Atmos. Chem. Phys. Discuss., 8, S9308–S9315, 2008
www.atmos-chem-phys-discuss.net/8/S9308/2008/


Atmospheric Chemistry and Physics Discussions



# *Interactive comment on* "$PM_{10}$ data assimilation over Europe with the optimal interpolation method" *by* M. Tombette et al.

**M. Tombette et al.**

Received and published: 20 November 2008

We thank the reviewer for her/his detailed comments that helped us in improving the paper quality. Besides, we carried out significant improvements in the paper, especially on the estimation of the error variances, as also suggested by the second reviewer. Hereafter, the comments of the referee are quoted in italic.

**General Comments**

1. *The paper could benefit from a more in-depth discussion of the motivation for pursuing aerosol assimilation in the introduction.*

*-There must be a better motivation to assimilate aerosol than the fact that aerosols are now a component of chemical transport model: please elaborate on that.*

The reviewer is right, the motivations of the study were partially reported. We added a

discussion about the models performance for gases and for aerosols. Also, we mentioned the performance of purely statistical models, which are essentially based on observations. In the introduction, the following text has been added: "Performance obtained with various models for $PM_{10}$ could be disappointing in comparison to the performance obtained for gas species. For example, the correlation of simulated concentrations with observed concentrations rarely exceeds 50% for hourly $PM_{10}$ over Europe (Van Loon et al., 2004), and the Root Mean Square Error (RMSE) is of the order of the concentration (i.e., 10 $\mu g\ m^{-3}$). On the contrary, ozone peaks forecasts, for example, show correlations that may exceed 70% or 80% and RMSEs around 20% of the concentrations. Meanwhile, performance obtained with purely statistical models for $PM_{10}$, which are mainly based on observational data, is much more impressive: the correlations in (Hooyberghs et al., 2005) are approximately 70%. It is therefore relevant to investigate data assimilation with physical models for aerosols, in order to take into account the information contained in measurements."

- *What do the authors mean by saying that the limit for $PM_{10}$ to exceed 50 $\mu g\ m^{-3}$ is set to 35 days per year? Who sets the limit and how is this limit enforced?*

This is a recommendation made by the European Commission. The directive associated to this limit is now mentioned in the text.

2. *The authors mention in the abstract that they would discuss the use of variational methods for assimilation in the conclusions but this item is missing.*

We removed the corresponding sentence.

3. *The conclusions are also too dismissive. It would be useful for the reader to hear more ideas from the authors on how they would improve their assimilation system or whether they propose alternatives to optimal interpolation, especially in regards to the temporally localised impact of the assimilation.*

In the conclusion, we already pointed out possible follow-ups (quoted from the paper):



**ACPD**

8, S9308–S9315, 2008

“

1. Using more sophisticated methods to build the background covariance matrix, such as methods based on statistical studies of the simulated fields (e.g. the Hollingsworth-Lönnberg method, (Daley, 1993));
2. Implementing inverse methods in order to improve the quality of input data (emissions) and/or parameterizations;
3. Assimilating observations of gases that are seldom measured but important for the formation of secondary inorganic species, like nitric acid ($HNO_3$) or ammonia ($NH_3$);
4. Assimilating observations of the aerosol chemical composition (nitrate, sulfate, ammonium, primary and organics); the bias existing for some species could then be lowered;
5. Assimilating optical data from a lidar network, which could improve the vertical distribution of aerosols and, as a result, improve the persistence of DA impacts over the domain. ”

*Also missing in the conclusions is the discussion of an operational feasibility of implementing inverse methods to improve the emissions.*

Before discussing the operational implementation of a joined state/parameter estimation, we should first test it in off-line mode. It is difficult to draw any reliable conclusion out of the study we performed with the state controlled alone.

The conclusion *needs to be expanded (see also general comments). The authors should discuss the merits and shortcomings of the optimal interpolation approach with respect to choosing other methods, as announced in the abstract.*

We removed from the abstract the announce about the comparison with other methods. This should be left to a more complete study where the other methods would be actually applied.

We simply added a note about this in the conclusion: “The OI method applied in this paper could be used in operational mode: the model is already running



**ACPD**

8, S9308–S9315, 2008

for real-time forecasts over Europe, in the context of tests on the Prév'air platform (http://www.prevair.org/), and the additional computational cost due to OI is slight. Applying other methods like 4D-Var (providing an adjoint model is available) or an ensemble Kalman filter (providing proper uncertainty information is available on inputs) would be much more demanding in terms of computational resources, and it would not necessarily improve the data assimilation efficiency if the state is to be solely controlled (for ozone, see Wu et al. 2008)"

4. *The style of the paper is, however, too synthetic, and there is an excessive use of parenthesis which makes some parts of the paper hard to read and interpret.*

We agree that the paper was difficult to follow. We dropped many brackets in order to improve the flow of the paper. We tried to extend the comments anywhere they could be ambiguous or little convincing as such.

noindent**Specific Comments**

1. *The choice for the observation error in the baseline assimilation experiment is disputable. Even if the intention of the authors is to ensure that the assimilation draws mainly to the observations, it is not realistic to choose an observation error that is lower than the instrument error. Later on, in section 6, the authors show that even with a larger error (case with $\alpha$ = 1), the impact of the assimilation is still positive. Moreover, the larger observation error does not impact at all the time extent of the influence of the data assimilation on the forecast, but only the amplitude of the improvement with respect to the free-running model.*

Yes, there is no need for an underestimation of the observational variance. In any case, the limiting factor remains the same: the impact of a change in the initial conditions vanishes rather quickly.

We completely reconsidered this aspect, also after the comment of the second reviewer. The variance for the observations is now determined based on the $\chi^2$ diagno-



**ACPD**

8, S9308–S9315, 2008

sis. And it is obviously much higher than the instrument error.

2. *Looking at equation (6) and Table 3, I cannot reconcile that in the comparison with the AirBase data the statistics for the single-species concentrations are not improved at all, while for PM10 (and PM2.5 ) they are. I would understand better if there was some aliasing of the increment in PM10 into an increment on one single species, namely the main contributing species to PM10 at the particular station, but table 3 shows no such as thing. Do the authors have an explanation for this behaviour other than the lower number of stations for the concentration comparison with respect to the PM10 comparison which could skew the statistics? Have the authors looked at the values of concentrations station by station? Given the low number of stations that should not be too hard, and it could provide some insights into this behaviour.*

The stations that provide the chemical data are different from the stations that provide the $PM_{10}$ data. It is therefore not possible to compare station by station the results for $PM_{10}$ and for a specific species. However, the reviewer is right, we can not claim that the statistics for single species are not improved. We corrected by introducing the following paragraph in Sect. 4: “The statistics for sulfate, chlorine and sodium are slightly better with DA. On the other hand, the statistics for nitrate are deteriorated. For ammonium, the statistics are stable. Actually, the model underestimates the $PM_{10}$ over the period, so DA tends to add material to the existing aerosol mass. As the repartition over the chemical species is homogeneous, DA tends to add mass to all species. Then, the species that were overestimated at first, like nitrate and ammonium, are even more overestimated. Sulfate was underestimated, so DA allows its concentration to be greater and the statistics are improved. The overestimation of the nitrate concentration plays a role in thermodynamic equilibrium, by reducing the mass of chlorine that condensates on the particles. The change of the sodium concentrations is essentially due to DA. The number of stations that provide measurements for chemical species is lower than the stations providing $PM_{10}$ observations and these stations maybe not taken into account for the $PM_{10}$ statistics. It is therefore difficult to draw conclusions



**ACPD**

8, S9308–S9315, 2008

about a general behavior. Moreover, the overestimation for nitrate concentrations is specific to winter conditions (Sartelet et al., 2008)."

3. (From the general comments:) *For example in the operational application, the authors show that the impact of the assimilation is only evident in the improved statistics for the the one-day forecast, whereas the two-day forecast is not much better than the free-running simulation. Is this a limit of the optimal interpolation technique or is it more intrinsic to the assimilation of concentration data only, regardless the assimilation approach?*

The following sentences have been added in the paper (Section 5): "The fact that DA with the OI method has some influence only during such a short period of time (one day in this experiment) is not only a limit of the OI method. Actually, in air quality models, concentrations are not much influenced by initial conditions. Also, it depends on several parameters (see section 6) and on the pollutant. For ozone, the influence is a bit longer (see Wu et al., 2008)."

**Technical Corrections**

The suggested technical corrections have been scrupulously applied, so we do not list them all.

- *What is the "granulometry" of the model?*

The word "granulometry" has been replaced by "size distribution" in the text.

- *What does "mechanism" means in this context? Please reword.*

"Mechanism" is the general term to refer to the considered chemical species and their chemical reactions. We refer to the paper introducing RACM for further details, and we now write: "The Regional Atmospheric Chemistry Mechanism (RACM, Stockwell et al., 1997) is used to simulate chemistry".

- *Page 9621, line 03: reformulate sentence - it is not clear what is meant by "the model*



**ACPD**

8, S9308–S9315, 2008



*shows the same uncertainties for all species".*

The sentence has changed to : "it is then supposed that the model uncertainties are equivalent for all species".

*- Explain or provide a reference why "the observations depend" on "organic species".*

These sentences have been added, with references : "Actually, the mass of low volatile material in aerosols is generally not or only partially measured by the instruments used in automatic stations. In most of $PM_{10}$ measurements instruments, the sampling are heated to measure the mass of dry particulate matter. For example, the sample temperature of a TEOM (Tapered Element Oscillating Microbalance), which received the certification of US-EPA, is 50°C. At this temperature the volatile material, such as organic aerosols, is evaporated (Allen et al., 1997; Smith et al., 1997; Salter and Parsons, 1999; Soutar et al., 1999; Green et al., 2001; Joseph et al., 2001; Charron et al., 2003). We assume that the observations slightly depend on organic species and then, we cannot assume that the difference in total mass is only attributed to organic species."

*- The sentence "For example, the EMEP...statistics" is not clear, please reformulate.*

The sentence has been modified to : "For example, the EMEP database only contains background stations. The data assimilated in this study could be provided by urban or suburban stations, that have higher concentrations. The forecast then tends to increase the concentrations accordingly, spoiling the statistics for rural stations as for the entire EMEP database."

*- Figure 1: from the text it is not clear what the "background" stations are. Does it mean that they are located in rural, low-traffic locations? Please explain.*

The following sentence has been added in Sect. 3.2 : "the station characteristics are specified: background (rural), suburban, urban, industrial and traffic. The scale at which a station is representative depends on the station type. In the assimilation



**ACPD**

8, S9308–S9315, 2008



procedure, the traffic and industrial stations, generally with high concentrations due to the proximity of sources, have been removed."

- *Figure 2: missing labels (a) and (b) in the plots. Please explain how the circle diameters are proportional to the statistical indicator.*

A legend has been added to Figure 2.

---

Interactive comment on Atmos. Chem. Phys. Discuss., 8, 9607, 2008.



**ACPD**

8, S9308–S9315, 2008

Interactive Comment