JOURNAL OF GEOPHYSICAL RESEARCH, VOL. 104, NO. D15, PAGES 18,715–18,737, AUGUST 20, 1999

Assimilation of photochemically active species and a case analysis of UARS data

Boris V. Khattatov,[1] John C. Gille,[1] Lawrence V. Lyjak,[1] Guy P. Brasseur,[1]
Victor L. Dvortsov,[2,3] Aidan E. Roche,[4] and Joe W. Waters[5]

Abstract. We present a short overview of applications of estimation theory in atmospheric chemistry and discuss some common methods of gridding and mapping of irregular satellite observations of chemical constituents. It is shown that these methods are unable to produce truly synoptic maps of short-lived photochemically active species due to insufficient temporal and spatial density of satellite observations. The only way to overcome this limitation is to supplement observations with prior independent information given, for instance, by atmospheric numerical models and/or climatologies. Objective approaches to combining such prior information with observations are commonly referred to as data assimilation. Mathematical basis of data assimilation known as optimal estimation equations is presented following *Lorenc* [1986]. Two particular techniques of data assimilation, the variational method and the extended Kalman filter, are briefly described, and their applications to time-dependent numerical photochemical models are discussed. We investigate validity of the linear approximation which is utilized in both methods, present time evolution of the linearization and covariance matrices, and discuss some of their properties. On the basis of ideas of *Fisher and Lary* [1995] we then employ a trajectory model and a photochemical box model for assimilation and mapping of the Upper Atmosphere Research Satellite (UARS) measurements of chemical species. The assimilation is performed using the variational technique and the extended Kalman filter, and results of both methods are presented and discussed.

1. Introduction

In order to characterize and to predict the evolution of the Earth's atmosphere one has to know its present state. The state of the atmosphere is defined by instantaneous global distributions of a number of physical parameters such as winds, temperature, concentrations of chemical constituents, moisture, radiation field, etc. One cannot envision being able to observe all the components of the atmospheric state simultaneously and at every point in space and with a perfect accuracy. It is inevitable then that to characterize the atmosphere some additional information has to be brought into consideration.

Such information can be introduced in two ways. It can be given by theoretical equations that allow one to deduce some components of the state from others or to derive them from a smaller set of model parameters. In this case the number of independent components of the state is reduced and the effective density of observations is increased. Also, the additional information can be given by independent, a priori estimates of the atmospheric parameters with their corresponding uncertainties. In the latter case one can think of the a priori estimates and their corresponding uncertainties as some additional "virtual measurements", as *Rodgers* [1976] suggested. Numerical models or climatologies can be used to obtain such virtual measurements.

Objective approaches to combining a priori knowledge about a physical system under consideration with available (usually sparse and irregular) observations are often referred to as data assimilation. The mathematical basis of data assimilation is the estimation or inverse problem theory. Inverse problem theory is an organized set of mathematical techniques for obtaining useful information about the physical world on the basis of observations [*Menke*, 1984]. In a conventional "forward" problem one uses a set of a priori parameters to predict the state of the physical system. In the "inverse" or estimation problem one attempts to use available observations of the state of the system to estimate poorly known model parameters and/or the state itself.

The ideas and methods of estimation and inverse problem theory can be applied to observations and modeling of almost anything. The meanings of the model, state, model parameters, and observations depend on a particular application. In atmospheric sciences, mathematical methods of estimation and inverse theory have long been used in numerical weather prediction, data retrievals from remote sensing experiments (particularly satellite measurements), and inverse modeling of surface fluxes of methane and carbon dioxide. A few recent publications briefly described below demonstrated that the same techniques can be very successfully applied for studies of atmospheric chemistry and trace gases distributions.

[1]National Center for Atmospheric Research, Boulder, Colorado.
[2]NOAA Aeronomy Laboratory, Boulder, Colorado.
[3]Also at Cooperative Institute for Research in Environmental Sciences, University of Colorado, Boulder, Colorado.
[4]Lockheed Research Laboratory, Palo Alto, California.
[5]Jet Propulsion Laboratory, Pasadena, California.

Paper number 1999JD900225.
0148-0227/99/1999JD900225$09.00

Austin [1992] made perhaps the first attempt to implement sequential assimilation of satellite measurements into a global chemistry-transport model on isentropic surfaces. Limb infrared monitor of the stratosphere (LIMS) data and model simulations were weighted to form initial conditions for the model for the next integration period. The weights were empirically determined from the relative confidence in the model results and observations.

Riishojgaard [1996] has used the four-dimensional variational technique to assimilate ozone data into a numerical weather prediction model. Since the photochemical lifetime of ozone is relatively long in the lower and middle atmosphere, some dynamical information can be extracted from its observations and used to improve the weather forecasts.

Levelt et al. [1996] have used a variation of the objective interpolation technique to assimilate the TIROS operational vertical sounder (TOVS) total ozone amount measurements using a two- dimensional, latitude-longitude, global advection model.

Eskes et al. [1999] assimilated the Global Ozone Monitoring Experiment (GOME) total column measurements in a two-dimensional advection model using a variational approach and provided estimates of the errors of the final analysis. Both *Levelt et al.* [1996] and *Eskes et al.* [1999] provide means to construct nearly synoptic maps from irregular and sparse (in time and space) satellite measurements.

Lyster et al. [1997] for the first time have used a two-dimensional transport model on isentropic surfaces and the Kalman filter technique to assimilate cryogenic limb array etalon spectrometer (CLAES)- and Halogen Occultation Experiment (HALOE)-measured CH_4. Although very computationally expensive in terms of both CPU and memory resources, their pioneering approach allows production of synoptic maps from irregularly distributed UARS measurements. No photochemical processes were taken into account, though, and thus this approach in its current form is valid only when applied to inert tracers. This approach should not be confused with the widely used so-called "Kalman filter mapping" [e.g., *Haggard et al.*, 1986; *Kohri*, 1981], which is often applied for gridding satellite-collected data such as the UARS 3B data (http://grid.gsfc.nasa.gov:8001/www/ l3b_description.html). *Lyster et al.* [1997] use the actual advection equations (i.e., a global transport model) to predict the evolution of the analysis covariance matrix.

Fisher and Lary [1995] have used variational data assimilation for assimilating and mapping the CLAES observations of O_3, NO_2, and HNO_3 using a fairly simple six-species, 19-reaction photochemical box model in conjunction with a trajectory model. This was the first application of data assimilation techniques for analysis of photochemically active species in the stratosphere.

Elbern et al. [1997] extended the variational method, suggested by *Fisher and Lary* [1995], to assimilation of various tropospheric trace gases using a box model which included a number of organic species. They have shown that even a very small number of observations can significantly influence and improve the model simulations.

Levelt et al. [1998] have assimilated UARS Microwave Limb Sounder (MLS) measurements of ozone in the global three-dimensional (3-D) chemistry-transport model using the 3D-Var approach. This work demonstrated that sequential assimilation using global photochemical models can be successfully used for mapping irregular satellite observations and filling in temporal and spatial gaps.

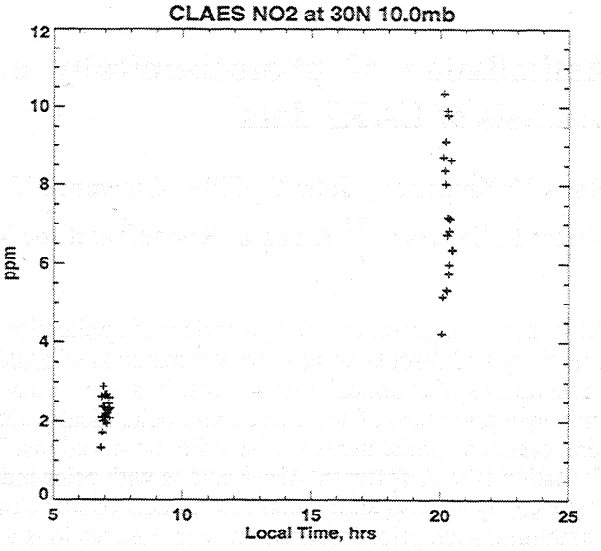

Figure 1. UARS cryogenic limb array etalon spectrometer (CLAES) NO_2 measurements for 1 day at 30° N at 10 mbar.

Interestingly, most of the referenced studies have concentrated on assimilation and analysis of satellite data. The global coverage of satellite measurements makes such measurements very attractive for producing global distributions of chemical species. However, the geometry of satellite orbits imposes strict limitations on the space and time sampling. This makes analysis of photochemically active species with a strong diurnal cycle particularly complicated. Consider, for example, daily UARS CLAES measurements of NO_2 at 10 mbar shown in Figure 1. At this altitude, NO_2 undergoes strong diurnal variations. One can see that all daily CLAES observations are grouped into two narrow clusters separated by about 12 hours. Obviously, the local time coverage of the CLAES data is insufficient to resolve the NO_2 diurnal cycle. The data assimilation methodology makes use of the additional constraints given by a priori knowledge and thus makes it possible to overcome some of the limitations of asynoptic satellite observations.

In this paper we describe two mathematical techniques, the variational data assimilation and the extended Kalman filter, and apply them to assimilation and mapping of the UARS-measured chemical species. The employed mathematical apparatus is not innovative by any means. Mathematical derivations presented here can be found in a number of publications [e.g., *Lorenc*, 1986]. However, applications of this methodology to photochemical models are fairly rare. The current work has been largely motivated by ideas introduced for the first time by *Fisher and Lary* [1995].

An overview of the existing conventional methods of satellite data mapping is given in the next section. Mathematical foundations of the employed data assimilation methodology are presented in section 3, followed by description of a photochemical box model used in the assimilation and description of the linear approximation to the model. In the remaining sections we present evolution of the linearization and error covariance matrices arising in both

methods and describe assimilation of the actual UARS observations.

2. Conventional Mapping Techniques

Atmospheric constituents exhibit variability on very different temporal and spatial scales. Insufficient temporal and spatial density of satellite observations generally does not allow analysis of constituent variations on all scales simultaneously. Well-established methods exist for mapping and analyzing the long-lived species. Some of the most often used methods include Fourier synoptic mapping [*Salby*, 1982a, b], trajectory mapping [*Morris et al.*, 1995; *Sutton et al.*, 1994], and the sequential estimation algorithm [*Rodgers*, 1976]. However, none of these techniques is directly suitable for analysis of satellite observations of diurnally varying species.

The synoptic mapping technique [*Salby*, 1982a, b] is a Fourier transform scheme modified for retrieving synoptic maps from asynoptic satellite-collected data. This technique works well for data that show relatively slow temporal variability and large-scale spatial features. This method has a major disadvantage: It requires continuous measurements along the satellite orbit. Hence this technique cannot be used for instruments that measure during only a portion of local time span. Although it is possible to adopt this technique to compute amplitudes of the diurnal variations of some species [*Sassi and Salby*, 1998] the full diurnal cycle cannot be resolved.

The trajectory mapping technique [*Morris et al.*, 1995; *Sutton et al.*, 1994] overcomes some of the limitations imposed by the satellite observational patterns through use of the observed winds and temperatures to "map" irregularly spaced observations made in the past onto regular grid points in the present. It is in some respect superior to the synoptic mapping since it introduces additional constraints through use of the observed dynamics. However, this method completely ignores chemical interactions between atmospheric constituents and is thus valid only on the timescales shorter than the photochemical lifetime of the analyzed constituents. It is therefore inappropriate for analysis of short-lived species. The data assimilation methodology presented here can be thought of as a combination of the trajectory mapping of *Morris et al.* [1995] and a photochemical box model.

The sequential estimation technique has been used in UARS data processing for producing the so-called level 3B data, which consist of sets of Fourier coefficients in longitude for every latitude and altitude level and some auxiliary information.

Level 3B UARS data were included in a number of scientific publications and are being used by the community. However, this technique also has serious limitations imposed by the asynoptic nature of satellite observations. Consider, for instance, Plate 1, showing a map of NO_2 distribution at approximately 10 mbar for May 3, 1993, produced from CLAES level 3B data for 1200 UT. At these altitudes, NO_2 has a strong diurnal cycle. If the map shown in Plate 1 were approximately synoptic, one should have been able to see increased concentrations of NO_2 at the nighttime longitudes and smaller concentrations at daytime longitudes. Clearly, the terminator is not present in this plot. There is very little longitudinal dependence in the distribution suggesting that all points at a given latitude correspond to approximately the same local time. Thus the sequential estimation technique applied to observations of diurnally varying chemical constituents yields results in which short temporal variability is aliased into what is interpreted as longer temporal scales, including the time mean component. Thus the technique is unable to produce a synoptic analysis of these constituents. Its results might be hard to interpret at best and misleading at worst. The extent to which diurnal variations contribute to the "zonal mean" and other Fourier harmonics is determined

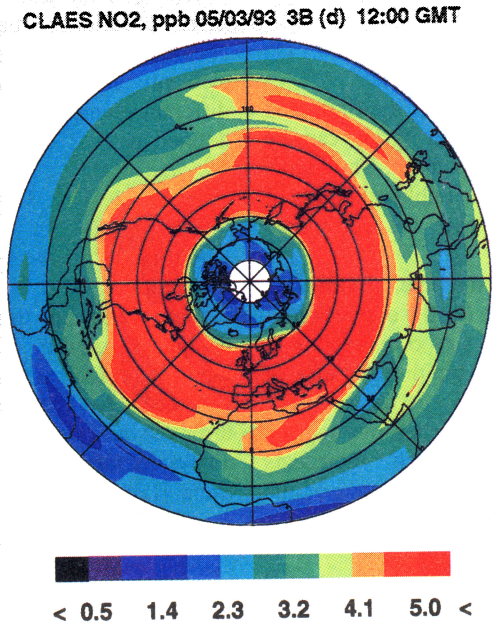

Plate 1. Map of UARS CLAES NO_2 (ppbv) produced from level 3B data for 1 day.

largely by the amplitude of these variations and is expected to be very significant for short-lived species such as NO_2 or ClO in the stratosphere and ozone in the upper stratosphere.

One can attempt to separate the local time variation by accumulating a number of days of observations at each latitude, with each day corresponding to a slightly different local time. For instance, in the case of the UARS the satellite orbit precesses by about 20 min of local time per day, and therefore about 36 days of observations for both ascending and descending portions of the orbit are needed to cover all local times. However, if the analysis is performed in this manner, the day-to-day geophysical variability cannot be distinguished from the local time variability. Therefore the day-to-day variations will be aliased into the local time variability and vice versa. These difficulties are inherent in the asynoptic satellite observations. In order to improve the situation some additional information needs to be taken into consideration.

From the above discussion one can conclude that the described techniques cannot produce a satisfactory analysis of the satellite-measured time-space distributions of species that exhibit temporal variations with timescales of about a day or shorter. The data assimilation methodology provides means to analyze various satellite observations of short-lived, as well as long-lived, constituents by supplementing measurements with our knowledge of atmospheric photochemistry and transport.

3. Mathematical Basis

Since applications of estimation and inverse problem theories in atmospheric chemistry are fairly rare, a short introduction to the employed mathematical framework is offered. Key concepts of the data assimilation methodology can be presented in a fairly concise mathematical form [e.g., *Lorenc*, 1986].

Consider a model of a physical system represented by operator (generally nonlinear) $\mathbf{M}$, and let vector $\mathbf{x}$ with dimension N_x be a set of input parameters for the model. These input parameters are used to predict the state of the system, vector $\mathbf{y}$ with dimension N_y:

$$\mathbf{y} = \mathbf{M}(\mathbf{x}) \tag{1}$$

We assume that dimensions of vectors $\mathbf{y}$ and $\mathbf{x}$ are finite and that therefore an appropriate discretization (grid) can be introduced in the real continuous physical space. The model state $\mathbf{y}$ is defined by values of the model variables at points of the grid. The discretization errors will be considered negligible in this discussion.

Let vector $\mathbf{y}_o$ contain observations of the state. Usually, the dimension of $\mathbf{y}_o$ is significantly less than N_y, the dimension of the model space. Moreover, the locations of observations in the real physical space can be different from the locations of the grid points at which the model space $\mathbf{y}$ is defined. The connection between $\mathbf{y}_o$ and $\mathbf{y}$ can be established through the so-called observational operator $\mathbf{H}$, which represents interpolation of the state variables from the locations of the grid points to the locations of the observations:

$$\mathbf{y}_o = \mathbf{H}(\mathbf{y}) \tag{2}$$

Everywhere in this discussion we assume that the interpolation errors associated with operator $\mathbf{H}$ are negligible. The results are easily extended to the case when this is not true [e.g., *Lorenc*, 1986]. Combining the above two equations, we get

$$\mathbf{y}_o = \mathbf{H}(\mathbf{M}(\mathbf{x})) \tag{3}$$

Our analysis problem is then to find the "best" value of $\mathbf{x}$, which inverts (3) for a given $\mathbf{y}_o$ allowing for observation errors and other prior information [*Lorenc*, 1986]. In most cases, dimensions of vectors $\mathbf{x}$ and $\mathbf{y}_o$ will be different, and this problem will be either overdetermined or underdetermined. Therefore inversion of (3) should be done in the statistical sense.

"Best" here means that the errors of the final analysis are minimal. An exact value of a physical quantity can rarely be determined. One can only say that this value lies within a certain range with a certain probability, and therefore all estimates of the best value of $\mathbf{x}$ or $\mathbf{x}_o$ obtained from the observed $\mathbf{y}_o$ are probabilistic in nature. A mathematically robust definition of best or optimal $\mathbf{x}$ is, for instance, the value corresponding to the maximum of the probability density function (PDF) of $\mathbf{x}$ given observations $\mathbf{y}_o$. This is the so-called maximum likelihood definition.

The exact shapes of the PDFs in both $\mathbf{x}$ and $\mathbf{y}$ spaces are generally unknown. In order to solve the posed analysis problem one needs to establish a relationship between the PDF of $\mathbf{x}$ and the PDF of $\mathbf{y}$. Formal transformation of PDFs by the model from the parameter space $\mathbf{x}$ to the model space y is described by the so-called Fokker-Kolmogorov equation [e.g., *Jazwinski*, 1970], which is impossible to solve in most practical applications. This is one of the reasons why simplifications are needed in order to be able to solve practical problems.

One simplification is that the probability density functions can be approximated by Gaussian functions:

$$\mathrm{PDF}(\mathbf{x}) \sim \exp[-0.5(\mathbf{x}-\mathbf{x}_t)^T \mathbf{C}^{-1}(\mathbf{x}-\mathbf{x}_t)] \tag{4}$$

where $\mathbf{x}_t$ is the true value of $\mathbf{x}$ and $\mathbf{C}$ is the corresponding error covariance matrix. Its diagonal elements are the uncertainties (standard deviations) of $\mathbf{x}$, and the off-diagonal elements represent correlation between uncertainties of different elements of vector $\mathbf{x}$. The covariance matrix $\mathbf{C}$ is defined as

$$\mathbf{C} = \langle (\mathbf{x}-\mathbf{x}_t)(\mathbf{x}-\mathbf{x}_t)^T \rangle \tag{5}$$

where angle brackets represent averaging over all available realizations of $\mathbf{x}$.

Here we assume that all probability density functions could be approximated by Gaussian functions. We also assume that there exist a prior, independent estimate of $\mathbf{x}$, or $\mathbf{x}_b$, often called the background, and the corresponding background error covariances $\mathbf{B}$. The solution minimizing the final analysis errors is given by a minimum of the following functional [*Lorenc*, 1986]:

$$J(\mathbf{x}) = [\mathbf{y}_o - \mathbf{H}(\mathbf{M}(\mathbf{x}))]^T (\mathbf{O}+\mathbf{F})^{-1} [\mathbf{y}_o - \mathbf{H}(\mathbf{M}(\mathbf{x}))] + [\mathbf{x}-\mathbf{x}_b]^T \mathbf{B}^{-1} [\mathbf{x}-\mathbf{x}_b] \tag{6}$$

here $\mathbf{O}$ is the observational error covariance matrix, $\mathbf{F}$ is the error covariance corresponding to operators $\mathbf{M}$ and $\mathbf{H}$, and $\mathbf{B}$

is the background error covariance matrix. They characterize our confidence in the measurements, the model and observation operator, and the a priori background estimate. $J(\mathbf{x})$ is often called the misfit or cost function.

Thus the solution of the optimal analysis problem can be written as

$$\mathbf{x}_o : J(\mathbf{x}_o) = \min\{J(\mathbf{x})\} \tag{7}$$

Once the optimal estimate of $\mathbf{x}$ is obtained, (1) can be used to derive the best estimate of the state $\mathbf{y}$.

Note that both types of independent information mentioned earlier in this section are included in (6). The information given by theoretical equations is contained in operator $\mathbf{M}$ and the a priori estimate of model parameters is given by $\mathbf{x}_b$. In the ultimate case, when one measures the state directly at the locations of the grid points, $\mathbf{M}$ and $\mathbf{H}$ are the identity operators and $\mathbf{x}=\mathbf{y}$. In the case when no a priori first guess is available, which effectively means that the background error covariances become infinitely large, the second term in (6) vanishes. Equations (6) and (7) are simply a generalization of the well-known fact that independent measurements should be added with weights inversely proportional to the square of the measurement errors to form an optimal estimate.

In practical applications one has to find an appropriate way to compute the error covariances and to minimize (6). In most cases, in order to be able to do that we need to introduce the linear approximation. In the linear approximation we assume that for small perturbations of the parameter vector $\Delta\mathbf{x}$ the following is approximately true:

$$\mathbf{M}(\mathbf{x}+\Delta\mathbf{x}) = \mathbf{M}(\mathbf{x}) + \mathbf{L}\Delta\mathbf{x} \tag{8}$$

Note that in this expression $\mathbf{L}$ is a matrix, while $\mathbf{M}$ is, in general, a nonlinear operator. Formally, $\mathbf{L}$ is a derivative of $\mathbf{M}$ with respect to $\mathbf{x}$:

$$\mathbf{L} = \frac{d\mathbf{M}}{d\mathbf{x}} \tag{9}$$

The linearization $\mathbf{L}$ of the original model M will be used in two ways. First, minimization of $J(\mathbf{x})$ often requires knowledge of the derivative of $J(\mathbf{x})$ with respect to $\mathbf{x}$. This, in turn, requires knowing $d\mathbf{M}/d\mathbf{x}$. In this discussion we assume that observational operator $\mathbf{H}$ is simply a linear interpolation from times/locations of the model grid points to times/locations of the observations. Generalization to the case of nonlinear $\mathbf{H}$ is straightforward [e.g., *Lorenc*, 1986, p. 1180].

Second, for small variations of $\mathbf{x}$ one can show (from equations (5) and (8)) that transformation of error covariance matrix $\mathbf{C}_x$ in the parameter space to the error covariance matrix $\mathbf{C}_y$ in the model space is as follows:

$$\mathbf{C}_y = \mathbf{L}\,\mathbf{C}_x\,\mathbf{L}^T \tag{10}$$

This, in turn, allows one to establish correspondence between PDF of $\mathbf{x}$ in the parameter space and PDF of $\mathbf{y}$ in the model space.

We will now limit the discussion to the case when $\mathbf{M}$ is used to predict the state of the system at some future time from past state estimates. An example of such model is a photochemical box model described later in this paper. Formally, in this case

$$\mathbf{x} = \mathbf{x}_t, \mathbf{y} = \mathbf{x}_{t+\Delta t} \tag{11}$$

and

$$\mathbf{x}_{t+\Delta t} = \mathbf{M}(t, \mathbf{x}_t) \tag{12}$$

As an example, assume that vector $\mathbf{x}$ represents the state of a time-dependent numerical photochemical model, i.e., concentrations of modeled species at model grid points in the atmosphere. In the case of a box model that includes N_s species, the dimension of vector $\mathbf{x}$ would be N_s, while in a multidimensional model that contains N_p spatial grid points, the dimension of $\mathbf{x}$ would be $N_s \times N_p$. These numbers can be quite large: For instance, a 3-D model with 20 vertical levels and a $5^\circ \times 5^\circ$ horizontal resolution contains 54,020 grid points. If 20 chemical species are included, then the dimension of $\mathbf{x}$ is 1,080,400. Significant computational difficulties arise when dealing with vectors and matrices of such high dimensions [e.g., *Lyster et al.*, 1997].

In the next sections we will apply the above mentioned concepts to the case of a time-dependent predictive model. Specifically, we will present two techniques for solving the analysis problem: the variational method and the sequential method, also known as the Kalman filter.

3.1. Variational Technique

In the variational method a minimization algorithm is used to find model initial conditions $\mathbf{x}$ that minimize a misfit between model results and observations for the whole analysis period. The analysis period is usually much longer than the model time step Δt.

Usually, there is more than one observation available inside the analysis interval, and (6) becomes

$$J(\mathbf{x}) = \sum_i [\mathbf{y}^i_o - \mathbf{H}^i(\mathbf{M}^i(\mathbf{x}))]^T (\mathbf{O}^i + \mathbf{F}^i)^{-1} [\mathbf{y}^i_o - \mathbf{H}^i(\mathbf{M}^i(\mathbf{x}))] + [\mathbf{x}-\mathbf{x}_b]^T \mathbf{B}^{-1} [\mathbf{x}-\mathbf{x}_b] \tag{13}$$

Here index i corresponds to a particular time t_i, $\mathbf{y}^i_o$ represents an observation at this time, $\mathbf{H}^i(\mathbf{M}^i(\mathbf{x}))$ is the model estimate of the state interpolated, if necessary, to the time/location of $\mathbf{y}^i_o$, and $\mathbf{O}^i$ and $\mathbf{F}^i$ are the error covariance matrices corresponding to i.

The model, observation, and the background error covariances are considered to be known a priori and do not change during the assimilation procedure. In practical applications of the variational method the background term is often neglected. This is justified if the background state $\mathbf{x}_b$ is poorly known or if enough observations are available to properly constrain the system.

The variational data assimilation technique can be thought of as a constrained least squares fit to a set of observations distributed over some period of time. The constraints are given by the model equations. The choice of the analysis period is somewhat arbitrary and is dictated by the frequency of the observations and the characteristic timescales of the modeled system. The solution inside the analysis interval is smooth, while usually there is a discontinuity at the ends of adjacent analysis periods. Figure 2 illustrates the general idea of the variational method.

Since $\mathbf{M}$ is almost always nonlinear, the explicit solution is rarely possible, and a minimization algorithm has to be used

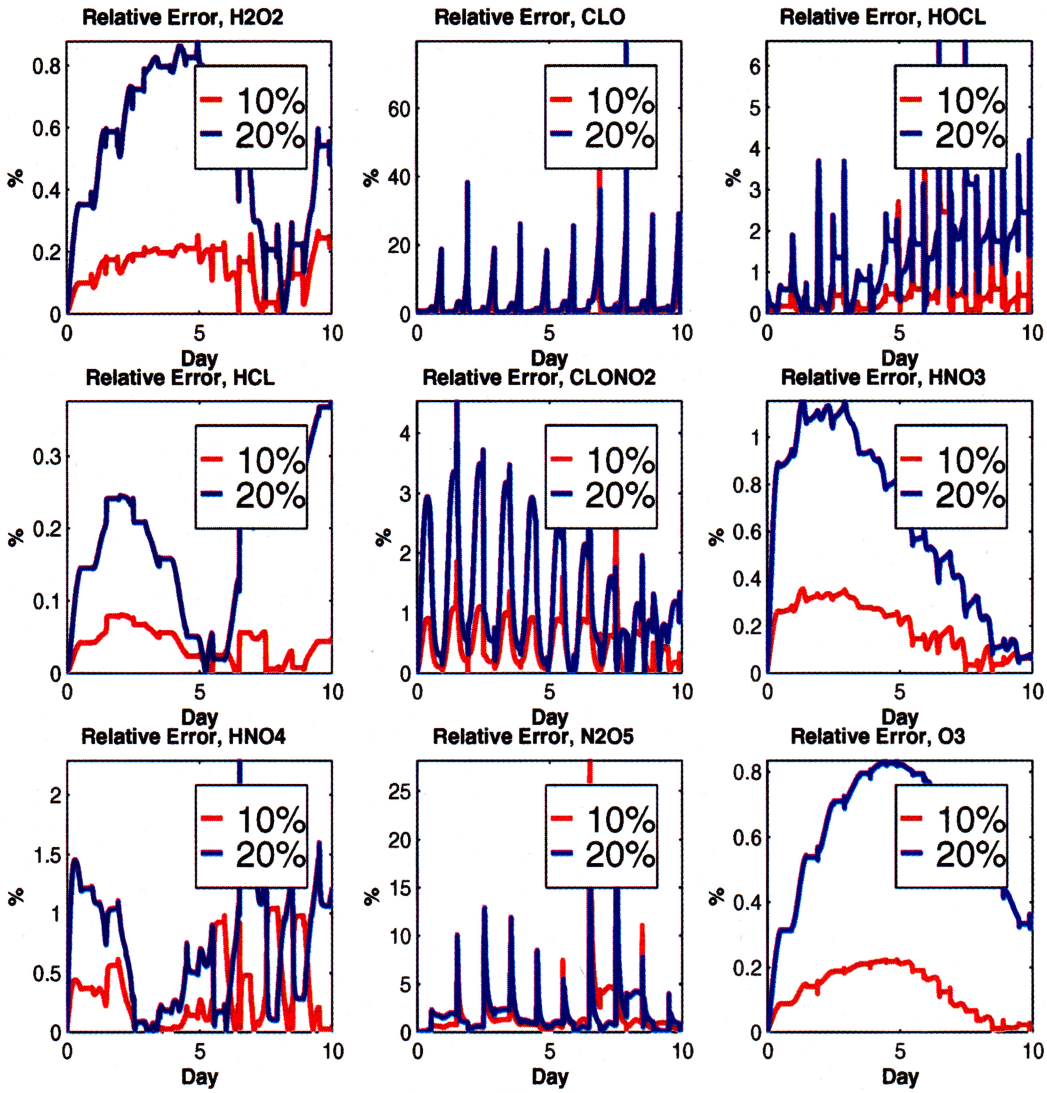

Plate 2. Error of the linear approximation as a function of time for a 10-day integration.

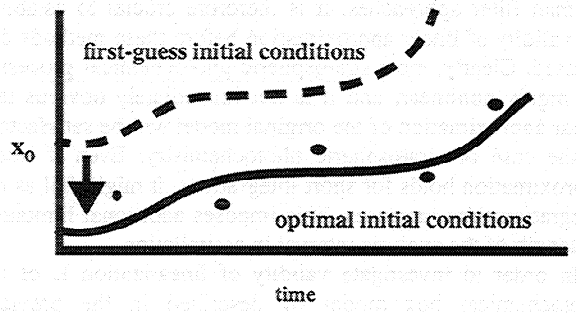

Figure 2. Illustration of the variational method. Initial conditions are modified in order to minimize the misfit between model simulations and observations.

for finding the optimal $\mathbf{x}$. Most minimization algorithms require knowledge of the gradient of the cost function $J(\mathbf{x})$ with respect to $\mathbf{x}$. The so-called adjoint method [e.g., *Talagrand and Courtier*, 1987; *Fisher and Lary*, 1995] is often used to compute $dJ(\mathbf{x})/d\mathbf{x}$. The adjoint method relies on the linearization of the model. This imposes additional requirements on the length of the analysis period, which should be short enough for the linear approximation to be valid.

Although it may be possible to compute the gradient $dJ(\mathbf{x})/d\mathbf{x}$ by analytically differentiating theoretical model equations, such approach might be unsatisfactory. In practice, the value of the cost function is obtained by numerically solving the discretized model equations rather than solving the theoretical model equations analytically. Thus the analytically obtained gradient may be inconsistent with the gradient of the numerically computed cost function, and the employed minimization algorithm might fail. To guarantee consistency of the gradient and the cost function, one should differentiate the numerical scheme that is used to integrate the model and compute $J(\mathbf{x})$. Modern techniques of automatic differentiation [e.g., *Griewank*, 1989] can be employed to facilitate this process. We should note, however, that in some applications an analytically computed gradient may suffice. An excellent discussion of the advantages and shortcomings of either approach is given by *Sirkes and Tziperman* [1997].

3.2. Sequential Technique (Kalman Filter)

In the sequential approach each observation is processed separately, and the analysis length is the time between two consecutive observations. Model forecast at the end of the analysis interval is considered to be the a priori background estimate $\mathbf{x}_b$. The optimal analysis equation is used to obtain the best estimate of $\mathbf{x}$ from the forecast value and a corresponding observation. The result is used as the model initial condition for the next analysis period [e.g., *Lorenc*, 1986; *Lyster et al.*, 1997]. Therefore $\mathbf{x}_t = \mathbf{x}_b$ is the model forecasted value of $\mathbf{x}$ at time t; $\mathbf{y}_o$ is the observation at time t; and $\mathbf{x}$ is the optimal estimate of model state at time t obtained from $\mathbf{x}_t$ and $\mathbf{y}_o$.

Equation (6) thus becomes

$$J(\mathbf{x}) = [\mathbf{y}_o - \mathbf{H}(\mathbf{x})]^T \mathbf{O}^{-1} [\mathbf{y}_o - \mathbf{H}(\mathbf{x})] + [\mathbf{x} - \mathbf{x}_t]^T \mathbf{B}_t^{-1} [\mathbf{x} - \mathbf{x}_t] \quad (14)$$

Here $\mathbf{B}_t$ is the forecast error covariance at time t. Note that $\mathbf{B}$ in (13) is also a forecast error covariance if the background is taken from a model forecast, as is often the case. Also, note that in (14) operator $\mathbf{M}$ is not explicitly included in the cost function. Since $\mathbf{M}$ is not included in the equation, an explicit solution is possible provided that $\mathbf{H}$ is linear. To minimize (14), one has to solve $dJ(\mathbf{x})/d\mathbf{x} = 0$. The solution is

$$\mathbf{x} = \mathbf{x}_t + \mathbf{K}(\mathbf{y}_o - \mathbf{H}\mathbf{x}_t) \quad (15)$$

$$\mathbf{K} = \mathbf{B}_t \mathbf{H}^T (\mathbf{H}\mathbf{B}_t \mathbf{H}^T + \mathbf{O})^{-1} \quad (16)$$

Matrix $\mathbf{K}$ is called the Kalman gain matrix.

At the end of each analysis period the model value ($\mathbf{x}_t$) and the corresponding observation ($\mathbf{y}_o$) are "mixed" with weights inversely proportional to their respective errors according to (15) and (16) to produce $\mathbf{x}$. Then the model is integrated forward in time starting from the obtained $\mathbf{x}$. Once an observation has been incorporated in the model, the analysis error covariance should be updated to reflect this. It can be shown that the new analysis error covariance is expressed as [*Lorenc*, 1986]

$$\mathbf{B} = \mathbf{B}_t - \mathbf{B}_t \mathbf{H}^T (\mathbf{H}\mathbf{B}_t \mathbf{H}^T + \mathbf{O})^{-1} \mathbf{H}\mathbf{B}_t \quad (17)$$

In the absence of observations, the model state is updated using (12), while evolution of the error covariance is obtained from the linearized model equations as in (10):

$$\mathbf{B}_{t+\Delta t} = \mathbf{L}\,\mathbf{B}_t\,\mathbf{L}^T \quad (18)$$

Equations (15)-(18) are often referred to as the extended Kalman filter. In the case of a linear model, $\mathbf{L}=\mathbf{M}$ and (15)-(18) will become the Kalman filter equations.

In reality, the forecast model is usually not perfect, and therefore equations (10), (12), and (18) should include additional terms due to the forecast model error. Everywhere in this work we assume that error is negligible relative to the observational error and is therefore omitted from the above equations.

Sequential analysis has a discontinuity whenever a new observation is encountered during forward model integration (Figure 3). If $\mathbf{x}$ is a function of both time and space, the four-dimensional analysis, as in the case of the variational technique when the analysis is done simultaneously in space and time, is replaced by a sequence of three-dimensional analyses performed at different times.

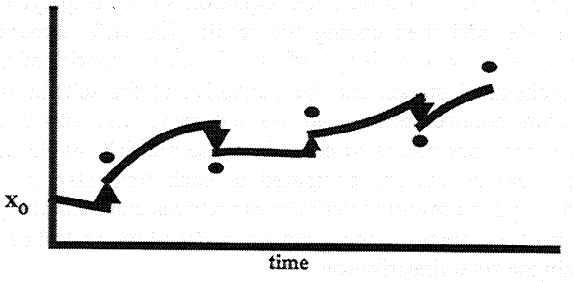

Figure 3. Illustration of the sequential method. A new estimate of the state is obtained each time an observation is encountered and the analysis contains discontinuities at these times. Then model integration continues starting from the new initial conditions.

4. Photochemical Model

A photochemical box model was specifically created for applications of the variational and extended Kalman filter methods to assimilation of satellite data in the stratosphere. The model uses adaptive-time-step Bulirsch-Stoer stiff solver with accuracy monitoring [*Press et al.*, 1996], similar to that described by *Fisher and Lary* [1995]. The photodissociation rates are computed by the delta-Eddington method implemented by *Dvortsov et al.* [1992]. The model automatically generates photochemical kinetic equations from a list of reactions and a list of species that the user wishes to include as well as the corresponding Jacobian and Hessian matrices. This is done by allocating special data structures in memory and filling them with pointers to arrays of species concentrations and photochemical reaction rates. When the right-hand side of the kinetic model equations or the Jacobian or Hessian need to be calculated in the model, algebraic operations are performed with components of these data structures according to rules inferred from a list of model species and a list of reactions. Thus adding new species and/or reactions to the model or altering the photochemical scheme is easy and does not require modification of the model code.

The photochemical scheme used in the data assimilation experiments includes over 50 chemical reactions, 16 photodissociation reactions, and the following species: H, OH, HO_2, H_2O_2, NO, NO_2, NO_3, N_2O_5, HNO_3, HNO_4, Cl, ClO, HOCl, HCl, $ClONO_2$, O, $O(^1D)$, O_3, CO, CH_4, N_2O, H_2, H_2O, and sulfate aerosol. Rates of the photochemical reactions were taken from *DeMore et al.* [1994]. Concentrations of CO, CH_4, N_2O, H_2, H_2O, and sulfate aerosol were held constant and typical for the altitude at which the calculations were performed.

For model validation we performed simulations with additional 16 species (OClO, Cl_2O_2, Cl_2, ClOO, CH_3, HCO, H_2CO, CH_3O, CH_3O_2, CH_3OOH, Br, BrO, HOBr, HBr, $BrONO_2$, and BrCl) and the corresponding extra photochemical reactions. Results of these simulations were compared with the Atmospheric Environmental Research (AER) box model [e.g., *Weisenstein et al.*, 1993] and an excellent agreement was demonstrated (M. Danilin, personal communication, 1998).

Linearization of the model, i.e., computer code, calculating derivatives of the final concentrations with respect to the initial concentrations, is required for implementation of both variational assimilation and the extended Kalman filter. This code was obtained by differentiating "by hand" the numerical scheme of the stiff differential equation solver employed in the model and then coding the result. The stiff numerical solver requires knowledge of the Jacobian matrix of the photochemical model, and the derivative of the scheme will therefore require the Hessian. As was mentioned above, the model provides means to compute these matrices as needed. The linearizations are computed at each time step in the model, and the resulting derivatives matrices corresponding to all the time steps in the integration are then multiplied to obtain the final linearization.

5. Linear Approximation

As was discussed above, linearization of a model used in the assimilation is central in both variational and extended Kalman filter approaches. It is therefore crucial to establish the validity of linear approximation before these methods can be used. Clearly, most stratospheric photochemical processes are highly nonlinear, and it is not immediately obvious that linear approximation of the original model will be satisfactory in the case of stratospheric photochemistry. Even if linear approximation holds for short integrations, it might fail as the integration time increases. This imposes additional limitation on length of the analysis interval in assimilation.

In order to investigate validity of linearization $\mathbf{L}$ of the photochemical box model $\mathbf{M}$ described in the previous section, we have performed full non-linear model integration starting from steady state initial conditions $\mathbf{x}_o$ and have linearized the model around the obtained model path. Formally,

$$\mathbf{x}(t) = \mathbf{M}(\mathbf{x}_o, t) \tag{19}$$

$$\mathbf{L}(t) = \frac{d\mathbf{x}(t)}{d\mathbf{x}_o} \tag{20}$$

Here vectors $\mathbf{x}$ and $\mathbf{x}_0$ contain concentrations of all variable species. Temperature and concentrations of nonvariable species were held constant with absolute values typical for 10 mbar. We then perturbed the initial species concentrations and performed calculations starting from the perturbed initial conditions $\mathbf{x}_0+\Delta\mathbf{x}$ using first the original nonlinear model:

$$\mathbf{x}_{NL}(t) = \mathbf{M}(\mathbf{x}_o + \Delta\mathbf{x}, t) \tag{21}$$

and then the linear approximation

$$\mathbf{x}_L(t) = \mathbf{M}(\mathbf{x}_o, t) + \mathbf{L}(t)\,\Delta\mathbf{x} \tag{22}$$

We then computed an error of linear approximation as a function of time as follows:

$$\varepsilon(t) = \frac{|\mathbf{x}_{NL}(t) - \mathbf{x}_L(t)|}{|\mathbf{x}_{NL}(t) + 10^{-20}|}$$

Value of 10^{-20} in the denominator was added to avoid division by zero when concentrations of some species disappear in diurnal cycle. The amplitude of perturbation $\Delta\mathbf{x}$ was set to 10% and 20% of the initial conditions $\mathbf{x}_o$. The shape of the perturbation was chosen at random, and several different perturbations were tried. Plots of $\varepsilon(t)$ are shown in Plate 2 for up to 10 days for some species. The relative error can be as high as 35% for brief periods of time for species experiencing strong diurnal cycles. Note that for some species (O_3 and HNO_3) the error is actually smaller at the end of a 10-day integration than in the middle. This can be perhaps explained by noting that the initial perturbation $\Delta\mathbf{x}$ pushes the chemical system away from its equilibrium state determined by initial conditions $\mathbf{x}_o$. During some time after initial perturbation the system experiences a nonlinear transition to a new equilibrium state. It appears that during this transition linear approximation does not hold as well as after the transition, when the system is in the vicinity of a new equilibrium state. In general, as appears from results shown in Plate 2, linear approximation holds surprisingly well for a 10-day integration. This behavior is probably due to the fact that concentrations of short-lived species undergoing nonlinear transformations are typically determined by concentrations of a few long-lived species such as O_3 or HCl and such parameters as total reactive chlorine and nitrogen. As long as

concentrations of these long-lived species do not vary significantly during analysis, we expect linear approximation to hold fairly well. During brief periods of time when short-lived species are not in photochemical equilibrium, at sunrise and sunset, the relative error increases rapidly, as seen in Plate 2. It appears therefore that validity of linear approximation is limited by the lifetime of long-lived constituents. For instance, we expect the linearization to fail when rapid photochemical production or destruction of ozone, such as that occurring in the ozone hole, takes place. It is also possible that linear approximation will not hold as well in the presence of high concentrations of background sulfate aerosol, for example, after heavy volcanic eruptions. These limitations should be kept in mind when interpreting our results.

While (23) represents relative difference between the linear and nonlinear models, it does not characterize the error in assuming linear dependency of the results on the size of initial perturbation $\Delta\mathbf{x}$. In order to estimate this error, we computed the ratio of the nonlinear and linear terms in the Taylor series expansion of $\mathbf{M}(\mathbf{x}_o+\Delta\mathbf{x}, t)$:

$$\delta(t)=\frac{|\mathbf{M}(\mathbf{x}_o+\Delta\mathbf{x},t)-\mathbf{M}(\mathbf{x}_o,t)-\mathbf{L}(t)\Delta\mathbf{x}|}{|\mathbf{L}(t)\Delta\mathbf{x}+10^{-20}|} \quad (24)$$

The results are shown in Plate 3. As one can see, the magnitude of $\delta(t)$ can be rather large. Particularly surprising is the rapid growth and decrease of the error in case of ozone simulation. Upon examination of the results we discovered that at the time the error reaches its maximum (day 7), the nonlinear as well as linear simulations, $\mathbf{x}_{NL}(t)$ and $\mathbf{x}_L(t)$, approach the basic state $\mathbf{M}(\mathbf{x}_o,t)$. Therefore, the magnitude of the linear term $\mathbf{L}(t)\Delta\mathbf{x}$ decreases drastically, while the magnitude of the nonlinear term $|\mathbf{M}(\mathbf{x}_o+\Delta\mathbf{x}, t) - \mathbf{M}(\mathbf{x}_o,t) - \mathbf{L}(t)\Delta\mathbf{x}|$ remains approximately the same (about 1% of the basic state, as seen in Plate 2). This leads to a rather intuitively obvious conclusion that linear approximation becomes invalid when the magnitude of $\mathbf{L}(t)$ becomes very small. Nevertheless, it appears that linear approximation remains reliable for up to about 3-4 days.

6. Evolution of the Linearization and Covariance Matrices

The photochemical box model $\mathbf{M}$ described above can be thought of as an algebraic transformation from N-dimensional space of constituent concentrations at present time to some future time, as illustrated in Figure 4. N is the length of the vector $\mathbf{x}$, i.e. the number of modeled species. As was demonstrated in the previous section, for small (<10-20%) increments $\Delta\mathbf{x}$ to the basic state, the nonlinear transformation given by the original model can be successfully approximated by linearization of the model:

$$\Delta\mathbf{x}_{t+\Delta t}=\mathbf{L}\Delta\mathbf{x}_t \quad (25)$$

The linearized model $\mathbf{L}$ is an $N\times N$ matrix, and therefore the transformation from present to future constituent space can be approximated as multiplication of matrix $\mathbf{L}$ by vector of concentration increments at the present time $\Delta\mathbf{x}_t$. This framework allows us to apply methods of linear algebra for quantitative investigation of some properties of the model.

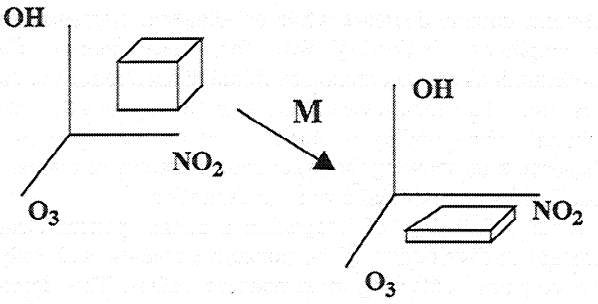

Figure 4. The photochemical model can be thought of as an algebraic transformation from the space of constituent concentrations at past time to future time.

Plate 4 presents normalized matrix $\mathbf{L}$ for Δt=0, 6, and 12 hours and 1, 3, and 6 days since the beginning of integration. The normalization was introduced to facilitate visualization of matrix $\mathbf{L}$, which is made difficult due to very large differences in the magnitudes of elements of $\mathbf{L}$. To introduce normalized linearization matrix, we assume that $\Delta\mathbf{x}_t$ and $\Delta\mathbf{x}_{t+\Delta t}$ can be written as follows:

$$\Delta\mathbf{x}_t=\mathbf{S}_x\mathbf{f}_t \quad (26)$$

$$\Delta\mathbf{x}_{t+\Delta t}=\mathbf{S}_x\mathbf{f}_{t+\Delta t} \quad (27)$$

where matrix $\mathbf{S}_x$ is a diagonal matrix, with its diagonal elements being the scaling factors for the concentrations. These diagonal elements are set equal to the concentrations of model species at time t, i.e., elements of $\mathbf{x}_t$. Elements of vectors $\mathbf{f}_t$ and $\mathbf{f}_{t+\Delta t}$ thus represent fractions of these concentrations.

Combining equations (25), (26), and (27), we get

$$\mathbf{f}_{t+\Delta t}=\mathbf{S}_x^{-1}\mathbf{L}\mathbf{S}_x\mathbf{f}_t \quad (28)$$

The normalized linearization matrix shown in Plate 4 is given by

$$\mathbf{L}_N=\mathbf{S}_x^{-1}\mathbf{L}\mathbf{S}_x \quad (29)$$

and

$$\mathbf{f}_{t+\Delta t}=\mathbf{L}_N\mathbf{f}_t \quad (30)$$

$\mathbf{L}_N$ therefore represents sensitivity of elements of $\mathbf{f}_{t+\Delta t}$ to elements of $\mathbf{f}_t$. Normalized linearization matrix is shown for all 18 modeled species as well as for sulfate aerosol, concentration of which is not changing in the model. The diagonal element of $\mathbf{L}_N$ corresponding to sulfate aerosol is therefore constant and always equals 1. The related row and column contain zeroes. At the beginning of integration, Δt=0 and $\mathbf{L}_N$ is simply the identity matrix. As time increases, the structure of $\mathbf{L}_N$ alters. Values of the diagonal elements change, and nonzero off-diagonal elements start appearing. As expected, for long-lived constituents such as ozone, HCl, or HNO_3, values of the diagonal elements remain close to 1, showing that their final concentrations are largely determined by their own initial concentrations. For short-lived species such as H, OH, or Cl, undergoing strong diurnal cycles, the diagonal

elements quickly decrease while off-diagonal elements grow in amplitude, indicating that for these species final concentrations are determined by initial concentrations of species other than themselves. Note that the eighteenth matrix column, corresponding to ozone, contains mainly nonzero elements, thus showing that final concentrations of almost all species depend on initial ozone concentration.

After a few days of integration a certain pattern clearly emerges in distribution of the nonzero elements, with only a few columns containing most nonzero values. This demonstrates that a relatively small number of mostly long-lived species or families, such as NO_x, determine concentrations of all constituents in the model at later times. This perhaps helps to understand why linear approximation described in the previous section holds so well.

We found that matrix **L** is not invertable for Δt of a few hours or longer. The reason for this is that the rank of **L**, i.e. number of linearly independent rows or columns, quickly decreases with time. The rank is shown in Plate 4 on top of each plot. As one can see, after just 6 hours the rank decreases from 19 to 11 and becomes 9 after 4 days of integration (day 4 is not shown). This means that, in general, only nine linear combinations of initial species concentrations completely define concentrations of all 19 constituents after 4 days. Formally, on day 4, matrix **L** represents a transformation from 19-dimensional space to 9-dimensional space. This is a multidimensional equivalent of multiplication by zero along some of the dimensions. No matter how large some concentration was initially, in a few hours or a few days its impact might be completely negligible. Clearly, this behavior is due to a strong diurnal cycle and the short lifetime of some species in the model.

An interesting consequence of this is that the past state of the modeled stratospheric chemical system can never be determined from present observations of the system since **L** cannot be inverted. On the other hand, it means that one does not have to know concentrations of all species to predict the state at some later time. For example, provided that the model is fairly realistic, only nine linear combinations of species concentrations spanning the orthogonal space of matrix **L** need to be known in order to predict concentrations of all 19 model constituents 4 days later. Although it is fairly straightforward to determine the orthogonal space of matrix **L** and therefore the defining linear combinations of the basis species concentrations, this has little practical value. In reality, most instruments measure concentrations of individual species and not their weighted sums. Perhaps a more relevant practical question is, What is the uncertainty of the model prediction given specified uncertainty of measurements of initial concentrations? A related question is, What are the measurement requirements that yield a specified uncertainty of the prediction? The first question is fairly easy to answer using time evolution of the error covariance matrix described above. The second problem is not so trivial and will contain multiple solutions. Careful investigation of this problem is beyond the scope of this paper.

Matrix **L** can be used to compute time evolution of the error covariances in the model according to (18). We have assumed that initial concentrations of all species in the model are known with 10% uncertainty and that these errors are uncorrelated. This means that for Δt=0 all diagonal elements of the relative error covariance matrix are equal to $0.1^2 = 0.01$ and all off-diagonal elements are zeros. The relative error covariance matrices are shown in Plate 5 at the same times as in Plate 4. Shortly after the beginning of integration, the original error covariance matrix structure breaks down and a large number of nonzero, both positive and negative, off-diagonal elements appear. The nonzero off-diagonal elements represent error correlation between different model constituents. Their appearance in the covariance matrix is expected, since initial error, for instance, of ozone concentration, will "propagate" to other constituents due to photochemical interactions in the model. At different portions of the diurnal cycle, some off-diagonal elements can change sign because of photochemical transformations between some constituents. For example, an uncertainty associated with ClO concentrations during daytime will be related to uncertainty of $ClONO_2$ concentrations during nighttime and vice versa. Values of the diagonal elements will, of course, always be positive. At later times, a stable distribution of nonzero off-diagonal elements appears in the covariance matrix. Such elements indicate strong photochemical links between various species (e.g., photochemical families).

Note that the values of diagonal elements or error variances associated with atomic oxygen and ozone decrease with time. This can be understood if we note that the modeled photochemical system is not enclosed and is slowly evolving to equilibrium with the environment. The equilibrium state is determined by external conditions such as the amount of solar radiation, temperature, pressure, etc. "Excessive" odd oxygen in the model is removed from the system through catalytic destruction cycles, and just the "right" amount of new odd oxygen is produced from molecular oxygen photodissociation. Since we assumed a perfect model, the uncertainty associated with odd oxygen asymptotically decreases with time as the model approaches its new equilibrium state.

In the next sections we will utilize linearization and error covariance matrices for assimilation of UARS data into a photochemical box model using the variational technique and the extended Kalman filter.

7. UARS Observations

In this paper we use data from two UARS instruments, the cryogenic limb array etalon spectrometer (CLAES) and the Microwave Limb Sounder (MLS). Specifically, we will consider CLAES observations of O_3, NO_2, HNO_3, $ClONO_2$, CH_4, and N_2O as well as MLS observations of ClO and H_2O. This particular choice of data represents a compromise between the importance of various chemical constituents in stratospheric photochemistry, the availability of data to the authors at the time when the described computations, the quality of data, and the efficiency of computations and programing. Thus this choice of data is not optimal. An extensive UARS validation campaign has been completed, and for description of uncertainties and quality of measurements of all UARS instruments the reader should refer to a special issue of the *Journal of Geophysical Research (101* (D6), 9539-10,476, April 30, 1996). Throughout this manuscript we use UARS level 3 AT data: nongridded, along the satellite track profiles of mixing ratios.

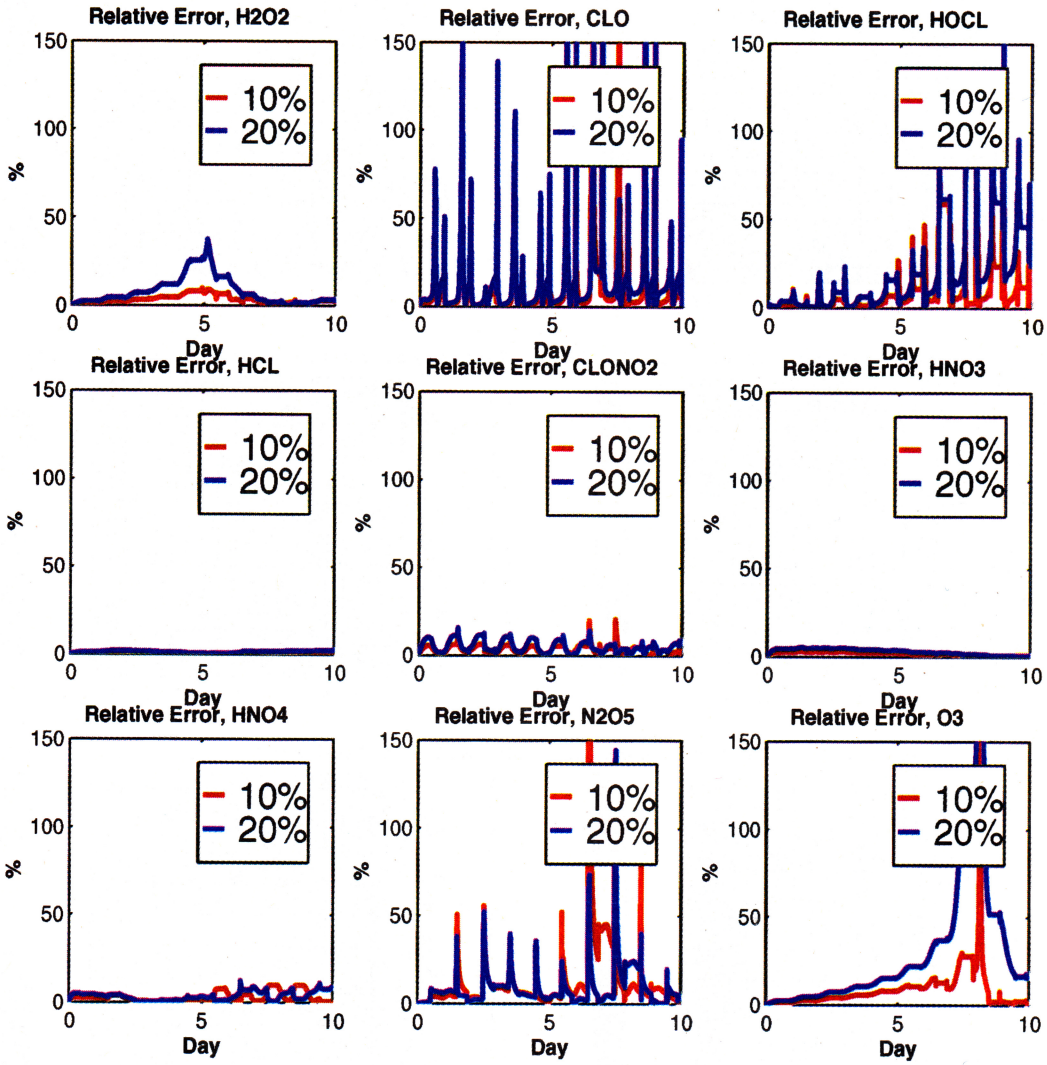

Plate 3. Ratio of the linear and nonlinear terms.

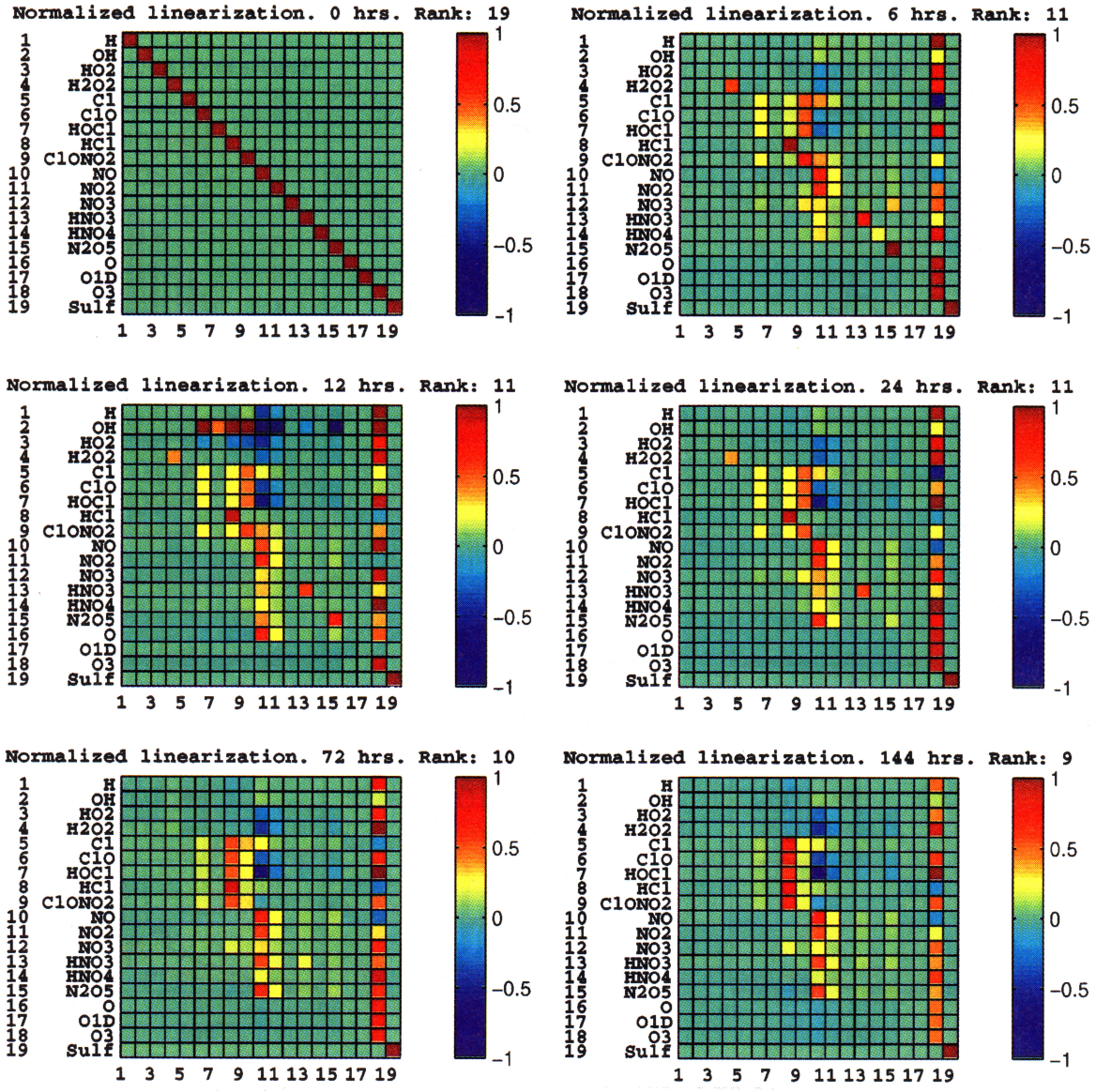

Plate 4. Normalized linearization of the photochemical box model

8. Trajectory Model and Data Selection

We have used the NASA Goddard Space Flight Center (GSFC) trajectory model described by *Newman et al.* [1988] to compute parcel trajectories on isentropic surfaces. We defined a grid with 5° latitude and 10° longitude spacing in the Northern Hemisphere and performed backward and forward trajectory computations for 24 hours, each starting from predefined grid points. Thus each trajectory arrives at its corresponding grid point simultaneously. All calculations were made at 840-K potential temperature surface, or approximately 10 mbar. The trajectory calculations described above have been used for all experiments with variational assimilation. For the extended Kalman filter analysis, we only performed the backward trajectory calculations, with each integration lasting 48 instead of 24 hours. In this case the length of analysis is the same as in the case of the variational method but all trajectories arrive at the predefined grid points at the end of the 48-hour interval rather than in the middle. This distinction is made because in a variational approach one expects to get the best results in the middle of the analysis period, while in the sequential approach the best results should be obtained at the end of the interval. Along with the horizontal coordinates the GSFC model provides temperatures along the trajectories at 30-min intervals. These temperatures were used in the box model calculations.

The choice of the day (May 3, 1993) and the pressure level (10 mbar) for the assimilation experiment is somewhat arbitrary and was dictated by the following general guidelines. The altitude level was chosen to be in the middle of the MLS and CLAES retrieval range, where retrieval errors are minimal. The model does not contain the (polar stratospheric clouds) PSC chemistry, so we chose to perform simulations in the Northern hemisphere, at the time when the PSC chemistry is not important and during relatively unperturbed dynamics conditions. At the present time we are extending the model and the assimilation scheme with a goal to systematically assimilate UARS observations on a few pressure levels and for different seasons.

After performing the described trajectory calculations, we searched for CLAES and MLS observations located within 1000 km and 1 hour from the trajectories. Measurements satisfying both these criteria were averaged together with weights inversely proportional to the measurements errors. These errors were assumed to grow proportionally to $\exp\{0.5(\Delta L/L)^2 + 0.5(\Delta T/T)^2\}$, where ΔL and ΔT are distances in space and time between a measurement and the corresponding trajectory point, L=300 km, and T=2 hours. The obtained averaged species concentrations were assigned to a point along each trajectory. Errors of individual measurements at the measurement location were obtained from the quality flag of the level 3AT data files. In most cases the quality indicator characterizes the lower boundary of the measurement error.

9. Variational Method

In the variational method we constructed the cost function from the UARS measurements and their uncertainties for each trajectory according to (13). We assumed that no background estimate of the initial concentrations ($\mathbf{x}_0$) exist, and therefore the last term in (13) was ignored. The model error covariance $\mathbf{F}$ was also ignored. We then minimized the cost function with respect to initial concentrations of all variable constituents in the model: H, OH, HO_2, H_2O_2, NO, NO_2, NO_3, N_2O_5, HNO_3, HNO_4, Cl, ClO, HOCl, HCl, $ClONO_2$, O, $O(^1D)$, and O_3. For each trajectory, concentrations of CH_4, N_2O, and H_2O were obtained from the UARS data and then simply interpolated in time along the trajectory. The analysis was done individually for each trajectory. The analysis interval is 48 hours, and the middle of this interval is when all trajectories arrive at the predefined grid points. Calculations were performed at the 840-K potential temperature surface or approximately 10 mbar.

For minimization of the cost function we employed the limited memory conjugate gradient technique described by *Buckley* [1994]. This algorithm requires knowledge of derivatives of the cost function. These derivatives were computed from the linearization matrices described above using the adjoint approach, similarly to *Fisher and Lary* [1995]. The inner product used in the minimization is the inner dot product. To account for different orders of magnitude of the concentrations of the species, a scaling similar to (26) was used.

The variational analysis consists of the following steps:

1. Compute the first-guess initial conditions by integrating the box model for 10 days at the location of the corresponding grid point using temperature, pressure, and concentrations of nonvariable species typical for that location.
2. Integrate the model starting from the current initial conditions and compute the cost function and its derivative from results of the model run and observations.
3. Pass the obtained values of the cost function and its gradient to the minimization routine, which returns a vector of new initial conditions.
4. Check for convergence. If convergence is not reached, return to step 2.

Typically, about 10 iterations are needed to decrease the cost function by a factor of 10-100 for each trajectory. This translates into 10 integrations of the box model and 10 integrations of the corresponding adjoint model. Once the convergence is reached, model concentrations corresponding to the middle of the integration are extracted and used to produce global distributions of modeled species. The CPU time required to complete the assimilation analysis for one trajectory on an IBM 590 RS 6000 is approximately 1 min.

An important issue in the variational approach is the shape of the cost function, in particular its smoothness (differentiability) and existence of multiple minima. If the cost function is not differentiable, then the minimization algorithm relying on the derivative of the function is likely to fail. If the cost function has multiple minima, then the minimization process might find a local minimum rather than a true, global minimum. It is practically impossible to numerically investigate all properties of the cost function due to its high dimensionality. In our case the cost function depends on 18 variables, and even storing a crudely discretized cost function is only marginally possible. Nevertheless, it might be beneficial to examine at least some cross-sections of the cost function. We computed a few cost function cross sections for several trajectories by varying initial concentrations of up to three species at a time,

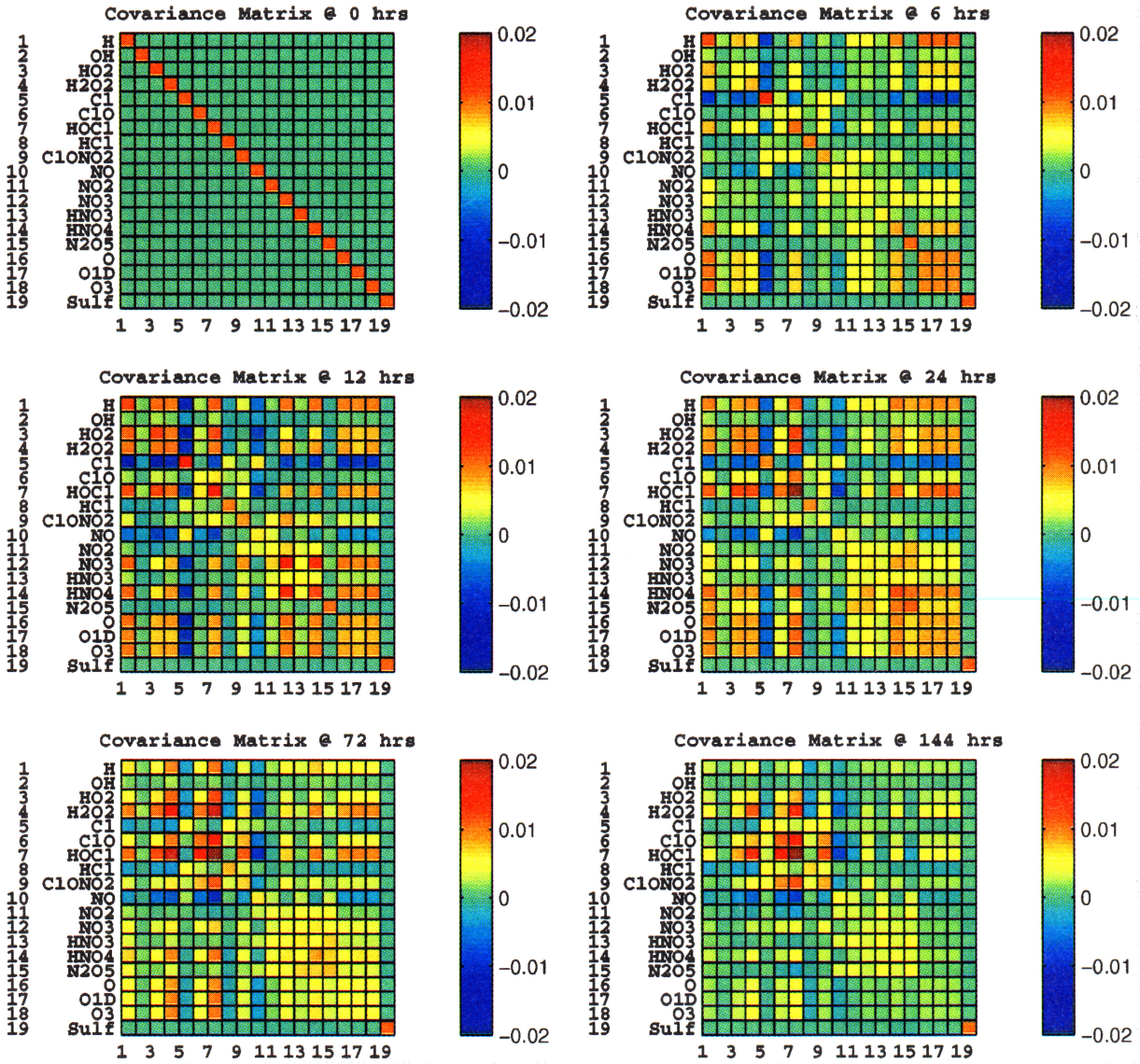

Plate 5. Relative error covariance matrix.

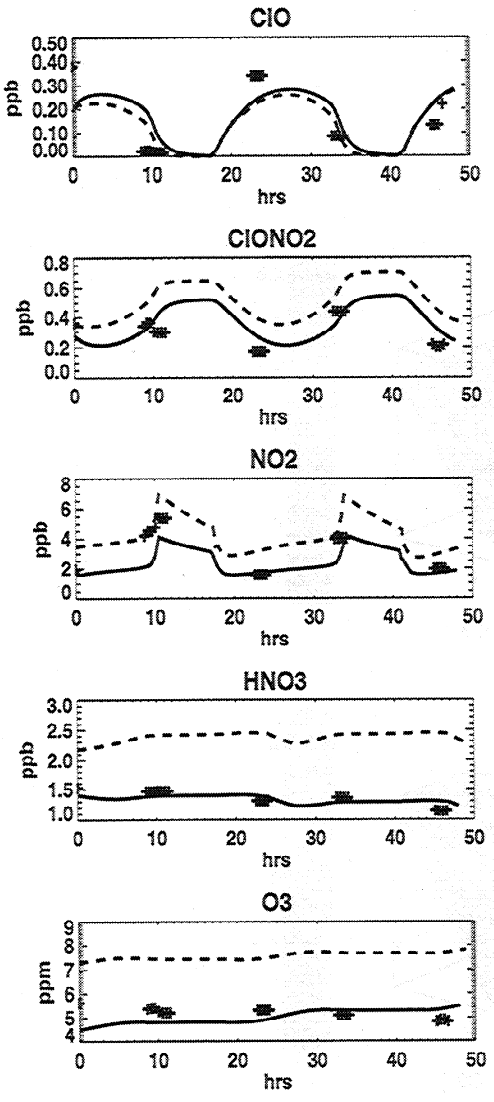

Figure 5. Results of variational assimilation for one trajectory. Dashed lines correspond to the very first box model integration with the first-guess initial conditions. The crosses are the UARS observations for the trajectory, and the solid lines are the final model integration starting from the optimized initial conditions.

including H, OH, NO, NO_2, NO_3, HNO_3, ClO, $ClONO_2$, and O_3. Some typical cross sections of the logarithm of the cost function are shown in Plate 6. In all investigated cases the resulting cost function cross sections were smooth and had either one global minima or a fold line, thus indicating that the cost function is not sensitive to changes in initial concentration of the species parallel to the trough line.

Results of the variational assimilation are presented in Figure 5 for one trajectory. The dashed line corresponds to the very first box model integration with the first-guess initial conditions. The crosses are the UARS observations for the trajectory, and the solid line is the final model integration starting from the optimized initial conditions. The agreement with observations is clearly better in the second case for all species except ClO. The variational method is in effect a constrained weighed least squares fit to all observations, where constraints are given by the model equations. The relative weight of observations depends on their uncertainties (see equation (13)). The reported error of the MLS ClO measurements is much higher than errors of observations of other species, and therefore ClO observations have relatively small influence on the results.

Global maps of some model species obtained by this method are shown in Plate 7 for 1200 UT on May 3, 1993. As one can see, the presented distributions are fairly realistic. Ozone concentrations are largest near the equator and decrease toward the pole, as expected. Distributions of short-lived species show clearly defined terminator, thus indicating that these distributions are truly instantaneous. Compare a map of NO_2 mixing ratio in Plate 7 with that in Plate 1. The latter plate is produced from the level 3B UARS data and shows very little longitudinal variation in NO_2 distribution, while the map in Plate 7 clearly indicates higher NO_2 mixing ratios in the nighttime regions and lower mixing ratios in the sunlit regions. Note that concentrations of species other than those observed by the UARS are presented in Plate 7 (e.g., O, NO, and Cl). Concentrations of these chemicals have been constrained by the available observations during assimilation.

In order to provide an approximate estimate of how well the assimilation results agree with observations, we calculated the root mean squares (RMS) deviations of our results from observations for each trajectory. Plots of these deviations are shown in Plate 8 for all five assimilated species. Note that the shown RMS differences should be interpreted not as the uncertainties of the assimilation results but rather as average deviations from UARS measurements. This might explain sporadic areas of high RMS differences seen in the plots: If a nearby observation has a large uncertainty, it will not affect the assimilation results, and the corresponding RMS difference might become quite large. The RMS differences are due to inability of the photochemical model to precisely fit all data points at the same time. This, in turn, can be caused by a few reasons. An obvious reason is that observations themselves have errors and therefore do not satisfy the model equations. Second, some processes occurring in the real atmosphere are not adequately described or simply not included in either photochemical or trajectory model. This includes possible errors in the rates of photochemical reactions. The third source of differences is the error of time and space interpolation introduced when selecting UARS observations for each trajectory. Estimating contribution of each source as well as determination of errors of the final analysis constitute a significant research project which is beyond the scope of this paper.

10. Extended Kalman Filter

For the extended Kalman filter analysis we performed backward trajectory calculations for 48 hours starting from the predefined grid points. This was done because in the sequential method, observations made after the time for which state estimates are sought do not influence the analysis at that time. To keep the number of observations for each trajectory approximately the same as in the variational method, we use the same length of the analysis interval (48 hours).

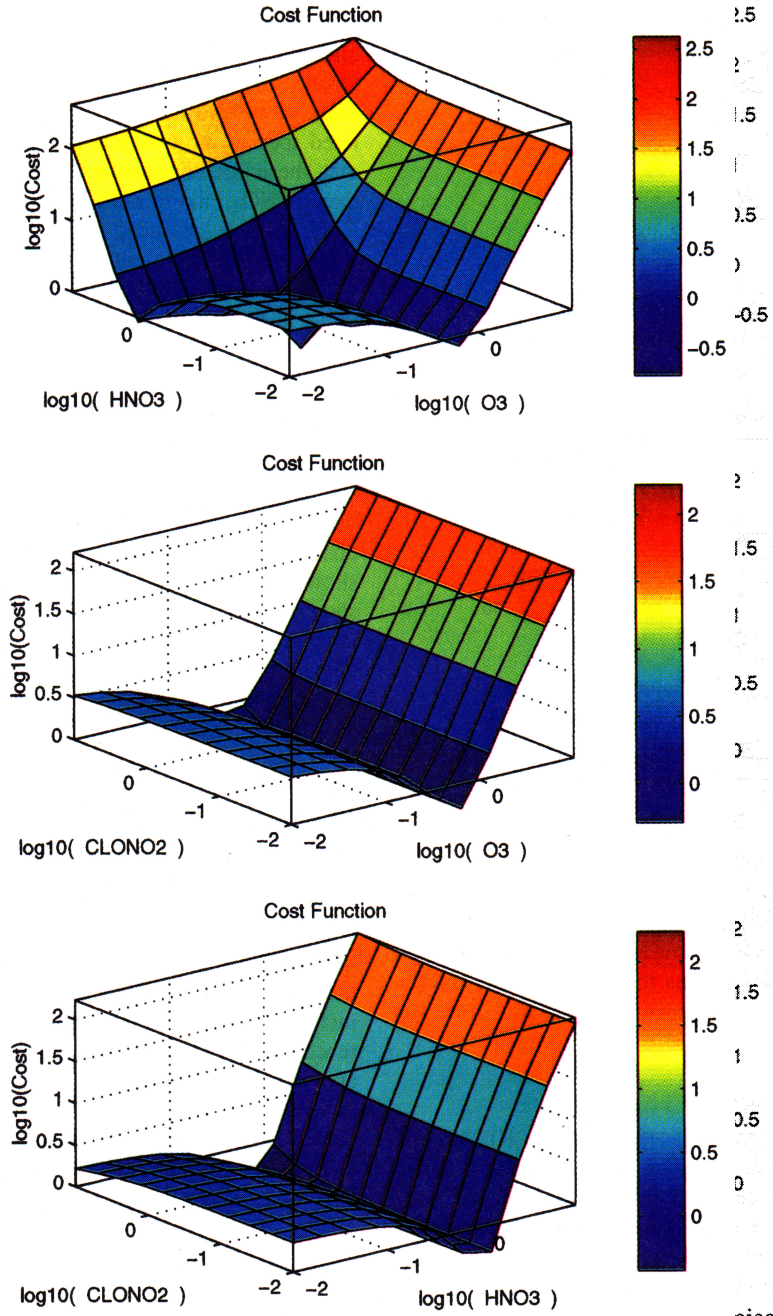

Plate 6. Sample cross sections of the cost function logarithm for one trajectory.

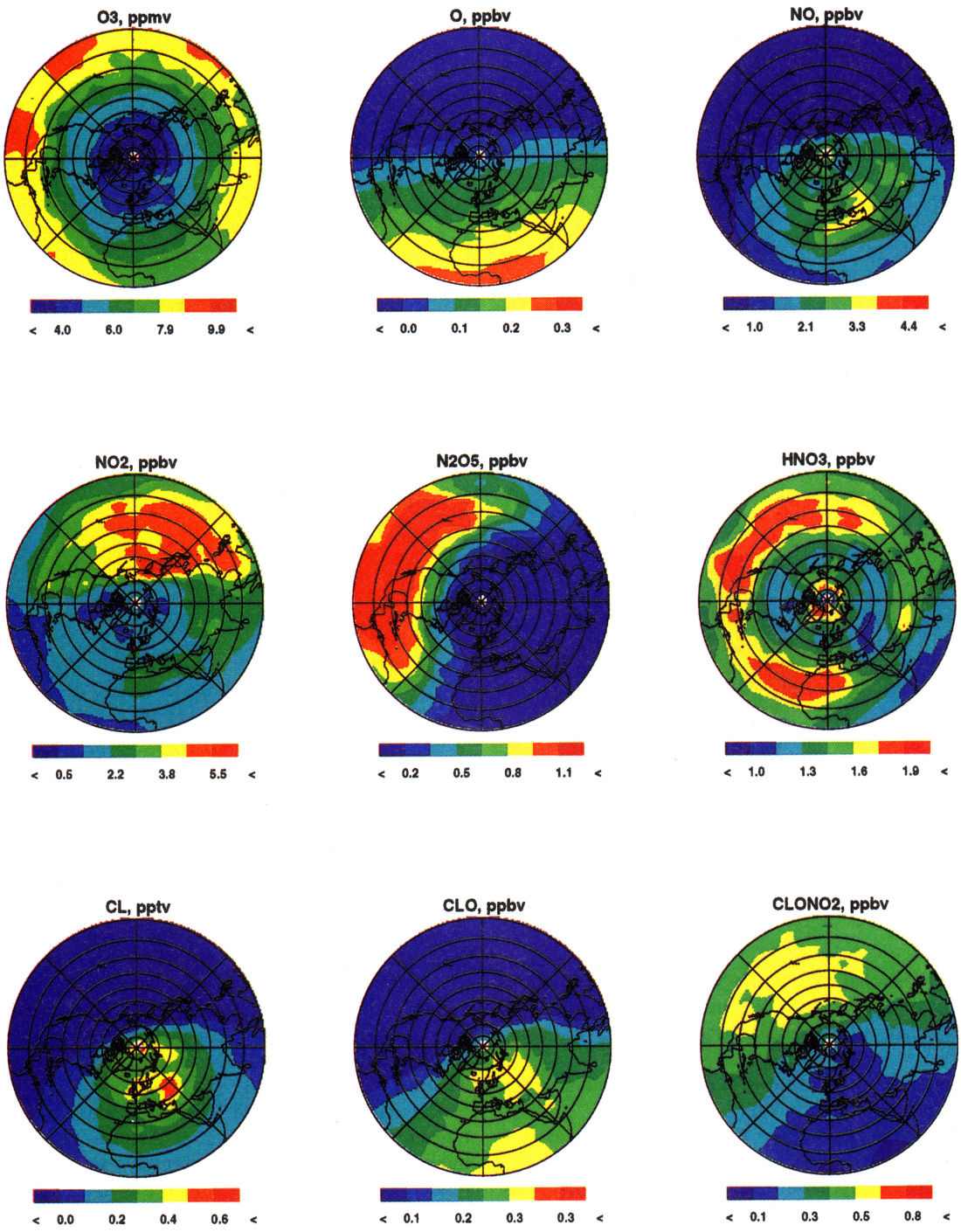

Plate 7. Synoptic maps of some species produced with the variational method for May 3, 1993, at 1200 UT.

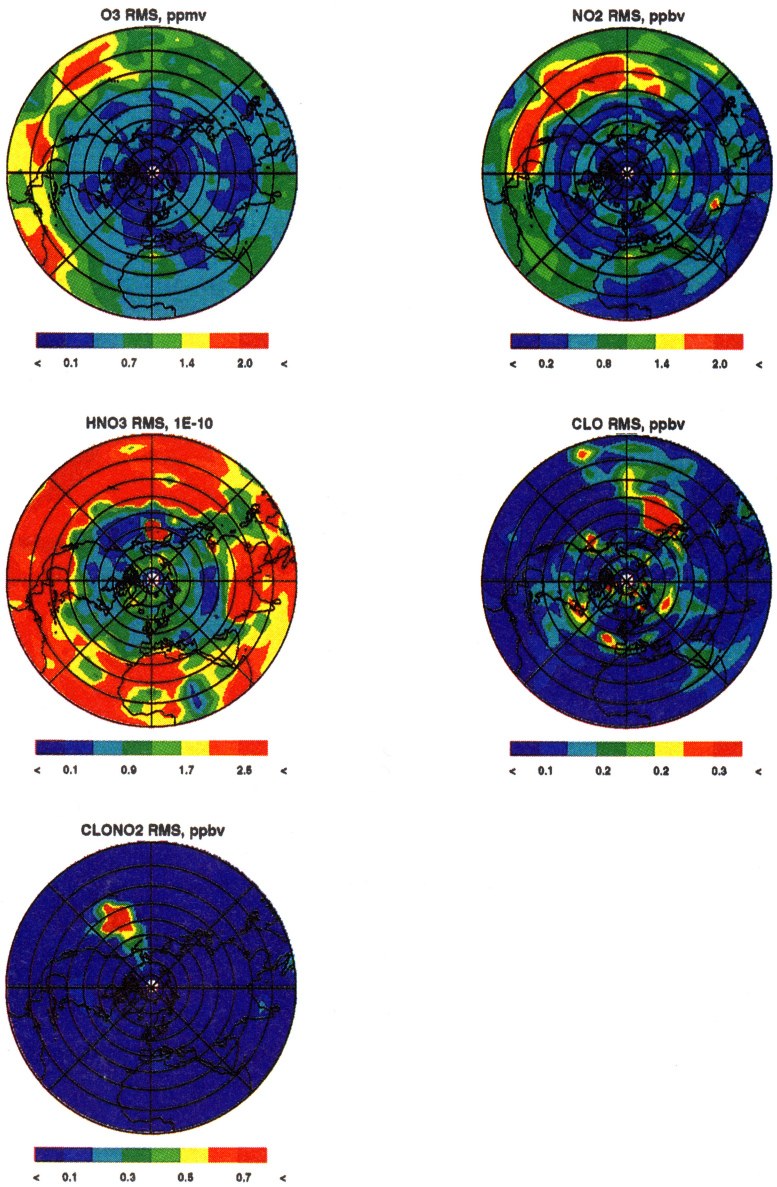

Plate 8. Averaged RMS errors for assimilated species in the variational method.

As in the variational case, vector **x** contains concentrations of 18 variable species included in the box model. The extended Kalman filter analysis consists of the following steps:

1. Compute initial value of **x** as in the variational approach and initialize the covariance matrix.
2. Integrate the model starting from the current value of **x** until time of the next observation.
3. Update the background error covariance matrix using (18).
4. Perform analysis using equations (15) and (16).
5. Stop if the end of the analysis period is reached. Otherwise, return to step 2.

The state variables in the extended Kalman filter calculations are concentrations of variable species. The initial state of the filter is the same as in the variational approach. The starting analysis error covariance matrix is diagonal, with variances corresponding to 10% of the initial concentrations. The CPU time required to complete the assimilation analysis for one trajectory on an IBM 590 RS 6000 is approximately 30 sec.

Results of the extended Kalman filter assimilation for one trajectory are shown in Figure 6. As one can see, the model mixing ratios are rather abruptly modified when an observation is encountered, similar to what is shown in Figure 3. Magnitude of the correction depends on the error of the observation and the current error covariance matrix. Note that for long-lived species with relatively small observational error, O_3 and HNO_3, the analysis fits observations much closer than in the case of variational assimilation. This behavior is expected since the model constraints are violated at the times of observations, thus allowing for more degrees

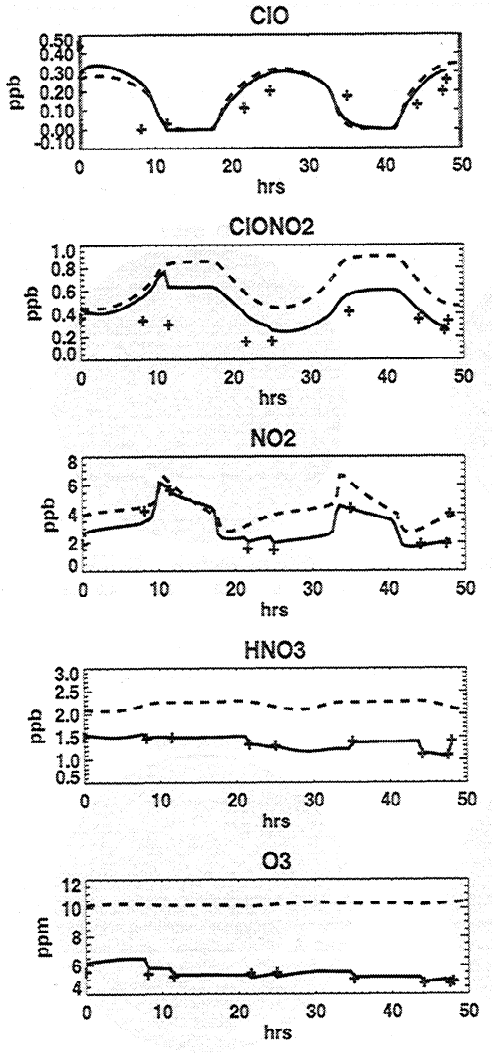

Figure 6. Results of the extended Kalman filter assimilation for one trajectory. Dashed lines correspond to the very first box model integration without incorporation of observations. The crosses are the UARS observations, and the solid lines are the extended Kalman filter results.

of freedom and a better fit. In the variational assimilation, analysis generally has discontinuities at the end of the analysis interval, while in the sequential method a discontinuity normally appears whenever a new observation is introduced. For species with relatively large observational errors, ClO and $ClONO_2$, the corrections are rather small.

A fairly common problem arising in applications of the (extended) Kalman filter is the so-called filter divergence. If the model error is assumed to be zero, the analysis may eventually diverge from observations since observations are rejected in favor of the model solution, which is then no longer constrained by reality. In our case, however, the divergence did not occur, probably due to a relatively short analysis window (2 days).

Global distributions of some species produced with the extended Kalman filter technique and the corresponding RMS differences for observed species are shown in Plate 9 and Plate 10. General features of the constituent distributions are fairly similar to those obtained in the variational method (Plate 7), but the absolute values of mixing ratios are noticeably different in some cases. At the same time, the RMS deviations (Plate 10) are smaller in the case of the extended Kalman filter assimilation, which is to be expected as discussed above. Possible reasons for the observed differences between the variational and extended Kalman filter analysis will be discussed in the next section.

11. Summary and Conclusions

Nearly global coverage, relatively high spatial density, and large volume of satellite-supplied measurements provide opportunities to study a wide range of atmospheric phenomena and make it possible, at least in principle, to constrain global atmospheric models with real data. At the same time, satellite measurements are usually irregular in both time and space while models operate on regular grid points. Orbital geometry imposes significant limitations on satellite sampling capabilities which may result in large data voids. In addition, short-lived atmospheric constituents experiencing variability on timescales of hours to days generally cannot be adequately sampled by satellite-based instruments. These features of satellite observations require unconventional methods of data analysis.

Analyses which rely on data assimilation methodology are potentially superior to conventional means of data processing in that they utilize our knowledge of the underlying physical processes governing the behavior of the physical system. This allows one to better utilize observations. For example, data assimilation provides ways to optimally fill in spatial and temporal gaps. This feature makes data assimilation very attractive for analysis of irregularly distributed, asynoptic or sparse observations, such as those collected by satellites, and for combining data sets from multiple instruments and platforms. In this paper we attempted to provide a general introduction to the related basic mathematical apparatus and considered applications of these mathematical techniques to a photochemical box model and analysis of the UARS chemical data.

It was demonstrated that in the case of a stratospheric photochemical box model the linear approximation essential to applicability of both the variational and the extended Kalman filter techniques holds fairly well for up to 10 days. This behavior can be explained by noting that concentrations of many short-lived species in the model are largely determined by concentrations of a few relatively long-lived species such as ozone and parameters such as total active chlorine or nitrogen.

Analysis of time evolution of linearization and error covariance matrices gives further evidence to an idea that the future evolution of the system is completely determined by a few linear combinations of species concentrations at the beginning of integration as the rank of linearization matrix decreases with increasing integration length. It appears that analysis of these matrices might provide interesting insights on interdependency of chemically active constituents in the model. The described calculations might be useful in determining uncertainty of model prediction given uncertainty of measurements. A related problem is determining

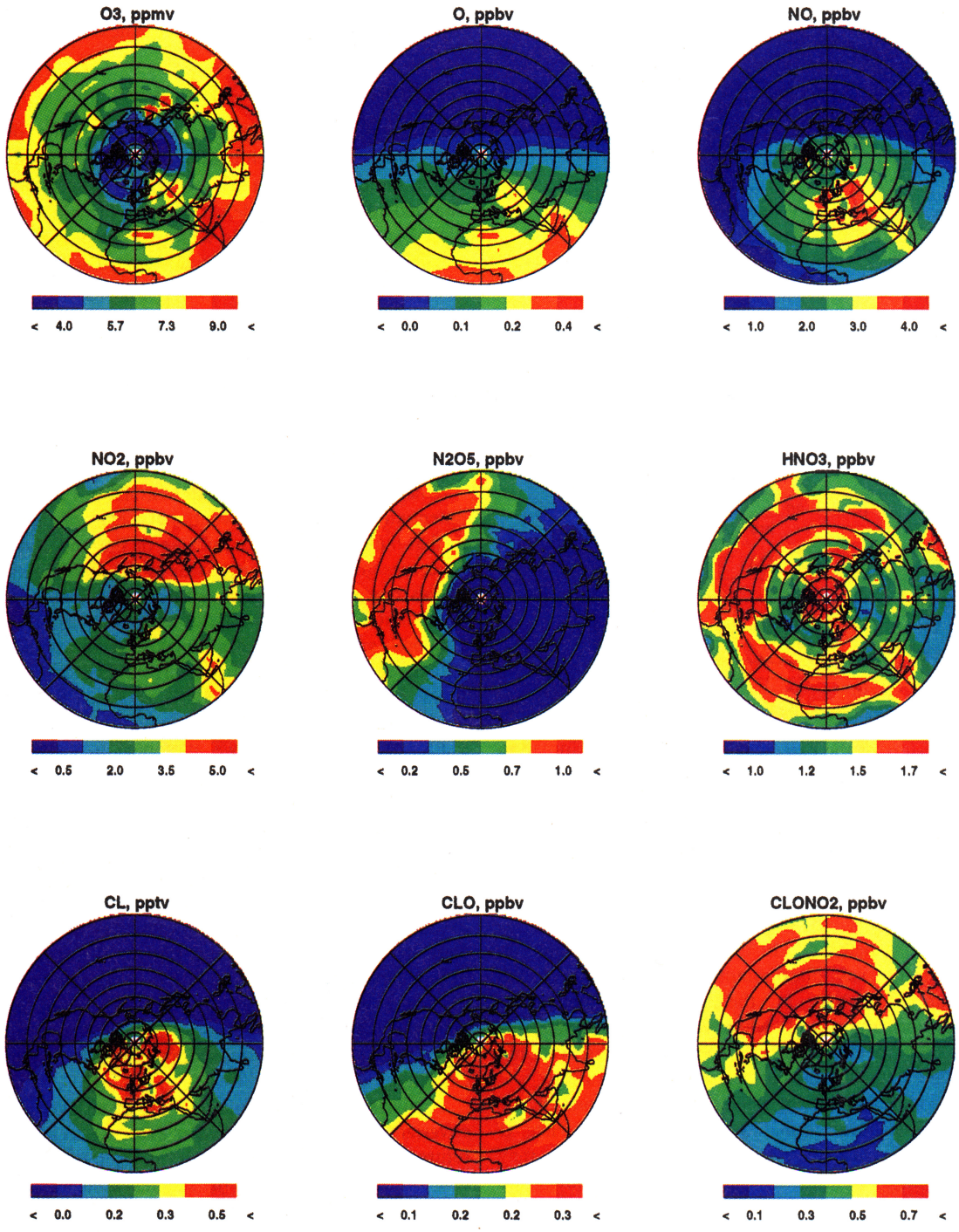

Plate 9. Synoptic maps of some species produced with the extended Kalman filter for May 3, 1993, at 1200 UT.

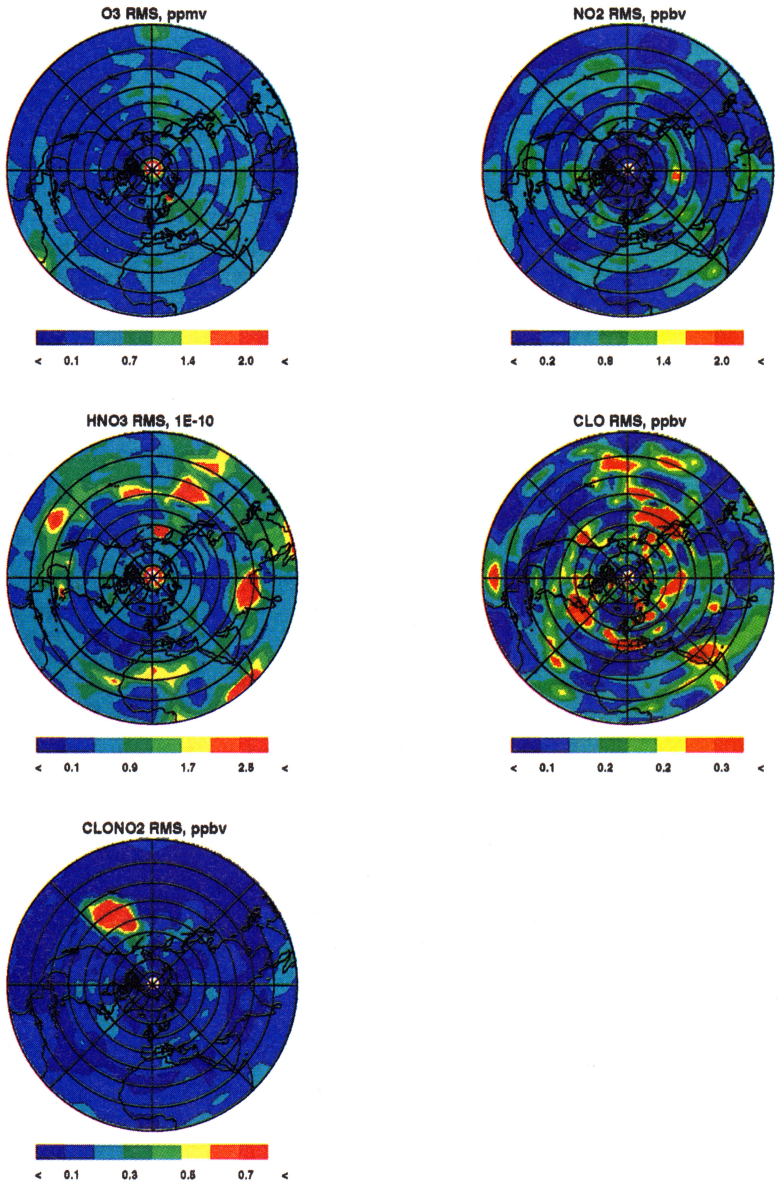

Plate 10. Averaged RMS errors for assimilated species in the extended Kalman filter method.

measurement requirements necessary to predict the future system state with a specified uncertainty. Note, however, that only the (extended) Kalman filter provides this opportunity as a part of the algorithm.

A few cross sections of the cost function appearing in the variational method have been computed, and it was shown that these cross sections are smooth and do not have multiple minima. This gives some confidence in the applicability of the variational approach to chemical data assimilation.

Synoptic maps of various species obtained with both variational assimilation and the extended Kalman filter display the same general features with a clearly defined terminator in distributions of constituents expected to have strong diurnal cycles. Absolute values of mixing ratios can be quite different in the two techniques. These differences can be attributed to a number of reasons.

Perhaps the most important reason is the fundamental difference in the basic principles of the variational and sequential techniques. In the variational method the 48-hour analysis (in our case) is the true solution of the model equations. Therefore it is implicitly assumed that errors due to uncertainties of the (photo)chemical reaction rates, trajectory calculations, missing photochemical processes, etc., are significantly smaller than the observational errors. The only set of adjustable parameters is the initial species concentrations. It is assumed that correct determination of these concentrations will allow the model to fit observations within the measurement error. Clearly, this assumption is not very realistic. However, as Figure 5 demonstrates, it results in a much better agreement of the analysis and observations than the model simulations alone. By assuming a perfect model, the variational method adjusts to both initial state errors and model error by adjusting only the initial state. Depending on the number and location of observations and the background conditions during analysis, minimization may or may not find the initial state that results in simulations fitting observations.

For instance, D. Lary (personal communication, 1998) found that it was impossible to fit a set of particular observations without including additional chemicals in the model. In such situations, the variational method provides a means to systematically validate models against observations by analyzing residual differences (cost function values) at the end of the minimization. If these differences are larger than typical observational errors, then either the model is unable to simulate the reality or observational errors are underestimated. Another way to validate the model in the framework of the variational approach is to consider the forecast made at the end of the analysis interval and compare the forecast value with independent observations.

The extended Kalman filter analysis, generally speaking, does not satisfy the model equations. Clearly, discontinuities seen in Figure 6 are unrealistic and can never be reproduced by solving the model photochemical equations for the 48 hours. This, however, does not make the sequential approach less valuable than the variational method, where a smooth solution might appear as a result of adherence to possibly incorrect model dynamics. The extended Kalman filter method might be more advantageous when the goal is to predict the state of the photochemical system in between just two consecutive observations or if it is needed to predict the state at some later time, when no observations are available. It is also possible that with some specification of model error the discontinuities seen in Figure 6 would be reduced. An important advantage of the extended Kalman filter technique is that it allows one to explicitly account for the model errors, provided, of course, that these errors are known. This aspect of the filter was not considered in our work.

One other possible reason for the observed differences is incorrect specification of observational error covariances. The observational errors are likely to be unrealistic and underestimated, as was mentioned in section 7. This will cause the extended Kalman filter to give more weight to the observations and lead to larger discontinuities. In addition, the sudden change of concentrations of the observed species in the model will introduce a "shock" in the system. As a result, some species might experience nonlinear transition to a new equilibrium state. This might affect our final results if the observation was encountered shortly before the trajectory arrives to the predefined grid point.

Note also that different sets of observations were used in each case due to different trajectory model initialization. Perhaps a better way to compare the two techniques would be to use the so-called Kalman smoother. In this method the usual extended Kalman filter analysis is followed by the analysis performed backward in time, thus propagating information from the future observations backward. In this case the same set of trajectories and observations could be used in both methods. This does not mean, however, that the Kalman smoother is preferable to the (extended) Kalman filter. Rather, the choice of the technique should be dictated by goals of a particular analysis.

The performed research demonstrates that both variational assimilation and the extended Kalman filter technique can be used for assimilation and mapping of satellite observations of a number of chemical species using relatively complex numerical photochemical models. It is very desirable to apply these techniques for producing consistent, global climatologies of atmospheric chemicals that are constrained by global satellite measurements.

Depending on the application, either method can be used for as long as estimates of the analysis errors are provided. If the goal is to simply produce instantaneous analysis that best agrees with observations and/or if the observational errors are relatively small, then the extended Kalman filter (or smoother) might be more advantageous than the variational technique. The sequential technique is also appropriate when analysis needs to be performed systematically, day after day, for prolonged periods of time. If the goal is to produce analysis over a moderately long finite period that best fits the observations within the limits imposed by model constraints or if the observational errors are large, the variational technique might be more suitable. The variational method will also be useful when one seeks to diagnose the quality of the model. In either case, it appears that more realistic estimates of corresponding observational errors are needed for practical applications of these techniques.

Acknowledgments. The authors are very grateful to Jean-Francois Lamarque, Bill Randel, Larry Horowitz, Liwen Pan, David Lary, Mike Fisher, Richard Menard, Valery Yudin, Pieternel Levelt, Bonnie Elizabeth Padgett, and Henk Eskes for fruitful discussion and valuable comments. We thank two anonymous reviewers for their detailed comments and help in improving the manuscript. The authors acknowledge use of the Goddard Space Flight Center trajectory model. Research of B.V.K. was supported by the NASA UARS project under NRA-97-MTPE-04, the NASA MOPITT project, and the National Center for Atmospheric Research. The National Center for Atmospheric Research is sponsored by the National Science Foundation.

References

Austin, J., Toward the four-dimensional assimilation of stratospheric chemical constituents, *J. Geophys. Res., 97*, 2569-2588, 1992.

Buckley, A. G., Algorithm 734: A Fortran 90 code for unconstrained nonlinear minimization, *Trans. Math. Software, 20*, 354-372, 1994.

DeMore, W. B., S. P. Sander, D. M. Golden, R. F. Hampson, R. E. Huie, M. J. Kurylo, C. J. Howard, A. R. Ravishankara, C. J. Kolb, and M. J. Molina, Chemical kinetic and photochemical data for use in stratospheric modeling: Evaluation No. 11 of the NASA Panel for Data Evaluation, *JPL Publ. 94-26*, 1994.

Dvortsov, V., S. Zvenigorodsky, and S. Smyshlyaev, On the use of Isaksen-Luther method of computing photodissociation rates in photochemical models, *J. Geophys. Res., 97*, 7593-7601, 1992.

Elbern, H., H. Schmidt, and A. Ebel, Variational data assimilation for tropospheric chemistry modeling, *J. Geophys. Res., 102*, 15,967-15,985, 1997.

Eskes, H. J., A. J. M. Piters, P. F. Levelt, M. A. Allaart, and H. M. Kelder, Variational assimilation of total ozone satellite data in a 2D lat-lon tracer-transport model, *J. Atmos. Sci.*, in press, 1999.

Fisher, M., and D. J. Lary, Lagrangian four dimensional variational data assimilation of chemical species, *Q. J. R. Meteorol. Soc., 121*, 1681-1704, 1995.

Griewank, A., On automatic differentiation, in *Mathematical Programming: Recent Developments and Applications*, pp. 83-108, Kluwer Acad., Norwell, Mass., 1989.

Haggard, K.V., E. E. Remsberg, W. L. Grose, J. M. Russell II, B. T. Marshall, G. Lingenfelser, Description of data on the Nimbus 7 LIMS map archive tape, *NASA Tech. Pap. 2553*, 1986.

Jazwinski, A. H., *Stochastic Processes and Filtering Theory*, Academic, San Diego, Calif., 1970.

Kohri, W.J., LRIR observations of the structure and propagation of the stationary planetary waves in the Northern Hemisphere during December, 1975, *NCAR/CT-63*, Natl. Cent. for Atmos. Res., Boulder, Colo., 1981.

Levelt, P. F., M. A. F. Alaart, and H. M. Kelder, On the assimilation of total-ozone satellite data, *Ann. Geophys., 14*, 1111-1118, 1996.

Levelt, P. F., B. V. Khattatov, J. C. Gille, G. P. Brasseur, X. X. Tie, and J. Waters, Assimilation of MLS ozone measurements in the

global three-dimensional chemistry-transport model ROSE, *Geophys. Res. Lett., 25*, 4493-4496, 1998.

Lorenc, A. C., Analysis methods for numerical weather prediction, *Q. J. R. Meteorol. Soc., 112*, 1177-1194, 1986.

Lyster, P. M., S. E. Cohn, R. Menard, L.-P. Chang, S.-J. Lin, and R. Olsen, An implementation of a two-dimensional filter for atmospheric chemical constituent assimilation on massively parallel computers, *Mon. Weather. Rev., 125*, 1674-1686, 1997.

Menke, W., *Geophysical Data Analysis: Discrete Inverse Theory*, Academic, San Diego, Calif., 1984.

Morris, G., et al., Trajectory mapping and applications to data from the Upper Atmosphere Research Satellite, *J. Geophys. Res., 100*, 16,491-16,505, 1995.

Newman, P. A., M. R. Schoeberl, R. A. Plumb, and J. E. Rosenfield, Mixing rates calculated from potential vorticity, *J. Geophys. Res., 93*, 5221-5240, 1988.

Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in Fortran 77*, Cambridge Univ. Press, New York, 1996.

Riishojgaard, L. P., On four-dimensional variational assimilation ozone data in weather prediction models, *Q. J. R. Meteorol. Soc., 122*, 1545-1557, 1996.

Rodgers, C., Statistical principles of inversion theory, in *Inversion Methods in Atmospheric Remote Sounding*, edited by A. Deepak, pp. 117-134, Academic, San Diego, Calif., 1976.

Salby, M. L., Sampling theory for asynoptic satellite observations, I. Space-time spectra, resolution, and aliasing, *J. Atmos. Sci., 39*, 2577-2600, 1982a.

Salby, M. L., Sampling theory for asynoptic satellite observations, II, Fast Fourier synoptic mapping, *J. Atmos. Sci., 39*, 2601-2614, 1982b.

Sassi, F., and M.L. Salby, Fast Fourier synoptic mapping of UARS data, *J. Geophys. Res., 103*, 10,885-19,898, 1998.

Sirkes, Z., and E. Tziperman, Finite difference of adjoint or adjoint of finite difference?, *Mon. Weather Rev., 125*, 3373-3378, 1997.

Sutton, R. T., H. Maclean, R. Swinbank, A. O'Neill, and F. W. Taylor, High-resolution stratospheric tracer fields estimated from satellite observations using Lagrangian trajectory calculations, *J. Atmos. Sci., 51*, 2995-3005, 1994.

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

Weisenstein, D. K., M. K. W. Ko, J. M. Rodriguez, and N.-D. Sze, Effects on stratospheric ozone from high-speed civil transport: Sensitivity to stratospheric aerosol loading, *J. Geophys. Res., 98*, 23,133-23,140, 1993.

G. P. Brasseur, J. C. Gille, B. V. Khattatov, and L. V. Lyjak, National Center for Atmospheric Research, P. O. Box 3000, Boulder, CO, 80307. (boris@ucar.edu)

V. L. Dvortsov, NOAA, Aeronomy Laboratory, 325 Broadway, Boulder, CO, 80303.

A. E. Roche, Lokheed Research Laboratory, 3251 Hanover Street, Palo Alto, CA, 94304.

J. W. Waters, Jet Propulsion Laboratory, 4800 Oak Grove Drive, Pasadena, CA, 91109.

(Received September 11, 1998; revised March 29, 1999; accepted April 2, 1999.)