Atmos. Chem. Phys. Discuss., 8, 9607–9640, 2008
www.atmos-chem-phys-discuss.net/8/9607/2008/

Atmospheric Chemistry and Physics Discussions

PM_{10} data assimilation over Europe with the optimal interpolation method

M. Tombette[1,2], **V. Mallet**[2,1], **and B. Sportisse**[1,2]

[1]CEREA, Joint Laboratory ENPC–EDF R&D, Université Paris-Est, Marne la Vallée, France
[2]INRIA, Paris-Rocquencourt Research Centre, France

Received: 26 March 2008 – Accepted: 9 April 2008 – Published: 27 May 2008

Correspondence to: M. Tombette (tombette@cerea.enpc.fr)

Published by Copernicus Publications on behalf of the European Geosciences Union.

Abstract

This paper presents experiments of PM_{10} data assimilation with the optimal interpolation method. The observations are provided by BDQA (Base de Données sur la Qualité de l'Air), whose monitoring network covers France. Two other databases (EMEP and AirBase) are used to evaluate the improvements in the analyzed state over one month (January, 2001) and for several outputs (PM_{10}, $PM_{2.5}$ and chemical composition). Then, the method is applied in operational conditions. The results show that the assimilation of PM_{10} observations significantly improves the one-day forecast for total mass (PM_{10} and $PM_{2.5}$). The errors on aerosol chemical composition are not reduced and are sometimes amplified by the assimilation procedure, which shows the need for chemical data. As the observations cover a limited part of the domain (France versus Europe) and as the method used for assimilation is sequential, we focus on the horizontal and temporal impacts of assimilation in the last part of this paper. To conclude, we discuss the perspectives, especially the use of a variational method for assimilation or the investigation of the sensitivity to a few choices (e.g., the error statistics, etc.).

1 Introduction

State-of-the-art models, in meteorology or in air quality, reasonably approximate the atmospheric state (meteorological fields and chemical composition). However, there are still a lot of uncertainties in modeling (Hanna et al., 1998; Mallet and Sportisse, 2006), and in particular for aerosols (Roustan et al., 2008), leading to important discrepancies with observational data.

Data assimilation (DA hereafter) makes use of measurements of parts of the model state in order to reduce the uncertainties in input data such as initial conditions or boundary conditions. In some cases (especially for air quality modeling) the purpose of the modeler may be the evaluation of the emissions fluxes, and not necessary the improvement of the forecast itself. This defines the so-called inverse modeling issue,

ACPD

8, 9607–9640, 2008

which is not addressed in this paper.

DA is based on several theories that provide the basis for many existing methods: statistical methods (e.g., the optimal interpolation), variational methods (3D- and 4D-Var) and sequential methods (Kalman filters). We refer to Bouttier and Courtier (2001) for an overview.

DA could be applied with different objectives: to produce an analysis, in other words to compute a field as close as possible to the "true" state; to improve the initial conditions in order to improve the forecasts; or to identify uncertain parameters (inverse modeling), such as the emission fluxes.

DA is a relatively recent research field in atmospheric chemistry (Austin, 1992; Fisher and Lary, 1995; Riishojgaard, 1996), whereas it has been widely applied in meteorology. However, numerous studies were carried out for specific gases and with measurements of diverse nature (in-situ, airborne, satellites). Ozone columns are assimilated with an optimal interpolation approach in Jeuken et al. (1999); Planet (1984), and with a 4D-Var approach in Riishojgaard (1996). In Levelt et al. (1998), ozone and carbon monoxide profiles are assimilated with the Mozart model. Elbern et al. (1997), Elbern et al. (2007) and Segers (2002) present several assimilation studies of terrestrial data with a 4D-Var approach and a sequential approach for the latter. Wu et al. (2008) compare four assimilation methods (optimal interpolation, ensemble Kalman filter, reduced-rank square root Kalman filter and 4DVar) for assimilation of ozone ground measurements.

Aerosol models are now a component of most of available chemistry-transport models (CTMs). It is therefore relevant to investigate data assimilation for aerosols. In the case of the aerosol model SIREAM (SIze-REsolved Aerosol Model, Debry et al., 2007) embedded in the POLYPHEMUS platform Mallet et al. (2007), which is used in this study, the assimilation method is constrained by the use of a thermodynamic model (ISORROPIA, Nenes et al., 1998). Indeed, there is currently no adjoint model for SIREAM: first, the thresholds set by ISORROPIA (discontinuous model due to phase transitions) raise a theoretical problem (definition of the derivatives at the discontinuity points);

second, the code of ISORROPIA has not been written so that it may be automatically differentiated (and in practice this leads to large difficulties). A variational approach is thus not conceivable at the moment.

Aerosol measurements seldom correspond to model state variables (which, for a size-resolved model, give the size distribution of the aerosol chemical composition). They are often aggregated data (PM_{10}, $PM_{2.5}$) or optical data (extinction coefficient, optical thickness). It is therefore not straightforward to assimilate this data. As the current standards for air quality regulation focus on PM_{10} mass and as the networks giving PM_{10} data are the most widely extended, we chose to assimilate these observations.

Since policies about particulate matter have mainly focused on PM_{10} up to now, it is important to accurately forecast their concentrations, primarily in urban areas. For example, in Europe, the limit is set to a maximum of 35 days per year during which PM_{10} exceeds $50\,\mu g\,m^{-3}$. However, the numerical models hardly reproduce the highest PM_{10} peaks, because of inaccurate descriptions of exceptional events (Saharan dust episodes, dust resuspension in dry conditions, etc.). DA could be useful in order to compensate for these model deficiencies.

In this paper, we use a simple method for PM_{10} DA, namely the optimal-interpolation method. This method was applied for the assimilation of aerosol optical thickness (Generoso et al., 2007; Collins et al., 2001). Variational methods were also investigated, like the 1D-Var method in Huneeus (2007) or the 4D-Var method in Benedetti and Fisher (2007), but for simplified aerosol models.

The objectives are to sketch answers to the following questions:

1. Is there an improvement for forecasts after the assimilation of PM_{10} hourly data over Europe? If so, to what time extent?

2. Does the analysis produced by the PM_{10} assimilation over a subdomain (here, France) result in an improvement over the remaining part of the domain (western Europe)?

3. Does the analysis produced by the PM_{10} assimilation improve the computed

chemical aerosol composition? Should the measured concentrations of precursor gases be assimilated?

The first question has no obvious answer. First, the thermodynamic equilibria between the aerosol and gaseous phases could annihilate the corrections after DA if the gaseous concentrations are not corrected meanwhile. Second, the aerosol residence time in the atmosphere is relatively low (a few days); particles are known for having mostly a regional effect. So, it may be necessary to perform a joint state-parameter estimation. For instance, it would be useful to make corrections on the aerosol emissions.

The second question is related to the timescales of atmospheric transport. It consists in determinating if the improvements in a part of the domain (where most of the measurements are available) can improve the simulation in places where the wind transports the air masses. This could be important for regions where only sparse data is available. The choice of the scale parameters L_h and L_v (see Eq. 5) might then be important, since they determine the distance at which the measurements have an influence in the analysis.

The last question deals with the effects of DA on chemistry. Assimilating a total mass only brings a coarse information about chemistry; we cannot, a priori, expect an improvement in the aerosol chemical composition, unless the errors are similar for all the species (which may not be realistic).

2 Methodology

2.1 Optimal interpolation

In the optimal interpolation method, the analysis is given by the best estimate (linear and unbiased) in the least-squares sense. The objective is to find the analysis state

vector $\boldsymbol{x}_a$, solution of the minimization problem

$$\boldsymbol{x}_a = \text{argmin}\, \mathcal{J}\,, \tag{1}$$

with $\mathcal{J}$ the following cost function:

$$\mathcal{J}(\boldsymbol{x}) = (\boldsymbol{x} - \boldsymbol{x}_b)^T \mathbf{B}^{-1} (\boldsymbol{x} - \boldsymbol{x}_b) + (\boldsymbol{y} - \mathbf{H}[\boldsymbol{x}])^T \mathbf{R}^{-1} (\boldsymbol{y} - \mathbf{H}[\boldsymbol{x}])\,. \tag{2}$$

$\boldsymbol{x}_b$ is the background state vector, or a priori state vector (i.e. the PM_{10} forecast provided by the model), $\boldsymbol{y}$ is the vector of observations (measured PM_{10}), and $\mathbf{H}$ is an interpolation function that maps the state $\boldsymbol{x}$ to the observational data. $\mathbf{B}$ and $\mathbf{R}$ are the matrices of error covariances, for background and observations respectively.

Upon minimization, $\boldsymbol{x}_a$ is given by:

$$\boldsymbol{x}_a = \boldsymbol{x}_b + \mathbf{K}\,(\boldsymbol{y} - \mathbf{H}(\boldsymbol{x}))\,, \tag{3}$$

$$\mathbf{K} = \mathbf{B}\mathbf{H}^T \left(\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R}\right)^{-1}\,, \tag{4}$$

where $\mathbf{K}$ is the so-called gain matrix.

The specification of the covariance matrices is decisive, because these matrices determine the corrections applied to the background field to coincide with the observations. The main parameters are the variances (diagonal terms), but the covariances are also important because they specify how the information should be distributed over the domain.

The variances of the observation errors can be evaluated by considering the characteristics of the instruments (supposed to be known). The covariances can be set to zero since the instruments errors are independent. This would not be the case for measurements from the same platform (radiosonde, airborne sensor or satellite), but this is reasonable for different instruments from different ground stations.

A first estimate of the background errors can be obtained by taking an arbitrary fraction of the climatological variance of the field itself. It is also conceivable to take an

estimate of the model-to-observation error (from the knowledge of the model uncertainty with respect to the observations). More complex methods, like the Hollingsworth-Lönnberg method, exist (Daley, 1991; Hollingsworth and Lönnberg, 1986).

The **B** matrix is difficult to determine explicitly. We therefore need a method to represent the covariances between the grid points. A classical method is the Balgovind approach: the covariance is a function of the horizontal and vertical distances (r_h and r_v respectively) between the two points of interest,

$$f(r_h, r_v) = \left(1 + \frac{r_h}{L_h} \exp\left(-\frac{r_h}{L_h}\right)\right) \times \left(1 + \frac{r_v}{L_v} \exp\left(-\frac{r_h}{L_v}\right)\right) \times v \,, \tag{5}$$

where L_h and L_v are two homogeneous influence radii and v is the variance estimate.

2.2 Redistribution over sections and chemical species

The correction applied to the simulated PM_{10} at ground is provided by the optimal-interpolation method. The controlled state of the model is thus the PM_{10} concentrations over the whole horizontal domain. Forecasted PM_{10} are computed by summing the concentrations of all aerosol species simulated over all sections (size discretization).

After DA, the analyzed PM_{10} are redistributed over the model variables following the initial chemical and size distributions. If $(PM_{10})^b$ and $(PM_{10})^a$ are the PM_{10} mass concentrations for the background and the analysis respectively, and $\left(P_i^j\right)^b$ and $\left(P_i^j\right)^a$ are the concentrations of the chemical species j in the section i for the background and the analysis respectively, then:

$$\left(P_i^j\right)^a = \frac{(PM_{10})^a}{(PM_{10})^b} \times \left(P_i^j\right)^b \,. \tag{6}$$

The underlying assumption is that the aerosol relative chemical composition and granulometry are well represented in the model.

3 Experimental setup

3.1 Simulations

The model used in this study is SIREAM, plugged to the chemistry-transport model Polair3D. SIREAM is a SIze-REsolved Aerosol Model, described in details in Debry et al. (2007). SIREAM includes 16 aerosol species: 3 primary species (mineral dust, black carbon and primary organic species), 5 inorganic species (ammonium, sulfate, nitrate, chloride and sodium) and 8 organic species managed by the SORGAM model (Schell et al., 2001). In the usual configuration, SIREAM includes 5 bins logarithmically distributed over the size range (0.01 μm, 10 μm). All these models are embedded in the POLYPHEMUS system, available at http://cerea.enpc.fr/polyphemus/ and which is described in Mallet et al. (2007).

The simulations presented hereafter (with or without assimilation) are at a continental scale, over Europe, and for one month (January, 2001). A first study, Sartelet et al. (2007), evaluates the model for the year 2001 with comparisons to three databases (also used in this study and described hereafter) and with respect to the performances of other models. The model shows a tendency to underestimate PM_{10} as the other models do in general, and to overestimate nitrate concentrations in wintertime. The model general configuration has the same features as in Sartelet et al. (2007). The main points are quoted hereafter.

The domain covers the area from 10.75° W to 22.75° E in longitude and from 34.75° N to 57.75° N in latitude, with a 0.5° step. There are five vertical layers: 0–50 m, 50–600 m, 600–1200 m, 1200–2000 m and 2000–3000 m.

The meteorological fields are interpolated from the outputs of the model of the European Center for Medium-range Weather Forecast (ECMWF, http://www.ecmwf.int/products/data/operational_system/). The ECMWF raw data has a resolution of 0.36° horizontally, 60 sigma-levels vertically and a timestep of 3 h.

The boundary conditions for aerosol species are interpolated from outputs of the GOddard Chemistry Aerosol Radiation and Transport model (GOCART, Chin et al.,

2000) for 2001.

The anthropogenic emissions for gases and aerosols are generated from the EMEP expert inventory for 2001 (available at http://www.emep.int/).

The chemical mechanism is RACM (Regional Atmospheric Chemistry Mechanism, Stockwell et al., 1997). Aerosol and gases are scavenged by dry deposition, rainout and washout. We take into account coagulation and condensation. Nucleation is not included because the diameters of nucleated particles (typically about 1nm) are lower than the lowest diameter bound of the model. Aqueous phase chemistry inside cloud droplets is also described (Variable Size Resolved Model VSRM, Fahey and Pandis, 2001; Strader et al., 1998).

3.2 Observational data for assimilation and comparison

In Sartelet et al. (2007), three databases are used for comparisons:

- the EMEP database, available on the EMEP Chemical Co-ordinating Centre (EMEP/CCC) web site at http://www.emep.int/;
- the AirBase database, available on the European Environment Agency (EEA) web site at http://air-climate.eionet.europa.eu/databases/airbase/;
- the BDQA database ("Base de Données Qualité de l'Air": the French data base for air quality that covers France).

The EMEP data is provided only on a daily basis. As a result, it will not be used for assimilation but for performance assessment.

The BDQA data will be used for assimilation in this study. The reasons for this choice is that it includes hourly concentrations and that the station types are specified, so that the traffic stations with too high concentrations can be removed. On the contrary, the types of the AirBase stations are not available, so AirBase data will only be used for comparisons, as the EMEP network.

Figure 1 shows the location of the BDQA stations on the model grid mesh, except the traffic and industrial stations whose high concentrations cannot be represented by our model (at 0.5° resolution).

4 Crosswise comparisons with other networks

In a first step, we would like to evaluate the improvements due to the assimilation process on other networks than the one used for assimilation. For this purpose, two simulations over one month (January, 2001) are carried out: one without DA (Model) and one with DA. For the simulation with DA, every hour, the model forecast (one-hour forecast) is modified by optimal interpolation with the data from BDQA stations. This gives an analyzed state (analysis), which becomes the initial condition for the next time step. Table 1 summarizes the fields compared in this section.

The Balgovind method is used to estimate the covariances for the background errors, with a scale parameter L_h set to 2 mesh cells (about 100 km). The background variance is set to $90\,\mu g^2 m^{-6}$, which derives from a RMSE of about $9\,\mu g\,m^{-3}$ for annual model-to-data comparisons (Sartelet et al., 2007). The matrix of observation covariances is assumed to be diagonal, with a variance set to $5\,\mu g^2 m^{-6}$. The error variance for observations is lower than the instrumental uncertainty[1]. In our study, the observations are assumed to be highly accurate because we aim at assessing the potential benefit of assimilation.

We first present statistics for the comparison with the AirBase data. The statistical measures are defined in the definition paragraph hereafter. The number of AirBase stations makes it possible to compute statistics for each country, as shown in Table 2. Figure 2 represents the RMSEs and the correlations with circles whose diameters are proportional to the statistical indicator. It is noteworthy that the statistics for PM_{10} are

[1] For a TEOM (Tapered Element Oscillating Microbalance), the observation error variance is about $25\,\mu g\,m^{-3}$.

globally improved: the global RMSE decreases from 21.4 to 19.3 $\mu g\,m^{-3}$ for the analysis, the correlation increases from 51.1% up to 58.9%, MFE decreases from 57.2% to 52.6% and MFB increases by 4%.

4.1 Definition

Let $\{o_i\}_{i=1,n}$ and $\{s_i\}_{i=1,n}$ be respectively the observed and the simulated concentrations. The Root Mean Square Error (RMSE), correlation, Mean Fractional Error (MFE) and Mean Fractional Bias (MFB) are defined below. RMSE is in $\mu g\ m^{-3}$, the other indicators are dimensionless.

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(o_i - s_i)^2}$$

$$\text{correlation} = \frac{\sum_{i=1}^{n}(o_i - \overline{o})(s_i - \overline{s})}{\sqrt{\sum_{i=1}^{n}(o_i - \overline{o})^2 \times \sum_{i=1}^{n}(s_i - \overline{s})^2}}$$

$$\text{MFE} = \frac{1}{n}\sum_{i=1}^{n}\frac{|s_i - o_i|}{(s_i + o_i)/2}$$

$$\text{MFB} = \frac{1}{n}\sum_{i=1}^{n}\frac{s_i - o_i}{(s_i + o_i)/2}$$

The statistics are clearly better over France (the RMSE decreases by more than 4 $\mu g\,m^{-3}$ and the correlation increases by more than 30%). This is obviously due to the fact that the BDQA stations are located in France. In addition, in France, the BDQA and AirBase networks share many stations. The statistics of border countries or near countries of northern Europe are improved, like for Belgium, Switzerland, Germany

(even if the correlation is deteriorated), Great Britain or the Netherlands. Countries that are relatively far from France, like Portugal, Poland or Slovakia show no changes in their statistics. On the contrary, the statistics for Spain (about 1 $\mu g\,m^{-3}$ increase for RMSE and 15% decrease for correlation) and Italy (equivalent RMSE but a decrease of 30% for correlation) are deteriorated. Section 6 will show that Italy and Spain are the countries with the highest the differences between the simulations with and without DA (if we except France). This remark raises the question of the distance over which the stations of southern France are representative of PM_{10} pollution in the direction toward Italy and Spain. It is possible that the Alps and the Pyrenees constitute a high barrier for aerosols at ground, barriers that the error statistics model does not take into account, with an overestimated L_h.

Table 3 shows that DA also improves the statistics for $PM_{2.5}$ (4 $\mu g\,m^{-3}$ decrease for RMSE and 24% increase for correlation), which could indicate that the a priori layout over the model bins is relatively reliable. On the other hand, the aerosol chemical composition is deteriorated (specially for ammonium and nitrate which is even more overestimated). The number of stations that provide measurements for chemical species is lower than the stations providing PM_{10} observations; it is therefore difficult to conclude on a general behavior. This highlights the need for more chemical measurements in the DA method presented here. The distribution on species could then be corrected by assimilation, while it is constant in this case.

Table 4 shows statistics for the comparisons with the EMEP data. Most statistics on PM_{10} for the simulation with DA (analysis and one-hour forecast) are deteriorated compared to the simulation without DA. Nevertheless, the simulated mean is better with assimilation because DA adjusts the underestimation. Note that the EMEP network includes very few stations in France, but much more in Spain where the results are spoiled by DA (see the results on the AirBase network). Moreover, EMEP stations are background stations, whereas the assimilated observations are measured at both background and urban or periurban stations. Some DA updates may not be consistent with background-concentrations levels. Besides, there may not be enough stations to

draw reliable conclusions.

The conclusion about the lack of improvement of the chemical speciation after DA remains the same with EMEP. For sulfate, the simulated mean agrees better with the measurements, while the RMSE is equivalent and the correlation decreases by 5%. All statistics are deteriorated for nitrate and ammonium, while the statistics do not change for sodium.

5 Operational forecast

In operational conditions, at time t_0, only the data for the previous days is available. It is possible to assimilate this data over a few days before t_0. The model results from t_0 to (t_0+1 day) are called one-day forecasts, the results from (t_0+1 day) to (t_0+2 days) are called two-day forecasts, etc. This operation can be repeated every day ("moving window"); a one-day forecast and a two-day forecast are then available every day.

Several five-day DA experiments are carried out: the BDQA data is assimilated during the first three days, then the model forecasts the next two days. The first experiment assimilates data from 1 January 2001 to 3 January 2001 and forecasts the days 4 January 2001 and 5 January 2001; the second experiment assimilates data from 2 January 2001 to 4 January 2001 and forecasts the days 5 January 2001 and 6 January 2001; and so on. Consequently, one-day forecasts are available from 4 January 2001 to 30 January 2001, and two-day forecasts are available from 5 January 2001 to 31 January 2001. Table 5 describes the simulations carried out in this section.

Table 6 summarizes the statistics of the model and of the one-day and two-day forecasts compared to BDQA data. It is noteworthy that, as expected, the one-day forecast clearly improves the statistics for PM_{10} and $PM_{2.5}$. The decrease of the RMSE value is 1.6 $\mu g\,m^{-3}$ for PM_{10} and 1.5 $\mu g\,m^{-3}$ for $PM_{2.5}$, that is, about 10%. The increase of the correlation is more than 10% for PM_{10} and $PM_{2.5}$. MFE and MFB are also markedly improved; the improvement in MFE brings the model to satisfy the performance objective of 50% defined by Boylan and Russell (2006) (see also Sartelet et al., 2007).

Since the optimal interpolation method only changes the initial conditions, two-day forecast shows a less obvious improvement. The model tends to its reference trajectory (without assimilation). Two-day forecast shows slightly better statistics, but the decrease of the RMSE is only 0.3 $\mu g\,m^{-3}$ and the increase of the correlation is 3%.

6 Temporal and horizontal impacts of DA

Because the optimal interpolation method only modifies the initial conditions, and not the model itself, we would like to evaluate the time and space scales for which DA affects the results. In this section, we carry out some tests over a shorter period, to estimate the effective time scales of DA (the data is assimilated over a period and then the control is "released" for the next days), and to find out the important parameters.

The configuration for the present simulations is the same as in the previous section, but only over the period from 1 January 2001 to 06 January 2001. As in the previous sections, the simulation without assimilation is compared to the other simulations with DA. For the simulations with DA, hourly data from all BDQA stations is assimilated from 1 January 2001 to 5 January 2001. The forecast starts on the 6 January 2001 at 00:00:00 UTC. Eight simulations are presented here: the reference test and seven alternatives. The different configurations are summarized in Table 7.

In this section, a higher variance for observations, equal to the background variance, is tested. The ratio $\alpha=v/r$ (where v is from Eq. 5 and, in our case, $R=rI$ where I is the identity matrix) is then equal to 1. The Balgovind method is used to represent the horizontal and vertical covariances for the background. The impact of the L_h and L_v parameters is tested. Horizontally, the reference L_h is taken equal to 2 grid cells (about 100 km), and is changed to 0.5 grid cell (about 25 km). Vertically, the reference test performs assimilation only in the first level at 25 m. Three other tests are carried out: respectively with a vertical L_v parameter equal to 200 m (with two controlled levels), 300 m (with three controlled levels) and 600 m (with three controlled levels).

The redistribution of analyzed PM_{10} on chemical species will also been investigated.

The different cases are the following:

- redistribution on all species (default): the model shows the same uncertainties for all species;
- redistribution of the corrections only on primary species: it is assumed that the uncertainties are mainly due to the emissions;
- redistribution of the corrections only on inorganic species: it is assumed that the uncertainties are mainly due to the condensation of inorganic species.

The redistribution only on the organic species is not tested. The reason is that the evaporation of organic species during the measurements is so important that we assume that the observations slightly depend on organic species.

Figures 3, 4 and 5 show the time evolution of the RMSE, the correlation and the simulated mean respectively, averaged over all BDQA stations for the different tests. Table 8 shows the associated statistics, over the first day of forecast (6 January 2001). All simulations with assimilation improve the RMSE and the correlation. Anyway, the figures show that the influence of DA lasts no more than a few hours. The RMSEs and the correlations are equivalent for all tests after 6 h. For the simulated mean, the typical time period before all simulations converge is 20 h. The inorganics simulation is almost the same as the reference test where all species are assimilated, whereas the primary simulation gives really different results for the simulated mean. Indeed, in the inorganics simulation, the transfer of the PM_{10} changes to the inorganic species will take place essentially in the fine mode (less than 2.5 μm diameter), where the major part of inorganics resides. The PM_{10} mass is also located in this mode, so the size distribution of corrected PM_{10} will be equivalent in both the reference and the inorganics simulations. For this reason, the scavenging (depending on the particle size), which could imply the differences in PM_{10} budget between simulations, is not impacted. Although the simulated mean for the primary simulation is the most different one from the simulation without DA at the beginning of the forecast, the RMSE is quite

the same. This probably means that, for this case, the errors of the model are mostly due to secondary species, rather than primary species. These errors could then be attributed essentially to physical processes rather than to emissions description for example. The simulation where α=1 is, as expected, rather close to the simulation without DA (see the evolution of the simulated mean). The simulations where one of the parameters L_h or L_v for the Balgovind method has changed give similar RMSEs on the BDQA network. However, the differences in the concentration mean remain significant for a longer period in the simulations where the number of levels used for assimilation and the vertical L_v are increased. This shows that DA should influence a part of the PM_{10} vertical profile, but increasing L_v over 200 m has a limited impact.

Figure 6 shows the maps of the absolute difference between the PM_{10} fields of simulation with DA and the same field without DA, averaged over one day after assimilation. As expected, the test with L_h=0.5 does not influence the PM_{10} at a long distance. Southern (Marseilles' region) and south-eastern France are the most impacted regions, showing that the concentrations of these regions are particularly badly reproduced by the model. This may be due to the fact that some stations in these regions are influenced by the mountains, by the Mediterranean circulation, and by large urban areas. Therefore, it is a region difficult to describe with a continental-scale simulation.

For the other tests, it is noteworthy that northern Spain is the region which is the most impacted by DA, showing that the corrections are transported at this location. Tests with L_h=2 show that northern Italy (Po Valley) is also greatly impacted. As L_v increases, the corrections seem to be transported over longer distances. It is noteworthy that the regions impacted by the inorganics and the primary tests are quite different. Actually, the inorganics test shows large differences over marine regions (see west of Corsica). Over marine areas, the changes in thermodynamic equilibria due to DA can be amplified by the presence of sea salt.

7 Conclusions

This paper shows that PM_{10} DA with the optimal interpolation method may be useful for one-day forecasts. In our tests, a mean decrease of 1.5 $\mu g\,m^{-3}$ for the RMSE and a 10% increase in the correlation are obtained. For longer forecast periods, the statistics are not improved because the concentrations converge too quickly to the trajectory without DA.

Moreover, the crosswise comparison with other networks than the one used for DA allows to evaluate the quality of the analysis for different network types. For example, the EMEP database only contains background stations and the assimilation of data from stations of other types spoils the statistics.

The background error covariances are important components of the assimilation. In this study, we used a simple parameterization with an influence radius L_h, supposed to be the same in the whole domain. This method results in spoiling the statistics in regions where topography would demand specific Balgovind lengths L_h.

For future works, the following studies could be initiated:

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, 1991);

2. Implementing inverse methods in order to improve the quality of input data (emissions) and/or parameterizations;

3. Assimilating observations of gases that are poorly measured but important for the formation of secondary inorganic species, like nitric acid (HNO_3) or ammoniac (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 withdrawn;

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.

Acknowledgements. This research is carried out in the ENPC-EDF R&D Joint Laboratory CEREA, and in the INRIA-ENPC joint project-team CLIME. The authors would like to thank Lin Wu for his help in the assimilation part of Polyphemus, and Marc Bocquet for scientific discussions. We also thank the BDQA, AirBase and EMEP networks for providing their data.

References

Austin, J.: Toward the four dimensional assimilation of stratospheric chemical constituents, J. Geophys. Res., 97, 2569–2588, 1992. 9609

Benedetti, A. and Fisher, M.: Background error statistics for aerosols, Q. J. Roy. Meteor. Soc., 133, 391–405, 2007. 9610

Bouttier, F. and Courtier, P.: Data assimilation concepts and methods, in: Meteorological Training Course Lecture Series, ECMWF, 2001. 9609

Boylan, J. and Russell, A.: PM and light extinction model performance metrics, goals, and criteria for three–dimensional air quality models, Atmos. Environ., 40, 4946–4959, 2006. 9619

Chin, M., Rood, R., Lin, S.-J., Muller, J. F., and Thompson, A. M.: Atmospheric sulfur cycle in the global model GOCART: Model description and global properties., J. Geophys. Res., 105, doi:10129/2000JD900384, 2000. 9614

Collins, W. D., Rasch, P. J., Eaton, B. E., Khattatov, B. V., Lamarque, J.-F., and Zender, C. S.: Simulating aerosols using a chemical transport model with assimilation of satellite aerosol retrievals: Methodology for INDOEX, J. Geophys. Res., 106, 7313–7336, 2001. 9610

Daley, R.: Atmospheric Data Analysis, Cambridge University Press, 1991. 9613, 9623

Debry, E., Fahey, K., Sartelet, K., Sportisse, B., and Tombette, M.: Technical note: A new SIze REsolved Aerosol Model, Atmos. Chem. Phys., 7, 1537–1547, 2007, http://www.atmos-chem-phys.net/7/1537/2007/. 9609, 9614

Elbern, H., Schmidt, H., and Ebel, A.: Variational data assimilation for tropospheric chemistry modeling, J. Geophys. Res., 102, 15 967–15 985, 1997. 9609

Elbern, H., Strunk, A., Schmidt, H., and Talagrand, O.: Emission rate and chemical state estimation by 4-dimensional variational inversion, Atmos. Chem. Phys., 7, 3749–3769, 2007, http://www.atmos-chem-phys.net/7/3749/2007/. 9609

Fahey, K. M. and Pandis, S. N.: Optimizing model performance: variable size resolution in cloud chemistry modeling, Atmos. Environ., 35, 4471–4478, 2001. 9615

Fisher, M. and Lary, D. J.: Lagrangian Four Dimensional Variational Data Assimilation of Chemical Species, Q. J. Roy. Meteor. Soc., 121, 1681–1704 Part A, 1995. 9609

Generoso, S., Bréon, F.-M., Balkanski, Y., Chevallier, F., Schulz, M., and Bey, I.: Assimilation of POLDER Aerosol Optical Thickness into the LMDz-INCA model: Implication for the Arctic aerosol burden, J. Geophys. Res., 112, doi:10.1029/2005JD006954, 2007. 9610

Hanna, S. R., Chang, J. C., and Fernau, M. E.: Monte Carlo estimates of uncertainties in predictions by a photochemical grid model (UAM-IV) due to uncertainties in input variables, Atmos. Environ., 32, 3619–3628, 1998. 9608

Hollingsworth, A. and Lönnberg, P.: The statistical structure of short-range forecast errors as determined from radiosonde data. Part I: the wind field, Tellus, 38A, 111–136, 1986. 9613

Huneeus, N.: Assimilation variationnelle d'observations satellitaires dans un modèle atmosphérique d'aérosols, Ph.D. thesis, Université des Sciences et Technologies de Lille, 2007. 9610

Jeuken, A., Eskes, H., Van Velthoven, P., Kelder, H., and Holm, E.: Assimilation of total ozone satellite measurements in a three-dimensional tracer transport model, J. Geophys. Res., 104, 5551–5563, 1999. 9609

Levelt, P., Khattatov, B., Gille, J., Brasseur, G., Tie, X., and Waters, J.: Assimilation of MLS ozone measurements in the global 3D CTM ROSE, Geophys. Res. Lett., 25, 4493–4496, 1998. 9609

Mallet, V. and Sportisse, B.: Uncertainty in a chemistry-transport model due to physical parametrizations and numerical approximations: An ensemble approach applied to ozone modeling, J. Geophys. Res., 111, doi:10.1029/2005JD006 149, 2006. 9608

Mallet, V., Quélo, D., Sportisse, B., Ahmed de Biasi, M., Debry, É., Korsakissok, I., Wu, L., Roustan, Y., Sartelet, K., Tombette, M., and Foudhil, H.: Technical Note: The air quality modeling system Polyphemus, Atmos. Chem. Phys., 7, 5479–5487, 2007, http://www.atmos-chem-phys.net/7/5479/2007/. 9609, 9614

Nenes, A., Pilinis, C., and Pandis, S.: ISORROPIA: A new thermodynamic model for inorganic multicomponent atmospheric aerosols, Aquat. Geochem., 4, 123–152, 1998. 9609

Planet, W.: Determination of total ozone amount from TIROS radiance measurements, J. Appl. Meteor., 23, 308–316, 1984. 9609

Riishojgaard, L.: On four-dimensional variational assimilation of ozone data weather prediction models, Q. J. Roy. Meteor. Soc., 122, 1545–1571, 1996. 9609

Roustan, Y., Sartelet, K. N., Tombette, M., Debry, E., Fahey, K. M., and Sportisse, B.: Simulation of aerosols and related species over Europe with the Polyphemus system. Part II: model uncertainty and sensitivity analysis for 2001, Tech. Rep., 2008-07, CEREA. 9608

Sartelet, K. N., Debry, E., Fahey, K. M., Roustan, Y., Tombette, M., and Sportisse, B.: Simulation of aerosols and gas-phase species over Europe with the Polyphemus system. Part I: model-to-data comparison for 2001, Atmos. Environ., 29, 6116–6131, 2007. 9614, 9615, 9616, 9619

Schell, B., Ackermann, I. J., and Haas, H.: Modeling the formation of secondary organic aerosol within a comprehensive air quality model system, J. Geophys. Res., 106, 28 275–28 293, 2001. 9614

Segers, A.: Data assimilation in atmospheric chemistry models using Kalman filtering, Ph.D. thesis, Delft University, 2002. 9609

Stockwell, W., Kirchner, F., and Kuhn, M.: A new Mechanism for regional chemistry modeling, J. Geophys. Res., 102, 25 847–25 879, 1997. 9615

Strader, R., Gurciullo, C., Pandis, S., Kumar, N., and Lurmann, F.: Development of a gas-phase chemistry, secondary organic aerosol and aquesous-phase chemistry modules for PM modelling, Technical report, STI, 1998. 9615

Wu, L., Mallet, V., Bocquet, M., and Sportisse, B.: A Comparison Study of Data Assimilation Algorithms for Ozone Forecasts, J. Geophys. Res., Tech Rep. 2008-06, CEREA. 9609

Table 1. Description of the European simulations for the period from 1 January 2001 to 31 January 2001.

Simulation	Data	Assimilated species	Assimilation period	Outputs
Without DA	–	–	–	Model
With DA	BDQA (all stations)	PM_{10}	1 to 31 January 2001	Analysis One-hour forecast

Table 2. RMSE, correlation, MFE and MFB (see Sect. 4.1) of the simulated PM_{10} without and with DA (for the model, the analysis and the one-hour forecast), computed with the observations from the AirBase network. The total is computed over the whole stations, without distinguishing the country. Countries are: Austria (AT), Belgium (BE), Switzerland (CH), Czech Republic (CZ), Germany (DE), Spain (ES), France (FR), Great Britain (GB), Ireland (IE), Italy (IT), the Netherlands (NL), Poland (PL), Portugal (PT), Slovenia (SI) and Slovakia (SK). Period: 1 January 2001 to 31 January 2001.

Stat	Country	AT	BE	CH	CZ	DE	ES	FR	GB	IE	IT	NL	PL	PT	SI	SK	Total
# stations		12	17	12	45	161	18	120	49	2	17	14	27	5	1	3	503
RMSE	Model	27.6	43.6	16.4	35.3	20.7	16.5	11.2	15.0	19.9	23.3	34.1	39.0	25.3	30.2	26.5	21.4
($\mu g\,m^{-3}$)	Analysis	27.9	33.7	15.7	35.5	19.5	17.3	6.6	13.4	19.1	23.7	29.0	39.1	25.1	30.3	26.5	19.3
	One-hour forecast	27.8	34.4	15.7	35.5	19.6	17.0	6.9	12.7	18.8	23.6	29.2	39.0	25.1	30.3	26.5	19.4
Correlation	Model	36.0	79.5	54.8	53.8	65.9	54.7	36.7	43.8	−3.4	35.7	65.8	46.6	−26.6	54.6	44.5	51.1
(%)	Analysis	16.7	88.3	54.2	45.2	62.4	39.1	71.9	70.4	40.4	0.9	74.6	45.2	−25.8	38.8	39.7	58.9
	One-hour forecast	18.8	89.7	53.9	46.2	63.2	43.5	69.4	70.5	35.6	4.1	77.3	45.1	−26.0	43.0	40.2	58.7
MFE	Model	83.1	58.4	53.3	78.1	54.7	90.5	45.2	40.4	56.0	59.8	49.5	86.5	61.5	89.0	83.5	57.2
(%)	Analysis	82.3	51.4	49.8	76.7	50.5	90.1	27.1	57.5	62.1	58.9	47.2	86.1	61.1	82.8	81.9	52.6
	One-hour forecast	82.5	51.3	49.9	77.0	50.1	90.5	28.3	48.0	56.8	58.7	46.7	86.1	61.3	84.0	82.1	51.9
MFB	Model	−60.6	−52.3	−8.1	−65.9	−37.5	−81.8	−9.7	−16.5	−24.9	−44.8	−41.0	−75.4	−6.1	−89.0	−81.8	−35.5
(%)	Analysis	−54.3	−46.2	9.4	−62.9	−29.9	−67.5	−0.2	−44.0	−46.7	−26.4	−37.7	−74.7	−9.5	-82.8	−80.2	−31.5
	One-hour forecast	−55.6	−47.1	6.7	−63.4	−30.5	−71.3	0.1	−34.4	−38.5	−29.2	−38.4	−74.8	−8.6	-84.0	−80.4	−30.9

Table 3. Statistics of the simulation results without and with DA (for the model, the analysis and the one-hour forecast) on AirBase network for different species. Period: from 1 January 2001 to 31 January 2001.

Species	Simulation	# stations	Obs. Mean $\mu g\ m^{-3}$	Sim. Mean $\mu g\ m^{-3}$	RMSE $\mu g\ m^{-3}$	Correlation %	MFE %	MFB %
$PM_{2.5}$	Model	10	18.8	17.7	12.5	61.4	45.4	8.3
	Analysis			20.4	8.7	85.8	38.2	13.8
	One-hour forecast			19.9	7.9	86.4	34.1	14.8
Sulfate	Model	11	2.4	1.7	2.2	62.1	64.1	3.9
	Analysis			1.9	2.0	60.3	63.5	5.7
	One-hour forecast			1.8	2.0	61.5	62.7	5.9
Nitrate	Model	8	4.4	7.8	4.2	66.3	71.9	68.8
	Analysis			8.7	6.6	66.3	74.3	70.9
	One-hour forecast			8.6	6.2	67.9	73.3	71.4
Ammon.	Model	8	2.2	2.6	1.3	81.6	52.9	34.0
	Analysis			3.0	1.7	77.8	53.9	36.3
	One-hour forecast			2.9	1.6	79.7	52.2	36.9
Chlore	Model	6	0.9	2.8	3.0	45.7	93.6	81.6
	Analysis			2.7	2.7	41.6	93.4	83.8
	One-hour forecast			2.7	2.7	41.8	93.0	84.8

Table 4. Statistics of the simulation results without and with DA (for the model, the analysis and the one-hour forecast) on the EMEP network for different species. Period: from 1 January 2001 to 31 January 2001.

Species	Simulation	# stations	Obs. Mean $\mu g\ m^{-3}$	Sim. Mean $\mu g\ m^{-3}$	RMSE $\mu g\ m^{-3}$	Correlation %	MFE %	MFB %
PM_{10}	Model	16	20.2	17.4	17.8	45.3	65.7	19.5
	Analysis			19.8	19.0	36.0	70.4	27.3
	One-hour forecast			19.4	18.7	36.7	69.0	26.5
$PM_{2.5}$	Model	7	20.4	17.3	17.2	57.0	68.8	19.9
	Analysis			21.1	18.8	52.4	70.0	30.0
	One-hour forecast			20.5	18.4	52.9	69.3	29.5
Sulfate	Model	55	2.3	1.2	1.9	52.5	63.5	−30.0
	Analysis			1.4	1.9	46.6	66.9	−22.7
	One-hour forecast			1.4	1.9	47.3	64.6	−22.1
Nitrate	Model	14	3.2	6.4	4.2	38.3	92.1	84.3
	Analysis			6.9	5.0	34.9	94.7	87.4
	One-hour forecast			6.8	4.9	35.2	94.3	87.2
Ammon.	Model	9	2.2	2.3	1.6	48.6	57.4	22.0
	Analysis			2.6	1.9	42.3	61.5	26.6
	One-hour forecast			2.6	1.8	43.0	60.6	26.5
Sodium	Model	3	1.2	3.2	3.0	60.6	88.3	81.3
	Analysis			3.2	2.9	62.9	87.4	78.8
	One-hour forecast			3.2	2.9	61.7	87.2	79.0

Table 5. Description of the European simulations used for the tests in operational forecasts and the outputs used for the comparison to observations. t_0 is a given day between 3 January 2001 and 30 January 2001. "d" stands for day.

Simulation	Data	Period of assimilation	Outputs
Without DA	–	–	Model
With DA	BDQA (all stations)	t_0-3d to t_0	One-day forecast (t_0 to t_0+1d) Two-day forecast (t_0+1d to t_0+2d)

Table 6. Statistics of the simulations (model, one-day forecast and two-day forecast) on the BDQA network for PM_{10} and $PM_{2.5}$. Period: 4 January 2001 to 30 January 2001.

Species	Simulation	# stations	Obs. Mean $\mu g\ m^{-3}$	Sim. Mean $\mu g\ m^{-3}$	RMSE $\mu g\ m^{-3}$	Correl. %	MFE %	MFB %
PM_{10}	Model	166	21.8	17.4	16.6	35.7	55.3	−9.2
	One-day forecast			18.7	15.0	47.6	49.5	−3.3
	Two-day forecast			17.7	16.3	38.5	53.3	−8.1
$PM_{2.5}$	Model	8	19.8	15.8	15.0	30.2	57.9	−10.3
	One-day forecast			16.9	13.5	44.0	44.0	−4.0
	Two-day forecast			16.0	14.7	33.1	57.1	−9.4

Table 7. Configuration of DA for the evaluation of the impact of the assimilation parameters on the forecasts.

Parameter	Reference	Alternative
L_h Balgovind	2 cells	0.5 cell
L_v Balgovind	30m	200m
L_v Balgovind	30m	300m
L_v Balgovind	30m	600m
Redistribution	all species	primary inorganics
$\alpha = v/r$	18	1

Table 8. Table of the scores of the different tests at BDQA stations for the first day forecast (6 January 2001).

Simulation	RMSE	Correlation	Concentration mean
Observations			14.1
No assimilation	11.1	54%	11.2
All species (Reference)	10.2	57%	11.3
Primary species	10.9	55%	12.0
Inorganic species	10.2	57%	11.4
L_h=0.5	10.2	57%	11.2
alpha=1	10.4	56%	11.0
L_v=200m	10.2	58%	12.0
L_v=300m	10.3	57%	12.0
L_v=600m	10.3	57%	12.0

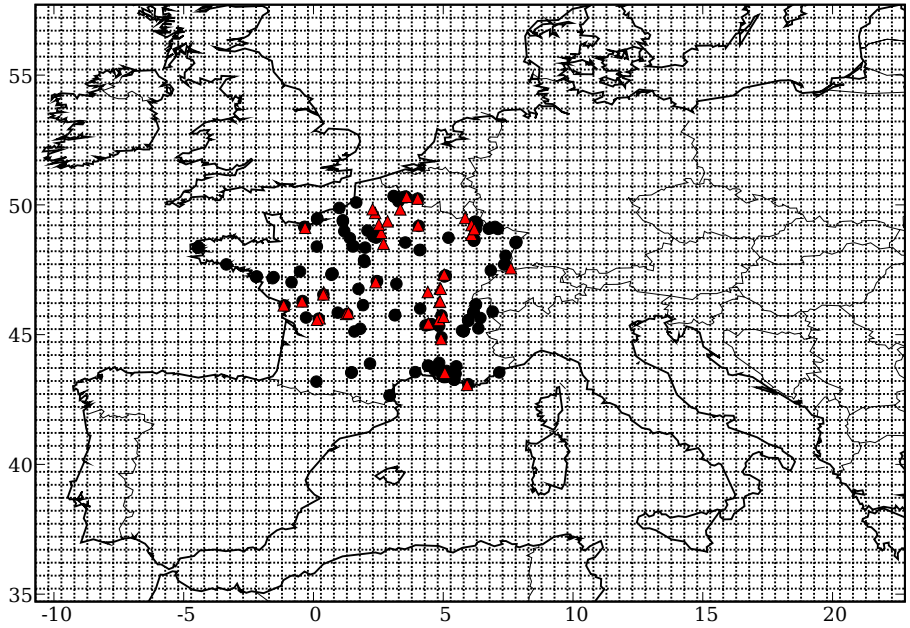

Fig. 1. Location of BDQA stations used for PM_{10} DA. The background stations are located with a red triangle, the other ones with a black point. The model grid is also shown.

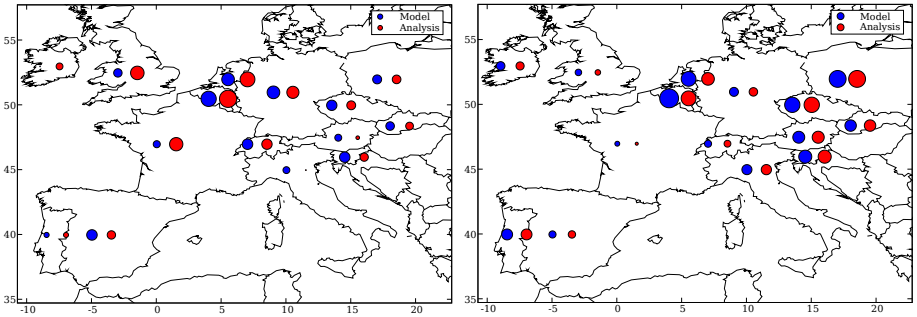

Fig. 2. Map of correlations **(a)** and RMSEs **(b)** for each country between the simulation and the AirBase observations for the simulation without DA (blue points) and for the one-hour forecast after assimilation (red points). The circle diameters are proportional to the statistical indicator.

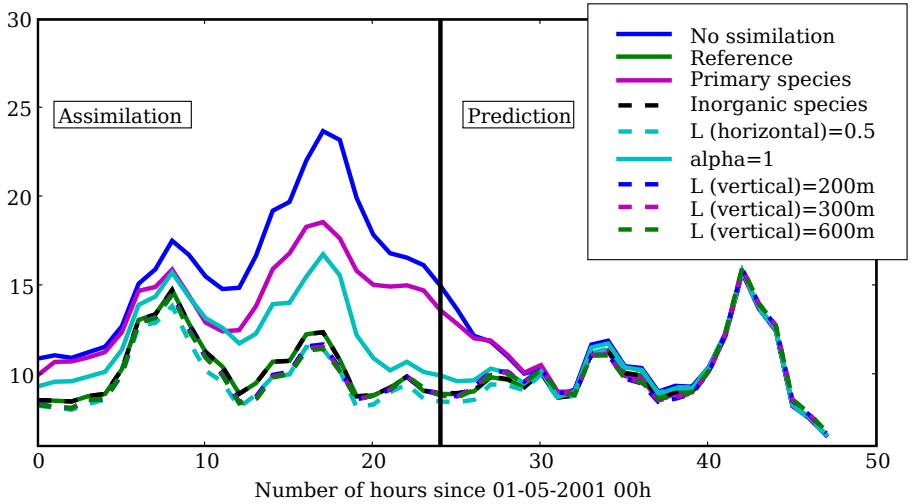

Fig. 3. Time evolution of the RMSE for the PM_{10} forecasts. The vertical line delimits the assimilation period from the prediction period.

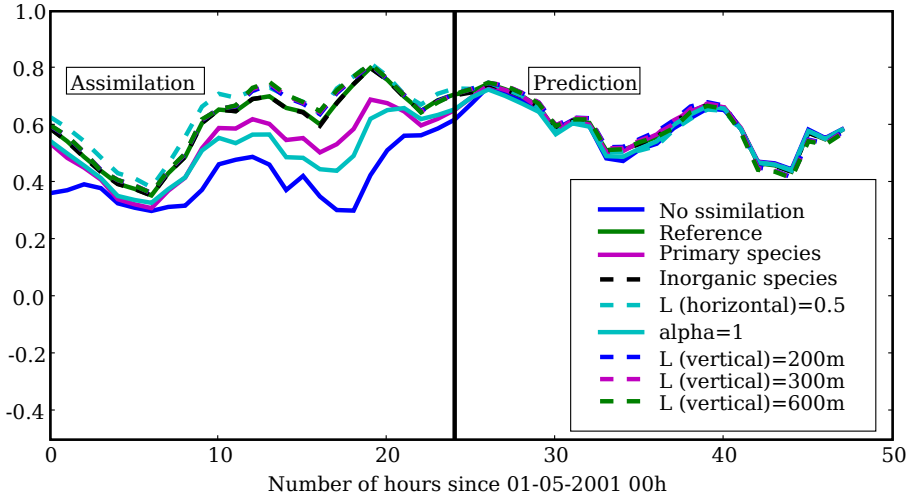

Fig. 4. Time evolution of the correlation for the PM_{10} forecasts. The vertical line delimits the assimilation period from the prediction period.

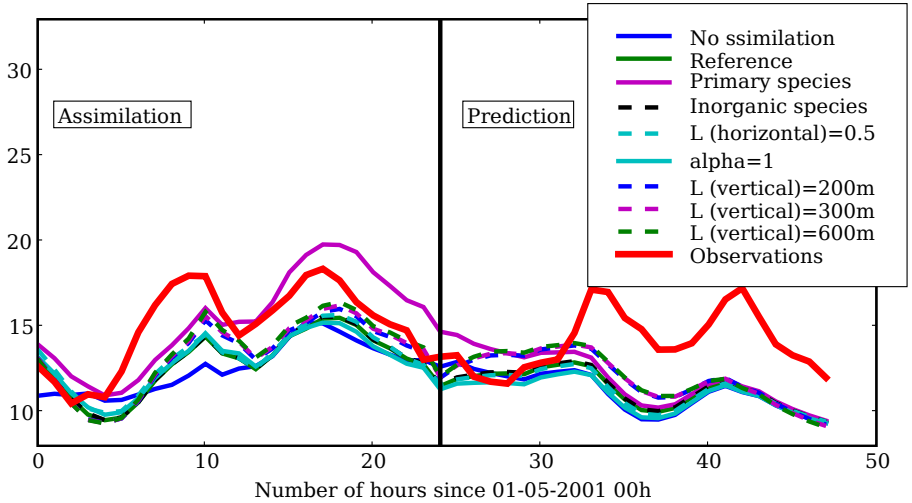

Fig. 5. Time evolution of the mean concentration for the PM_{10} forecasts. The vertical line delimits the assimilation period from the prediction period.

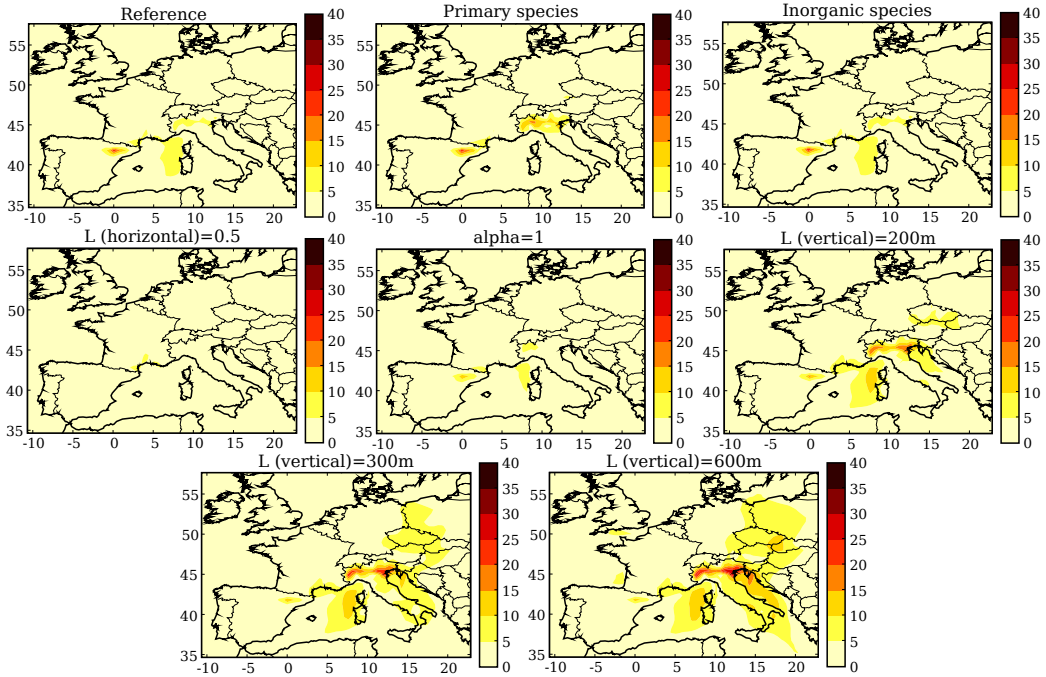

Fig. 6. Maps of the absolute difference of the PM_{10} fields. Comparison between the simulation without assimilation and the simulation with assimilation for the eight tests over the day 6 January 2001.