}}.
\end{aligned} \tag{A.17}$$

Taking the conditional expectation $\langle \quad \mid \mathcal{M}_k^o = \mathbf{m}_k^o \rangle$, we get $\mathbf{M}_k\mathbf{P}_k^a\mathbf{M}_k^{\mathrm{T}} = \mathbf{M}_k(\mathbf{M}_k\mathbf{P}_k^a)^{\mathrm{T}}$ for the first term on the rhs of (A.17). Similarly, the evaluation of the last term on the rhs of (A.17) is obtained as follows:

$$\begin{aligned}
&\langle (\mathbf{f}_k(\boldsymbol{\mu}_k^a) \circ \boldsymbol{\epsilon}_k^q)(\mathbf{f}_k(\boldsymbol{\mu}_k^a) \circ \boldsymbol{\epsilon}_k^q)^{\mathrm{T}} \mid \mathcal{M}_k^o = \mathbf{m}_k^o \rangle \\
&\quad = \mathbf{f}_k(\boldsymbol{\mu}_k^a)\mathbf{f}_k^{\mathrm{T}}(\boldsymbol{\mu}_k^a) \circ \langle \boldsymbol{\epsilon}_k^q(\boldsymbol{\epsilon}_k^q)^{\mathrm{T}} \mid \mathcal{M}_k^o = \mathbf{m}_k^o \rangle \\
&\quad = \mathbf{f}_k(\boldsymbol{\mu}_k^a)\mathbf{f}_k^{\mathrm{T}}(\boldsymbol{\mu}_k^a) \circ \mathbf{C}_k^q,
\end{aligned} \tag{A.18}$$

where we used the fact that $\boldsymbol{\mu}_k^a \equiv \langle \boldsymbol{\mu}_k^t \mid \mathcal{M}_k^o = \mathbf{m}_k^o \rangle$ and that $\boldsymbol{\epsilon}_k^q$ is independent of the observations. Finally, the conditional expectation of the two cross terms in (A.17) vanishes as in the standard Kalman filter derivation (Jazwinski 1970, section 3.9), as a result of assuming that $\boldsymbol{\epsilon}_k^q$ is white with zero mean, independent of the observations, and independent of the initial error. Thus

$$\mathbf{P}_{k+1}^f = \mathbf{M}_k(\mathbf{M}_k\mathbf{P}_k^a)^{\mathrm{T}} + \mathbf{Q}_k, \tag{A.19}$$

where

$$\mathbf{Q}_k \equiv \mathbf{f}_k(\boldsymbol{\mu}_k^a)\mathbf{f}_k^{\mathrm{T}}(\boldsymbol{\mu}_k^a) \circ \mathbf{C}_k^q.$$

The density $p(\boldsymbol{\mu}_{k+1}^t \mid \mathcal{M}_k^o = \mathbf{m}_k^o)$ is Gaussian since $\boldsymbol{\mu}_{k+1}^t$ is the sum of two Gaussian densities, $\mathbf{M}_k\boldsymbol{\mu}_k^t$ and $\mathbf{f}_k(\boldsymbol{\mu}_k^a) \circ \boldsymbol{\epsilon}_k^q$.

REFERENCES

Bacmeister, J. T., S. D. Eckermann, L. Sparling, K. R. Chan, M. Loewenstein, and M. H. Proffitt, 1997: Analysis of intermittency in aircraft measurements of velocity, temperature and atmospheric tracers using wavelet transforms. *Gravity Wave Processes and their Parametrization in Global Climate Models*, K. P. Hamilton, Ed., NATO ASI Ser., Vol. 1, Springer-Verlag, 85–102.

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

Bloom, S. C., L. L Takacs, A. M. da Silva, and D. Ledvina, 1996: Data assimilation using incremental analysis updates. *Mon. Wea. Rev.*, **124,** 1256–1271.

Bryson, A. E., and Y.-C. Ho, 1975: *Applied Optimal Control*. Hemisphere Publishing, 479 pp.

Cohn, S. E., 1993: Dynamics of short-term univariate forecast error covariances. *Mon. Wea. Rev.*, **121,** 3123–3148.

——, 1997: An introduction to estimation theory. *J. Meteor. Soc. Japan* (Special Issue), **75,** 257–288.
——, and D. P. Dee, 1988: Observability of discretized partial differential equations. *SIAM. J. Numer. Anal.,* **25,** 586–617.
——, and R. Todling, 1996: Approximate data assimilation schemes for stable and unstable dynamics. *J. Meteor. Soc. Japan,* **74,** 63–75.
——, A. da Silva, J. Guo, M. Sienkiewicz, and D. Lamich, 1998: Assessing the effects of data selection with the DAO Physical-space Statistical Analysis System. *Mon. Wea. Rev.,* **126,** 2913–2926.
Courtier, P., J.-N. Thépaut, and A. Hollingsworth, 1994: A strategy for operational implementation of 4D-Var. using an incremental approach. *Quart. J. Roy. Meteor. Soc.,* **120,** 1367–1388.
——, and Coauthors, 1998: The ECMWF implementation of the three dimensional variational assimilation (3D-Var). Part I: Formulation. *Quart. J. Roy. Meteor. Soc.,* **124,** 1783–1807.
Daley, R., 1991: *Atmospheric Data Analysis.* Cambridge University Press, 457 pp.
——, 1992: The lagged innovation covariance: A performance diagnostics for atmospheric data assimilation. *Mon. Wea. Rev.,* **120,** 178–196.
——, 1993: Estimating observation error statistics for atmospheric data assimilation. *Ann. Geophys.,* **11,** 634–647.
——, 1997: Error propagation and observability for the constituent transport equation in steady, non-divergent, two-dimensional flow. *Numerical Methods in Atmospheric and Oceanic Modeling. The André J. Robert Memorial Volume,* C. A. Lin, R. Laprise, and H. Ritchie, Eds., NRC Research Press and Canadian Meteorological and Oceanographic Society, 323–351.
——, and R. Ménard, 1993: Spectral characteristics of Kalman filter systems for atmospheric data assimilation. *Mon. Wea. Rev.,* **121,** 1554–1565.
Dee, D. P., 1995a: Testing the perfect-model assumption in variational data assimilation. *Proc. Second Int. Symp. on Assimilation of Observations in Meteorology and Oceanography,* Tokyo, Japan, World Meteorological Organization, 225–228.
——, 1995b: On-line estimation of error covariance parameters for atmospheric data assimilation. *Mon. Wea. Rev.,* **123,** 1128–1145.
——, and A. da Silva, 1999: Maximum-likelihood estimation of forecast and observation error covariance paramteres. Part I: Methodology. *Mon. Wea. Rev.,* **127,** 1822–1834.
——, G. Gaspari, C. Redder, L. Rukhovets, and A. da Silva, 1999: Maximum-likelihood estimation of forecast and observation error covariance paramteres. Part II: Applications. *Mon. Wea. Rev.,* **127,** 1835–1849.
Gaspari, G., and S. E. Cohn, 1999: Construction of correlation functions in two and three dimensions. *Quart. J. Roy. Meteor. Soc.,* **125,** 723–757.
Houtekamer, P. L., J. Derome, H. Ritchie, and H. L. Mitchell, 1996: A system simulation approach to ensemble prediction. *Mon. Wea. Rev.,* **124,** 1225–1242.
Jazwinski, A. H., 1970: *Stochastic Processes and Filtering Theory.* Academic Press, 376 pp.
Joiner, J., and A. M. da Silva, 1998: Efficient methods to assimilate remotely-sensed data based on information content. *Quart. J. Roy. Meteor. Soc.,* **124,** 1669–1694.
Lin, S.-J., and R. B. Rood, 1996: Multidimensional flux-form semi-Lagrangian transport schemes. *Mon. Wea. Rev.,* **124,** 2046–2070.
Lyster, P. M., S. E. Cohn, R. Ménard, L.-P. Chang, S.-J. Lin, and R. G. Olsen, 1997: Parallel implementation of a Kalman filter for constituent data assimilation. *Mon. Wea. Rev.,* **125,** 1674–1686.
Marks, C. J., and C. D. Rodgers, 1993: A retrieval method for atmospheric composition from limb measurements. *J. Geophys. Res.,* **98,** 14 939–14 953.
Ménard, R., 1994: Kalman filtering of Burger's equation and its application to atmospheric data assimilation. Ph.D. thesis, Stormy Weather Group Scientific Report MW-100, McGill University, Montreal, PQ, 211 pp. [Available online at ftp://hera.gsfc.nasa.gov/pub/papers/menard/thesis.]
——, and R. Daley, 1996: The application of Kalman smoother theory to the estimation of 4DVAR error statistics. *Tellus,* **48,** 221–237.
——, and L.-P. Chang, 2000: Stratospheric assimilation of chemical tracer observations using a Kalman filter. Part II: χ^2-validated results and analysis of variance and correlation dynamics. *Mon. Wea. Rev.,* **128,** 2672–2686.
——, P. M. Lyster, L.-P. Chang, and S. E. Cohn, 1995: Middle-atmosphere assimilation of UARS constituent data using Kalman filtering: Preliminary results. *Proc. Second Int. Symp. on Assimilation of Observations in Meteorology and Oceanography,* Tokyo, Japan, World Meteorological Organization, 235–238.
Mitchell, H. L., and R. Daley, 1997a: Discretization error and signal/error correlation in atmospheric data assimilation. Part I: All scales resolved. *Tellus,* **49A,** 32–53.
——, and ——, 1997b: Discretization error and signal/error correlation in atmospheric data assimilation. Part II: The effect of unresolved scales. *Tellus,* **49A,** 54–73.
Nastrom, G. D., W. H. Jasperson, and K. S. Gage, 1986: Horizontal spectra of atmospheric tracers measured during the global atmospheric sampling program. *J. Geophys. Res.,* **91,** 13 201–13 209.
Ngan, K., and T. G. Shepherd, 1997: Comments on some recent measurements of anomalous steep N_2O and O_3 tracer spectra in the stratospheric surf zone. *J. Geophys. Res.,* **102,** 24 001–24 004.
Park, J. H., and Coauthors, 1996: Validation of Halogen Occultation Experiment CH_4 measurements from UARS. *J. Geophys. Res.,* **101,** 10 183–10 203.
Pfaendtner, J., S. Bloom, D. Lamich, M. Seablom, M. Sienkiewicz, J. Stobie, and A. da Silva, 1995: Documentation of the Goddard Earth Observing System (GEOS) Data Assimilation System—Version 1. NASA Tech. Memo. 104606, Vol. 4. [Available from Goddard Space Flight Center, Greenbelt, MD 20771.]
Polavarapu, S., M. Tanguay, R. Ménard, and A. Staniforth, 1996: The tangent linear model for semi-Lagrangian schemes: Linearizing the process of interpolation. *Tellus,* **48,** 74–95.
Reber, C. A., 1993: The Upper Atmosphere Research Satellite (UARS). *Geophys. Res. Lett.,* **20,** 1215–1218.
Riishøjgaard, L. P., and E. Källén, 1997: On the correlation between ozone and potential vorticity for large-scale Rossby waves. *J. Geophys. Res.,* **102,** 8793–8804.
Roche, A. E., J. B. Kumer, J. L. Mergenthaler, G. A. Ely, W. G. Uplinger, J. F. Potter, T. C. James, and L. W. Sterritt, 1993: The Cryogenic Limb Array Etalon Spectrometer (CLAES) on UARS: Experiment description and performance. *J. Geophys. Res.,* **98,** 10 763–10 775.
——, and Coauthors, 1996: Validation of CH4 and N2O measurements by the Cryogenic Limb Array Etalon Spectrometer instrument on the Upper Atmosphere Research Satellite. *J. Geophys. Res.,* **101,** 9679–9710.
Rodgers, C. D., 1990: Characterization and error analysis of profiles retrieved from remote sounding measurements. *J. Geophys. Res.,* **95** (D5), 5587–5595.
Russell, J. M., III, and Coauthors, 1993: The Halogen Occultation Experiment. *J. Geophys. Res.,* **98,** 10 777–10 797.
Schoeberl, M. R., and L. R. Lait, 1992: Conservative-coordinate transformations for atmospheric measurements. *The Use of EOS for Studies of Atmospheric Physics, Proceedings of the International School of Physics Enrico Fermi,* J. C. Gille and G. Visconti, Eds., North-Holland, 419–431.
Shapiro, R., 1970: Smoothing, filtering, and boundary effects. *Rev. Geosphys. Space Phys.,* **8,** 359–387.

Sorenson, H. W., and J. E. Sacks, 1971: Recursive fading memory filtering. *Inform. Sci.,* **3,** 101–119.
Strang, G., 1968: On the construction and comparison of difference schemes. *SIAM J. Numer. Anal.,* **5,** 506–517.
Tarantola, A., 1987: *Inverse Problem Theory. Methods for Data Fitting and Model Parameter Estimation.* Elsevier Science, 613 pp.
Todling, R., and S. E. Cohn, 1994: Suboptimal schemes for atmospheric data assimilation based on the Kalman filter. *Mon. Wea. Rev.,* **122,** 2530–2557.
Waugh, D. W., and Coauthors, 1997: Three-dimensional simulations of long-lived tracers using winds from MACCM2. *J. Geophys. Res.,* **102** (D17), 21 493–21 513.
Xu, Q., 1996: Generalized adjoint for physical processes with parametrized discontinuities. Part I: Basic issues and heuristic examples. *J. Atmos. Sci.,* **53,** 1123–1142.