17th International Conference on
Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes
9-12 May 2016, Budapest, Hungary

OPTIMIZING INITIAL VALUES AND EMISSION FACTORS ON MESOSCALE AIR QUALITY MODELLING USING 4D-VAR DATA ASSIMILATION

Isabel Ribeiro[1], Zoi Paschalidi[1] and Hendrik Elbern[1,2]

[1] Rhenish Institute for Environmental Research (RIU-EURAD) at the University of Cologne, Germany.
[2]Research Centre of Jülich, Institute for Energy and Climate Change – Troposphere (IEK-8), Germany.

Abstract: The use of models to analyse the complex atmospheric physical and chemical processes deals with significant uncertainties of key parameters and input information, most prominently emission rates and chemical background information. In this frame, the four-dimensional variational (4D-var) data assimilation scheme for gas-phase and aerosols, encoded in the EURopean Air pollution Dispersion – Inverse Model (EURAD-IM), was used to improve concentrations predictions of gaseous and aerosol species in the troposphere through corrections on emission rates jointly with initial values. The Po valley region was defined as a case study due to the availability of PEGASOS campaign observational data measured by the Zeppelin NT instruments. Observed data from the campaign, ground stations and satellite retrievals were assimilated by the EURAD-IM system, resulting in a model performance improvement not only at the surface level but also in height within the planetary boundary layer (PBL).

Key words: *Data assimilation, inverse modelling, initial values and emission factors optimization*

INTRODUCTION

The complexity of atmospheric models has been rising with the increase of computational power and scientific knowledge, leading to higher spatial and temporal resolutions as well as more sophisticated physical and chemical processes. However, to obtain a good prediction it is also needed to start from reasonable initial values estimation. The provision of a set of optimum initial values (IV) by employing as much observational data as possible is the classical goal of data assimilation, combining observations with physical and chemical knowledge of atmospheric processes encoded in the numerical models (Kalnay, 2003).

Besides the initial values, the application of space-time-dimensional variational data assimilation (4D-var DA) also allows the estimation of emission factors (EF) due to the comprised inverse modelling technique (Elbern et al., 2007). The 4D-var DA algorithm propagates the model forward and backward in time fitting the model simulation to the set of observations, distributed in a predefined time interval (assimilation window). By this way, the model is able to calculate the system state that lays in the minimum distance between model and observations: the analysis state. The consistency of this system state is guaranteed by the inverse (or adjoint) simulation of the emitted species and their products. The sophisticated 4D-var DA scheme for gas-phase and aerosols included in the EURAD-IM (Elbern et al., 2007), is used to define an optimized set up to apply 4D-var assimilation of ground based measurements and satellite retrievals data. The optimal 4D-var application is the key to quantitatively estimate anthropogenic and biogenic pollutant concentration patterns, as well as to understand their interactions at a given air-shed, by the provision of improved IV and EF of gas-phase and aerosol species.

The EURAD-IM is an Eulerian mesoscale chemical transport model (CTM) that includes the Modal Aerosol Dynamics Model for Europe (MADE, Ackermann et al., 1998) to simulate the aerosol dynamics. The meteorological driver that is applied in the EURAD-IM is the Weather Research and Forecasting Model (WRF, Skamarock et al., 2008). For this work, the Po valley region was defined as a case study due to its high density of anthropogenic emissions and frequent occurrence of stagnant meteorological conditions that promote recurrent episodes of high air pollutant concentrations. The optimized model set up was used to assimilate data measured by the Zeppelin NT borne instruments of the PEGASOS campaign over the same region, during 10-12.07.2012 (Li et al., 2014). The added value from the combination of PEGASOS campaign data by the 4D-var DA scheme is the provision of improved model results in the boundary layer. The corrections provided by the observed data to the numerical modelling system are explored in this work.

4D-VAR DATA ASSIMILATION TECHNIQUE

Data assimilation methods use all the available information about the system to provide an as accurate as possible and consistent image of a system's state at a given time (the optimal guess). Within 4D-var DA the analysis problem is formulated as a minimisation problem using variational calculus. The definition of a cost function (J, Eq. 1) is necessary, aiming the calculation of the distance between the model simulation and the observations, during a predefined time interval (assimilation window) (Kalnay, 2003). J is derived from the properties of the mapping between the model space and the observational space (the forward model $\boldsymbol{H}$) and from the prognostic model $\boldsymbol{M}$ itself. Due to the fact that 4D-var algorithm propagates information forward and backward in time, it is regarded a smoother, fitting a model simulation to a set of observations distributed in a predefined assimilation window.

$$\begin{aligned} J(x_0, e_0) &= J_{iv} + J_{obs} + J_{ef} \\ &= \frac{1}{2}\left([x_0 - x_b]^T\ \mathbf{B}^{-1}[x_0 - x_b] + \sum_{i=0}^{N}\left([HM_i x_0 - y_1]R^{-1}[HM_i x_0 - y_1]\right) + [e_0 - e_b]^T\ K^{-1}[e_0 - e_b] \right) \end{aligned} \quad (1)$$

Encoded on the EURAD-IM modelling system, the 4D-var DA scheme is able to optimize IV and EF benefiting from their individual impact on the model evolution (Elbern et al., 2007). The IV play an important role in the beginning of the assimilation window, while the EF optimization represents an influence through all the assimilation window. Consequently, an optimal result is given by joint optimisation of both parameters. In this sense, the total cost (J) results from the sum of three individual costs: the background cost of the initial state of the chemical constituents (J_{iv}), the observational costs (J_{obs}) and the cost of the emission inventory, expressed as EF (J_{ef}). J_{iv} and J_{ef} are, respectively, calculated taking into account the error covariance matrixes of the initial state of the chemical constituents ($\boldsymbol{B}$), the emissions ($\boldsymbol{K}$) and the observations ($\boldsymbol{R}$).

METHODOLOGY AND CASE STUDY DESCRIPTION

The 4D-var DA algorithm described in the previous section is applied over the Po valley region, during the Pan-European Gas-AesoSOls Climate Interaction Study (PEGASOS) flight campaign, on 10-12 July 2012. This case study is partly motivated by the necessity to evaluate the 4D-var data assimilation analysis performance in a highly resolved grid over polluted areas, such as Po valley, as well as the new opportunities provided by the airborne data to the model in terms of: distinguishing emission patterns, investigating the vertical distribution of trace gases and the PBL dynamics. To this end, the current study applied the 4D-var data assimilation method by EURAD-IM. For that, a sequence of three domains (Figure 1) is used starting with the coarse grid covering Europe, followed by the central Europe and finally the Po valley area, with 15×15, 5×5 and 1×1 km^2 of horizontal resolution, respectively. Further, the optimisation of EF in nested grids of 1×1 km^2 resolution is studied, addressing the issue of the

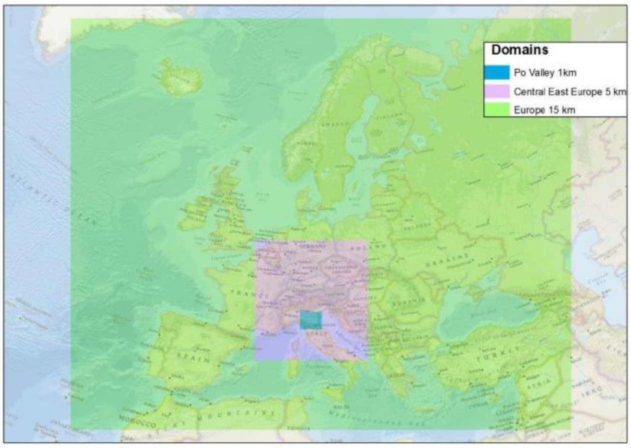

Figure 1. The sequence of nests in EURAD-IM. The coarse grid is Europe with 15×15 km^2 resolution which includes two nest domains: the central Europe and the area of Po Valley in Italy with 5×5 km^2 and 1×1 km^2 resolution, respectively.

representatively of observations such as NO_2. Taking advantage of the high quality airship campaign measurements, special focus is given on the analysis of the vertical mixing in the PBL with EURAD-IM.

The set of observations used here consists of ground based and satellite measurements as well as PEGASOS campaign data. AirBase measurements of NO_2, O_3, NO, SO_2 and CO were combined with the O_3 and CO observations from MOPITT and NO_2 tropospheric columns of GOME-2. The vertical profiles of NO_2, O_3 and CO from the flight campaign over the rural supersite of San Pietro Capofiume (SPC) were also used.

For aerosols, the 4-Dvar scheme included in the EURAD-IM modelling system is at an earlier stage of development than for the gas-phase. Currently, it only includes the IV optimization, the EF optimization being in a development stage. The study here presented is focused on the European domain (15x15 km^2) and comprehends assimilation of above-mentioned species plus PM10, PM2.5 and NH_3 from the AirBase network as well as aerosol optical depth (AOD) retrievals from MODIS satellite. Based on this case study, the performance of the EURAD-IM 4D-var data assimilation scheme with joint optimisation of IV and EF for gas-phase, and the IV optimization for aerosols, is assessed in the following section.

CASE STUDY ANALYSIS

The accuracy of the EURAD-IM 4D-var data assimilation scheme is assessed by highly resolving grids, together with the joint optimisation of IV and EF for gas-phase. Figure 2 underlines the benefit from the joint optimisation, illustrating the differences between the analysis result and the model's first guess in case of NO_2 concentrations and the NOx emission factors correction, during the first day of assimilation (10.07.2012). The increased optimization of NOx emission factors, up to a factor of 4, correct the general underestimation of the model for the NO_2 concentration. Moreover, as the domain is 1 km grid spacing, emission patterns can be resolved and the problem of the representatively of NO_2 observations is diminished in comparison with less resolved grids (not shown here).

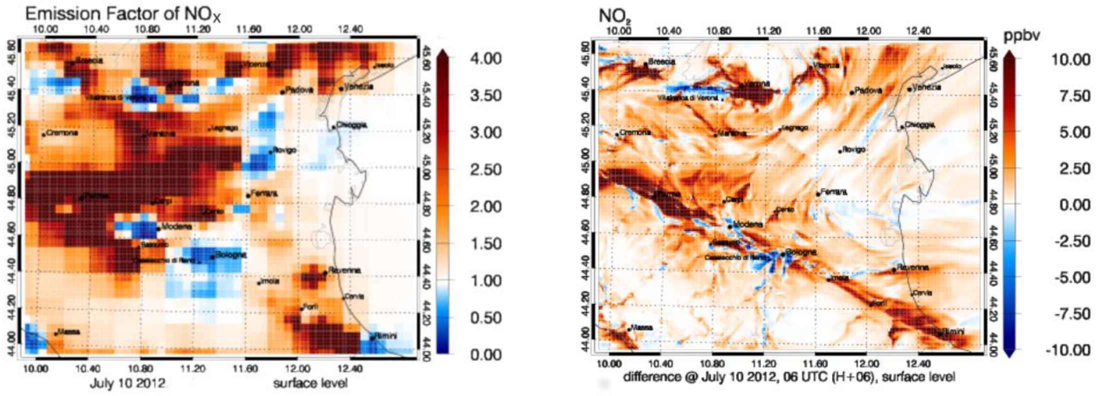

Figure 2. Differences between the analysis and the background for NO_2 during morning rush hour (06:00 UTC) over Europe (left panel) and over the Po valley area (right panel), on 10.07.2012.

Validating the analysis output for each resolution, a comparison with independent observations took place. Figure 3 depicts the timeseries of NO_2 and O_3 concentrations for the analysis result of the three different domains against measurements of two stations in Italy that are not included into the assimilation procedure. For both the emitted NO_2 and its product O_3, the optimisation is more efficient for the finest nest grid (blue curve). The good representativeness of NO_2 observations for the 1 km resolved grid is obvious, fact that also influences the analysis of O_3, since the system maintains its chemical consistency. The afternoon peak of NO_2 analysed concentrations in case of the second nest (blue curve) presumably comes from the assimilation of observations from other stations over areas with higher traffic activity, fact that is not possible to recognised in case of the less resolved grid.

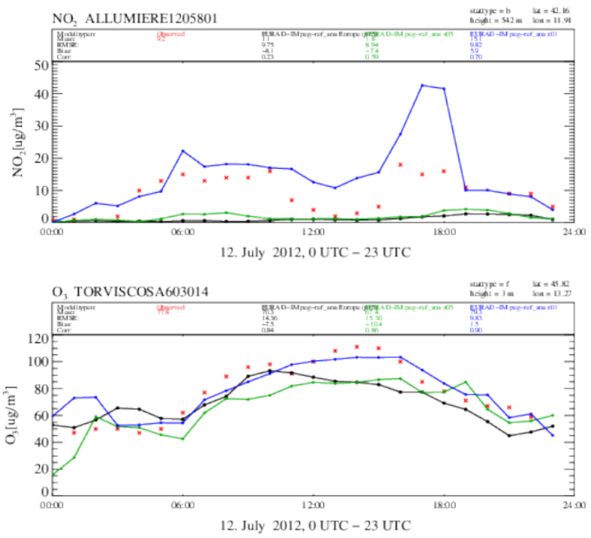

Figure 3. Timeseries of NO_2 (upper panel) and O_3 (lower panel) concentrations regarding the three domains, for two ground stations non-assimilated observations (Allumiere and Troviscosa, in Italy). Observations are given in red; the analysis are given in black for the European domain, in green and blue for central Europe and Po valley, respectively.

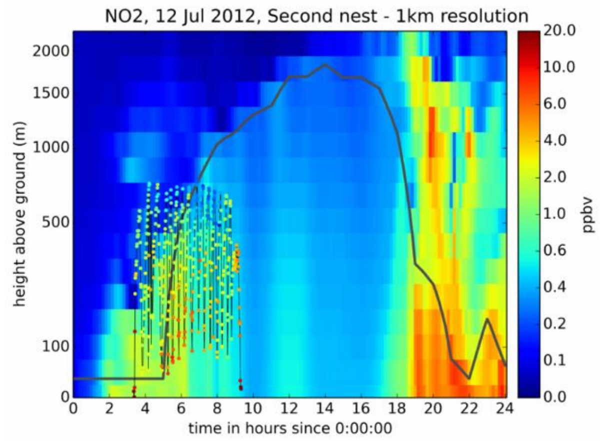

Figure 4. HovmØller plot. Timeseries of PBL and the vertical NO_2 concentrations at SPC: observations from the flight campaign (dots) and model results from concerning the Po valley domain (background).

Figure 4 depicts the HovmØller plot for the assimilation of the PEGASOS airborne observations, on 12.07.2017, over the SPC supersite, taking into account the finest nest (1×1 km^2 of spatial and 25 seconds of temporal resolution). Compared with the analysed concentrations (background colour), there is a match with them and the airborne data in upper altitudes (500-700m) from the beginning of the flight until around 8:00, as well as at close to 300 m until 6:00. On the other hand, in the model's mixed layer, the observed NO_2 concentrations are higher than the analysed ones, up to 300-400 m. The Zeppelin's observations capture clearly the layered structure of the PBL (see also Li et al., 2014), however differently than the model. In other words, the campaign data underline that the model calculates a higher PBL than it is measured. Although there is a correction of the analysed concentrations towards the observations, this does not influence the calculation of the PBL by the model. Thus, during the model analysis the mixing takes place in higher altitudes.

Regarding the aerosols, the optimal dimension of the assimilation window was tested, taking into account only IV optimization for both gas-phase and aerosols. In other words, how many hours of observations are needed to provide improved model information with as less computing time consumption as possible? Two ranges of assimilation windows were tested: 12h and 6h which allowed to verify that the shortest assimilation window provides better analysis reducing the root mean square error in 8%. Nevertheless, the better analysis achieved was still far of the observations (observed mean value is 19.5 µg.m^{-3} while for analysis (6h) it is 6.7 µg.m^{-3}). As aerosols formation is highly dependent of gas-phase, the previous best analysis (6h assimilation window) was compared against the analysis from 24h of assimilation window with optimization of IV (gas-phase and aerosols) and EF from gas-phase. Combining gas-phase with aerosol optimisation improved the analysis in 22% (Figure 5).

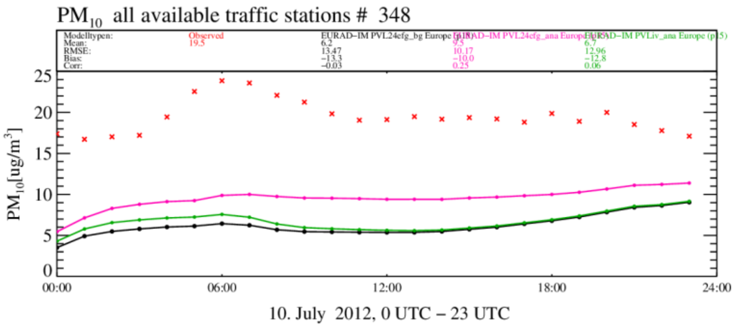

Figure 5. Timeseries for PM10 concerning the average of all AirBase station with traffic influence over the European domain. Observations are given in red, background in black; analyses with 6h of assimilation window in green and with joint optimization of aerosol and gas-phase are in pink.

CONCLUSIONS

The joint optimisation of IV of the chemical constituents and EF has been successfully employed for a 24 hours assimilation window, letting the better-known diurnal emission profiles be considered as strong constraint. The high resolution nesting technique is shown to face the representatively of NO_2 observations in the finest grid, being able to identify traffic emissions and more accurate emission patterns. Beyond that, it was verified that rich campaign measurements have been essential to the model analysis as they give a more detailed insight than the routine data to the horizontal and vertical dispersion of the emissions in polluted areas. Regarding aerosols, the IV optimisation has shown as effective, the joint EF optimisation is crucial to an accurate prediction, however.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge the computing time granted by the John von Neumann Institute for Computing (NIC) and provided on the supercomputer JURECA at Jülich Supercomputing Centre (JSC).

REFERENCES

Ackermann, I.J., Hass, H., Memmesheimer, M., Ebel, A., Binkowski, F.S., Shankar, U., 1998. Modal aerosol dynamics model for Europe. *Atmos. Environ*. **32**, 2981–2999. doi:10.1016/S1352-2310(98)00006-5

Elbern, H., Strunk, A., Schmidt, H., Talagrand, O., 2007. Emission rate and chemical state estimation by 4-dimensional variational inversion. *Atmos. Chem. Phys*. **7**, 3749–3769. doi:10.5194/acp-7-3749-2007

Kalnay, E., 2003. Atmospheric modeling, data assimilation, and predictability, *Annals of Physics*. doi:10.1256/00359000360683511

Li, X., Rohrer, F., Hofzumahaus, A., Brauers, T., Häseler, R., Bohn, B., Broch, S., Fuchs, H., Gomm, S., Holland, F., Jäger, J., Kaiser, J., Keutsch, F.N., Lohse, I., Lu, K., Tillmann, R., Wegener, R., Wolfe, G.M., Mentel, T.F., Kiendler-Scharr, A., Wahner, A., 2014. Missing gas-phase source of HONO inferred from Zeppelin measurements in the troposphere. *Science* **344**, 292–6. doi:10.1126/science.1248999

Skamarock, W.C., Klemp, J.B., Dudhia, J., Gill, D.O., Barker, D.M., Huang, X.Y., Wang, W., Powers, J.G., 2008. A Description of the Advanced Research WRF Version 3.