Q. J. R. Meteorol. Soc. (1995), **121**, pp. 1681–1704

Lagrangian four-dimensional variational data assimilation of chemical species

By M. FISHER*[1] and D. J. LARY[2]

[1]*Meteorological Office, UK*
[2]*Cambridge University, UK*

(Received 15 March 1994; revised 16 December 1994)

SUMMARY

For the first time, the method of four-dimensional variational data assimilation is applied to the analysis of chemically active trace species. By combining observations with a numerical model to analyse simultaneously several species over a period of a few days, the analysis method is able to exploit information which is not available to conventional analysis techniques. Moreover, effective use can be made of asynoptic observations even for species which have strong diurnal cycles. Synoptic analyses are produced. A Lagrangian approach is adopted, allowing a separation of dynamics and chemistry which considerably reduces the computational expense of the method.

KEYWORDS: Atmospheric chemistry Data analysis Lagrangian model Stratosphere Variational data assimilation

1. INTRODUCTION

The analysis of chemical trace species has received little attention in comparison with the analysis of meteorological variables. Current methods tend to treat species independently, ignoring the complex balances which exist between species. Moreover, the large diurnal variations in the concentrations of many species are either accounted for in very simple ways, or avoided by analysing concentrations at fixed local time.

Salby (1982a, 1982b) presented a temporal and spatial interpolation method for producing synoptic analyses from asynoptic satellite data. This method is well suited to the analysis of slowly varying fields, but is not suitable for the analysis of species with short photochemical timescales such as NO_2 (Salby 1987). In an alternative approach, Haggard *et al.* (1988) used a Kalman filter to analyse retrievals of water vapour and NO_2 from the LIMS instrument (Gille and Russell 1984) on board the Nimbus-7 satellite. Their method was based on a linear model of the variation of species concentration along the orbital track of the satellite and took advantage of the sun synchronous orbit of Nimbus 7 to produce analyses which are valid for fixed local time.

Austin (1992) analysed observations of chemical species from the LIMS instrument by inserting observations of O_3, HNO_3, NO_2, and H_2O into a two-dimensional chemical model. Austin's method—a simple form of data assimilation—has several advantages. The use of a chemical model allows knowledge of atmospheric chemistry to be incorporated into the analysis procedure. The model is capable of predicting accurately the time evolution of species over short periods, allowing information from asynoptic observations to be propagated to a common synoptic time. With Austin's method it is also possible to analyse several species simultaneously.

Although a significant step forward, direct insertion of observations into a chemical model has some disadvantages. First, the analysed concentrations do not satisfy the equations of the model since the equations are modified by the addition of forcing terms which 'nudge' the analysed concentrations towards the observed concentrations. Second, observations may be 'rejected'. For example, if the daytime concentration of NO_2 is perturbed by insertion of an observation, the model will rapidly adjust the perturbed value to restore photochemical equilibrium with other species in the NO_x family. In meteorological data assimilation, rejection is minimized by balancing the increments to the model (Bengtsson

* Corresponding author, current address: European Centre for Medium-Range Weather Forecasts, Shinfield Park, Reading, Berkshire RG2 9AX, UK.

and Gustafsson 1971). A similar procedure is possible when assimilating chemical species. For example, given an observation of one species in a photochemical family, equilibrium equations can be used to calculate consistent changes to other species within the family. The success of this approach depends on the accuracy of the assumption of photochemical equilibrium. A particular problem is that the partitioning of species within a family may depend on the concentrations of species which are outside the family. The balancing process may introduce errors if the concentrations of these species are poorly known. Direct insertion may also be incapable of correctly inducing the concentration for such 'controlling' species even when the partitioning of species within the family is known.

The Upper Atmosphere Research Satellite (UARS) has provided a rich set of observations of trace species concentrations in the middle atmosphere (see, for example, Reber 1993). The asynoptic nature of the observations, the precessing orbit of the satellite and the ability of the observing instruments to produce collocated observations of several species highlight the shortcomings of conventional analysis techniques. In this paper, we present an analysis method which is well suited to UARS observations, and which overcomes the shortcomings of direct insertion data assimilation. Our method of analysis uses the technique of four-dimensional, variational data assimilation (4D-Var) which was developed in the context of the analysis of meteorological variables by several authors, including Lewis and Derber (1985), LeDimet and Talagrand (1986), and Talagrand and Courtier (1987). The method may be regarded as an application of the theory of optimal control (Lions 1971). 4D-Var is currently being considered by a number of centres for the operational analysis of meteorological observations for numerical weather prediction. The method has been applied to the analysis of humidity, treated as a passive tracer, by Andersson *et al.* (1992). A comprehensive bibliography for the subject has been given by Courtier *et al.* (1993).

4D-Var seeks to produce an analysis which fits a set of observations taken over a period of time, subject to the strong constraint that the evolution of the analysed quantities is governed by a deterministic model. By imposing the equations of the model as strong constraints, the analysis problem is reduced to that of determining initial values for the model such that the subsequent evolution minimizes a measure of the fit to the observations.

The model used here is a photochemical 'box' model. That is, the model simulates the evolution of chemical trace species for a number of independent air parcels whose trajectories are assumed to be known *a priori*. The analyses of dynamical and chemical variables proceed separately, thus allowing a considerable reduction in computational cost since it is unnecessary then to model the entire three-dimensional domain, or to include a dynamical model in the iterative analysis procedure. However, separation of the chemical and dynamical analyses prevents the use of some useful information. Specifically, observations of chemical species contain information about the wind and temperature distributions which is ignored in our analysis scheme.

We have evaluated our analysis method in two ways, which we discuss in sections 4 and 6, respectively. In section 4 we apply the technique to the analysis of concentrations of trace species for the idealized case of a single, stationary air parcel. Observations are taken from an integration of the photochemical model. In section 6 we apply the method to the analysis of retrievals of concentrations of trace species from measurements made by the MLS and CLAES instruments on board the UARS satellite. We demonstrate that the method is capable of analysing such observations accurately.

The analysis method and the chemical model are described in sections 2 and 3. In section 5 we discuss the ability of four-dimensional variational analysis to make use of information which is not available to conventional analysis methods or to simpler methods of data assimilation. Conclusions are given in section 7.

2. Analysis method

4D-Var expresses the analysis problem as the minimization of a cost functional, $\mathcal{J}$, defined as

$$\mathcal{J} = \frac{1}{2}(\mathbf{x}_b - \mathbf{x}_0)^T\mathbf{B}^{-1}(\mathbf{x}_b - \mathbf{x}_0) + \frac{1}{2}\sum_{n=0}^{N}(\mathbf{y}_n - \mathbf{s}_n)^T\mathbf{R}_n^{-1}(\mathbf{y}_n - \mathbf{s}_n). \quad (1)$$

Here, $\mathbf{x}_0$ is the vector of initial parcel concentrations, $\mathbf{x}_b$ is an independent estimate of the initial parcel concentrations, and $\mathbf{B}$ is the covariance matrix of expected errors in $\mathbf{x}_b$. The expression $(\mathbf{x}_b - \mathbf{x}_0)^T\mathbf{B}^{-1}(\mathbf{x}_b - \mathbf{x}_0)$ is generally called the 'background term' of the cost functional, and $\mathbf{x}_b$ is called the background.

The vector $\mathbf{y}_n$ in Eq. (1) consists of all observations which are considered valid at timestep n; $\mathbf{s}_n$ is a vector of 'model equivalents' of the observations. That is, each element of $\mathbf{s}_n$ is an estimate of the corresponding element of $\mathbf{y}_n$, based, in our application of the method, on the parcel concentrations at timestep n. In the analyses we present here, $\mathbf{s}_n$ is a linear function of parcel concentrations, i.e.

$$\mathbf{s}_n = \mathbf{H}_n\mathbf{x}_n,$$

where $\mathbf{x}_n$ is the vector of concentrations of all species for all parcels at timestep n, and $\mathbf{H}_n$ is an 'observation operator'. We leave further discussion of the calculation of $\mathbf{s}_n$ until later.

The matrix $\mathbf{R}_n$ is the covariance matrix for the random errors in the term $(\mathbf{y}_n - \mathbf{s}_n)$ which would be expected given a perfect analysis. That is, $\mathbf{R}_n$ accounts for the random errors in the observations and the 'representativeness errors' (Lorenc 1986) introduced in simulating the observations.

The strong constraints of the model equations are incorporated into the analysis by regarding $\mathcal{J}$ as a function of the initial concentrations only—i.e. as a function of $\mathbf{x}_0$. Concentrations at subsequent times are determined by integrating the model equations forward in time. This procedure produces two major simplifications. First, it replaces a constrained minimization problem with an unconstrained problem. (Numerical algorithms for unconstrained minimization are considerably more efficient and less prone to problems of ill-conditioning than are algorithms for constrained minimization.) Second, the number of independent variables is reduced by a factor of $N + 1$.

The analysis scheme uses a descent algorithm to produce a convergent sequence of estimates of the vector $\mathbf{x}_0$ which minimizes the cost functional. The algorithm requires the calculation of the gradient of the cost functional with respect to $\mathbf{x}_0$. This is evaluated by integrating the adjoint of the tangent linear equations for the model. These equations may be derived in a number of ways. Talagrand and Courtier (1987) derived the equations using the theory of adjoint operators. A derivation in terms of Lagrange multipliers is also possible (Daley 1991). We present the following derivation primarily to clarify our discussion in section 4.

Define the functional $\mathcal{J}_m$ in the form

$$\mathcal{J}_m = \frac{1}{2}\sum_{n=m}^{N}(\mathbf{y}_n - \mathbf{s}_n)^T\mathbf{R}_n^{-1}(\mathbf{y}_n - \mathbf{s}_n). \quad (2)$$

Now consider an infinitesimal variation, $\delta\mathbf{x}_0$, in the initial concentrations. At each subsequent step m of the model, there will be corresponding infinitesimal variations, $\delta\mathbf{x}_m$ and $\delta\mathcal{J}_m$ in the concentrations and in the functional $\mathcal{J}_m$.

$\mathcal{J}_m$ depends only on concentrations at step m and later. Since these concentrations are uniquely determined by the equations of the model and the concentrations for any

step $i \leqslant m$, it is legitimate to regard $\mathcal{J}_m$ as a function of the concentrations at step i only. By definition, the gradient of $\mathcal{J}_m$ with respect to the concentrations at step i satisfies the equation

$$\delta \mathcal{J}_m = (\nabla_{\mathbf{x}_i} \mathcal{J}_m)^{\mathrm{T}} \delta \mathbf{x}_i. \tag{3}$$

Suppose that $\nabla_{\mathbf{x}_m} \mathcal{J}_m$ is known. We then wish to calculate $\nabla_{\mathbf{x}_{m-1}} \mathcal{J}_{m-1}$, and hence, by induction, to calculate

$$\nabla_{\mathbf{x}_0} \mathcal{J} = \nabla_{\mathbf{x}_0} \mathcal{J}_0 + \nabla_{\mathbf{x}_0} \frac{1}{2} \left\{ (\mathbf{x}_b - \mathbf{x}_0)^{\mathrm{T}} \mathbf{B}^{-1} (\mathbf{x}_b - \mathbf{x}_0) \right\}. \tag{4}$$

From the definition of $\mathcal{J}_m$, we have

$$\nabla_{\mathbf{x}_{m-1}} \mathcal{J}_{m-1} = \nabla_{\mathbf{x}_{m-1}} \mathcal{J}_m + \frac{1}{2} \nabla_{\mathbf{x}_{m-1}} \left\{ (\mathbf{y}_{m-1} - \mathbf{s}_{m-1})^{\mathrm{T}} \mathbf{R}_{m-1}^{-1} (\mathbf{y}_{m-1} - \mathbf{s}_{m-1}) \right\}. \tag{5}$$

The ease with which the second term in Eq. (5) may be evaluated depends on the complexity of the method of simulating observations. To evaluate the first term in Eq. (5), note that for an arbitrary infinitesimal variation in $\mathbf{x}_{m-1}$, we have by Eq. (3),

$$\delta \mathcal{J}_m = (\nabla_{\mathbf{x}_{m-1}} \mathcal{J}_m)^{\mathrm{T}} \delta \mathbf{x}_{m-1}$$

and

$$\delta \mathcal{J}_m = (\nabla_{\mathbf{x}_m} \mathcal{J}_m)^{\mathrm{T}} \delta \mathbf{x}_m.$$

Hence

$$(\nabla_{\mathbf{x}_{m-1}} \mathcal{J}_m)^{\mathrm{T}} \delta \mathbf{x}_{m-1} = (\nabla_{\mathbf{x}_m} \mathcal{J}_m)^{\mathrm{T}} \delta \mathbf{x}_m. \tag{6}$$

If the equations of the model are written as

$$\mathbf{x}_m = \mathcal{M}_{m-1} (\mathbf{x}_{m-1}) \tag{7}$$

then, for infinitesimal variations,

$$\delta \mathbf{x}_m = \mathbf{M}_{m-1} \delta \mathbf{x}_{m-1} \tag{8}$$

where $\mathbf{M}_{m-1}$ is a matrix whose elements are the partial derivatives of the elements of $\mathcal{M}_{m-1}$ with respect to the elements of $\mathbf{x}_{m-1}$.

Substituting Eq. (8) into Eq. (6) gives

$$(\nabla_{\mathbf{x}_{m-1}} \mathcal{J}_m)^{\mathrm{T}} \delta \mathbf{x}_{m-1} = (\nabla_{\mathbf{x}_m} \mathcal{J}_m)^{\mathrm{T}} (\mathbf{M}_{m-1} \delta \mathbf{x}_{m-1}) \tag{9}$$

$$= (\mathbf{M}_{m-1}^{\mathrm{T}} \nabla_{\mathbf{x}_m} \mathcal{J}_m)^{\mathrm{T}} \delta \mathbf{x}_{m-1}. \tag{10}$$

Since this equation holds for arbitrary $\delta \mathbf{x}_{m-1}$, we must have

$$\nabla_{\mathbf{x}_{m-1}} \mathcal{J}_m = \mathbf{M}_{m-1}^{\mathrm{T}} \nabla_{\mathbf{x}_m} \mathcal{J}_m. \tag{11}$$

Hence, by Eq. (5),

$$\nabla_{\mathbf{x}_{m-1}} \mathcal{J}_{m-1} = \mathbf{M}_{m-1}^{\mathrm{T}} \nabla_{\mathbf{x}_m} \mathcal{J}_m + \frac{1}{2} \nabla_{\mathbf{x}_{m-1}} \left\{ (\mathbf{y}_{m-1} - \mathbf{s}_{m-1})^{\mathrm{T}} \mathbf{R}_{m-1}^{-1} (\mathbf{y}_{m-1} - \mathbf{s}_{m-1}) \right\}. \tag{12}$$

This is the adjoint tangent linear (ATL) equation. Given $\nabla_{\mathbf{x}_m} \mathcal{J}_m$ the equation allows $\nabla_{\mathbf{x}_{m-1}} \mathcal{J}_{m-1}$ and so, by induction, $\nabla_{\mathbf{x}_0} \mathcal{J}_0$, to be calculated. Once $\nabla_{\mathbf{x}_0} \mathcal{J}_0$ is known, the

required gradient of the cost functional with respect to the initial conditions is given by Eq. (4).

Starting the induction at step N we have

$$\nabla_{\mathbf{x}_N} \mathcal{J}_N = \frac{1}{2} \nabla_{\mathbf{x}_N} \left\{ (\mathbf{y}_N - \mathbf{s}_N)^{\mathrm{T}} \mathbf{R}_N^{-1} (\mathbf{y}_N - \mathbf{s}_N) \right\} .$$

The algorithm used to minimize $\mathcal{J}$ is as follows.

1. Start with an initial guess for $\mathbf{x}_0$.
2. Integrate the photochemical model to give $\mathbf{x}_n$, for $n = 1...N$.
3. Evaluate $\mathcal{J}$. If the value of $\mathcal{J}$ is small enough then STOP.
4. Iterate the ATL equations to calculate $\nabla_{\mathbf{x}_0} \mathcal{J}$.
5. Use a descent algorithm to find a better guess at $\mathbf{x}_0$ (i.e. a guess for which $\mathcal{J}$ is reduced).
6. GOTO 2

3. The chemical model

The model used in this study is a new model by Lary called AutoChem (Lary *et al.* 1995). The model is explicit and uses the adaptive-timestep Bulirsch–Stoer time integration scheme (Stoer and Bulirsch 1980) with error monitoring designed by Press *et al.* (1992) for stiff systems of equations. The integration scheme is as accurate as the often used Gear (1971) package, but faster. Photolysis rates are calculated using full spherical geometry and multiple scattering as described by Lary and Pyle (1991a,b) after Meier *et al.* (1982) and Anderson (1983). For this study, photolysis rates were updated every fifteen minutes. A single ozone profile, representative of northern hemisphere middle latitudes during winter, was used for all photolysis calculations.

An extremely useful feature of the model is the existence of a code-generation program which automates the process of writing numerical chemical models. Given a set of reaction data-bases for bimolecular, trimolecular, photolysis and heterogeneous reactions the program automatically writes the Fortran code for calculating the time derivatives and the Jacobian matrix required by the numerical integration program. This enables new reaction schemes, covering different chemical species or rates, or both, to be implemented without the need for manual coding. The user specifies which reactants and products should be included. Reactions which are just upper-limit estimates, with unknown products of the required species, or reactions which are endothermic by more than a given amount can be automatically excluded if required.

Our analysis method integrates the adjoint of the tangent linear equations for the model in addition to integrating the model. The number of code changes required to implement new reactions in the analysis scheme is about double the number required to implement the reactions in the model alone. By modifying the code generation program to write code automatically for the adjoint tangent linear model, we have extended to the analysis scheme the ability to implement new reactions without manual coding. We intend to exploit this ability in future studies to determine the effect on the accuracy of the analysis of including or excluding various species and reactions.

For this study, a simple reaction scheme was chosen deliberately to provide a better understanding of the characteristics of the analysis method. Table 1 lists the reactions and rate coefficients used. Six species are integrated, namely: $O(^3P)$, O_3, NO, NO_2, NO_3 and N_2O_5. Seven bimolecular reactions, five trimolecular reactions and seven photolysis reactions are included. The rate constants for the reactions were taken from DeMore *et al.*

(1992). Although simple, the reaction scheme is sufficiently complete to describe the main photochemical processes influencing the integrated species over periods of a few days.

TABLE 1. THE REACTION SCHEME USED IN THIS STUDY. THE RATES OF ALL THESE REACTIONS ARE TAKEN FROM DEMORE *et al.* (1992)

$O(^3P)$	+	O_3	$\longrightarrow$	O_2	+	O_2			k_{b1}
$O(^3P)$	+	NO_2	$\longrightarrow$	NO	+	O_2			k_{b2}
$O(^3P)$	+	NO_3	$\longrightarrow$	O_2	+	NO_2			k_{b3}
O_3	+	NO	$\longrightarrow$	NO_2	+	O_2			k_{b4}
NO	+	NO_3	$\longrightarrow$	NO_2	+	NO_2			k_{b5}
O_3	+	NO_2	$\longrightarrow$	NO_3	+	O_2			k_{b6}
NO_2	+	NO_3	$\longrightarrow$	NO	+	NO_2	+	O_2	k_{b7}
$O(^3P)$	+	O_2	$\xrightarrow{M}$	O_3					k_{t1}
$O(^3P)$	+	NO	$\xrightarrow{M}$	NO_2					k_{t2}
$O(^3P)$	+	NO_2	$\xrightarrow{M}$	NO_3					k_{t3}
NO_2	+	NO_3	$\xrightarrow{M}$	N_2O_5					k_{t4}
N_2O_5			$\xrightarrow{M}$	NO_2	+	NO_3			k_{t5}
O_2	+	$h\nu$	$\longrightarrow$	$O(^3P)$	+	$O(^3P)$			j_1
O_3	+	$h\nu$	$\longrightarrow$	O_2	+	$O(^3P)$			j_2
NO_2	+	$h\nu$	$\longrightarrow$	NO	+	$O(^3P)$			j_3
NO_3	+	$h\nu$	$\longrightarrow$	NO	+	O_2			j_4
NO_3	+	$h\nu$	$\longrightarrow$	NO_2	+	$O(^3P)$			j_5
N_2O_5	+	$h\nu$	$\longrightarrow$	NO_3	+	NO_2			j_6
N_2O_5	+	$h\nu$	$\longrightarrow$	NO_3	+	NO	+	$O(^3P)$	j_7

4. IDEALIZED ANALYSIS FOR A SINGLE PARTICLE

In this section we consider a single parcel which, it is supposed, remains at a constant longitude, latitude, pressure and temperature. The solid lines in Fig. 1 show the evolution over a period of 36 hours of species concentrations for a parcel at position 0°E, 43.7°N and pressure 5 hPa. The temperature of the parcel was kept at 228 K throughout the integration. The model was initialized at 12 UTC, and initial concentrations were chosen to be typical of air at this latitude and pressure. For brevity, we refer to this integration as integration A.

A second integration, B, was performed, marked by the dashed lines in Fig. 1. For this integration, the initial concentrations of all species were double those for integration A.

To test the analysis procedure, the concentrations of O_3 and NO_2 at each model timestep were extracted from integration A and used as observations. These species correspond to those used in the analysis of UARS data referred to in section 6, and represent the species for which accurate measurements are currently available. The observation operators, $\mathbf{H}_n$, for this analysis were simply 6×2 matrices of ones and zeros which picked out the O_3 and NO_2 values from the vector of model values at each timestep. The covariance matrices, $\mathbf{R}_n$, were specified as constant and diagonal with elements proportional to the square of the initial concentrations of O_3 and NO_2. The initial concentrations of integration B were used for the first guess. The background term was not included in the cost function. This ensured that any success in analysing the correct concentrations was due to the influence of the observations.

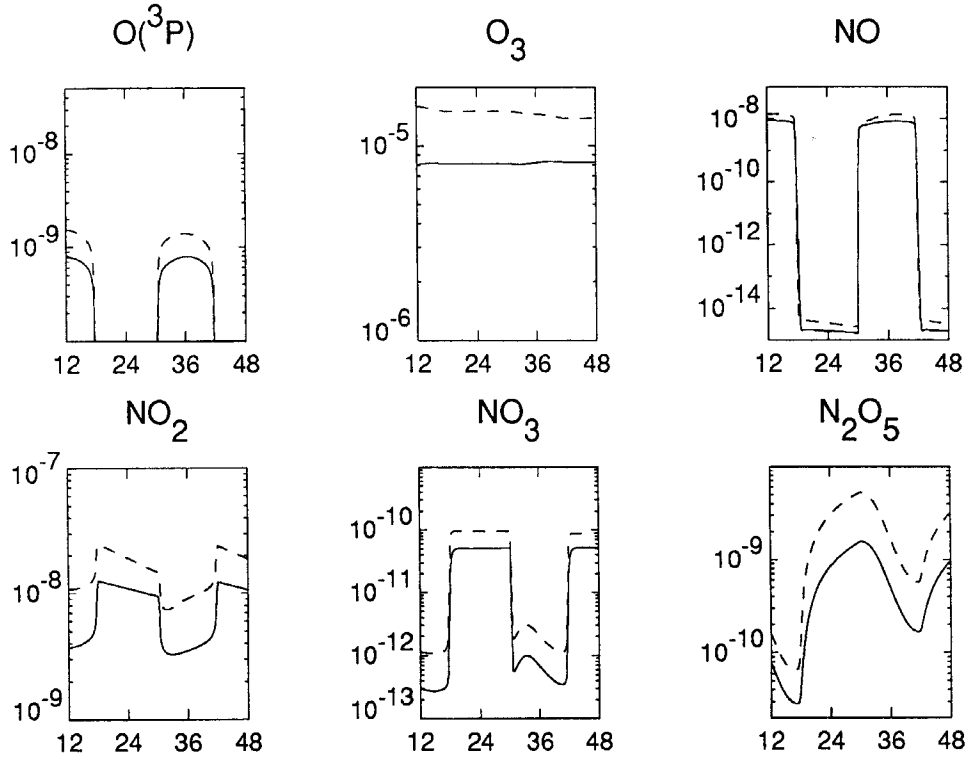

Figure 1. Model integrations for a single, stationary parcel. Solid lines show integration A, dashed lines show integration B. In each plot, volume mixing ratio is plotted as a function of time in hours past midnight on the initial day. Note the different scales for the ordinates.

The minimization algorithm used for the analysis was the Polak–Ribière (Polak 1971) variant of the conjugate gradient descent algorithm. To avoid problems due to negative concentrations, the minimization was performed with respect to the logarithms of the initial concentrations. This also acts as a simple preconditioning by non-dimensionalizing the control variables. The approximate line minimizations required at each iteration of the algorithm were performed by fitting a quadratic function to the value of the cost functional and the component of its gradient along the descent direction, and to the value of the cost functional at a single trial point. Evaluating the cost functional at the minimum of the quadratic was found to achieve adequate descent for most iterations and requires just two evaluations of the cost functional and one evaluation of its gradient. Ten iterations of the analysis procedure were found to be sufficient to achieve good convergence.

Figure 2 shows the relative differences in percentage between an integration using the analysed initial concentrations and integration A. Analysed values of the two observed species are well within 1% of the 'observations' throughout the analysis period. In addition the analysis procedure has accurately analysed the unobserved species, with the exception of the first six hours when N_2O_5 and NO_3 are poorly analysed, and for short periods just after sunset when the NO concentration is in error by about 4%.

The relative difference between the true (i.e. integration A) and analysed initial concentrations of N_2O_5 is 100%, indicating that the analysis has not modified the first-guess concentration. There is no observation of N_2O_5 to constrain the initial concentration and the decay of N_2O_5 into NO_2 (the primary way in which N_2O_5 interacts with the other species

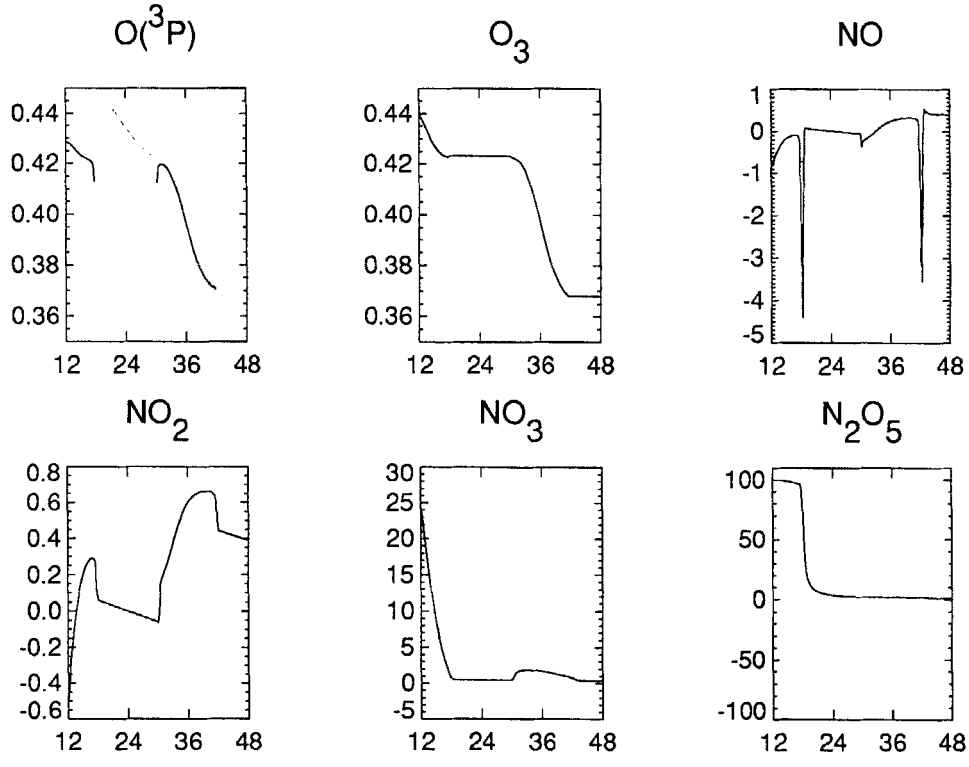

Figure 2. Relative differences in percentage plotted as a function of time in hours past midnight between true and analysed concentrations for an analysis using observations of O_3 and NO_2 only. Errors for $O(^3P)$ are plotted only when the volume mixing ratio of $O(^3P)$ is greater than 10^{-20}.

of the model during the day) is on the timescale of hours. The concentration of N_2O_5 at the initial time (noon) is small. As soon as the sun sets, N_2O_5 production commences and the concentration starts to increase. The rate of production, via reaction k_{b6} followed by k_{t4}, is dependent on the O_3 concentration and on the temperature. Since the temperature is known and the O_3 concentration is accurately analysed, the rate of production of N_2O_5 is accurately determined and the relative difference between the true and analysed N_2O_5 concentrations decreases.

The relative difference between the true and analysed initial concentrations of NO_3 is about 25%, however the daytime concentration of NO_3 is small. During the day, NO_3 has a photochemical time constant of a few seconds and is therefore in photochemical equilibrium. The main sources of NO_3 are photolysis of N_2O_5, the reaction of ozone with NO_2 and the reaction of $O(^3P)$ with NO_2. At the start of the integration the N_2O_5 concentration is relatively large. Consequently, N_2O_5 is an important factor in determining the concentration of NO_3. The large initial error in NO_3 reflects the large error in N_2O_5. During the first few hours of integration, the concentration of N_2O_5 decreases markedly while that of NO_2 increases. As a result, the importance of N_2O_5 in determining the NO_3 concentration diminishes, as does the error in NO_3.

Close to sunset, the destruction of NO is faster than its production. The sharp peaks in differences between true and analysed NO reflect the more rapid loss of NO due to the overestimate of O_3 in the analysis. This is supported by the fact that the second peak, when the O_3 concentration is closer to the true value, is of smaller magnitude than the first.

The 'observations' used to produce the analysis presented above were exact. Observations derived from satellite data, on the other hand, contain both random and systematic errors. To determine the sensitivity of the analysed concentrations to errors in the 'observed' concentrations, a series of analyses was done for which the observed concentrations were modified. Three analyses were done to determine the sensitivity to systematic errors in the observations. For these analyses, the observations of O_3 ,NO_2 or of both species were increased by 10%. The relative changes induced in the analysed concentrations are shown, respectively, by the dotted, dashed and dot-dashed lines in Fig. 3. In all three cases, changes in analysed concentrations are less than 20%, except at sunset when changes in the concentration of NO briefly reach around 45%. The increase in analysed ozone concentration, due to a systematic increase in the O_3 observations, is less than the increase in the observations, indicating that the analysed ozone concentration is constrained by information extracted from the observations of NO_2. It is likely that this information derives from the diurnal cycle of NO_2.

Although not a conclusive test, the results of these analyses suggest that, with this chemical scheme, the relative biases in the analyses of both the observed and the unobserved species are of the same order as the systematic biases in the observations.

A further analysis was carried out to test the sensitivity of the analysed concentrations to random noise in the 'observed' concentrations. For this analysis, a random amount was

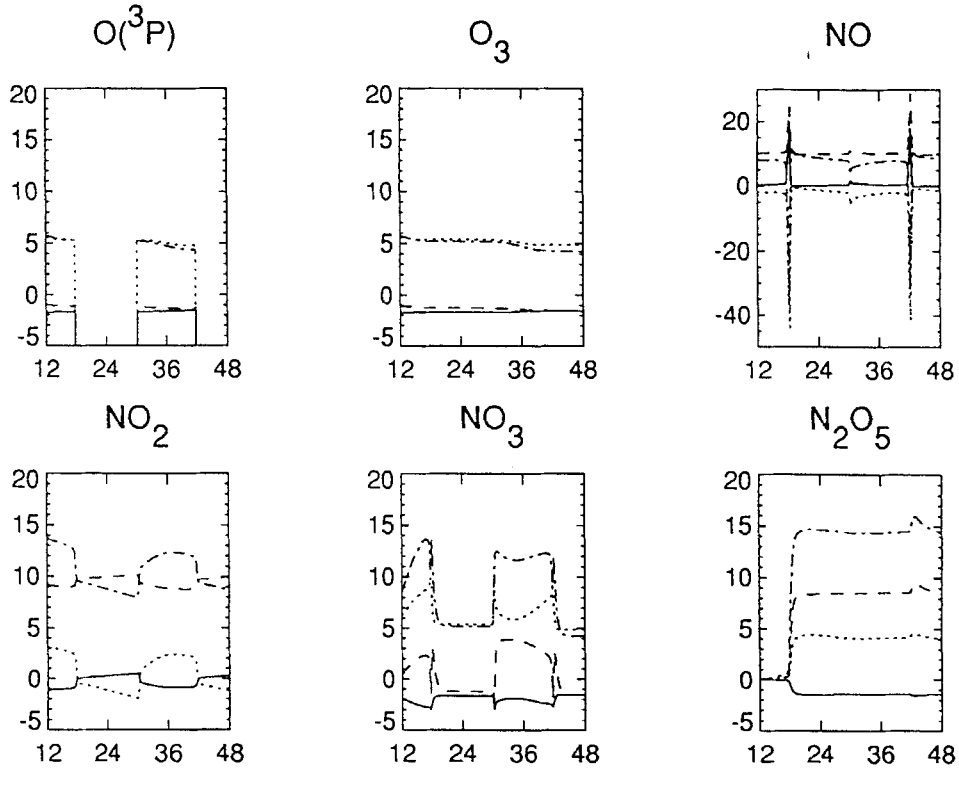

Figure 3. Relative differences between analysed concentrations using perturbed observations and an analysis using unperturbed observations. For three analyses, observations of O_3 (dotted lines), NO_2 (dashed lines) or of both species (dot-dashed lines) were increased by 10% throughout the analysis period. For the fourth analysis (solid lines) random noise was added to the observations. Differences are plotted in percentage as a function of time in hours after midnight.

added to each observation of O_3 and NO_2. The random perturbation to each species was uniformly distributed with a zero mean value and a maximum absolute value equal to 10% of the initial concentration. The relative differences between the analyses using noisy and unperturbed observations are shown by the solid lines in Fig. 3. The analysis has clearly filtered the noise from the observations.

5. Influence of observations

The analyses presented in section 4 illustrate one major advantage of the variational analysis scheme presented in this paper, namely that observations of one species can convey useful information about other species. The propagation of information from species to species results from the non-diagonal nature of the matrices $\mathbf{M}_m$ in the adjoint tangent linear equations. Since these equations are linear, the effect of each observation on the gradient of the functionals $\mathcal{J}_m$ is additive and it is possible to examine the propagation of information between species by examining the influence of single observations.

Equation 3 relates infinitesimal changes in $\mathbf{x}_i$ to the corresponding changes in $\mathcal{J}_m$ via its gradient. The term $\nabla_{\mathbf{x}_m} \mathcal{J}_m$ may therefore be interpreted as indicating the sensitivity of $\mathcal{J}_m$ to small changes in the concentrations at step m (Rabier *et al.* 1993).

Consider the case in which there is a single observation at step $n \geqslant m$. Then $\mathcal{J}_m = \mathcal{J}$ and $\nabla_{\mathbf{x}_m} \mathcal{J}_m$ indicates the sensitivity of the cost functional to changes in $\mathbf{x}_m$. However, the information provided by the gradient of the cost function requires careful interpretation. In particular, the photochemical balances between species in the model imply that not all changes to $\mathbf{x}_m$ are possible. For example, it is not possible to modify the initial conditions of the model to achieve a change in $\mathbf{x}_m$ for which O_3 and $O(^3P)$ are not in photochemical equilibrium. Further, in a full application of the analysis method, the possible changes to $\mathbf{x}_m$ will be constrained by the requirement to fit observations at earlier steps. In view of these difficulties, we prefer to interpret $\nabla_{\mathbf{x}_m} \mathcal{J}_m$ as indicating the sensitivity of the cost functional to changes in the *initial* concentrations for an analysis for which the step m is the initial step.

In order that species may be compared, whose concentrations differ by orders of magnitude, it is convenient to consider the effect of relative rather than absolute changes in concentration. To do this, note that

$$\delta \mathcal{J}_m = \left\{ \left(\nabla_{\mathbf{x}_m} \mathcal{J}_m \right)_i (\mathbf{x}_m)_i \right\} \frac{(\delta_{\mathbf{x}_m})_i}{(\mathbf{x}_m)_i}, \tag{13}$$

where the notation $(\cdot)_j$ denotes the jth element of a vector. Thus $\left(\nabla_{\mathbf{x}_m} \mathcal{J}_m \right)_i (\mathbf{x}_m)_i$ measures the sensitivity of the cost functional to small relative changes in concentration for species i.

We define an 'influence function', $\gamma_{i,j,m,n}$, as

$$\gamma_{i,j,m,n} = \frac{\left(\nabla_{\mathbf{x}_m} \mathcal{J}_m \right)_i (\mathbf{x}_m)_i}{\left(\nabla_{\mathbf{x}_n} \mathcal{J}_n \right)_j (\mathbf{x}_n)_j}.$$

The denominator provides a normalization, so that for $m = n$ the value of the influence function is unity for $i = j$. (For $m = n$ and $i \neq j$, the influence function is zero.) The value of the influence function for species i indicates the sensitivity of the fit to the observation to small relative changes in initial species concentrations. A large absolute

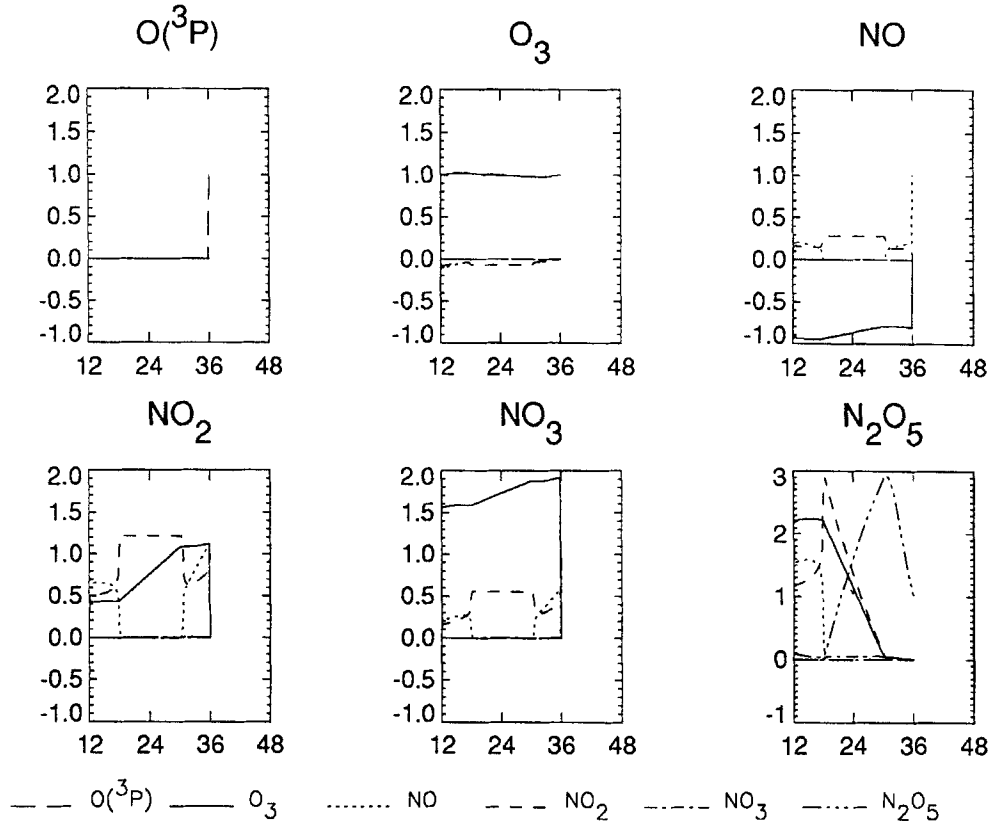

Figure 4. Graphs of the influence function, $\gamma_{i,j,m,n}$. Each panel shows the influence function values as a function of time in hours after midnight resulting from a single observation of one species at noon. Note the different scaling for the ordinate in the graph for N_2O_5.

value of the influence function for species i due to an observation of species j indicates that observations of species j play an important part in determining the analysed initial values for species i.

Figure 4 shows the influence function evaluated for integration A. Each plot corresponds to a different observed species. In each case, a single observation was made at local noon (36 hours on the time axis). Values of the influence function for $O(^3P)$ are not shown since they are negligible for observations of all species; except, by definition, for an observation of $O(^3P)$ at the time of the observation. This is because changing the initial concentration of $O(^3P)$ has negligible effect on the subsequent course of the integration. $O(^3P)$ equilibrates rapidly, but in doing so does not significantly affect the concentrations of other species. Note however that, although the analysis scheme is unable to determine an initial value for $O(^3P)$, its concentration at all subsequent times will be accurately determined through photochemical equilibrium with O_3. All species in the model except $O(^3P)$ have some influence on the analysed initial values of one or more unobserved species, showing that these observations contain information about unobserved species.

The influence function for an observation of O_3 shows that the main factor determining the degree to which the integration fits the observation is the initial value of O_3. In the photochemical scheme used, the time constant for ozone at this level is quite long. Consequently, the concentrations of other species have a relatively small role in determining

the ozone concentration during the integration. A small sensitivity to the concentrations of NO and NO_2 is reflected in the values of the influence function for these species.

The graphs of the influence function for observations of NO and NO_2 show a number of interesting features. The partitioning of NO_x is controlled primarily by the ozone concentration. This is reflected in the large values of the influence function for ozone. During the day NO and NO_2 are, to good approximation, in photochemical equilibrium. A change to the initial value of either species will quickly equilibrate, producing a change in the total NO_x concentration. Since the concentrations of the two species are comparable, a particular relative change in the initial concentration of either species will have a similar effect on the cost functional. As a result, the influence function values for both species are similar during the day. At night NO is rapidly converted to NO_2. Even a large relative change in the night-time NO concentration will produce a negligible change in NO_x. During the following day, therefore, the concentration of NO is insensitive to small changes in the night-time NO concentration, but depends strongly on the concentrations of NO_2 and O_3. The large influence of NO_2 measurements on the concentration of ozone explains why, in Fig. 3, unbiased observations of NO_2 were able to reduce the bias in the analysed ozone concentration significantly.

The sensitivity of NO_3 to initial ozone, NO and NO_2 concentrations is similar to that of NO and NO_2. However, neither NO nor NO_2 is sensitive to the initial concentration of NO_3.

The values of the influence function for an observation of N_2O_5 are high, with marked diurnal cycles for most species. As well as the diurnal cycle, there is an increase in the values of the influence function as the period between the initial time and the time of the observation increases. This is because N_2O_5 is a reservoir species with a time constant of hours, but whose concentration is influenced in the long term by the concentrations of most other species in the model. It is likely that the usefulness of observations of N_2O_5 and other reservoir species for 4D-Var will depend on the length of the analysis period.

Figure 5 shows the influence functions for observations made at midnight (48 hours on the time axis). The influence function for a night-time ozone observation is similar to that for a daytime observation and reflects the long photochemical time constant for O_3 at this level. The influence functions for night-time observations of NO and N_2O_5 are also similar to those for daytime observations.

A night-time observation of $O(^3P)$ produces a very large value of the influence function for several species, in particular for ozone. However, it should be noted that the concentration of $O(^3P)$ during the night is effectively zero. Useful measurements of $O(^3P)$ are unlikely to be possible at night. Moreover, the night-time concentration of $O(^3P)$ in the model is dominated by numerical rounding errors, making comparison with observations impossible. The influence function for a night-time observation of NO_3 is remarkably similar to that for O_3. For both NO_3 and O_3 observations, the influence function for O_3 remains at about unity throughout the analysis period, whereas for other species it is small. The photochemical behaviour of NO_3 and O_3, however, is markedly different. The photochemical time constant for ozone is long, so that the concentration at the time of the observation is determined mainly by the concentration at the initial time. NO_3 on the other hand has a short photochemical time constant. However, the night-time concentration of NO_3 is, to a good approximation, proportional to the concentration of O_3. A change in the initial concentration of ozone therefore produces nearly identical relative changes in the concentrations of ozone and NO_3.

The partition between daytime values of NO_2 and NO is largely determined by the concentration of O_3. This is reflected in the opposite signs for the values of the influence function for ozone due to observations of NO_2 and NO at noon. Observations which

indicate that the analysed value of NO_2 should be increased, whereas that of NO should be decreased (or vice versa) would combine to have a strong effect on the analysed ozone concentrations. Thus the analysis method is able to extract useful information from the partitioning of species within a family.

The diurnal cycles of some species can also convey useful information about the concentrations of other species. For example, the main cause of the large difference between daytime and night-time concentrations of NO_2 is the rapid photolysis of NO_2 to produce NO at dawn, and the recombination of NO to form NO_2 at dusk. The difference between daytime and night-time observations of NO_2 therefore provides information on the partitioning between NO_2 and NO, and hence on the ozone concentration. The ability of 4D-Var to make use of such information is reflected in the opposite signs for the influence function for ozone due to noon and midnight observations of NO_2.

The influence function provides a useful tool with which to diagnose the behaviour of the analysis system. However, it is clear from the example of daytime $O(^3P)$ that, while a small value of the influence function for a species indicates an insensitivity of the cost functional to changes in initial concentration, it does not preclude accurate analysis of the species at later times. This is also illustrated by the differences between the analysed and 'true' NO_3 and N_2O_5 concentrations plotted in Fig. 2. Both species show large errors in analysed initial concentrations, as expected from the small values of influence function

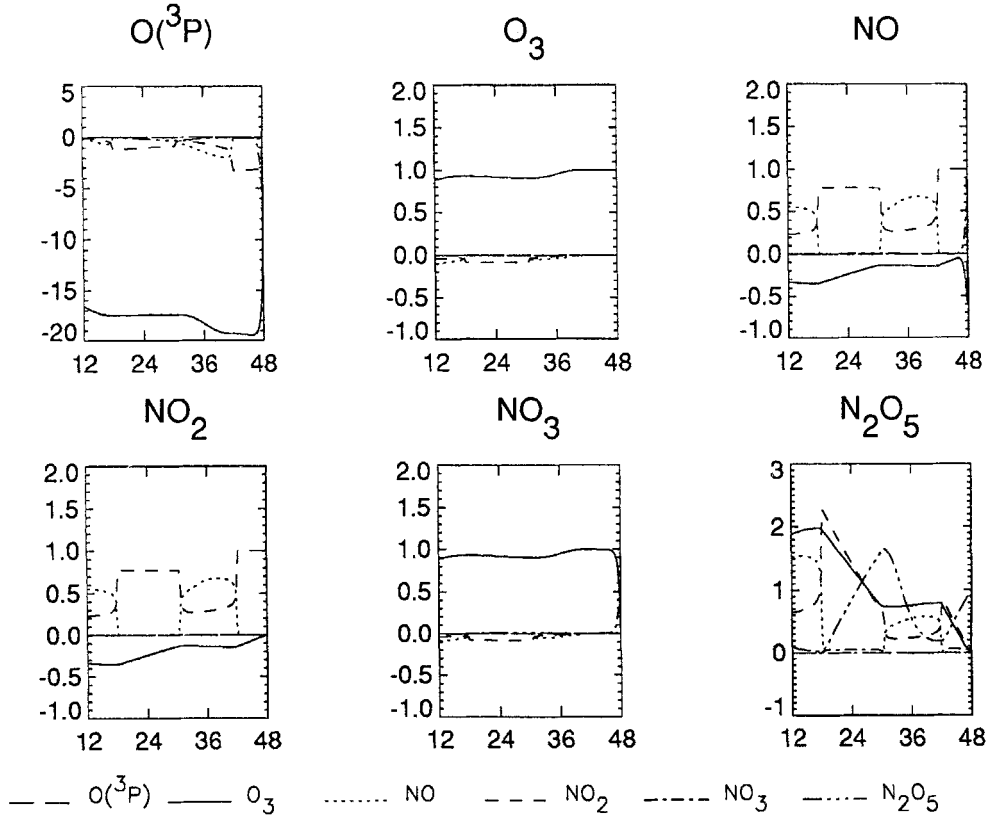

Figure 5. Graphs of the influence function as for Fig. 4, but for observations at midnight.

for these species for O_3 and NO_2 observations. Nevertheless, both species are analysed accurately for most of the analysis period.

6. Analysis of UARS data

In this section, we present a variational analysis for the 1100 K isentropic surface of observations made by instruments on board the UARS satellite. Several of the instruments measure one or more of the species in our model. We have chosen to use O_3 retrievals (version 3) from the 205 GHz channel of the MLS instrument and retrievals of NO_2 (version 6) from the CLAES instrument. Retrievals of temperature are available from these instruments and are collocated with the species retrievals. The temperature retrievals from each instrument were used to calculate concentrations on the 1100 K isentropic surface to provide the observations required by the analysis scheme. A one-hour analysis timestep was used. That is, consecutive values of n in Eq. 1 correspond to times one hour apart. At each step, observations were considered valid for use at that step if they were taken within a half hour of the model time for the step.

Isentropic trajectories for 1716 parcels were calculated using a fourth-order Runge–Kutta trajectory scheme. The scheme has been adapted from that described by Fisher *et al.* (1993) to allow advection of particles on an isentropic surface. Winds and temperatures from the UK Meteorological Office stratospheric analyses (Swinbank and O'Neill 1993) were used to perform the horizontal advection and to locate the 1100 K isentropic surface. Forward and backward trajectory calculations from 12 UTC on 10 January 1992 were performed and combined to produce trajectories covering the period 00 UTC on 9 January to 00 UTC on 11 January. At 12 UTC on 10 January the parcels were located at the points of a regular polar stereographic grid with an approximate horizontal resolution of 370 km.

Figure 6 shows winds and temperatures on the 1100 K isentropic surface for 12 UTC on 10 January 1992, calculated from the UK Meteorological Office stratospheric analyses. The pressure of the 1100 K surface varies between approximately 3.6 hPa (where the temperature is 220 K) and 8.3 hPa (where the temperature is 280 K). The polar vortex is displaced from the pole along longitude 0°E and air from low latitudes moves polewards and downwards as it is transported around the vortex and warms adiabatically as it descends. The use of observations which are not collocated with the air parcels of the model makes the calculation of the model's equivalents of the observations more complicated than for the idealized analysis described in section 4. We have used a rather crude, but simple, method to simulate the observations of mixing ratio on the isentrope. Each observation is regarded as representing, to within observational error, the average mixing ratio over a finite area. For each observation, the equivalent for the model is calculated by averaging the mixing ratios for all parcels which lie within the observation area. To the extent to which the spatial distribution of particles within each area may be regarded as random, this gives an unbiased estimate of the mean mixing ratio in the area. The variance of the estimate is simply the ratio of the variance of the sampled field to the number of particles within the area.

The observation operator requires no statistical information about the spatial correlations of the fields of mixing ratio. A particular advantage is that, for non-overlapping observation areas and an assumption of random distribution of particles within each area, the representativeness errors are spatially uncorrelated. The contributions of errors of representativeness to the covariance matrices $\mathbf{R}_n$ are therefore diagonal. For simplicity, the observation errors were also assumed to be spatially uncorrelated and observation errors for O_3 were assumed to be uncorrelated with observation errors for NO_2 . The covariance matrices $\mathbf{R}_n$ were therefore diagonal.

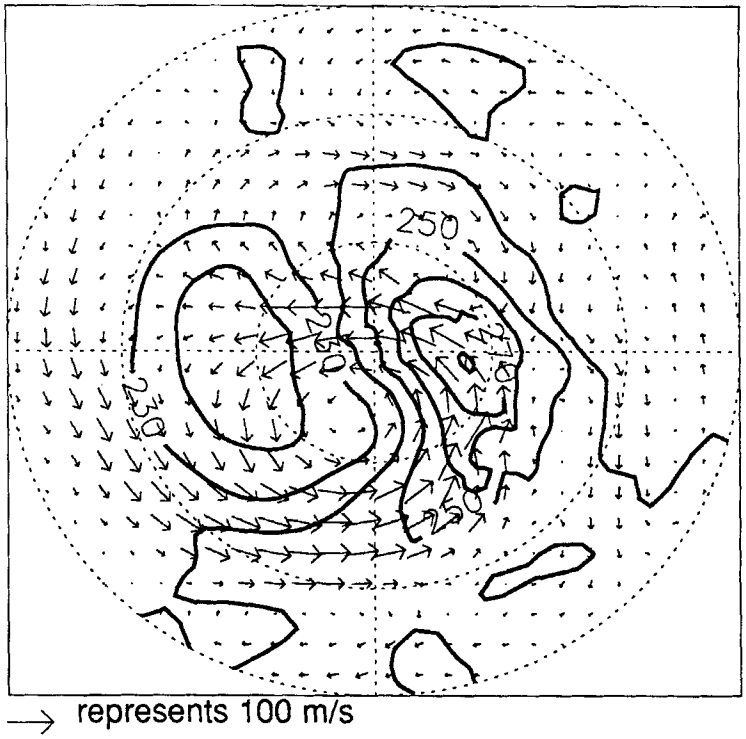

Figure 6. Winds and temperatures on the 1100 K isentropic surface for 12 UTC on 10 January 1992. Circles are drawn at 10°N, 30°N and 60°N. The Greenwich meridian extends downwards from the centre of the chart.

Each 'observation area' was defined as the region delimited by $\lambda_0 \pm \delta\lambda$ and $\phi_0 \pm \delta\phi$ where λ_0 and ϕ_0 are the nominal longitude and latitude of the observation. The values of $\delta\lambda$ and $\delta\phi$ were chosen to give an area covering 800 km in the meridional direction and 2000 km in the zonal direction at the equator. The observation areas defined in this way are considerably larger than the true averaging areas of the instruments and represent a compromise between representing the characteristics of the observations accurately and reducing sampling errors by increasing the number of particles in the observation areas.

Estimation of the observation errors is complicated both by the use of unrealistically large observation areas and by the need to account for errors in the interpolation onto the 1100 K isentrope. We did not attempt to estimate the additional errors introduced by these factors and merely assumed rather large values for the variances of observation error. Standard deviations of 4.5 ppmv for ozone measurements and 10 ppbv for NO_2 measurements were used.

The specification of the background term is complicated by the irregular positioning of the air parcels at the initial time. The number of particles is sufficiently large that explicit construction of the background-error covariance matrix, its inversion and multiplication by the vector of initial parcel concentrations are computationally expensive operations. Instead, the background term was approximated by the expression

$$(\mathbf{x}_b - \mathbf{x}_0)^T\mathbf{D}^{-1}(\mathbf{x}_b - \mathbf{x}_0) + \alpha\{\mathbf{x}_b - \mathbf{x}_0 - \mathbf{S}(\mathbf{x}_b - \mathbf{x}_0)\}^T\mathbf{D}^{-1}\{\mathbf{x}_b - \mathbf{x}_0 - \mathbf{S}(\mathbf{x}_b - \mathbf{x}_0)\}. \quad (14)$$

Here, $\mathbf{D}$ is a diagonal matrix whose non-zero elements are the variances of background error and the matrix $\mathbf{S}$ is a smoothing operator. Multiplication by $\mathbf{S}$ was implemented by

replacing each parcel concentration by a weighted average of nearby parcel concentrations for the same species. For smooth departures of $\mathbf{x}_0$ from the background, the second term in Eq. (14) is negligible, since the smoothing operator has little effect on such fields. For small-scale departures from the background, the term $\mathbf{S}(\mathbf{x}_b - \mathbf{x}_0)$ is small. Such departures are penalized by a factor $(1 + \alpha)$ more heavily than the largest scales. A particular advantage of this approximation to the background term is that, for any smoothing operator, the effective background error covariance matrix is guaranteed to be positive definite.

Values for $\mathbf{x}_b$ were calculated as follows. First, typical mid-January concentrations of all species were interpolated from a two-dimensional isentropic model (Kinnersley and Harwood 1993) onto the initial positions of the air parcels. The concentrations were then allowed to adjust to local conditions of pressure, temperature and solar illumination by integrating the chemical model used for the analysis for one diurnal cycle with the positions of the air parcels kept constant.

The minimization algorithm used for the UARS analysis was the limited memory quasi-Newton scheme M1QN3 due to Gilbert and Lemaréchal (1989). This scheme was found to be considerably more efficient than the simple conjugate gradient scheme used in section 4. Forty iterations of the analysis procedure were performed starting from first-guess initial values which were identical to the background values. Figure 7 shows the value of the cost functional, relative to the initial cost, for each iteration. Convergence is rapid for the first few iterations, after which the cost function appears to approach an asymptotic value of about 0.4 times its initial value. The relative change in cost between the 39th and 40th iterations is less than 0.1%, indicating that satisfactory convergence has been achieved.

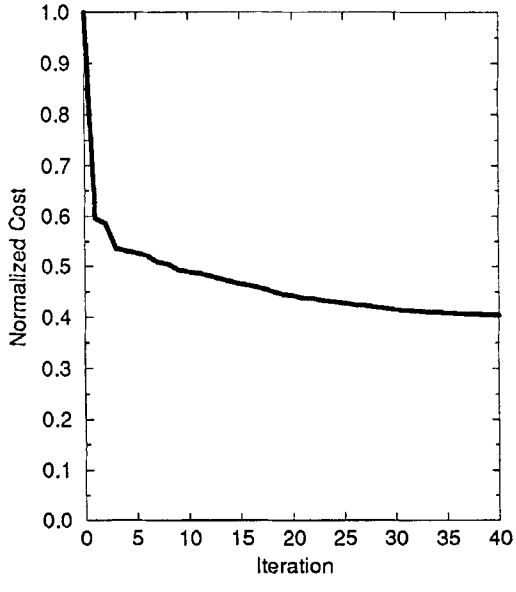

Figure 7. Convergence of the cost functional during the analysis of UARS observations.

Figure 8 shows synoptic maps of the volume mixing ratios for all species in the model for 12 UTC on 10 January 1992, taken from an integration using the analysed initial concentrations for 00 UTC on 9 January. Also shown are mixing ratios for NO_x and NO_y (defined in our restricted reaction scheme as $NO+NO_2+NO_3+2N_2O_5$). At this time, the particles are coincident with the points of a regular polar stereographic grid.

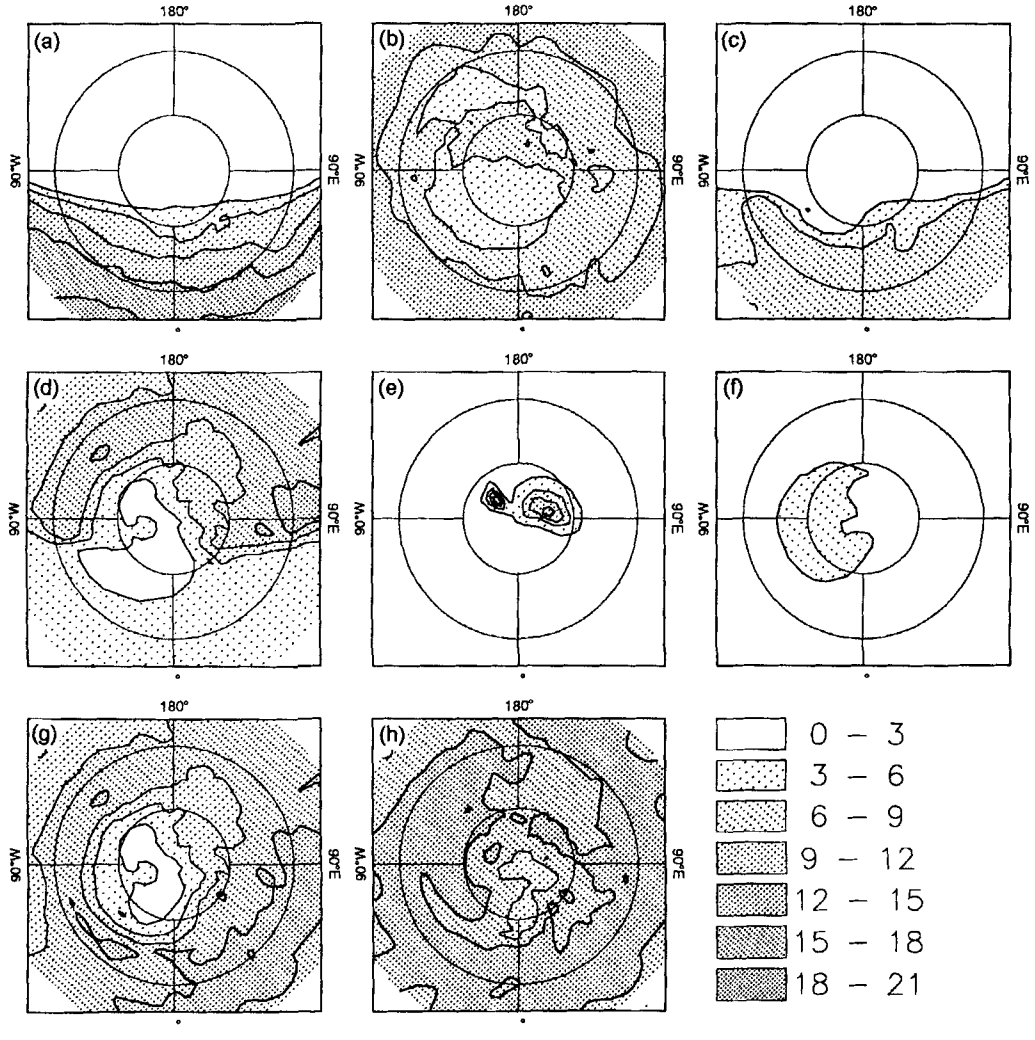

Figure 8. Analysed mixing ratios for the 1100 K isentrope for 12 UTC on 10 January 1992. (a) $O(^3P)$ (pptv), (b) O_3 (ppmv), (c) NO (ppbv), (d) NO_2 (ppbv), (e) NO_3 (pptv), (f) N_2O_5 (ppbv), (g) $NO+NO_2$ (ppbv), (h) $NO+NO_2+NO_3+2N_2O_5$ (ppbv). The circles plotted on each map represent 30°N and 60°N.

The concentrations of all the species of the model were modified by the analysis procedure. This may be seen by comparing Fig. 8 with Fig. 9, which shows mixing ratios for 12 UTC on 10 January 1992 for a model integration initialized with the first-guess concentrations on 00 UTC 9 January 1992.

It can be seen from Fig. 8 that 4D-Var has captured the diurnal cycles of $O(^3P)$, NO NO_2 and NO_3 . The terminator can clearly be seen running across the plots, the lower portions of which are in sunlight (the analysis is for 12 UTC).

The effect of the strong cross-polar flow (see Fig. 6) can be seen in several of the concentration fields. The O_3, NO_2 and NO_x minima are all shifted from the pole in the direction of the cross-polar flow. Notably, the NO_2 field has a pool of low NO_2 ,which is surrounded by the Noxon cliff (Noxon 1979), displaced in the direction of the cross-polar flow. Along the 90°E meridian, the cliff occurs at approximately 70°N whereas along the 90°W meridian it occurs at approximately 45°N.

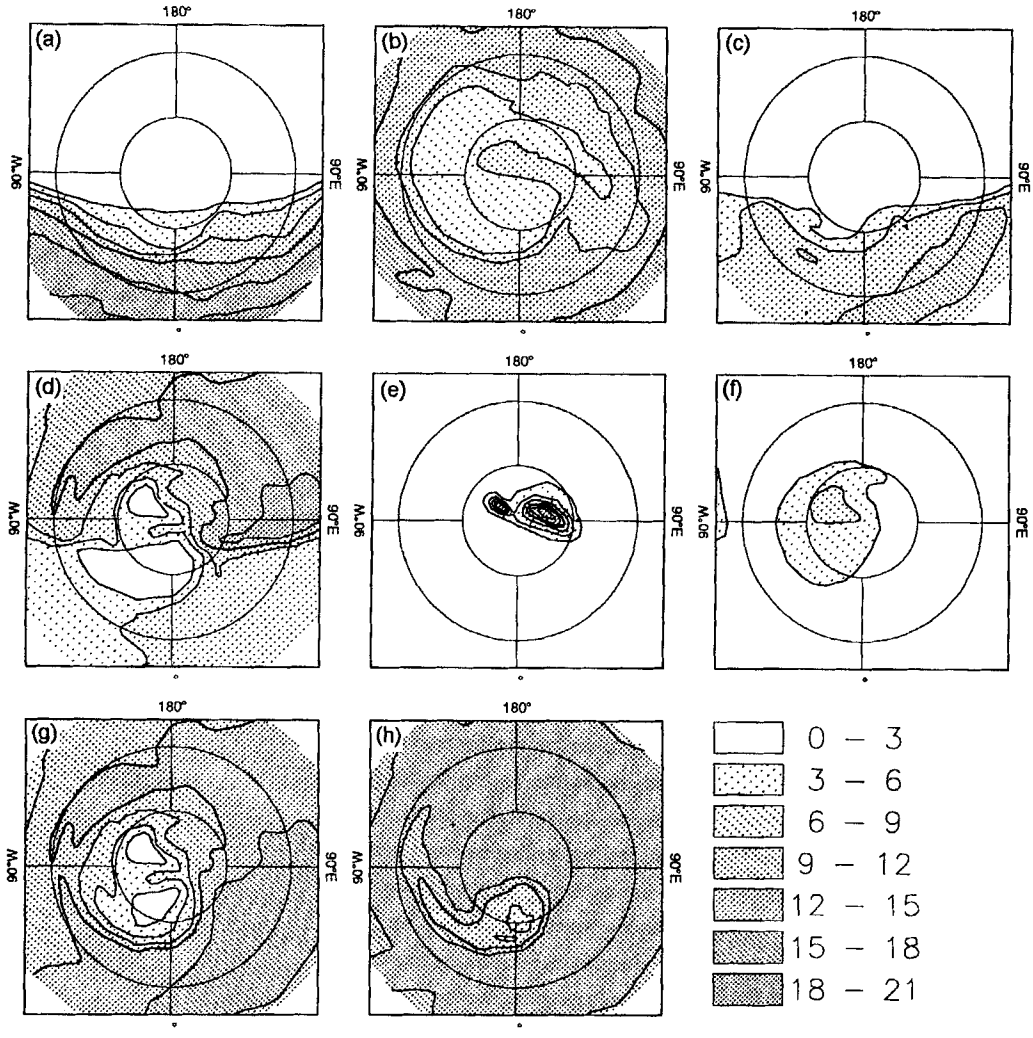

Figure 9. Mixing ratios for the 1100K isentrope for 12 UTC on 10 January 1992 for a model integration initialized with the first-guess concentrations for 00 UTC on 9 January 1992. The species and units for each panel are the same as for Fig. 8.

Advection of lower latitude air, rich in NO_y (Fig. 8(h)) enhances the magnitude of the Noxon cliff at around position 45°N, 90°W. This feature is particularly notable since the largest NO_2 concentrations and the steepest latitudinal NO_2 gradients would normally be expected just after sunset (near longitude 90°E at 12 UTC).

The analyses of the unobserved species in the model are in accordance with what would be expected, given the physical state of the atmosphere. For example, the effect of the pool of warm air centred around position 70°N, 90°E is clearly seen in the NO_3 field. Within and downwind of the warm pool the concentration of NO_3 has been considerably enhanced. As would be expected, the peak N_2O_5 concentration is in the polar night region. Since the chemical timescale of N_2O_5 is of the order of hours (as opposed to minutes for NO_2), the region of lowest N_2O_5 is rotated with respect to the terminator.

To assess the accuracy of the analysis, the analysed concentrations at the nominal time of each observation were interpolated to the locations of observations throughout

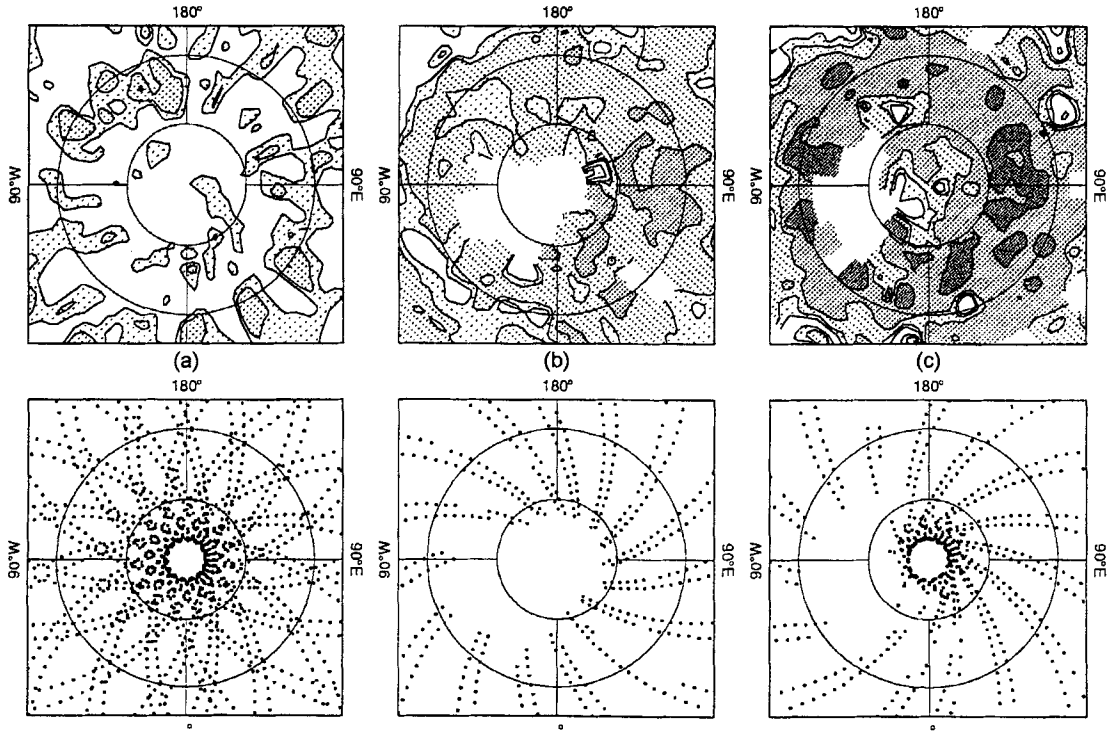

Figure 10. Differences between observed and analysed mixing ratios. (a) O_3 (ppmv), (b) daytime NO_2 (ppbv), (c) night-time NO_2(ppbv). Unshaded areas indicate differences in the range ± 0.5 units. Light, medium and dark shaded areas indicate differences in the ranges ± 1, ± 2 and more than ± 4 units, respectively. The locations of the observations used for the analysis are shown in the lower plots.

the analysis period. The differences between observed and analysed mixing ratios were then calculated. Figure 10 shows the differences for O_3, and for daytime and night-time observations of NO_2. Differences between observed and analysed O_3 have a mean value of -0.31 ppmv and a standard deviation of 0.66 ppmv. For NO_2 the mean difference is -1.12 ppbv for daytime observations and -2.19 ppbv for night-time observations. The standard deviation of NO_2 differences is 0.69 ppbv for daytime observations and 1.5 ppbv for night-time observations. The differences between observed and analysed concentrations are well within the assumed uncertainty in the measurements, suggesting that our estimated variances of observation error were too pessimistic. For ozone, the main differences occur at low latitudes. This may reflect the large observation areas used in this region. Differences between observed and analysed NO_2 concentrations are correlated with low-latitude air. This systematic error may result from the use of a simplified chemical scheme. The negative bias in the NO_2 differences is due to a consistent bias between the background and observed concentrations.

7. Discussion and conclusions

For the first time, 4D variational assimilation has been used to produce a synoptic analysis of chemical species from asynoptic satellite data. The method provides many useful insights. The analysis presented for the 1100 K surface at 12 UTC on 10 January 1992 realistically captured several interesting features, such as a displaced Noxon cliff

due to a strong cross-polar flow, enhanced NO_3 concentrations within and downstream of a warm pool of air associated with adiabatic descent, the advection of low-latitude air rich in NO_y to higher latitudes, and the sharp concentration gradients associated with the terminator for $O(^3P)$, NO, NO_2 and NO_3.

The analysis method has significant advantages over methods used hitherto. The analysis makes use of and is consistent with the temperature and wind distributions. A full diurnal cycle is produced for all species included in the model. Full use is made of asynoptic observations and the information contained in the partitioning and diurnal variation of chemical species is used in the analysis process. The ability to make use of such indirect information allows analyses to be produced for species which are not directly measured.

Although the chemical and dynamical assimilations used in this study were performed separately, they can be seen to be qualitatively consistent. Separating the analysis of chemical species from that of the dynamical variables simplifies the analysis problem considerably. The corresponding reduction in computational cost allowed all the analyses presented in this paper to be performed on a workstation. It is likely, however, that by applying 4D-Var to a combined chemical and dynamical model, useful information about the distributions of temperature and horizontal and vertical wind components could be extracted from observations of chemical species. This information would arise from the interaction of dynamics and chemistry in the model.

The gradient of the cost functional, expressed in this paper by the influence function, provides useful information about the sensitivity of the cost functional to changes in the initial conditions. However, this information must be carefully interpreted. In particular, our results demonstrate that a poor analysis of the initial concentration for a given species does not preclude accurate analysis of the species at subsequent times.

One question which we have left open is the degree to which the gradient calculated by the adjoint tangent linear model represents gradients at nearby points. Any practical minimization algorithm must use the gradient to make a finite step towards the minimum of the cost function. If the gradient is a rapidly varying function of species concentration, then only a very small step will be possible, leading to poor convergence. The requirement for a slowly varying gradient is equivalent to requiring that small, but finite, variations satisfy the tangent linear Eq. (8) to good approximation. This sets a practical limit on the length of the analysis period. The success of our analyses suggests that the tangent linear equation is valid over a 48-hour period for finite variations in at least some important directions in phase space.

Integrations presented in section 4 suggest that the systematic errors in our analysis of unobserved species are likely to be of the same order as the errors in the observed species, while random errors may be significantly reduced. A more complete estimation of analysis error should attempt to calculate at least the diagonal elements of the covariance matrix of analysis error. However, this is currently an area of active research which we consider to be outside the scope of this paper.

We recognize that improvements could be made in the statistical aspects of our analysis. We have not made use of information about the variation of observational error from observation to observation or its spatial and inter-species correlations, and we considerably overestimated the variances of the background and observational errors. The observation operator could be improved by using a better representation of the spatial averaging of the instrument, and by using information from parcels which are outside the observation area.

The model used for this paper is explicit. This type of model was used because the adjoint tangent linear equations are easily derived. However, it would be possible to derive these equations for a family model. This would significantly reduce the computational cost

of the method both through the increased speed of the model and through the reduction in the number of independent variables in the analysis problem. With these improvements, global 'real-time' analysis of chemical species in four dimensions would become feasible on current computers.

ACKNOWLEDGEMENTS

We thank the following people for their assistance: Dr J. Waters, Dr L. Froidevaux and the MLS team for the ozone data; Dr A. Roche and the CLAES team for the NO_2 data; Dr A. O'Neill for suggesting the use of trajectory methods in data assimilation; Mr A. Scaife for his work in adapting the trajectory code to isentropic coordinates, and Mr H. Maclean for calculating the trajectories. This paper was supported by grants from NERC UGAMP and from the Earth Observation Programme Board grants panel, and was in part funded by CEC STEP contracts STEP016 and EV5V-CT91-0014.

APPENDIX

The integration schemes for the chemical and ATL equations

At each timestep (typically, every 15 minutes) the integration scheme calculates $\mathbf{x}_n$ from $\mathbf{x}_{n-1}$ using several intermediate steps. The number of intermediate steps, L, is determined by monitoring the accuracy of the solution.

Schematically,

$$\mathbf{x}_n = \hat{\mathbf{x}}_L$$

where $$\hat{\mathbf{x}}_l = \hat{\mathcal{M}}_{l-1}(\hat{\mathbf{x}}_{l-1}) \quad \text{for } l = 1...L$$

and $$\hat{\mathbf{x}}_0 = \mathbf{x}_{n-1}$$

where $\hat{\mathcal{M}}_{l-1}$ represents a single intermediate step.

Each intermediate step calculates $\hat{\mathbf{x}}_l$ as a weighted sum of a number of trial values calculated by integrations with different step lengths, viz.

$$\hat{\mathbf{x}}_l = \sum_{k=1}^{K} c_k \, \tilde{\mathbf{x}}_k(\hat{\mathbf{x}}_{l-1}).$$

The coefficients c_k extrapolate the trial values to the solution for zero step length and the number, K, of trial values is determined by monitoring the convergence of the sum.

Each trial value is calculated using M_k steps of a semi-implicit midpoint method with step length h_k:

$$\tilde{\mathbf{x}}_k \equiv \mathbf{z}_{M_k} + \Delta_{M_k}$$

where $$(\mathbf{I} - h_k \, \mathbf{J}_0)\Delta_{M_k} = h_k \mathbf{f}_{M_k} - \Delta_{M_k - 1}$$

$$\mathbf{z}_{j+1} = \mathbf{z}_j + \Delta_j \quad \text{for } j = 0...M_k - 1$$

$$(\mathbf{I} - h_k \mathbf{J}_0)(\Delta_j - \Delta_{j-1}) = 2(h_k \, \mathbf{f}_j - \Delta_{j-1}) \quad \text{for } j = 1...M_k - 1$$

$$(\mathbf{I} - h_k \, \mathbf{J}_0)\Delta_0 = h_k \, \mathbf{f}_0$$

$$\mathbf{z}_0 \equiv \hat{\mathbf{x}}_{l-1}$$

$$\mathbf{f}_j = \left(\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}t}\right)_{\mathbf{x}=\mathbf{z}_j}$$

$$\mathbf{J}_0 = \left(\frac{\partial \mathbf{f}_0}{\partial \mathbf{x}_0}\right).$$

The elements of the matrix $\mathbf{M}_n$ of the ATL model are given by

$$[\mathbf{M}_n]_{i,j} = \frac{\partial(\mathbf{x}_{n+1})_i}{\partial(\mathbf{x}_n)_j}.$$

Substituting for $\mathbf{x}_{n+1}$ and $\mathbf{x}_n$ and applying the product rule for differentiation, gives

$$\mathbf{M}_n^{\mathrm{T}} = \left(\frac{\partial \hat{\mathcal{M}}_0}{\partial \hat{\mathbf{x}}_0}\right)^{\mathrm{T}} \left(\frac{\partial \hat{\mathcal{M}}_1}{\partial \hat{\mathbf{x}}_1}\right)^{\mathrm{T}} \cdots \left(\frac{\partial \hat{\mathcal{M}}_{L-2}}{\partial \hat{\mathbf{x}}_{L-2}}\right)^{\mathrm{T}} \left(\frac{\partial \hat{\mathcal{M}}_{L-1}}{\partial \hat{\mathbf{x}}_{L-1}}\right)^{\mathrm{T}}.$$

Thus a single step of the ATL model may be broken down into L intermediate steps corresponding to the intermediate steps of the forward model but in reverse order.

Note that since L is chosen adaptively, the cost functional and its gradient are discontinuous for those values of $\mathbf{x}_n$ at which the scheme decides to change the step length. This has the potential of causing problems for the minimization algorithm. On the other hand, since the integration scheme monitors the accuracy of the solution, the change in the solution resulting from a change in step size is small. In practice, we have encountered no problems resulting from the use of an adaptive step-size integration scheme.

Substituting for $\hat{\mathcal{M}}_l$ gives

$$\left(\frac{\partial \hat{\mathcal{M}}_l}{\partial \hat{\mathbf{x}}_l}\right) = \sum_{k=1}^{K} c_k \left(\frac{\partial \tilde{\mathbf{x}}_k}{\partial \hat{\mathbf{x}}_l}\right).$$

Substituting for $\tilde{\mathbf{x}}_k$ gives a set of equations for the matrix $(\partial \tilde{\mathbf{x}}_k / \partial \hat{\mathbf{x}}_l)$. Denote by $\hat{x}_i$ the ith element of $\hat{\mathbf{x}}_l$. Then the ith column of $(\partial \tilde{\mathbf{x}}_k / \partial \hat{\mathbf{x}}_l)$ is given by

$$\begin{aligned}
\frac{\partial \tilde{\mathbf{x}}_k}{\partial \hat{x}_i} &= \frac{\partial \mathbf{z}_{M_k}}{\partial \hat{x}_i} + \frac{\partial \boldsymbol{\Delta}_{M_k}}{\partial \hat{x}_i} \\
(\mathbf{I} - h_k \mathbf{J}_0)\frac{\partial \boldsymbol{\Delta}_{M_k}}{\partial \hat{x}_i} &= h_k \frac{\partial \mathbf{f}_{M_k}}{\partial \hat{x}_i} + h_k \left(\frac{\partial \mathbf{J}_0}{\partial \hat{x}_i}\right) \boldsymbol{\Delta}_{M_k} - \frac{\partial \boldsymbol{\Delta}_{M_k - 1}}{\partial \hat{x}_i} \\
\frac{\partial \mathbf{z}_{j+1}}{\partial \hat{x}_i} &= \frac{\partial \mathbf{z}_j}{\partial \hat{x}_i} + \frac{\partial \boldsymbol{\Delta}_j}{\partial \hat{x}_i} \\
(\mathbf{I} - h_k \mathbf{J}_0)\left(\frac{\partial \boldsymbol{\Delta}_j}{\partial \hat{x}_i} - \frac{\partial \boldsymbol{\Delta}_{j-1}}{\partial \hat{x}_i}\right) &= 2\left\{h_k \frac{\partial \mathbf{f}_j}{\partial \hat{x}_i} + \frac{h_k}{2}\left(\frac{\partial \mathbf{J}_0}{\partial \hat{x}_i}\right)(\boldsymbol{\Delta}_j - \boldsymbol{\Delta}_{j-1}) - \frac{\partial \boldsymbol{\Delta}_{j-1}}{\partial \hat{x}_i}\right\} \\
(\mathbf{I} - h_k \mathbf{J}_0)\frac{\partial \boldsymbol{\Delta}_0}{\partial \hat{x}_i} &= h_k \frac{\partial \mathbf{f}_0}{\partial \hat{x}_i} + h_k \left(\frac{\partial \mathbf{J}_0}{\partial \hat{x}_i}\right) \boldsymbol{\Delta}_0 \\
\frac{\partial \mathbf{z}_0}{\partial \hat{x}_i} &= \frac{\partial \hat{\mathbf{x}}_l}{\partial \hat{x}_i}.
\end{aligned}$$

The term $\partial \mathbf{f}_j / \partial \hat{x}_i$ may be evaluated using the product rule

$$\begin{aligned}
\frac{\partial \mathbf{f}_j}{\partial \hat{x}_i} &= \frac{\partial \mathbf{f}_j}{\partial \mathbf{z}_j}\frac{\partial \mathbf{z}_j}{\partial \hat{x}_i} \\
&= \mathbf{J}_j \frac{\partial \mathbf{z}_j}{\partial \hat{x}_i}.
\end{aligned}$$

This completes the description of the integration scheme for the ATL equations. The values of $\mathbf{z}_j$, $\boldsymbol{\Delta}_j$ and $\hat{\mathbf{x}}_l$ generated during the integration of the chemical model must be stored

for use later when integrating the ATL equations. In addition to these values, information must be stored on the step lengths used in the forward calculation.

The scheme requires the calculation of the matrices $\mathbf{J}_j$ and $(\partial \mathbf{J}_0 / \partial \hat{x}_i)$. Code to calculate these matrices is generated by symbolic differentiation of the chemical equations during the code generation step of the model.

References

Anderson, D. E. 1983 The troposphere to stratosphere radiation field at twilight: A spherical model, *Planet. Space Sci.*, **31**(12), 1517–1523

Andersson, E., Pailleux, J., Thépaut, J.-N., Eyre, J. R., McNally, A. P., Kelly, G. A. and Courtier, P. 1992 Use of radiances in 3D/4D variational data assimilation, Pp. 123–156 in *Proc. ECMWF workshop on variational assimilation with special emphasis on three-dimensional aspects, 9-12 November 1992*

Austin, J. 1992 Towards the four-dimensional assimilation of stratospheric chemical constituents, *J. Geophys. Res.*, **97d2**, 2569–2588

Bengtsson, L. and Gustafsson, N. 1971 An experiment in the assimilation of data in dynamical analysis. *Tellus*, **23**, 328–336

Courtier, P., Derber, J., Errico, R., Louis, J.-F., and Vukićević, T. 1993 Important literature on the use of adjoint, variational methods and the Kalman filter in meteorology. *Tellus*, **45a**, 342–357

Daley, R. 1991 *Atmospheric data analysis*, Cambridge Univ. Press

DeMore, W. B., Howard, C. J., Sander, S. P., Ravishankara, A. R., Golden, D. M., Kolb, C. E., Hampson, R. F., Molina, M. J. and Kurylo, M. J. 1992 Chemical kinetics and photochemical data for use in stratospheric modeling. Evaluation number 10. *JPL Publ.*, **92-20.**

Fisher, M., O'Neill, A. and Sutton, R. 1993 Rapid descent of mesospheric air into the stratospheric polar vortex. *Geophys. Res. Lett.*, **20**(12), 1267–1270

Gear, C. W. 1971 Chapter 9 in *Numerical initial value problems in ordinary differential equations.* Englewood Cliffs NJ: Prentice-Hall

Gilbert, J. Ch. and Lemaréchal, C. 1989 Some numerical experiments with variable storage quasi-Newton algorithms. *Math. Program.*, **45**, 407–435.

Gille, J. C. and Russell III, J. M. 1984 The limb infrared monitor of the stratosphere: experiment description, performance and results. *J. Geophys. Res.*, **89**, 5125–5140

Haggard, K. V., Marshall, B. T., Kurzeja, R. J., Remsberg, E. E. and Russell III, J. M. 1988 'Description of data on the Nimbus 7 LIMS map archive tape, water vapour and nitrogen dioxide'. NASA Tech. Pap., 2761

Kinnersley, J. S. and Harwood, R. S. 1993 An isentropic two-dimensional model with an interactive parametrization of dynamical and chemical planetary-wave fluxes. *Q. J. R. Meteorol. Soc.*, **119**, 1167–1194

Lary, D. J. and Pyle, J. A. 1991a Diffuse radiation, twilight and photochemistry—I. *J. Atmos. Chem.*, **13**, 373–392

1991b Diffuse radiation, twilight and photochemistry—II. *J. Atmos. Chem.*, **13**, 393-406

Lary, D. J., Chipperfield, M. P. and Toumi, R. 1995 The impact of the reaction $OH+ClO \rightarrow HCl+O_2$ on polar ozone photochemistry. *J. Atmos. Chem.*, **21**, 61–79

LeDimet, F. and Talagrand, O. 1986 Variational algorithms for analysis and assimilation of meteorological observations: theoretical aspects. *Tellus*, **38A**, 97–110

Lewis, J. and Derber, J. 1985 The use of adjoint equations to solve a variational adjustment problem with advective constraints. *Tellus*, **37**, 309–327

Lions, J. 1971 *Optimal control of systems governed by partial differential equations.* Springer Verlag, Berlin

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

Meier, R. R., Anderson, D. E. and Nicolet, M. 1982 The radiation field in the troposphere and stratosphere from 240 nm to 1000 nm: general analysis. *Planet. Space Sci.*, **30**, 923–933

Noxon, J. F. 1979 Stratospheric NO_2, 2. Global behaviour. *J. Geophys. Res.*, **84**, 5067–5076

Polak, E. 1971 *Computational methods in optimisation: a unified approach.* Academic Press, New York

Press, W. H., Teukolsky, S. A., Vetterling, W. T. and Flannery, B. P. 1992 *Numerical recipes in Fortran—The art of scientific computing.* Second ed. Cambridge University Press

Reber, C. A. 1993 The upper atmosphere research satellite (UARS). *Geophys. Res. Lett.*, **20**, 1215–1218

Rabier, F., Courtier, P., Herveou, M., Strauss, B. and Persson, A. 1993 Sensitivity of forecast error to initial conditions using the adjoint model. *ECMWF Tech. Memo. 197*

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

1982b Sampling theory for asynoptic satellite observations, II: Fast Fourier synoptic mapping. *J. Atmos. Sci.*, **39**, 2601–2614

1987 Irregular and diurnal variability in asynoptic measurements of stratospheric trace species. *J. Geophys. Res.*, **92**, 14 781–14 805

Stoer, J. and Bulirsch, R. 1980 Chapter 7 in *Introduction to numerical analysis.* Springer Verlag, New York

Swinbank, R. and O'Neill, A. 1993 A stratosphere-troposphere data assimilation system. *Mon. Weather Rev.*, **122**, 686–702

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