Inverse modeling of emissions using the CMAQ adjoint model

Jaroslav Resler, Kryštof Eben, Pavel Juruš and Pavel Krč

E-mail: resler@cs.cas.cz

Institute of Computer Science
Academy of Sciences of the Czech Rep.
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.

Motivation

Chemistry transport models often show substantial bias and errors due to insufficient specification of inputs, lack of information in higher levels of troposphere, simplified chemistry etc.

Possible directions towards an improvement of performance:
- Optimization of the emission inventory and emission model.
- Using the information contained in satellite data. They have better spatial coverage but about two overpasses per day only. The essential feature of satellite data is that they bring an important information about vertical distribution of species.

Means to achieve these tasks:
Data assimilation and inverse modeling of emission

Adjoint of the CMAQ CTM

- Implemented in Caltech and Virginia Tech
- Includes adjoint code for gas phase (chemical mechanism CB4)
- Implementation by similar means as in STEM

A.Sandu et al.: Adjoint Sensitivity Analysis of Regional Air Quality Models (J. of Computational Physics 204 (2005))
A. Hakami et al.: The Adjoint of CMAQ (Environ. Sci. Technol., 41 (2007))
A. Sandu et al.: Direct and adjoint sensitivity analysis of chemical kinetic systems with KPP (Atmospheric Environment 37 (2003))

Our contribution:
- Finalized parallelization of the code
- Fixed some bugs
- Implementation of the observation operator for satellite data

What is still missing:
- Some scientific processes
- More chemical mechanisms (SAPRC99, CB5)
- Aerosol processes
- Parallelization of the L-BFGS-B code
- Standard build process

Effectivity of the parallelization

The parallelized version of the adjoint and 4DVar code was tested for both correctness and performance in different configurations. Parallel efficiency of the code is about 0.8 for 16 parallel MPI processes. This test was performed for a 4DVar run on coarse assimilation domain on a Linux computer with 4xAMD Opteron 8350, 16GB RAM.

4DVar design

- We follow partially the approach of H. Elbern
  H.Elbern et al.: Emission rate and chemical state estimation by 4-dimensional variational inversion Atmos. Chem. Phys., 7, 3749–3769, 2007
- Emissions optimized by one multiplicative factor for each gridpoint and day
- We minimize a discrete cost function

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

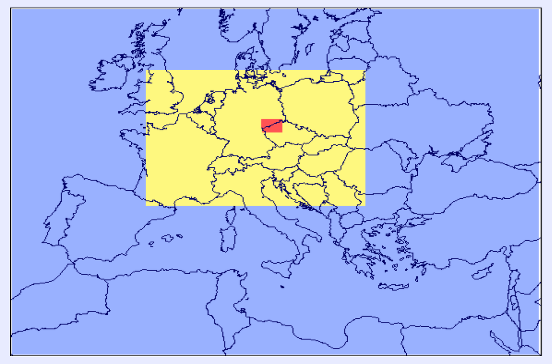
where
c_0, c_B are optimized and first guess concentrations in time t_0
e, e_B are optimized and first guess emission multiplicative factors
$c = M(c_0,e,t)$ are the modeled concentrations
H is the observation operator
y are the available observations, both satellite-retrieved columns and in situ observations
B, K and R and the covariance matrices for initial conditions, emission factors and observations errors

Gradient with respect to (c_0,e):

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

- Minimization of the cost function based on L-BFGS-B
  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, pp. 550-560 (1997)
- Covariance matrices in the cost function are taken diagonal.
  Some experiments with covariance matrices constructed by means of a diffusion operator done.
- Every single day starts from optimized initial conditions and emission factors from the previous day.

Setup of the experiment

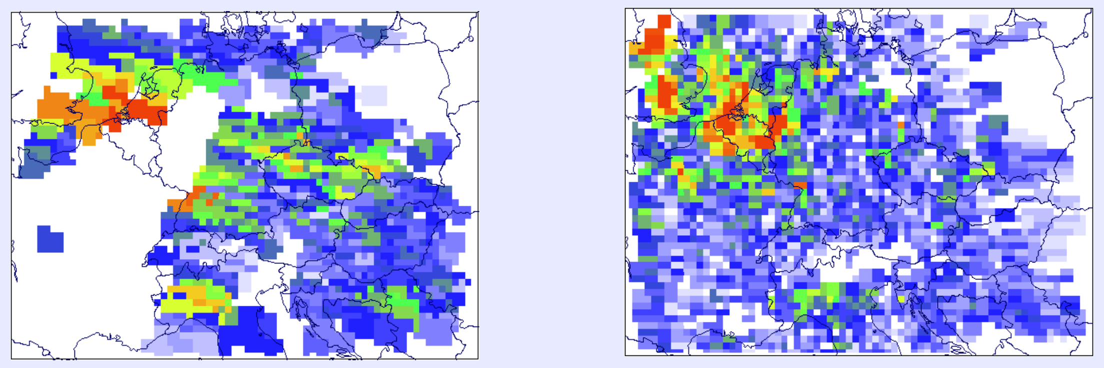


The domains used for the experiment

- Model pair WRF-CMAQ (adjoint version)
- 8 days from June 28 to July 5, 2008
- 3 nested domains (the large one for BC)
- Resolution 27km and 3km (red)
  (in WRF runs additionally an intermediate 9km domain)
- Fine domain at Czech-German border:
  - a heavily polluted area in the past (coal mining, chemical industry, electricity power stations)
  - large changes in emission sources in recent years probably not fully reflected in EMEP emission inventories
  - complex orography
- Emission model based on EMEP inventories with some enhancements for Central Europe.

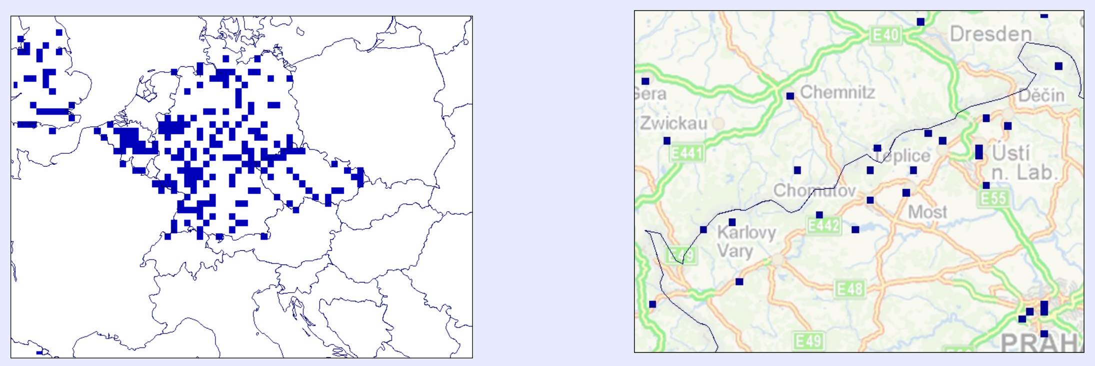
Tropospheric columns from GOME2 and OMI

We use the tropospheric columns of NO_2 retrieved using DOAS approach (data provided by TEMIS service of ESA). Linearized observation operator for satellite observations H and its adjoint H^T is constructed in every observation time step by using numerical integration, pixel and layer interpolation and applying the averaging kernel operator.



Tropospheric NO2 columns at July 1 2008 retrieved from GOME2 and OMI.
GOME2 track about 10:24 CET (left), OMI track about 13:24 CET (right)

In-situ observations



Ground level observations from CZ, DE, B, GB *(provided by CHMI, UBA and EEA)*
- Coarse domain - 280 background stations included (left)
- Fine domain - 30 background and industrial stations included (right)

Results

Decrease of the cost function

After the first two days, the decrease of the cost function is rather small.

Day	Date	BG CF	Opt. CF	Ratio(%)
1	28.6.2008	180693	56533	31
2	29.6.2008	63779	53148	83
3	30.6.2008	58806	53890	92
4	01.7.2008	88471	84801	96
5	02.7.2008	316840	274541	87
6	03.7.2008	91861	90591	99
7	04.7.2008	62442	56775	91
8	05.7.2008	48790	46070	94

Coarse domain, 27km resolution

Day	Date	BG CF	Opt. CF	Ratio(%)
1	28.6.2008	270607	21819	8
2	29.6.2008	33477	29504	88
3	30.6.2008	32717	27916	85
4	01.7.2008	40288	34889	87
5	02.7.2008	27839	27706	99

Fine domain, 3km resolution

Performance of the forecast

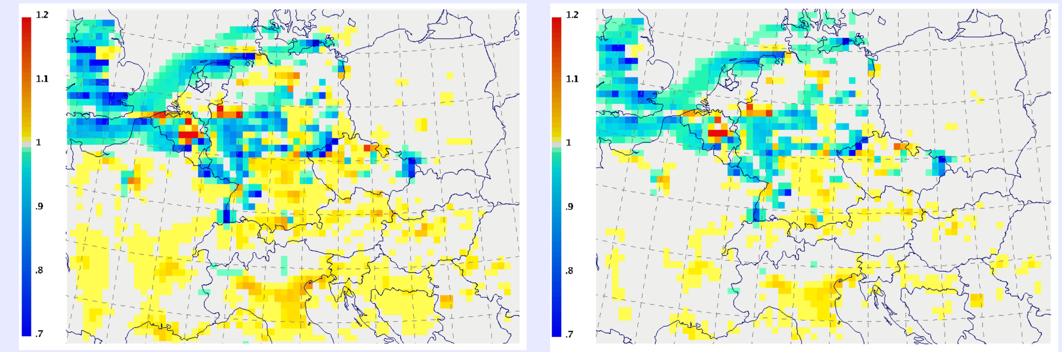
A forecast of next day's concentrations was run each day, starting from optimized initial concentrations and emission factors obtained by data assimilation for the previous day.

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

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 optimized 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/m3.

Optimized emission factors

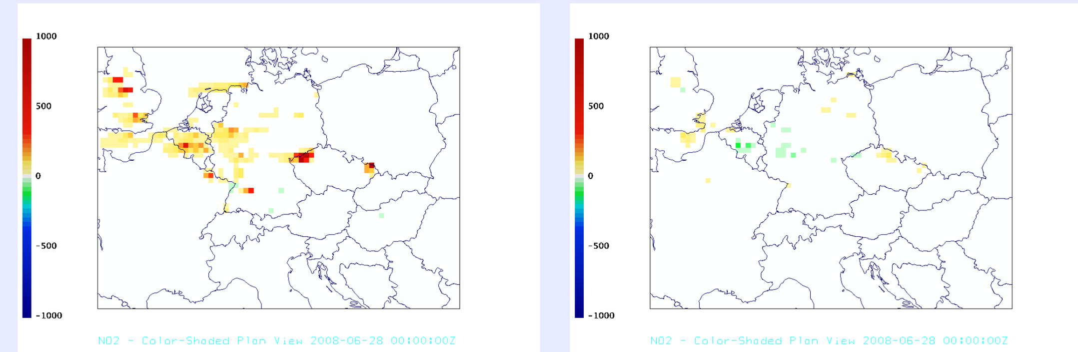
After the first two days of assimilation, the emission factors converged to a fairly stable solution. The changes in some regions (e.g. France, Italy, Austria) can be attributed to satellite observations only.



NO2 emission optimization factors on July 03 2008 (left)
The average of the emission opt. factors from June 29 to July 05 2008 (right)

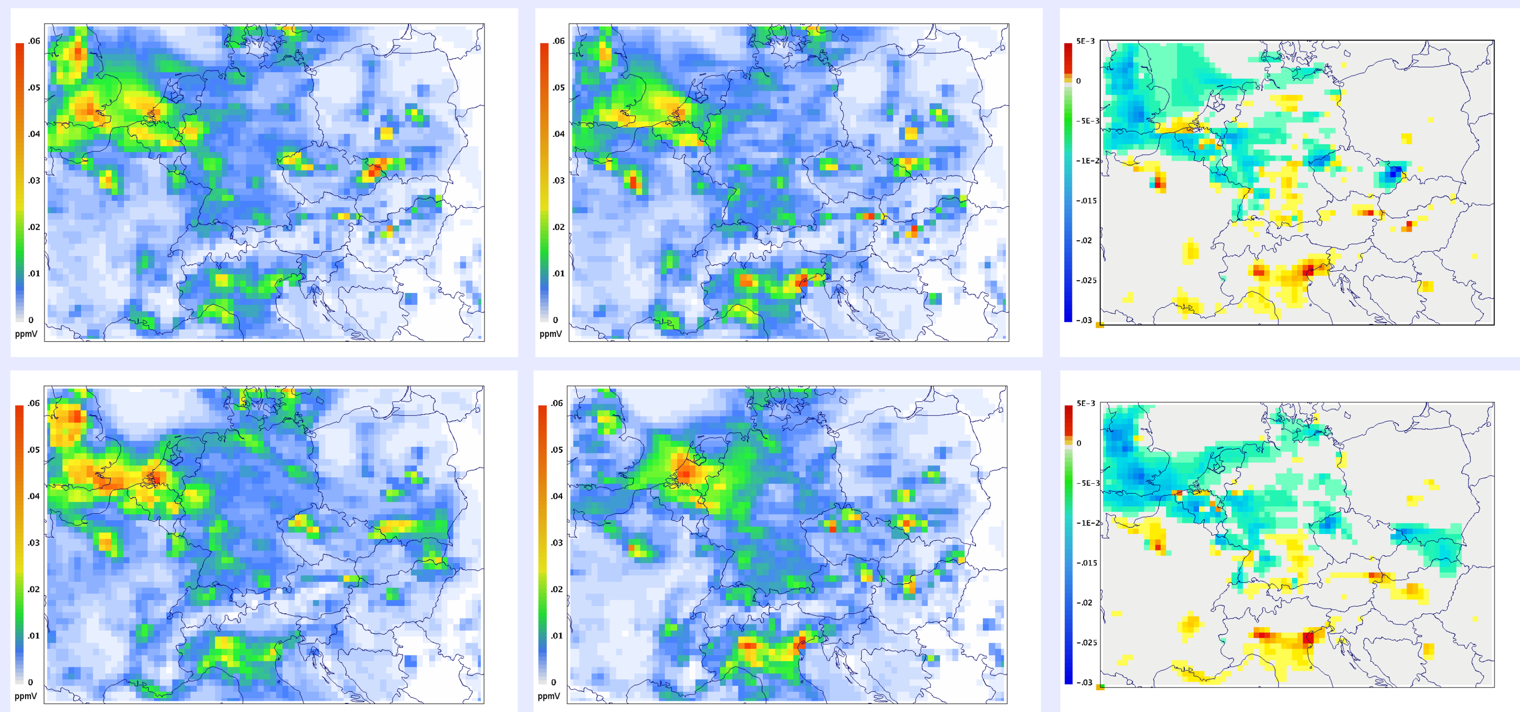
Gradient of the cost function

Gradient of the cost function with respect to the emission factors at first iteration is steep at the first day of the assimilation and falls down during the following days (the maximum of the ratio is almost 100:1).



Gradient of the CF with respect to emission parameters at first iteration.
28.06.2008 (left), 01.07.2008 (right).
Remark: An unified scale is used. The peak values in the left map are 6880 and in the right one 78.

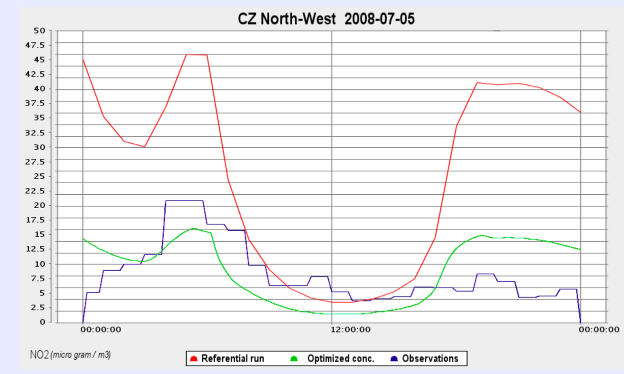
NO_2 concentrations



Ground level NO2 concentrations (ppmV) on July 1 2008 (upper) and July 5 2008 (lower) at 20:00
Referential free run (left), optimized run (middle) and differences (right)

Ground station diurnal profile

The multiplicative emission correction factors do not provide enough freedom for modeling time profiles properly. A more general parameterization is necessary, but the stability of the solution must be kept.

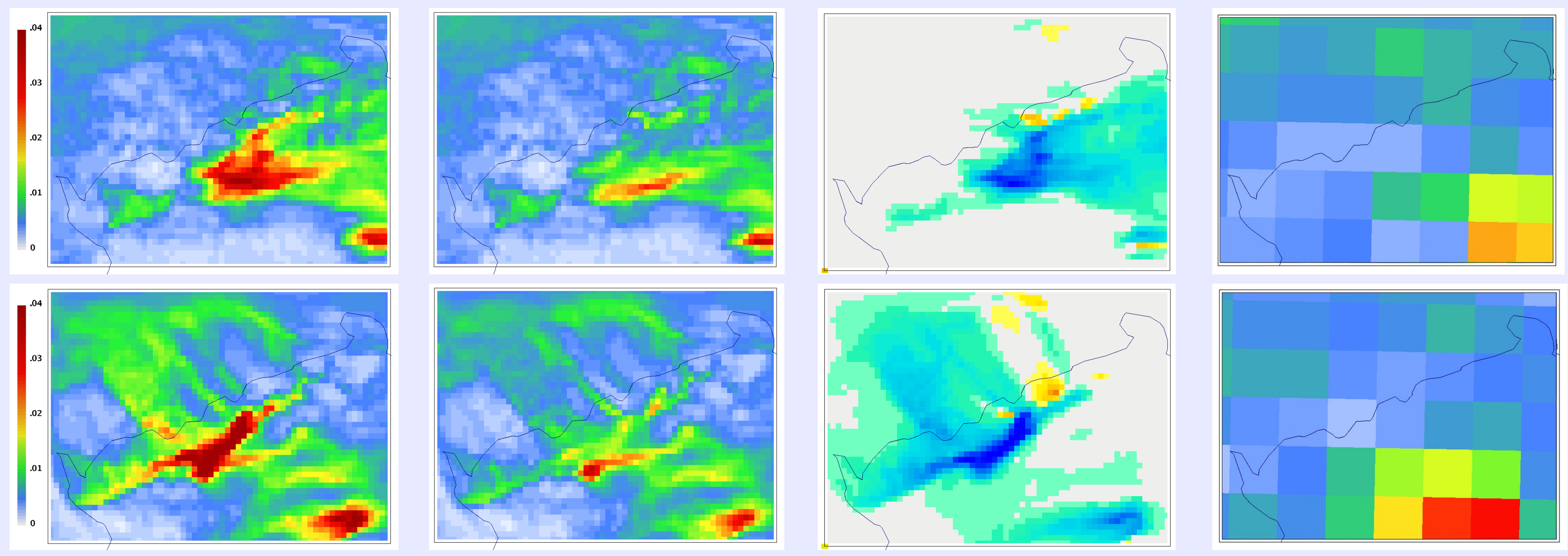


Concentrations on obs. station Litoměřice
(North-West Bohemia, the strongest changes induced by the assimilation).

Fine domain

Ground level concentrations

The fine domain covers a region with large spatial variability in concentrations of NO_2 and complex orography. Fine resolution gives better opportunity for modeling and optimization of NO_2. The emissions in this region decreased in last years which is not fully reflected in the emission inventory.



Ground level NO2 concentrations (ppmV) at June 28 (upper) and July 1 2008 (lower)
From left to right: Referential free run, optimized run, differences of opt. run and ref. run, zoom of the optimized run on coarse domain

Conclusions and outlook

The assimilation experiment shows the stability and good performance of this particular setting of the 4DVar method. The satellite data seem to be useful for constraining emissions and initial conditions even in the short term and fine resolution. The improvement in forecasting NO_2 concentrations due to assimilation of in situ and satellite observations is moderate so far. Longer experiments, more flexible parameterization of the emission corrections, adaptations of the emission model itself and a deeper investigation of complementarity between in-situ and satellite observations are required before using the technique in operational forecasting.

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".