# Assimilation of ENVISAT stratospheric trace gas observations into the new SACADA global chemistry circulation model

J. Schwinger[1], H. Elbern[1], R. Botchorisvili[2]
www.riu.uni-koeln.de
[1]Rheinisches Institut für Umweltforschung an der Universität zu Köln
[2]Institut Algorithmen und Wissenschaftliches Rechnen, Fraunhofer-Gesellschaft

## Objectives and Methodology

A new four-dimensional variational (4D-var) data assimilation system for stratospheric trace gas observations is being developed by the project consortium SACADA. The basic idea of 4D-var is to minimise a scalar cost function $J$:

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

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 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, which 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 $J$ w.r.t. $\boldsymbol{x}(t_i)$ backwards in time to deliver the gradient 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. The German Weather Service's global forecast model (GME) serves as an online meteorological driver for the GCCM. The icosahedral grid structure (**Fig. 1**), the horizontal transport and the parallelisation strategy (**Fig. 2**) are adopted from GME. The stratospheric chemistry module accounts for 148 gas phase and 7 heterogeneous reactions involving 43 stratospheric constituents. The background error covariance matrix (BECM) **B** is modelled using a diffusion approach, yielding a quasi-gaussian correlation of background errors between neighbouring grid points (**Fig. 3**), while **R** is taken to be diagonal. The minimisation procedure uses a quasi-newton (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, consuming a wallclock runtime of six hours for the assimilation of data from a 24h interval.

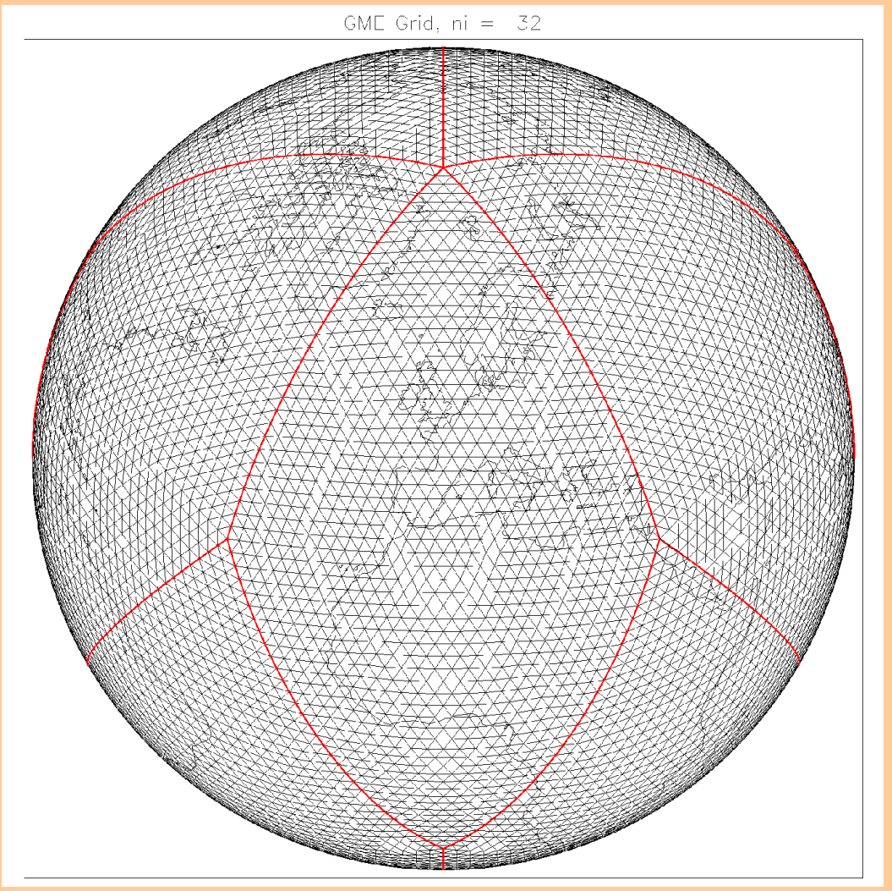

Figure 1: Icosahedral grid in the configuration used by the SACADA GCCM. The resulting mesh size is 220-260 km (~T80, or 2.5° resolution), yielding 10,242 grid points per level. In the vertical there are 42 level from the ground up to 0.1 hPa.

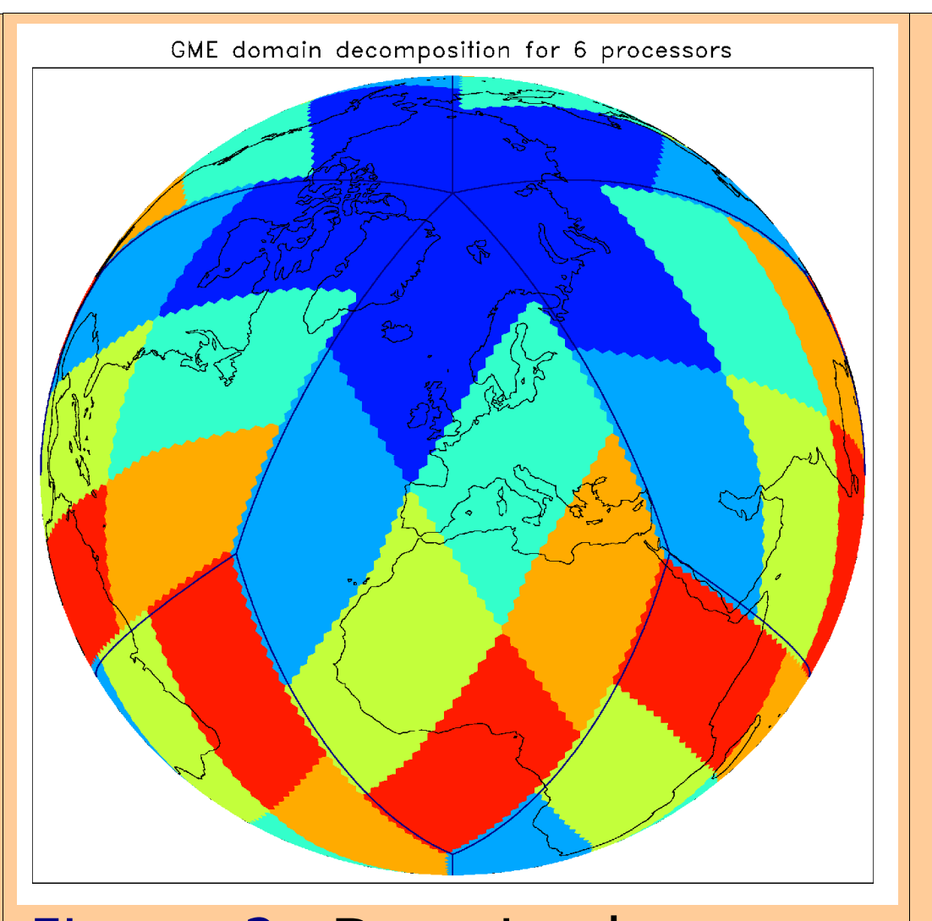

Figure 2: Domain decomposition for six processors: Two triangles of the icosahedron are combined to form a diamond. Each processor works on one rectangular section of all diamonds. This is a simple but effective strategy for load balancing.

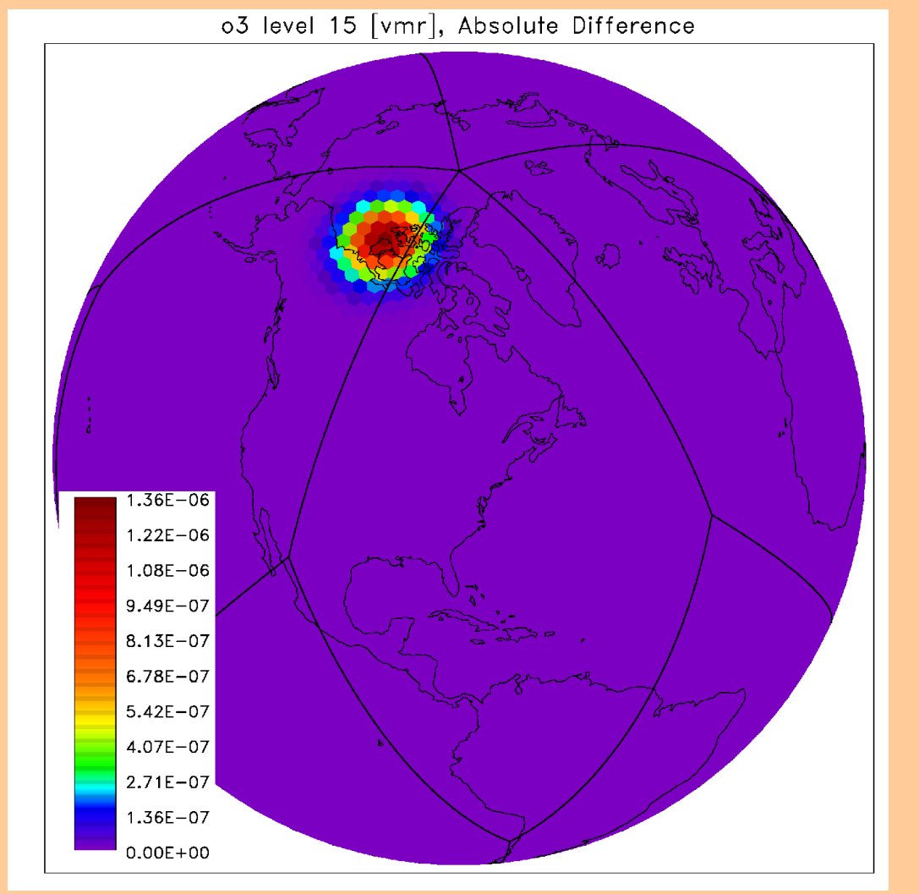

Figure 3: Analysis increment caused by a single ozone observation. The increment is spread out to neighbouring grid points due to the correlation of background error. The BECM is modelled using a diffusion approach.

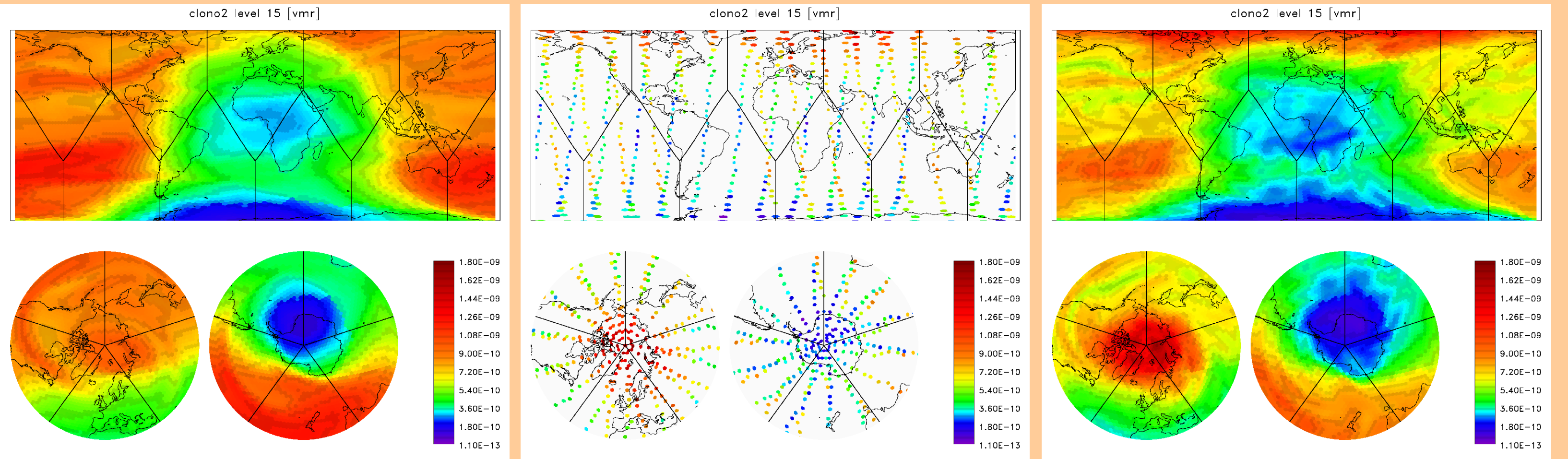

**Figure 4:** $ClONO_2$ assimilation results for Nov. 13, 2003, 12:00 UTC at 7.6 hPa (~33 km): Results of the control run are shown (left) together with the available $ClONO_2$ observations for that day (centre) and the corresponding analysis (right). Compared to the control run, features of the analysed atmospheric state are the lower night-time volume mixing ratios of $ClONO_2$ and enhanced chlorine nitrate inside the arctic polar vortex.

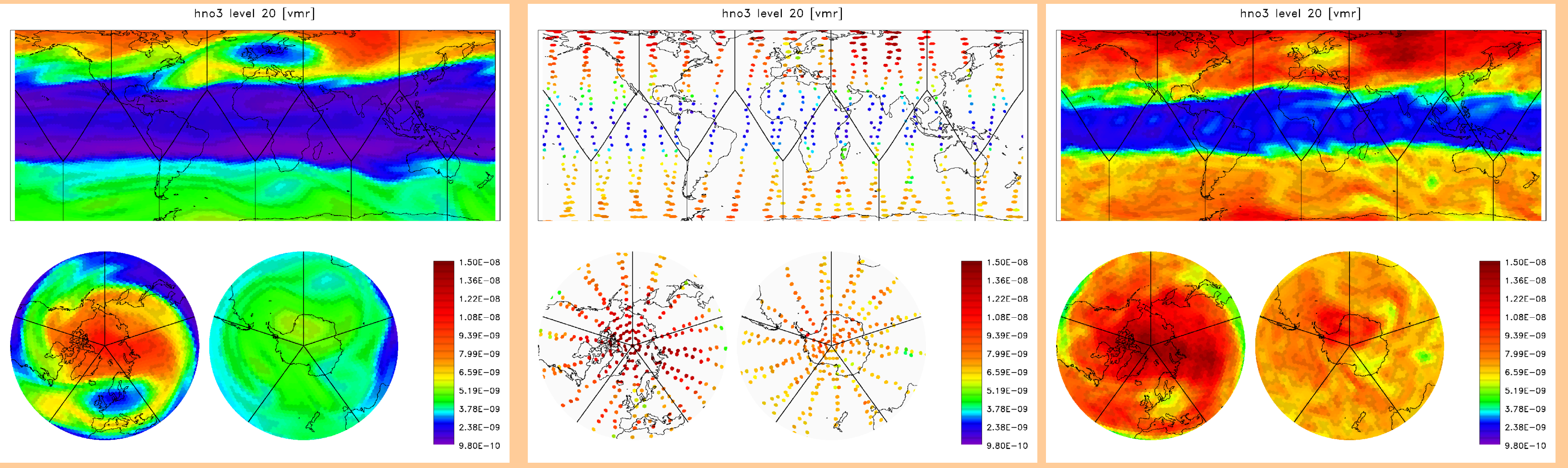

**Figure 5:** $HNO_3$ assimilation results for Nov. 13, 2003, 12:00 UTC at 28 hPa (~24 km): Results of the control run are shown (left) together with the available $HNO_3$ observations for that day (centre) and the corresponding analysis (right). Generally, the control run (without data assimilation) shows too low $HNO_3$ volume mixing ratios, especially in the southern hemisphere.

## Experimental Set-up

To demonstrate the performance of the new data assimilation system, a period of 25 days during Oct./Nov. 2003 was selected for assimilation. Observational data from the MIPAS sensors on ENVISAT have been provided by IMK (Forschungszentrum Karlsruhe) comprising the species $O_3$, $CH_4$, $N_2O$, $HNO_3$, $NO_2$, $H_2O$, $ClONO_2$, $N_2O_5$, CFC-11 and CFC-12. An initial chemical state for Oct. 21, 2003 was derived from the SOCRATES 2d model. Observational data from each day was consecutively assimilated into the GCCM from Oct. 21 to Nov. 14, 2003, using the analysis from the previous day as the background field for next day's assimilation. A control run (model run without any data assimilation) was accomplished for the same period of time. For the BECM parametrisation, a background error of 100% and a correlation length of 600 km have been assumed for the first six days of the period, to allow a rapid forcing of the model towards the observed atmospheric state. These values have been reduced to 50% and 300 km for the remaining days, because these values are more realistic, once the background is already in reasonable agreement with the true atmospheric state.

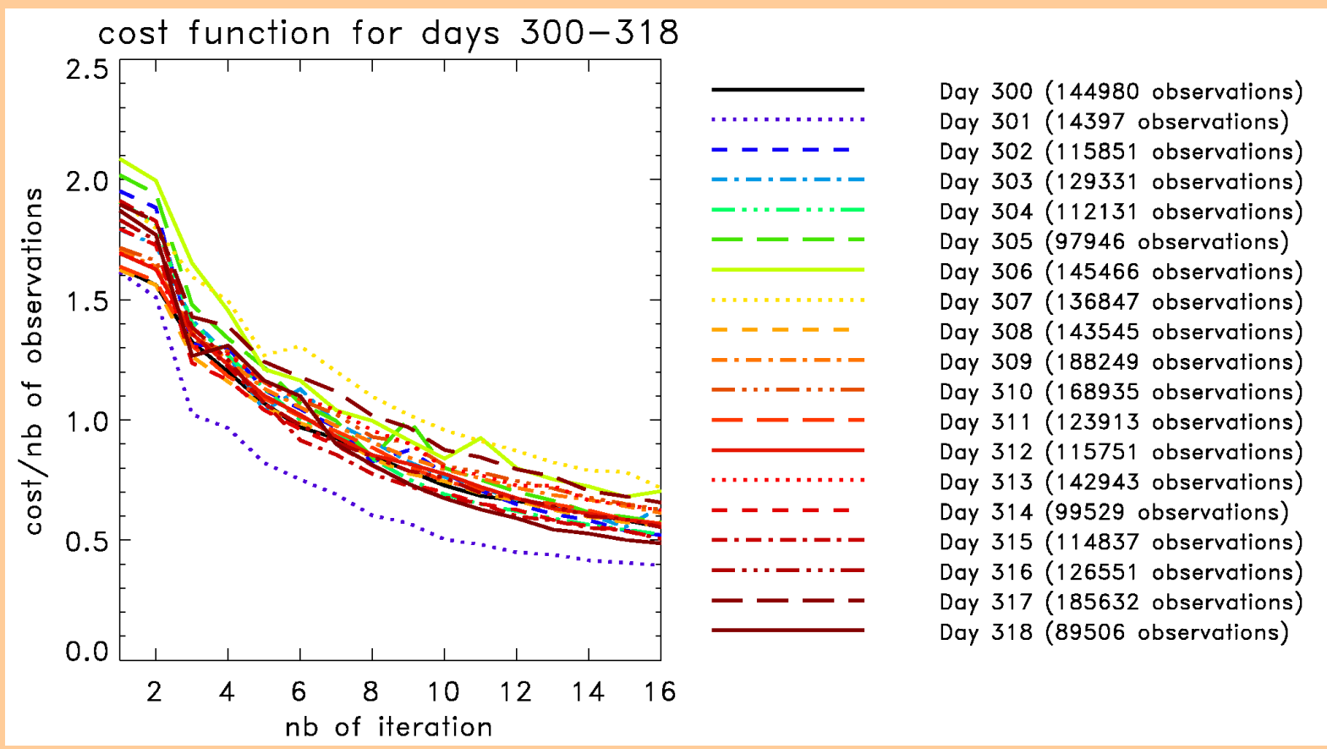

Figure 6: Decrease of the cost function during the iterative process of minimisation for Oct. 27 - Nov. 14, 2003. The cost function shows a very low day to day variability, indicating a stable performance of the system. The theoretically expected optimum value of 0.5 at the minimum of the cost function (at iteration 16 = analysis) is slightly exceeded. This is due to the coarse estimates of BECM parameters and/or possible misspecifications of observational errors.

## Results

The assimilation procedure significantly reduces the discrepancy between the model state and observations as demonstrated in **Fig. 8**, while chemical consistency in terms of the underlying chemistry scheme is maintained. The analysis of ClONO2 and HNO3 for Nov. 13 is shown in **Figs. 4** and **5** together with the corresponding atmospheric state that is produced by the model without data assimilation (control run). The corrections that are applied to the background state are demonstrated in **Fig. 7** for selected profiles of MIPAS-IMK retrievals. The assimilation system shows a stable performance over the whole period, and BECM parameters seem to be reasonably estimated (**Figs. 6, 7**).

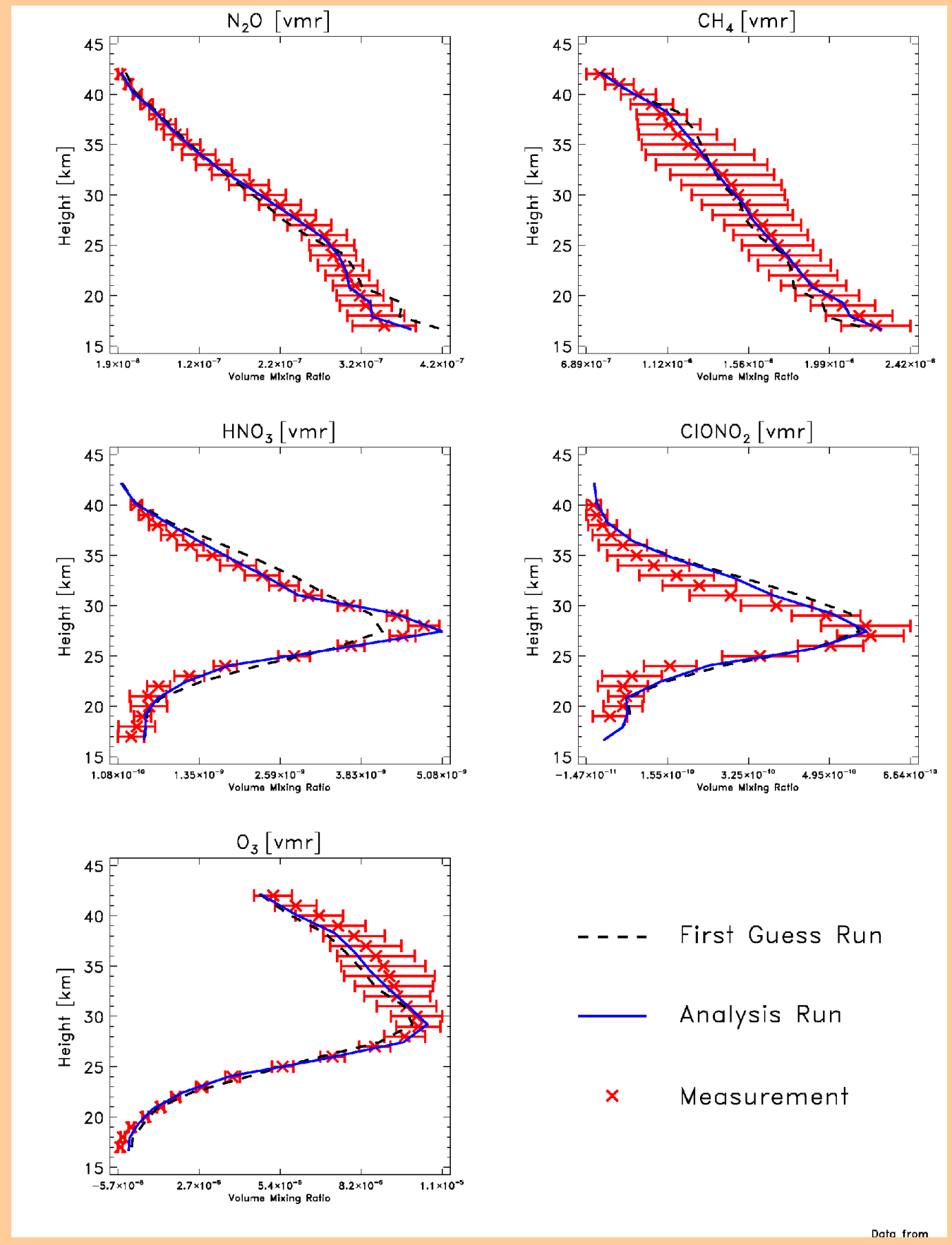

Figure 7: Assimilated Profiles above the Indian Ocean (57°E/2°N) for Nov. 14, 2003 and corresponding observations (blue line = analysis model state, black line = background). The model background state is already in good agreement with observations after 24 days of consecutive data assimilation, thus the corrections are relatively small.

## Conclusions/Future Work

It has been demonstrated that the application of the new 4D-var system can significantly reduce the discrepancy between our knowledge of the atmospheric state as described by a model and observations from space borne sensors. It can be expected that the performance of the system will further improve by applying a more sophisticated BECM formulation. Validation of the system using data from independent sources together with satellite observations will follow.

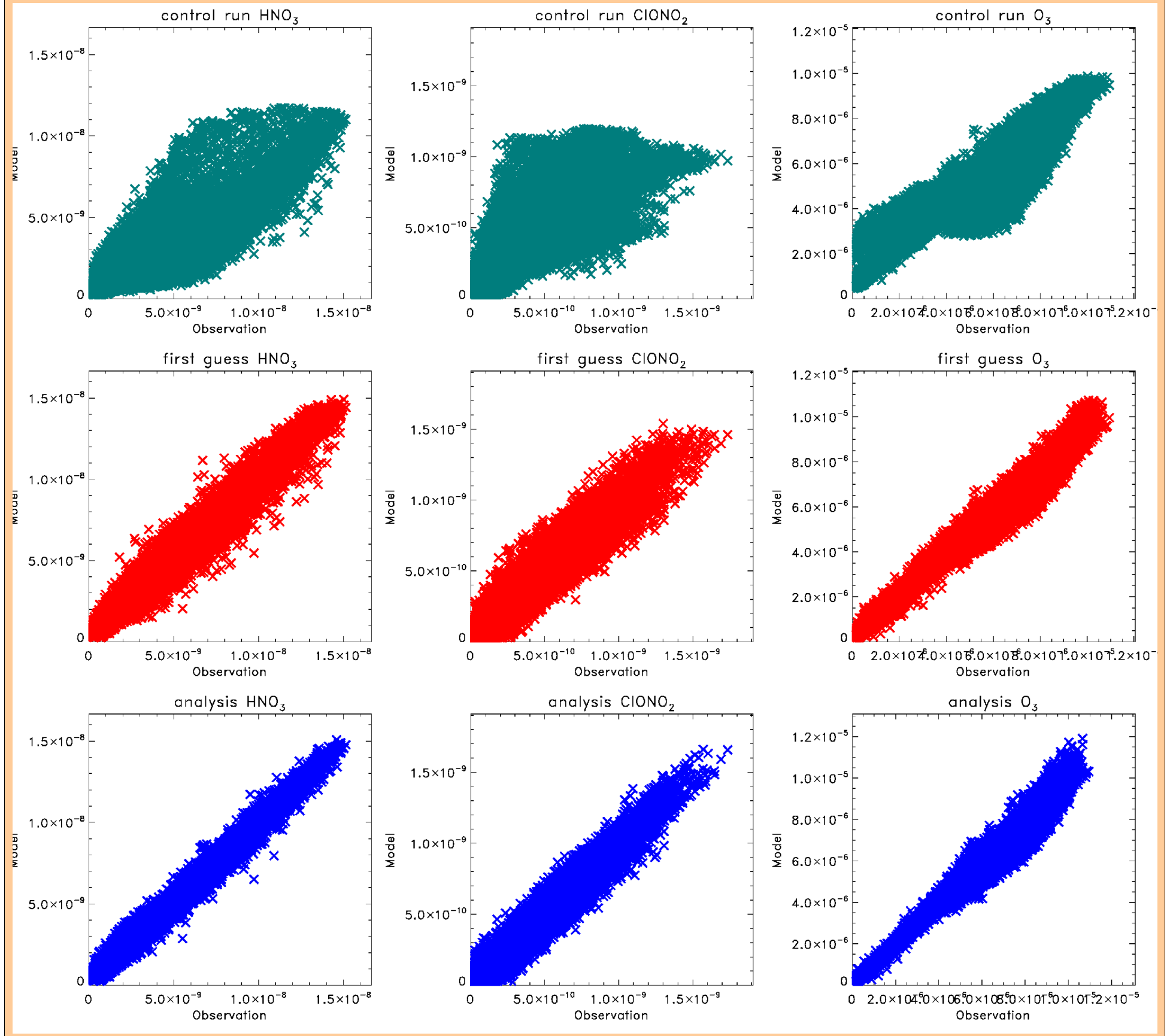

Figure 8: Scatterplots model versus observations; control run (green), background (red) and analysis (blue) for Nov. 13, 2003. The difference between background and analysis scatter represents the gain due to the assimilation of one day's data, while the model state's improvement over the whole case study period can be identified by comparison with the control run scatter.

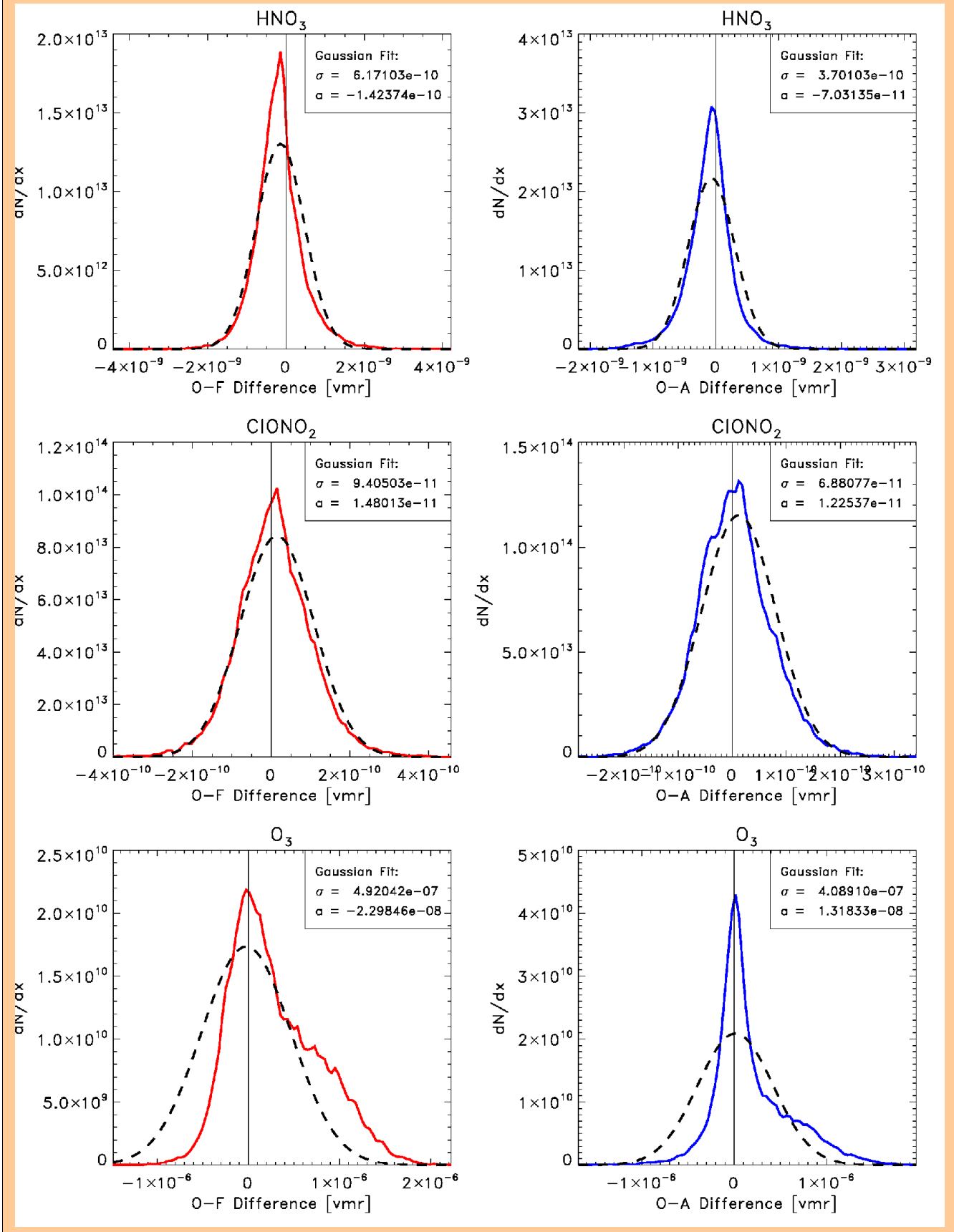

Figure 9: Observation-Forecast (red) and Observation-Analysis (blue) distributions. Biases are generally small, but the shape of the distributions is steeper than the corresponding gaussian distribution (black lines), which is a well known phenomenon in data assimilation systems.

### References:

**T. von Clarmann et al.**, Experimental evidence of perturbed odd hydrogen and chlorine chemistry after the October 2003 solar proton events, ***J. Geophys. Res.***, submitted Feb. 2005, under revision.

**M. Lopez-Puertas et al.**, Observation of $NO_x$ Enhancement and Ozone Depletion in the Northern and Southern Hemispheres After the October-November 2003 Solar Proton Events, ***J. Geophys. Res.***, submitted Feb. 2005, under revision.

**M. Lopez-Puertas et al.**, $HNO_3$, $N2O_5$, and $ClONO_2$ Enhancements after the October-November 2003 Solar Proton Events, ***J. Geophys. Res.***, submitted Feb. 2005, under revision.

**J. Schwinger, H. Elbern and R. Botchorishvili**, A new 4-Dimensional Variational Assimilation system applied to ENVISAT MIPAS observations, Proceedings of the Envisat Symposium, Salzburg 2004.

### 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;
German Weather Service (DWD) for the provision of their GME software; Sabine Pott at SCAI-FhG for the preparation of meteorological initial data.
The SACADA project is funded by the German Federal Ministry of Education and Research (BMBF) in the framework of AFO-2000.