Implementation of a parallel 4D-variational chemistry data-assimilation scheme

Hendrik Elbern  
Institute for Geophysics and Meteorology, Project EURAD, University of Cologne, Cologne, Germany

Hauke Schmidt  
Institute for Geophysics and Meteorology, Project EURAD, University of Cologne, Cologne, Germany

Adolf Ebel  
Institute for Geophysics and Meteorology, Project EURAD, University of Cologne, Cologne, Germany

1. Introduction

The problem of exploiting observations scattered in time for the analysis of the state of the atmosphere has been considered in meteorology for initial value determination since the mid-1980s, e.g. Le Dimet and Talagrand (1986). While there is now a growing literature on that subject in meteorology and oceanography, applications in atmospheric chemistry data assimilation are still very rare. The feasibility of the variational calculus for the comprehensive tropospheric gas phase mechanism RADM2 (Stockwell et al., 1990) has been shown by Elbern et al. (1997).

Up to now, the parameters of interest were the initial values. However, the underlying control theory admits for further parameters to be optimized. For example, emphasis may also be placed on lateral boundary values, emission rates, and deposition velocities. The objectives of the present paper are:

1. to make a step beyond one-dimensional variational assimilation and to demonstrate the feasibility of the four-dimensional variational data assimilation technique (4D-var) for comprehensive gas phase chemistry transport models;
2. to describe the parallel implementation; and
3. to give an account of first experiences with the analysis skill and encountered preconditioning problems.

2. Model description

The EURAD CTM2 is a comprehensive tropospheric Eulerian model operating on the mesoscale – α. A full description of the European Air pollution Dispersion model (EURAD) CTM2, which is an offspring of the Regional Acid Deposition Model RADM2 (Chang et al., 1987) may be found in Hass (1991). In its standard configuration the model domain encompasses the area from the eastern North Atlantic to the Black Sea and from northern Africa to central Scandinavia. The model grid configuration is 33 × 27 × 3 in west-east, south-north, and vertical directions, respectively. The chemistry transport model calculates the transport, diffusion, and gas phase transformation of about 60 chemical species with 158 reactions. For the gas phase chemistry, a semi-implicit and quasi-steady state approximation method (QSSA) is applied for the numerical solution of the stiff ordinary differential equation system as proposed by Hesstvedt et al. (1978). Horizontal and vertical transport is simulated by a fourth order (Bott, 1989) scheme. The processes are calculated sequentially by symmetric operator splitting, when stepping from t to t + Δt. The following sequence is implemented

\[ c_i^{t+\Delta t} = T_h T_z D_z T_z T_h \circ c_i^t \]

Here T and D denote transport and diffusion operators in horizontal (h) or vertical (z) direction, respectively, which applies a fixed dynamic time step Δt of ten minutes.

Unfortunately reproduction of some figures in this article presented problems of clarity. However, they have been included as a demonstration of data presentation as described in the article.
In contrast to the solvers for advection and diffusion, the chemistry time step $\Delta t_c$ of the stiff ordinary differential equation solver is highly variable in time.

### 3. Variational data assimilation

The following cursory description of the variational calculus focuses on the practical requirements of implementation, namely the storage problem. A more comprehensive exposition of the 4D-var method may be found in Talagrand and Courtier (1987) for meteorological modeling and in Elbern et al. (1997) for an atmospheric chemistry application. A first example of adjoint sensitivity (1997) for an atmospheric chemistry application of (2).

Data assimilation procedures seek to find an initial model state which assures an optimal compliance between observations and an ensuing model integration. An objective measure to quantify the difference between measurements and model state is conveniently defined by a distance function as follows:

$$J(x(t)) = \frac{1}{2} (x_b - x(t)) B^{-1} (x_b - x(t)) + \frac{1}{2} \int_{t_0}^{t_1} (\dot{x}(t) - x(t))^2 \, dt$$

where $J$ is a scalar functional defined on the time interval $t_0 \leq t \leq t_N$ dependent on the vector valued state variable $x \in H$ with $H$ denoting a Hilbert space. The first guess or background values $x_b$ are defined at $t = t_0$, and $B$ is the covariance matrix of the estimated background error. The observations are denoted $\dot{x}$ and the observation and representativeness errors are included in the covariance matrix $O$.

Let the differential equation of the model

$$\dot{x}(t) = M(x)$$

where $M$ acts as a generally non-linear operator defining uniquely the state variable $x(t)$ at time $t$, after an initial state $x(t_0)$ is provided. The linear perturbation equation, giving the evolution of a small deviation $\delta x(t)$ from a model state $x(t)$, then reads $\frac{d\delta x}{dt} = M' \delta x$, where $M'$ is the tangent linear model of $M$.

The evolution of an initial perturbation $\delta x(t_0)$ at time $t_0$ to a perturbation state $\delta x(t_n)$ at time $t_n$ can be formally written by a sequence of stepwise integration operators $R(t_n, t_{n-1})$ as

$$\delta x(t_n) = R(t_n, t_{n-1}) R(t_{n-1}, t_{n-2}) \ldots R(t_1, t_0) \delta x(t_0)$$

with $R(t_n, t_{n-1})$ being a sufficiently accurate numerical operator for the stepwise calculation of (2).

Let $\lambda$ and $M' \lambda(t)$ be the variables and model adjoint to $\delta x$ and $M'$, respectively. It can be shown that the adjoint differential equation then reads

$$-\frac{d\lambda(t)}{dt} - M' \lambda(t) = O^{-1} (\dot{x}(t) - x(t)).$$

with the right hand side being the weighted observational forcing.

Finally, with $S$ being the adjoint analog to $R$, the gradient of the cost function $J$ is given by

$$\nabla J(x(t)) = \sum_{m=0}^{N} \dot{S}(t_0, t_1) \dot{S}(t_1, t_2) \ldots \dot{S}(t_{m-1}, t_m) O^{-1} (\dot{x}(t) - x(t)).$$

The backward integration (4) may be regarded as a sequence of linear operators, linearized around time step states and all intermediate computational steps obtained during the forward sweep of the generally non-linear forward model. For the proper construction of the linear adjoint operators these quantities must be readily available in reverse order.

The ensuing requirements of computer memory pose a principal problem in adjoint calculus of complex models. In atmospheric chemistry modelling this is even more the case, as the number of constituent state variables are one order of magnitude higher than in the case of meteorological implementations.

An assessment of storage and calculation complexity of the EURAD-CTM can be based on the following estimates in terms of orders of magnitude: spatial dimension of the grid $O(N_x \cdot N_y \cdot N_z) \approx 10^4 - 10^5$, number of constituents $O(N_c) = 10^2$, dynamic time steps $O(T_d) = 10^2$, number of operator calls per dynamic time step $O(M) = 10$, number of chemistry time steps within a dynamic time step $O(L) = 10^2$, and number of intermediate results in the code $O(L) = 10$. From runtime measurements it can be observed that the time of calculation of an adjoint piece of code $T_{ad}$ takes about twice as much as the forward code $T_{fw}$. Table 1 comprises the storage and runtime estimates in terms of $T_{fw}$ for three storage strategies.

<table>
<thead>
<tr>
<th>Storage Strategy</th>
<th>$T_{fw}$ (s)</th>
<th>$T_{ad}$ (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Operator-wise</td>
<td>100</td>
<td>200</td>
</tr>
<tr>
<td>Forward</td>
<td>50</td>
<td>100</td>
</tr>
<tr>
<td>Forward + Reverse</td>
<td>100</td>
<td>200</td>
</tr>
</tbody>
</table>

A hypothetically total storage requires only one forward integration, but is far beyond available memory and disk capacities. This also holds for operator-wise storage. Hence, the dynamic stepwise approach is adopted in this study, which requires computing time of about five forward integrations.
In the present implementation, a Cray T3E with 128 MByte local storage per processor using a horizontal equal area partitioning approach is applied (Elbern, 1997).

On the basis of these computational and storage resources, a single iteration step includes: first, a forward model run with storage of all dynamic time steps \( t_0, \ldots, t_T \) on disk. Then the backward adjoint integration from \( t_T \) to \( t_0 \) consists of the following operations for each step from \( t_l \) to \( t_{l-1} \), \( l = T, \ldots, 1 \): a forward integration from \( t_l \) to \( t_{l-1} \) with storage of all operator split intermediate states, in main memory, followed by a backward integration from \( t_{l-1} \) to \( t_l \). Each non-linear adjoint operator requires the calculation of its forward version with internal storage of intermediate states. Finally, a call of the L-BFGS minimization routine (Nocedal, 1980) within the parallel implementation completes the iteration step. By this procedure, the non-linear advection and vertical diffusion operators and the chemistry solver are integrated three times in forward direction per a single backward integration. A schematic of this procedure in a parallel application is presented in Figure 1. Fourth order Bott’s advection scheme requires a three grid point overlap for the forward mode and a six grid point overlap in the adjoint mode in the parallel implementation; as shown in Figure 2.

![Figure 1](image_url)

**Figure 1**
Storage and recalculation strategy for the parallel adjoint implementation. Left panel: storage and retrieval for the full dynamic time step states of the entire assimilation window. Right panel: computation sequence, storage and retrieval during a single dynamic time step for the intermediate operator split states. See text for further details.
4. Results of identical twin experiments

The capability of the adjoint method is taxed within the framework of identical twin experiments. With this method artificial “observations” are produced by a preceding reference model integration based on initial values to be analysed by the subsequent assimilation procedure. The skill of the 4D-var method can then be estimated by comparing the analysis with initial values withheld from the reference run. The assimilation time interval spans six hours, starting at 06.00 local time in the centre of the integration domain at midsummer conditions. The length of the assimilation interval is limited by available computing time. The wind field is defined to form two vortices, a cyclonic one in the eastern, and an anticyclonic one in the western integration domain, with logarithmically increasing wind speed aloft.

In the experiment, two circular emission areas are introduced, in the north-western and the south-eastern part of the integration domain. Only ozone observations of the lowest level are provided to the algorithm. Figure 3 exhibits a sequence of assimilation runs aiming at the acceleration of the minimization process.

The key issue of a fast convergence to the minimum is the pertinent preconditioning of the minimization problem. From a theoretical viewpoint, it is shown by Rabier and Courtier (1992) that the inverse of the Hessian matrix of the cost function (1) is the analysis error covariance matrix, which gives the desired preconditioning. However, the latter is unknown and its computation is a problem of its own (Barkmeijer et al., 1998). Nevertheless, first improvements in constructing error covariances of species relative to each other are presented in the sequel.

The top left panel displays the reference state of the surface ozone concentrations at the beginning of the assimilation interval. The emission areas of anthropogenic pollutants and the horizontal distribution of their plumes is clearly visible. The first guess state is given in the right top panel by a uniform concentration field. The next panels...
present results of four assimilation runs with 30 iterations each, which differ in a simple preconditioning factor (stretching of the co-ordinate) of the ozone components with respect to all other species. In the case without preconditioning (medium left panel) the analysis skill is quite favourable within 30 iterations. Minor degradations can be identified in the area south of the northwestern emission area. A fully satisfying analysis is obtained with factor ten (medium right panel) where nearly no differences with the “true” state can be identified. In the case of factors 100 and 1,000 (lower left and lower

Figure 3
Identical twin experiments with different preconditioning for ozone with respect to all other constituents. Displayed quantity is surface ozone. Left upper panel: reference state (“truth”) of the surface level and depicted ozone concentrations are provided. Right upper panel: first guess, medium left: only logarithmic prescaling, medium right: scaling factor 10; lower left: scaling factor 100; lower right: scaling factor 1,000.
right panel, respectively), the analysis is fully unsatisfying after the executed 30 iterations.

Another major benefit which may be obtained from proper scaling can be found at height levels where no observations were available. Information transfer from the observed surface layer to elevated height levels is only due to the vertical transport and diffusion operators. As a consequence of under-determination, adaptation to the set of observations is possible not only by modification of unobserved chemical species but also by mixing and transport downward of

**Figure 4**

Identical twin experiments with different preconditioning for ozone with respect to all other constituents. Left upper panel: reference state (“truth”) of the third level, scaled to be 230m, but depicted ozone observations are not provided, right upper panel: first guess, medium left: only logarithmic prescaling, medium right: scaling factor 10; lower left: scaling factor 100; lower right: scaling factor 1,000.
the observed tracer. Although the minimum of the cost function may be unique in a reasonably bounded domain of the phase space, an elongated extremum reflecting the poor preconditioning of the minimization problem may severely hamper the approximation to a sufficient vicinity of the true state. In the case of a proper choice of the preconditioning weights, a satisfying analysis of those height levels should be possible, where the air mass exchange with the observed level is intense enough. Figure 4 displays the analysis result for ozone at height level three, representing an elevation of about 230m. Again the superiority of weighting factor ten (panel d) compared to all other scaling factors is evident. Hence, in practice, expedient prescaling is not only a matter of saving iterations but also an issue of skilful analysis.

A complementary view on that problem is given in Figure 5, where the minimization progress is given in terms of the cost functions of the different prescaling runs. The highest convergence rate is given with prescaling factor ten. For the other cases, minimization proceeds at nearly the same rate, irrespective of a skilful analysis, as in the cases of factors 100 and 1,000. A closer look into the analyses of the \( \text{NO}_x \) concentrations reveals that, in the latter case, high \( \text{NO} \) levels reduce the surface ozone very shortly after model start and hence reduce the cost function, while the ozone concentrations at the analysis time 06.00 UTC are erroneous. In fact it turns out that only the concentrations of \( [\text{O}_3] = [\text{O}_x] + [\text{NO}_2] \) are skilfully analysed from the very start of the assimilation interval. A straightforward remedy of this problem would be a small shift to an earlier initial state, while maintaining the analysis time.

In Figure 6, time series of the assimilation run with prescaling factor ten are presented for a location in the plume of the southeastern emission area at surface grid point (20,13). Each panel includes the “true” state in terms of observations (dashed lines), first guess run (dotted lines), and the analysis results after 30 iterations (bold lines). The upper left panel displays the time series of the surface ozone, which is, one recalls, the only observed quantity. As expected, it is clearly visible that the reference curve and the analysis curve are in very close agreement. A nearly similar picture is presented in the upper right panel, where the ozone analysis for the third layer (230m) is presented. All information obtained from the surface observations is propagated upwards by vertical transport and diffusion, which suffices to analyze ozone properly. The lower panels exhibit the analyses for \( \text{NO} \) (left) and \( \text{NO}_2 \) (right) at the surface on the basis of ozone observations mentioned above. Major discrepancies are visible for the first hour of the assimilation window, where the concentration levels are significantly off the chemical equilibrium. Taking notice of the different scale in the upper left panel, it may be observed that the analysis of ozone is about 1ppbV higher than the “measurement” at 06.00 UTC. Misfits between the three surface analyses may therefore be partly interpreted as an initialization problem with correctly analyzed \( \text{O}_x \) levels, as mentioned above.

5. Conclusions

In general, the adjoint method appears to be a valuable tool for chemistry data assimilation and also in the realm of atmospheric chemistry transport modeling. However, several precautionary remarks should be kept in mind when interpreting the assimilation results. Minimization should be accelerated by preconditioning to attain a better optimization. The analysis at the starting point of the assimilation interval is not necessarily a good approximation to the true state, even if the cost function is small and the fit to observations later in the assimilation window is good. In this case, there is an undue chemical imbalance at the starting point, which relaxes towards the true state very quickly. This is the chemical analog to the “initialization problem” in meteorological...
data analysis, where fast gravity waves obstruct direct usage of the assimilation product. In the case of only ozone observations, a further pitfall is given by ambiguous analysis results of precursor species like NO\textsubscript{x} and VOCs, where various combinations may account for observed ozone variations (Sillman et al., 1990). Hence, a good first guess or a priori knowledge is required. This is the reason why a sound suite of experiments is required to form a statistical basis for reliable background estimates. Nevertheless, the ability of the 4D-var method to extract information from observations is unprecedented as compared to other methods within the limits of today’s computing resources.

References


The authors would like to thank O. Talagrand, LMD-ENS, for many helpful discussions. They also thank J. Nocedal, Northwestern University, for making the L-BFGS code available. The adjoint code of the dynamic operators was developed by use of the AMC software, version 3.88, obtained from Ralf Giering, MPI of Meteorology, Hamburg. The authors gratefully acknowledge the computational support by the Central Institute for Applied Mathematics (ZAM) of the Research Centre Jülich and the Regional Computing Centre of the University of Cologne (RRZK). Financial support was provided by the European Union’s Climate and Environmental program project RIFTUZ under grant ENV4-CT95-0025, the Troposphärenforschungsschwerpunkt of the German Ministry for Education, Science, Research, and Technology (BMBF) under grant 07TF 10/LT1-C.1, and the Ministry for Science and Education and Health.