Presented at the 7th Annual CMAS Conference, Chapel Hill, NC, October 6-8, 2008

INVERSE MODELING OF EMISSIONS USING THE CMAQ ADJOINT MODEL

Jaroslav Resler*, Kryštof Eben, Pavel Juruš and Pavel Krč

Institute of Computer Science
Academy of Sciences of the Czech Republic, Prague, Czech Republic

Summary

The CMAQ adjoint model has been used to set up a data assimilating system based on the 4DVar method. Ground-level observations of NO_2 and tropospheric NO_2 columns retrieved from the satellite instruments GOME2 and OMI have been assimilated into the model. The 4DVar method has been used to optimize both initial conditions and emission factors. Simulations and forecasting experiments have been performed on mesoscale domains with different horizontal resolution.

Introduction

We used for our simulations the model pairs WRF-CMAQ and WRF-CAMx. Both pairs have been configured for three nested domains with horizontal resolutions 27, 9 and 3 km. As a preliminary work, a comparison of retrieved columns with their model counterparts has been done with the pair WRF-CAMx. [Eben et al., 2007]. The information contained in satellite columns and its potential for data assimilation has been investigated. Besides data assimilation, several possibilities for using this information have been found. In particular, model bias and stations with low representativeness can be detected from averaged differences of the observed and modeled columns. Another positive contribution of satellite columns is their lower sensitivity to daily cycles of NO_2 concentrations, which enables better tracking of transport phenomena.

In earlier work [Eben et al. 2005] it was found that correction of model concentrations achieved by assimilation of in situ observations is effective if we want to improve e.g. an exposure index estimated from long-term off-line simulations. For the purpose of prediction this kind of data assimilation has a limited benefit as long as model bias and errors in emission inputs are present. Another cause of this behavior is the lack of information for higher levels of troposphere.

In order to improve the forecast performance of the model, further sources of information are required and emission constraining appears to be necessary. Since the latter is hard to achieve by ensemble techniques, a CTM capable of adjoint modeling had to be selected. The adjoint operator for the model CMAQ was being developed by the CMAQ community during recent years ([Hakami, 2007]). We have finalized the parallelization of the adjoint operator. We also contributed several technical improvements to the adjoint code so as to enable its use for real cases. Finally, observation operators for satellite columns and their adjoint have been implemented. Thus the 4DVar method could be used. We have done some adaptations in order to constrain both initial conditions and emissions.

Assimilation of satellite columns and in-situ observations

We used a similar approach as [Elbern et al., 2007]. As a first step, the variability in emissions has been roughly parameterized by allowing for an emission factor, specific for each gridpoint. This leads to the discrete formulation of the cost function

$$J(c_0,e)=(c_0-c_B)^T B^{-1}(c_0-c_B)+(e-e_B)^T K^{-1}(e-e_B)+\sum_{i=1}^{N}(y_i-H(M(c_0,e,t_i)))^T R^{-1}(y_i-H(M(c_0,e,t_i)))$$

where

- c_0, c_B are optimized and first guess concentrations in time t_0
- e, e_B are optimized and first guess emission multiplicative factors
- $c = M(c_0,e,t)$ are the modeled concentrations
- H is the observation operator
- y are the available observations, both satellite-retrieved columns and in situ observations

* *Corresponding author: Jaroslav Resler, Institute of Computer Science, Academy of Sciences of the Czech Republic, Pod Vodarenskou vezi 2, 182 07 Prague 8, Czech Republic, e-mail: resler@cs.cas.cz*

- B, K and R and the covariance matrices for initial conditions, emission factors and observations errors

The cost function J is minimized utilizing the L-BFGS-B algorithm [Zhu 1997] and the gradient of J with respect to joint variable (c_0, e) can be expressed as

$$\nabla_{(c_0,e)} J(c_0,e) = B^{-1}(c_0 - c_B) + K^{-1}(e - e_B) + \sum_{i=1}^{N} M^T H^T R^{-1}(y_i - H(M(c_0, e, t_i)))$$

This gradient is calculated by means of the adjoint method. For easier problem constraining we apply a logarithmic transformation of variables

$$u(c_0, e) = (\ln(c_0), \ln(e))$$

The variable u can be considered as unconstrained.

The implementation of the adjoint in CMAQ is done by similar means as in the model STEM (see [Sandu et al., 2003]). An experimental CMAQ adjoint code was implemented in California Institute of Technology and Virginia Polytechnic Institute. This code contains the adjoint model for gas phase processes for the mechanism CB4. We have parallelized it in order to be able to process larger domains and real cases and we implemented our 4DVar assimilation above this adjoint code.

The observation operator H has two separate versions, one for treating tropospheric columns obtained by the retrieval process and the other one for handling in-situ observations. The first one is computed by numerical integration and accounts for the spatial intersection of the satellite instrument pixel with grid cells and for the influence of the retrieval process given by the averaging kernel operator [Eskes and Boersma, 2003]. The second one is straightforward. If there are more than one station in a grid cell, an average of all measurements has been taken for y.

Similarly the matrix R is assembled from two blocks. The estimates of standard deviations of the retrieved NO_2 columns are based on values supplied by the TEMIS service of the European Space Agency (www.temis.nl). These values are multiplied by a factor (0.3 in this experiment) which takes into consideration smaller time variability of these values in comparison with ground values and thus longer time representativity of these values. As for the block corresponding to the in-situ observations, the base standard deviation of the errors for the stations is calculated as 20% of the NO_2 average over all stations in the domain during a day. Representativeness errors are taken into account by multiplying this constant by a factor which depends on the type of the station. After some test runs, the multiplicative coefficients were set up to 20, 10 and 1 for urban, suburban and rural stations respectively. This rough approach will be generalized in the future. Only background stations have been taken into account. The matrices B a K are taken diagonal, some tests were done with covariance matrices constructed by means of a diffusion operator.

The assimilation experiment

In our experiments we used three nested domains (see Fig. 1). The outer domain encompasses most parts of Europe (horizontal res. 27km). It was used for obtaining realistic initial and boundary conditions for the assimilation run. The assimilation experiment was performed on a subwindow of the outer domain with 72x52 gridpoints. In some simulations, an intermediate domain (not shown in the figure, res. 9km) was used. It was also used for all WRF nested runs. The fine 3km resolution domain covers north-west part of the Czech Republic and adjacent regions of Germany (62x46 gridpoints). The emission data are based on the EMEP[1] inventory, the emission model was the same as in [Eben et al., 2005].

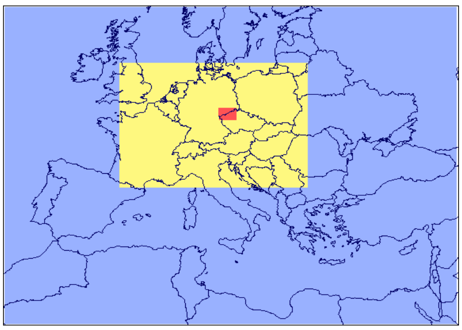


Fig. 1: Simulation domains: outer domain, assimilation domain, fine domain.

Tropospheric NO_2 columns are retrieved from measurements obtained by satellite instruments OMI and GOME2 (provided by the TEMIS service). The data contain all other necessary information, in particular the averaging kernel operator, the air mass factor etc. The necessary meteorological variables for constructing of the observation operator are taken from the WRF output files processed by MCIP3.

The overpass of the satellites above our domains occurs usually twice daily in the interval from 9am to 3pm. In Fig. 2 there is a typical example of the tropospheric column data (white squares represent cloudy conditions).

1 *EMEP – European Monitoring and Evaluation Programme, www.emep.int*

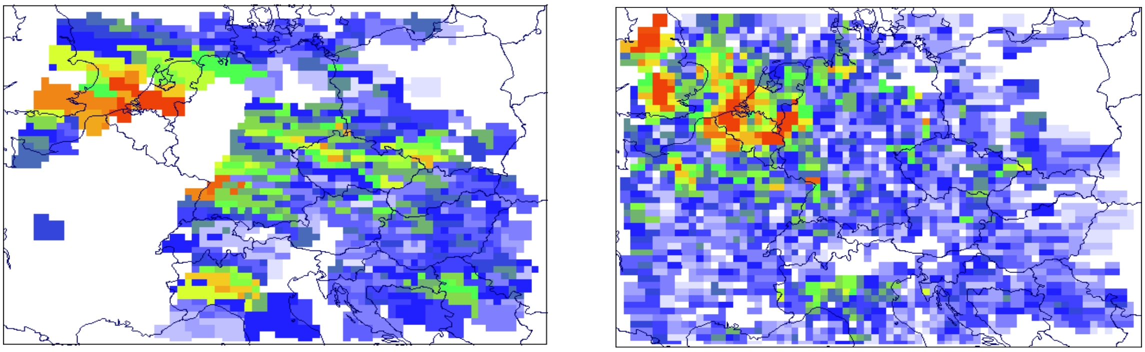


Fig. 2: An example of retrieved tropospheric column data from GOME2 (left) and OMI(right) for the coarse assimilation domain.

Altogether 280 background monitoring stations have been included into the experiment.[2] Their location in coarse and fine domain is in Fig. 3.

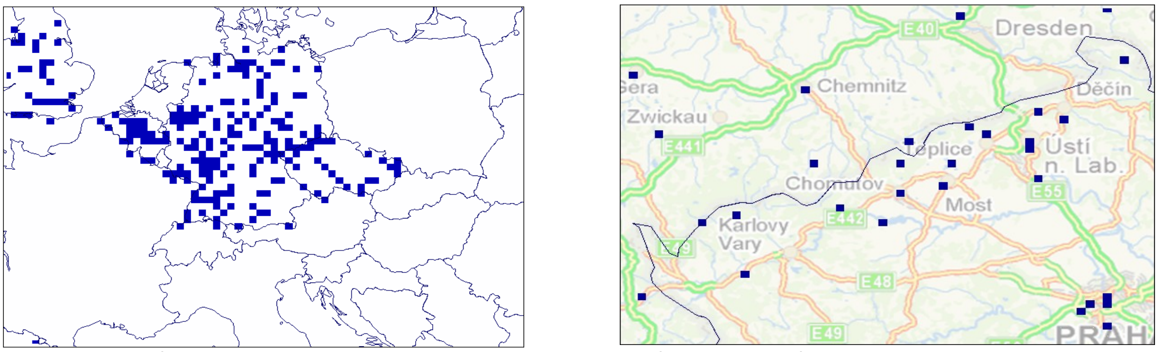
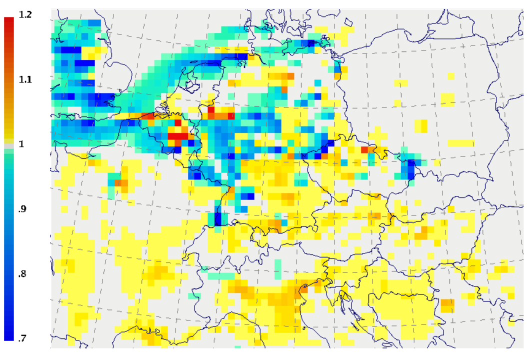


Fig. 3: Location of ground-level stations in the coarse domain (left) and in the fine domain (right)

We selected for our experiments an assimilation period of eight days from June 28 to July 5. A free run of CMAQ, 21 days on the outer domain, was performed to obtain reasonable initial and boundary conditions.
For each day sequentially, an assimilation run has been performed, assimilating both NO_2 columns and in-situ observations. A one-day-ahead forecast has also been computed, using emission factors and initial conditions obtained by 4DVar for the previous day.

Results

During the first two days of assimilation, the estimated emission factors reached a fairly stable solution and since the third day the cost function showed only a decrease of a few percent, even though the number of observed satellite pixels (and thus the value of the cost function) varies due to changing cloud coverage. The map of emission factors for one day and the average from the entire assimilation period is shown in Fig. 4 and 5.



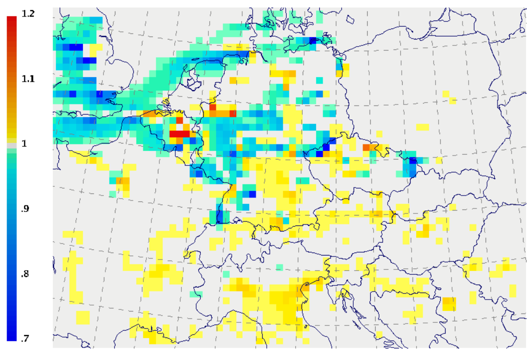
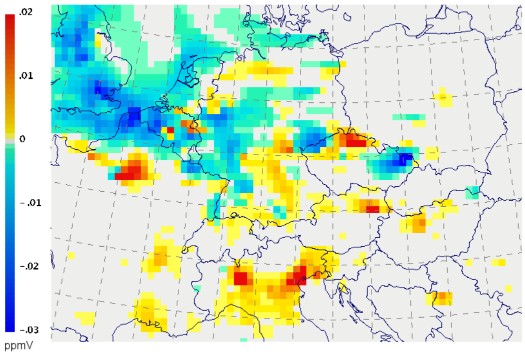


Fig. 4: Optimized emission factors for July 3, 2008
Fig. 5: The average of the optimized emission factors from June 29 to July 5, 2008

2 *The selection of data has been dictated by their availability. We have used data from the UK and Belgium (supplied by courtesy of the European Environment Agency), Germany (provided by the Umweltbundesamt www.uba.de) and Czech republic (provided by the Czech Hydrometeorological Institute).*

Fig. 6 depicts the changes between the free run and concentrations corrected by 4DVar. It is seen that both in situ observations and satellite columns contribute to the corrections. In particular, the corrections in France and Italy are induced by satellite observations only. In Fig. 7 there is a sample map of the gradient of the cost function with respect to the emission factors.



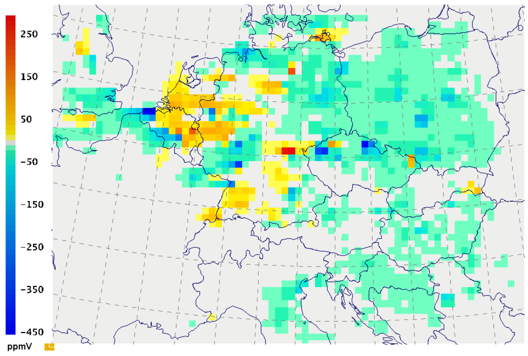


Fig. 6: Differences between optimized and referential concentrations of NO_2 for July 2, 2008 20:00
Fig. 7: Emission factors gradient for the first iteration for July 3, 2008

Table 1 contains results of the one-day ahead forecast, compared with these of the free run. It is seen that the forecast from assimilated values outperforms the free run, but the bias of the forecast is larger. In Fig. 8 we see one particular forecast for the fine domain, compared with the corresponding free run and the zoom of the coarse domain. Assimilation helps here to identify and localize sources in a better resolution (beyond the downscaling due to better orography and meteorology), even though the initial emission inventory is coarse.

No. of observations	*Free run mean residual*	*Forecast mean residual*	*Free run mean absolute res.*	*Forecast mean absolute res.*
58567	2.3	12.1	20.15	17.7

Tab 1. Mean residuals (i.e. differences of hourly observed value at a station and the model value) and absolute residuals for the free run and forecast from assimilated initial conditions and parameters. We excluded small values of NO_2 from the evaluation, so that only values of NO_2 larger than 20, either in the observation or in the model, enter the evaluation. All values are in μg/m^3.

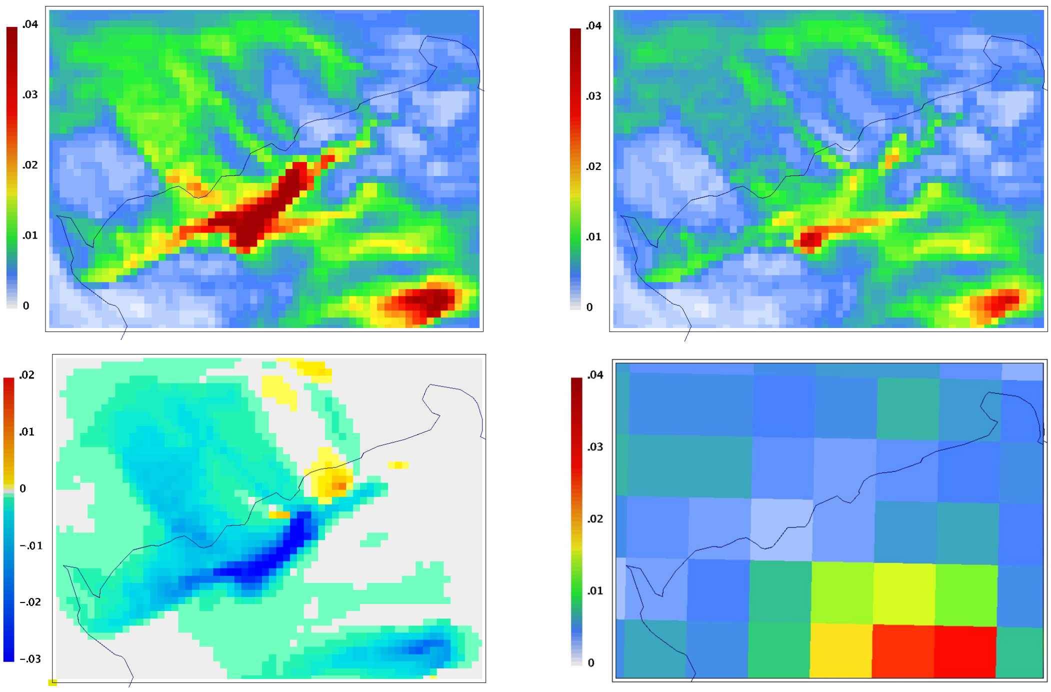


Fig. 8: Referential (upper left) and optimized (upper right) concentrations in ppmV for July 1, 2008 20:00 in the fine domain. The differences of the optimized and referential concentrations (lower left) and the optimized concentrations in the zoomed coarse domain (lower right).

In Fig. 9 there is an example of the time profile of concentrations for one of the stations in the polluted area, where the downward trend in emissions occurred during recent years. Although using an emission factor reduces the error substantially, the parameterization is too crude and a generalization is to be made, while keeping stability of the solution.

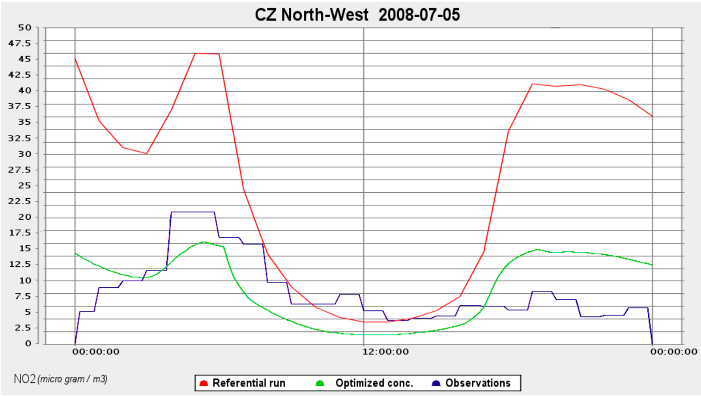


Fig. 9: Concentrations on obs. station Litoměřice (North-West Bohemia, the strongest changes by the assimilation).

Conclusions and outlook

The assimilation experiment shows stability and good performance of the 4DVar method in our setting. A greater flexibility of parameterization of emission corrections will be needed since a simple multiplicative factor cannot achieve a better fit of time profiles. A subsequent statistical analysis of the relation between corrected emission fields and parameters of the emission model should lead to the adaptation of the emission model itself.

The improvement in forecasting NO_2 concentrations due to assimilation of in situ and satellite observations is moderate so far, but evident. Also, a deeper investigation of complementarity between in-situ and satellite observations, is required. This would result in a better description of vertical profiles of NO_2. Finally, joint assimilation of other important species, in particular ozone, as well as more precise modeling of spatial covariances could improve the allover performance of the model.

References

[Eben et al. 2005] K.Eben, P.Jurus, J.Resler, M.Belda, E.Pelikan, B.C.Kruger and J.Keder: An ensemble Kalman filter for short term forecasting of tropospheric ozone concentrations, Q. J. R. Meteorol. Soc., 2005

[Eben et al. 2007] Eben, K., Resler,J., Juruš,P. , Belda M.,: A comparison of tropospheric NO_2 columns retrieved from satellite instruments with model values simulated by the mesoscale model CAMx. Proc. of the 2nd ACCENT symposium, Urbino, July 2007, 5p. (in print).

[Elbern et al. 2007] H. Elbern, A. Strunk, H. Schmidt and O. Talagrand: Emission rate and chemical state estimation by 4-dimensional variational inversion, Atmos. Chem. Phys., 7, 3749–3769, 2007

[Eskes et al. 2003] Eskes, H. J., and K.F. Boersma: Averaging kernels for DOAS total-column satellite retrievals, Atmos. Chem. Phys., 3, 1285-1291, 2003.

[Sandu et al 2005] Adrian Sandu, Dacian Daescu, Gregory Carmichael and Tianfeng Chai: Adjoint Sensitivity Analysis of Regional Air Quality Models, J. of Computational Physics 204 (2005), 222 – 252

[Sandu et al 2003] Adrian Sandu, Dacian N. Daescu, Gregory R. Carmichael: Direct and adjoint sensitivity analysis of chemical kinetic systems with KPP: Part I—theory and software tools, Atmospheric Environment 37 (2003) 5083–5096

[Hakami 2007] A. Hakami et al.: The Adjoint of CMAQ (Environ. Sci. Technol., 41 (2007),7807 -7817

[Zhu 1997] C. Zhu, R. H. Byrd and J. Nocedal: L-BFGS-B, FORTRAN routines for large scale bound constrained optimization, ACM Transactions on Mathematical Software, Vol 23, Num. 4 (1997), pp. 550-560.

Acknowledgement This work was supported by the Grant Agency of the Academy of Sciences of the Czech Rep. (grant No. 1ET400300414, framework „Information Society“), by the grant No. SP/1a4/107/07 of the Ministry of Environment of the Czech Rep. and by the Institutional research plan AV02 10300504 Computer Science for the Information Society “Models, Algorithms, Applications”.