

A NEW 4-DIMENSIONAL VARIATIONAL ASSIMILATION SYSTEM APPLIED TO ENVISAT MIPAS OBSERVATIONS

J. Schwinger¹, H. Elbern¹, and R. Botchorishvili^{2,3}

¹Rhenish Institute for Environmental Research, University of Cologne, 50923 Köln, Germany, Email: js@eurad.uni-koeln.de, he@eurad.uni-koeln.de

²Fraunhofer-Institute for Algorithms and Scientific Computing (SCAI), 53754 Sankt Augustin, Germany, Email: ramaz.botchorishvili@scai.fraunhofer.de

³Vekua Institute of Applied Mathematics, Tbilisi State University, 380043 Tbilisi, Georgia

ABSTRACT

A four-dimensional variational (4D-var) data assimilation system intended for operational application to stratospheric trace gas observations is being developed. The Kernel of this new system is a novel stratospheric global chemistry circulation model (GCCM) and its adjoint version. The German Weather Service's global forecast model (GME) with its icosahedral discretisation scheme serves as an online meteorological driver for the GCCM. As a first application of the new system, ENVISAT MIPAS data has been assimilated for a selected period. It can be demonstrated that the assimilation procedure gains a considerable improvement over legacy model runs, as the discrepancies between observations and the model are significantly reduced, while chemical consistency is maintained by the computationally costly, but efficient 4D-var technique.

Key words: variational data assimilation, ENVISAT, MIPAS.

1. INTRODUCTION

Satellite data from space borne platforms such as ENVISAT and others provide unique possibilities of global measurements of stratospheric trace gases. To make optimal use of this data, that is scattered in time and space and that originates from different sensors, a data assimilation system with the ability to produce chemical consistent synoptic maps of stratospheric constituents is needed. Advanced spatio-temporal data assimilation methods provide a powerful technique to combine observations, statistical information, and three-dimensional chemistry circulation models into an analysis which theoretically has the desired property of a "Best Linear Unbiased Estimate" (BLUE) of the stratosphere's chemical state. There are two families of algorithms, which comply with the BLUE property while making use of models:

the four-dimensional variational data assimilation techniques (4D-var) and the Kalman filter methods. In atmospheric chemistry Fisher and Lary [1] pioneered the 4D-var method in a multiple trajectory example with a reduced stratospheric chemistry mechanism. A first 4D-var system with complex tropospheric chemistry in an Eulerian chemistry transport model context was presented by Elbern and Schmidt [2, 3], where surface in situ measurements from ozone were assimilated. It could be demonstrated that the assimilation based analysis results, which involve all species of the CTM in the 4D-var context, lead to an improved ozone forecast for the following day. The first full stratospheric chemical 4D-var assimilation system was developed and applied to CRISTA data by Errera and Fonteyn [4].

In the present paper, the new data assimilation system SACADA (Synoptic Analyses for Chemical constituents with Advanced Data Assimilation) will be described. Intended for routine operation at the German Remote Sensing Data Centre (DFD), SACADA, at its final stage, is designed to comply with the following requirements:

- i optimal use of temporally and spatially scattered satellite retrievals from all types of sensors and for all available species included in a state of the art stratospheric chemistry mechanism, as well as occasionally available balloon borne measurements,
- ii maintenance of chemical consistency in terms of the chemistry mechanism applied, envisaged to result in the
- iii ability to extend the assimilation based analyses to unobserved species, which are strongly chemically coupled with observed constituents,
- iv cross validation between satellite data from different sensors and the analysis,
- v preservation of dynamical consistency, especially ensuring correctness of vertical winds, and
- vi preserving numerical efficiency for daily routine analyses to be made available and archived in near

real time, including parallel computing and portability.

The usage of either the 4D-var or Kalman filter technique enables an improvement in the knowledge of concentration levels of species which are not directly observed (item i). The consistency of dynamics (item v) will be achieved by online coupling of the meteorological driver and the chemistry transport module to a General Chemistry Circulation Model (GCCM).

The next section describes the assimilation system concepts, aiming to meet the specifications, and details the implementation issues of both the assimilation algorithm and the model numerics. In section 3 the first results are presented and the final section 4 gives an outlook to further finalisation tasks.

2. METHODOLOGY

As the SACADA data assimilation system is built from scratch, the following three fundamental system features were selected to fulfil the demands:

- The design requirements i-iv enumerated above, clearly limit the choice between the Kalman filter and the 4D-var method. In the former case, only severely complexity-reduced versions are admissible, even with future compute resources. Furthermore, for the Kalman filter, there is no guarantee that the spatio-temporal trajectory of the analysis is necessarily continuous at ingestion times of observations or that unnatural chemical imbalances are suppressed. In contrast, the 4D-var method maintains continuous system trajectories throughout the selected assimilation intervals, which is critical for archiving purposes, and furthermore facilitates to approximate concentration levels of a multitude of unobserved species. It is therefore opted to apply the 4D-var approach, although this involves a time consuming development and application of the model adjoint to the chemistry-transport module.
- As CTMs are usually driven offline by external meteorological analyses, enforcing temporal and spatial interpolation, the vertical winds are often poorly represented. To comply with requirement v, it has therefore been decided, to keep the meteorological driver model online in the assimilation system.
- Most computational burden is encountered at each grid point, due to the solution of the chemical mechanism's ordinary differential equation and its adjoint. This is an incentive for the use of parallel compute platforms. On the other hand, traditional model grid structures suffer from the poleward convergence of grid points, introducing unwanted high zonal resolution. As numerical efficiency is a key issue, the application of legacy grid structure and spectral methods is declined, in favour of an icosahedral grid. In the case of the latter, an almost isotropic grid point distribution can be preserved all

over the globe, thereby offering an avenue to satisfy efficiency demands claimed in item vi.

2.1. The 4D variational data assimilation

The basic idea of 4D-var is to minimise a scalar cost-function J , that measures the distance between a GCCM model run and the observations within a predefined time-span (also referred to as assimilation interval, and selected as 24 hours in this study) on the one hand, and an appropriate background field on the other:

$$J(\mathbf{x}_0) = J^b + J^o = \frac{1}{2} [\mathbf{x}_0 - \mathbf{x}^b]^T \mathbf{B}^{-1} [\mathbf{x}_0 - \mathbf{x}^b] + \frac{1}{2} \sum_{i=0}^N [\mathbf{H}M_i(\mathbf{x}_0) - \mathbf{y}_i]^T \mathbf{R}^{-1} [\mathbf{H}M_i(\mathbf{x}_0) - \mathbf{y}_i]. \quad (1)$$

Here \mathbf{x}_0 is the model state at $t = t_0$, \mathbf{x}^b is a background model state, usually obtained from a preceding forecast or assimilation run. \mathbf{y}_i is the vector of available observations at $t = t_i$. \mathbf{H} is a linear operator that maps from model space to the observation space while M_i is the non-linear model that integrates the initial concentrations \mathbf{x}_0 forward in time to yield the concentrations \mathbf{x}_i at $t = t_i$. For a proper weighting of the reliability of information that is contained in the observations and in the background fields, covariances of all quantities have to be specified as accurately as possible by means of the covariance matrices \mathbf{B} and \mathbf{R} (where \mathbf{R} also includes the error of representativeness of the observation for the model grid cell).

In order to find the minimum of J , the gradient of the cost-function with respect to the initial concentrations \mathbf{x}_0 is needed:

$$\nabla_{\mathbf{x}_0} J = \mathbf{B}^{-1} [\mathbf{x}_0 - \mathbf{x}^b] + \sum_{i=0}^N \mathbf{M}_i^* \mathbf{H}^T \mathbf{R}^{-1} [\mathbf{H}\mathbf{x}_i - \mathbf{y}_i]. \quad (2)$$

\mathbf{M}_i^* is the adjoint model operator that maps the gradient of the cost-function w.r.t. \mathbf{x}_i backwards in time to deliver the gradient of J w.r.t. the initial concentrations \mathbf{x}_0 . For a more detailed description of the variational assimilation technique for atmospheric chemistry applications, see [2] or [5] for example.

A possibility of an a posteriori validation of the assimilation result is given by χ^2 -evaluation of the final cost-function value J after assimilation. As demonstrated by Talagrand [6], a necessary but not sufficient condition is $J/N = 1/2$, where N is the number of available observations.

2.2. System Description

The data assimilation system consists of the meteorological driver GME, the chemistry transport model and its ad-

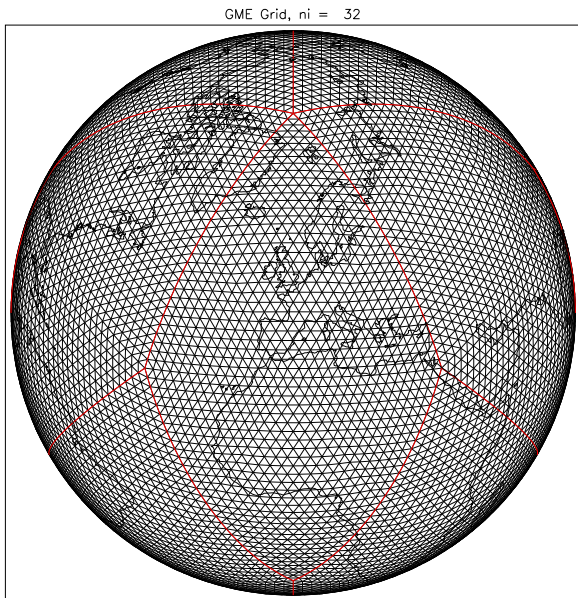


Figure 1. Icosahedral grid in a resolution of $n_i = 32$ as used for this study. The resulting mesh-size is ~ 250 km, corresponding to a resolution of $\sim 2.5^\circ$ or T80. The Borders of the ten diamonds (see text) are marked by red lines.

joint, the minimisation module and the covariance module. These components will be described with emphasis placed on the new SACADA GCCM.

2.2.1. Icosahedral Grid and Parallelisation

The new SACADA GCCM was designed for highest efficiency in order to meet the requirements of the computational demanding task of four-dimensional variational data assimilation. The icosahedral meteorological model, grid structure and the definition of operators is adopted from the German Weather Service's GME, which is described in detail in [7]. The icosahedral grid is constructed as follows: An icosahedron, i. e. the highest Platonic body with 20 equilateral triangles, is placed into a sphere with the 12 touching vertices connected to their neighbours along the sphere. The resulting great circle sections are equally subdivided into a number of n_i intervals to form an almost regular grid (see Fig. 1). This approach results in a mesh with virtually perfect constant mesh-size all over the globe, avoiding the inefficient non-uniform resolution common to conventional latitude-longitude grids. With $n_i = 32$ there are 10,242 grid-points per level and the model comprises 42 levels in the vertical, ranging from the earth's surface to 0.1 hPa. Following the focus placed on the stratosphere, the vertical resolution here and at tropopause height levels is lower than 2 km. To save CPU time, the chemistry and adjoint chemistry computations are currently limited to 16 levels from 100 hPa (~ 16 km) to 2 hPa (~ 42 km).

As each grid-point has six nearest neighbours (five neighbours at the twelve vertices of the original icosahedron), the area of representativeness is a hexagon (a pentagon at

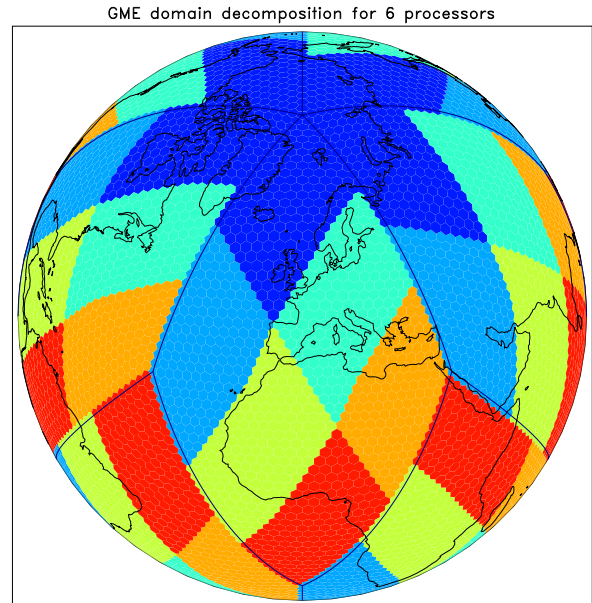


Figure 2. Domain decomposition for 6 processors: Each diamond is divided into six sub-domains. One processor handles one sub-domain per diamond.

the twelve special points). To obtain a rectangular data structure, two adjacent spherical triangles are combined to form a diamond, partitioning the grid into ten logically rectangular sub-grid domains (see Fig. 1). The meteorological driver part, as preserved from the hydrostatic GME, is kept online in the assimilation system, offering an identical spatial and temporal discretisation for both, the meteorological driver and the CTM module with its adjoint. This is esteemed as especially valuable for improved simulation of both types of vertical transports, high tropical up-drafts into the lower stratosphere, and tropopause exchanges, such as those induced by folds.

To facilitate the use of the model on parallel computers a diamond-wise domain decomposition is performed as shown in Fig. 2: Each processor works on one portion of each diamond. This is a simple yet effective strategy to achieve a good load balancing between processors. Sub-domains have a *halo* of two rows and columns of grid-points that have to be exchanged among processors each GCCM time-step using the MPI interface.

2.2.2. Chemistry Transport Module and its Adjoint

The set of reactions that are considered in the SACADA GCCM comprises 148 gas phase and 7 heterogeneous reactions on PSC (ice and NAT) and sulphate aerosol surfaces. The reader should refer to [8] for a detailed description of the reaction mechanism including the treatment of heterogeneous reactions. An evaluation of this scheme in comparison with other state of the art mechanisms has been presented by Krämer et al. [9]. The current implementation uses updated values for gas phase reaction rates taken from [10]. Furthermore, to accommodate the operation of the adjoint model, a 2nd order Rosenbrock scheme with adaptive step-size control as de-

scribed in [11] is used to solve the chemical kinetic system, with the consequence of abandoning the chemical "family concept". The KPP chemical solver tool was taken for code generation [12]. A semi-Lagrange scheme with second order interpolation has been adopted from GME to solve the horizontal transport of trace gas constituents, while the vertical transport is solved by means of a simple and efficient upwind algorithm.

The adjoint model operator M^* which is needed for the computation of the cost function's gradient $\nabla_{x_0} J$ with respect to the initial concentrations x_0 (see Eq. 2) was constructed starting from the forward code detailed above, as described in [13]. This approach alleviates the extensive and cumbersome coding work, because the task can be partly automated by using adjoint compilers like TAMC [14] and *Odyssée* [15].

In its present version, no simplifications have been introduced into neither the SACADA CTM module nor its adjoint. Namely, the inversion of the Jacobian is performed at each adaptive time step of the Rosenbrock solver.

2.2.3. Covariance Modelling and Minimisation

To specify the covariance matrices \mathbf{R} and \mathbf{B} in (1) and (2) as accurate as possible is the most critical part in the implementation of any data assimilation scheme. As far as the observation error covariances are concerned, these have to be delivered by the instrument group. Even more important especially in the case of sparse data, are the background error covariances, because the information that is contained in the observational data is spread out to neighbouring grid-points by means of \mathbf{B} . However, as the dimension of \mathbf{B} is $M \times M$, where M is the dimension of the model state vector, the storage demands, as well as the fact that the extensive statistical knowledge required is not at hand, prohibit a straightforward implementation of \mathbf{B} . Instead, an univariate background covariance operator may be modelled by means of a diffusion approach, as shown in [16]. In the framework of this study a homogeneous, isotropic, and univariate horizontal correlation was assumed with an ad hoc correlation length of 300 km. A more advanced 3-d correlation model, that is based on rigorous statistics and allows for inhomogeneous and anisotropic correlation length is currently under development.

The quasi-newton limited memory L-BFGS algorithm devised in [17] and [18] is applied for minimisation. Introducing a background error covariance matrix with non-zero off-diagonal elements, renders the minimisation problem more ill-conditioned. The minimisation is done in incremental space as described in [16], opening the possibility to use \mathbf{B} as a generally efficient preconditioner.

2.2.4. Synthesis of Assimilation System Modules

The observational data from MIPAS, enclosed by each 24 hours assimilation window, is assimilated by a single iterative assimilation procedure. The prologue is formed by

the meteorological driver model (GME) run over one day, with the meteorological data stored for each grid point and time-step. Compared to the chemistry model run and its adjoint, consumed CPU time and storage of the meteorological prologue is negligible. The kernel operation of the chemical assimilation part is, following the adjoint concept, an iterative process (running on 16 stratospheric levels for the purpose of this study), each step consisting of the

1. forward direct chemistry run with storage of the concentrations prior to the horizontal transport and prior to the chemistry calculation for each time step,
2. the adjoint backward run with recovery of the stored values, and
3. the minimisation, resulting in improved initial values for the chemical constituents.

Following experience made so far, 15 iterations suffice for a significant improvement of initial volume mixing ratio values. Wall clock time expiration for this assimilation scheme is about 7 hours on a PC-cluster consisting of 6 AMD Athlon 2 Ghz Processors. Hence, one can conclude that the demands of the 4D-var assimilation procedure in near real time operation, stay well within the limits of today's affordable computational power.

3. RESULTS

3.1. Observability Test Using Artificially Generated Data

In order to explore the potential and limits of 4D-var data assimilation applied to the particular problem of assimilating ENVISAT MIPAS and SCIAMACHY observations, a suite of so called "observability tests" was conducted with artificially generated observational data. To this end, the standard assimilation procedure was preceded by a reference model run which represents the *true state* of the atmosphere within this experiment. Volume mixing ratios from selected species at selected grid-points at different time-steps can be picked out and declared a *measurement*. Disturbed reference run initial values, i. e. reference run initial values multiplied by some factor, serve as a first guess for the assimilation scheme. Based on this experimental setup, the standard assimilation procedure as described in section 2.2.4 is started with the objective of recovering the true atmospheric state (that is the known reference state). By varying the set of selected species and/or the selection of observation time and location, the response of the assimilation scheme to a wide range of different observation scenarios can be studied.

Here we present the results of two experiments that aim to prove the promised ability of 4D-var data assimilation to analyse unobserved species at least in the idealised context of these observability tests. To this end, two data sets of "observed" species have been defined:

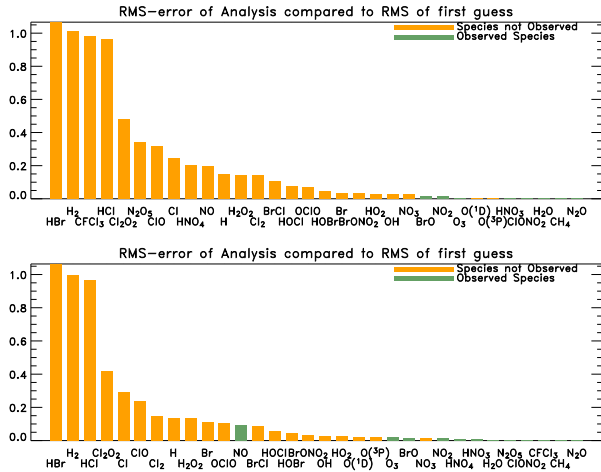


Figure 3. Results of observability tests; the analysis-run RMS error divided by first-guess run RMS (see text) is shown for each stratospheric constituent. Observed species have green coloured bars. Upper panel: Results for data set 1. Lower panel: Results for data set 2.

1. **Standard Data Set + ClONO₂ + BrO**, consisting of the operationally observed species O₃, H₂O, CH₄, N₂O, HNO₃ and NO₂ extended by ClONO₂ and BrO.
2. **Extended Data Set** comprising all species of data set 1 plus NO, HNO₄, CFC-11 and N₂O₅

Observations were assumed 8.2 hours (model time step 50) after simulation start time (00:00 UTC) and it was assumed that these observations took place at all model grid-points. However, nighttime observation of BrO and NO had to be discarded, because these trace gases almost vanish during night time in the stratospheric region of interest. The first-guess initial values have been derived from the reference run initial values by multiplying them with a factor of 1.9, resulting in a first-guess which overestimates the atmospheric concentrations of all trace gases by 90%. To quantify the success of the assimilation scheme, two RMS errors are defined:

$$\text{RMS}_{\text{fg}} = \frac{\sqrt{\sum_{k=1}^M (x_k^{\text{fg}} - x_k^{\text{rf}})^2}}{M} \quad \text{and} \quad (3)$$

$$\text{RMS}_{\text{an}} = \frac{\sqrt{\sum_{k=1}^M (x_k^{\text{an}} - x_k^{\text{rf}})^2}}{M},$$

where x_k^{fg} , x_k^{an} and x_k^{rf} are the first-guess, analysis and reference volume mixing ratio values of a trace gas under consideration at grid-point k . Thus these RMS errors measure the misfit between the atmosphere's *true state* (reference run values) and first-guess and analysis-run respectively. The results for the two experiments are summarised in Fig. 3. The analysis-run RMS error divided by first-guess run RMS, is shown for each stratospheric constituent, expressing the normalised residual error. A value of 1.0 means that no improvement has been achieved during the course of the assimilation procedure, while a value of 0.0 stands for a perfect recovery

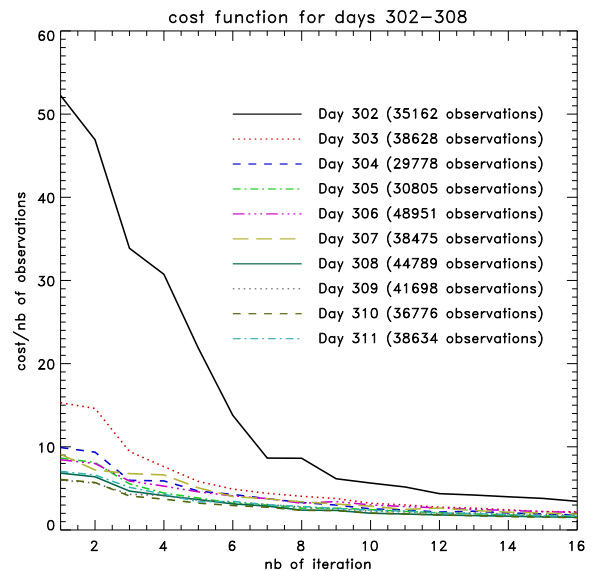
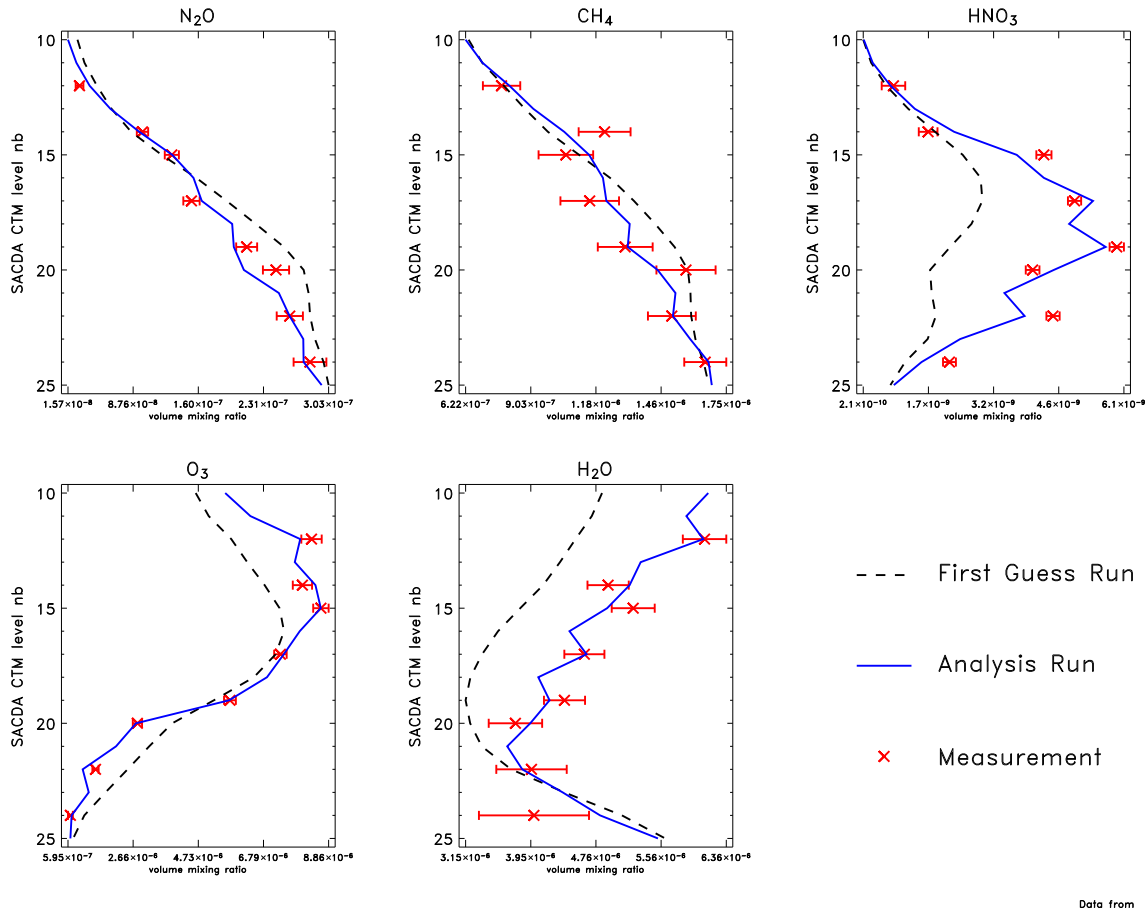


Figure 4. Evolution of the cost function during the course of the optimisation procedure; the value of the cost function normalised by the number of observations versus iteration number, is shown for all days of the assimilation period.

of the *true atmospheric state* as represented by the reference run. A considerable improvement can be claimed for many species that are not directly observed in both cases. Furthermore it is remarkable, that the standard data set enriched by ClONO₂ and BrO works almost as well as the extended data set. It can therefore be concluded, that the 4D-var method has a large potential to provide for useful estimates of a large suit of unobserved species. In real world applications, it should be observed, that a sound preconditioning is prerequisite to exploit this favourable property.

3.2. Assimilation of ENVISAT MIPAS observations

To demonstrate the performance of the new data assimilation system, a period of ten days in October/November 2003 was selected for assimilation. Operational retrievals from the MIPAS sensor in their presently latest version (version 4.61) have been used for assimilation. The set of observed species comprises the six major stratospheric trace gases O₃, H₂O, CH₄, N₂O, HNO₃ and NO₂. An initial first-guess for day 302 (Oct. 29, 2003) was produced by using output of the SOCRATES 2-D model [19] and a 144 hours spin-up model run to relax the 2-D model state towards a chemical equilibrium. Starting from this point, observational data for 24 hours was consecutively assimilated into the GCCM for days 302–311 (Oct. 29–Nov. 7, 2003), using the analysis from the previous day as a first-guess and background field. This choice of background ensures that the information that entered into the model through the assimilation of observations the day before, is retained and prior information is accumulated. If the observational data is sparsely distributed, as it is the case with ENVISAT MIPAS data, this strategy, in combination with an accurate background covariance modelling, is es-



Data from

Figure 5. Assimilated profiles for Oct. 29, 2003 (solid blue lines) and corresponding first-guess profiles (dotted black lines) together with available measurements and observational error following data specifications. Profile location is 32°N and 36°O .

essential for a meaningful global analysis of stratospheric trace gases. Note that for this first test of the new assimilation system, some coarse estimates for the background covariances have been made. The background error was set globally to 100% and the correlation of background errors between two grid-points was assumed to be quasi Gaussian (see [16]) with a correlation length of 300 km. Both these parameter values are ad hoc choices due to the present lack of adequate statistical knowledge. The observational errors are taken from the operational MIPAS data set, and covariances between measurements are not considered in this study, resulting in a diagonal error covariance matrix \mathbf{R} . A control model run (model run without any data assimilation) was accomplished for the same period of time.

In Fig. 4 the evolution of the cost function (normalised by the number of observations N) is shown for all days of the period under consideration. The almost monotonically decreasing first-guess value of J/N (iteration 1) clearly indicates that the gain of additional information that enters the model each day through the assimilation of observations is accumulated. An a posteriori inspection of the cost function value for the analysis (iteration 16), reveals decreasing values from 3.5 down to 1.5, missing the targeted 0.5 (see section 2.1) by a factor of three. A closer investigation of the variances and

covariances yields the result that the contribution of the background term in Eq. 1 to the final value of J is very small (below 2%), leading to the conclusion that the estimated background error of 100% is rather too large than too small. On the other hand the observational errors that come from the operational data sets seem to be overly optimistic: Approximately one third of all data has a relative error smaller than 5%, and two thirds has a relative error smaller than 10%. There are even a few values with a relative error below 1% in the data sets (these values are excluded from the assimilation procedure because they severely hamper the minimisation algorithm). Thus, we conclude that the deviation of J/N from the theoretically expected value 0.5 is mainly caused by over-optimistic small observational errors in the operational data.

Fig. 5 shows profiles of assimilated trace gas constituents, together with the first-guess (background) profiles and observations for the first day of the assimilation period (Oct. 29, 2003). The overall result is very satisfying: The analysed profiles meet the observations significantly better than the first-guess does. Note, that some analysed profiles, notably the HNO_3 -profile, appear to be very variable from one model level to another. This shortcoming is due to the fact that presently there are no vertical background error covariances specified. In the case of better known background error covariance statistics and

with a respective implementation of this knowledge in form of a proper 3-d \mathbf{B} -operator, the assimilation system would produce smoother profiles. As already mentioned in section 2.2.3 a sophisticated 3-d implementation of \mathbf{B} is currently under development.

Assimilation results for Nov. 7, 2003 compared with observational data and the result of the control run, i. e. the model run which was not upgraded by any data assimilation after its start 10 days before, are shown in Fig. 6. In the case of O_3 at the 7.6 hPa model level (left column of Fig. 6), the a priori model state (control run) was not too far from observations, but the information about the observed lower ozone values in the arctic region is reflected in the assimilation result only. A much more impressive improvement was achieved for HNO_3 at the 28 hPa model level (right column of Fig. 6) during the course of the assimilation procedure. Volume mixing ratios are globally too low in the a priori model state compared to the MIPAS observations, while the analysis shows no such discrepancy any more. Further examination of other height levels and species show similar favourable assimilation skills.

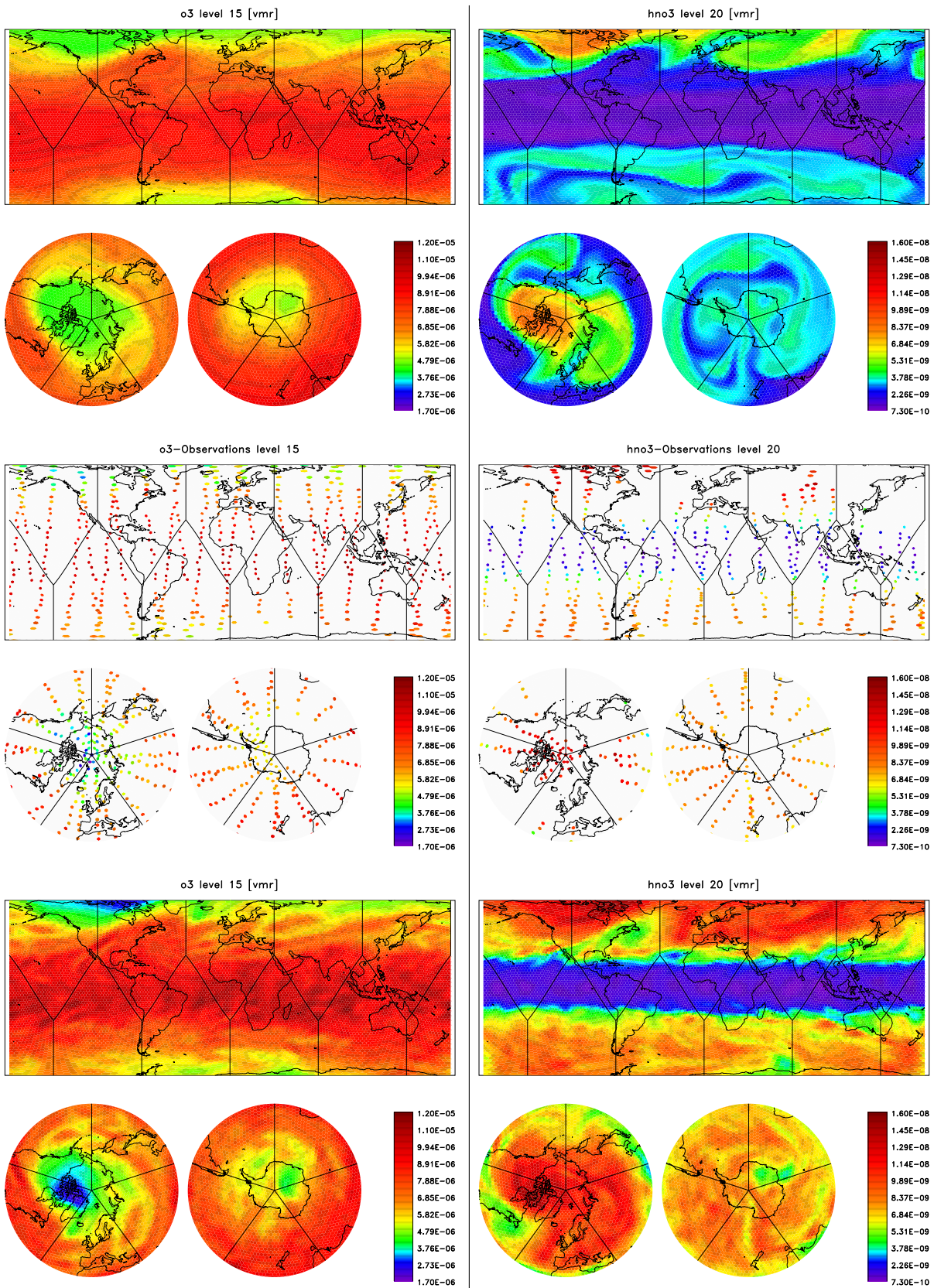


Figure 6. Results for Nov. 7, 2003; the top panels show the volume mixing ratio of O_3 at 7.6 hPa (left column) and HNO_3 at 28 hPa (right column) at 12:00 UTC for the control run (no assimilation). The ENVISAT MIPAS observations for the same day, 00:00 UTC–24:00 UTC are displayed below. The bottom panels finally shows the analysis for 12:00 UTC, obtained after consecutive assimilation of ten days' observational data.

4. SUMMARY

In total the first results obtained with the new SACADA 4D-var chemical data assimilation system demonstrate both, a fully satisfying skill, and efficiency with respect to its present state of development. First work on assimilation of more comprehensive MIPAS data and SCIAMACHY LIMB and occultation data, as well as column data, has partly begun. A thorough routine testing with the compilation of Analysis–Observation error statistics will replace the present case study based evaluation. Future work will also focus on a non-homogeneous and anisotropic covariance formulation in three dimensions. The computational challenges include the formulation of the problem as an operator, to circumvent the storage of an inhibiting large matrix. Further, an extension of the heterogeneous chemistry is planned. Progress in micro-processor technology will allow for a grid refinement, resulting in a grid resolution of about 170 km.

ACKNOWLEDGEMENTS

We are especially grateful for support from the following groups, persons and institutions: German Weather Service (DWD) for the provision of their GME software; Dr. Johannes Hendricks, for helpful advice with the stratospheric reaction scheme; Sabine Pott at SCAI-FhG for the preparation of meteorological initial data; Thilo Erbetseder at DFD-DLR for the provision of pre-processed ENVISAT MIPAS observational data. The SACADA project is funded by the German Federal Ministry of Education and Research (BMBF) in the framework of AFO-2000.

REFERENCES

1. Fisher, M. and Lary, D. Lagrangian four-dimensional variational data assimilation of chemical species. *Quart. J. Roy. Meteor. Soc.*, 121(527):1681–1704, 1995.
2. Elbern, H. and Schmidt, H. A four-dimensional variational chemistry data assimilation scheme for Eulerian chemistry transport modeling. *J. Geophys. Res.*, 104(D15):18,583–18,598, 1999.
3. Elbern, H. and Schmidt, H. Ozone Episode Analysis by Four-Dimensional Variational Chemistry Data Assimilation. *J. Geophys. Res.*, D4(106):3569 – 3590, 2001.
4. Errera, Q. and Fonteyn, D. Four-dimensional variational chemical assimilation of CRISTA stratospheric measurements. *J. Geophys. Res.*, 2001.
5. Wang, K.-Y., Lary, D. J., Shallcross, D. E., Hall, S. M., and Pyle, J. A. A review on the use of the adjoint method in four-dimensional atmospheric-chemistry data assimilation. *Quart. J. Roy. Meteor. Soc.*, 127:2181–2204, 2001.
6. Talagrand, O. A posteriori evaluation and verification of analysis and assimilation algorithms. In *Workshop on Diagnosis of Data Assimilation Systems*. European Centre for Medium-range Weather Forecasts Reading, England, 2-4 November, 1998.
7. Majewski, D., Liermann, D., Prohl, P., Ritter, B., Buchhold, M., Hanisch, T., Paul, G., Wergen, W., and Baumgardner, J. The Operational Global Icosahedral-Hexagonal Gridpoint Model GME: Description and High-Resolution Tests. *Month. Weath. Rev.*, 130:319–338, 2001.
8. Hendricks, J., Baier, F., Günter, G., Krüger, B., and Ebel, A. Stratospheric ozone depletion during the 1995-1996 Arctic winter: 3-D simulations on the potential role of different PSC types. *Ann. Geophysicae*, 19, 2001.
9. Krämer, M. et al. Intercomparison of stratospheric chemistry models under polar vortex conditions. *J. Atmos. Chem.*, 45:51–77, 2003.
10. Sander, S. P. et al. Chemical Kinetics and Photochemical Data for Use in Atmospheric Studies. JPL Publication 02-25, Jet Propulsion Laboratory, February 2003.
11. Verwer, J., Spee, E., Blom, J., and Hundsdorfer, W. A Second Order Rosenbrock Method Applied to Photochemical Dispersion Problems. Technical Report MAS R9717, Centrum voor Wiskunde en Informatica, 1997.
12. Sandu, A., Daescu, D. N., and Carmichael, G. R. Direct and Adjoint Sensitivity Analysis of Chemical Kinetic Systems with KPP: Part I – Theory and Software Tools. *Atmos. Environ.*, 37:5083–5096, 2003.
13. Giering, R. and Kaminski, T. Recipes for Adjoint Code Construction. *ACM Trans. Mathematical Software*, 24(4):437–474, December 1998.
14. Giering, R. *Tangent linear and Adjoint Model Compiler, Users manual TAMC Version 5.2*, September 1999.
15. Faure, C. and Papegay, Y. Odyssée Version 1.7 Users Guide. Technical Report 0224, Institut National de Recherche en Informatique et en Automatique, September 1998.
16. Weaver, A. and Courtier, P. Correlation Modelling on the Sphere Using a Generalized Diffusion Equation. *Quart. J. Roy. Meteor. Soc.*, 127:1815–1846, 2001.
17. Nocedal, J. Updating Quasi-Newton Matrices With Limited Storage. *Math. Comput.*, 35(151):773–782, July 1980.
18. Liu, D. C. and Nocedal, J. On the limited memory BFGS method for large scale optimization. *Math. Programming*, 45:503–528, 1989.
19. *SOCRATES, an Interactive 2-D Model of the Middle Atmosphere: Model Description*, 2004. <http://acd.ucar.edu/models/SOCRATES/>.