



HAL
open science

PM10 data assimilation over Europe with the optimal interpolation method

Marilyne Tombette, Vivien Mallet, Bruno Sportisse

► **To cite this version:**

Marilyne Tombette, Vivien Mallet, Bruno Sportisse. PM10 data assimilation over Europe with the optimal interpolation method. *Atmospheric Chemistry and Physics*, 2009, 9 (1), pp.57-70. 10.5194/acp-9-57-2009 . hal-00328328v2

HAL Id: hal-00328328

<https://hal.science/hal-00328328v2>

Submitted on 9 May 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

PM₁₀ data assimilation over Europe with the optimal interpolation method

M. Tombette^{1,2}, V. Mallet^{2,1}, and B. Sportisse^{1,2}

¹CEREA, joint laboratory ENPC – EDF R&D, Université Paris-Est, Marne la Vallée, France

²INRIA, Paris-Rocquencourt research center, France

Received: 26 March 2008 – Published in Atmos. Chem. Phys. Discuss.: 27 May 2008

Revised: 27 November 2008 – Accepted: 29 November 2008 – Published: 7 January 2009

Abstract. This paper presents experiments of PM₁₀ 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 January 2001 and for several outputs (PM₁₀, PM_{2.5} and chemical composition). The method is then applied in operational-forecast conditions. It is found that the assimilation of PM₁₀ observations significantly improves the one-day forecast of total mass (PM₁₀ and PM_{2.5}), whereas the improvement is non significant for the two-day forecast. The errors on aerosol chemical composition are sometimes amplified by the assimilation procedure, which shows the need for chemical data. Since the observations cover a limited part of the domain (France versus Europe) and since the method used for assimilation is sequential, we focus on the horizontal and temporal impacts of the assimilation and we study how several parameters of the assimilation system modify these impacts. The strategy followed in this paper, with the optimal interpolation, could be useful for operational forecasts. Meanwhile, considering the weak temporal impact of the approach (about one day), the method has to be improved or other methods have to be considered.

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 atmospheric components (Hanna et al., 1998; Mallet and Sportisse, 2006), in particular aerosols (Roustan et al., 2008), leading to substantial discrepancies with observational data.

Data assimilation (DA hereafter) makes use of observations 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 may be the evaluation of the emissions fluxes, and not necessary the improvement of the forecast itself (Quélo et al., 2005; Chang et al., 1997). This defines the so-called *inverse modeling* issue, which is not addressed in this paper.

There are many existing approaches to DA: statistical methods (e.g., *optimal interpolation*, OI hereafter), 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, such as the emission fluxes.

DA is a relatively recent research field in atmospheric chemistry (Austin, 1992; Fisher and Lary, 1995; Riishojgaard, 1996), even if 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 OI approach in Jeuken et al. (1999) and Planet (1984), and with a 4D-Var approach in Riishojgaard (1996). In Levelt et al. (1998), ozone and carbon monoxide profiles are assimilated within 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



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

and a sequential approach for the latter. Wu et al. (2008) compare four assimilation methods for assimilation of ozone ground measurements: OI, ensemble Kalman filter, reduced-rank square root Kalman filter and 4D-Var.

Aerosol models are now a component of most of available chemistry-transport models (CTMs). Performance obtained for PM₁₀ with various models could be disappointing in comparison to the performance obtained for gas species. For example, the correlation of simulated versus observed concentrations rarely exceeds 50% for hourly PM₁₀ over Europe (Van Loon et al., 2004), and the Root Mean Square Error (RMSE) is of the order of the concentration values, i.e. 10 µg 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 PM₁₀, 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. In this paper, the aerosol model is SIREAM – SIze-REsolved Aerosol Model, Debry et al. (2007) – embedded in the POLYPHEMUS platform (Mallet et al., 2007). The assimilation method is partially determined by the constraints originated from the thermodynamic model – ISORROPIA, Nenes et al. (1998). Indeed, there is currently no adjoint model for SIREAM. First, the thresholds set by ISORROPIA, which is a discontinuous model because of its phase transitions, raise a theoretical problem: how to define the derivatives at the discontinuity points? Second, the code of ISORROPIA has not been written so that it may be automatically differentiated. A variational approach is thus not practical 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 (PM₁₀, PM_{2.5}) or optical data (extinction coefficient, optical thickness). It is therefore not straightforward to assimilate this data. A reason for the choice of assimilating PM₁₀ data in this study is that the networks giving PM₁₀ data are the most widely extended. Moreover, since policies about particulate matter have mainly focused on PM₁₀ up to now, it is important to accurately forecast its concentrations, primarily in urban areas. For example, in Europe, the limit is set to a maximum of 35 days per year during which PM₁₀ exceeds 50 µg m⁻³ (directive of the European Commission 1999/30/CE). However, the numerical models sometimes miss important events, e.g., because of the lack of description of some emission sources, like for example wildfires (Hodzic et al., 2007). DA could be useful in order to compensate for these model deficiencies.

In this paper, we use a simple method for PM₁₀ DA, namely the OI method. This method was applied for the assimilation of aerosol optical thickness (Generoso et al., 2007; Collins et al., 2001). Variational methods for aerosol assim-

ilation have also been investigated, like the 1D-Var method in Huneeus (2007) or the 4D-Var method in Benedetti and Fisher (2007), but for simplified aerosol models.

This paper addresses the following questions:

1. Are forecasts improved after the assimilation of PM₁₀ hourly observations over Europe? If so, to what time extent?
2. Do the analyses produced by the PM₁₀ assimilation over a sub-domain (here, France) result in an improvement over the remaining part of the domain (western Europe)?
3. Does the analysis produced by the PM₁₀ assimilation improve the computed chemical aerosol composition?

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 at the same time. Second, the aerosol residence time in the atmosphere is of the order of few days and particles are known for having mostly a regional effect. For this reason, 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 aims to determine whether 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 a realistic assumption.

2 Methodology

2.1 Optimal interpolation

In the OI method, observations, as soon as available, are used to produce an *analysis*. This analysis is supposed to be a better estimate of the true state: it replaces the current state of the model and it therefore serves as initial conditions for the next model iterations. This procedure is repeated at the frequency of measurements.

The analysis is given by the best estimate (linear and unbiased) in the least-squares sense. The analysis, or analyzed state vector, \mathbf{x}_a , is solution of the minimization problem

$$\mathbf{x}_a = \operatorname{argmin} \mathcal{J}, \quad (1)$$

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

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

\mathbf{x}_b is the background state vector, or a priori state vector (i.e., the PM₁₀ forecast provided by the model), \mathbf{y} is the vector of observations (measured PM₁₀), and \mathbf{H} is an interpolation function that maps the state \mathbf{x} to the observational data. \mathbf{B} and \mathbf{R} are the matrices of error covariances, for background and observations, respectively.

Upon minimization and under the assumption of the linearity of \mathbf{H} , \mathbf{x}_a is given by:

$$\mathbf{x}_a = \mathbf{x}_b + \mathbf{K} (\mathbf{y} - \mathbf{H}[\mathbf{x}_b]), \quad (3)$$

where \mathbf{K} is the so-called *gain matrix* defined as

$$\mathbf{K} = \mathbf{B}\mathbf{H}^T (\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1}. \quad (4)$$

During the assimilation period, the process is repeated each time observations are available. So, at time h , the assimilation produces an analysis $\mathbf{x}_a^{(h)}$, which serves as new model state. Starting from that new state, the model computes a forecast $\mathbf{x}_f^{(h+1)}$ at time $h+1$. This forecast is actually called “background”, \mathbf{x}_b , in previous paragraphs. The assimilation takes place to compute $\mathbf{x}_a^{(h+1)}$; and so on and so forth.

2.2 Specification of the error covariance matrices

The specification of the covariance matrices is crucial, because these matrices determine the corrections to be applied to the background field in order to better match the truth. 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.

A first estimate of the background error variances 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. More complex methods, like the Hollingsworth-Lönnberg method, exist (Daley, 1993; Hollingsworth and Lönnberg, 1986).

It is impossible to accurately approximate all coefficients of \mathbf{B} . We therefore need a simplified representation of the covariances between the grid points. A classical method is the Balgovind approach (Balgovind et al., 1983) according to which the covariance is a function of the horizontal and vertical distances (r_h and r_v respectively) between the two points of interest,

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

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

The covariances between the observation errors are set to zero since the instruments errors are independent; so \mathbf{R} should be diagonal. This may not be the case for measurements from the same platform (radiosonde, airborne sensor or satellite), but it is a reasonable assumption for instruments located at different ground stations. We also assume that all stations have the same error variance: $\mathbf{R} = r\mathbf{I}$ where \mathbf{I} is the identity matrix and r the error variance.

The value of the observation error variance r is determined with a χ^2 diagnosis (e.g., Ménard et al., 2000). Let the forecast at time h be $\mathbf{x}_f^{(h)}$, and the observation at time h be $\mathbf{y}^{(h)}$. Under the usual assumptions of data assimilation, the departure $\mathbf{d}^{(h)} = \mathbf{y}^{(h)} - \mathbf{H}\mathbf{x}_f^{(h)}$ (called “innovation”) should have the covariance matrix $\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R}$. If that is true, the scalar

$$\chi_h^2 = \mathbf{d}^{(h)T} (\mathbf{H}\mathbf{B}\mathbf{H}^T + \mathbf{R})^{-1} \mathbf{d}^{(h)} \quad (6)$$

should be equal, on average, to the number of observations (at time h). Let N_h be that number of observations. In the previous equations, we assumed that N_h is constant since \mathbf{H} and \mathbf{R} do not depend on time. In practice, N_h is not constant because not the same stations will be available at two different times.

In practice, the χ^2 diagnosis should lead to check that χ_h^2/N_h is about 1. We determined the value of r so that χ_h^2/N_h remains reasonably close to 1. This can be a difficult optimization because the innovations $\mathbf{d}^{(h)}$ depend on r .

2.3 Redistribution over sections and chemical species

The correction applied to the simulated PM₁₀ at ground is provided by the OI method. The controlled variable is thus the PM₁₀ concentrations over the whole horizontal domain, in the first layer only (in the reference configuration) or in more layers. Forecast PM₁₀ are computed by summing the concentrations of all aerosol species simulated over all sections (size discretization).

After DA, the analyzed PM₁₀ concentrations are redistributed over the model variables following the initial chemical and size distributions. If $(\text{PM}_{10})^b$ and $(\text{PM}_{10})^a$ are the PM₁₀ mass concentrations for the background and the analysis respectively, and $(P_i^j)^b$ and $(P_i^j)^a$ are the concentrations of the chemical species j in the section i for the background and the analysis respectively, then:

$$(P_i^j)^a = \frac{(\text{PM}_{10})^a}{(\text{PM}_{10})^b} \times (P_i^j)^b. \quad (7)$$

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

3 Experimental setup

3.1 Simulations

The aerosol model used in this study is SIREAM, plugged to the chemistry-transport model Polair3D. SIREAM stands for SIZe-RESolved Aerosol Model, and is 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 predicted with the Secondary Organic Aerosol Model – SORGAM, 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 described in Mallet et al. (2007).

The simulations presented here with and without assimilation are carried out at a continental scale, over Europe, and for one month (January 2001). A previous study, Sartelet et al. (2007), evaluated the model configuration for the year 2001 with comparisons to three databases (also used in this study and described hereafter) and with respect to the performance of other CTMs used in Europe (Chimere, EMEP, ...). Polyphemus shows a tendency to underestimate PM₁₀, and to overestimate nitrate concentrations in wintertime. Other models also show similar behaviour. The configuration of the simulations of this paper is essentially the same as in Sartelet et al. (2007). The main characteristics of the configuration are summarized below.

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 provided by the European Centre for Medium-range Weather Forecast (ECMWF, http://www.ecmwf.int/products/data/operational_system/). The ECMWF raw data for 2001 is that of the 12-h forecast cycles starting from analyzed fields, and this 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 Regional Atmospheric Chemistry Mechanism – RACM, Stockwell et al. (1997) – is used to simulate chemistry. 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 1 nm) are below the lowest diameter bound of the model. Aqueous-phase

chemistry inside cloud droplets is also taken into account – Variable Size Resolved Model, VSRM, Fahey and Pandis (2001) and Strader et al. (1998).

3.2 Observational data for assimilation and comparison

In this paper, just like in Sartelet et al. (2007), three databases are used for comparisons:

- the EMEP (European Monitoring and Evaluation Programme) 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 sur la Qualité de l’Air”: the French national data base for air quality that covers France).

EMEP data is provided only on a daily basis. Hence, in this paper, it is not used for assimilation but for performance assessment.

On the contrary, BDQA data is used for assimilation. The reasons for this choice is that it includes hourly concentrations and that 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 procedure, the traffic and industrial stations, generally with high concentrations due to the proximity of sources, have been removed.

Since the characteristics of the AirBase stations are not available, AirBase data will only be used for validation, as the EMEP network. Also note that AirBase compiles several European databases, including BDQA.

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

3.3 Data-assimilation parameters

The covariances for the background errors are assumed to be in Balgovid form, with a scale parameter L_h set to 1 mesh cell (about 50 km). The background variance is set to 200 $\mu\text{g}^2\text{m}^{-6}$, which derives from a RMSE of about 15 $\mu\text{g}\text{m}^{-3}$ for annual model-to-observation comparisons (Sartelet et al., 2007).

We performed several simulations, over January 2001, with different observation error variances r in order to satisfy the χ^2 criterion – see Sect. 2.2. We finally selected $r \simeq 49 \mu\text{g}^2\text{m}^{-6}$. This value is significantly greater than the variance of the measurement error which is probably less than 25 $\mu\text{g}^2\text{m}^{-6}$. With this variance for observation error and with the 200 $\mu\text{g}^2\text{m}^{-6}$ variance for background, the time evolution of χ_h^2/N_h is shown in Fig. 2. On average the ratio

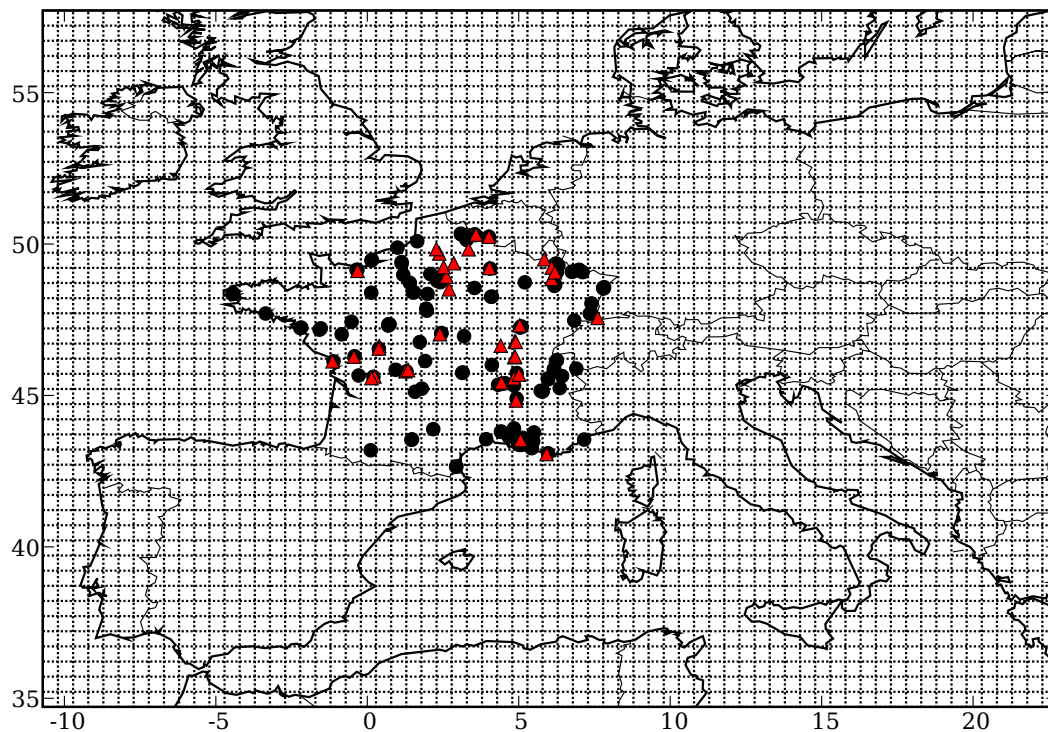


Fig. 1. Location of BDQA stations used for PM₁₀ DA. The background stations are located with a red triangle, the other ones with a black point. The model grid is also shown.

χ_h^2/N_h is almost 1, as expected. The ratio shows no clear temporal trend, which is also desirable. Nevertheless, there are significant time variations, even if there are 107 observations per hour (on average) and therefore 2568 observations per day – so the samples seem to be large enough for the averaged values to be statistically significant. This suggests that further error modeling would be useful.

4 Comparisons with other networks

As a first step, we 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 is modified by OI, using the data from BDQA stations. This produces a sequence of analyzed states (or analyses), which are thereafter compared to hourly observations.

The statistical measures to evaluate the results are: the Root Mean Square Error (RMSE), the correlation, the Mean Fractional Error (MFE), the Mean Fractional Bias (MFB). Let $\{o_i\}_{i=1,n}$ and $\{s_i\}_{i=1,n}$ be respectively the observed and

the simulated concentrations. The RMSE – (in $\mu\text{g m}^{-3}$) – and the other indicators (dimensionless) are defined as:

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

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

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

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

where $\bar{o} = \frac{1}{n} \sum_{i=1}^n o_i$ and $\bar{s} = \frac{1}{n} \sum_{i=1}^n s_i$.

We first present statistics for the comparison with the AirBase data. The large number of AirBase stations makes it possible to compute statistics for each country, as shown in Table 1. Figure 3 represents the RMSEs and the correlations with circles whose diameters are proportional to the statistical indicator, e.g., the higher the value of RMSE or the correlation, the larger the diameter of the circle. It is noteworthy that the statistics for PM₁₀ are globally improved: the global

Table 1. RMSE, correlation, MFE and MFB of the simulated PM₁₀ 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 all 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: 01 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 ($\mu\text{g m}^{-3}$) | 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 |
| | Analysis | 27.6 | 36.5 | 15.1 | 35.4 | 19.6 | 16.5 | 6.5 | 13.3 | 19.5 | 22.9 | 31.1 | 39.1 | 25.4 | 30.2 | 26.5 | 19.6 |
| | One-hour forecast | 27.6 | 37.1 | 15.3 | 35.3 | 19.8 | 16.4 | 6.9 | 13.5 | 19.5 | 22.9 | 31.3 | 39.1 | 25.3 | 30.2 | 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 | 35.0 | 91.8 | 59.7 | 53.9 | 70.2 | 54.4 | 72.0 | 64.7 | 6.3 | 27.3 | 78.1 | 46.5 | -26.4 | 54.0 | 44.4 | 63.8 |
| | One-hour forecast | 35.6 | 92.3 | 59.1 | 54.2 | 70.0 | 54.8 | 69.3 | 62.9 | 5.6 | 27.7 | 78.7 | 46.8 | -26.5 | 54.5 | 44.7 | 63.0 |
| 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 | 83.1 | 52.3 | 48.2 | 77.9 | 51.1 | 90.8 | 26.8 | 40.8 | 55.5 | 57.7 | 46.4 | 86.5 | 61.5 | 88.3 | 83.5 | 51.2 |
| | One-hour forecast | 83.0 | 52.8 | 48.8 | 77.9 | 51.1 | 90.5 | 28.3 | 40.0 | 55.1 | 57.9 | 46.8 | 86.5 | 61.4 | 88.4 | 83.4 | 51.5 |
| 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 | -60.0 | -49.2 | 4.4 | -65.6 | -33.8 | -77.2 | -0.0 | -25.6 | -26.8 | -36.3 | -39.7 | -75.4 | -6.2 | -88.3 | -81.7 | -32.2 |
| | One-hour forecast | -60.0 | -49.7 | 2.4 | -65.6 | -34.0 | -78.0 | 0.5 | -23.3 | -26.0 | -37.5 | -40.2 | -75.4 | -6.1 | -88.4 | -81.7 | -32.0 |

Table 2. 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 to 31 January 2001.

| Species | Simulation | # stations | Obs. Mean $\mu\text{g m}^{-3}$ | Sim. Mean $\mu\text{g m}^{-3}$ | RMSE $\mu\text{g m}^{-3}$ | Correlation % | MFE % | MFB % |
|-------------------|-------------------|------------|--------------------------------------|--------------------------------------|------------------------------|------------------|----------|----------|
| PM _{2.5} | Model | 10 | 18.8 | 17.7 | 12.5 | 61.4 | 45.4 | 8.3 |
| | Analysis | | | 19.9 | 8.4 | 87.7 | 36.1 | 17.7 |
| | One-hour forecast | | | 19.5 | 9.1 | 87.2 | 34.8 | 11.4 |
| Sulfate | Model | 11 | 2.4 | 1.7 | 2.2 | 62.1 | 64.1 | 3.9 |
| | Analysis | | | 1.8 | 2.0 | 63.5 | 62.4 | 4.9 |
| | One-hour forecast | | | 1.8 | 2.0 | 64.0 | 62.4 | 4.8 |
| Nitrate | Model | 8 | 4.4 | 7.8 | 4.2 | 66.3 | 71.9 | 68.8 |
| | Analysis | | | 8.2 | 4.9 | 70.1 | 72.6 | 71.0 |
| | One-hour forecast | | | 8.2 | 4.8 | 70.4 | 72.5 | 70.7 |
| Ammon. | Model | 8 | 2.2 | 2.6 | 1.3 | 81.6 | 52.9 | 34.0 |
| | Analysis | | | 2.7 | 1.3 | 83.3 | 51.6 | 36.2 |
| | One-hour forecast | | | 2.8 | 1.2 | 83.8 | 51.4 | 36.0 |
| Chlorine | Model | 6 | 0.9 | 2.8 | 3.0 | 45.7 | 93.6 | 81.6 |
| | Analysis | | | 2.7 | 2.8 | 39.8 | 93.9 | 85.6 |
| | One-hour forecast | | | 2.7 | 2.7 | 39.9 | 94.7 | 79.7 |

RMSE decreases from 21.4 to 19.6 $\mu\text{g m}^{-3}$ for the analysis, the correlation increases from 51.1% up to 63.8%, MFE decreases from 57.2% to 51.2% and MFB decreases by 3%.

The statistics are clearly better over France: the RMSE decreases by more than 4 $\mu\text{g m}^{-3}$ 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 most border- or nearby-countries (Belgium, Switzerland, Germany, Great Britain or the Netherlands) are also improved with the exception of the MFB for Great Britain. Countries that are relatively far from France, like Portugal, Poland or Slovakia show no changes in their statistics. Also, the statistics for Spain do not change, certainly because there are few stations in the south-west of

France, and the scale parameter L_h is not large enough to influence Spain. On the contrary, RMSE, MFE and MFB for Italy are improved, but the correlation decreases by 8%. Indeed, the number of BDQA stations in the south-eastern part of France, which could influence Italy, is significant. This remark raises the question of the distance over which the stations of southern France are representative of PM₁₀ pollution in the direction toward Italy. It is possible that the Alps constitute for aerosols at ground a high barrier, that the error statistics model does not take into account, with an overestimated L_h .

Table 2 shows that DA also improves the statistics for PM_{2.5} (4 $\mu\text{g m}^{-3}$ decrease for RMSE and 26% increase for correlation), which could indicate that the a priori layout over the model bins is relatively reliable.

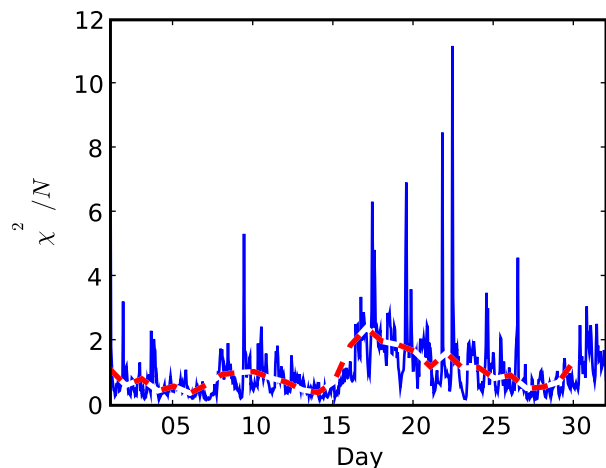
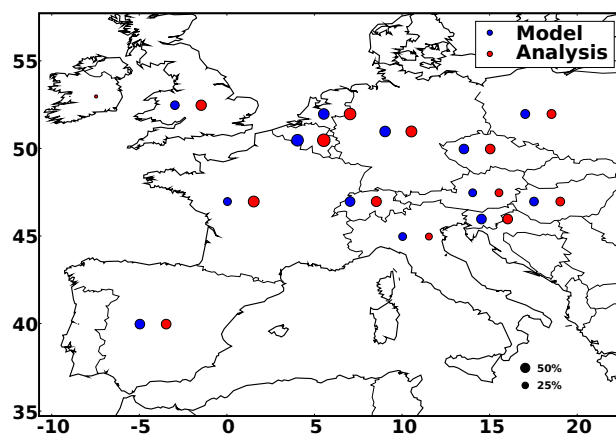


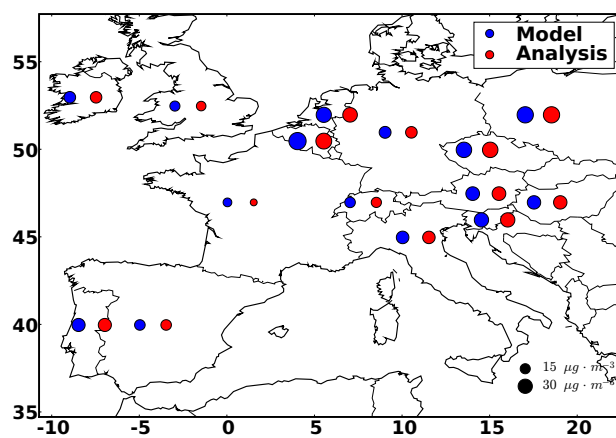
Fig. 2. Time evolution (over January 2001) of the quantity χ_h^2/N_h defined with Eq. 6 in Sect. 2.2. In red, the daily average of χ_h^2 divided by the daily average of N_h . The mean over January of χ_h^2/N_h is 1.03.

Table 3 shows statistics for the comparisons with the EMEP data. Most statistics on PM₁₀ 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 globally increases the concentrations, and thus lessens the underestimation over all stations. EMEP stations are background stations, whereas the assimilated observations are measured at both background and urban or suburban stations. Some DA updates might not be consistent with background concentrations levels. Besides, there may not be enough stations to draw reliable conclusions.

The statistics for sulfate, chlorine and sodium are slightly better with DA. On the other hand, the statistics for nitrate are worse. For ammonium, the statistics are stable. Actually, the model underestimates the PM₁₀ 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 PM₁₀ observations and these stations maybe not taken into account for the PM₁₀ statistics. It is therefore difficult to draw conclusions about a general behavior. Moreover, the overestimation for nitrate concentrations is specific to winter conditions (Sartelet et al., 2007).



(a) Correlations



(b) RMSEs

Fig. 3. 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.

Despite this remark, these results highlight the need for more chemical measurements in the DA method presented here. The partitioning in different species could then be corrected by assimilation, while it is constant here. At the moment, without chemical data, a more deeper knowledge of the uncertainties on modeled concentrations for each aerosol component would certainly improve the system. Actually, the repartition after DA could be changed according to tendencies in uncertainties.

5 Operational forecast

In operational conditions, at time t_0 , only the data for the past is available. It is possible to assimilate the past data over a

Table 3. 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 to 31 January 2001.

| Species | Simulation | # stations | Obs. Mean $\mu\text{g m}^{-3}$ | Sim. Mean $\mu\text{g m}^{-3}$ | RMSE $\mu\text{g m}^{-3}$ | Correlation % | MFE % | MFB % |
|-------------------|-------------------|------------|--------------------------------------|--------------------------------------|------------------------------|------------------|----------|----------|
| PM ₁₀ | Model | 16 | 20.2 | 17.4 | 17.8 | 45.3 | 65.7 | 19.5 |
| | Analysis | | | 18.9 | 18.3 | 44.2 | 66.2 | 24.2 |
| | One-hour forecast | | | 18.7 | 18.1 | 50.5 | 66.5 | 23.9 |
| PM _{2.5} | Model | 7 | 20.4 | 17.3 | 17.2 | 57.0 | 68.8 | 19.9 |
| | Analysis | | | 20.4 | 18.7 | 55.4 | 68.8 | 28.4 |
| | One-hour forecast | | | 19.9 | 18.2 | 55.5 | 68.3 | 28.0 |
| Sulfate | Model | 55 | 2.3 | 1.2 | 1.9 | 52.5 | 63.5 | -30.0 |
| | Analysis | | | 1.4 | 1.9 | 50.7 | 63.7 | -22.3 |
| | One-hour forecast | | | 1.3 | 1.9 | 50.7 | 63.3 | -23.1 |
| Nitrate | Model | 14 | 3.2 | 6.4 | 4.2 | 38.3 | 92.1 | 84.3 |
| | Analysis | | | 6.6 | 4.5 | 38.1 | 93.1 | 85.8 |
| | One-hour forecast | | | 6.5 | 4.5 | 38.0 | 93.0 | 85.6 |
| Ammon. | Model | 9 | 2.2 | 2.3 | 1.6 | 48.6 | 57.4 | 22.0 |
| | Analysis | | | 2.5 | 1.7 | 47.8 | 58.5 | 24.3 |
| | One-hour forecast | | | 2.4 | 1.7 | 47.7 | 58.3 | 24.1 |
| Sodium | Model | 3 | 1.2 | 3.2 | 3.0 | 60.6 | 88.3 | 81.3 |
| | Analysis | | | 3.2 | 2.9 | 62.1 | 88.1 | 80.8 |
| | One-hour forecast | | | 3.2 | 2.9 | 61.1 | 88.1 | 80.8 |

Table 4. Description of the European simulations carried out in operational-forecast conditions and of their outputs that are compared to observations. t_0 is a given day between 3 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+1 d) Two-day forecast (t_0+1 d to t_0+2 d) |

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”); one-day forecast and two-day forecast are then available every day.

Several five-day DA experiments were carried out: the BDQA data are assimilated every hour during the first three days, after which the model runs freely and produces forecasts for the next two days. The first experiment assimilates data from 1 to 3 January 2001 and forecasts the days 4 and 5 January 2001; the second experiment assimilates data from 2 to 4 January 2001 and forecasts the days 5 and 6 January

2001; and so on. Consequently, one-day forecasts are available from 4 to 30 January 2001, and two-day forecasts are available from 5 to 31 January 2001. Table 4 describes the simulations carried out.

Table 5 summarizes the performance of the model without assimilation, and of the one-day and two-day forecasts, when compared to BDQA observations. It is noteworthy that, as expected, the one-day forecast clearly shows better statistics for PM₁₀ and PM_{2.5} than the simulation without assimilation. The decrease of the RMSE is $1.5 \mu\text{g m}^{-3}$ for PM₁₀ and $1.4 \mu\text{g m}^{-3}$ for PM_{2.5}, that is, about 10%. The increase of the correlation is more than 10% for PM₁₀ 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 OI method only changes the initial conditions, the model tends after all to its reference trajectory (without assimilation). Here, the two-day forecasts show a less obvious improvement. Two-day forecasts show slightly better statistics than the free-running model, but the decrease of the RMSE is only $0.2 \mu\text{g m}^{-3}$ and the increase of the correlation is 2%. The time period after which the corrections due to DA become essentially ineffective are discussed in the next section.

Figures 4, 5, and 6 show the daily evolution (averaged over the period 4 to 30 January 2001) of the RMSE, the correlation and the mean concentrations respectively for the model

Table 5. Statistics of the simulations (model, one-day forecast and two-day forecast) on the BDQA network for PM₁₀ and PM_{2.5}. Period: 4 to 30 January 2001.

| Species | Simulation | # stations | Obs. Mean $\mu\text{g m}^{-3}$ | Sim. Mean $\mu\text{g m}^{-3}$ | RMSE $\mu\text{g m}^{-3}$ | Correl. % | MFE % | MFB % |
|-------------------|------------------|------------|--------------------------------------|--------------------------------------|------------------------------|--------------|----------|----------|
| PM ₁₀ | Model | 156 | 21.8 | 17.4 | 16.6 | 35.7 | 55.3 | -9.2 |
| | One-day forecast | | | 18.7 | 15.1 | 46.8 | 49.9 | -3.5 |
| | Two-day forecast | | | 17.7 | 16.4 | 37.9 | 54.6 | -8.2 |
| PM _{2.5} | Model | 8 | 19.8 | 15.8 | 15.0 | 30.2 | 57.9 | -10.3 |
| | One-day forecast | | | 16.9 | 13.6 | 43.2 | 53.3 | -3.9 |
| | Two-day forecast | | | 15.9 | 14.8 | 32.3 | 57.4 | -9.4 |

Table 6. Configuration of DA for the evaluation of the impact of the assimilation parameters on the forecasts.

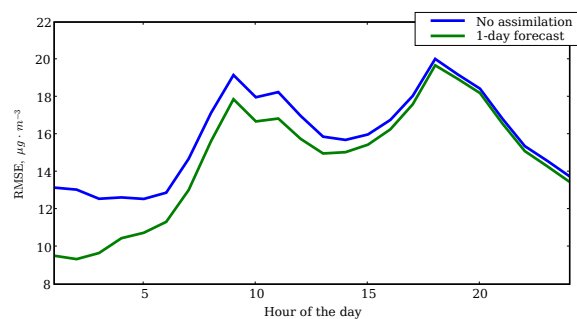
| Parameter | Reference | Alternative |
|-----------------|-------------|--------------------|
| L_h Balgovind | 2 cells | 0.5 cell |
| L_v Balgovind | 30 m | 200 m |
| L_v Balgovind | 30 m | 300 m |
| L_v Balgovind | 30 m | 600 m |
| Redistribution | all species | primary inorganics |
| $\alpha=v/r$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

without assimilation and for the one-day forecast. These figures underline the tendency of the assimilation procedure to be almost ineffective after 24 h of forecast. Actually, after 12:00 UTC, the differences in RMSE and in mean concentration are lower than $1 \mu\text{g m}^{-3}$, and the difference in correlation is about 2%.

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 Sect. 6 – and on the pollutant. For ozone, the influence is a bit longer – see Wu et al. (2008).

6 Sensitivity to DA parameters

Because the OI method only modifies the initial conditions, and not the model itself, one may want to evaluate the time and space scales for which DA affects the results. Tests with different configurations were carried out over a shorter period to estimate the effective time scales of the DA impact. The data is assimilated over a period and then the model is running without assimilation for the remaining days. The aim of these tests is to find out key parameters.

**Fig. 4.** Hourly evolution (averaged over the 04-01-2001 to 30-01-2001 period) of the RMSE for the PM₁₀ forecast without assimilation (blue line) and for the one-day forecast (green line).

The configuration for the present simulations is the same as in the previous section, over the period from 1 to 6 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 BDQA stations is assimilated from 1 to 5 January 2001. The forecast starts on the 6 January 2001 at 00:00 UTC. Eight simulations are presented here: the reference test and seven alternatives. The different configurations are summarized in Table 6.

In one experiment, the variance for observation errors is equal to the background variance: if $\mathbf{R}=r\mathbf{I}$ where \mathbf{I} is the identity matrix and r a scalar variance, the ratio $\alpha=v/r$ (with v from Eq. 5) is then equal to 1. Note that the results of the OI method are the same for all pairs (r, v) such that the ratio r/v is constant (hence equal to 1) – the actual value of the variance $r=v$ has no impact.

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, L_h is taken equal to one grid cell (about 50 km) in the reference simulation, and it is changed to two grid cells (about 100 km). Vertically, the reference test performs assimilation only in the first model level (centered at 25 m); hence the results are

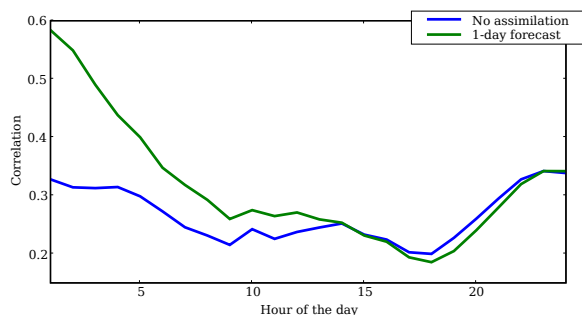


Fig. 5. Hourly evolution (averaged over the 04-01-2001 to 30-01-2001 period) of the correlation for the PM₁₀ forecast without assimilation (blue line) and for the one-day forecast (green line).

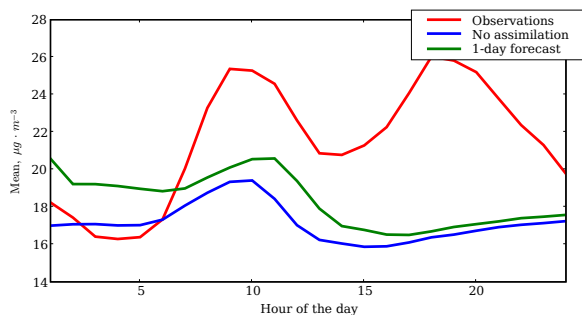


Fig. 6. Hourly evolution (averaged over the 04-01-2001 to 30-01-2001 period) of the mean concentration for the PM₁₀ observations (red line), for the forecast without assimilation (blue line) and for the one-day forecast (green line).

independent of L_v . Three other tests are carried out: respectively with a vertical L_v parameter equal to 200 m with two controlled model levels, 300 m with three controlled levels, and 600 m with three controlled levels.

The redistribution of analyzed PM₁₀ on chemical species is also investigated. The following cases are considered:

- redistribution on all species (default): it is assumed that the model uncertainties are equivalent 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. 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 PM₁₀ measurement instruments, the samplings are heated to measure the mass of dry particulate matter. For example, the sample temperature of a

Table 7. Table of the scores of the different tests at BDQA stations for the first day forecast (6 January 2001).

| Simulation | RMSE | Correlation Mean | Concentration |
|-------------------------|------|------------------|---------------|
| Observations | | | 14.1 |
| No assimilation | 13.9 | 30% | 14.8 |
| All species (Reference) | 11.3 | 50% | 14.4 |
| Primary species | 12.5 | 52% | 15.1 |
| Inorganic species | 11.3 | 50% | 14.4 |
| $L_h=2$ | 11.5 | 44% | 14.3 |
| $\alpha=1$ | 11.5 | 52% | 14.3 |
| $L_v=200\text{m}$ | 11.1 | 50% | 14.0 |
| $L_v=300\text{m}$ | 11.1 | 53% | 13.9 |
| $L_v=600\text{m}$ | 11.1 | 48% | 13.7 |

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; Josef 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.

Figures 7, 8 and 9 show the time evolution of the RMSE, the correlation and the simulated mean respectively, averaged over BDQA stations for the different tests. Table 7 shows the associated statistics, over the forecast day (6 January 2001). All simulations with assimilation improve the RMSE and the correlation. However, 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 figure shows that there still could be a difference after 24 h. The *inorganics* simulation is almost the same as the reference test where all species are assimilated (RMSEs and correlations are equal), whereas the *primary* simulation gives very different results for the simulated mean. In the *inorganics* simulation, the transfer of the PM₁₀ changes to the inorganic species takes place essentially in the fine mode (less than 2.5 µm diameter), where the major part of inorganics resides. The PM₁₀ mass is also located in this mode, so the size distribution of corrected PM₁₀ will be equivalent in both the *reference* and the *inorganics* simulations. For this reason, the scavenging which depends on particle size and which could therefore explain the differences in the PM₁₀ budgets between the simulations, is not affected. The RMSE for the *primary* simulation is the highest of the tests with DA. This likely 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 the

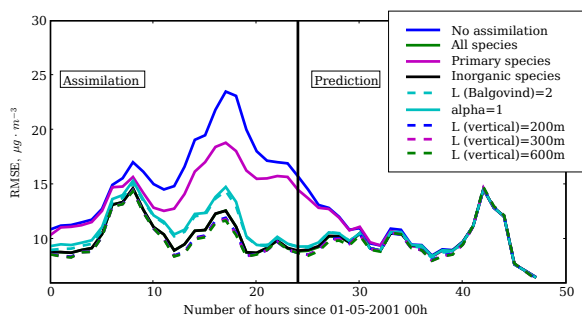


Fig. 7. Time evolution of the RMSE for the PM₁₀ forecasts. The vertical line delimits the assimilation period from the prediction period.

emissions description. The simulation where $\alpha=1$ is close to the simulation without DA, particularly in the evolution of the 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 if the number of controlled levels and the parameter L_v are increased. This shows that DA can partially influence the PM₁₀ vertical profile; but increasing L_v above 200 m has a limited impact.

Figure 10 shows the maps of the absolute differences between the PM₁₀ fields of the simulations with DA and the same field without DA, averaged over the first forecast day. As expected, the test with $L_h=2$ shows differences in regions that were not affected with the reference assimilation settings. The northern part of Italy is more affected. In the reference assimilation experiment, the most impacted region is in the north of Spain. There, the highest difference is located in one grid cell, but it does not affect the statistics since there is no station in that cell. Besides this point, southern (Marseille region) and south-eastern France are the most impacted regions, showing that the concentrations in these regions are particularly badly reproduced by the model. This may be due to the fact that a few 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 represent with a continental-scale simulation.

It is noteworthy that the regions impacted by the *inorganics* and the *primary* tests are rather 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 PM₁₀ DA with the OI method may be useful for one-day forecasts. In our tests, a mean decrease of

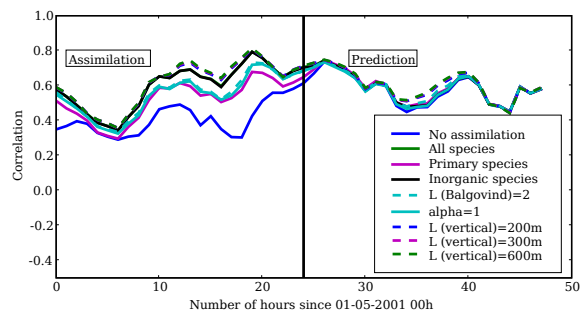


Fig. 8. Time evolution of the correlation for the PM₁₀ forecasts. The vertical line delimits the assimilation period from the prediction period.

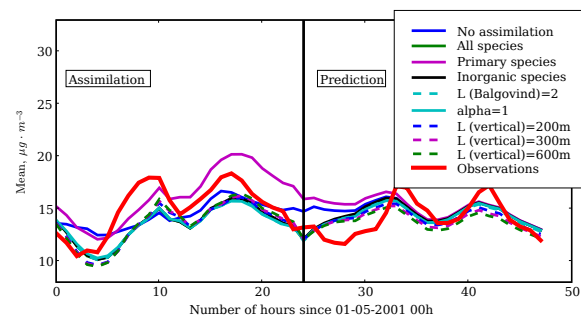


Fig. 9. Time evolution of the mean concentration for the PM₁₀ forecasts. The vertical line delimits the assimilation period from the prediction period.

1.5 $\mu\text{g m}^{-3}$ 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.

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. On the contrary, the data assimilated in this study includes observations from urban or suburban stations, which have much higher concentrations. The forecasts thus tend to increase the concentrations accordingly, spoiling the statistics for rural stations as those of the EMEP database.

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 worse statistics in regions where topography would require specific decorrelation lengths L_h .

The sensitivity tests show that, in this specific study, the uncertainties on the condensation process might be greater than the uncertainties on the emissions. These uncertainties could originate from uncertainties in the concentrations of condensable gas species or in the modeling of the condensation process itself.

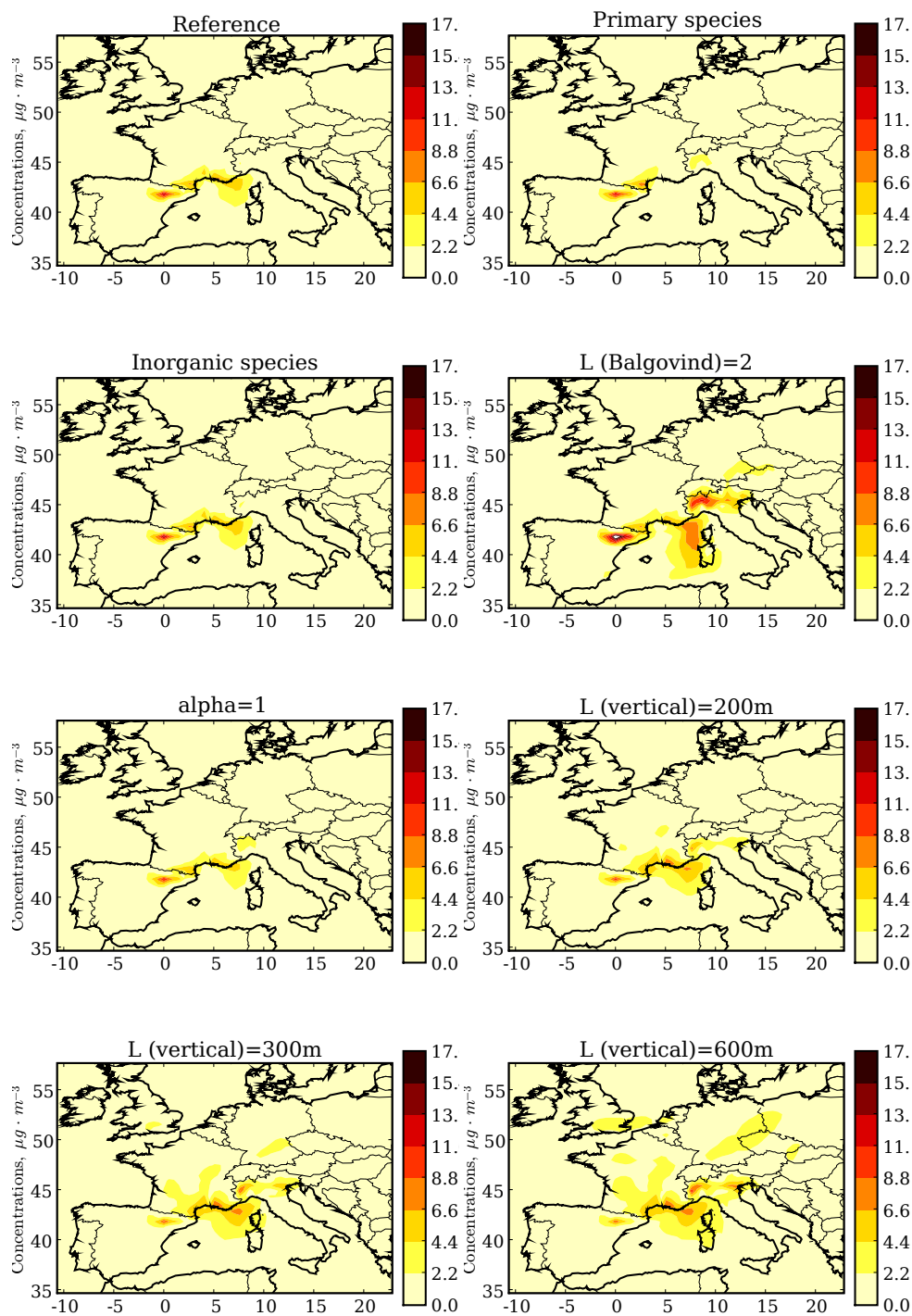


Fig. 10. Maps of the absolute difference of the PM₁₀ fields (in $\mu\text{g} \cdot \text{m}^{-3}$). Comparison between the simulation without assimilation and the simulation with assimilation for the eight tests over the day 6 January 2001.

The OI method applied in this paper could be used in operational mode: the model is already running for real-time forecasts over Europe, in the context of tests on the Prév'air platform (<http://www.prevoir.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).

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 (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 (HNO₃) or ammonia (NH₃);
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.

Acknowledgements. This research is carried out in the ENPC-EDF R&D joint laboratory CERECA, 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.

Edited by: W. Lahoz

References

- Allen, G., Sioutas, C., Koutrakis, P., Reiss, R., Lurmann, F., and Roberts, P.: Evaluation of the TEOM method for measurement of ambient particulate mass in urban areas, *J. Air Waste Manage.*, 47, 682–689, 1997.
- Austin, J.: Toward the four dimensional assimilation of stratospheric chemical constituents, *J. Geophys. Res.*, 97, 2569–2588, 1992.
- Balgovind, R., Dalcher, A., Ghil, M., and Kalnay, E.: A Stochastic-Dynamic Model for the Spatial Structure of Forecast Error Statistics, *Mon. Weather Rev.*, 111, 701–722, 1983.

- Benedetti, A. and Fisher, M.: Background error statistics for aerosols, *Q. J. Roy. Meteor. Soc.*, 133, 391–405, 2007.
- Bouttier, F. and Courtier, P.: Data assimilation concepts and methods, in: *Meteorological Training Course Lecture Series*, ECMWF, 2001.
- 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.
- Chang, M. E., Hartley, D. E., Cardelino, C., Haas-Laursen, D., and Chang, W.-L.: On using inverse methods for resolving emissions with large spatial inhomogeneities, *J. Geophys. Res.*, 102, 16023–16036, 1997.
- Charron, A., Harrison, R., Moorcroft, S., and Booker, J.: Quantitative interpretation of divergence between PM₁₀ and PM_{2.5} mass measurement by TEOM and gravimetric (Partisol) instruments, *Atmos. Environ.*, 38, 415–423, 2003.
- 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, 24671–24688, 2000.
- 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.
- Daley, R.: *Atmospheric Data Analysis*, Cambridge University Press, 1993.
- 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/>.
- Elbern, H., Schmidt, H., and Ebel, A.: Variational data assimilation for tropospheric chemistry modeling, *J. Geophys. Res.*, 102, 15967–15985, 1997.
- 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/>.
- Fahey, K. M. and Pandis, S. N.: Optimizing model performance: variable size resolution in cloud chemistry modeling, *Atmos. Environ.*, 35, 4471–4478, 2001.
- 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.
- 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, D02311, doi:10.1029/2005JD006954, 2007.
- Green, D., Fuller, G., and Barratt, B.: Evaluation of TEOM correction factors for assessing the EU Stage I limit values for PM₁₀, *Atmos. Environ.*, 35, 2589–2593, 2001.
- 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.
- Hodzic, A., Madronich, S., Bohn, B., Massie, S., Menut, L., and Wiedinmyer, C.: Wildfire particulate matter in Europe during summer 2003: Meso-scale modeling of smoke emissions,

- transport and radiative effects, *Atmos. Chem. Phys.*, 7, 4043–4064, 2007, <http://www.atmos-chem-phys.net/7/4043/2007/>.
- 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.
- Hooyberghs, J., Mensink, C., Dumont, G., Fierens, F., and Brasseur, O.: A neural network forecast for daily average PM₁₀ concentrations in Belgium, *Atmos. Environ.*, 39, 3279–3289, 2005.
- Huneus, N.: Assimilation variationnelle d'observations satellitaires dans un modèle atmosphérique d'aérosols, Ph.D. thesis, Université des Sciences et Technologies de Lille, 2007.
- 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.
- Josef, C., Gunnar, D., Wolfgang, K., Thomas, T., and Joachim, H.: PM_{2.5} measurements in ambient aerosol: comparison between Harvard Impactor (HI) and the Tapered Element Oscillating Microbalance (TEOM) system, *The Science of The Total Environment*, 278, 191–197, 2001.
- 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.
- 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:10.1029/2005JD006149, 2006.
- 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/>.
- Ménard, R., Cohn, S. E., Chang, L.-P., and Lyster, P. M.: Assimilation of stratospheric chemical tracer observations using a Kalman filter. Part I: Formulation, *Mon. Weather Rev.*, 128, 2654–2671, 2000.
- Nenes, A., Pilinis, C., and Pandis, S.: ISORROPIA: A new thermodynamic model for inorganic multicomponent atmospheric aerosols, *Aquatic Geochem.*, 4, 123–152, 1998.
- Planet, W.: Determination of total ozone amount from TIROS radiance measurements, *J. Appl. Meteor.*, 23, 308–316, 1984.
- Quélo, D., Mallet, V., and Sportisse, B.: Inverse modeling of NO_x emissions at regional scale over Northern France. Preliminary investigation of the second-order sensitivity, *J. Geophys. Res.*, 110, D24310, doi:10.1029/2005JD006151, 2005.
- Riishojgaard, L.: On four-dimensional variational assimilation of ozone data weather prediction models, *Q. J. Roy. Meteor. Soc.*, 122, 1545–1571, 1996.
- Roustan, Y., Sartelet, K. N., Tombette, M., Debry, E., Fahey, K. M., and Sportisse, B.: Simulation of aerosols and gas-phase species over Europe with the Polyphemus system. Part II: Sensitivity analysis., Technical report no 2008-7, CERE, Atmos. Environ., 2008.
- Salter, L. and Parsons, B.: Field trials of the TEOM and the Partisol for PM₁₀ monitoring in China clay area, Cornwall, UK, *Atmos. Environ.*, 33, 2111–2114, 1999.
- 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.
- 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, 28275–28293, 2001.
- Segers, A.: Data assimilation in atmospheric chemistry models using Kalman filtering, Ph.D. thesis, Delft University, The Netherlands, 2002.
- Smith, S., Stribley, T., Barratt, B., and Perryman, C.: Determination of Partisol, TEOM, ACCU and cascade impactor instruments in the London Borough of Greenwich, *Clean Air*, 27, 70–73, 1997.
- Soutar, A., Watt, M., Cherrie, J., and Seaton, A.: Comparison between a personal PM₁₀ sampling head and the tapered element oscillating microbalance (TEOM) system, *Atmos. Environ.*, 33, 4373–4377, 1999.
- Stockwell, W., Kirchner, F., and Kuhn, M.: A new mechanism for regional chemistry modeling, *J. Geophys. Res.*, 102, 25847–25879, 1997.
- Strader, R., Gurciullo, C., Pandis, S., Kumar, N., and Lurmann, F.: Development of a gas-phase chemistry, secondary organic aerosol and aqueous-phase chemistry modules for PM modelling, Technical report, STI, 1998.
- Van Loon, M., Roemer, M., and Bultjes, P.: Model inter-comparison in the framework of the review of the Unified EMEP model, Tech. rep., TNO Environment Energy and Process Innovation, TNO report – R2004/282, 2004.
- Wu, L., Mallet, V., Bocquet, M., and Sportisse, B.: A comparison study of data assimilation algorithms for ozone forecasts, *J. Geophys. Res.*, 113, D20310, doi:10.1029/2008JD009991, 2008.