Assimilation of ENVISAT MIPAS and SCIAMACHY observations into a new stratospheric global chemistry circulation model – first results

J. Schwinger[1], H. Elbern[1], R. Botchorisvili[2] and D. Klasen[1]

www.riu.uni-koeln.de

[1]Rheinisches Institut für Umweltforschung an der Universität zu Köln

[2]Institut Algorithmen und Wissenschaftliches Rechnen, Fraunhofer-Gesellschaft

Objectives

Present and forthcoming satellite sensors offer an unprecedented surveillance capacity of the changing atmosphere. The SACADA project aims to devise an operational chemical data assimilation system to estimate synoptic fields of constituents. Observing the general objective to upgrade data to information, the following items are aspired:

i. optimal exploitation of satellite retrievals of various trace gases scattered in space and time and obtained from different sensors,

ii. provision for chemical consistency of the analysis products,

iii. extension to the estimation of not observed species, which are chemically coupled to observed species,

iv. portability of the system between parallel platforms,

v. highest possible computational efficiency.

The system is designed to become operational at the German Remote Sensing Data Center at DLR.

Methodology

To comply with these requirements a four-dimensional variational (4D-var) approach has been selected for SACADA. 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 timespan (1 day) on the one hand and an appropriate background field on the other hand:

$$J(\boldsymbol{x}_0) = J_b + J_o =$$

$$\frac{1}{2}\left[\boldsymbol{x}_0 - \boldsymbol{x}_b\right]^T \mathbf{B}^{-1}\left[\boldsymbol{x}_0 - \boldsymbol{x}_b\right] + \frac{1}{2}\sum_{i=0}^{N}\left[\mathbf{H}M_i(\boldsymbol{x}_0) - \boldsymbol{y}_i\right]^T \mathbf{R}^{-1}\left[\mathbf{H}M_i(\boldsymbol{x}_0) - \boldsymbol{y}_i\right]$$

Here $\boldsymbol{x}_0$ is the model state at $t=t_0$, $\boldsymbol{x}_b$ is an appropriate background model state and $\boldsymbol{y}_i$ 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 $\boldsymbol{x}_0$ forward in time to yield the concentrations $\boldsymbol{x}(t_i)$ at $t=t_i$.

For a proper weighting of the information that is contained in the observations and in the background, covariances of all quantities have to be specified as accurate as possible by means of the covariance matrices $\mathbf{B}$ and $\mathbf{R}$ (where $\mathbf{R}$ also contains the error of model representativeness). In order to find the minimum of J the gradient of the cost function with respect to the initial concentrations $\boldsymbol{x}_0$ is needed:

$$\nabla_{\boldsymbol{x}_0} J = \mathbf{B}^{-1}\left[\boldsymbol{x}_0 - \boldsymbol{x}_b\right] + \sum_{i=0}^{N}\mathbf{M}_i^{*}\mathbf{H}^T\mathbf{R}^{-1}\left[\mathbf{H}\boldsymbol{x}(t_i)) - \boldsymbol{y}_i\right]$$

$\mathbf{M}^*_i$ is the adjoint model operator that maps the gradient of the cost function w.r.t. $\boldsymbol{x}(t_i)$ backwards in time to deliver the gradient of J w.r.t. the initial concentrations $\boldsymbol{x}_0$.

System Description

Kernel of the new system is a novel stratospheric global chemistry circulation model (GCCM) and its adjoint version, with the grid design adopted from the global weather forecast model (GME) of German Weather Service. The GME serves as an online

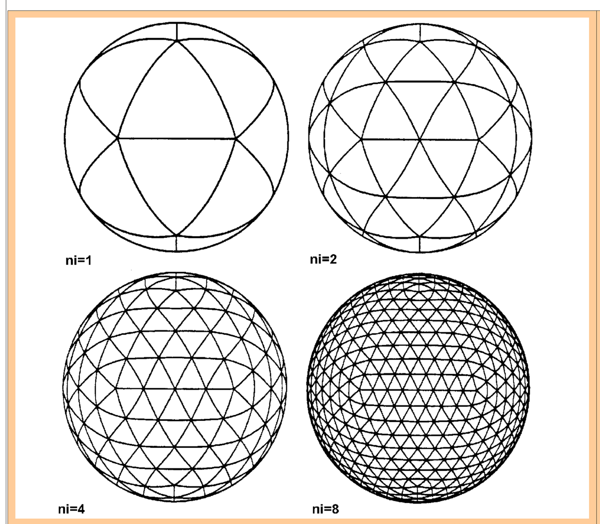


Figure 1: Construction of icosahedral grid. An icosahedron is placed into a sphere and adjacent vertices are connected. The resulting great circle arcs are subdivided into ***ni*** intervals to form a regular grid. The area of representativeness is a hexagon and a pentagon for the twelve vertices of the original icosahedron. (Source of figure: DWD)

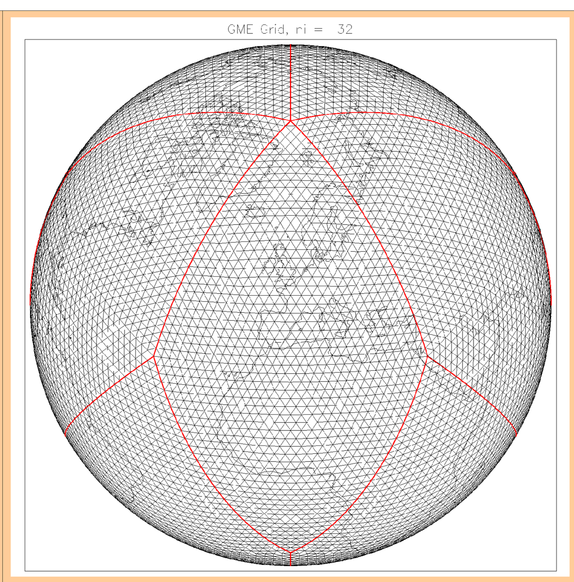


Figure 2: Icosahedral grid in the configuration that is currently used by the SACADA GCCM (ni=32). The resulting mesh size is 220-260 km (~T80, or 2.5° resolution), yielding 10,242 gridpoints per level. In the vertical there are 42 level from the ground up to 0.1 hPa.

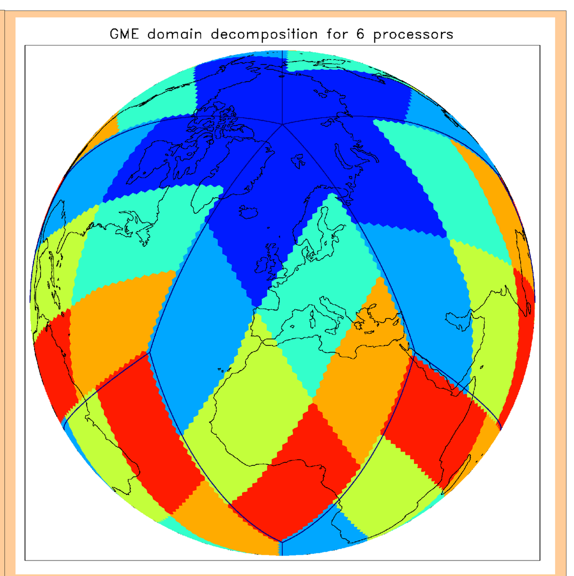


Figure 3: Domain decomposition for six processors: Two triangles of the icosahedron are combined to form a diamond. A single processor works on one rectangular section of each diamond. This is a simple but effective strategy for load balancing.

meteorological driver for the GCCM where the icosahedral grid structure (see Figures 1 and 2), the horizontal transport and the parallelisation strategy (see figure 3) are adopted from GME. The stratospheric chemistry module accounts for 148 gas phase and 7 heterogeneous reactions involving 43 stratospheric constituents. A 2nd order Rosenbrock solver with adaptive stepsize control is used to solve the system of chemical reactions. The minimisation procedure uses a quasi-newton method (L-BFGS algorithm) to iteratively find new initial values which better fit the model to the given observations. **Computational aspects:** The system is currently running on a PC-cluster with six AMD 2.4 GHz processors, reaching a wallclock runtime (for 15 iterations, data assimilation on levels 10-25) of approximately seven hours.

Experimental Set-up

To demonstrate the performance of the new data assimilation system, a period of one week in October/November 2003 was selected for assimilation. Observational data from the MIPAS sensors on ENVISAT were provided by IMK (Forschungszentrum Karlsruhe) comprising the species O_3, CH_4, N_2O, HNO_3, NO_2, H_2O, $ClONO_2$ and N_2O_5. Additionally, a special data set from SCIAMACHY solar occultation measurements (O_3 and NO_2) was provided by IfE (University of Bremen). An initial first guess for day 302 (Oct. 29, 2003) was produced by using output of the SOCRATES 2d model and a 144 hours spin-up model run to expand towards a chemical equilibrium. Starting from this point, observational data for 24 hours was consecutively assimilated into the GCCM for the days 302-308 (Oct. 29 - Nov. 4, 2003), using the analysis from the previous day as a first guess for the next day assimilation. A control model (model run without data assimilation) was accomplished for the same period of time. Note that for this first test of the new assimilation system some unvalidated estimates for the covariances have been made.

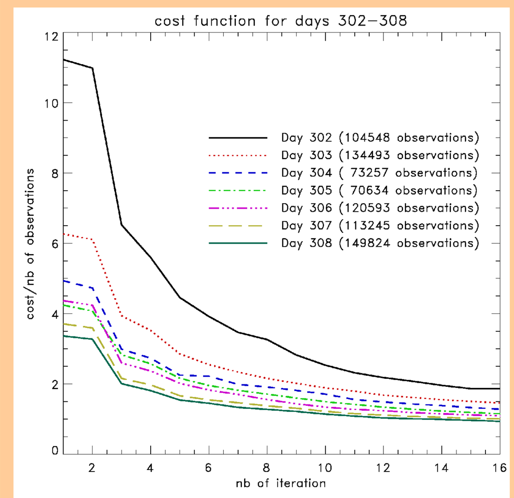


Figure 4: Evolution of cost function during the course of the assimilation procedure and the total number of available observations for each day. The cost function was divided by the total number of observations for each day to make the values comparable.

Results

The assimilation procedure significantly reduces the discrepancy between the model state and observations from day to day as more information comes into the model (see Figure 4). Note that the spatial coverage of the data is low in equatorial regions

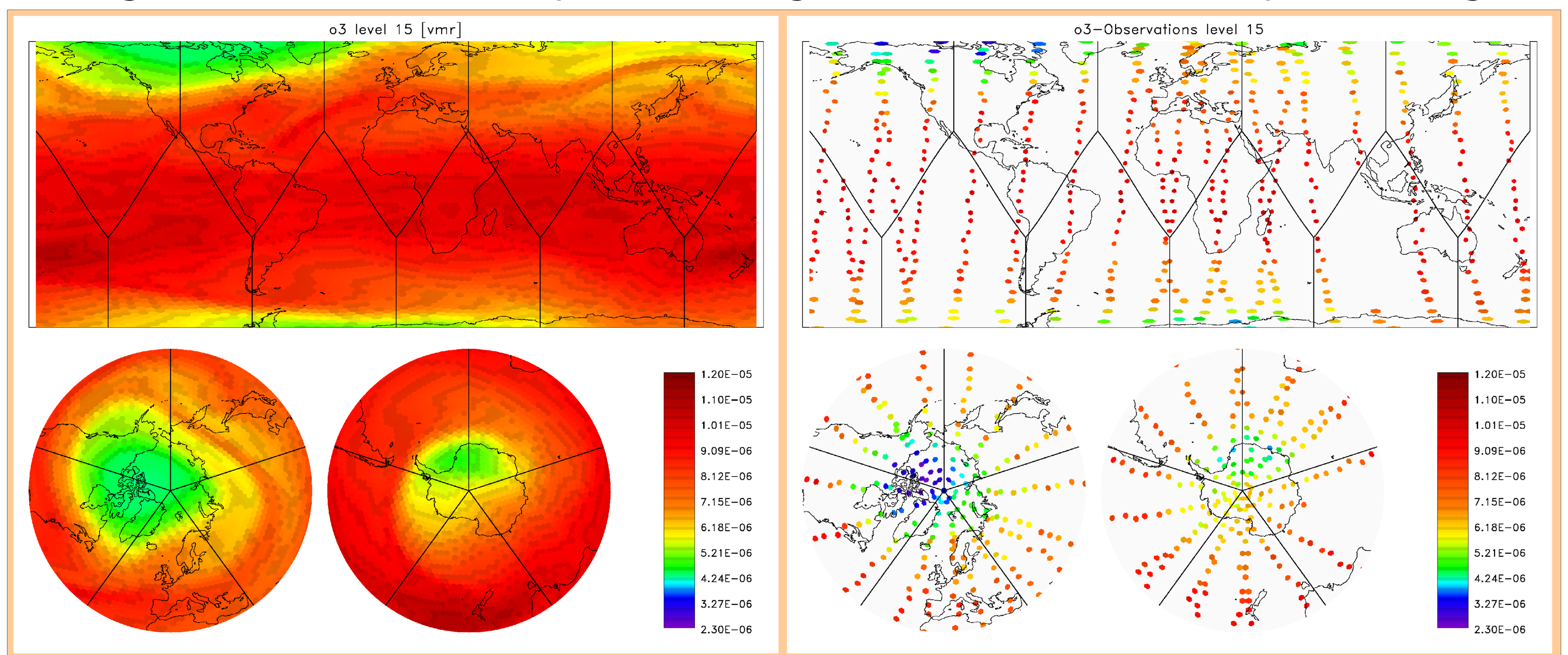
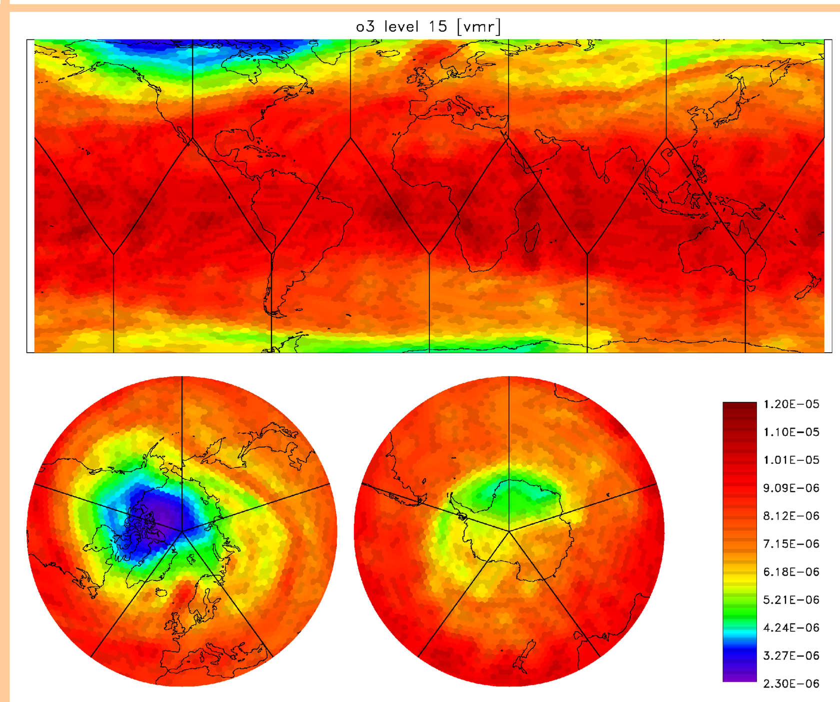


Figure 5: Assimilation results for day 308, after daily application of the assimilation system for a period of one week. For comparison the results of the control run are shown (upper panel left hand side) together with the final assimilation result for day 308, 12:00 UTC for ozone at 7.6 hPa (~33 km, lower panel right hand side). The available ozone observations for day 308 are shown at the right hand side, upper panel. Note that day 308 was a day with a very good data coverage compared to some of the preceding days (see figure 4). Globally, 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.



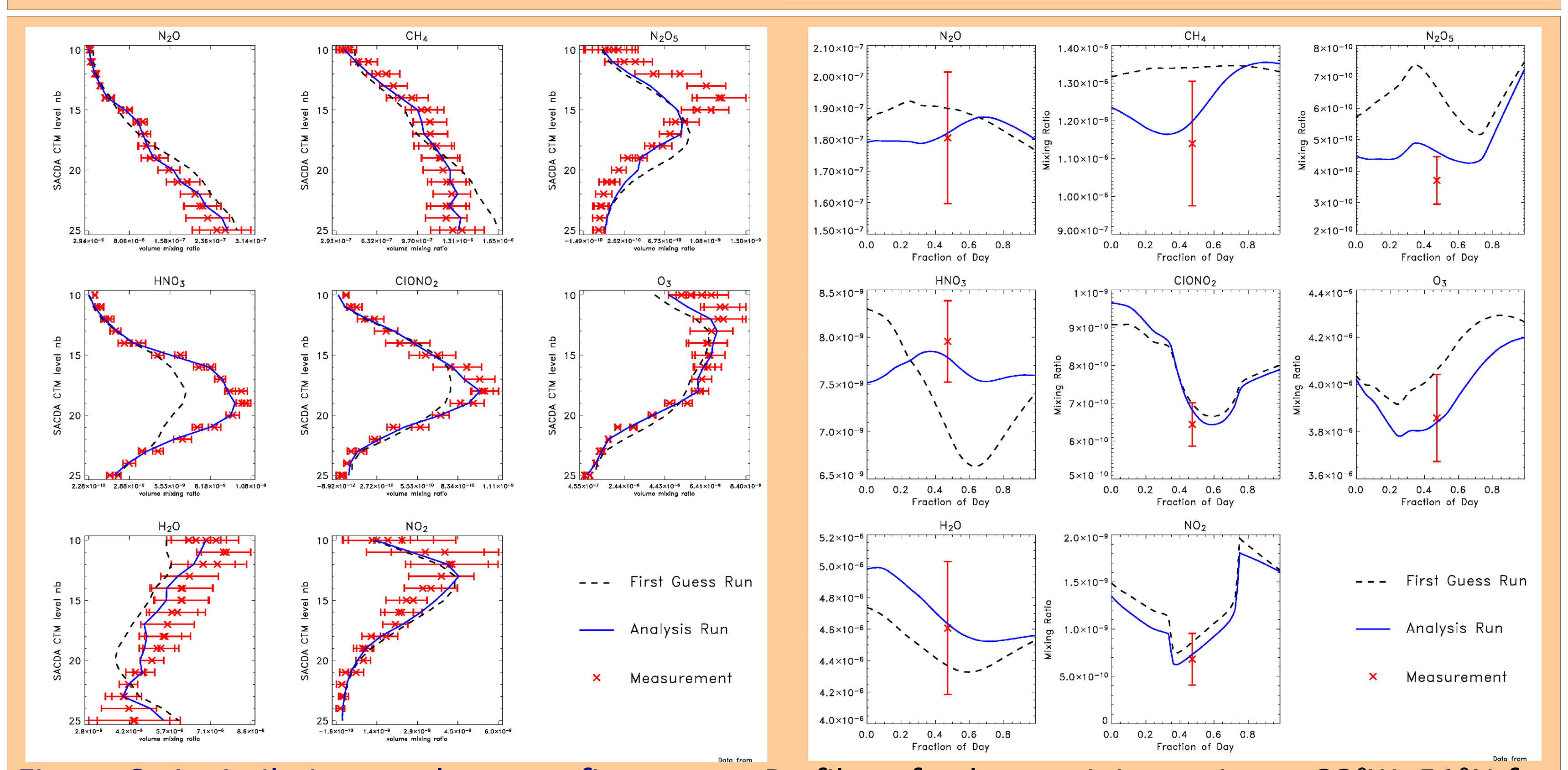


Figure 6: Assimilation result versus first guess. Profiles of volume mixing ratios at 33°W, 51°N for day 303 and corresponding observations (left hand side). Timeseries of volume mixing ratio at 17°W, 52°N, 25 hPa for day 308 and corresponding observations. The blue line is the analysis model state while the black dotted line gives the first guess.

(see Figure 5), so there are still unobserved air masses at low latitudes even after a few days of assimilation. The final assimilation result for ozone (model state at day 308, 12:00 UTC) is displayed in Figure 5. Compared to the control run the most prominent benefit from data assimilation is the correction of too high ozone values in the arctic region. The analysis of the vertical structure of trace gas profiles is generally very good, as can be seen in Figure 6.

Conclusions/Future Work

It has been demonstrated that the application of the new 4D-var system can significantly reduce the discrepancy of our knowledge of the atmospheric state as described by a model and observations of stratospheric trace gases from space borne sensors. It can be expected that the performance of the system will further improve by applying a more sophisticated covariance formulation, especially for the background covariances $\mathbf{B}$. Validation of the system using data from campaign flights together with satellite observations will follow.

Acknowledgements

We are especially grateful for support by the following groups, persons and institutions:
M. Kiefer and G. Stiller at IMK-FZK for the provision of their MIPAS observational data product; J. Meyer and H. Bovensmann at IFE, University of Bremen for the provision of nonoperational solar occultation observational data; German Weather Service (DWD) for the provision of their GME software; Sabine Pott at SCAI-FhG for the preparation of meteorological initial data.